Categorical Probability and Stochastic Dominance in Metric Spaces

Submitted PhD Thesis

Paolo Perrone*

Max Planck Institute for Mathematics in the Sciences
Leipzig, Germany

5th July 2018

*Correspondence: perrone [at] mis.mpg.de

Abstract

In this work we introduce some category-theoretical concepts and techniques to study probability distributions on metric spaces and ordered metric spaces. The leading themes in this work are *Kantorovich duality* [Vil09, Chapter 5], *Choquet theory* [Win85, Chapter 1], and the categorical theory of *monads and their algebras* [Mac00, Chapter VI].

Categorical Probability. In Chapter 1 we give an overview of the concept of a *probability monad*, first defined by Giry [Gir82].

Probability monads can be interpreted as a categorical tool to talk about *random elements* of a space. Given a space X, we can assign to it a space PX, which extends X by allowing extra elements, *random elements*. We can consider these random elements as *formal convex combinations, or mixtures, of elements of X*. For example, the law of a fair coin flip is 1/2 "heads" + 1/2 "tails". Of course, in general, such mixtures are given by integrals rather than just sums. Probability monads allow to iterate the construction, and talk about the space PPX of *random elements with random law.* Given such an element of PPX, one can always *integrate* it to obtain a simple probability measure in PX. In other words, integration always defines a map $E : PPX \to PX$.

Spaces where the convex combinations can be *actually evaluated*, so that they are well-defined operations, are called *algebras of the probability monad.* These are the spaces, for example $\mathbb{R}$, where one can take *expectation values* of random variables. The set {"heads", "tails"} is *not* an algebra of the monad: there is no element, or *deterministic* state which correspond to "halfway between heads and tails".

As it is known, to every monad corresponds an adjunction. For probability monads, this adjunction can be interpreted in terms of Choquet theory [Win85, Chapter 1]: given any object X and any algebra A, there is a natural bijection between maps $X \to A$ and *affine* maps $PX \to A$.

The Kantorovich Monad. In Chapter 2 we define a probability monad on the category of complete metric spaces and 1-Lipschitz maps called the *Kantorovich monad*, extending a previous construction due to van Breugel [vB05]. This monad assigns to each complete metric space X its *Wasserstein space* PX, which is itself a complete metric space [Vil09].

It is well-known [Vil09, Chapter 6] that finitely supported probability measures with rational coefficients, or *empirical distributions of finite sequences*, are dense in the Wasserstein space. This density property can be translated into categorical language as a *universal property* of the Wasserstein space PX, namely, as a *colimit* of a diagram

involving certain powers of X. The monad structure of P, and in particular the integration map E, is uniquely determined by this universal property, without the need to define it in terms of integrals or measure theory. In some sense, the universal property makes the integration map *inevitable*, it arises directly from the characterization of P in terms of finite powers.

We prove that the algebras of the Kantorovich monad are exactly the *closed convex subsets of Banach spaces.* In the spirit of categorical probability, these can be interpreted as the *complete metric spaces with a well-defined notion of convex combinations.* The "Choquet adjunction" that we obtain is then the following: *given a complete metric space X and a Banach space A, there is a natural bijection between short maps $X \to A$ and short affine maps $X \to A$.*

In the end of the chapter we show that both the integration map $E : PPX \to PX$ and the marginal map $\Delta : P(X \times Y) \to PX \times PY$ are *proper maps.* This means in particular that the set of probability measures *over* a Wasserstein space PX which integrate to a given measure $p \in PX$ is always compact, and analogously, that the set of couplings of any two probability measures p and q is compact as well. As a consequence, on every complete metric space, every Kantorovich duality problem admits an optimal solution.

Stochastic Orders. In Chapter 3 we extend the Kantorovich monad of Chapter 2 to metric spaces equipped with a partial order. The order is inherited by the Wasserstein space, and is called the *stochastic order.* Differently from most approaches in the literature, we define a compatibility condition of the order with the *metric itself*, rather then with the topology it induces. We call the spaces with this property *L-ordered spaces.*

On L-ordered spaces, the stochastic order induced on the Wasserstein spaces satisfies itself a form of Kantorovich duality: given two measures p, q, we can say that $p \leq q$ if and only if they admit a coupling r such that for all the points (x, y) in the support of r we have $x \leq y$. An interpretation is that *there exists a transport plan that moves the mass only upwards in the order, not downwards.* Alternatively, we can say that $p \leq q$ if and only if for all *monotone* 1-Lipschitz functions $\int_X f\,dp \leq \int_X f\,dq$.

This Kantorovich duality property implies that the stochastic order on L-ordered spaces is always a partial order, i.e. it is antisymmetric.

The Kantorovich monad of Chapter 2 can be extended naturally to the category of L-ordered metric spaces. We prove that its algebras are the closed convex subsets of *ordered* Banach spaces, i.e. Banach spaces equipped with a partial order induced by a closed cone. The integration map on ordered Banach spaces is always monotone, and we prove that it is even *strictly* monotone: if $p \leq q$ for the stochastic order and p and

q have the same expectation value, then $p = q$. This generalizes a result which is long known for the real line.

We can consider the category of L-ordered metric spaces as *locally posetal 2-categories*, with the 2-cells given by the pointwise order of the functions. This gives an order-theoretical version of the "Choquet adjunction": given an L-ordered complete metric space X and an ordered Banach space A, there is a natural *isomorphism of partial orders* between short *monotone* maps $X \to A$ and short affine *monotone* maps $X \to A$.

Moreover, in this 2-categorical setting, we can describe *concave and convex maps* categorically, exactly as the *lax and oplax morphisms of algebras.*

Convex Orders. In Chapter 4 we study a different order between probability measures, which can be interpreted as *pointing in the direction of increasing randomness.*

We have seen that probability monads can be interpreted in terms of formal convex combinations, and that their algebras can be interpreted as spaces where such convex combinations can be evaluated. Here we develop a *new* categorical formalism to describe *operations evaluated partially.* For example, "5+4" is a partial evaluation of the sum "2+3+4". We prove that partial evaluations for the Kantorovich monad, or *partial expectations*, define a closed partial order on the Wasserstein space PA over every algebra A, and that the resulting ordered space is itself an algebra.

We prove that, for the Kantorovich monad, these partial expectations correspond to conditional expectations *in distribution.* This implies that the partial evaluation order is equivalent to the order known in the literature as the *convex* or *Choquet order* [Win85].

A useful consequence of this equivalence and of the fact that the integration map E is proper is that *bounded monotone nets in the partial evaluation order always converge.* This fact can be interpreted as a result of *convergence in distribution for martingales and inverse martingales over general Banach spaces.*

Given an algebra A, we can compare the partial evaluation order and the stochastic order on PA. We show that the two orders are *transverse*, in the sense that every two probability distributions comparable for both orders are necessarily equal. We can also combine the two orders to form a new order, which we call the *lax partial evaluation order.* The space PA with this order also forms an algebra.

Finally, we study the relation between these partial evaluation orders and convex functions. As is well-known [Win85], the Choquet order is dual to convex functions. We know from Chapter 3 that convex functions are the oplax morphisms of algebras. This is not a coincidence: as we show, the partial evaluation order and convex functions are related by the "ordered Choquet adjunction" of Chapter 3. This permits to characterize the partial evaluation order in terms of a universal property, as an *oplax codescent object* [Lac02]. From this universal property we can derive a general duality result valid on *all ordered Banach spaces*, which says that over every ordered Banach space A, the lax partial evaluation order is dual to *monotone convex functions.* In other words, for every two probability measures p and q over A, $\int f dp \leq \int f dq$ for all convex monotone functions f *if and only if* $p \preceq_l q$ for the lax partial evaluation order. As far as we know, this result in its full generality is new.

Sources. Part of this work is contained in the papers [FP17] and [FP18]. The rest will appear in two papers which are currently in preparation.

Acknowledgements

There are a number of people that I would like to thank, who helped me with this thesis, with my PhD, and with my mathematical work in general:

- I would like to thank Prof. Jürgen Jost for the trust, the support, and the guidance during these years, and for all the deep and insightful discussions.
- I would especially like to thank Slava Matveev and Sharwin Rezagholi for the very fruitful and interesting discussions, from which came some core questions that this work tries to answer.
- I would also like to thank for all the interesting discussions Renan Assimos, Kostas Zemas, Ramón Urquijo Novella, Caio Teodoro de Magalhães Alves, Gerardo Sosa, Jim Portegies, Alexei Kudryashov, Sarah Berkemer, Pradeep Banerjee, and all the mathematicians at Max Planck Institute with whom I have shared my thoughts and ideas, and who shared their ideas with me, during the course of my PhD.
- I would like to thank Franco Rota, Alessandro Meani, Rory Lucyshyn-Wright, Dorette Pronk, Tarmo Uustalu, Emily Riehl, Michael Lambert, Pierre Cagne, Maxime Lucas, and all the people I have met in the category theory community, for all the help that they gave me on topics which were completely new to me.
- I want to especially thank the community of the nLab for creating such an amazing resource, on which I have learned way more than I ever have on any textbook.
- I would like to thank Prof. Nihat Ay for guiding me and supporting me during the first phase of my PhD, and for all the work that we have done together.
- I would like to thank Antje Vandenberg for all the precious support, the time and the patience which she gave to me, and to all the people in our department.
- Last, but by far not least, I want to immensely thank Tobias Fritz for helping me greatly in this research, for introducing me to category theory, and most of all, for helping me find again my strong passion for mathematics and conceptual thinking.

Paolo Perrone.

Contents

Introduction

This work is about some applications of category theory to probability theory. In the past, category theory has not been applied to probability as much as to other fields, such as algebraic geometry and topology. However, there are many areas of probability and analysis in which category theory can be applied or at least in which it can give systematic understanding of some of the structures involved.

The categorical and the analytic way of thinking are quite different: analysis concerns itself with approximations, quantitative results, and estimates. Category theory, on the other hand, involves discrete and qualitative statements: it studies the features of an object which can be explained in terms of the interplay that it has with other objects (universal properties). Since these two ways of reasoning are so different, the same problem can be hard in terms of category theory but easy in terms of analysis, or vice versa. Category theory, therefore, can give *alternative techniques* which complement the traditional analytic techniques. Combining both ways of reasoning can be very powerful: there are results which could be much harder to prove using only one of the two approaches. Somewhat dually, analysis can also be useful to category theory: many concepts which naturally arise in analysis and probability can lead to new categorical concepts.

In particular, this work contains the following new results which are also of interest outside of category theory:

- Theorem 2.6.7 shows that the integration map over every Wasserstein space is proper whenever the underlying space is complete. In particular, given any probability measure p of finite first moment, the space of measures which integrate to p is always compact.

- Theorem 2.6.13 shows that the *assignment of marginals* is a proper map as well. In particular, given probability measures p and q of finite first moment, the space of their couplings is compact. This seems to be known under some hypotheses on Polish spaces [Vil09], we give a proof that works for all complete metric spaces.

- Proposition 3.2.5 gives a density result for the stochastic order. It says that over every ordered metric space, the stochastic order is the topological closure of the order induced by finitely supported measures.

- L-ordered spaces (Definition 3.3.1), which we introduce in this work, are a class of spaces where the metric and the order are compatible in a strong way. Theorem 3.3.3 says that on such spaces the stochastic order satisfies a Kantorovich duality property. Theorem 3.3.9 shows that because of this, the stochastic order on any L-ordered space is always a partial order (i.e. antisymmetric).

- Proposition 3.5.11 says that on every ordered Banach space, the integration map is strictly monotone. In other words, two probability measures p and q over an ordered Banach space such that $p \leq q$ in the stochastic order have equal expectation if and only if they are equal. This generalizes a result long known for the stochastic order over the real line.

- The concept of *partial evaluation* (Definition 4.1.1) can be of interest in many settings also outside of probability theory. In particular, it always satisfies a "law of total evaluation" (Proposition 4.1.2) analogous to the law of total expectation of random variables.

- Theorem 4.2.14 extends an earlier result of Winkler [Win85, Theorem 1.3.6] in a metric setting to possibly unbounded spaces, and proves that laws of random variables which have a conditional expectation relation are instances of partial evaluations in the sense of Definition 4.1.1.

- Theorem 4.2.18 gives a result of convergence for all bounded monotone nets in the Choquet order, valid on all Banach spaces. It can be thought of as a theorem of convergence in distribution for martingales and inverse martingales. The theorem extends an earlier result also due to Winkler [Win85, Theorem 2.4.2] in a metric setting to possibly unbounded spaces.

- Corollary 4.3.1 says that over every ordered Banach space, the Choquet order and the stochastic order are *transverse*: any two probability distributions are comparable for both orders if and only if they are equal.

- Finally, Theorem 4.4.9 and Corollary 4.4.10 say that over every ordered Banach space, the composition of the stochastic order and the Choquet order is dual to *monotone convex functions*. In other words, for every two probability measures p, q over an ordered Banach space, $\int f\,dp \leq \int f\,dq$ for all convex monotone functions f *if and only if* there exists p' such that $p \leq p'$ for the stochastic order and $p' \preceq_c q$ for the Choquet order. This result generalizes previous results known for unordered Banach spaces [Win85, Theorem 1.3.6] and for the real line [RS70].

Just as well, the study of Wasserstein spaces in this work leads to the following new concepts and results of categorical nature:

- Theorem 2.2.18 says that the density of finitely supported measures in Wasserstein spaces can be given a categorical meaning in terms of a colimit characterization, which can be extended to the Kantorovich functor itself.

- The whole of Section 2.3 shows that the universal property of the Kantorovich functor determines its monad structure uniquely, and the monad can be written as the colimit of a *graded monad*, in a way similar to how finitary monads are obtained from Lawvere theories in the category of sets.

- Theorem A.2.1 shows that under some hypotheses, Kan extensions of lax monoidal functors are themselves lax monoidal. This is an instance of the theory of *algebraic Kan extensions* [Kou15, Web16], for the case of the 2-monad of monoidal categories.

- Theorem 3.5.6 gives a categorical definition of (convex subsets of) ordered Banach spaces, as algebras of the ordered Kantorovich monad defined in this work.

- The *2-coseparator* of Definition 3.5.15 is a possible extension of the concept of coseparator to a locally posetal 2-category. Corollary 3.5.17 shows that the Hahn-Banach theorem can be interpreted in terms of such a concept.

- Theorem 3.5.19, as far as we know, gives the first categorical characterization of *concave and convex functions*, as lax and oplax morphisms of the Kantorovich monad.

- The concept of partial evaluation (Definition 4.1.1), which is introduced to study probability distributions, can be of interest for every monad. It has a natural operational interpretation, and it raises many general questions (for example, whether partial evaluations can always be composed).

- Finally, Theorem 4.4.3 establishes a correspondence between partial evaluations and lax morphisms of algebras. In the case of the Kantorovich monad, this corresponds precisely to the duality theory between convex functions and the Choquet order, however, the correspondence works in all locally posetal 2-categories. This is shown to be an instance of the theory of *lax codescent objects* [Lac02], which are therefore intimately related to the concept of partial evaluations. For the Kantorovich monad, the universal property of lax codescent objects determines the partial evaluation order uniquely.

Outline

The chapters of this work are organized as follows:

- In Chapter 1 we give an overview of the basic concepts of categorical probability;
- In Chapter 2 we define and study the *Kantorovich monad* on the category of complete metric spaces;
- In Chapter 3 we extend the Kantorovich monad to *ordered metric spaces*, in terms of the stochastic order;
- In Chapter 4 we use the concepts developed in the previous chapters to study orders of *increasing randomness*, and their duality theory.

Part of this work is contained in the papers [FP17] and [FP18]. The rest will appear in two papers which are currently in preparation.

1. Categorical probability

"Categorical probability" is a collection of categorical structures and methods which can be applied to probability, measure theory, and mathematical statistics [Law62, Gir82, JP89]. This first chapter is intended as an overview of the basic constructions of categorical probability which are used in the rest of this work.

Throughout this chapter, and this work, the notation PX will denote the *space of probability distributions over a space X*. Probability measures on a space X can be thought of as *laws of random variables*, or *of random elements* of X. A central theme in probability and statistics is that one is not only interested in random variables, but also in *random variables whose law is also random*, or "random random variables", with their law in PPX. This happens, for example, when probability distributions have to be estimated, and so they come themselves with some *likelihood*.

Example 1.0.1. Suppose that you have two coins in your pocket. Suppose that one coin is fair, with "heads" on one face and "tails" on the other; suppose the second coin has "heads" on both sides. Suppose now that you draw a coin randomly, and flip it.

We can sketch the probabilities in the following way:

$$\begin{tikzcd}
 & & & ? \arrow[dll, "1/2"'] \arrow[drr, "1/2"] & & & \\
 & \text{coin 1} \arrow[dl, "1/2"'] \arrow[dr, "1/2"] & & & & \text{coin 2} \arrow[dl, "1"'] \arrow[dr, "0"] & \\
\text{heads} & & \text{tails} & & \text{heads} & & \text{tails}
\end{tikzcd}$$

Let X be the set $\{$"heads", "tails"$\}$. A coin gives a *law* according to which we will obtain "heads" or "tails", so it determines an element of PX. Since the choice of coin is also random (we also have a *law on the coins*), the law on the coins determines an element of PPX.

By averaging, the resulting overall probabilities are

$$\begin{tikzcd}
 & ? \arrow[dl, "3/4"'] \arrow[dr, "1/4"] & \\
\text{heads} & & \text{tails}
\end{tikzcd}$$

In other words, the "average" or "composition" can be thought of as an assignment $E : PPX \to PX$, from laws of "random random variables" to laws of ordinary random variables.

The space of probability measures PX is usually larger, in some sense, than the underlying space X. The space PPX of *probability distributions over the space of probability distributions* is even larger, which in practice can be hard to work with (for example, PPX is infinite-dimensional even if X is finite). There are mainly two ways to address this issue:

- In *parametric statistics*, one restricts to a family of probability distributions in a specific form, parametrized by a smaller set A, usually a region of $\mathbb{R}^n$. Instead of looking at distributions over PX, which are themselves elements of PPX, one can look at a map $A \to PX$, a *Markov kernel*, and let a probability distribution on A determine, via its pushforward, a probability measure on PX.
- Another approach is to work in a setting where the space of probability measures PX inherits many properties from the underlying space X, and so it can be studied in the same way. This is for example the case for *Wasserstein spaces*, which are widely used in optimal transport and related fields [Vil09].

As we will see, the two approaches listed above are often *equivalent* in a very formal sense (an equivalence of categories). More on that in 1.1.3.

In category theory there is a theory to systematically treat recursive constructions like that of X, PX, PPX, et cetera, and to keep track of the interplay of the different levels: the theory of *monads*. In particular, *probability monads* have been specifically introduced to accomplish this task, and they are arguably the most important concept in categorical probability. The first probability monad was introduced by Giry [Gir82], and the first ideas about its structure can be traced back to Lawvere [Law62]. We will give an overview of the concept of a probability monad in Section 1.1.

Probability theory is not only about spaces of probability measures, but also, and mostly, about the *interactions* and *propagations* of different random variables. This can be treated categorically in terms of *monoidal categories and monoidal functors* [FP18]. We will give an overview of how this works in Section 1.2.

There is another very important aspect of probability theory that can be addressed categorically, namely *stochastic processes*. This will not be treated in this work. However, some results of Chapter 4 have direct applications to *martingales*, such as Theorem 4.2.14 and Theorem 4.2.18.

Outline

- In Section 1.1 we introduce the concept of a *probability monad*. In 1.1.1 we give an interpretation in terms of *spaces of generalized elements*, and in 1.1.2 we give an interpretation in terms of *spaces of formal expressions*. In 1.1.3 we show how the adjunction associated to a probability monad can be connected with the *Choquet theory* of convex spaces.

- In Section 1.2 we explain how to talk about joint and marginal probability distributions in terms of monoidal structures. In 1.2.1 we introduce semicartesian monoidal categories and affine monads, and explain why they are a good setting for categorical probability. In 1.2.2 we introduce bimonoidal monads, and explain the role they play in categorical probability. In 1.2.3 we show how the monoidal structure allows to form *convolutions* of random variable in a very general way. Finally, in 1.2.4 we show how to obtain a "category of probability spaces", or "of random elements", from a category equipped with a probability monad.

This chapter is motivational, it is an informal introduction to the basic concepts that are used later on, and defined in full rigor. Most of the content of Section 1.2 is part of the paper [FP18].

1.1. Probability monads

A central concept in most categorical approaches to probability theory is that of a *probability monad*, first introduced by Giry [Gir82]. Probability monads give a systematic way of talking about "probability measures over probability measures", and of the interactions between the different levels. The term "probability monad", also introduced by Giry, is not a technical term; it simply means a monad whose interpretation is that of a "space of probability distributions", more on that below. A detailed list containing most of the probability monads in the literature, together with their main properties, can be found in [Jac17].

Monads are very general concepts, which model many different constructions in mathematics as well as in computer science, and which can be interpreted and motivated in many ways. Below is the general category-theoretical definition. To motivate this definition and its usage in probability, we will then focus on two aspects of the theory of monads: its interpretation in terms of *spaces of generalized elements* in 1.1.1, and its interpretation in terms of *spaces of formal expressions* in 1.1.2. Technically, the interpretations given there are helpful for all monads which have a monic unit. Most monads in the literature have a monic unit, in particular, all probability monads do (at least all those listed in [Jac17], as well as all the ones defined in this work).

Definition 1.1.1. *Let* C *be a category. A* monad on C *consists of:*

- *A functor* $T : \mathsf{C} \to \mathsf{C}$*;*
- *A natural transformation* $\eta : \mathrm{id}_{\mathsf{C}} \Rightarrow T$ *called* unit*;*
- *A natural transformation* $\mu : TT \Rightarrow T$ *called* composition *or* multiplication*;*

such that the following diagrams commute, called "left and right unitality" and "associativity", respectively:

$$\begin{tikzcd} T \arrow[r, Rightarrow, "T\eta"] \arrow[dr, equal] & TT \arrow[d, Rightarrow, "\mu"] \\ & T \end{tikzcd} \qquad \begin{tikzcd} T \arrow[r, Rightarrow, "\eta T"] \arrow[dr, equal] & TT \arrow[d, Rightarrow, "\mu"] \\ & T \end{tikzcd} \qquad \begin{tikzcd} TTT \arrow[r, Rightarrow, "T\mu"] \arrow[d, Rightarrow, "\mu T"] & TT \arrow[d, Rightarrow, "\mu"] \\ TT \arrow[r, Rightarrow, "\mu"] & T \end{tikzcd} \tag{1.1.1}$$

Let's see now how we can interpret this definition in practice.

1.1.1. Monads modeling spaces of generalized elements

A first interpretation of the theory of monads can be summarized in the following way: *a monad is like a consistent way of extending spaces to include generalized elements and generalized functions of a specific kind.*

More in detail, Definition 1.1.1 can be translated in the following way. A functor $T : \mathsf{C} \to \mathsf{C}$ consists of the following data:

(a) To each space X, we assign a new space $T(X)$, or more briefly TX, which we think of as an extension of X, containing the "generalized elements" of X.

(b) Given two spaces X, Y and a function $f : X \to Y$, the function f can be extended to generalized elements of X, and it will output generalized elements of Y. In other words, f defines a function $Tf : TX \to TY$, by "extension". This assignment should preserve identities and composition.

For example, consider the category of sets and functions. Given a set X, its power set $\mathcal{P}X$ can be considered an extension of X. Given sets X and Y and a function $f : X \to Y$, we get automatically a function $\mathcal{P}f : \mathcal{P}X \to \mathcal{P}Y$, the direct image. It maps each $A \in \mathcal{P}X$, which is a subset of X, to the subset of Y given by the image of A under f. Since this assignment is uniquely specified by f, usually this map is denoted again by f, i.e. it is customary to write $f(A) \subseteq Y$, or sometimes f_*. However, technically it is a different map, from *subsets* of X to *subsets* of Y, and the subsets are treated as "generalized elements".

Strictly speaking, elements of X are not subsets. However, each element $x \in X$ defines a subset canonically: the singleton $\{x\}$. In other words, there is an embedding $X \to \mathcal{P}X$. This is part of a natural transformation $\eta : \mathrm{id}_{\mathsf{C}} \Rightarrow T$, called "unit", which in general consists of the following data:

(a) To each X we give a map $\eta_X : X \to TX$, usually an embedding. The interpretation is that TX, the extension, includes the old space X.

(b) For each $f : X \to Y$, the extended function Tf must agree with f on the "old elements", i.e. the elements coming from X via η. In other words, this diagram has to commute:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\eta_X"] & Y \arrow[d, "\eta_Y"] \\ TX \arrow[r, "Tf"] & TY \end{tikzcd} \tag{1.1.2}$$

Whenever this does not lead to ambiguity, we will drop the subscript on the components of the natural transformations. For example, we will write $\eta : X \to TX$ instead of $\eta_X : X \to TX$.

In the case of the power set, the inclusion $X \to \mathcal{P}X$ given by singletons makes diagram (1.1.2) commute: given $x \in X$ and $f : X \to Y$, the direct image of a singleton is exactly the singleton containing the result. In symbols: $f(\{x\}) = \{f(x)\}$.

The "composition" natural transformation, $\mu : TT \Rightarrow T$, is a bit more involved, and it is the most important piece of structure:

(a) To each X we give a map $\mu_X : TTX \to TX$. The interpretation is that given a *generalized generalized* (twice) element, there is a coherent way of obtaining a *generalized* (once) element, "simplifying" the double generalization to just a single generalization.

(b) For each $f : X \to Y$, simplifying before or after applying f gives the same result. In other words, this diagram has to commute:

$$\begin{tikzcd} TTX \arrow[r, "TTf"] \arrow[d, "\mu_X"] & TTY \arrow[d, "\mu_Y"] \\ TX \arrow[r, "Tf"] & TY \end{tikzcd} \tag{1.1.3}$$

The motivation for this map will be given shortly. Let's first continue our example: in the case of the power set, $\mathcal{P}\mathcal{P}X$ contains *subsets of subsets* of X. Given a subset of subsets of X, there is a canonical way of obtaining a subset of X: via the union. For example, if $x, y, z \in X$, a subset of subsets has the form:

$$\mathcal{A} = \{\{x,y\},\{y,z\},\{\}\} \in \mathcal{P}\mathcal{P}X.$$

From the element above, we can take the union of the subsets contained in it, which is:

$$\bigcup_{A \in \mathcal{A}} A = \{x,y,z\} \in \mathcal{P}X.$$

This gives an assignment $\cup : \mathcal{P}\mathcal{P}X \to \mathcal{P}X$. Given a function $f : X \to Y$, commutativity of the diagram (1.1.3) says that the union of the direct images is the direct image of the union. In symbols:

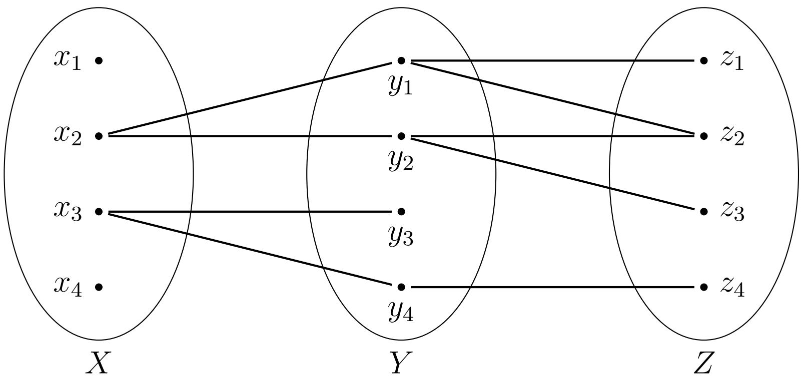
$$f\left(\bigcup_{A \in \mathcal{A}} A\right) = \bigcup_{A \in \mathcal{A}} f(A).$$

Given a monad T, we can not only talk about generalized elements, but also of generalized *functions*. Given spaces X and Y, we can form functions which as output take generalized elements of Y. That is, functions $k : X \to TY$.

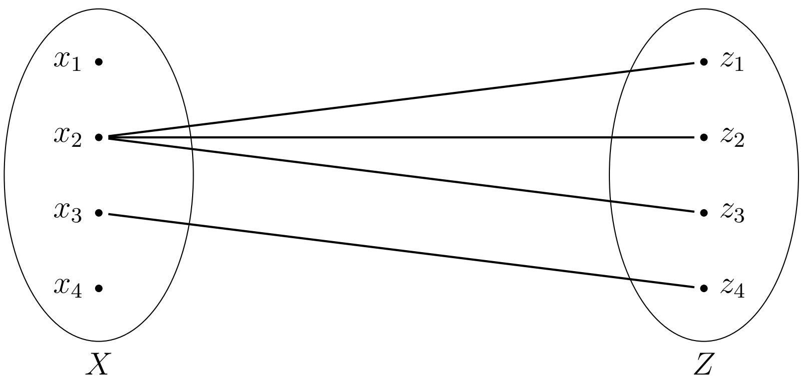
Definition 1.1.2. *Let (T, η, μ) be a monad on a category C. A* Kleisli morphism *of T from X to Y is a morphism $k : X \to TY$ of C.*

In mathematics it often happen that one would like to obtain a function from X to Y from some construction (for example, a limit), but the result is not always well-defined or unique, or not always existing. Allowing more general functions sometimes solves the problem, that is, replacing Y with the extension TY. Generalized functions include ordinary functions in the same way as generalized elements include ordinary elements: via the map η. A function $f : X \to Y$ defines uniquely a map $X \to TY$ given by $\eta \circ f$. Note that this is *different* from extending an existing $f : X \to Y$ to TX: we are not extending an existing function to generalized elements, we are allowing more general functions on X which *take values* in elements of TY which may not come from Y. In particular, a generalized element can be seen as a constant generalized function.

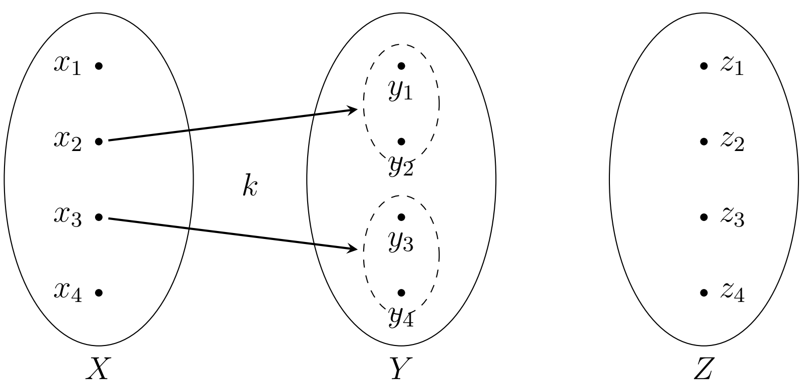
In the case of the power set, Kleisli morphisms, or generalized maps, are precisely *relations*: given sets X and Y, a map $k : X \to \mathcal{P}Y$ assigns to each element of X a *subset* of Y, i.e. it is a multi-valued function. Relations can be composed. Given relations $k : X \to \mathcal{P}Y$ and $h : Y \to \mathcal{P}Z$, as in the following picture:



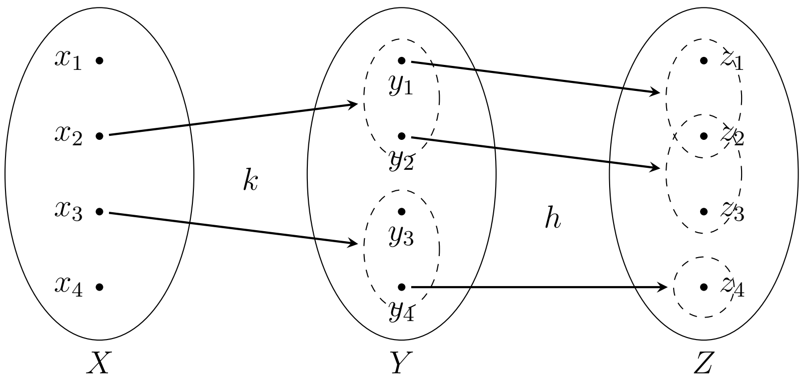
we can compose the two and forget about Y, obtaining a relation $X \to \mathcal{P}Z$:



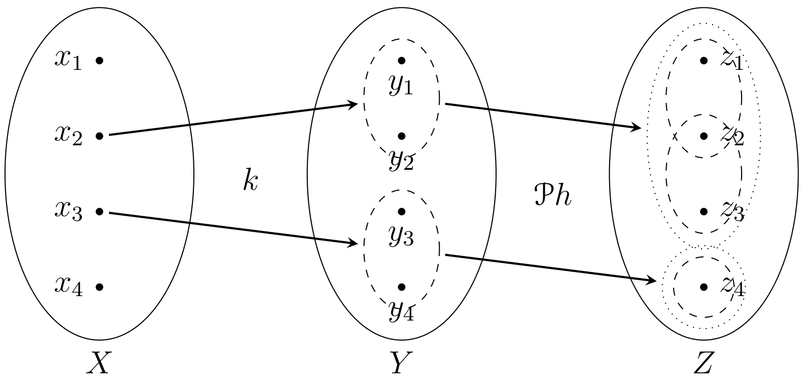
What happened formally is that we have first applied $k : X \to \mathcal{P}Y$, which assigns to each $x \in X$ a subset of Y:



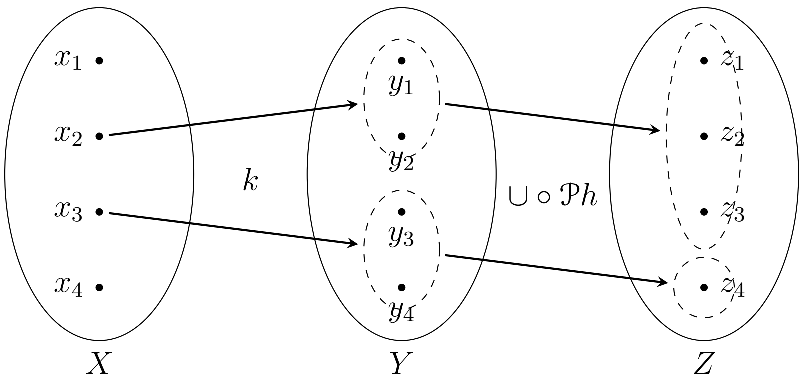
Then we have applied h to *elementwise to each subset in the image of* k:



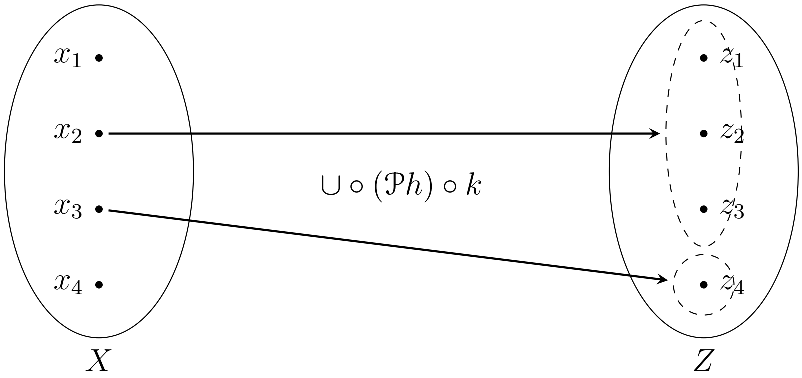
In other words, we have taken the *direct image* of $h : Y \to \mathcal{P}Z$, which we know is the map $\mathcal{P}h : \mathcal{P}Y \to \mathcal{P}\mathcal{P}Z$. Technically, to each subset of Y we have a *subset of subsets* of Z, which contains the images of h:



Now for each subset of Y, we take the *union* of the subsets in its image:



thereby obtaining the composite relation $X \to \mathcal{P}Z$:



More in general, Kleisli morphisms can be composed, and the map μ plays the role that the union played in the power set case.

Definition 1.1.3. *Let* (T, η, μ) *be a monad on a category* C. *Let* $k : X \to TY$ *and* $h : Y \to TZ$. *We define the* Kleisli composition *of* k *and* h *to be the morphism* $(h \circ_{kl} k) : X \to TZ$ *given by:*

$$X \xrightarrow{k} TY \xrightarrow{Th} TTZ \xrightarrow{\mu} TZ. \tag{1.1.4}$$

In other words, the Kleisli composition permits to compose generalized functions from X to Y with generalized functions from Y to Z to give generalized functions from X to Z. The names "unit" and "composition" can be motivated by the facts that the map η is like the identity the for Kleisli composition, and that the map μ allows to define the Kleisli composition itself. More motivation will be given in 1.1.2.

The conditions (1.1.1) are motivated by the following:

- The left unitality condition, for each $k : X \to TY$, gives a commutative diagram

$$\begin{tikzcd} X \arrow[r, "k"] & TY \arrow[r, "T\eta"] \arrow[dr, "\mathrm{id}"'] & TTY \arrow[d, "\mu"] \\ & & TY \end{tikzcd} \tag{1.1.5}$$

  which means that $\eta \circ_{kl} k = k$, i.e. η behaves like a left identity for the Kleisli composition;

- The right unitality condition, together with the naturality of η, for each $k : X \to TY$, gives a commutative diagram

$$\begin{tikzcd} & TX \arrow[r, "Tk"] & TTY \arrow[d, "\mu"] \\ X \arrow[ur, "\eta"] \arrow[r, "k"] & TY \arrow[ur, "\eta"] \arrow[r, "\mathrm{id}"] & TY \end{tikzcd} \tag{1.1.6}$$

  which means that $k \circ_{kl} \eta = k$, i.e. η behaves like a right identity for the Kleisli composition;

- The associativity square, together with naturality of μ, gives for each $\ell : W \to TX$, $k : X \to TY$, and $h : Y \to TZ$ a commutative diagram

$$\begin{tikzcd} W \arrow[r, "\ell"] & TX \arrow[r, "Tk"] & TTY \arrow[r, "TTh"] \arrow[d, "\mu"] & TTTZ \arrow[r, "T\mu"] \arrow[d, "\mu"] & TTZ \arrow[d, "\mu"] \\ & & TY \arrow[r, "Th"] & TTZ \arrow[r, "\mu"] & TZ \end{tikzcd} \tag{1.1.7}$$

which means that $h \circ_{kl} (k \circ_{kl} l) = (h \circ_{kl} k) \circ_{kl} l$, i.e. the Kleisli composition is associative.

In other words, Kleisli morphisms form themselves a category, which we can think of as "having as morphisms the generalized maps".

Definition 1.1.4. *Let (T, η, μ) be a monad on a category C. The* Kleisli category *of T, denoted by C_T, is the category whose:*

- *Objects are the objects of C;*
- *Morphisms are the Kleisli morphisms of T;*
- *Identities are given by the units $\eta : X \to TX$ for each object X;*
- *Composition is given by Kleisli composition.*

We have basically proven that the power set forms a monad on the category of sets and functions, and that its Kleisli category is the category of sets and *relations.* Once again, the interpretation is that the power set "forms spaces of generalized elements in a coherent way", and that the associated "coherent generalization" of functions is relations.

A more detailed account of this interpretation of monads, with a rigorous definition of *extension system* can be found in [MW10], and from a more computer-scientific point of view in [PP02].

The first idea behind a probability monad is now the following: *probability measures and stochastic maps behave like generalized elements and generalized functions.* Let X be a space (measurable, topological, metric, etc.). The space PX of suitably regular probability measures on X can be thought of as containing "random elements of X", or (laws of) random variables. The usual elements of X define elements of PX via Dirac measures, which we can think of as "deterministic". This plays the role of the unit map η. Elements of PPX can be thought of *random variables whose law is also random*, or "random random variables" (see the beginning of this chapter). Given a random variable with random law, we can average it, or *simplify* it, to get a simple random variable, exactly as in Example 1.0.1. This plays the role of the composition map μ.

More in detail, in a category C of suitably regular measurable spaces and functions, a probability monad P on C has the following interpretation:

(a) It assigns to each space X a space PX of probability measures on X, and to each map $f : X \to Y$ the map $(Pf) : PX \to PY$ given by the push-forward of probability measures (which we sometimes denote f_*);

(b) For each space X, it gives an inclusion map $\delta : X \to PX$ which maps each element $x \in X$ to the Dirac measure δ_x;

(c) For each space X, it gives an averaging map $E : PPX \to PX$ which maps each measure $\mu \in PPX$ to the measure $E\mu \in PX$ given by the integral:

$$A \mapsto (E\mu)(A) := \int_{PX} p(A)\, d\mu(p) \tag{1.1.8}$$

for each measurable subset $A \subseteq X$.

Given two spaces X and Y, Kleisli morphisms $X \to PY$ correspond to stochastic maps (or Markov kernels). The identity stochastic map is exactly the delta map $\delta : X \to PX$. The Kleisli composition is defined in terms of the composition map E, which as we know from equation (1.1.8) inserts an integral. More explicitly, given $k : X \to PY$ and $h : Y \to PZ$, the composition (1.1.4) gives us

$$(h \circ_{kl} k) : x \mapsto (h \circ_{kl} k)_x = (E \circ (Ph) \circ k)_x$$

which maps a measurable set $A \subseteq Z$ to

$$(h \circ_{kl} k)_x(A) = \int_{PZ} p(A)\, d(h_* k_x)(p) = \int_Y h_y(A)\, dk_x(y). \tag{1.1.9}$$

This is the famous *Chapman-Kolmogorov formula.* Therefore the Kleisli composition of stochastic maps is exactly the usual composition of Markov kernels.

The first ideas on how to use category theory to extend functions to stochastic maps can be traced back to Lawvere [Law62]. It was first formalized in terms of monads by Giry [Gir82]. The interpretation given above is common to all probability monads in the literature, however the details of how this is carried out vary depending on the context. In particular, one needs to select:

- the right notion of space (for example, measurable, topological, metric);

- the right notion of maps (for example, measurable, continuous, short);
- *which* probability measures are allowed in PX (for example, inner regular, or compactly supported, or of finite moments),

so that the functor and natural transformations are all well-defined. For example, for most of this work we will work in the category of complete metric spaces and short maps. We will then need to make sure that:

- If X is a complete metric space, PX is constructed in such a way that it is itself a complete metric space;
- For each complete metric space X, the maps δ and E described above are well-defined, and short;

and so on. The analytical details of how this is attained are explained in Chapter 2.

1.1.2. Monads modeling spaces of formal expressions

Another interpretation of the theory of monads, also of interest for probability, is that *a monad is like a consistent choice of spaces of formal expressions of a specific kind.*

The key word here is "formal". Intuitively, a formal expression is an "operation which has not been performed". Think of the difference between "3+2" and "5". A formal expression can always be *written*, however its *result* may not be defined. For example, one could write "$a+b$", where a and b are elements of some set X which has no addition defined.[1] The expression does not return any element of X as a result, however it can still be written. The main utility of formal expressions is that, even if they cannot be evaluated, *formal expressions of formal expressions can be reduced to formal expressions.* For example,

$$(a + b + c) + (a + b + d)$$

can be reduced to

$$2a + 2b + c + d,$$

even if the latter expression remains formal. We cannot sum elements of a generic set, but we *can* sum formal sums of them, and the result will be again a formal sum. In other words, formal sums of elements of a set X *do* have a well-defined sum operation, they form a commutative monoid called the *free commutative monoid over X*. We will denote such a monoid by FX.

[1] By, "addition", here we mean an operation satisfying the axioms of a commutative monoid, like the addition of natural numbers. We will always denote such an operation, formal or not, by "+".

Suppose that now we have another set Y and a function $f : X \to Y$. We automatically get a function from formal sums of elements of X to formal sums of elements of Y by just "extending linearly". For example:

$$a + b + 2c \mapsto f(a) + f(b) + 2f(c). \tag{1.1.10}$$

We can then interpret Definition 1.1.1 in the following way. A functor $T : \mathsf{C} \to \mathsf{C}$ consists of the following assignments:

(a) To each space X, we assign a new space TX, which we think of as containing "formal expressions of elements of X of a specific kind" (for example, formal sums).

(b) Given two spaces X and Y and a function $f : X \to Y$, we get a function $Tf : TX \to TY$, which we think of as "evaluated pointwise", or as before, "extended linearly". This assignment should preserve identity and composition.

Formal expressions can be interpreted as generalized elements, and vice versa, so this interpretation and the one given in 1.1.1 are compatible (and there are many more).

In the case of formal sums, any element x can be considered a (trivial) formal sum. An analogous property is required in general in the definition of a monad, via the unit: a natural transformation $\eta : \mathrm{id}_{\mathsf{C}} \Rightarrow T$ consists of the following data:

(a) To each X we have a map $\eta : X \to TX$, with the interpretation that each element of X defines a (trivial) formal expression;

(b) For each $f : X \to Y$, Tf has to agree with f on the elements coming from X. That is, x is mapped to $f(x)$ both as an element of X, and as a trivial formal expression.

As we have seen, formal sums of formal sums can be reduced to just formal sums. This in general is encoded in the composition, a natural transformation $\mu : TT \Rightarrow T$, which consists of the following data:

(a) To each X, we have a map $\mu : TTX \to TX$ which we think of as a rule of evaluation of the nested formal expression, or of "removing the parentheses", as we have seen for formal sums;

(b) For each $f : X \to Y$, applying μ before or after applying f elementwise does not change the result.

The conditions (1.1.1) mean respectively the following, for each space X:

(a) Given a formal expression of formal expressions, if the first formal expression is trivial, then the simplification is also trivial. For formal sums, this says that a formal sum of trivial formal sums (i.e. given by single elements) is evaluated to just the formal sum of the elements For example, $(x) + (y) + (z)$ is evaluated to $x + y + z$.

(b) Given a formal expression of formal expressions, if the *second* formal expression is trivial, then the simplification is also trivial. For formal sums, this says that a trivial formal sum containing only one formal sum is evaluated to the formal sum it contains. For example, $(x + y + z)$ is evaluated to $x + y + z$.

(c) Given a formal expression of formal expressions of formal expressions (three times), there is really only one way of simplifying the expression. For example, the expression in the top left corner of the following diagram can be simplified in these two equivalent ways:

$$\begin{tikzcd} ((a+b)+(a+c)) \arrow[r, mapsto, "T\mu"] \arrow[d, mapsto, "\mu T"] & (2a+b+c) \arrow[d, mapsto, "\mu"] \\ (a+b)+(a+c) \arrow[r, mapsto, "\mu"] & 2a+b+c \end{tikzcd}$$

Intuitively, we can first remove the inner parentheses and then the outer ones, or vice versa, and the result will not change.

There are however spaces where the operations specified by the monad T *are* defined. For example, in an actual commutative monoid A (say, natural numbers with addition) the additions can be actually evaluated. An *algebra* of a monad, more generally, is precisely a space which is closed under the operations specified by the monad. Here is the category-theoretical definition.

Definition 1.1.5. *Let (T, η, μ) be a monad on a category C. An* algebra *of T, or T-algebra, consists of:*

- *An object A of C;*
- *A morphism $e : TA \to A$ of C,*

such that the following diagrams commute, called "unit" and "composition", respectively:

$$\begin{tikzcd} A \arrow[r, "\eta"] \arrow[dr, "\mathrm{id}"'] & TA \arrow[d, "e"] \\ & A \end{tikzcd} \qquad \begin{tikzcd} TTA \arrow[r, "Te"] \arrow[d, "\mu"] & TA \arrow[d, "e"] \\ TA \arrow[r, "e"] & A \end{tikzcd} \tag{1.1.11}$$

Let's see what this means in our interpretation. We have first of all an object A, which we think of as being closed under the operation specified by T. For example, a commutative monoid, closed under additions. Then we have a map $e : TA \to A$, which we can think of as actually *evaluating* the expression, turning it from formal to actual. For example, $2 + 3 \mapsto 5$.

- The unit diagram of the algebra says that if we evaluate a trivial expression, i.e. one simply coming from an element of A, the result is that element. For example, the evaluation of the trivial formal sum a gives as result again a;

- If we have a formal expression of formal expressions, we can either first remove the parentheses and then evaluate the result, or first evaluate the content of the parentheses, remove them, and then evaluate the resulting expression. The composition diagram says that the result will be the same. For example, the expression in the top left corner of the following diagram can be evaluated in these two equivalent ways:

$$\begin{tikzcd} (2+3)+(1+2) \arrow[r, mapsto, "Te"] \arrow[d, mapsto, "\mu"] & 5+3 \arrow[d, mapsto, "e"] \\ 2+3+1+2 \arrow[r, mapsto, "e"] & 8 \end{tikzcd}$$

The algebras of the "formal sum monad" described above, which is usually called the *free commutative monoid monad*, can be proven to be exactly the commutative monoids. Just as well, there is a monad on the category of sets whose algebras are groups, a monad for rings, and so on. The algebras of the power set monad are precisely the join-complete semilattices.

Consider now two commutative monoids A and B. Not every function between them respects addition. The function f preserves the addition if and only if evaluating expressions before or after applying f does not change the result. For example, if $f(a + b) = f(a) + f(b)$. In other words, f preserves additions if an only if the following diagram commutes:

$$\begin{tikzcd} TA \arrow[r, "Tf"] \arrow[d, "e"] & TB \arrow[d, "e"] \\ A \arrow[r, "f"] & B \end{tikzcd} \tag{1.1.12}$$

where we have denoted the structure maps of A and B both as e (but they are two different maps).

Definition 1.1.6. *Let (A, e) and (B, e) be T-algebras of a monad T on C. A morphism of T-algebras, or T-morphism, is a morphism $f : A \to B$ of C such that diagram* (1.1.12) *commutes.*

The category of T-algebras and T-morphisms is called the Eilenberg-Moore category of T *and it is denoted by C^T.*

For example, the Eilenberg-Moore category of the free monoid monad is the category of commutative monoids and monoid homomorphisms. The category of groups and group morphisms is the Eilenberg-Moore category of the monad of groups, and so on.

We have seen that given any set X, the set of formal sums FX is always a commutative monoid, called the free commutative monoid: sums of formal sums can be "evaluated" to formal sums. In other words, FX is an F-algebra, with as structure map exactly $\mu : FFX \to FX$. This is a general phenomenon: the space of formal expressions is automatically an algebra with the map μ as structure map.

Definition 1.1.7. *Let (T, η, μ) be a monad on a category C. A* free T-algebra *is an algebra of the form (TX, μ_X) for some object X of C.*

Every object X of C gives rise to a free algebra. The unit and composition diagrams for the algebra (TX, μ) are exactly the left unitality and associativity diagrams for T at the object X. The name "free", which generalizes the case of free commutative monoids, will be motivated in 1.1.3.

Here is now the link with probability theory: *probability measures behave like formal convex combinations*, or *formal mixtures*. Consider a coin flip, where "heads" and "tails" both have probability 1/2. Then *in some sense*, this is a convex combination of "heads" and "tails". The word "*formal*" here is the key: the set {"heads", "tails"} is not a convex space, so one can't really take actual mixtures of its elements, just as for sums in the example above. However, one can embed {heads, tails} into the space

$$\big\{\lambda \text{ ``heads''} + (1-\lambda) \text{ ``tails''} \mid \lambda \in [0,1]\big\},$$

using the map "heads" $\mapsto$ 1 "heads" + 0 "tails", "tails" $\mapsto$ 0 "heads" + 1 "tails". In *this* new space, one can actually take convex combinations: 1/2 "heads" + 1/2 "tails" is now actually a convex combination of the extremal elements. In general one does not only take finite convex combinations, but rather integrals with respect to normalized measures, so we are talking about *generalized* mixtures, in the sense of Choquet theory [Win85]. The interpretation is nevertheless the same:

- Given an object X, which we can think of a set of possible (deterministic) states, we can form an object PX, which contains "formal mixtures" of elements of X;

- Every function $f : X \to Y$ gives a function $Pf : PX \to PY$ by pointwise evaluation, or linear extension;
- X is embedded into PX via a map $\delta : X \to PX$ which maps an element $x \in X$ to the trivial formal convex combination x;
- Formal mixtures of formal mixtures can be evaluated using the map $E : PPX \to PX$: for example, in Example 1.0.1 we have

$$\frac{1}{2}\bigg(\frac{1}{2}\text{ “heads” } + \frac{1}{2}\text{ “tails”}\bigg) + \frac{1}{2}(1\text{ “heads” } + 0\text{ “tails”}) \\ \longmapsto \frac{3}{4}\text{ “heads” } + \frac{1}{4}\text{ “tails”}.$$

There are however spaces, like for example $\mathbb{R}$, where one *can* take actual mixtures. These correspond exactly to the *algebras* of P. In other word, a P-algebra is a convex space of some sort, a space which is closed under mixture operations (usually, a convex subset of some vector space). Taking expectation values is one of the most important operations in probability theory: the spaces where this can be done are precisely the algebras of a probability monad. The P-morphisms, the maps preserving the P-algebra structure, are precisely the *affine maps*, or *mixture-preserving*. In other words, given P-algebras (A, e) and (B, e), a P-morphism is a map $f : A \to B$ such that

$$f\big(\lambda\, a + (1-\lambda)\, a'\big) = \lambda\, f(a) + (1-\lambda)\, f(a')$$

for every $\lambda \in [0, 1]$ and $a, a' \in A$, or, more generally,

$$f\bigg(\int_A a\, dp(a)\bigg) = \int_A f(a)\, dp(a)$$

for every probability measure $p \in PA$.[2]

Again, the details of how this is carried out in practice vary, depending on the choice of category, of monad, and so on. So in particular, one may get different sorts of “convex spaces”. The probability monad that we present in Chapter 2, the Kantorovich monad, has as algebras precisely the closed convex subsets of Banach spaces (see 2.4.3). Another example in the literature is the Radon monad on the category of compact Hausdorff spaces: its algebras are precisely the compact convex subsets of locally convex topological vector spaces [Świ74, Kei08].

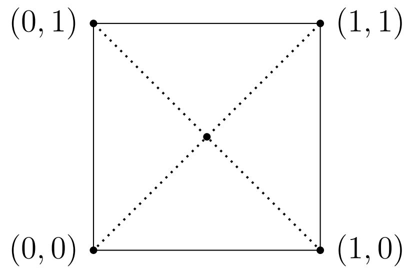
[2]It turns out that one can treat categorically also *convex* maps, not just affine maps. This is done in 3.5, and to the best of the author’s knowledge it was never done before.

Given any space X, possibly not convex, one can always form the *free P-algebra PX*, or the *free convex space over X*. Mixtures in those spaces are given by the integration map $E : PPX \to PX$. As explained in the next section, these spaces are in some sense "simplices".

1.1.3. Adjunctions, Choquet theory, stochastic matrices

Among commutative monoids, the *free ones*, i.e. those in the form FX for some set X, have a special property: their elements can be written in a unique way. For example, take the set $X = \{x, y\}$. Then the element $x + y \in FX$ is different from the element $x+x$. This is not true for all commutative monoids. For example in the natural numbers modulo 2, (which is an abelian group, and so in particular a commutative monoid), 1 can be equivalently written as $1 + 1 + 1$, and so on.

Just as well, consider a generic convex space, for example, the square in $\mathbb{R}^2$ in the following picture:



Not every point in the square can be obtained *uniquely* as a convex combination of extremal points: for example, the center of the square $(1/2, 1/2)$ can be obtained as $\frac{1}{2}(0, 0) + \frac{1}{2}(1, 1)$, as well as $\frac{1}{2}(0, 1) + \frac{1}{2}(1, 0)$.

However, if we take a *simplex*, for example a triangle, every element corresponds to a *unique* convex combination of its extremal points. In other words, in a simplex, and in a free commutative monoid, *there is a one-to-one correspondence between elements of the space and allowed operations on the generating set.* This property is usually called *freeness.* We have interpreted this as "the elements of the space TX are precisely formal expressions on X". Independently from the interpretation, the following is always true:

Proposition 1.1.8. *Let (T, η, μ) be a monad on a category C. Let X be an object of C, and (A, e) a T-algebra. Then there is a natural bijection*

$$\mathsf{C}(X, A) \cong \mathsf{C}^{\mathsf{T}}(TX, A) \tag{1.1.13}$$

between morphisms $X \to A$ of C, and T-morphisms $TX \to A$.

In the language of category theory, *every monad gives rise to an adjunction.* This is a standard result [Mac00, Chapter VI]. Depending on the choice of the monad, this gives rise to important correspondences in mathematics, for example:

- If T is the vector space monad on sets, Proposition 1.1.8 says that every linear map (T-morphism) from a vector space TX with a basis X to another vector space A is uniquely determined by its action on the elements of the basis. In finite dimension, this means precisely that a linear map is uniquely specified by a matrix;

- For a probability monad P, Proposition 1.1.8 says that *every map from a space X to a convex space A* (for example $\mathbb{R}$) *can be uniquely extended to a mixture-preserving map $PX \to A$*, and moreover that *every such mixture-preserving map arises in this way.*

In analogy with the finite case, it is customary to call a space PX the *simplex* over X. In the language of Choquet theory [Win85, Chapter 1], Proposition 1.1.8 says that *every affine function on a simplex is uniquely determined by its action on the extreme points*, and that conversely any function on the extreme points of a simplex can arise in this way. This is a rigorous way of saying that *the simplices are the free convex spaces.*

Suppose now that A is as well a free algebra, i.e. $A = PY$ for some space Y. Then maps $X \to PY$ are precisely stochastic maps. So Proposition 1.1.8 implies that a stochastic map $X \to PY$ is uniquely specified by a mixture-preserving ("linear") map $PX \to PY$. If X and Y are finite sets, this is exactly a stochastic matrix. So Proposition 1.1.8 is a generalization of the known correspondence between Markov kernels and stochastic maps.

Corollary 1.1.9. *Let (T, η, μ) be a monad on a category C. There is an equivalence of categories between the Kleisli category C_T of T, and the full subcategory of the Eilenberg-Moore category C_T whose objects are precisely the free T-algebras.*

For probability monads, this means that by the correspondence above, the category whose morphisms are stochastic maps is equivalent to the category whose objects are *simplices*, and whose morphisms are mixture-preserving maps.

1.2. Joints and marginals

We have see in Section 1.1 that we can talk about random elements categorically in terms of a probability monad. Given a category C, whose objects X we think of as spaces of possible states or outcomes, we can form spaces PX which can be thought of as containing random states or outcomes.

A central theme of probability theory is that random variables can form *joints and marginals*, and that joints may exhibit either independence, or statistical interaction of some kind. For this to make sense in C, we need C to be a *monoidal category.* A monoidal category [Mac00, Chapters VII and XI] is intuitively a category whose objects can be "glued together to form new objects". That is, given spaces X and Y, we can form a new object $X \otimes Y$, which we can think of as containing "composite states", or "joint states". This new object is conventionally called "tensor product" and denoted with the symbol $\otimes$ in analogy with the tensor product of vector spaces, however it may in practice look very different from the tensor product of vector spaces. One category may admit many monoidal structures, satisfying different properties, depending on which behavior one wants to model. For example, typical monoidal categories are:

- Sets with the cartesian product;
- Sets with the disjoint union;
- Vector spaces with the tensor product;
- Vector spaces with the direct sum.

In probability, the "joint states" are usually elements of the cartesian product, so $X \otimes Y$ has as underlying set the cartesian product of the underlying sets of X and Y. However, as an object of the category C, $X \otimes Y$ may *not* be the categorical product of X and Y, i.e. C does not need necessarily to be cartesian monoidal. For example, the monoidal structure that we define in 2.1.1 on the category of complete metric spaces is not cartesian.

In order to form joint and marginal distributions, we need P to interact well with the monoidal structure. This interaction is best modeled in terms of a *bimonoidal structure* of the monad, as we have explained in detail in the paper [FP18]. Here we give an overview of the main ideas, since some of those concepts are needed in the rest of the work (in particular, Sections 2.5 and 3.4.2). As it is well-known, the probability of the product is not the same as the product probability, so P does not directly (or *strongly*) preserve monoidal products: in general, $P(X \otimes Y) \ncong PX \otimes PY$. However, there are maps between the two spaces which make P compatible with product in a weaker sense, which (as we show in [FP18]) captures the ideas of statistical interaction and independence. In particular:

- A *monoidal* or *lax monoidal structure for the monad P* is that given two probability measures $p \in PX$ and $q \in PY$, one can canonically define a probability measure $p \otimes q \in P(X \otimes Y)$, the "product distribution". This is not the only possible joint distribution that p and q have, but it can be obtained without additional knowledge (of their correlation).

- An *opmonoidal* or *oplax monoidal structure for the monad P* formalizes the dual intuition, namely that given a joint probability distribution $r \in P(X \otimes Y)$ we canonically have the marginals on PX and PY as well. A bimonoidal structure is a compatible way of combining the two structures, in a way consistent with the usual properties of products and marginals in probability.

- The interplay between the monoidal and opmonoidal structures gives a notion of *stochastic independence* which works for general monads, and which for probability monads is equivalent to the usual notion of stochastic independence.

The interested reader is referred to the paper [FP18].

1.2.1. Semicartesian monoidal categories and affine monads

Definition 1.2.1. *A* semicartesian monoidal category *is a monoidal category in which the monoidal unit* 1 *is a terminal object.*

For probability theory, this is a very appealing structure of a category, because the object 1 can be interpreted as a trivial space, having only one possible element, or only one possible state. In other words, the object 1 would have the property that for every object X, $X \otimes 1 \cong X$ (monoidal unit), so that tensoring with 1 does not increase the number of possible states, and moreover there is a unique map $! : X \to 1$ (terminal object), which we can think of as "forgetting the state of X". cartesian monoidal categories are in particular semicartesian. Not every monoidal category of interest in probability theory is cartesian, but most of them are semicartesian. The categories of metric spaces used in the rest of this work are in particular semicartesian monoidal, as are all the categories listed in the paper [Jac17].

Semicartesian monoidal categories have another appealing feature for probability: every tensor product space comes equipped with natural projections onto its factors:

$$X \otimes Y \xrightarrow{\mathrm{id} \otimes !} X \otimes 1 \xrightarrow{\cong} X,$$

$$X \otimes Y \xrightarrow{! \otimes \mathrm{id}} 1 \otimes Y \xrightarrow{\cong} Y,$$

which satisfy the universal property of the product projections if and only if the category is cartesian monoidal. These maps are important in probability theory, because they give the *marginals*. Since these projections are automatically natural in X and Y, a semicartesian monoidal category is always equivalently a *tensor category with projections* in the sense of [Fra01, Definition 3.3]; see [Lei16] for more background.

Suppose now that P is a probability monad on a semicartesian monoidal category C. Since we can interpret the unit 1 as having only one possible (deterministic) state, it is tempting to say that just as well there should be only one possible random state: if there is only one possible outcome, then there is no real randomness. In other words, it is appealing to require that $P(1) \cong 1$. A monad with this condition is called *affine*. Most monads of interest for probability are indeed affine (in particular, again, all the ones listed in [Jac17]).

A last requirement on the monoidal structure in order to talk about probability is *symmetry*: since there is no real difference between joints on $X \otimes Y$ and joints on $Y \otimes X$, we want the category to be *symmetric monoidal*, and the monad to be compatible with the symmetry.

In the rest of this chapter, and of this work, we will always work in a symmetric semicartesian monoidal category with an affine probability monad. These conditions simplify the treatment a lot, while keeping most other conceptual aspects interesting. By the remarks above, they seem to be the right framework for classical probability theory. The definition of monoidal, opmonoidal, and bimonoidal monads can however be given for general braided monoidal categories: the interested reader can find them in Appendix A.1.

1.2.2. Bimonoidal monads and stochastic independence

Let P be an affine probability monad on a strict symmetric semicartesian monoidal category C. In this setting, a *monoidal structure* for the functor P amounts to a natural map $\nabla : PX \otimes PY \to P(X \otimes Y)$ with associativity and unitality conditions. The probabilistic interpretation is the following: given $p \in PX$ and $q \in PY$, there is a canonical (albeit not unique) way of obtaining a joint in $P(X \otimes Y)$, namely the product probability. Technically we also should need a map $1 \to P(1) \cong 1$, but due to our affineness assumption, such a map can only be the identity. The associativity condition now says that it should not matter in which way we multiply first, i.e. the following diagram must commute for all objects $X, Y, Z \in \mathsf{C}$:

$$\begin{tikzcd}
(PX \otimes PY) \otimes PZ \arrow[r, "\cong"] \arrow[d, "\nabla_{X,Y} \otimes \mathrm{id}"] & PX \otimes (PY \otimes PZ) \arrow[d, "\mathrm{id} \otimes \nabla_{Y,Z}"] \\
P(X \otimes Y) \otimes PZ \arrow[d, "\nabla_{X \otimes Y, Z}"] & PX \otimes P(Y \otimes Z) \arrow[d, "\nabla_{X, Y \otimes Z}"] \\
P((X \otimes Y) \otimes Z) \arrow[r, "\cong"] & P(X \otimes (Y \otimes Z))
\end{tikzcd}$$

so that there is really just one way of forming a product of three probability distributions. The unitality conditions say that the product distribution of some $p \in PX$ with the unique measure on 1 should be essentially the same as just p.

An *opmonoidal structure* for the functor P amounts to a natural map $\Delta : P(X \otimes Y) \to PX \otimes PY$, which we can interpret as taking a joint probability measure $r \in P(X \otimes Y)$, and returning the pair of marginals $(r_X, r_Y) \in PX \otimes PY$. Again, technically we also need a map $P(1) \to 1$, but again in this setting such a map can only be the identity. We have, dually, a coassociativity condition, a commutative diagram:

$$\begin{tikzcd}
P((X \otimes Y) \otimes Z) \arrow[r, "\cong"] \arrow[d, "\Delta_{X \otimes Y, Z}"] & P(X \otimes (Y \otimes Z)) \arrow[d, "\Delta_{X, Y \otimes Z}"] \\
P(X \otimes Y) \otimes PZ \arrow[d, "\Delta_{X,Y} \otimes \mathrm{id}"] & PX \otimes P(Y \otimes Z) \arrow[d, "\mathrm{id} \otimes \Delta_{Y,Z}"] \\
(PX \otimes PY) \otimes PZ \arrow[r, "\cong"] & PX \otimes (PY \otimes PZ)
\end{tikzcd}$$

The probabilistic interpretation is that, just as for the product probability, it does not matter in which order we take marginalize the different variables. Analogously, we have also counitality conditions, which say that the marginal distribution of some $p \in P(X \otimes 1)$ on the first factor (or of some $p \in P(1 \otimes X)$ on the second factor) is essentially just p again.

The monoidal and opmonoidal structure should interact to form a *bimonoidal structure* [AM10] for the functor P. To have that, we have first of all some unit-counit conditions, which in our setting are trivially satisfied, since they only involve maps to 1. But more importantly, the following bimonoidality (or distributivity) condition needs to hold, i.e. the following diagram has to commute:

$$\begin{tikzcd}[column sep=tiny] & P(W \otimes X) \otimes P(Y \otimes Z) \arrow[dl, "\nabla_{W\otimes X, Y\otimes Z}"'] \arrow[dr, "\Delta_{W,X} \otimes \Delta_{Y,Z}"] & \\ P(W \otimes X \otimes Y \otimes Z) \arrow[d, "\cong"'] & & PW \otimes PX \otimes PY \otimes PZ \arrow[d, "\cong"] \\ P(W \otimes Y \otimes X \otimes Z) \arrow[dr, "\Delta_{W\otimes Y, X\otimes Z}"'] & & PW \otimes PY \otimes PX \otimes PZ \arrow[dl, "\nabla_{W,Y} \otimes \nabla_{X,Z}"] \\ & P(W \otimes Y) \otimes P(X \otimes Z) & \end{tikzcd} \tag{1.2.1}$$

where the center of the diagram on the right is a swap of PX and PY. The probabilistic interpretation is roughly the following: if we take a joint measure on $W \otimes X$ and a joint measure on $Y \otimes Z$, and then form their product measure, then in the resulting coupling, W will be independent from Y and X will be independent from Z. It is analogous to the *first graphoid axiom of stochastic independence* [PP85], with trivial conditioning, which says that if a random variable X is independent from the joint (Y, Z), then it is also independent from Y alone. More details on the relation between bimonoidal structures and stochastic independence can be found in [FP18, Section 4].

An important consequence of diagram (1.2.1) is that *correlation can be forgotten, but not created*. Consider two spaces X and Y. Then given a joint distribution $r \in P(X \otimes Y)$, we can form the marginals $r_X \in PX$ and $r_Y \in PY$. If we try to form a joint again, via the product, the correlation is lost. Vice versa, instead, if we have two marginals, form their joint, and then divide them again into marginals, we expect to get our initial random variables back.

Proposition 1.2.2. *Let X, Y be objects of a symmetric semicartesian monoidal category C. Let $P : \mathsf{C} \to \mathsf{C}$ be a bimonoidal endofunctor, with $P(1) \cong 1$. Then $\Delta \circ \nabla = \mathrm{id}_{PX \otimes PY}$. In particular, $PX \otimes PY$ is a retract of $P(X \otimes Y)$.*

The proposition above is proved in [FP18, Proposition 4.1]. It is a special case of a standard result about the so-called *normal* bimonoidal functors, which can be found for example in [AM10, Section 3.5].

We can say even more about the structure of joints and marginals: the whole monad structure should respect the bimonoidal structure of P, i.e. $\delta : X \to PX$ and $E : PPX \to PX$ should commute with the operations of taking products and marginals. In other words, we are saying that δ and E should be *bimonoidal natural transformations*. In more concrete terms, it means that the delta over the pair $(x, y) \in X \otimes Y$ is the product of the deltas over $x \in X$ and $y \in Y$, and vice versa that the marginals of a product delta are precisely the deltas over the projections. The same can be said about the average map E: the product of the average is the average of the product, and the marginals of an average are the averages of the marginals. These last conditions may seem a bit obscure, but they come up naturally in probability: see as an example the case of the Kantorovich monad (Section 2.5). These conditions can be summarized in the fact that P is a bimonoidal monad.

Definition 1.2.3. *A* bimonoidal monad (P, δ, E) *is a monad whose functor is a bimonoidal functor, and whose unit and composition are bimonoidal natural transformations.*

The general, diagrammatic definitions are given in Appendix A.1.

1.2.3. Algebra of random variables

A corollary of the so-called "law of the unconscious statistician" is that given a function $f : X \to Y$ and a random variable on X with law $p \in PX$, the law of the image random variable under f will be the push-forward of p along f. In categorical terms, this simply means that P is a functor, and that the image random variable has law $(Pf)(p)$, where $Pf : PX \to PY$ is given by the push-forward.

The bimonoidal structure of P comes into play whenever we have functions to and from product spaces. Consider a morphism $f : X \otimes Y \to Z$. Given random variables X and Y, we can form an image random variable on Z in the following way: first we form the joint on $X \otimes Y$ using the monoidal structure, and then we form the image under f. In other words, in terms of laws we perform the following composition:

$$PX \otimes PY \xrightarrow{C} P(X \otimes Y) \xrightarrow{f_*} PZ. \tag{1.2.2}$$

For maps in the form $g : X \to Y \otimes Z$ we can proceed analogously by forming the marginals, using the opmonoidal structure:

$$PX \xrightarrow{f_*} P(Y \otimes Z) \xrightarrow{m} PY \otimes PZ. \tag{1.2.3}$$

This way, together with associativity and coassociativity, one can form functions to and from arbitrary products of random variables.

Whenever we have an internal structure, like an internal monoid, this way we can extend the operations on the random elements, via *convolution*. For example, if X is a monoid, then also PX becomes a monoid, using $PX \otimes PX \to P(X \otimes X) \to PX$ for the multiplication. The analogous statements apply for coalgebraic structures. In other words, the bimonoidal structure allows to have an *algebra (and coalgebra) of random variables* whenever the deterministic variables form an internal algebraic structure. For example, if as monoid we take the real line with addition, as convolution algebra we get the usual convolution of probability measures. We notice that such a convolution algebra is a monoid (with the neutral element given by the Dirac delta at zero), but *not* a group: only the *monoid* structure is inherited, in general.

1.2.4. Categories of probability spaces

In the literature, many categorical treatments of probability theory are in categories whose objects are probability spaces, or fixed probability measures on a space, rather than categories with a probability monad [Fra01, Sim18]. In particular, two types of categories are of interest:

- Probability spaces as objects, and measure-preserving maps as morphisms;
- Probability spaces as objects, and stochastic maps (or conditionals) as morphisms.

Both categories can be formed from a probability monad in a canonical way. First of all, measure-preserving maps are the same as the morphisms in a suitable arrow category:

Definition 1.2.4. *Let C be a category with terminal object 1 and P a probability monad on C. Then the category $\mathsf{Prob(C)}$ is defined to be the co-slice category $1/P$. In other words:*

- *Objects of $\mathsf{Prob(C)}$ are objects X of C together with arrows $1 \to PX$ of C;*
- *Morphisms of $\mathsf{Prob(C)}$ are maps $f : X \to Y$ of C which makes the diagram*

$$\begin{tikzcd} & 1 \arrow[dl] \arrow[dr] & \\ PX \arrow[rr, "Pf"] & & PY \end{tikzcd}$$

commute.

In analogy with the category of elements, we can interpret $\mathsf{Prob(C)}$ as a *category of random elements*, or of probability spaces. The objects can be interpreted as elements of PX, i.e. probability measures on X, and the morphisms can be interpreted as maps preserving the selected element in the space of measures, i.e. measure-preserving maps.

Analogously, by replacing C with the Kleisli category C_T (whose morphisms, as seen in 1.1.1, can be thought of as stochastic maps), we get a category whose objects are probability spaces, and whose morphisms are stochastic maps. We denote such a category $\mathsf{Stoch}(\mathsf{C})$.

Under some mild assumptions, if C has a semicartesian monoidal structure (as we explained in 1.2, we can transfer that structure to the category of random elements, with a construction analogous to that of Section 1.2.3.

Definition 1.2.5. *Let C be a semicartesian monoidal category and P an affine probability monad on C with monoidal structure ∇. We define the following monoidal structure on $\mathsf{Prob}(\mathsf{C})$: given $p : 1 \to PX$ and $q : 1 \to PY$, we define $p \otimes_\nabla q : 1 \to P(X \otimes Y)$ to be the composition:*

$$1 \cong 1 \otimes 1 \xrightarrow{p \otimes q} PX \otimes PY \xrightarrow{\nabla} P(X \otimes Y).$$

and for morphisms we proceed analogously.

With a slight abuse, but in agreement with the probability literature, we will denote the product probability by $p \otimes q$ instead of $p \otimes_\nabla q$.

This way $(\mathsf{Prob}(\mathsf{C}), \otimes)$ is a semicartesian monoidal category, with the unit $1 \to 1$ isomorphic to the terminal object. This generalizes the construction given in Section 3.1 therein (in which the base category Meas is cartesian monoidal). The same can be done for $\mathsf{Stoch}(\mathsf{C})$.

It is worth noting that, even if C is cartesian monoidal, in general $\mathsf{Prob}(\mathsf{C})$ and $\mathsf{Stoch}(\mathsf{C})$ will not be *cartesian* monoidal, but only semicartesian. In other words, the product of probability spaces does *not* satisfy the universal property of a categorical product, and the reason is that *uniqueness* fails: given two probability spaces (X, p) and (Y, q), there are many possible measures on $X \times Y$ whose marginals are p and q, respectively. The fact that $p \otimes q$ is the "canonical" choice is not enough to make it the categorical product. However, even if $p \otimes q$ does not satisfy a universal property, the intuition that $p \otimes q$ "has the same amount of information as p and q separately" can be made precise by means of the bimonoidal nature of probability monads (thanks to Proposition 1.2.2). The details can be found in [FP18, Section 4].

2. The Kantorovich Monad

In this chapter we will define and study a particular probability monad on the category of complete metric spaces, the *Kantorovich monad*. It was introduced by van Breugel [vB05] in 2005 on the category of compact and on the category of 1-bounded complete metric spaces, and we extended it in [FP17] to all complete metric spaces.

The basic idea is that given a metric space X, as space of probability measures PX one can take the *1-Wasserstein space* over X, which is itself a metric space, sharing many properties with the underlying space (like compactness or completeness). The 1-Wasserstein distance (or Kantorovich-Rubinstein distance) has the necessary convexity properties which make this assignment part of a monad.

It is well-known [Vil09, Chapter 6] that finitely supported probability measures with rational coefficients, or *empirical distributions of finite sequences*, are dense in the Wasserstein space. This allows to define Wasserstein spaces in terms of a universal property, as a colimit. Moreover, finite sequences can be nested (to form sequences of sequences), and so spaces of finite sequences form naturally a structure similar to a monad, called a *graded monad* [FKM16]. We prove that, under suitable assumptions, the colimit of a graded monad gives rise to a monad. This allows us to define the monad structure of P, especially the integration map, *directly in terms of this colimit construction*, without the need to use measure theory (but the resulting map will be the same). The monad structure obtained this way is compatible with the formations of joints and marginals, and it has appealing geometric properties which allow to talk about some disintegration results in a purely categorical way.

Outline

- In Section 2.1 we introduce the main mathematical constructions that we use in this chapter: the categories Met and CMet of (complete) metric spaces and short maps, and the Radon probability measures on them with finite first moment. We prove (Theorem 2.1.3) that such measures are equivalently linear, positive, Scott-continuous functionals on the space of Lipschitz functions. Using this, in Section 2.1.4 we introduce the Wasserstein metric, and we show the functoriality of the Wasserstein space construction (Lemma 2.1.14), resulting in the *Kantorovich functor* P.

- In Section 2.2 we prove (Theorem 2.2.18 and Corollary 2.2.20) that the Wasserstein spaces and the Kantorovich functor can be obtained as colimits of the spaces of finite sequences and of their associated *power functors*, defined in 2.2.1, and that the universal arrow is given by the empirical distribution map, which we define in 2.2.2.
- In Section 2.3 we prove that P has a monad structure (Theorem 2.3.8), which arises naturally from its colimit characterization, given the particular graded monad structure of the power functors (Theorems 2.3.1 and 2.3.2). This can be interpreted as a Kan extension in the 2-category MonCat of monoidal categories and lax monoidal functors (Theorem 2.3.3).
- In Section 2.4 we study the algebras of P. We show (Theorem 2.4.2) that the algebras are equivalently convex 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 2.4.3).
- In Section 2.5 we prove (Theorem 2.5.17) that P has a symmetric bimonoidal monad structure, where the monoidal and opmonoidal parts have the operational meaning of forming product probabilities from given marginals, and of forming marginal probabilities from given joints, in agreement with the discussion in Section 1.2.
- Finally, in Section 2.6 we prove a lifting property for the integration map E (Proposition 2.6.5), with which we show that E is proper (Theorem 2.6.7). This allows in turn to state a disintegration-like result (Theorem 2.6.9).

The content of this chapter is mostly contained in the paper [FP17], with the exception of Section 2.5, which is contained in [FP18, Section 5], and Section 2.6, which will be part of a paper currently in preparation.

2.1. Wasserstein spaces

The basic idea behind Wasserstein spaces is the following: given a metric space X, and a set PX of suitably regular probability measures on X, we want to equip PX with a metric compatible with the metric of X. If, as in 1.1.1, we view PX as an extension of X in which X sits embedded via the Dirac delta map $\delta : X \to PX$, it is natural to require that the metric on PX makes δ an isometric embedding, i.e. for all $x, y \in X$,

$$d_{PX}(\delta_x, \delta_y) = d_X(x, y). \tag{2.1.1}$$

This requirement makes Wasserstein metrics different from other metrics for probability measures (such as the total variational distance), in that *point measures over neighboring points are themselves neighboring*, even if they have no actual overlap. So Wasserstein metrics keep track nontrivially of the distance and topology of the underlying space.

Clearly, (2.1.1) is not enough to determine the metric uniquely, we need to see how the metric works when the measures are nontrivial. Consider three points $x, y_1, y_2 \in X$, and the probability measures $p = \delta_x$ and $q = \frac{1}{2}\,\delta_{y_1} + \frac{1}{2}\,\delta_{y_2}$. We would like $d_{PX}(p, q)$ to lie between $d_X(x, y_1)$ and $d_X(x, y_2)$. A possible choice is

$$d_{PX}(p, q) = \frac{1}{2}\,d_X(x, y_1) + \frac{1}{2}\,d_X(x, y_2). \tag{2.1.2}$$

This can be interpreted in the following way: half the mass of p has to be moved from x to y_1, and the other half from x to y_2. Therefore the total cost of transport is

$$\text{amount of mass} \cdot \text{distance} + \text{amount of mass} \cdot \text{distance} = \frac{1}{2}\,d_X(x, y_1) + \frac{1}{2}\,d_X(x, y_2).$$

For this reason, the distance obtained from the choice (2.1.2) is sometimes called the *earth mover's distance*. Another interpretation, in line with the formal convex combinations of 1.1.2, would be that *the distance between formal convex combinations is the convex combination of the distances.*

If p also is nontrivial, for example $p = \frac{1}{2}\,\delta_{x_1} + \frac{1}{2}\,\delta_{x_2}$, there are at least two possible ways of moving the mass between p and q: moving the mass at x_1 to y_1 and the mass at x_2 to y_2, or moving the mass at x_1 to y_2 and the mass at x_2 to y_1, in pictures:

$$\begin{tikzcd} x_2 \arrow[r] & y_2 \\ x_1 \arrow[r] & y_1 \end{tikzcd} \qquad \text{and} \qquad \begin{tikzcd} x_2 \arrow[dr] & y_2 \\ x_1 \arrow[ur] & y_1 \end{tikzcd}$$

or even a combination of the two. In this case, the distance will be the optimal choice between these possibilities, that is:

$$d(p, q) = \min_{\sigma \in S_2} \big(d(x_1, y_{\sigma(1)}) + d(x_2, y_{\sigma(2)}) \big).$$

Since we are optimizing an affine functional and all the possibilities form a convex set, it is sufficient to optimize over the extreme points (see Proposition 2.2.10), which are permutations (in this case, of y_1 and y_2). This procedure specifies the metric uniquely, as we will show in Section 2.2. The resulting distance is called the *1-Wasserstein distance*, or *Kantorovich-Rubinstein distance*.[1] We will define it rigorously in 2.1.4.

Another possible choice alternative to (2.1.2), with more "Euclidean" or "Riemannian" properties, is

$$d_{PX}(p,q) = \sqrt{\frac{1}{2}\, d_X(x,y_1)^2 + \frac{1}{2}\, d_X(x,y_2)^2}. \tag{2.1.3}$$

This gives the so-called 2-Wasserstein distance. The same can be done for any positive p, in analogy with the L^p norms. In this work, we will only work with $p = 1$.

A treatment of the Wasserstein spaces and their interpretation in terms of optimal transport can be found for example in [Vil09].

2.1.1. Categorical setting

There are two categories that are of primary interest to us. The first one is the monoidal category Met, where:

- Objects are metric spaces, which we will refer to as "spaces";
- Morphisms are short maps (or 1-Lipschitz maps), i.e. functions $f : X \to Y$ such that for all $x, x' \in X$:
$$d_Y(f(x), f(x')) \le d_X(x, x')\ ; \tag{2.1.4}$$
- As monoidal structure, we define $X \otimes Y$ to be the set $X \times Y$, equipped with the ℓ^1-product metric:
$$d_{X\otimes Y}\big((x,y),(x',y')\big) := d_X(x,x') + d_Y(y,y')\ . \tag{2.1.5}$$

The second one is its full subcategory CMet, consisting of *complete metric spaces* and short maps.

The choice of these morphisms and monoidal structure can be partially motivated by the following remarks:

- 1-Lipschitz maps, as opposed for example to just continuous maps, are sensitive precisely to distances, and not just to the underlying topology. In particular, in CMet the isomorphisms are precisely the *isometries*, and the extremal monomorphisms are precisely the *isometric embeddings*. This allows us in 2.2 to state a density result categorically, as a colimit.

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

- In order to still retain finite distances between the measures, one is forced to choose between restricting the spaces to just the bounded ones, and restricting the maps to just the Lipschitz or 1-Lipschitz ones (see Remark 2.1.15). All the structural functions of use in probability theory (like those arising from the formation of joints, marginals, integrals, etc.) are 1-Lipschitz, provided one chooses the right metrics. Thus the restriction to 1-Lipschitz maps seems to be the most convenient choice;

- From the categorical perspective, metric spaces and short maps can be considered particular enriched categories and functors [Law73, Law86]. In this view, one can see that the above monoidal structure is closed both in Met and in CMet, where in both cases the exponential object Y^X is the space of short maps $X \to Y$ with the supremum distance [Law73, Section 2].

Further motivation will be given in 2.1.4, in 2.2, and in 2.5. Other choices of base categories for probability monads appearing in the literature can be found for example in [Jac17].

2.1.2. Analytic setting

Here we fix the analytic setting of the rest of this work. The following definitions will be needed in particular in this section, in Section 2.2, where we prove our colimit characterization by density, and in Chapter 3.

Every metric space is in particular a topological space, and so also a measurable space with the Borel σ-algebra. All our probability measures are Radon, i.e. Borel measures which are tight (equivalently, inner regular).

For $X \in \mathsf{Met}$, we write $\mathrm{Lip}(X)$ for the space of Lipschitz functions $X \to \mathbb{R}$, where $\mathbb{R}$ carries its usual Euclidean metric. Every Lipschitz function is a scalar multiple of an element of $\mathsf{Met}(X, \mathbb{R})$, i.e. a short map $X \to \mathbb{R}$. We expect that working with the latter space, or even just with $\mathsf{Met}(X, \mathbb{R}_+)$, would be the way to go for achieving further abstraction. However, currently we prefer to work with $\mathrm{Lip}(X)$, which has the added convenience of being a vector space.

2.1.3. Finite first moments and a representation theorem

In order to define our Wasserstein spaces, we first have to define probability measures of finite first moment, which are precisely those for which every Lipschitz function has an expectation value.

Definition 2.1.1. *Let $X \in \mathsf{Met}$ 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.

Lemma 2.1.2. *The following are equivalent for a probability measure p on $X \in \mathsf{CMet}$:*

(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 $f \in \mathrm{Lip}(X)$ 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)⇒(b): if the integral of a nonnegative function is finite, then the integrand is finite at at least one point.
- (b)⇒(c): For all $z \in X$, and for y as in (b), we have:

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

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

- (c)⇒(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)⇒(a): Since the distance is short in both arguments, 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. □

So from now on, we write PX for the set of probability measures on X with finite first moment. Below, we will equip this set itself with a metric, but for now it is just a set. As we also discuss in more detail below, pushing forward measures along a short map $f : X \to Y$ defines a function $Pf : PX \to PY$ which makes P into a functor.

A general theme is that measures are specified by how they act on functions by integration, e.g. as in the definition of the Daniell integral or in the Riesz representation theorem. We will now get to an analogous result for PX. Concretely, every $p \in PX$ defines a linear functional $\mathbb{E}_p : \mathrm{Lip}(X) \to \mathbb{R}$ given by mapping every function to its expectation value,
$$f \longmapsto \mathbb{E}_p(f) := \int_X f(x)\,dp(x). \tag{2.1.6}$$
We can thus consider $\mathbb{E}$ as a map $\mathbb{E} : PX \to \mathrm{Lip}(X)^*$ 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 $\mathrm{Lip}(X)$ 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 $\mathrm{Lip}(X)$ indexed by a directed set I, such that $f_\alpha \leq f_\beta$ if $\alpha \leq \beta$. It is *bounded* if there is $g \in \mathrm{Lip}(X)$ such that $f_\alpha \leq g$ for all α; this is a stronger condition than pointwise boundedness, since the bound itself must be Lipschitz. It is easy to that if f_α is short for every α, then $\sup_\alpha f_\alpha$ is also short. Then the formation of the expectation value satisfies
$$\mathbb{E}_p\Big(\sup_\alpha f_\alpha\Big) = \sup_\alpha \mathbb{E}_p(f_\alpha),$$

by Lebesgue's monotone convergence theorem.

And these properties indeed characterize integration against a measure, as per the following representation theorem similar to [Edg98, Theorem 2.4.12]:

Theorem 2.1.3. *Let $X \in \mathsf{Met}$. 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 $\phi : \mathrm{Lip}(X) \to \mathbb{R}$ with the following properties:*

- *Positivity: $f \geq 0$ implies $\phi(f) \geq 0$;*
- *τ-smoothness: if $(f_\alpha)_{\alpha \in I}$ is a bounded monotone net in $\mathrm{Lip}_1(X)$, then*
$$\phi(\sup_\alpha f_\alpha) = \sup_\alpha \phi(f_\alpha).$$
- *Normalization: $\phi(1) = 1$.*

The concept of τ-smoothness is also known as *Scott continuity* in the context of domain theory, or *normality* in the context of von Neumann algebras.

Proof. We already know that every $\mathbb{E}_p$ has these properties. Furthermore, it is shown in [Edg98, Theorem 2.4.12] that integration defines a bijection between probability measures on X and σ-smooth positive linear functionals on bounded Lipschitz functions, where σ-smoothness is the special case of τ-smoothness for indexing set $\mathbb{N}$. So starting with ϕ, we therefore obtain a probability measure p which represents ϕ on bounded Lipschitz functions. For an arbitrary $f \in \mathrm{Lip}(X)$ with $f \geq 0$ and $b \in \mathbb{R}_+$, consider the function given by truncating f at b,

$$f_b(x) := \min\{b, f(x)\},$$

so that $(f_b)_{b \in \mathbb{R}_+}$ is a bounded monotone net of bounded Lipschitz functions with supremum f. Thus by monotone convergence of the integral and the assumed τ-smoothness of ϕ, we have

$$\int f(x)\, dp(x) = \sup_b \int f_b(x)\, dp(x) = \sup_b \phi(f_b) = \phi(f),$$

which proves both that p has finite first moment, and that it represents ϕ on all positive Lipschitz functions, which is enough. It is clear that this construction is inverse to the map $p \mapsto \mathbb{E}_p$. □

A notion that will be useful in the rest of this work is the notion of *dual pair*, or *dual system* [AT07, Definition 8.6]. We repeat the definition for convenience.

Definition 2.1.4. *A* dual system *is a pair of vector spaces* (L, L') *equipped with a bilinear mapping* $\langle \cdot, \cdot \rangle : L \times L' \to \mathbb{R}$ *such that:*

- *If* $\langle x, x' \rangle = 0$ *for all* $x' \in L'$*, then* $x = 0$*;*
- *If* $\langle x, x' \rangle = 0$ *for all* $x \in L$*, then* $x' = 0$*.*

Intuitively, the spaces L and L' need to separate the points of each other by means of the pairing $\langle \cdot, \cdot \rangle$. The pairing induces locally convex topologies on both spaces which are dual to each other (see [AT07, Section 8.2]). Theorem 2.1.3 now says precisely the following:

Corollary 2.1.5. *Let X be a metric space. Let $M(X)$ be the set of signed Radon measures of finite first moment. Then by Theorem 2.1.3, the spaces* $\mathrm{Lip}(X)$ *and* $M(X)$ *together with the integration*

$$(f, \mu) \longmapsto \int f \, d\mu$$

form a dual pair.

Definition 2.1.6. *We will call the dual system* $(\mathrm{Lip}(X), M(X))$ *given above the* dual system over X.

When we do not talk about this dual system, we will always use only probability measures, normalized and nonnegative. We collect another property for future use, which relies crucially on the nonnegativity of a measure:

Lemma 2.1.7. *Let $p \in PX$ and $f : X \to Y$ 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. $\square$

2.1.4. Construction of the Wasserstein space

A central theme of this work is the celebrated *Kantorovich duality* [Vil09, Chapter 5]. The following formulation can be obtained from [Vil09, Theorem 5.10] together with [Vil09, Particular Case 5.4])

Theorem 2.1.8 (Kantorovich duality)**.** *Let X be a Polish space. Let p and q be Radon probability measures on X, and let $c : X \otimes X \to \mathbb{R}_+$ be a lower-semicontinuous function satisfying the triangle inequality. Then we have an equality:*

$$\inf_{r \in \Gamma(p,q)} \int_{X \times X} c(x,y)\, dr(x,y) = \sup_f \bigg(\int_X f dq - \int_X f\, dp \bigg), \tag{2.1.7}$$

where the infimum is taken over the space $\Gamma(p,q)$ of couplings *between p and q, and where $f : X \to \mathbb{R}$ varies over functions which have finite integral with both measures p and q, and such that $f(y) - f(x) \le c(x,y)$ for all $x, y \in X$.*

The form of Kantorovich duality that we will always use in this work is the following:

Corollary 2.1.9. *Let X be a complete metric space. Let $c : X \otimes X \to \mathbb{R}_+$ be a lower-semicontinuous function bounded above by the distance, and which satisfies the triangle inequality. Let $p, q \in PX$. Then there is an equality*

$$\inf_{r \in \Gamma(p,q)} \int_{X \times X} c(x,y)\, dr(x,y) = \sup_f \bigg(\int_X f dq - \int_X f\, dp \bigg), \tag{2.1.8}$$

where $f : X \to \mathbb{R}$ varies over functions such that $f(y) - f(x) \le c(x,y)$ for all $x, y \in X$.

Proof of the Corollary. Let now X be a complete metric space. We know that the support of the Radon measures p and q is separable. Denote now by $\tilde{X}$ the union of the supports of p and q. It is by construction separable, as it is the union of separable sets, and closed, since it is the union of closed sets. Therefore it is complete, and so Polish. Moreover, the supremum is taken over maps f such that $f(y) - f(x) \le c(x,y) \le d(x,y)$, i.e. they are short. Short maps can always be extended from a closed subset to the whole space in the following way: given $f : \tilde{X} \to \mathbb{R}$, we define $f' : X \to \mathbb{R}$ to be

$$f'(x) := \sup_{y \in \tilde{X}} \big(f(y) - d(x,y) \big).$$

Therefore the supremum over such short maps $f : X \to \mathbb{R}$ can be equivalently taken over maps $f : \tilde{X} \to \mathbb{R}$. We can then apply Theorem 2.1.8 to get:

$$\begin{aligned} \inf_{r \in \Gamma(p,q)} \int_{X \times X} c(x,y)\, dr(x,y) &= \inf_{r \in \Gamma(p,q)} \int_{\tilde{X} \times \tilde{X}} c(x,y)\, dr(x,y) \\ = \sup_{f : \tilde{X} \to \mathbb{R}} \bigg(\int_{\tilde{X}} f dq - \int_{\tilde{X}} f\, dp \bigg) &= \sup_{f : X \to \mathbb{R}} \bigg(\int_X f dq - \int_X f\, dp \bigg). \end{aligned}$$

Since p and q have finite first moments, moreover, the integral of such short f with both measures will always exist. $\square$

We can now define the Wasserstein spaces, which we will use in the rest of this work.

Definition 2.1.10. *Let $X \in \mathsf{CMet}$. The* Wasserstein space PX *is the set of Radon probability measures on X with finite first moment, with metric given by the* Wasserstein distance*, or* Kantorovich-Rubinstein distance*, or* earth mover's distance*:*

$$d_{PX}(p,q) := \inf_{r \in \Gamma(p,q)} \int_{X \times X} d_X(x,y)\, dr(x,y) \tag{2.1.9}$$

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

Applying Kantorovich duality, one can also characterize the Wasserstein metric as

$$d_{PX}(p,q) = \sup_{f:X\to\mathbb{R}} \left| \int f(x)\, d(p-q)(x) \right| = \sup_{f:X\to\mathbb{R}} (\mathbb{E}_p[f] - \mathbb{E}_q[f]), \tag{2.1.10}$$

where the sup is taken over all the short maps [Vil09, AGS05], which we think of as the 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:

Lemma 2.1.11. *Let $\delta(x_0)$ the Dirac measure at some $x_0 \in X$.[2] Then*

$$d(\delta(x_0), p) = \int d(x_0, x)\, dp(x). \tag{2.1.11}$$

Proof. The only possible 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,

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

So in particular, condition (2.1.1) is satisfied: the Kantorovich distance on PX really extends the distance of X.

[2] We will sometimes write $\delta(x_0)$ instead of δ_{x_0}, implying the map $\delta : X \to PX$. A rigorous definition of this map is given in 2.3.3.

Theorem 2.1.12 ([Edg98, Theorems 2.5.14 and 2.5.15])**.** *Let $X \in \mathsf{CMet}$. Then PX is also a complete metric space.*

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

Lemma 2.1.13. *If $f : X \to Y$ is an isometric embedding, then so is $Pf : PX \to PY$.*

Proof. This follows from the duality formula (2.1.10) together with the fact that for $X \subseteq Y$, every 1-Lipschitz function $g : X \to \mathbb{R}$ can be extended to Y, e.g. via

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

We would like the construction $X \mapsto PX$ to be functorial in X, and this indeed turns out to be the case. For $f : X \to Y$, we define $Pf : PX \to PY$ to be given by 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 $g : Y \to \mathbb{R}$,

$$\mathbb{E}_{f_*p}(g) = \int_Y g(y)\, d(f_*p)(y) = \int_X g(f(x))\, dp(x) = \mathbb{E}_p(g \circ f), \tag{2.1.12}$$

which can be occasionally useful. While preservation of composition and identities are clear, there are still two small things to check in order to establish functoriality:

Lemma 2.1.14. *Let $f : X \to Y$ be short, and $p \in PX$. Then,*

(a) f_*p has finite first moment as well;

(b) $f_* : PX \to PY$ is short.

Proof. (a) For $g : Y \to \mathbb{R}$ any Lipschitz map, we have $\mathbb{E}_{f_*p}(g) = \mathbb{E}_p(g \circ f) < \infty$ by (2.1.12) and by assumption.

(b)

$$\begin{aligned} d_{PY}\big(f_*p, f_*q\big) &= \sup_{g:Y\to\mathbb{R}} (\mathbb{E}_{f_*p}(g) - E_{f_*q}(g)) = \sup_{g:Y\to\mathbb{R}} (\mathbb{E}_p(g \circ f) - E_q(g \circ f)) \\ &\le \sup_{h:X\to\mathbb{R}} (\mathbb{E}_p(h) - \mathbb{E}_q(h)) = d_{PX}\big(p, q\big). \end{aligned} \qquad \square$$

Thus we have a functor $P : \mathsf{Met} \to \mathsf{Met}$. By Theorem 2.1.12, P restricts to an endofunctor of CMet, which we also denote by P. This is the functor that we will work with from now on. We call it the *Kantorovich functor*, in accordance with [vB05].

Remark 2.1.15. Proposition 2.1.14 does not work if we allow f to be more generally continuous: f_*p may in that case have infinite first moment, and so it would not define an element of PY. So in that case, P would not be a functor.

2.2. Colimit characterization

It is well-known that finitely supported measures with rational coefficients are dense in PX [Bas15, Proposition 1.9]. Since those measures are specified by powers of X up to permutations, one can obtain PX as the Cauchy completion of the space of symmetrized powers, provided that one equips such space with the right metric. In this section we want to give a *categorical treatment* of this density result: in the category CMet, PX can be obtained by a universal property, as the *colimit* of a diagram of powers of X. This is in turn used to give a characterization of the *functor P itself* as a colimit of certain *power functors*. We will use this universal property in Section 2.3 to show that P, constructed in this way, has a canonical monad structure.

2.2.1. Power functors

For $X \in \mathsf{Met}$ 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 relative to those of the latter space,

$$d_{X^n}\big((x_1,\dots,x_n),(y_1,\dots,y_n)\big) := \frac{d_X(x_1,y_1)+\ldots+d_X(x_n,y_n)}{n}. \tag{2.2.1}$$

One way to motivate this renormalization is that the diagonal map $X \to X^{\otimes n}$ is not short[3], while the diagonal map $X \to X^n$ is an isometric embedding which we call the *n-copy embedding*. Another motivation is given in [FP17, Appendix A], where it is shown how Met is a pseudoalgebra of the simplex operad in such a way that the power X^n is precisely the uniform n-ary "convex combination" of X with itself.

Now 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 σ. The elements of X_n are therefore multisets $\{x_1,\dots,x_n\}$. The quotient metric is explicitly given by

$$d_{X_n}\big(\{x_1\dots x_n\},\{y_1\dots y_n\}\big) := \min_{\sigma\in S_n} \frac{1}{n}\sum_{i=1}^{n} d_X(x_i,y_{\sigma(i)}), \tag{2.2.2}$$

since this is exactly the minimal distance between the two fibers of the quotient map $q_n : X^n \to X_n$, and these distances already satisfy the triangle inequality. Due to this formula, the composite $X \to X^n \to X_n$ is also an isometric embedding, which we call the *symmetrized n-copy embedding* $\delta_n : X \to X_n$. It is clear that the assignments $X \mapsto X^n$

[3] This is related to the fact that the symmetric monoidal category $(\mathsf{Met},\otimes)$ is semicartesian, but not cartesian.

and $X \mapsto X_n$ are functorial in $X \in \mathsf{Met}$, so that we have functors $(-)^n : \mathsf{Met} \to \mathsf{Met}$ and $(-)_n : \mathsf{Met} \to \mathsf{Met}$. The quotient map is a natural transformation $q_n : (-)^n \Rightarrow (-)_n$.

There is a simple alternative way to write the metric (2.2.2) that makes the connection with the Wasserstein distance (2.1.9):

Lemma 2.2.1.

$$d_{X_n}(\{x_i\}, \{y_i\}) = \min_A \frac{1}{n} \sum_{i,j} A_{ij}\, d(x_i, y_j), \tag{2.2.3}$$

where A ranges over all bistochastic matrices[4]*.*

Proof. This is upper bounded by (2.2.2) since every permutation matrix is bistochastic; conversely, every bistochastic matrix is a convex combination of permutation matrices (Birkhoff–von Neumann theorem), so that the linear optimization of (2.2.3) attains the optimum on one of these. □

Lemma 2.2.2. *If $f : X \to Y$ is an isometric embedding, then so are $f^n : X^n \to Y^n$ and $f_n : X_n \to Y_n$.*

Proof. Clear. □

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 $x_{(-)} : S \to X$ equipped with the rescaled ℓ^1-metric,

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

The idea is that the points of X^S are finite samples indexed by a set of observations S, and a function $x_{(-)} : S \to X$ 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 $X^T \to X^S$ for arbitrary injective $S \to T$, corresponding to semicartesianness of $(\mathsf{Met}, \otimes)$. But due to the rescaling by $\frac{1}{|S|}$, the functoriality now is quite different:

Lemma 2.2.3. *Whenever $\phi : S \to T$ has* fibers of uniform cardinality*, we have an isometric embedding $- \circ \phi : X^T \to X^S$.*

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

[4]We recall that a bistochastic matrix is a square matrix whose row and columns all sum to one.

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

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

□

Definition 2.2.4. *Let* FinUnif *be the monoidal category where:*

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

$$|\phi^{-1}(t)| = |S|/|T| \quad \forall t \in T. \tag{2.2.4}$$

- *The monoidal structure is given by cartesian product*[5]*.*

In particular, 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 FinUnif is precisely that subcategory of FinSet which contains the measure-preserving maps.

In the following, we either use the powers X^S for finite sets $S \in$ FinUnif, or equivalently the X^n. In the latter case, we take the n to be the objects of a skeleton of FinUnif indexed by positive natural numbers n. By equivalence of categories, we are free to choose whatever picture fits our current context more adequately.

We write $X^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \to \mathsf{CMet}$ for the *power functor* of Lemma 2.2.3.

Definition 2.2.5. *Let* N *be the monoidal poset of positive natural numbers $\mathbb{N} \setminus \{0\}$ ordered by reverse divisibility, so that a unique morphism $n \to m$ exists if and only if $m|n$, and monoidal structure given by multiplication.*

[5]This is not the categorical product. In fact, FinUnif does not have any nontrivial products, but it is semicartesian monoidal.

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

In analogy with the power functor $X^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \to \mathsf{CMet}$, we can also consider the *symmetrized power functor* $X_{(-)} : \mathsf{N}^{\mathrm{op}} \to \mathsf{CMet}$ which takes $n \in \mathsf{N}$ to X_n, and the unique morphism $m \to mn$, or $m|mn$, goes to the embedding $X_{m|mn} : X_m \to X_{mn}$ given by n-fold repetition on multisets,

$$\{x_1, \dots, x_m\} \longmapsto \{x_1, \dots x_m, \ \dots, \ x_1, \dots, x_m\}. \tag{2.2.5}$$

which is clearly natural in X. One can also consider this as arising from diagrams of the form

$$\begin{tikzcd} X^T \arrow[r, "X^\phi"] \arrow[d] & X^S \arrow[d] \\ X_{|T|} \arrow[r, "X_{|T|\,|\,|S|}"] & X_{|S|} \end{tikzcd} \tag{2.2.6}$$

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

Lemma 2.2.6. *$X_{m|mn} : X_m \to X_{mn}$ is an isometric embedding.*

Proof. Let $\{x_i\}$, $\{y_i\} \in X_m$. Then using Lemma 2.2.1, we can write

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

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

Thus we have a functor $X_{(-)} : \mathsf{N}^{\mathrm{op}} \to \mathsf{CMet}$ that lands in the subcategory of complete metric spaces and isometric embeddings.

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

$$q_S\big(x_{(-)}\big) = \{x_s \ : \ s \in S\} \in X_{|S|}. \tag{2.2.7}$$

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

In this way, we obtain a natural transformation $q : X^{(-)} \Rightarrow X_{|-|}$ between functors $\mathsf{FinUnif}^{\mathrm{op}} \to \mathsf{CMet}$.

Lemma 2.2.7. *Via q, the functor $X_{(-)} : \mathsf{N}^{\mathrm{op}} \to \mathsf{Met}$ is the left Kan extension of $X^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \to \mathsf{Met}$ along $|-|^{\mathrm{op}}$. Likewise with CMet in place of Met.*

Proof. Again because $\mathsf{CMet} \subseteq \mathsf{Met}$ is reflective, it is enough to prove this for Met. There it follows from the universal property of the quotient map q. We have in diagrams

$$\begin{tikzcd} \mathsf{FinUnif}^{\mathrm{op}} \arrow[rr, "X^{(-)}"] \arrow[dr, "|-|"'] & \arrow[d, phantom, "\Downarrow q"] & \mathsf{Met} \\ & \mathsf{N}^{\mathrm{op}} \arrow[ur, "X_{(-)}"'] & \end{tikzcd}$$

Consider now another functor K and natural transformation α as in

$$\begin{tikzcd} \mathsf{FinUnif}^{\mathrm{op}} \arrow[rr, "X^{(-)}"] \arrow[dr, "|-|"'] & \arrow[d, phantom, "\Downarrow \alpha"] & \mathsf{Met} \\ & \mathsf{N}^{\mathrm{op}} \arrow[ur, "K"'] & \end{tikzcd}$$

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

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

and we need to find a factorization

$$\begin{tikzcd} X^S \arrow[r, "q"] \arrow[dr, "\alpha_S"'] & X_{|S|} \arrow[d, dashed, "u_{|S|}"] \\ & K(|S|) \end{tikzcd} \tag{2.2.8}$$

for some $u : X_{(-)} \Rightarrow K$. By naturality of α with respect to automorphisms $\sigma : S \to S$, we know that α_S is invariant under precomposing by X^σ. Therefore it factors uniquely across 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 $m, n \in \mathsf{N}$, the diagram

$$\begin{tikzcd} X_m \arrow[r, "X_{m|mn}"] \arrow[d, "u_m"] & X_{mn} \arrow[d, "u_{mn}"] \\ K(m) \arrow[r, "K(m|mn)"] & K(mn) \end{tikzcd}$$

commutes. This follows from the fact that $|-| : \mathsf{FinUnif} \to \mathsf{N}$ is also full, so that the morphism $X_{m|mn}$ is the image of some morphism in $\mathsf{FinUnif}$, together with naturality of α and the definition (2.2.8). ∎

In effect, this argument is very similar to using the coend formula for pointwise Kan extensions, which however does not exactly apply since Met is not cocomplete (is missing coproducts).

It is also not hard to see that if X is complete, then so is every X^S. And since $\mathsf{CMet} \subseteq \mathsf{Met}$ is a reflective subcategory, the same applies to all X_n. Thus we also have endofunctors $(-)^S : \mathsf{CMet} \to \mathsf{CMet}$ and $(-)_n : \mathsf{CMet} \to \mathsf{CMet}$.

2.2.2. Empirical distributions

Definition 2.2.8. *Let $X \in \mathsf{Met}$. For $S \in \mathsf{FinUnif}$, the* empirical distribution *is the map $i^S : X^S \to PX$ which assigns to each S-indexed family $x_{(-)} \in X^S$ the uniform probability measure,*

$$i^S(x_{(-)}) := \frac{1}{|S|} \sum_{s \in S} \delta(x_s). \tag{2.2.9}$$

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

Definition 2.2.9. *For $n \in \mathsf{N}$, the* symmetric empirical distribution *is the map $i_n : X_n \to PX$ given by assigning to each multiset $\{x_1, \dots, x_n\} \in X_n$ the corresponding uniform probability measure,*

$$i_n(\{x_1 \dots x_n\}) := \frac{\delta(x_1) + \dots + \delta(x_n)}{n}. \tag{2.2.10}$$

The empirical distribution has less information than the original sequence. However, the only information lost is precisely the ordering, as the following proposition shows:

Proposition 2.2.10. *$i_n : X_n \to PX$ is an isometric embedding for each X and n.*

Proof. For $\{x_i\}, \{y_i\} \in X_n$, let $N_{xy} := \{1_x, \dots, n_x\} \amalg \{1_y, \dots, n_y\}$ be a finite pseudometric space with distances such that the canonical map $N_{xy} \to X$ is an isometric embedding, which means in particular that $d(i_x, j_y) = d_X(x_i, y_j)$. In the commutative square

$$\begin{tikzcd} N_{xy,n} \arrow[r, "i_n"] \arrow[d] & PN_{xy} \arrow[d] \\ X_n \arrow[r, "i_n"] & PX \end{tikzcd}$$

both vertical arrows are isometric embeddings by Lemmas 2.2.2 and 2.1.13. 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 (2.2.3),

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

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 (2.1.9). □

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

Lemma 2.2.11. *Let $n, m \in \mathsf{N}$, and $X \in \mathsf{CMet}$. Then the following diagram commutes:*

$$\begin{tikzcd} X_m \arrow[rr, "X_{m|mn}"] \arrow[dr, "i_m"'] & & X_{mn} \arrow[dl, "i_{mn}"] \\ & PX & \end{tikzcd} \tag{2.2.11}$$

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

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

Therefore the symmetric empirical distribution i_n is natural in n. It follows that the empirical distribution i^S is natural in S.

2.2.3. Universal property

Definition 2.2.12. *Let X be a complete metric space, and consider the symmetric empirical distribution embeddings $i_n : X_n \to PX$ for each $n \in \mathsf{N}$. We write $I(X)$ for the union of their images,*

$$I(X) := \bigcup_{n \in \mathsf{N}} i_n(X_n) \subseteq PX\ . \tag{2.2.12}$$

Lemma 2.2.13. *$I(X)$ is the colimit of the functor $X_{(-)} : \mathsf{N}^{\mathrm{op}} \to \mathsf{Met}$, and also of the functor $X^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \to \mathsf{Met}$, with the i_n and the i^S forming the universal cocones.*

Proof. By Lemma 2.2.7, it is enough to prove this for $X_{(-)}$. So let the $\{g_n : X_n \to Y\}$ form a cocone, i.e. a family of short maps such that $g_m = g_{mn} X_{m|mn}$. Since the $i_n : X_n \to I(X)$ are jointly epic by definition of $I(X)$, there can be at most one map $I(X) \to Y$ that is a morphism of cocones. Concerning existence, every point of $I(X)$ is of the form $i_n(\{x_i\})$ 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 for the following reason: if $i_n(\{x_i\}) = i_m(\{x'_j\})$, 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 $i_n : X_n \to I(X)$ 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.

Theorem 2.2.14. *Let X be a metric space. Then $I(X)$ is dense in PX.*

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

Lemma 2.2.15. *Let X be a compact metric space. Then $I(X)$ is dense in PX.*

Proof. We will show that arbitrary finite supported probability measures are dense in PX; this is enough since each of these is a convex combination of δ's, and we land in $I(X)$ by choosing rational approximations for the coefficients of such a convex combination.

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. Considering 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 sequence 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

$$p_\varepsilon := \sum_{i=1}^{k} p(A_i)\, \delta(y_i)\ . \tag{2.2.13}$$

is finitely supported. In order to witness that it is close to p, we choose a convenient joint,

$$m := \sum_{i=1}^{k} p|_{A_i} \otimes \delta(y_i), \tag{2.2.14}$$

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

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

as was to be shown. $\square$

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

Lemma 2.2.16. *Let $p, q_1, q_2 \in PX$ and $\lambda \in [0,1]$. Then*

$$d_{PX}\big(\lambda q_1 + (1-\lambda)p,\ \lambda q_2 + (1-\lambda)p\big) = \lambda\, d_{PX}(q_1, q_2). \tag{2.2.15}$$

Proof. This follows immediately from the duality (2.1.10), 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

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

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

Lemma 2.2.17. *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. With $B(x_0, \rho)$ the closed ball of radius $\rho > 0$, 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

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

By decomposing

$$p = p|_{B(x_0,\rho)} + p|_{X \backslash B(x_0,\rho)} \tag{2.2.16}$$

we can compute

$$d_{PX}(p, p') \overset{(2.2.15)}{=} p(X \backslash B(x_0,\rho))\, d_{PX}\left(\delta(x_0), \frac{p|_{X \backslash B(x_0,\rho)}}{p(X \setminus B(x_0,\rho))}\right)$$

$$\overset{(2.1.11)}{=} \int_{X \setminus B(x_0,\rho)} d(x, x_0)\, dp(x)$$

$$= \int_X d(x, x_0)\, dp(x) - \int_{B(x_0,\rho)} d(x, x_0)\, dp(x).$$

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

$$f_\rho(x) := \begin{cases} d(x, x_0) & \text{if } d(x, x_0) \leq \rho, \\ 0 & \text{otherwise.} \end{cases} \tag{2.2.17}$$

which converges pointwise to $d(-, x_0)$ as $\rho \to \infty$. By monotone convergence, this term therefore 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 $d_{PX}(p, p') \to 0$ as $\rho \to \infty$, and the approximating measure p' is boundedly supported.

For the second part, we therefore assume that $\mathrm{diam}(X) < \infty$. 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 using

$$p' := p|_K + p(X \setminus K)\, \delta(x_0), \tag{2.2.18}$$

By decomposing

$$p = p|_K + p|_{X \setminus K}, \tag{2.2.19}$$

we can compute

$$d_{PX}(p, p') \overset{(2.2.15)}{=} p(X \setminus K)\, d\left(\frac{p_{X \setminus K}}{p(K)}, \delta(x_0) \right) \overset{(2.1.11)}{=} p(X \setminus K)\, \mathrm{diam}(X),$$

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

Theorem 2.2.14 then follows as a corollary.

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

Theorem 2.2.18. *The space PX is the colimit of the functor $X_{(-)} : \mathsf{N}^{\mathrm{op}} \to \mathsf{CMet}$, and also of the functor $X^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \to \mathsf{CMet}$.*

Proof. Use Lemma 2.2.13 together with the previous Theorem 2.2.14, 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. □

Remark 2.2.19. This result relies again crucially on the choice of morphisms, i.e. the short maps. For continuous maps, in particular, the above does not hold: a continuous map defined on a dense subset does not always extend to a continuous map on the completion.

Since colimits over $\mathsf{FC}^{\mathrm{op}}$ or N^{op} in a category of functors into Met or CMet are computed pointwise[6], this implies that the Wasserstein space construction in the form of the object $P \in [\mathsf{CMet}, \mathsf{CMet}]$, is the colimit of the power functor construction:

Corollary 2.2.20. *The empirical distributions are the universal arrows:*

(a) Consider the functor $(=)_{(-)} : \mathsf{N}^{\mathrm{op}} \to [\mathsf{CMet}, \mathsf{CMet}]$ *mapping* $n \in \mathsf{N}$ *to the symmetrized power functor* $X \mapsto X_n$*. Then* $P \in [\mathsf{CMet}, \mathsf{CMet}]$ *is the colimit of* $(=)_{(-)}$*, with universal cocone given by the symmetric empirical distributions* $i_n : (-)_n \Rightarrow P$*.*

(b) Consider the functor $(=)^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \to [\mathsf{CMet}, \mathsf{CMet}]$ *mapping* $S \in \mathsf{FinUnif}$ *to the power functor* $X \mapsto X^S$*. Then* $P \in [\mathsf{CMet}, \mathsf{CMet}]$ *is the colimit of* $(=)^{(-)}$*, with universal cocone given by the empirical distributions* $i^S : (-)^S \Rightarrow P$*.*

Remark 2.2.21. Unfortunately, there is a small size issue here: since CMet is not equivalent to a small category—for example because there are complete metric spaces of arbitrary cardinality—the endofunctor category $[\mathsf{CMet}, \mathsf{CMet}]$ is not even locally small. One can fix this by uncurrying, using $(=)_{(-)} : \mathsf{N}^{\mathrm{op}} \times \mathsf{CMet} \to \mathsf{CMet}$ and $(=)^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \times \mathsf{CMet} \to \mathsf{CMet}$, as in the theory of graded monads developed in [FKM16].

2.3. Monad structure

The main result of this section is that the functor P is part of a monad, with unit and composition defined in a way analogous to the Giry monad [Gir82]. It was proven in [vB05] that the restriction of P on compact metric spaces carries a monad structure. In the spirit of categorical probability theory (see Section 1.1), the composition map E is given by integrating a measure on measures to a measure, and the unit δ by assigning to each points its Dirac measure.

An appealing feature of our Kantorovich functor is that its monad structure can be constructed directly from the colimit characterization in terms of the power functors defined in Section 2.2. In particular, the integration map E is obtained uniquely by a

[6] Technically, this relies on the fact that such limits always exists in Met and CMet. For the latter, this follows from the former and $\mathsf{CMet} \subseteq \mathsf{Met}$ being a reflective subcategory.

universal property, without the need to define it in terms of integrals or measure theory. In some sense, the universal property makes the integration map *inevitable*, coming directly from the characterization of P in terms of finite powers.

More technically, we use the fact that the power functors carry the structure of a monad *graded* by $\mathsf{FinUnif}^{\mathrm{op}}$, in the sense of a lax monoidal functor[7] into the endofunctor category $[\mathsf{CMet}, \mathsf{CMet}]$, and similarly for the symmetrized power functors in terms of N^{op}.

2.3.1. The power functors form a graded monad

As we will see next, the functor $(=)^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \to [\mathsf{CMet}, \mathsf{CMet}]$ has a canonical strong monoidal structure with respect to the monoidal structure on $\mathsf{FinUnif}$ inherited by the cartesian product of sets. We assume the monoidal structure to be strict for notational convenience.

Concerning the unit, there is a canonical transformation $\delta : 1_{\mathsf{CMet}} \Rightarrow (=)^1$ with components given by the identity isomorphisms $X \cong X^1$. For the composition, we use the currying maps $E^{S,T} : (X^S)^T \cong X^{S\times T}$. So 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 $\{x_{ij}\}_{i\in S,\, j\in T}$. Intuitively, an element of $(X^S)^T$ is a "double list", or "matrix", and from it we can canonically obtain a "list" or "vector" of length $|S \times T|$, i.e. an element of $X^{S\times T}$ by "flattening". A straightforward computation shows that $E^{S,T}$ indeed preserves distances, since distances add up across all components i and j in get rescaled by $|S| \cdot |T|$ in both cases. It is also clear that $E^{S,T}$ is natural in X. This $E^{S,T}$ is the map that, once we take the colimit, will become the integration map E.

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.

Theorem 2.3.1. *The above structure transformations δ and $E^{-,-}$ equip the functor $(=)^{(-)}$ with a strong monoidal structure, meaning that the following diagrams commute for all $X \in \mathsf{CMet}$:*

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

- *The unit triangles*

$$
\begin{tikzcd}
X^S \arrow[r, "\delta"] \arrow[dr, equal] & (X^S)^1 \arrow[d, "E^{S,1}"] \\
& X^{S\times 1}
\end{tikzcd}
\qquad
\begin{tikzcd}
X^S \arrow[r, "\delta^S"] \arrow[dr, equal] & (X^1)^S \arrow[d, "E^{1,S}"] \\
& X^{1\times S}
\end{tikzcd}
\tag{2.3.1}
$$

- *The associativity square*

$$
\begin{tikzcd}
((X^R)^S)^T \arrow[r, "E^{S,T}"] \arrow[d, "(E^{R,S})^T"] & (X^R)^{S\times T} \arrow[d, "E^{R,S\times T}"] \\
(X^{R\times S})^T \arrow[r, "E^{R\times S,T}"] & X^{R\times S\times T}
\end{tikzcd}
\tag{2.3.2}
$$

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.

2.3.2. 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 $q_n : X^n \to X_n$ is the universal map which coequalizes the action of the symmetric group S_n by 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

$$
\begin{tikzcd}
(X^m)^n \arrow[r, "q_n"] \arrow[d, "(q_m)^n"] & (X^m)_n \arrow[d, "(q_m)_n"] \\
(X_m)^n \arrow[r, "q_n"] & (X_m)_n
\end{tikzcd}
\tag{2.3.3}
$$

which commutes by naturality of q_n. The left arrow has a universal property as well: $(q_m)^n$ 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. This is because $(X^m)^n$ and $(X_m)^n$ are rescalings of the monoidal powers $(X^m)^{\otimes n}$ and $(X_m)^{\otimes n}$, and the monoidal tensor preserves colimits. It then 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 $E_{m,n} : (X_m)_n \to X_{mn}$ by the universal property of the $S_m \wr S_n$-quotient map $(X^m)^n \to (X_m)_n$ as the unique morphism which makes

$$\begin{tikzcd} (X^m)^n \arrow[r, "E^{m,n}"] \arrow[d] & X^{mn} \arrow[d, "q_{mn}"] \\ (X_m)_n \arrow[r, "E_{m,n}"] & X_{mn} \end{tikzcd} \tag{2.3.4}$$

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 no longer 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.

Theorem 2.3.2. *The above structure transformations δ and $E_{-,-}$ equip the functor $(=)_{(-)}$ with a lax monoidal structure, meaning that the following diagrams commute for all $X \in \mathsf{CMet}$:*

- *The unit triangles*

$$\begin{tikzcd} X_m \arrow[r, "\delta"] \arrow[dr, equal] & (X_m)_1 \arrow[d, "E_{m,1}"] \\ & X_{m\times 1} \end{tikzcd} \qquad \begin{tikzcd} X_m \arrow[r, "\delta_m"] \arrow[dr, equal] & (X_1)_m \arrow[d, "E_{1,m}"] \\ & X_{1\times m} \end{tikzcd} \tag{2.3.5}$$

- *The associativity square*

$$\begin{tikzcd} ((X_\ell)_m)_n \arrow[r, "E_{m,n}"] \arrow[d, "(E_{\ell,m})_n"] & (X_\ell)_{mn} \arrow[d, "E_{\ell,mn}"] \\ (X_{\ell m})_n \arrow[r, "E_{\ell m,n}"] & X_{\ell mn} \end{tikzcd} \tag{2.3.6}$$

Proof. We reduce this to Theorem 2.3.1. Only the associativity square is nontrivial.

By reasoning similar to (2.3.3), composing the quotient maps results in a unique epimorphism $((X^\ell)^m)^n \to ((X_\ell)_m)_n$. In fact, we get a cube:

$$\begin{tikzcd}
& ((X^\ell)^m)_n \arrow[rr, "(q_m)_n"] \arrow[dd, "((q_l)^m)_n"] & & ((X^\ell)_m)_n \arrow[dd, "((q_l)_m)_n"] \\
((X^\ell)^m)^n \arrow[ur, "q_n"] \arrow[rr, "(q_m)^n"] \arrow[dd, "((q_l)^m)^n"] & & ((X^\ell)_m)^n \arrow[ur, "q_n"] \arrow[dd, "((q_l)_m)^n"] & \\
& ((X_\ell)^m)_n \arrow[rr, "(q_m)_n"] & & ((X_\ell)_m)_n \\
((X_\ell)^m)^n \arrow[ur, "q_n"] \arrow[rr, "(q_m)^n"] & & ((X_\ell)_m)^n \arrow[ur, "q_n"] &
\end{tikzcd} \tag{2.3.7}$$

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

$$\begin{tikzcd}
& ((X_\ell)_m)_n \arrow[rr, "E_{m,n}"] \arrow[dd, "(E_{\ell,m})_n"] & & (X_\ell)_{mn} \arrow[dd, "E_{\ell,mn}"] \\
((X^\ell)^m)^n \arrow[ur] \arrow[rr, "E^{m,n}"] \arrow[dd, "(E^{\ell,m})^n"] & & (X^\ell)^{mn} \arrow[ur] \arrow[dd, "E^{\ell,mn}"] & \\
& (X_{\ell m})_n \arrow[rr, "E_{\ell m,n}"] & & X_{\ell mn} \\
(X^{\ell m})^n \arrow[ur] \arrow[rr, "E^{\ell m,n}"] & & X^{\ell mn} \arrow[ur] &
\end{tikzcd} \tag{2.3.8}$$

where the unlabeled diagonal arrows are the quotient maps discussed previously, and we need to show that the back face commutes. The bottom and right faces commute by (2.3.4). The top face also commutes, thanks to

$$\begin{tikzcd}
((X^\ell)^m)^n \arrow[r, "E^{m,n}"] \arrow[d] & (X^\ell)^{mn} \arrow[d] \\
((X_\ell)^m)^n \arrow[r, "E^{m,n}"] \arrow[d] & (X_\ell)^{mn} \arrow[d] \\
((X_\ell)_m)_n \arrow[r, "E_{m,n}"] & ((X_\ell)_m)_n
\end{tikzcd}$$

and similarly for the left face. Finally, commutativity of the front face is by Theorem 2.3.1. Therefore since $((X^\ell)^m)^n \to ((X_\ell)_m)_n$ is epi, this implies that the back face commutes as well. $\square$

We can also consider the N^{op}-graded monad $(=)_{(-)}$ as the universal N-graded monad that one obtains from the $\mathsf{FinUnif}^{\mathrm{op}}$-graded monad $(=)^{(-)}$ by change of grading along $\mathsf{FinUnif}^{\mathrm{op}} \to \mathsf{N}^{\mathrm{op}}$. In fact, this follows by the things that we have proven so far:

Theorem 2.3.3. *Let* MonCat *be the bicategory of monoidal categories, lax monoidal functors, and monoidal transformations. Then the lax monoidal functor* $(=)_{(-)} : \mathsf{N}^{\mathrm{op}} \to [\mathsf{CMet}, \mathsf{CMet}]$ *is the left Kan extension in* MonCat *of* $(=)^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \to [\mathsf{CMet}, \mathsf{CMet}]$ *along* $\mathsf{FinUnif}^{\mathrm{op}} \to \mathsf{N}^{\mathrm{op}}$.

Proof. By Lemma 2.2.7, this Kan extension works in Cat, and it is clear that $\mathsf{FinUnif}^{\mathrm{op}} \to \mathsf{N}^{\mathrm{op}}$ is strong monoidal and essentially surjective. In order to apply Theorem A.2.1, it remains to check two things: first, that the transformation $q : (=)^{(-)} \to (=)_{(-)}$ is monoidal, which boils down to the diagram

$$\begin{tikzcd} (X^m)^n \arrow[r, "E^{m,n}"] \arrow[d, "q \circ q"] & X^{mn} \arrow[d, "q"] \\ (X_m)_n \arrow[r, "E_{m,n}"] & X_{mn} \end{tikzcd}$$

which is (2.3.4) again. Second, that $q \otimes q$ is an epimorphism in the functor category $[\mathsf{FinUnif}^{\mathrm{op}} \times \mathsf{FinUnif}^{\mathrm{op}}, [\mathsf{CMet}, \mathsf{CMet}]]$, which follows from the fact that even every individual double quotient map $(X^m)^n \to (X_m)_n$ is an epimorphism. □

2.3.3. The monad structure on the Kantorovich functor

Now that we have shifted the graded monad structure from $\mathsf{FinUnif}^{\mathrm{op}}$ to N^{op}, we shift it one step further and crush it down to a lax monoidal functor $1 \to [\mathsf{CMet}, \mathsf{CMet}]$, i.e. to an ungraded monad on CMet whose underlying functor is P.

We define the unit and composition maps in terms of the power functors and the empirical distributions.

Definition 2.3.4. *For* $X \in \mathsf{CMet}$ *and* $n \in N$*, The* Dirac delta embedding *is the composite*

$$X \xrightarrow{\delta} X^1 \xrightarrow{i_1} PX,$$

which we also denote by δ.

Proposition 2.2.10 implies that δ is an isometric embedding. As a composite of natural transformation, we also have naturality $\delta : \mathbb{1} \Rightarrow P$. Before getting to the composition, we need another bit of preparation. A *sifted category* is a category S such that S-indexed colimits in Set commutes with finite products. In this sense, they generalize directed and filtered categories. N^{op} is trivially sifted thanks to being directed. However, the category $\mathsf{FinUnif}^{\mathrm{op}}$ itself is not sifted, for example since the spans

$$\begin{tikzcd} & S \arrow[dl, equal] \arrow[dr, equal] & \\ S & & S \end{tikzcd} \qquad \begin{tikzcd} & S \arrow[dl, equal] \arrow[dr, "\alpha"] & \\ S & & S \end{tikzcd}$$

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

Lemma 2.3.5. *Both the power functors* $(-)^S$ *and the symmetric power functors* $(-)_n$ *preserve sifted colimits.*

Proof. Let D be the sifted category indexing the colimits under consideration.

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

Similarly to the quotient maps $(X^m)^n \to (X_m)_n$ in (2.3.3), we have a commutative square

$$\begin{tikzcd} (X_m)_n \arrow[r, "(i_m)_n"] \arrow[d, "i_n"] & (PX)_n \arrow[d, "i_n"] \\ P(X_m) \arrow[r, "Pi_m"] & PPX \end{tikzcd} \tag{2.3.9}$$

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

Proposition 2.3.6. *PPX is the colimit of both*

(a) the $(X_m)_n$ with universal cocone given by the $i_n \circ (i_m)_n = Pi_m \circ i_n$ for $m, n \in \mathsf{N}^{\mathrm{op}}$;

(b) the subdiagram of this formed by the $(X_n)_n$ for $n \in \mathsf{N}^{\mathrm{op}}$.

While measures on spaces of measures are often quite delicate to handle, this results gives a concrete way to work with them in terms of finite data only. Although we do not 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 N^{op} is sifted. For the first, the lemma tells us that the $(i_m)_n : (X_m)_n \to (PX)_n$ form a colimiting cocone for each n; the claim then follows from the construction of a colimit over $\mathsf{N}^{\mathrm{op}} \times \mathsf{N}^{\mathrm{op}}$ by first taking the colimit over the first factor and then over the second. $\square$

Lemma 2.3.7. *For $X \in \mathsf{CMet}$, there is a unique morphism $E : PPX \to PX$ such that*

$$\begin{tikzcd} (X_m)_n \arrow[r, "E_{m,n}"] \arrow[d] & X_{mn} \arrow[d] \\ PPX \arrow[r, "E"] & PX \end{tikzcd} \tag{2.3.10}$$

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

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

Proof. This amounts to showing that the $i_{mn} \circ E_{m,n}$ form a cocone to which the universal property of Proposition 2.3.6 applies. Since every morphism in N is a divisibility relation, this corresponds to commutativity of the two diagrams

$$\begin{tikzcd} (X_m)_n \arrow[rr, "(X_{m|\ell m})_n"] \arrow[d, "E_{m,n}"'] & & (X_{\ell m})_n \arrow[d, "E_{\ell m,n}"] \\ X_{mn} \arrow[rr, "X_{mn|\ell mn}"] \arrow[dr, "i_{mn}"'] & & X_{\ell mn} \arrow[dl, "i_{\ell mn}"] \\ & PX & \end{tikzcd} \qquad \begin{tikzcd} (X_m)_n \arrow[rr, "(X_m)_{n|\ell n}"] \arrow[d, "E_{m,n}"'] & & (X_m)_{\ell n} \arrow[d, "E_{m,\ell n}"] \\ X_{mn} \arrow[rr, "X_{mn|\ell mn}"] \arrow[dr, "i_{mn}"'] & & X_{\ell mn} \arrow[dl, "i_{\ell mn}"] \\ & PX & \end{tikzcd}$$

for every $\ell \in \mathsf{N}$. The upper squares commute by naturality of E in its two arguments in N, and the triangles by Lemma 2.2.11. $\square$

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

Let's show why this 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 $f : X \to \mathbb{R}$ into

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

This map makes diagram (2.3.10), since for all $\{\{x_{11}, \dots, x_{m1}\}, \dots, \{x_{1n}, \dots, x_{mn}\}\}$ in $(X^M)^N$, by linearity of the integral:

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

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

2.3.4. 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 2.3.2.

Theorem 2.3.8. *(P, δ, E) is a monad on* CMet*. In other words, we have commutative diagrams:*

$$\begin{tikzcd} P \arrow[r, Rightarrow, "P\delta"] \arrow[dr, equal] & PP \arrow[d, Rightarrow, "E"] & P \arrow[l, Rightarrow, "\delta P"'] \arrow[dl, equal] \\ & P & \end{tikzcd} \tag{2.3.11}$$

and:

$$\begin{tikzcd} PPP \arrow[r, Rightarrow, "PE"] \arrow[d, Rightarrow, "EP"] & PP \arrow[d, Rightarrow, "E"] \\ PP \arrow[r, Rightarrow, "E"] & P \end{tikzcd} \tag{2.3.12}$$

We call then 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 $X \in \mathsf{CMet}$. Because of the universal property of P, E_n, E and i, we have the following.

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

$$\begin{tikzcd} & PX \arrow[rr, "\delta"] \arrow[drr, equal] & & PPX \arrow[d, "E"] \\ X_m \arrow[ur] \arrow[r, "(X_m)_{1|n} \circ \delta"] \arrow[dr, "X_{m|mn}"'] & (X_m)_n \arrow[urr] \arrow[d, "E_{m,n}"] & & PX \\ & X_{mn} \arrow[urr] & & \end{tikzcd} \tag{2.3.13}$$

Now:

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

$$\begin{tikzcd} X_m \arrow[r, "\delta"] \arrow[dr, equal] & (X_m)_1 \arrow[r, "(X_m)_{1|n}"] \arrow[d, "E_{m,1}"] & (X_m)_n \arrow[d, "E_{m,n}"] \\ & X_m \arrow[r, "X_{m|mn}"'] & X_{mn} \end{tikzcd} \tag{2.3.14}$$

  which commutes by the left unit diagram of Theorem 2.3.2, together with naturality of $E_{m,-}$;

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

$$\begin{tikzcd}
X_m \arrow[r, "\delta"] \arrow[d, "i_m"] & (X_m)_1 \arrow[r, "(X_m)_{1|n}"] \arrow[d, "(i_m)_1"] & (X_m)_n \arrow[d, "(i_m)_n"] \\
PX \arrow[r, "\delta"'] & (PX)_1 \arrow[r, "(PX)_{1|n}"'] & (PX)_n \arrow[r, "i_n"'] & PPX
\end{tikzcd} \tag{2.3.15}$$

  which commutes by naturality of δ and $(-)_{1|n}$;
- The right face commutes by Lemma 2.3.7;
- The left bottom face commutes by the naturality of the empirical distribution.

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:

$$\begin{tikzcd}
& PX \arrow[r, "P\delta"] \arrow[dr, equal] & PPX \arrow[d, "E"] \\
X_n \arrow[ur] \arrow[r, "(X_{1|m}\circ\delta)_n"] \arrow[dr, "X_{m|mn}"'] & (X_m)_n \arrow[ur] \arrow[d, "E_{m,n}"] & PX \\
& X_{mn} \arrow[ur] &
\end{tikzcd} \tag{2.3.16}$$

Now:

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

$$\begin{tikzcd}
X_n \arrow[r, "(\delta)_n"] \arrow[dr, equal] & (X_1)_n \arrow[r, "(X_{1|m})_n"] \arrow[d, "E_{1,n}"] & (X_m)_n \arrow[d, "E_{m,n}"] \\
& X_n \arrow[r, "X_{n|mn}"'] & X_{mn}
\end{tikzcd} \tag{2.3.17}$$

  which commutes by the right unit diagram of Theorem 2.3.2, together with naturality of $E_{-,n}$;
- The top face can be decomposed as the following diagram:

$$\begin{tikzcd}
X_n \arrow[r, "(\delta)_n"] \arrow[d, "i_n"] & (X_1)_n \arrow[r, "(X_{1|m})_n"] \arrow[d, "i_n"] & (X_m)_n \arrow[d, "i_n"] \\
PX \arrow[r, "P\delta"'] & P(X_1) \arrow[r, "P(X_{1|m})"'] & P(X_m) \arrow[r, "Pi_m"'] & PPX
\end{tikzcd} \tag{2.3.18}$$

  which commutes by naturality of i_n;

- The right face commutes again by Lemma 2.3.7;
- The left bottom face commutes again by the naturality of the empirical distribution.

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:

$$\begin{tikzcd}
& PPPX \arrow[rr, "PE"] \arrow[dd, "E"] & & PPX \arrow[dd, "E"] \\
((X_\ell)_m)_n \arrow[ur] \arrow[rr, "(E_{\ell,m})_n"] \arrow[dd, "E_{m,n}"] & & (X_{\ell m})_n \arrow[ur] \arrow[dd, "E_{\ell m,n}"] & \\
& PPX \arrow[rr, "E"] & & PX \\
(X_\ell)_{mn} \arrow[ur] \arrow[rr, "E_{\ell,mn}"] & & X_{\ell mn} \arrow[ur] &
\end{tikzcd} \tag{2.3.19}$$

where the map $((X_\ell)_m)_n \to PPPX$ is uniquely obtained in the same way as the map $((X^\ell)^m)^n \to ((X_\ell)_m)_n$ in the proof of Theorem 2.3.2, using naturality of i instead of q. Now:

- The front face is just the associativity square of Theorem 2.3.2;
- The top face can be decomposed as:

$$\begin{tikzcd}
((X_\ell)_m)_n \arrow[r] \arrow[d, "(E_{\ell,m})_n"] & (PPX)_n \arrow[r, "i_n"] \arrow[d, "(E)_n"] & PPPX \arrow[d, "PE"] \\
(X_{\ell m})_n \arrow[r] & (PX)_n \arrow[r, "i_n"] & PPX
\end{tikzcd} \tag{2.3.20}$$

  which commutes by Lemma 2.3.7, and by naturality of i_n;
- The left, right, and bottom faces commute by Lemma 2.3.7.

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. $\square$

In analogy with Theorem 2.3.3, we can now conclude that P as a monad is exactly what one obtains upon taking the $\mathsf{FinUnif}^{\mathrm{op}}$-graded monad $(=)^{(-)}$ or the N^{op}-graded monad $(=)_{(-)}$ and "crushing them down" universally to an ungraded monad:

Theorem 2.3.9. *As a lax monoidal functor,* $P : 1 \to [\mathsf{CMet}, \mathsf{CMet}]$ *is the Kan extension in* MonCat

(a) of $(=)^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \to [\mathsf{CMet}, \mathsf{CMet}]$ *along* $! : \mathsf{FinUnif}^{\mathrm{op}} \to 1$*, and*

(b) of $(=)_{(-)} : \mathsf{N}^{\mathrm{op}} \to [\mathsf{CMet}, \mathsf{CMet}]$ *along* $! : \mathsf{N}^{\mathrm{op}} \to 1$*,*

with respect to the empirical distributions as the universal transformation.

Together with Corollary 2.2.20 and Theorem 2.3.3, this means that we have a diagram

$$\begin{tikzcd}
\mathsf{FinUnif}^{\mathrm{op}} \arrow[d, "|-|"'] \arrow[drr, "(=)^{(-)}"] \arrow[dr, phantom, "\Leftarrow q"] & & \\
\mathsf{N} \arrow[rr, "(=)_{(-)}"] \arrow[d, "!"'] \arrow[dr, phantom, "\Downarrow i_{(-)}"] & & {[\mathsf{CMet}, \mathsf{CMet}]} \\
1 \arrow[urr, "P"'] & &
\end{tikzcd}$$

in which all 2-cells are Kan extensions, both in Cat and in MonCat.

Proof. By composition of Kan extensions and Theorem 2.3.3, it is enough to prove the second item. In order to apply Theorem A.2.1, it remains to check two things: first, that the transformation $i_{(-)} : (=)_{(-)} \Rightarrow P$ is monoidal, which boils down to the diagram

$$\begin{tikzcd}
(X_m)_n \arrow[r, "E_{m,n}"] \arrow[d, "i_n \circ (i_m)_n"'] & X_{mn} \arrow[d, "i_{mn}"] \\
PPX \arrow[r, "E"] & PX
\end{tikzcd}$$

which is (2.3.10) again. Second, that $i \otimes i$ is an epimorphism in the functor category $[\mathsf{N}^{\mathrm{op}} \times \mathsf{N}^{\mathrm{op}}, [\mathsf{CMet}, \mathsf{CMet}]]$, which follows from the fact that for every X, the maps $(X_m)_n \to PPX$ are jointly epic. $\square$

Moreover, the uniqueness of the monoidal Kan extension A.2.1 implies that the monad structure on P is the only one which makes the empirical distribution maps into a morphism of graded monads.

2.4. Algebras

In this section we will study the algebras of the Kantorovich monad. Following the intuition of Section 1.1, P-algebras are spaces A which are closed under mixtures, or convex combinations, weighted by measures of PA.

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

$$\begin{tikzcd} A \arrow[r, "\delta"] \arrow[dr, equal] & PA \arrow[d, "e"] \\ & A \end{tikzcd} \qquad\qquad \begin{tikzcd} PPA \arrow[r, "Pe"] \arrow[d, "E"] & PA \arrow[d, "e"] \\ PA \arrow[r, "e"] & A \end{tikzcd}$$

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

$$\begin{tikzcd} PA \arrow[r, "Pf"] \arrow[d, "e_A"] & PB \arrow[d, "e_B"] \\ A \arrow[r, "f"] & B \end{tikzcd}$$

commutes. We also say that f is *P-affine*. The Eilenberg-Moore category 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 $e = E : PPX \to PX$. The Kleisli category CMet_P is the full subcategory of CMet^P on the free algebras. Its morphisms are the short maps $X \to PY$ for complete metric spaces X,Y, which correspond bijectively and naturally to P-affine maps $PX \to PY$, so that it naturally contains CMet as a subcategory (see Section 1.1).

As sketched in 1.1.1, the Kleisli morphisms should be thought of as *stochastic maps* or *Markov kernels*. An important difference between other approaches to categorical probability theory and the one developed by van Breugel [vB05] and now here 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 $f : X \to Y$, it is generally not possible to write p as the image of the pushforward $(Pf)(p)$ under a Kleisli morphism $PY \to PX$, because the resulting map may not be short. However, many cases in which one would classically like to use conditional expectations can be treated categorically using different approaches, see Section 2.6 and Chapter 4.

In this section, we will give equivalent characterizations of the P-algebras and their category. We will again exploit the colimit characterization, to show that algebras are equivalently spaces that are closed under *finite midpoints*. 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 similarly appealing 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 [Dob06].

2.4.1. Convex spaces

A set together with an abstract notion of convex combinations satisfying 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, as 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 composition by the formation of the expected measure,

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

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

Definition 2.4.1. *A* convex space *is a set A equipped with a family of binary operations $c : [0,1] \times A \times A \to A$, 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:*

$$\nu = \begin{cases} \frac{\lambda(1-\mu)}{1-\lambda\mu} & \text{if } \lambda,\mu \neq 1; \\ \text{any number in } [0,1] & \text{if } \lambda = \mu = 1. \end{cases} \tag{2.4.1}$$

The category of convex spaces has as morphisms those maps $f : A \to B$ *such that*

$$\begin{tikzcd} A \times A \arrow[r, "f \times f"] \arrow[d, "c_\lambda"] & B \times B \arrow[d, "c_\lambda"] \\ A \arrow[r, "f"] & B \end{tikzcd} \tag{2.4.2}$$

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

In the following, we freely make use of the equivalence between this definition of convex space and algebras of the convex combinations monad $C : \mathsf{Set} \to \mathsf{Set}$.

2.4.2. Equivalent characterizations of algebras

Theorem 2.4.2. *The following structures are equivalent on a complete metric space* A, *in the sense that there is an equivalence of categories over* CMet.

(a) A P*-algebra structure;*

(b) A short map $e_n : A_n \to A$ *for each* $n \in \mathsf{N}$, *such that* $e_1 = \delta^{-1}$, *and such that the diagrams*

$$\begin{tikzcd} A_m \arrow[rr, "A_{m|mn}"] \arrow[dr, "e_m"'] & & A_{mn} \arrow[dl, "e_{mn}"] \\ & A & \end{tikzcd} \qquad \begin{tikzcd} (A_m)_n \arrow[r, "(e_m)_n"] \arrow[d, "E_{m,n}"] & A_n \arrow[d, "e_n"] \\ A_{mn} \arrow[r, "e_{mn}"] & A \end{tikzcd} \tag{2.4.3}$$

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

$$\begin{tikzcd} A_n \arrow[r, "f_n"] \arrow[d, "e_n"] & B_n \arrow[d, "e_n"] \\ A \arrow[r, "f"] & B \end{tikzcd} \tag{2.4.4}$$

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

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

$$\begin{tikzcd} A^T \arrow[rr, "A^\phi"] \arrow[dr, "e^T"'] & & A^S \arrow[dl, "e^S"] \\ & A & \end{tikzcd} \qquad \begin{tikzcd} (A^S)^T \arrow[r, "(e^S)^T"] \arrow[d, "E^{S,T}"] & A^T \arrow[d, "e^T"] \\ A^{S\times T} \arrow[r, "e^{S\times T}"] & A \end{tikzcd} \tag{2.4.5}$$

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

$$\begin{tikzcd} A^S \arrow[r, "f^S"] \arrow[d, "e^S"] & B^S \arrow[d, "e^S"] \\ A \arrow[r, "f"] & B \end{tikzcd} \tag{2.4.6}$$

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

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

$$d\big(c_\lambda(x,z), c_\lambda(y,z)\big) \leq \lambda\, d(x,y), \tag{2.4.7}$$

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] also can be interpreted as establishing the equivalence between (a) and (d). Second, in (c) and (b), these structures differ from the *graded algebras* in the sense of [FKM16, Definition 1]: for a graded algebra, the algebra morphisms would have to be of type $(A_m)_n \to A_{mn}$ and $(A^S)^T \to A^{S\times T}$, respectively.

It will follow from Theorem 2.4.3 that in structures of type (d), the inequality (2.4.7) necessarily holds with equality.

Proof. We first apply the universal properties from before to show that the structures of type (a), (c) and (b) are equivalent, using the universal properties from before.

- (b)⇔(c): By composing with the quotient maps $A^S \to A_{|S|}$, the $(e_n)_{n\in\mathsf{N}}$ determine morphisms $e^S : A^S \to A$, and conversely by the universal property. The equivalence between the triangles in (2.4.3) and (2.4.5) follows from $e_{|S|} = e^S \circ q$ and the diagram (2.2.6). The equivalence of (2.4.4) and (2.4.6) is by the same reason.

  It remains to verify the equivalence of the squares in (2.4.3) and (2.4.5). This follows by a cube similar to (2.3.8),

$$
\begin{tikzcd}
& (A_m)_n \arrow[rr, "(e_m)_n"] \arrow[dd, "E_{m,n}"] & & A_n \arrow[dd, "e_n"] \\
(A^m)^n \arrow[ur] \arrow[rr, "(e^m)^n"] \arrow[dd, "E^{m,n}"] & & A^n \arrow[ur] \arrow[dd, "e^n"] & \\
& A_{mn} \arrow[rr, "e_{mn}"] & & A \\
A^{mn} \arrow[ur] \arrow[rr, "e^{mn}"] & & A \arrow[ur, equal] &
\end{tikzcd}
$$

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.

- (a)⇔(b): This works similarly. By the universal property of PA, the cocone defined by the first diagram in (2.4.3) is equivalent to a short map $e : PA \to A$. The equivalence between the square in (2.4.3) and the composition square of a P-algebra follow by considering the cube

$$
\begin{tikzcd}
& PPA \arrow[rr, "Pe"] \arrow[dd, "E"] & & PA \arrow[dd, "e"] \\
(A_m)_n \arrow[ur] \arrow[rr, "(e_m)_n"] \arrow[dd, "E_{m,n}"] & & A_n \arrow[ur] \arrow[dd, "e_n"] & \\
& PA \arrow[rr, "e"] & & A \\
A_{mn} \arrow[ur] \arrow[rr, "e_{mn}"] & & A \arrow[ur, equal] &
\end{tikzcd}
\tag{2.4.8}
$$

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

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

$$
\begin{tikzcd}
\mathsf{CMet} \arrow[r, "P"] \arrow[d, "U"] & \mathsf{CMet} \arrow[d, "U"] \\
\mathsf{Set} \arrow[r, "C"] \arrow[ur, Rightarrow, "\eta"] & \mathsf{Set}
\end{tikzcd}
$$

where U is the forgetful functor, and η is the natural transformation with $\eta : CUX \to UPX$ 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 CMet to C-algebras in Set. In other words, every P-algebra is a convex space in a canonical way.

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

$$\begin{aligned} & d\big(c_\lambda(x,z), c_\lambda(y,z)\big) \\ &\qquad = d(e(\lambda\delta(x) + (1-\lambda)\delta(z)), e(\lambda\delta(y) + (1-\lambda)\delta(z))) \\ &\qquad \le d(\lambda\delta(x) + (1-\lambda)\delta(z), \lambda\delta(y) + (1-\lambda)\delta(z)) \\ &\qquad \overset{\text{Lemma 2.2.16}}{=} \lambda\, d(\delta(x), \delta(y)) = \lambda\, d(x,y). \end{aligned}$$

- (d)⇒(c): Intuitively, the e^S correspond to taking convex combinations with equal weights, and commutativity of (2.4.5) follow from the equations satisfied by taking convex combinations in any convex space. To make this formally precise, it is most convenient to consider a convex space as a model of the Lawvere theory $\mathsf{FinStoch}^{\mathrm{op}}$. Considering $\mathsf{FinUnif}$ as a subcategory $\mathsf{FinUnif} \subseteq \mathsf{FinSet} \subseteq \mathsf{FinStoch}$, defining maps $u_S : 1 \to S$ in $\mathsf{FinStoch}$ which pick out the uniform distribution on each finite set S results in commutativity of the two diagrams

$$\begin{tikzcd} & 1 \arrow[dl, "u_S"'] \arrow[dr, "u_T"] & \\ S \arrow[rr, "\phi"] & & T \end{tikzcd} \qquad \begin{tikzcd} 1 \arrow[r, "u_{S\times T}"] \arrow[d, "u_T"] & S \times T \arrow[d, equal] \\ T \arrow[r, "u_{S\times T}"] & S \times T \end{tikzcd}$$

for every $S, T \in \mathsf{FinUnif}$ and $\phi \in \mathsf{FinUnif}(S,T)$. Thus given a convex space A as a model of $\mathsf{FinStoch}^{\mathrm{op}}$, the u_S become maps $e^S : A^S \to A$ satisfying the required equations, and every affine map between convex spaces will make (2.4.6) commute. What is not a priori clear is that the e^S are short; but this follows from (2.4.7), two applications of which give

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

which generalizes to

$$d\left(e\left(\sum_i \lambda_i \delta(x_i) \right), e\left(\sum_i \lambda_i \delta(y_i) \right) \right) \le \sum_i \lambda_i d(x_i, y_i) \tag{2.4.9}$$

by decomposing a general convex combination into a sequence of binary ones and using induction. Shortness of e^S is now 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)⇒(d)⇒(c), one recovers the underlying (c)-structure of A. To see that the composite functor given by (d)⇒(b)⇔(a)⇒(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 $\lambda \in (0,1)$, but this is surprisingly tricky. First, as λ varies, this distance is bounded; this is 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 use a sufficiently small rational $\varepsilon > 0$, as well as rational $\nu \in \left(\frac{\lambda-\varepsilon}{1-\varepsilon}, \frac{\lambda}{1-\varepsilon}\right)$, and put $z := c_\nu(x,y) = c'_\nu(x,y)$. Then

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

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 $\varepsilon \to 0$. □

It seems plausible that P-algebras also coincide with the *metric mean-value algebras* of [vBHMW05, Definition 6], when the requirement of 1-boundedness is dropped.

2.4.3. Algebras as closed convex subsets of Banach spaces

If E is a Banach space and $A \subseteq E$ is a closed convex subset, then A is a convex space which carries a metric

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

with respect to which it is complete. These two structures interact via the metric compatibility inequality (2.4.7),

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

which even holds with equality. Therefore by Theorem 2.4.2(d), A is a P-algebra $e : PA \to A$ in a canonical way. In particular, we can therefore *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 for us the expectation value of any Banach-space valued random variable probability measure with finite first moment on any other complete metric space.

So let $\mathsf{ConvBan}$ be the category whose objects are closed convex subsets of Banach spaces $A \subseteq E$, and whose morphisms $f : (A \subseteq E) \to (B \subseteq F)$ are the short affine maps $f : A \to B$.[8] We then have a canonical functor $\mathsf{ConvBan} \to \mathsf{CMet}^P$ 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 to a closed convex subset of a Banach space. We therefore obtain that P-algebras and closed convex subsets of Banach spaces are the same concept:

Theorem 2.4.3. *The functor* $\mathsf{ConvBan} \to \mathsf{CMet}^P$ *is an equivalence of categories.*

As a corollary, since to every monad there corresponds an adjunction, we have a Choquet-like adjunction for *possibly noncompact spaces* in the spirit of 1.1.3:

Corollary 2.4.4. *There is a natural bijection*

$$\mathsf{COMet}(X, A) \cong \mathsf{ConvBan}(PX, A). \tag{2.4.10}$$

In practice, this means the following. A short, monotone, map $f : X \to A$ from a complete metric space X to an convex space (P-algebra) A is uniquely determined by the affine extension it defines, as an affine map on probability measures

$$p \longmapsto \int_X f\, dp,$$

i.e. the P-morphism given by the composition

$$PX \xrightarrow{Pf} PA \xrightarrow{e} A, \tag{2.4.11}$$

and every affine map $PX \to A$ can be written in this form, as the affine extension of a map $f : X \to A$. Equivalently, any affine map $\tilde{f} : PX \to A$ is uniquely determined by its restriction on the extreme points of the simplex, i.e. the composition

$$X \xrightarrow{\delta} PX \xrightarrow{\tilde{f}} A, \tag{2.4.12}$$

[8] One might be tempted to define morphisms to be equivalence classes of short affine maps $f : E \to F$ 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 $A \to F$ can in general not be extended to a short (or even merely continuous) affine map $E \to F$.

of which it is the affine extension. As it is easily checked, the operations $f \mapsto e \circ (Pf)$ and $\tilde{f} \mapsto \tilde{f} \circ \delta$ are inverse to each other, forming the natural bijection (2.4.10).

The same can be said about Lipschitz maps with arbitrary constant (but not in general about just continuous functions, if the spaces are not bounded). Different variants of this result are known in the literature as *noncompact Choquet theory*, see for example [Win85, Chapter 1].

We will refer to this adjunction, and to analogous ones in similar categories, as the "Choquet adjunction".

2.5. Bimonoidal structure

We can now define product joints and marginals, which will equip P with a bimonoidal structure in the way sketched in Section 1.2 (and described more in detail in [FP18]).

Definition 2.5.1. *Let $p \in PX, q \in PY$. We denote $p \otimes q$ the joint probability measure on $X \otimes Y$ defined by:*

$$\int_{X \otimes Y} f(x,y)\, d(p \otimes q)(x,y) := \int_{X \otimes Y} f(x,y)\, dp(x)\, dq(y).$$

Let now $r \in P(X \otimes Y)$. We denote (r_X) the marginal probability on X defined by:

$$\int_X f(x)\, dr_X(x) := \int_{X \otimes Y} f(x)\, dr(x,y).$$

The marginal on Y is defined analogously.

It is straightforward to check that the functionals defined in Definition 2.5.1 are positive, linear, and Scott-continuous, therefore they specify uniquely Radon probability measures of finite first moment.

In the rest of this section we will show that the joints and marginals in Definition 2.5.1 equip the Kantorovich monad on CMet with a bimonoidal monad structure (Theorem 2.5.17).

2.5.1. Monoidal structure

Definition 2.5.2. *Let $X, Y \in \mathsf{CMet}$. We define the map $\nabla : PX \otimes PY \to P(X \otimes Y)$ as mapping $(p,q) \in PX \otimes PY$ to the joint $p \otimes q \in P(X \otimes Y)$.*

Proposition 2.5.3. $\nabla : PX \otimes PY \to P(X \otimes Y)$ *is short.*

Therefore, ∇ is a morphism of CMet.

Remark 2.5.4. This would not be the case if we had taken as monoidal structure for CMet the cartesian product: for the product metric, ∇ is Lipschitz, but in general not 1-Lipschitz.

In order to prove Proposition 2.5.3, first a useful result:

Proposition 2.5.5. *Let $f : X \otimes Y \to \mathbb{R}$ be short. Let $p \in PX$. Then the function*

$$\left(\int_X f(x, -) \, dp(x) \right) : Y \to \mathbb{R}$$

is short as well.

Proof of Proposition 2.5.5. First of all, $f : X \otimes Y \to \mathbb{R}$ being short means that for every $x, x' \in X, y, y' \in Y$:

$$|f(x, y) - f(x', y')| \leq d(x, x') + d(y, y').$$

Now:

$$\begin{aligned}
&\left| \int_X f(x, y) \, dp(x) - \int_X f(x, y') \, dp(x) \right| \\
&= \left| \int_X \big(f(x, y) - f(x, y') \big) \, dp(x) \right| \\
&\leq \int_X |f(x, y) - f(x, y')| \, dp(x) \\
&\leq \int_X \big(d(x, x) + d(y, y') \big) \, dp(x) \\
&= \int_X d(y, y') \, dp(x) \\
&= d(y, y').
\end{aligned}$$

□

Proof of Proposition 2.5.3. To prove that ∇ it is short, let $p, p' \in PX, q, q' \in PY$. Then

$$\begin{aligned}
&d\big(\nabla(p, q), \nabla(p', q')\big) \\
&= d\big(p \otimes q, p' \otimes q'\big) \\
&= \sup_{f : X \otimes Y \to \mathbb{R}} \int_{X \otimes Y} f(x, y) \, d(p \otimes q - p' \otimes q')(x, y)
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{f:X\otimes Y\to\mathbb{R}} \int_{X\otimes Y} f(x,y)\, d\big(p\otimes q - p'\otimes q + p'\otimes q - p'\otimes q'\big)(x,y) \\
&= \sup_{f:X\otimes Y\to\mathbb{R}} \int_{X\otimes Y} f(x,y)\, d\big((p-p')\otimes q + p'\otimes (q-q')\big)(x,y) \\
&= \sup_{f:X\otimes Y\to\mathbb{R}} \int_X \left\{ \int_Y f(x,y)\, dq(y) \right\} d(p-p')(x) \\
&\qquad + \int_Y \left\{ \int_X f(x,y)\, dp'(x) \right\} d(q-q')(y) \\
&\leq \sup_{g:X\to\mathbb{R}} \int_X g(x) d(p-p')(x) + \sup_{h:Y\to\mathbb{R}} \int_Y h(y) d(q-q')(y) \\
&= d(p,p') + d(q,q') \\
&= d\big((p,q),(p',q')\big),
\end{aligned}
$$

where by replacing the partial integral of f by g we have used Proposition 2.5.5. □

The fact that ∇ equips P with a monoidal structure now follows directly from the naturality and associativity of the product probability construction (as sketched in Section 1.2). In other words, the proofs of the next three statements can be adapted to most other categorical contexts in which the map ∇ is of a similar form.

Proposition 2.5.6. *$\nabla : PX \otimes PY \to P(X \otimes Y)$ is natural in X and Y.*

Proof. By symmetry, it suffices to show naturality in X. Let $f : X \to Z$. We need to show that this diagram commutes:

$$
\begin{array}{ccc}
PX \otimes PY & \xrightarrow{\nabla_{X,Y}} & P(X\otimes Y) \\
\big\downarrow {\scriptstyle f_*\otimes \mathrm{id}} & & \big\downarrow {\scriptstyle (f\otimes\mathrm{id})_*} \\
PZ \otimes PY & \xrightarrow{\nabla_{Z,Y}} & P(Z\otimes Y)
\end{array}
$$

Now let $p \in PX, q \in PY$, and $g : Z \otimes Y \to \mathbb{R}$. Then

$$
\begin{aligned}
\int_{Z\otimes Y} f(z,y)\, d\big((f\otimes \mathrm{id})_*\nabla_{X,Y}(p,q)\big)(z,y) &= \int_{X\otimes Y} g(f(x),y)\, d(\nabla_{X,Y}(p,q))(x,y) \\
&= \int_{X\otimes Y} g(f(x),y)\, dp(x)\, dq(y) \\
&= \int_{Z\otimes Y} g(z,y)\, d(f_*p)(z)\, dq(y)
\end{aligned}
$$

$$= \int_{Z\otimes Y} g(z,y)\, d\big((f_*p)\otimes q\big)(z,y)$$

$$= \int_{Z\otimes Y} g(z,y)\, d\big(\nabla_{Z,Y}\circ (f_*\otimes \mathrm{id})(p,q)\big)(z,y).$$

□

Proposition 2.5.7. $(P, \mathrm{id}_1, \nabla)$ *is a symmetric lax monoidal functor* $\mathsf{CMet} \to \mathsf{CMet}$.

Proof. Since both maps are natural, we only need to check the coherence diagrams. Since the unitor is just the identity at the terminal object, the unit diagrams commute. The associativity diagram at each X, Y, Z

$$\begin{tikzcd} PX\otimes PY\otimes PZ \arrow[r, "\mathrm{id}\otimes\nabla_{Y,Z}"] \arrow[d, "\nabla_{X,Y}\otimes\mathrm{id}"] & PX\otimes P(Y\otimes Z) \arrow[d, "\nabla_{X,Y\otimes Z}"] \\ P(X\otimes Y)\otimes PZ \arrow[r, "\nabla_{X\otimes Y,Z}"] & P(X\otimes Y\otimes Z) \end{tikzcd}$$

gives for $(p,q,r)\in PX\otimes PY\otimes PZ$ on one path

$$(p,q,r)\mapsto (p\otimes q, r)\mapsto (p\otimes q)\otimes r,$$

and on the other path

$$(p,q,r)\mapsto (p, q\otimes r)\mapsto p\otimes (q\otimes r).$$

The product of probability distributions is now associative, as a simple calculation can show.

The symmetry condition is straightforward. □

Proposition 2.5.8. (P, δ, E) *is a symmetric monoidal monad.*

Proof. We know that $(P, \mathrm{id}_1, \nabla)$ is a lax monoidal functor. We need to check now that δ and E are monoidal natural transformations. Again we only need to show the commutativity with the multiplication, since the unitor is trivial. For $\delta : \mathrm{id}_{\mathsf{CMet}} \Rightarrow P$ we need to check that this diagram commute for each X, Y:

$$\begin{tikzcd} X\otimes Y \arrow[r, "\delta\otimes\delta"] \arrow[dr, "\delta"'] & PX\otimes PY \arrow[d, "\nabla_{X,Y}"] \\ & P(X\otimes Y) \end{tikzcd}$$

which means that for each $x\in X, y\in Y$ $\delta_x\otimes\delta_y = \delta_{(x,y)}$, which is easy to check (the delta over the product is the product of the deltas). For $E : PP \Rightarrow P$ we first need to find the multiplication map $\nabla^2_{X,Y} : PPX \otimes PPY \to PP(X \otimes Y)$ (the unit is just twice the deltas, and the unit diagram again trivially commutes). This map is given by

$$P(PX) \otimes P(PY) \xrightarrow{\nabla_{PX,PY}} P(PX \otimes PY) \xrightarrow{(\nabla_{X,Y})_*} P(P(X \otimes Y))$$

and more explicitly, if $\mu \in PPX, \nu \in PPY$, and $f : P(X \times Y) \to \mathbb{R}$,

$$\begin{aligned}
\int_{P(X\otimes Y)} f(r)\, d\big(\nabla^2_{X,Y}(\mu,\nu)\big)(r) &= \int_{P(X\otimes Y)} f(r)\, d\big((\nabla_{X,Y})_* \circ \nabla_{PX,PY}(\mu,\nu)\big)(r) \\
&= \int_{P(X\otimes Y)} f(r)\, d\big((\nabla_{X,Y})_*(\mu \otimes \nu)\big)(r) \\
&= \int_{PX\otimes PY} f(\nabla_{X,Y}(p,q))\, d(\mu\otimes\nu)(p,q) \\
&= \int_{PX\otimes PY} f(p\otimes q)\, d\mu(p)\, d\nu(q).
\end{aligned}$$

Now we have to check that this map makes this multiplication diagram commute:

$$\begin{tikzcd}
PPX \otimes PPY \arrow[r, "E_X \otimes E_Y"] \arrow[d, "\nabla^2_{X,Y}"] & PX \otimes PY \arrow[d, "\nabla_{X,Y}"] \\
PP(X \otimes Y) \arrow[r, "E_{X\otimes Y}"] & P(X \otimes Y)
\end{tikzcd}$$

Now let $\mu \in PPX, \nu \in PPY$, and $g : X \times Y \to \mathbb{R}$. We have, using the formula for ∇^2 found above,

$$\begin{aligned}
&\int_{X\otimes Y} g(x,y)\, d\big(\nabla_{X,Y} \circ (E_X, E_Y)(\mu,\nu)\big)(x,y) = \\
&= \int_{X\otimes Y} g(x,y)\, d\big(\nabla_{X,Y}(E\mu, E\nu)\big)(x,y) \\
&= \int_{X\otimes Y} g(x,y)\, d\big(E\mu \otimes E\nu\big)(x,y) \\
&= \int_{PX\otimes PY} \left\{ \int_{X\otimes Y} g(x,y)\, dp(x)\, dq(y) \right\} d\mu(p)\, d\nu(q) \\
&= \int_{PX\otimes PY} \left\{ \int_{X\otimes Y} g(x,y)\, d(p\otimes q)(x,y) \right\} d\mu(p)\, d\nu(q) \\
&= \int_{P(X\times Y)} \left\{ \int_{X\otimes Y} g(x,y)\, dr(x,y) \right\} d(\nabla^2_{X,Y}(\mu,\nu))(r)
\end{aligned}$$

$$= \int_{X \otimes Y} g(x, y) \, d\big(E_{X \otimes Y} \circ \nabla^2_{X,Y}(\mu, \nu)\big)(x, y).$$

Therefore the diagram commutes, and (P, δ, E) is a monoidal monad. □

We know that a monoidal monad is the same as a commutative monad, and therefore obtain:

Corollary 2.5.9. *P is a commutative strong monad, with strength $X \otimes PY \to P(X \otimes Y)$ given by:*

$$(x, q) \mapsto \delta_x \otimes q \in P(X \otimes Y).$$

2.5.2. Opmonoidal structure

We now turn to the analogous statements for the marginals, and show that they equip P with an opmonoidal structure.

Definition 2.5.10. *Let $X, Y \in \mathsf{CMet}$. We define the map $\Delta : P(X \otimes Y) \to PX \otimes PY$ as mapping $r \in P(X \otimes Y)$ to the pair of marginals $(r_X, r_Y) \in PX \otimes PY$.*

Proposition 2.5.11. *$\Delta : P(X \otimes Y) \to PX \otimes PY$ is short.*

Therefore Δ is a morphism of CMet.

Just as in the case of joints, to prove Proposition 2.5.11 we first prove the following useful result.

Proposition 2.5.12. *Let $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ be short. Then $(f + g) : X \otimes Y \to \mathbb{R}$ given by $(x, y) \mapsto f(x) + g(y)$ is short.*

Proof of Proposition 2.5.12. Let $x, x' \in X$ and $y, y \in Y$. Then

$$\begin{aligned} |f(x) + g(y) - f(x') - f(y')| &\le |f(x) - f(x')| + |g(y) - g(y')| \\ &\le d(x, x') + d(y, y') = d\big((x, y), (x', y')\big). \end{aligned}$$

□

Proof of Proposition 2.5.11. To prove that Δ is short, let $p, q \in P(X \otimes Y)$, and denote p_X, p_Y, q_X, q_Y their marginals. Then:

$$d\big(\Delta(p), \Delta(q)\big) = d\big((p_X, p_Y), (q_X, q_Y)\big) = d(p_X, q_X) + d(p_Y, q_Y)$$

$$= \sup_{f : X \to \mathbb{R}} \int_X f(x) \, d(p_X - q_X)(x) + \sup_{g : Y \to \mathbb{R}} \int_Y g(y) \, d(p_Y - q_Y)(y)$$

$$
\begin{aligned}
&= \sup_{f:X\to\mathbb{R}} \int_{X\otimes Y} f(x)\, d(p-q)(x,y) + \sup_{g:Y\to\mathbb{R}} \int_{X\otimes Y} g(y)\, d(p-q)(x,y) \\
&= \sup_{f:X\to\mathbb{R}} \sup_{g:Y\to\mathbb{R}} \int_{X\otimes Y} \Big(f(x) + g(y) \Big)\, d(p-q)(x,y) \\
&\le \sup_{h:X\otimes Y\to\mathbb{R}} h(x,y)\, d(p-q)(x,y) \\
&= d_{P(X\otimes Y)}(p,q),
\end{aligned}
$$

where by replacing $f+g$ with h we have used Proposition 2.5.12. □

Again, the following statements follow just from the properties of marginals, and their proofs can be adapted to most other categorical contexts provided that Δ is of a similar form.

Proposition 2.5.13. $\Delta : P(X \otimes Y) \to PX \otimes PY$ *is natural in* X, Y.

Proof. By symmetry, it suffices to show naturality in X. Let $f : X \to Z$. We need to show that this diagram commutes:

$$
\begin{tikzcd}
P(X\otimes Y) \arrow[r, "\Delta_{X,Y}"] \arrow[d, "(f\otimes \mathrm{id})_*"] & PX\otimes PY \arrow[d, "f_*\otimes \mathrm{id}"] \\
P(Z\otimes Y) \arrow[r, "\Delta_{Z,Y}"] & PZ\otimes PY
\end{tikzcd}
$$

Let now $p \in P(X \otimes Y)$. We have to prove that:

$$\Delta_{Z,Y} \circ (f\otimes \mathrm{id})_* p = (f_* \otimes \mathrm{id}) \circ \Delta_{X,Y}(p).$$

On one hand:

$$
\begin{aligned}
(f_* \otimes \mathrm{id}) \circ \Delta_{X,Y}(p) &= (f_* \otimes \mathrm{id})(p_X, p_Y) \\
&= (f_* p_X, p_Y).
\end{aligned}
$$

On the other hand, let $h : Z \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ be short. Then:

$$
\begin{aligned}
\int_Z h(z)\, d(((f\otimes \mathrm{id})_* p)_Z)(z) &= \int_{Z\otimes Y} h(z)\, d((f\otimes \mathrm{id})_* p)(z,y) \\
&= \int_{X\otimes Y} h(f(x))\, dp(x,y) \\
&= \int_X h(f(x))\, dp_X(x)
\end{aligned}
$$

$$= \int_Z h(z)\, d(f_* p_X)(x),$$

and:

$$\begin{aligned} \int_Y g(y)\, d(((f \otimes id)_* p)_Y)(y) &= \int_{Z\otimes Y} g(y)\, d((f \otimes id)_* p)(z,y) \\ &= \int_{X\otimes Y} g(y)\, dp(x,y) \\ &= \int_Y g(y)\, dp_Y(y), \end{aligned}$$

so the two components are again $(f_* p_X, p_Y)$. □

Proposition 2.5.14. *The marginal map together with the trivial counitor defines a symmetric oplax monoidal functor* $(P, \mathrm{id}_1, \Delta)$.

Proof. We already have naturality of the maps, and the counitor is trivial, we just have to check coassociativity. Namely, that the following diagrams commutes for each X, Y, Z:

$$\begin{tikzcd} P(X\otimes Y\otimes Z) \arrow[r, "\Delta_{X\otimes Y,Z}"] \arrow[d, "\Delta_{X,Y\otimes Z}"'] & P(X\otimes Y)\otimes P(Z) \arrow[d, "\Delta_{X\otimes Y}\otimes \mathrm{id}"] \\ P(X)\otimes P(Y\otimes Z) \arrow[r, "\mathrm{id}\otimes\Delta_{Y\otimes Z}"'] & P(X)\otimes P(Y)\otimes P(Z) \end{tikzcd}$$

Now given $p \in P(X \otimes Y \otimes Z)$, we get:

$$(\Delta_{X\otimes Y} \otimes \mathrm{id}) \circ \Delta_{X\otimes Y,Z}(p) = (\Delta_{X\otimes Y} \otimes \mathrm{id})(p_{XY}, p_Z) = (p_X, p_Y, p_Z),$$

and:

$$(\mathrm{id} \otimes \Delta_{Y\otimes Z}) \circ \Delta_{X,Y\otimes Z}(p) = (\mathrm{id} \otimes \Delta_{Y\otimes Z})(p_X, p_{YZ}) = (p_X, p_Y, p_Z),$$

since there is only one way of forming marginals.

The symmetry condition is again straightforward. □

Proposition 2.5.15. (P, δ, E) *is a symmetric opmonoidal monad.*

Proof. We know that $(P, \mathrm{id}_1, \Delta)$ is an oplax monoidal functor. We need to check now that δ and E are comonoidal natural transformations. Again we only need to show the commutativity with the comultiplication, since the counitor is trivial. For $\delta : \mathrm{id}_{\mathsf{CMet}} \Rightarrow P$ we need to check that this diagram commute for each X, Y:

$$\begin{tikzcd} X\otimes Y \arrow[r, "\delta"] \arrow[dr, "\delta\otimes\delta"'] & P(X\otimes Y) \arrow[d, "\Delta_{X,Y}"] \\ & PX\otimes PY \end{tikzcd}$$

which means that for each $x \in X, y \in Y$, $(\delta_{(x,y)})_X = \delta_x$ and $(\delta_{(x,y)})_Y = \delta_y$, which is again easy to check (the marginals of a delta are the deltas at the projections). For $E : PP \Rightarrow P$ we first need to find the comultiplication map $\Delta^2_{X,Y} : PP(X \otimes Y) \to PPX \otimes PPY$ (the unit is just twice the deltas, and the unit diagram again trivially commutes). This map is given by:

$$P(P(X \otimes Y)) \xrightarrow{(\Delta_{XY})_*} P(PX \otimes PY) \xrightarrow{\Delta_{PX,PY}} P(PX) \otimes P(PY)$$

and more explicitly, if $\mu \in P(P(X \otimes Y))$, and $f : PX \to \mathbb{R}$ and $g : PY \to \mathbb{R}$ are short:

$$\begin{aligned}\int_{PX} f(p)\, d\big(((\Delta_{XY})_*\mu)_{PX}\big)(p) &= \int_{PX\otimes PY} f(p)\, d\big(((\Delta_{XY})_*\mu)_{PX}\big)(p,q) \\ &= \int_{P(X\otimes Y)} f(r_X)\, d\mu(r)\end{aligned}$$

since g only depends on PX, and analogously:

$$\int_{PY} g(q)\, d\big(((\Delta_{XY})_*\mu)_{PY}\big)(q) = \int_{P(X\otimes Y)} f(r_Y)\, d\mu(r).$$

We have to check that this map makes this multiplication diagram commute:

$$\begin{array}{ccc} PP(X\otimes Y) & \xrightarrow{E_{X\otimes Y}} & P(X\otimes Y) \\ \big\downarrow{\scriptstyle \Delta^2_{X,Y}} & & \big\downarrow{\scriptstyle \Delta_{X,Y}} \\ PPX \otimes PPY & \xrightarrow{E_X \otimes E_Y} & PX \otimes PY \end{array}$$

Now let $\mu \in P(P(X \otimes Y))$, and $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ short. We have, using the formula for Δ^2 found above:

$$\begin{aligned}\int_X f(x)\, d((E_{X\otimes Y}\mu)_X)(x) &= \int_{X\otimes Y} f(x)\, d(E_{X\otimes Y}\mu)(x,y) \\ &= \int_{P(X\otimes Y)} \left\{ \int_{X\otimes Y} f(x)\, dr(x,y) \right\} d\mu(r) \\ &= \int_{P(X\otimes Y)} \left\{ \int_X f(x)\, d(r_X)(x) \right\} d\mu(r) \\ &= \int_{PX\otimes PY} \left\{ \int_X f(x)\, dp(x) \right\} d\big((\Delta_{XY})_*\mu\big)(p,q) \\ &= \int_{PX} \left\{ \int_X f(x)\, dp(x) \right\} d\big(((\Delta_{XY})_*\mu)_{PX}\big)(p)\end{aligned}$$

$$= \int_X f(x)\, d\big(E_X((\Delta_{XY})_*\mu)_{PX}\big)(x),$$

and analogously:

$$\int_Y g(y)\, d((E_{X\otimes Y}\mu)_Y)(y) = \int_Y f(y)\, d\big(E_Y((\Delta_{XY})_*\mu)_{PY}\big)(y),$$

which means:

$$\begin{aligned}
\Delta_{X,Y} \circ E_{X\otimes Y}\mu &= (E_X \otimes E_Y) \circ \Delta_{PX,PY}(\Delta_{XY})_*\mu) \\
&= (E_X \otimes E_Y) \circ (\Delta_{PX,PY} \circ (\Delta_{XY})_*)\mu \\
&= (E_X \otimes E_Y) \circ \Delta^2_{X,Y}\mu.
\end{aligned}$$

Therefore the diagram commutes, and (P, δ, E) is an opmonoidal monad. $\square$

2.5.3. Bimonoidal structure

The lax and oplax monoidal structure interact to give a bimonoidal structure. The following statement also follows just from the properties of joints and marginals.

Proposition 2.5.16. *P is a symmetric bilax monoidal functor.*

Proof. We already know that P is lax and oplax. We only need to check the compatibility diagrams between the two structures. The unit diagrams are trivial, because the unitors are trivial. The bimonoidality diagram:

$$\begin{tikzcd}
& P(W\otimes X)\otimes P(Y\otimes Z) \arrow[dl, "\nabla_{W\otimes X, Y\otimes Z}"'] \arrow[dr, "\Delta_{W,X}\otimes\Delta_{Y,Z}"] & \\
P(W\otimes X\otimes Y\otimes Z) \arrow[d, "\cong"'] & & P(W)\otimes P(X)\otimes P(Y)\otimes P(Z) \arrow[d, "\cong"] \\
P(W\otimes Y\otimes X\otimes Z) \arrow[dr, "\Delta_{W\otimes Y, X\otimes Z}"'] & & P(W)\otimes P(Y)\otimes P(X)\otimes P(Z) \arrow[dl, "\nabla_{W,Y}\otimes\nabla_{X,Z}"] \\
& P(W\otimes Y)\otimes P(X\otimes Z) &
\end{tikzcd}$$

says that given $p \in P(W\otimes X), q \in P(Y \otimes Z)$:

$$\Delta_{W\otimes Y, X\otimes Z} \circ \nabla_{W\otimes X, Y\otimes Z}(p,q) = (\nabla_{W,Y}\otimes\nabla_{X,Z}) \circ (\Delta_{W,X}\otimes\Delta_{Y,Z})(p,q)$$

Now on one hand:

$$(\nabla_{W,Y}\otimes\nabla_{X,Z}) \circ (\Delta_{W,X}\otimes\Delta_{Y,Z})(p,q) = (\nabla_{W,Y}\otimes\nabla_{X,Z})(p_W, p_X, q_Y, q_Z)$$

$$= (p_W \otimes q_Y, p_X \otimes q_Z).$$

On the other hand:

$$\Delta_{W\otimes Y, X\otimes Z} \circ \nabla_{W\otimes X, Y\otimes Z}(p,q) = \Delta_{W\otimes Y, X\otimes Z}(p\otimes q).$$

The marginal of $p \otimes q$ on $W \otimes Y$ is, by Fubini's theorem, let $f : W \otimes Y \to \mathbb{R}$:

$$\begin{aligned} \int_{W\otimes Y} f(w,y)\, d((p\otimes q)_{WY})(w,y) &= \int_{W\otimes X\otimes Y\otimes Z} f(w,y)\, d(p\otimes q)(w,x,y,z) \\ &= \int_{W\otimes X\otimes Y\otimes Z} f(w,y)\, dp(w,x)\, dq(y,z) \\ &= \int_{W\otimes Y} f(w,y)\, dp_W(w)\, dq_Y(y) \\ &= \int_{W\otimes Y} f(w,y)\, d(p_W \otimes q_Y)(w,y), \end{aligned}$$

and similarly the marginal on $X \otimes Z$ is given by $p_X \otimes q_Z$. In other words, if the pairs are independent, the components from different pairs are also independent. It follows that P is bilax monoidal. □

The main result then just follows as a corollary:

Theorem 2.5.17. *The Kantorovich monad is a symmetric bimonoidal monad, with monoidal structure given by the product joint, and opmonoidal structure given by the marginals.*

By Proposition 1.2.2, we therefore have:

Corollary 2.5.18. $\Delta_{X,Y} \circ \nabla_{X,Y} = \mathrm{id}_{PX\otimes PY}$. *Therefore, the inclusion ∇ of product measures into general joints, is an isometric embedding for the Kantorovich metric, and its image is a retract of the space of all joints.*

2.6. Lifting and disintegration results

The main goal of this section is to prove that E is a proper map, i.e. its inverse image maps compact sets to compact sets (Theorem 2.6.7). This result will allow us to prove straightforwardly some lifting results for probability measures without requiring disintegration theorems.

In 2.6.1 we look at the behavior of E on the supports. We will find that the inverse image of E never increases the support of a measure. In 2.6.2 we will prove a lifting criterion for E, analogous to that of a fibration, or of a submersion, but for metric spaces. We use these results in 2.6.3 to prove that E is a proper map (Theorem 2.6.7). In 2.6.4, we show why Theorem 2.6.7 implies a sort of disintegration theorem, Theorem 2.6.9. Finally, in 2.6.5 we apply the same technique to prove that the marginal map Δ is proper as well, which implies that the space of couplings of any two fixed probability measures is always compact.

2.6.1. Expectations and supports

Let X be a complete metric space, and $p \in PX$. The *support* of p is the set of points of X whose neighborhoods have positive measure. We give here an alternative characterization, which will be useful later.

Proposition 2.6.1. *Let $X \in \mathsf{CMet}$, let $p \in PX$, and let $x \in X$. Denote by $B(\varepsilon, x)$ the ball of radius ε centered at x. The following conditions are equivalent:*

(a) For every $\varepsilon > 0$, $p(B(\varepsilon, x)) > 0$;

(b) For every $\varepsilon > 0$,

$$\int_X \phi_{x,\varepsilon}(y)\, dp(y) > 0,$$

where:

$$\phi_{x,\varepsilon}(y) := \max\{0\,,\, \varepsilon - d(x,y)\};$$

(c) For every $\varepsilon > 0$, and for every short map $f : X \to \mathbb{R}_+$, such that $f(y) > 0$ for every $y \in B(\varepsilon, x)$,

$$\int_X f\, dp > 0.$$

Proof. We first notice that $\phi_{x,\varepsilon}$ is short, bounded above by ε, and zero outside an ε-neighborhood of x.

- $(a) \Rightarrow (b)$:

$$\begin{aligned} \int_X \phi_{x,\varepsilon}(y)\, dp(y) &\geq \int_{B(\varepsilon/2,x)} \phi_{x,\varepsilon}(y)\, dp(y) \\ &\geq \inf_{y \in B(\varepsilon/2,x)} \phi_{x,\varepsilon}(y) \cdot p(B(\varepsilon/2, x)) \\ &= \varepsilon/2 \cdot p(B(\varepsilon/2, x)) > 0. \end{aligned}$$

- $(b) \Rightarrow (a)$: Assume $\varepsilon < 1$. So

$$
\begin{aligned}
p(B(\varepsilon,x)) &\geq \int_{B(\varepsilon,x)} \varepsilon \, dp \\
&\geq \int_{B(\varepsilon,x)} \phi_{x,\varepsilon}(y)\, dp(y) \\
&= \int_X \phi_{x,\varepsilon}(y)\, dp(y) > 0.
\end{aligned}
$$

- $(b) \Rightarrow (c)$: Let f be such a function and set $\delta := f(x) > 0$. Since f is short $f(x) - f(y) \leq d(x,y)$, so that for every $y \in B(x,\delta)$,

$$f(y) \geq f(x) - d(x,y) = \delta - d(x,y) = \phi_{x,\delta}(y).$$

  Now

$$
\begin{aligned}
\int_X f(y)\, dp(y) &\geq \int_{B(x,\delta)} f(y)\, dp(y) \\
&\geq \int_{B(x,\delta)} \phi_{x,\delta}(y)\, dp(y) \\
&= \int_X \phi_{x,\delta}(y)\, dp(y) > 0.
\end{aligned}
$$

- $(c) \Rightarrow (b)$: $\phi_{x,2\varepsilon}$ is short, and strictly positive on $B(x,\varepsilon)$.

□

We denote the set of points satisfying any of the condition above by $\mathrm{supp}(p)$. Such a set is always closed. Denote by HX the set of closed sets of X. The support gives a function $\mathrm{supp} : PX \to HX$. We could equip HX with a metric, for example the Hausdorff metric; however, to the best of our knowledge there is no interesting metric that makes the map supp short, or even continuous.[9] So in the following beware: supp is *not* a morphism of CMet.

Proposition 2.6.2. *Let $X \in$ CMet, $x \in X$, and $\mu \in PPX$. Let $p \in PX$ be in the support of μ. Then*

$$\mathrm{supp}(p) \subseteq \mathrm{supp}(E\mu). \tag{2.6.1}$$

[9] The support map does have a continuity-like property, namely *Scott-* or *lower-semicontinuity* for the inclusion order in HX. This will however not be pursued in this work.

Proof. Let $x \in \operatorname{supp}(p)$. By Proposition 2.6.1 we have that for every $\varepsilon > 0$,

$$\int_X \phi_{x,\varepsilon}\, dp > 0.$$

Now let $\delta > 0$. Since $\phi_{x,\varepsilon}$ is short, for every $q \in B(\delta, p)$, we have as well that

$$\left| \int_X \phi_{x,\varepsilon}\, dq - \int_X \phi_{x,\varepsilon}\, dp \right| \leq d(p,q) < \delta,$$

so that:

$$\int_X \phi_{x,\varepsilon}\, dq > \int_X \phi_{x,\varepsilon}\, dp - \delta,$$

which by taking δ small enough, is positive. Therefore the map

$$q \mapsto \int_X \phi_{x,\varepsilon}\, dq, \tag{2.6.2}$$

which is short, is strictly positive on $B(p, \delta)$. Since p is in the support of μ, by Proposition 2.6.1, the integral of 2.6.2 is strictly positive, i.e.

$$\int_{PX} \left(\int_X \phi_{x,\varepsilon}(y)\, dq(y) \right) d\mu(q) > 0,$$

but the r.h.s. above is equal to

$$\int_X \phi_{x,\varepsilon}\, d(E\mu), \tag{2.6.3}$$

so that again by Proposition 2.6.1, $x \in \operatorname{supp}(E\mu)$. □

Corollary 2.6.3. *Let $\mu \in PPX$, and let $E\mu$ be supported on $Y \subseteq X$. Then μ is supported on PY, i.e. on the measures which are themselves supported on Y.*

2.6.2. Metric lifting

There is a lifting criterion for E, which is a metric analogue of the homotopy lifting property: given $p, q \in PX$ with distance less than r, and given a preimage $\mu \in E^{-1}(p)$, then there is a $\nu \in E^{-1}(q)$ with $d(\mu, \nu) < r$.

To prove the statement, we will use the colimit characterization of P given in Sections 2.2 and 2.3. This allows to prove the result first for finite sequences, where the proof is only combinatorics, and then to extend it by density to the fully general case.

Proposition 2.6.4. *Let $\bar{\mu} = \{\{\bar{\mu}_{m,n}\}_{m\in M}\}_{n \in N} \in (X^M)^N$ and $\bar{q} = \{\bar{q}_{m,n}\}_{m\in M, n\in N} \in X^{MN}$. Suppose that*

$$d\big(i \circ E^{M,N}(\bar{\mu}), i(\bar{q})\big) < r, \tag{2.6.4}$$

where $i : X^{MN} \to PX$ denotes the empirical distribution. Then there exists $\bar{\nu} \in (X^M)^N$ such that $E^{M,N}(\bar{\nu}) = (\bar{q}) \circ \sigma$ for some permutation $\sigma \in S_{M'N'}$, and $d(\bar{\mu}, \bar{\nu}) < r$.

Proof. By the formula (2.2.2) together with the fact that i_n is an isometric embedding (Proposition 2.2.10), condition (2.6.4) is equivalent to say that

$$\min_{\sigma \in S_{MN}} \frac{1}{|MN|} \sum_{(m,n)\in MN} d(\bar{\mu}_{m,n}, \bar{q}_{\sigma(m,n)}) < r,$$

which means that there exists a $\bar{\sigma} \in S_{MN}$ such that

$$\frac{1}{|MN|} \sum_{(m,n)\in MN} d(\bar{\mu}_{m,n}, \bar{q}_{\bar{\sigma}(m,n)}) < r. \tag{2.6.5}$$

Let now

$$\bar{\nu} := \{\{\bar{\nu}_{m,n}\}\} := \{\{\bar{q}_{\bar{\sigma}(m,n)}\}\}.$$

Then (2.6.5) implies that

$$\begin{aligned} d(\bar{\mu}, \bar{\nu}) &= \frac{1}{|M|} \sum_{m\in M} \left(\frac{1}{|N|} \sum_{n\in N} d(\bar{\mu}_{m,n}, \bar{q}_{\bar{\sigma}(m,n)}) \right) \\ &= \frac{1}{|MN|} \sum_{(m,n)\in MN} d(\bar{\mu}_{m,n}, \bar{q}_{\bar{\sigma}(m,n)}) < r. \end{aligned}$$

□

By density, we get a similar statement for general probability measures:

Proposition 2.6.5 (Metric lifting)**.** *Let $X \in \mathsf{CMet}$. Let $\mu \in PPX$, $q \in PX$, and suppose $d(E\mu, q) < r$. Then there exists $\nu \in PPX$ such that $E\nu = q$, and $d(\mu, \nu) < r$.*

Proof. By density, for any $\delta > 0$, by we can find $N, M \in \mathsf{FinUnif}$, $\bar{\mu} = \{\{\bar{\mu}_{m,n}\}_{m\in M}\}_{n\in N} \in (X^M)^N$ and $\bar{q} = \{\bar{q}_{m,n}\}_{m\in M, n\in N} \in X^{MN}$ such that $d(i(\bar{\mu}), \mu) < \delta$ in PPX and $d(i(\bar{q}), q) < \delta$ in PX. We have that

$$\begin{aligned} d\big(i \circ E^{M,N}(\bar{\mu}), i(\bar{q})\big) &= d\big(E \circ i(\bar{\mu}), i(\bar{q})\big) \\ &\le d\big(E \circ i(\bar{\mu}), E\mu\big) + d\big(E\mu, q\big) + d\big(q, i(\bar{q})\big) \\ &< \delta + r + \delta. \end{aligned}$$

By Proposition 2.6.4, there exists $\bar{\nu} \in (X^M)^N$ such that $E^{M,N}(\bar{\nu}) = (\bar{q}) \circ \sigma$ for some permutation $\sigma \in S_{MN}$, and $d(\bar{\mu}, \bar{\nu}) < 2\delta + r$. This implies that:

$$\begin{aligned} d\big(\mu, i(\bar{\nu})\big) &\le d\big(\mu, i(\bar{\mu})\big) + d\big(i(\bar{\mu}), i(\bar{\nu})\big) \\ &\le d\big(\mu, i(\bar{\mu})\big) + d\big(\bar{\mu}, \bar{\nu}\big) \end{aligned}$$

$$< 3\delta + r,$$

so that by choosing δ suitably small,

$$d\big(\mu, i(\bar{\nu})\big) < r. \tag{2.6.6}$$

We can now repeat this process for smaller and smaller δ. We use a sequence $\{\bar{q}_j\}$ with $q_j \in X^{M_j N_j}$ for some $M_j, N_j \in \mathsf{FinUnif}$ suitably large, such that $\{i(q_j)\}$ is Cauchy in PX, tending to q in PX arbitrarily fast. We get a sequence $\{\bar{\nu}_j\}$ with $\bar{\nu}_j \in (X^{M_j})^{N_j}$, such that for all $h \leq j$:

$$\begin{aligned} d\big(\bar{i}(\nu_h), i(\bar{\nu}_j)\big) &\leq \sum_{k=h}^{j-1} d\big(i(\bar{\nu}_k), i(\bar{\nu}_{k+1})\big) \\ &\leq \sum_{k=h}^{j-1} d\big(i(\bar{q}_k), i(\bar{q}_{k+1})\big) \\ &\leq \sum_{k=h}^{j-1} \Big(d\big(i(\bar{q}_k), q\big) + d\big(i(\bar{q}_{k+1}), q\big)\Big) \\ &\leq 2\sum_{k=h}^{\infty} d\big(i(\bar{q}_k), q\big). \end{aligned}$$

By choosing $\{\bar{q}_j\}$ such that $d(i(\bar{q}_j), q) \leq r \cdot 2^{-j}$, we get that $\{i(\bar{\nu}_j)\}$ must be Cauchy, and therefore by completeness converge to some $\nu \in PPX$. We then have:

$$\begin{aligned} E(\nu) &= \lim_{j\to\infty} E \circ i(\bar{\nu}_j) \\ &= \lim_{j\to\infty} i \circ E^{M,N}(\bar{\nu}_j) \\ &= \lim_{j\to\infty} i(\bar{q}_j) = q, \end{aligned}$$

and by (2.6.6):

$$d(\mu, \nu) = \lim_{j\to\infty} d\big(\mu, i(\bar{\nu}_j)\big) < r.$$

□

There is also an intermediate result, which will be useful later.

Proposition 2.6.6. *Let $M, N \in \mathsf{FinUnif}$, let $\mu \in PPX$, and $\bar{q} = \{\bar{q}_{m,n}\}_{m\in M, n\in N} \in X^{MN}$. Suppose that $d(E\mu, i(\bar{q})) < r$. Then there exist M', N' multiples of M, N, $\bar{\nu} \in (X^{M'})^{N'}$ and $\tilde{q}$ representing $\bar{q}$ in $X^{M'N'}$ (via some diagonal embedding), such that $E^{M,N}(\bar{\nu}) = (\bar{q}) \circ \sigma$ for some permutation $\sigma \in S_{M'N'}$, and $d(\mu, i(\bar{\nu})) < r$.*

Proof. Let $\mu, \bar{q}$ be as in the hypothesis. By density, for any $\delta > 0$, by possibly picking larger $N, M \in \mathsf{FinUnif}$, we can find $\bar{\mu} = \{\{\bar{\mu}_{m,n}\}_{m \in M}\}_{n \in N} \in (X^M)^N$ such that $d(i(\bar{\mu}), \mu) < \delta$. Now

$$\begin{aligned} d\big(i \circ E^{M,N}(\bar{\mu}), i(\bar{q})\big) &= d\big(E \circ i(\bar{\mu}), i(\bar{q})\big) \\ &\le d\big(E \circ i(\bar{\mu}), E\mu\big) + d\big(E\mu, i(\bar{q})\big) \\ &\le d\big(i(\bar{\mu}), \mu\big) + d\big(E\mu, i(\bar{q})\big) \\ &< \delta + r. \end{aligned}$$

By Proposition 2.6.4, there exists $\bar{\nu} \in (X^M)^N$ such that $E^{M,N}(\bar{\nu}) = \bar{q}$, and $d(\bar{\mu}, \bar{\nu}) < \delta + r$. In other words, we are saying that for every $\delta > 0$ we can find $\bar{\mu}, \bar{\nu} \in (X^M)^N$ such that

$$E^{M,N}\{\{\bar{\nu}_{m,n}\}\} = \{\bar{\nu}_{mn}\} = \{q_{\sigma(m,n)}\},$$

and

$$\begin{aligned} d(\mu, i(\bar{\nu})) &\le d(\mu, i(\bar{\mu})) + d(i(\bar{\mu}), i(\bar{\nu})) \\ &< \delta + d(\bar{\mu}, \bar{\nu}) < \delta + r. \end{aligned}$$

By choosing δ suitably small, we obtain the assertion. $\square$

2.6.3. Properness of expectation

He we will prove that the integration map E is proper, i.e. its preimage maps compact sets to compact sets. This result will have important applications in 2.6.4 and in Chapter 4.

Theorem 2.6.7. *Let $X \in \mathsf{CMet}$.*

(a) Let $p \in PX$. Then $E^{-1}(p) \subseteq PPX$ is compact.

(b) Let $K \subseteq PX$ be compact. Then $E^{-1}(K) \subseteq PPX$ is compact as well.

In other words, E is a proper map.

Proof. (a) Let $p \in PX$. Then by density, for every $\varepsilon > 0$ there exists a p_ε with compact support K_ε, and such that $d(p, p_\varepsilon) < \varepsilon/2$. By Proposition 2.6.5, then for every $\mu \in E^{-1}(p)$ we can find some μ_ε such that $d(\mu, \mu_\varepsilon) < \varepsilon/2$ and $E\mu_\varepsilon = p_\varepsilon$. By Corollary 2.6.3, μ_ε is supported on $P(K_\varepsilon)$, which is itself compact, and which does not depend on μ varying in $E^{-1}(p)$. In other words, the whole $E^{-1}(p)$ is contained within an $\varepsilon/2$-neighborhood of $PP(K_\varepsilon)$. By compactness, for every $\varepsilon > 0$, $PP(K_\varepsilon)$ can be covered by a finite number of balls of radius $\varepsilon/2$. Then $E^{-1}(p)$ can be covered by a finite number of balls of radius ε, i.e. it is totally bounded. Since E is continuous, $E^{-1}(p)$ is closed. Therefore $E^{-1}(p)$ is compact.

(b) Again, we just need to show total boundedness. Since K is compact, for every $\varepsilon > 0$ there exists a finite $(\varepsilon/2)$-net $\{p_n\}$ covering K (i.e. every element $k \in K$ is within distance $\varepsilon/2$ from $\{p_n\}$). Take now the finite collection of sets $\{E^{-1}(p_n)\}$. By (a), we know that they are all compact, and by Proposition 2.6.5 we know that every element $\mu \in E^{-1}(K)$ is within distance $\varepsilon/2$ from some element of $\cup_n E^{-1}(p_n)$. Now the set $\cup_n E^{-1}(p_n)$ is a finite union of compact sets, so it is compact, and in particular it can be covered by finitely many balls of radius $\varepsilon/2$. This implies that for every $\varepsilon > 0$, the whole $E^{-1}(K)$ can be covered by finitely many balls of radius ε, i.e. it is totally bounded.

□

Corollary 2.6.8. *Let $X \in \mathsf{CMet}$. Let $\{\mu_i\}$ be a (generic) sequence in PPX, such that $\{E\mu_i\}$ forms a Cauchy sequence in PX whose limit we denote p. Then $\{\mu_i\}$ admits an accumulation point $\mu \in PPX$ (which then necessarily satisfies $E\mu = p$).*

Proof of the corollary. Let $K := \{E\mu_1, E\mu_2, \dots, E\mu_i, \dots, p\} \subseteq PX$, which is compact since $\{E\mu_i\}$ tends to p. By Theorem 2.6.7, its inverse image $E^{-1}(K)$ is compact as well. Now $\{E\mu_i\}$ takes values inside $E^{-1}(K)$, and so it must have at least one accumulation point μ. Since E is continuous, then $E\mu = p$. □

2.6.4. Existence of disintegrations

Here we prove the following "disintegration" result:

Theorem 2.6.9. *Let $f : X \to Y$. Consider the following naturality square:*

$$\begin{tikzcd} PPX \arrow[r, "PPf"] \arrow[d, "E"'] & PPY \arrow[d, "E"] \\ PX \arrow[r, "Pf"'] & PY \end{tikzcd}$$

Let $p \in PX$ and $\nu \in PPY$ such that $(Pf)\,p = E\nu$ in PY. Then there exists $\mu \in PPX$ such that

$$E\mu = p \quad \textit{and} \quad (PPf)\,\mu = \nu. \tag{2.6.7}$$

The intuition is that we can find a "disintegration" μ of p by looking at how f_*p is "disintegrated" into ν. To prove the theorem, again we first prove an analogous result for empirical distributions of finite sequences, and then proceed to the general case by density.

Lemma 2.6.10. *Let $f : X \to Y \in \mathsf{CMet}$, and $M, N \in \mathsf{FinUnif}$. Consider the following naturality square of symmetric powers:*

$$\begin{tikzcd}
(X_{|M|})_{|N|} \arrow[r, "(f_{|M|})_{|N|}"] \arrow[d, "E_{|M|,|N|}"'] & (Y_{|M|})_{|N|} \arrow[d, "E_{|M|,|N|}"] \\
X_{|MN|} \arrow[r, "f_{|MN|}"'] & Y_{|MN|}
\end{tikzcd}$$

Let $x \in X_{|MN|}$ and $y \in (Y_{|M|})_{|N|}$ such that $f_{|MN|}(x) = E_{|M|,|N|}(y)$ in $Y_{|MN|}$. Then there exists $w \in (X_{|M|})_{|N|}$ such that $E_{|M|,|N|}(w) = x$ and $(f_{|M|})_{|N|}(w) = y$.

Proof of Lemma 2.6.10. Consider the following commutative diagram:

$$\begin{tikzcd}
(X^M)^N \arrow[rrr, "(f^M)^N"] \arrow[dr, "q"] \arrow[ddd, "E^{M,N}"', "\cong"] & & & (Y^M)^N \arrow[dl, "q"'] \arrow[ddd, "E^{M,N}", "\cong"'] \\
& (X_{|M|})_{|N|} \arrow[r, "(f_{|M|})_{|N|}"] \arrow[d, "E_{|M|,|N|}"'] & (Y_{|M|})_{|N|} \arrow[d, "E_{|M|,|N|}"] & \\
& X_{|MN|} \arrow[r, "f_{|MN|}"'] & Y_{|MN|} & \\
X^{MN} \arrow[ur, "q"] \arrow[rrr, "f^{MN}"'] & & & Y^{MN} \arrow[ul, "q"']
\end{tikzcd}$$

where the maps q are the respective quotients maps. The hypothesis is equivalent to saying that there exist $\{x_{mn}\} \in X^{MN}$ (with $q\{x_{mn}\} = x$) and $\{\{y_m\}_n\} \in (Y^M)^N$ (with $q\{\{y_m\}_n\} = y$) such that for some permutation $\sigma \in S_{MN}$, $f^{MN}\{x_{\sigma(m,n)}\} = E^{M,N}\{\{y_m\}_n\}$. But then by possibly permuting the components of $\{x_{mn}\}$, we have $\{x_{mn}\} \in X^{MN}$ and $\{\{y_m\}_n\} \in (Y^M)^N$ such that $f^{MN}\{x_{mn}\} = E^{M,N}\{\{y_m\}_n\}$. Take now $w := q \circ (E^{M,N})^{-1}\{x_{mn}\}$. We have that

$$\begin{aligned}
E_{|M|,|N|}(w) &= E_{|M|,|N|} \circ q \circ (E^{M,N})^{-1}\{x_{mn}\} \\
&= q \circ E^{M,N} \circ (E^{M,N})^{-1}\{x_{mn}\} \\
&= q\{x_{mn}\} = x,
\end{aligned}$$

and

$$\begin{aligned}
(f_{|M|})_{|N|}(w) &= (f_{|M|})_{|N|} \circ q \circ (E^{M,N})^{-1}\{x_{mn}\} \\
&= q \circ (f^M)^N \circ (E^{M,N})^{-1}\{x_{mn}\} \\
&= q \circ (E^{M,N})^{-1} \circ f^{MN}\{x_{mn}\} \\
&= q \circ (E^{M,N})^{-1} \circ E^{M,N}\{\{y_m\}_n\} \\
&= q \circ \{\{y_m\}_n\} = y.
\end{aligned}$$

□

We can now proceed to prove the main statement.

Proof of Theorem 2.6.9. Consider the following commutative diagram:

$$\begin{tikzcd}
(X_{|M|})_{|N|} \arrow[rrr, "(f_{|M|})_{|N|}"] \arrow[dr, "i"'] \arrow[ddd, "E_{|M|,|N|}"'] & & & (Y_{|M|})_{|N|} \arrow[dl, "i"] \arrow[ddd, "E_{|M|,|N|}"] \\
& PPX \arrow[r, "PPf"] \arrow[d, "E"'] & PPY \arrow[d, "E"] & \\
& PX \arrow[r, "Pf"'] & PY & \\
X_{|MN|} \arrow[ur, "i"] \arrow[rrr, "f_{|MN|}"'] & & & Y_{|MN|} \arrow[ul, "i"']
\end{tikzcd}$$

By density, we can find sequences $\{M_j\}, \{N_j\}$ in FinUnif and $\{p_j\}, \{\bar{\nu}_j\}$, with $p_j \in X_{M_j N_j}$ and $\bar{\nu}_j \in (Y_{|M_j|})_{|N_j|}$ for all j, and such that $i(p_j) \to p$ and $i(\bar{\nu}_j) \to \nu$. Since i is an isometric embedding, this means that

$$\begin{aligned}
d\big(f_{|M_j N_j|}(p_j), E_{|M_j|,|N_j|}(\bar{\nu}_j)\big) &= d\big(i \circ f_{|M_j N_j|}(p_j), i \circ E_{|M_j|,|N_j|}(\bar{\nu}_j)\big) \\
&= d\big((Pf) \circ i(p_j), E \circ i(\bar{\nu}_j)\big) \\
&\to d\big((Pf)p, E\nu\big) = 0.
\end{aligned}$$

By Proposition 2.6.6, we can then find a sequence $\{\nu_j\}$ with $\nu_j \in (Y_{|M_j|})_{|N_j|}$ and $i(\nu_j) \to \nu$, such that in addition $f_{|M_j N_j|}(p_j) = E_{|M_j|,|N_j|}(\nu_j)$ for all j. By Lemma 2.6.10, for each j there exists a $\mu_j \in (X_{|M_j|})_{|N_j|}$ such that $E_{|M_j|,|N_j|}(\mu_j) = p_j$ and $(f_{|M_j|})_{|N_j|}(\mu_j) = \nu_j$. Consider now the sequence $\{i(\mu_j)\}$ in PPX. First of all we have that

$$E \circ i(\mu_j) = i \circ E_{|M_j|,|N_j|}(\mu_j) = i(p_j) \to p,$$

and

$$(PPf) \circ i(\mu_j) = i \circ (f_{|M_j|})_{|N_j|}(\mu_j) = i(\nu_j) \to \nu,$$

so that any accumulation point of $\{i(\mu_j)\}$ satisfies the requirements (2.6.7). By Corollary 2.6.8, we know that at least one such accumulation point exists. $\square$

2.6.5. Properness of the marginal map

The same technique that we used to prove that E is a proper map can be used to show that the marginal map $\Delta : P(X \otimes Y) \to PX \otimes PY$ of Section 2.5 is proper as well. This implies in particular that the space $\Gamma(p, q)$ of couplings of two fixed probability measures p and q is always compact, and therefore the optimal coupling is always attained. This statement seems to be known at least for Polish spaces [Vil09], our result works on all complete metric spaces.

Proposition 2.6.11 (Metric lifting)**.** *Let $X, Y \in \mathsf{CMet}$. Let $r \in P(X \otimes Y)$, $(p, q) \in PX \otimes PY$, and suppose $d(\Delta r, (p, q)) < r$. Then there exists $s \in P(X \otimes Y)$ such that $\Delta(s) = (p, q)$, and $d(r, s) < r$.*

First of all, an analogous statement for finite empirical distributions. Denote $\Delta^N : (X \otimes Y)^N \to X^N \otimes Y^N$ to be the map

$$\{(x_n, y_n)\}_{n\in N} \longmapsto \big(\{x_n\}_{n\in N}, \{y_n\}_{n\in N}\big). \tag{2.6.8}$$

Proposition 2.6.12. *Let $\bar{r} = \{\bar{r}_n\}_{n\in N} \in (X \otimes Y)^N$ and $(\bar{p}, \bar{q}) = (\{\bar{p}_n\}_{n\in N}, \{\bar{q}_n\}_{n\in N}) \in X^N \otimes Y^N$. Suppose that*

$$d\big(i \circ \Delta^N(\bar{r}), (i(\bar{p}), i(\bar{q}))\big) < r, \tag{2.6.9}$$

where $i : X^N \otimes Y^N \to PX \otimes PY$ denotes the empirical distribution applied twice (i.e. it is short for $i \otimes i$). Then there exists $\bar{s} \in (X \otimes Y)^N$ such that $\Delta^N(\bar{s}) = (\bar{p} \circ \sigma, \bar{q} \circ \sigma')$ for some permutations $\sigma, \sigma' \in S_n$, and $d(\bar{r}, \bar{s}) < r$.

Proof. Denote explicitly $\bar{r}_n := (x_n, y_n)$ for all $n \in N$. By the formula (2.2.2) together Proposition 2.2.10, condition (2.6.9) is equivalent to say that

$$\min_{\sigma,\sigma' \in S_N} \frac{1}{|N|} \sum_{n\in N} \big(d(x_n, p_{\sigma(n)}) + d(y_n, q_{\sigma'(n)})\big) < r,$$

which means that there exist $\bar{\sigma}, \bar{\sigma}' \in S_N$ such that

$$\frac{1}{|N|} \sum_{n\in N} \big(d(x_n, p_{\bar{\sigma}(n)}) + d(y_n, q_{\bar{\sigma}'(n)})\big) < r. \tag{2.6.10}$$

Let now

$$\bar{s} := \{\bar{s}_n\} := \{(p_{\bar{\sigma}(n)}, q_{\bar{\sigma}'(n)})\} \in (X \otimes Y)^N.$$

Then (2.6.10) implies that

$$d(\bar{r}, \bar{s}) = \frac{1}{|N|} \sum_{n \in N} \big(d(x_n, p_{\bar{\sigma}(n)}) + d(y_n, q_{\bar{\sigma}'(n)}) \big) < r.$$

□

We can now prove the statement by density:

Proof of Proposition 2.6.11. Let r, p, q be as in the hypothesis. By density, for any $\delta > 0$, we can find $N \in \mathsf{FinUnif}$, $\bar{r} = \{\bar{r}_n\}_{n \in N} \in (X \otimes Y)^N$ and $(\bar{p}, \bar{q}) = (\{\bar{p}_n\}_{n \in N}, \{\bar{q}_n\}_{n \in N}) \in X^N \otimes Y^N$ such that $d(i(\bar{r}), r) < \delta$ in $P(X \otimes X)$ and $d\big((i(\bar{p}), i(\bar{q})), (p, q)\big) < \delta$ in $PX \otimes PX$. We have that

$$\begin{aligned} d\big(i \circ \Delta^N(\bar{r}), i(\bar{p}, \bar{q})\big) &= d\big(\Delta \circ i(\bar{r}), i(\bar{p}, \bar{q})\big) \\ &\le d\big(\Delta \circ i(\bar{r}), \Delta(r)\big) + d\big(\Delta(r), (p, q)\big) + d\big((p, q), i(\bar{p}, \bar{q})\big) \\ &< \delta + r + \delta. \end{aligned}$$

Then by Proposition 2.6.12 there exists $\bar{s} \in (X \otimes Y)^N$ such that $\Delta^N(\bar{s}) = (\bar{p} \circ \sigma, \bar{q} \circ \sigma')$ for some permutations $\sigma, \sigma' \in S_n$, and $d(\bar{r}, \bar{s}) < 2\delta + r$. This implies that:

$$\begin{aligned} d\big(r, i(\bar{s})\big) &\le d\big(r, i(\bar{r})\big) + d\big(i(\bar{r}), i(\bar{s})\big) \\ &\le d\big(r, i(\bar{r})\big) + d\big(\bar{r}, \bar{s}\big) \\ &< 3\delta + r, \end{aligned}$$

so that by choosing δ suitably small,

$$d\big(r, i(\bar{s})\big) < r. \tag{2.6.11}$$

We can now repeat this process for smaller and smaller δ. We use a sequence $\{(\bar{p}_j, \bar{q}_j)\}$ with $p_j \in X^{N_j}$ and $q_j \in Y^{N_j}$ for some $N_j \in \mathsf{FinUnif}$ suitably large, such that $\{i(\bar{p}_j)\}$ and $\{i(\bar{q}_j)\}$ are Cauchy in PX and PY, tending to p and q, respectively, arbitrarily fast. For example, choose the sequences in such a way that $d(i(\bar{p}_j), p) \le r \cdot 2^{-j}$ and $d(i(\bar{q}_j), q) \le r \cdot 2^{-j}$. We get a sequence $\{\bar{s}_j\}$ with $\bar{s}_j \in (X \otimes Y)^{N_j}$, such that for all $h \le j$:

$$d\big(i(\bar{s}_h), i(\bar{s}_j)\big) \le \sum_{k=h}^{j-1} d\big(i(\bar{s}_k), i(\bar{s}_{k+1})\big)$$

$$\leq \sum_{k=h}^{j-1} d\big(i(\bar{p}_k), i(\bar{p}_{k+1})\big) + d\big(i(\bar{q}_k), i(\bar{q}_{k+1})\big)$$

$$\leq \sum_{k=h}^{j-1} \Big(d\big(i(\bar{p}_k), p\big) + d\big(i(\bar{p}_{k+1}), p\big)\Big) + \Big(d\big(i(\bar{q}_k), q\big) + d\big(i(\bar{q}_{k+1}), q\big)\Big)$$

$$\leq 2 \sum_{k=h}^{\infty} d\big(i(\bar{p}_k), p\big) + d\big(i(\bar{q}_k), q\big).$$

By choosing $\{(\bar{p}_j, \bar{q}_j)\}$ such that $d(i(\bar{p}_j), p) \leq r \cdot 2^{-j}$ and $d(i(\bar{q}_j), q) \leq r \cdot 2^{-j}$, we get that $\{i(\bar{s}_j)\}$ must be Cauchy, and therefore by completeness converge to some $s \in P(X \otimes Y)$. We then have:

$$\Delta(s) = \lim_{j\to\infty} \Delta \circ i(\bar{s}_j) = \lim_{j\to\infty} i \circ \Delta^N(\bar{s}_j)$$

$$= \lim_{j\to\infty} (i(\bar{p}_j), i(\bar{q}_j)) = (p, q),$$

and by (2.6.11):

$$d(r, s) = \lim_{j\to\infty} d\big(r, i(\bar{s}_j)\big) < r.$$

□

We are ready to prove the main statement.

Theorem 2.6.13. *Let $X, Y \in \mathsf{CMet}$.*

(a) Let $(p, q) \in PX \otimes PY$. Then $\Delta^{-1}(p) \subseteq P(X \otimes Y)$ is compact.

(b) Let $K \subseteq PX \otimes PY$ be compact. Then $\Delta^{-1}(K) \subseteq P(X \otimes Y)$ is compact as well.

In other words, Δ is a proper map.

Proof. (a) Let $(p, q) \in PX \otimes PY$. Then by density, for every $\varepsilon > 0$ there exist $p_\varepsilon \in PX$ and $q_\varepsilon \in PY$ with compact support K_ε and H_ε, respectively, and such that $d(p, p_\varepsilon) < \varepsilon/4$ and $d(q, q_\varepsilon) < \varepsilon/4$. By Proposition 2.6.11, then for every $r \in \Delta^{-1}(p, q)$ we can find some r_ε such that $d(r, r_\varepsilon) < \varepsilon/2$ and $\Delta(r_\varepsilon) = (p_\varepsilon, q_\varepsilon)$. Now r_ε must be supported on (a subset of) $K_\varepsilon \times H_\varepsilon$, which is itself compact, and which does not depend on r varying in $\Delta^{-1}(p, q)$. In other words, the whole $\Delta^{-1}(p, q)$ is contained within an $\varepsilon/2$-neighborhood of $P(K_\varepsilon \times H_\varepsilon)$. By compactness, for every $\varepsilon > 0$, $P(K_\varepsilon \times H_\varepsilon)$ can be covered by a finite number of balls of radius $\varepsilon/2$. Then $\Delta^{-1}(p, q)$ can be covered by a finite number of balls of radius ε, i.e. it is totally bounded. Since Δ is continuous, $\Delta^{-1}(p, q)$ is closed. Therefore $\Delta^{-1}(p, q)$ is compact.

(b) Again, we just need to show total boundedness. Since K is compact, for every $\varepsilon > 0$ there exists a finite $(\varepsilon/2)$-net $\{(p_n, q_n)\}$ covering K (i.e. every element $k \in K$ is within distance $\varepsilon/2$ from $\{(p_n, q_n)\}$). Take now the finite collection of sets $\{\Delta^{-1}(p_n, q_n)\}$. By (a), we know that they are all compact, and by Proposition 2.6.11 we know that every element $r \in \Delta^{-1}(K)$ is within distance $\varepsilon/2$ from some element of $\cup_n \Delta^{-1}(p_n, q_n)$. Now the set $\cup_n \Delta^{-1}(p_n, q_n)$ is a finite union of compact sets, so it is compact, and in particular it can be covered by finitely many balls of radius $\varepsilon/2$. This implies that for every $\varepsilon > 0$, the whole $r \in \Delta^{-1}(K)$ can be covered by finitely many balls of radius ε, i.e. it is totally bounded.

□

Corollary 2.6.14. *Given $p, q \in PX$, the set of coupling $\Gamma(p,q) = \Delta^{-1}(p,q)$ is compact. Therefore the infimum appearing in the Kantorovich duality formula is actually a minimum:*

$$\min_{r \in \Gamma(p,q)} \int_{X \times X} c(x,y)\, dr(x,y) = \sup_f \left(\int_X f dq - \int_X f\, dp \right). \tag{2.6.12}$$

3. Stochastic Orders

In this chapter we extend the Kantorovich monad of Chapter 2 to metric spaces equipped with a partial order. The order induced this way on the Wasserstein spaces will itself satisfy a form of Kantorovich duality.

The study of orders on spaces of probability measures induced by orders on the underlying space is of interest in many mathematical disciplines, and it is known under different names. In decision theory and in mathematical finance one talks of *first order stochastic dominance* of random variables [Fis80]. In probability theory, the common name is the *usual stochastic order* [Leh55, SS07]. Most of the theory, in this sense, is specifically for *real-valued* random variables, where the order is an answer to the question of *when a random variable is statistically larger than another one.* There are mainly three ways to define such an order: given two probability measures p, q on the same ordered space X,

(a) $p \leq q$ if and only if p assigns less measure than q to all upper sets;

(b) $p \leq q$ if and only if there exists a coupling entirely supported on the order relation

$$\{(x,y) \in X \times X | x \leq y\};$$

(c) $p \leq q$ if and only if for all monotone functions $f : X \to \mathbb{R}$ of a certain class (for example, continuous),

$$\int f\, dp \leq \int f\, dq.$$

A possible interpretation of the first condition is that *the mass of the measure p is overall placed lower in the order compared to q*. A possible interpretation of the second condition, in terms of optimal transport, is that there exists a transport plan from p to q such that no mass is moved lower in the order. These two approaches are in most cases proven to be equivalent by means of Strassen's theorem [Str65, Theorem 11]. An intepretation of the third condition is that for any choice of utility function compatible with the order, the expected utility with measure p will be less than the expected utility with measure q. The equivalence of this third approach to the other two has been long known in the literature for probability measures on $\mathbb{R}$. To the best of the authors' knowledge, it was first stated for general regular topological spaces by Edwards [Edw78]. In this chapter we show that, for a large class of spaces, this can be thought of as an instance of Kantorovich duality (see 3.3.1). While it is easy to see that the stochastic order over any partially ordered space is reflexive and transitive, antisymmetry seems to be a long-standing question [Law17, HLL18]. We will show in this work that antisymmetry indeed holds for a large class of metric spaces, including all Banach spaces (see 3.1.1).

From the point of view of categorical probability, the first probability monad on ordered spaces, and specifically on continuous domains, was defined by Jones and Plotkin [JP89], and called the *probabilistic powerdomain*. In more recent years, Keimel [Kei08] studied another probability monad for ordered spaces, the *Radon monad over compact ordered spaces*. He gave a complete characterization of its algebras, which are the compact convex subsets of locally convex topological vector spaces, with the order given by a closed positive cone.

In this chapter, we study the interplay between metric and order on ordered Wasserstein spaces. We show how to make the interpretation of the order in terms of "moving the mass upward" precise in terms of a colimit characterization of the order, generalizing a result of Lawson [Law17]. We also prove that the algebras for the ordered Kantorovich monad are exactly the closed convex subsets of Banach spaces, equipped with a closed positive cone. Moreover, we give a categorical characterization of *convex maps* between ordered convex spaces as exactly the *oplax* morphism of algebras.

Ordered metric spaces are closely related to *Lawvere metric spaces* [Law73, Law86], which are generalizations of metric spaces to asymmetric distances. Such objects already incorporate a partial order structure in terms of zero distances. A treatment of probability monads on Lawvere metric spaces, and the related Kantorovich duality theory, has been initiated by Goubault-Larrecq [GL17]. In this chapter we work with ordinary metric spaces; however, the duality theory and the interplay between metric and order can be interpreted in terms of Lawvere distances.

Outline.

- In Section 3.1 we define the categories of ordered metric spaces. In 3.1.1 we will give the definition of the usual stochastic order, and of ordered Wasserstein spaces.

- In Section 3.2 we will show that the ordered Wasserstein space satisfies a colimit characterization, or density result (Proposition 3.2.5), in analogy with the colimit characterization of unordered Wasserstein spaces given in 2.2.

- In Section 3.3 we define and study a particular class of ordered spaces, which we call *L-ordered spaces*, in which the order is compatible with the metric in a particular way. In 3.3.1 we show that this property allows to express the stochastic order in terms of Kantorovich duality (Theorem 3.3.3), and in 3.3.2 we prove, using Kantorovich duality, that the order is antisymmetric (Corollary 3.3.9).

- In Section 3.4 we will define and study the monad structure of the *ordered Kantorovich monad*. In 3.4.2 we prove (Theorem 3.4.6) that the formation of joints and marginals equips the ordered Kantorovich monad with a bimonoidal structure, just like in the unordered case (Section 2.5). In 3.4.3 we prove (Proposition 3.4.8) that the stochastic order satisfies a lifting property analogous to the metric lifting property of 2.6.2.

- In 3.5 we prove that the algebras of the ordered Kantorovich monad are precisely closed convex subsets of ordered Banach spaces (Theorem 3.5.6). The structure maps, as in the unordered case, are given by integration, and in 3.5.1 we show that these maps are strictly monotone, fully generalizing a result that is long known in the real-valued case (Proposition 3.5.11). In 3.5.2 we show that, if one considers the category of ordered metric spaces as a locally posetal 2-category, the Choquet adjunction (2.4.10) can be strengthened to an isomorphism of partial orders. In 3.5.3 we show, again using the 2-categorical approach, that the lax and oplax morphisms of algebras are precisely the concave and convex maps (Theorem 3.5.19).

- In Section 3.6 we define the "exchange law" as an even stronger compatibility condition between metric and order. We show that the spaces satisfying these property are necessarily L-ordered (Proposition 3.6.3), and we show that if a space X satisfies the exchange law, then its Wasserstein space does too (Proposition 3.6.5). This will be useful to study the orders that we encounter in Chapter 4.

Most of the material in this chapter will be part of a paper which is currently in preparation.

3.1. Ordered Wasserstein spaces

Definition 3.1.1. *An* ordered metric space *is a metric space X equipped with a partial order relation whose graph $\{\leq\} \subseteq X \otimes X$ is closed.*

The closure condition is a sort of continuity for the order relation: if we have sequences $\{x_i\}$ and $\{y_i\}$ in X tending to x and y, respectively, and such that $x_i \leq y_i$ for definitively all i, then necessarily $x \leq y$. Intuitively, the order can be *approximated* by sequences.

In analogy with the monoidal category Met from Section 2.1.1, we put:

Definition 3.1.2. *The symmetric monoidal category* OMet *has:*

- *As objects, ordered metric spaces;*
- *As morphisms, monotone, short maps;*
- *As monoidal structure* $\otimes$*, the* ℓ^1*-product, with the product order, and together with the obvious symmetric monoidal structure isomorphisms.*

There exists an essentially surjective forgetful functor $U : \mathsf{OMet} \to \mathsf{Met}$ with a left adjoint (the discrete order).

We are moreover interested in *complete* metric spaces.

Definition 3.1.3. *The category* COMet *is the full subcategory of* OMet *whose objects are ordered metric spaces which are complete as metric spaces.*

3.1.1. The stochastic order

Definition 3.1.4. *Let* $X \in \mathsf{OMet}$*. For any* $p, q \in PX$*, the* stochastic order *relation* $p \leq q$ *holds if and only if there exists a coupling of* p *and* q *entirely supported on the graph* $\{\leq\} \subseteq X \otimes X$*.*

This is a standard notion, see for example [HLL18]. A possible interpretation, as sketched in the introduction, is that the mass of p can be moved so as to form the distribution q in a way such that every unit of mass is only moved upwards in the order (if at all).

As sketched in the introduction, the stochastic order can be defined in several equivalent ways. The following equivalence result is a special case of [Kel84, Proposition 3.12], which holds even for arbitrary topological spaces equipped with a closed partial order[1].

Theorem 3.1.5 (Kellerer)**.** *Let* $X \in \mathsf{OMet}$*, and let* $p, q \in PX$*. Then* $p \leq q$ *if and only if* $p(C) \leq q(C)$ *for every closed upper set* $C \subseteq X$*.*

In contrast to Definition 3.1.1, transitivity of the order relation is immediate from this alternative characterization.

Upon applying Theorem 3.1.5 to the order itself and then again to the opposite order, it also follows that $p \leq q$ holds if and only if $p(U) \leq q(U)$ for all *open* upper sets U.

[1]Such a space is automatically Hausdorff [Nac65, Proposition 2].

3.2. Colimit characterization

Just as in the unordered case (Section 2.2), PX can be obtained as a colimit of spaces of finite sequences. Here we want to prove that the *order* structure of PX also arises in this way, as the closure of the order between the finite empirical sequences. A possible interpretation, which gives an additional characterization of the stochastic order, is the following: *$p \leq q$ if and only if p and q can be approximated arbitrarily well by empirical distributions of finite sequences $\{x_i\}$ and $\{y_i\}$, such that up to permutation, $x_i \leq y_i$ for all i, i.e. to obtain q from p each unit of mass is moved upward in the order.*

We construct the finite sequences in a functorial way in 3.2.1. We construct the empirical distribution map as a natural transformation in 3.2.2, and prove the order density result in (3.2.3).

3.2.1. Power functors

Let's first define the ordered version of the power functors of 2.2.1.

Definition 3.2.1. *Let $X \in \mathsf{COMet}$ and N be a finite set. We denote by X^N the N*-fold cartesian power*, or more briefly just* power*, of X:*

- *Its elements are functions $N \to X$, or equivalently tuples $(x_n)_{n\in N}$ of elements of X indexed by elements of N;*
- *Its metric is defined to be:*

$$d\big((x_n)_{n\in N}, (y_n)_{n\in N}\big) := \frac{1}{|N|} \sum_{n\in N} d(x_n, y_n);$$

- *Its order is the product order: $(x_n) \leq (y_n)$ if and only if $x_n \leq y_n$ for all $n \in N$.*

Given $X \in \mathsf{OMet}$, the powers X^- form again a functor $\mathsf{FinUnif}^{\mathrm{op}} \to \mathsf{OMet}$:

Proposition 3.2.2. *Let $\phi : M \to N$ be a map in $\mathsf{FinUnif}$, and consider the map $X^\phi : X^N \to X^M$ defined in 2.2.1. Then X^ϕ is an isometric* order *embedding.*

Proof. We know from Lemma 2.2.3 that X^ϕ is an isometric embedding. For the order part, first of all, $(x_{\phi(m)})_{m\in M} \leq (y_{\phi(m)})_{m\in M}$ if an only if for all $m \in M$, $x_{\phi(m)} \leq y_{\phi(m)}$. Since ϕ is surjective, this is equivalent to $x_n \leq y_n$ for all $n \in N$, which in turn means exactly that $(x_n)_{n\in N} \leq (y_n)_{n\in N}$. □

Since the forgetful functor $\mathsf{OMet} \to \mathsf{Met}$ is faithful, all these constructions are again natural. We then have a functor $(-)^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \otimes \mathsf{OMet} \to \mathsf{OMet}$, or by currying, equivalently we consider the functor $(-)^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \to [\mathsf{OMet}, \mathsf{OMet}]$. The curried functor is strongly monoidal, where the monoidal structure of the functor category $[\mathsf{OMet}, \mathsf{OMet}]$ is given by functor composition. If we restrict to *complete* ordered metric spaces, the powers are complete as well, and we get a strong monoidal functor $(-)^{(-)} : \mathsf{FinUnif}^{\mathrm{op}} \to [\mathsf{COMet}, \mathsf{COMet}]$.

3.2.2. Empirical distribution

Consider now the empirical distribution map defined in 2.2.2, mapping $(x_n)_{n\in N} \in X^N$ to the probability measure

$$\frac{1}{|N|} \sum_{n\in N} \delta_{x_n}.$$

We know that this assignment gives a short, natural map $i^N : X^N \to PX$, We want to show that it is monotone. Just as it is not an isometric embedding, but it is one up to permutation (formula (2.2.2) together with Proposition 2.2.10), we show that it is as well an order embedding up to permutation.

Lemma 3.2.3 (Splitting Lemma)**.** *Let $X \in \mathsf{OMet}$. Let $(x_n) \in X^N$ and $(y_m) \in X^M$. Then $i_N(x_n) \le i_M(y_m)$ if and only if there exist a set K and maps $\phi : K \to N$ and $\psi : K \to M$ in $\mathsf{FinUnif}$ such that $X^\phi(x_n) \le X^\psi(y_m)$.*

Proof. The homonymous statement in [GHK+03, Proposition IV-9.18] implies[2] in particular that for two finitely supported measures ("simple valuations") $\zeta = \sum_n r_n \delta_{x_n}$ and $\xi = \sum_m s_m \delta_{y_m}$, we have $\zeta \le \xi$ if and only if there exists a matrix of entries $t_{n,m} \in [0, \infty)$ such that:

(a) $t_{n,m} > 0$ only if $x_n \le y_m$;

(b) $\sum_m t_{n,m} = r_n$;

(c) $\sum_n t_{n,m} \le s_m$.

In our case, $\zeta := i_N(x_n)$ and $\xi := i_M(y_m)$ are normalized, so condition (c) can be strengthened to an equality. Since all r_n and s_m are rational, the $t_{n,m}$ can also be chosen to be rational if they exist. By finiteness, we can find a common denominator d for all its entries, so that the matrix $(t_{n,m})$ can be written as the empirical distribution of an

[2] The stochastic order considered there coincides with ours if one takes the topology on X to be given by the open upper sets of X.

element of $X^{M\otimes N\otimes D}$, where $|D| = d$. Therefore we can fix $K = M \otimes N \otimes D$. Conditions (b) and (c) together with naturality of the empirical distribution imply that we can find the desired maps ϕ and ψ, and condition (a) then says that $X^{\phi}(x_n) \leq X^{\psi}(y_m)$. $\square$

Corollary 3.2.4. *Let $X \in \mathsf{OMet}$. Let $(x_n), (y_n) \in X^N$. Then $i_N(x_n) \leq i_N(y_n)$ if and only if there exists a permutation $\sigma : N \to N$ such that for each $n \in N$, $x_n \leq y_{\sigma(n)}$.*

Proof. The "if" direction is clear. For "only if", we assume $i_N(x_n) \leq i_N(y_n)$. Then the matrix $(t_{n,m})$ constructed as in the proof of Lemma 3.2.3 is bistochastic, and therefore a convex combination of permutations by the Birkhoff–von Neumann theorem. Choosing any permutation which appears in such a convex combination works, thanks to property (a). $\square$

3.2.3. Order density

We are now finally ready to state an order-theoretical equivalent of Theorem 2.2.14. First we need a density result, which works for general metric spaces.

Proposition 3.2.5. *Let $X \in \mathsf{OMet}$. Let $p \leq q$ in PX. Then there exists a sequence $\{N_j\}_{j\in\mathbb{N}}$ in $\mathsf{FinUnif}$, and $\{\bar{p}_j\}, \{\bar{q}_j\}$ such that:*

- *$\bar{p}_j, \bar{q}_j \in X^{N_j}$ for all j;*
- *$i(\bar{p}_j) \to p$ and $i(\bar{q}_j) \to q$ in PX;*
- *$\bar{p}_j \leq \bar{q}_j$ in the order of X^{N_j} for all j.*

In other words, the order of PX is the closure of the order induced by the image of all the empirical distributions. Or equivalently, any two probability measures in stochastic order can be approximated arbitrarily closely by uniform finitely supported measures which are also stochastically ordered.

This result generalizes Lawson's recent [Law17, Theorem 4.8], who has also found applications of this type of result to generalizations of operator inequalities.

Proof. Consider the set

$$I(X) := \bigcup_{N\in\mathsf{FinUnif}} X_{|N|} \subseteq PX, \tag{3.2.1}$$

where $X_{|N|}$ is the quotient of X^N under permutations of the components, the "symmetrized power functor" of 2.2.1. This set is dense in PX, and we equip it with the smallest ordering relation which makes the canonical maps $X^N \to I(X)$ monotone; by

Lemma 3.2.3, this is equivalently the restriction of the stochastic order from PX to $I(X)$.

Let now $p, q \in PX$, and suppose $p \leq q$. By Corollary 3.1.5, there exists a joint r on $X \otimes X$ supported on $\{\leq\}$ with marginals p and q. Now consider $\{\leq\} \subseteq X^2$, and construct the subset $I(\{\leq\})$ in the same way as $I(X)$ was constructed for X. Again, the set $I(\{\leq\})$ is dense in $\{\leq\}$. This means that for every $\varepsilon > 0$, we can find a $\bar{r} \in I(\{\leq\})$ such that $d(r, \bar{r}) < \varepsilon$. Let now $\bar{p}, \bar{q}$ be the marginals of $\bar{r}$. Since the marginal projections are short (Proposition 2.5.11), $d(p, \bar{p}) < \varepsilon$ and $d(q, \bar{q}) < \varepsilon$. Moreover, again by Corollary 3.1.5, since $\bar{r}$ is supported on $\{\leq\}$, $\bar{p} \leq \bar{q}$. By taking ε smaller and smaller, we get the desired Cauchy sequence. $\square$

Corollary 3.2.6. *PX is the colimit of $X^{(-)} : \mathsf{FinUnif} \to \mathsf{COMet}$, with colimit components given by the empirical distribution maps $i_N : X^N \to PX$.*

Proof. We already know that PX is the colimit as a metric space. We only need to show that given any commutative cocone indexed by N, i.e. made up of triangles

$$\begin{tikzcd} X^N \arrow[r, "\phi"] \arrow[dr, "f_N"'] & X^M \arrow[d, "f_M"] \\ & Y \end{tikzcd} \tag{3.2.2}$$

where each cocone component f_N is monotone, then also the unique short map u in

$$\begin{tikzcd} X^N \arrow[d, "i"'] \arrow[dr, "f_N"] & \\ PX \arrow[r, dashed, "u"'] & Y \end{tikzcd} \tag{3.2.3}$$

is monotone. Now let $p \leq q$. By Proposition 3.2.5, we can find sequences $\{N_j\}$ in $\mathsf{FinUnif}$, and $\{\bar{p}_j\}, \{\bar{q}_j\}$ such that:

- $\bar{p}_j, \bar{q}_j \in X^{N_j}$ for all j;
- $i(\bar{p}_j) \to p$ and $i(\bar{q}_j) \to q$;
- $\bar{p}_j \leq \bar{q}_j$ in the order of X^{N_j} for all j.

Since u is short, it is in particular continuous. By the commutativity of (3.2.3),

$$u(p) = u\big(\lim_j i(\bar{p}_j)\big) = \lim_j u \circ i(\bar{p}_j) = \lim_j f_{N_j}(\bar{p}_j),$$

and just as well $u(q) = \lim_j f_{N_j}(\bar{q}_j)$. Now for all j, $\bar{p}_j \leq \bar{q}_j$, and since all the f_{N_j} are monotone, $f_{N_j}(\bar{p}_j) \leq f_{N_j}(\bar{q}_j)$. By the closure of the order on Y, we then have that

$$u(p) = \lim_j f_{N_j}(\bar{p}_j) \leq \lim_j f_{N_j}(\bar{q}_j) = u(q),$$

which means that u is monotone. $\square$

3.3. L-ordered spaces

In this section we study a stronger compatibility condition between the metric and the order. So far, we have required the order relation to be *closed*, which is a merely topological property. Instead, we now define a property that depends nontrivially *on the metric itself.*

Definition 3.3.1. *Let X be an ordered metric space. We say that X is* L-ordered *if for every $x, y \in X$ the following conditions are equivalent:*

- *$x \leq y$;*
- *for every* short, monotone *function $f : X \to \mathbb{R}$, $f(x) \leq f(y)$.*

The condition is similar to the following property of the metric, which all spaces have:

$$d(x, y) = \sup_{f:X\to\mathbb{R}} f(x) - f(y), \tag{3.3.1}$$

where the supremum is taken over all short maps. The intuition is that on L-ordered spaces, short functions, which are the functions that are enough to determine the metric, are also enough to determine the order.

For all ordered metric spaces, the first condition in Definition 3.3.1 implies the second. The converse does not always hold, as the following counterexample shows.

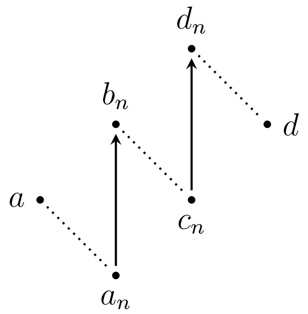
Example 3.3.2. Consider the space X containing four different sequences

$$\{a_n\}, \ \{b_n\}, \ \{c_n\}, \ \{d_n\},$$

and two extra points a, d with:

- $\{a_n\}$ tending to a, with $d(a_n, a) = \frac{1}{n}$ for all n;
- $\{d_n\}$ tending to d, with $d(d_n, d) = \frac{1}{n}$ for all n;
- $a_n \leq b_n$ for all n,
- $c_n \leq d_n$ for all n,
- $d(b_n, c_n) = \frac{1}{n}$ for all n (but the two sequences are not Cauchy);
- All other distances equal to 1;
- No points other than those indicated above are related by the order, in particular $a \not\leq d$.

With this definition, the only two nontrivial Cauchy sequences are $\{a_n\}$ and $\{d_n\}$, therefore the space is complete, and the order is closed, so $X \in \mathsf{COMet}$. We can sketch the space in the following picture, where the dotted lines are distances $1/n$, and the arrows denote the order:



Now consider a short, monotone function $X \to \mathbb{R}$. We have that:

$$\begin{aligned} f(a) &= \lim_{n\to\infty} f(a_n) \leq \lim_{n\to\infty} f(b_n) \\ &= \lim_{n\to\infty} f(c_n) \leq \lim_{n\to\infty} f(d_n) = f(d), \end{aligned}$$

however, $a \not\leq d$.

In any case, many ordered spaces of interest in mathematics are L-ordered, for example, all ordered Banach spaces (see Section 3.5).

We call $\mathsf{L\text{-}OMet}$ and $\mathsf{L\text{-}COMet}$ the full subcategories of OMet and COMet, respectively, whose spaces are L-ordered.

3.3.1. Kantorovich duality for the order structure

L-ordered spaces allow to study the order using Kantorovich duality. In particular, on an L-ordered space we have a dual characterization of the order in terms of duality to Lipschitz functions. We want to prove the following theorem.

Theorem 3.3.3. *Let $X \in \mathsf{L\text{-}COMet}$. Let $p, q \in PX$. Then $p \leq q$ if and only if for every short monotone map $f : X \to \mathbb{R}$,*

$$\int f\, dp \leq \int f\, dq. \tag{3.3.2}$$

We will prove the theorem using Kantorovich duality. As cost function, we use a quantity which is sensitive to the metric, as well as the order.

Definition 3.3.4. *Let X be an ordered metric space. We define the following quantity, which we call* L-distance*:*

$$d_L(x,y) := \sup_{f:X\to\mathbb{R}} (f(x) - f(y)), \tag{3.3.3}$$

where the supremum is taken over all short, monotone maps.

This quantity can be interpreted as a *Lawvere metric* compatible with the order (see [Law73, Law86], as well as the treatment in [GL17]). More intuitively, the L-distance is to short monotone maps as the usual distance is to short maps, as the following remark shows.

Remark 3.3.5. Let X and Y be ordered metric spaces, and let $f : X \to Y$ be short and monotone. Then

$$\begin{aligned} d_L\big(f(x), f(x')\big) &= \sup_{g:Y\to\mathbb{R}} (g(f(x)) - g(f(y))) \\ &\le \sup_{h:X\to\mathbb{R}} (h(x) - h(y)) = d_L(x,x'). \end{aligned}$$

Here are some useful properties satisfied by d_L, which make it suitable for Kantorovich duality.

Proposition 3.3.6. *Let X be an ordered metric space, not necessarily L-ordered. The L-distance satisfies the following properties:*

(a) For all $x, y \in X$ such that $x \le y$, we have $d_L(x,y) = 0$. In particular, $d_L(x,x) = 0$.

(b) If (and only if) X is L-ordered, $d_L(x,y) = 0$ implies $x \le y$ for all x, y in X.

(c) d_L satisfies the triangle inequality: for every $x, y, z \in X$,

$$d_L(x,z) \le d_L(x,y) + d_L(y,z).$$

So in particular d_L is a quasi-metric (not necessarily symmetric).

(d) d_L is bounded above by the metric: for all x, y in X, $d_L(x,y) \le d(x,y)$.

(e) d_L is lower-semicontinuous in both arguments.

Proof of Proposition 3.3.6.

(a) If $x \le y$, then for all short monotone functions f, we have $f(x) - f(y) \le 0$. The supremum is attained by $f = 0$.

(b) Suppose that X is L-ordered. If

$$d_L(x, y) = \sup_{f: X \to \mathbb{R}} (f(x) - f(y)) = 0,$$

for all short, monotone maps $f : X \to \mathbb{R}$,

$$f(x) - f(y) \leq 0,$$

which means $f(x) \leq f(y)$. Since X is L-ordered, then $x \leq y$.

Suppose now that X is *not* L-ordered. Then there exist $x \not\leq y$ such that for all short monotone $f : X \to \mathbb{R}$, $f(x) \leq f(y)$. But then

$$d_L(x, y) = \sup_{f: X \to \mathbb{R}} (f(x) - f(y)) \leq 0,$$

and again the supremum is attained by $f = 0$.

(c) Let $x, y, z \in X$. Then

$$\begin{aligned} d_L(x, z) &= \sup_{f: X \to \mathbb{R}} (f(x) - f(z)) \\ &= \sup_{f: X \to \mathbb{R}} (f(x) - f(y) + f(y) - f(z)) \\ &\leq \sup_{f: X \to \mathbb{R}} (f(x) - f(y)) + \sup_{f': X \to \mathbb{R}} (f'(y) - f'(z)) \\ &= d_L(x, y) + d_L(y, z). \end{aligned}$$

(d) For all $x, y \in X$,

$$\begin{aligned} d_L(x, y) &= \sup\{(f(x) - f(y)),\ f \text{ short and monotone}\} \\ &\leq \sup\{(f(x) - f(y)),\ f \text{ short}\} = d(x, y). \end{aligned}$$

(e) d_L is defined as a pointwise supremum of continuous functions, therefore it is lower-semicontinuous. □

We are now ready to prove the theorem:

Proof of Theorem 3.3.3. Let $p, q \in PX$. Suppose that for all short, monotone $f : X \to \mathbb{R}$,

$$\int f \, dp \leq \int f \, dq,$$

or in other words,

$$\sup_{f:X\to\mathbb{R}} \left(\int f\,dp - \int f\,dq \right) = 0,$$

where the supremum is taken over all short, monotone maps. Short monotone maps are precisely those that are bounded by d_L, which is lower-semicontinuous and satisfies the triangle inequality by Proposition 3.3.6. Therefore we can apply Kantorovich duality (Corollary 2.1.9) to obtain:

$$0 = \sup_{f:X\to\mathbb{R}} \left(\int f\,dp - \int f\,dq \right) = \min_{r\in\Gamma(p,q)} \int_{X\otimes X} d_L(x,y)\,dr(x,y)$$

where the minimizing r exists (Corollary 2.6.14). In other words, there exists a coupling r entirely supported on

$$\{d_L(x,y) = 0\}.$$

Since X is L-ordered, all the points in the set above are contained in $\{\leq\}$. So r is supported on $\{\leq\}$, which means that $p \leq q$.

Conversely, if such a coupling r exists, then again by Kantorovich duality,

$$\sup_{f:X\to\mathbb{R}} \left(\int f\,dp - \int f\,dq \right) = 0.$$

□

From this characterization it is easy to see that the order is closed and transitive. Antisymmetry will be proven shortly.

Corollary 3.3.7. *Let X be an L-ordered metric space. Then PX is L-ordered too.*

Proof. Given a short, monotone map $f : X \to \mathbb{R}$, the assignment

$$p \mapsto \int f\,dp$$

is short and monotone as a map $PX \to \mathbb{R}$ By Theorem 3.3.3, this determines the order. Therefore PX is L-ordered. □

3.3.2. Antisymmetry

Here we prove that the stochastic order on any L-ordered space is a partial order, i.e. it is antisymmetric. It is apparently an open question whether antisymmetry holds over every order metric space. The property is known to be true for compact spaces [Edw78], and for particular cones in Banach spaces [HLL18].

For L-ordered spaces, we can prove antisymmetry using a Kantorovich duality argument, encoded in the following statement.

Proposition 3.3.8. *Let X be an L-ordered metric space. Let $p, q \in PX$, and suppose that $p < q$ strictly. Then there exists a short monotone $f : X \to \mathbb{R}$ such that*

$$\int f\, dq > \int f\, dp \quad \textit{strictly.}$$

Proof. Suppose that $p \leq q$ but $p \neq q$. Then there exists a coupling r supported on the relation $\{\leq\}$, which cannot be supported only on the diagonal $D := \{(x, x)\}$. In other words, there exists a point $(\bar{x}, \bar{y})$ in the support of r with $\bar{x} < \bar{y}$ strictly, and every open neighborhood of $(\bar{x}, \bar{y})$ has strictly positive measure. Since X is L-ordered, and since $\bar{y} \not\leq \bar{x}$, there exists a short, monotone map $f : X \to \mathbb{R}$ such that $f(\bar{y}) > f(\bar{x})$ strictly. We can then choose an open neighborhood U of $(\bar{x}, \bar{y})$ which is disjoint from the diagonal, and on which the function

$$(x, y) \mapsto f(y) - f(x)$$

is strictly positive. Therefore,

$$\begin{aligned} \int f\, dq - \int f\, dp &= \int_{X \otimes X} \big(f(y) - f(x)\big)\, dr(x, y) \\ &\geq \int_U \big(f(y) - f(x)\big)\, dr(x, y) > 0 \end{aligned}$$

strictly, which in turn means that

$$\int f\, dq > \int f\, dp.$$

□

Corollary 3.3.9. *Let X be an L-ordered metric space. Then the stochastic order on PX is antisymmetric.*

Proof of the Corollary. Let $p, q \in PX$, and suppose that both $p \leq q$ and $q \leq p$ in the stochastic order. Then necessarily

$$\int f\, dp = \int f\, dq$$

for all short monotone maps $f : X \to \mathbb{R}$. By Proposition 3.3.8, then, it must be that $p = q$. □

3.4. The ordered Kantorovich monad

In Chapter 2 we showed that in the unordered case, P carries a monad structure, whose algebras are the closed convex subsets of Banach spaces. Here we show that the monad structure can be lifted to the category $\mathsf{L\text{-}COMet}$. The easiest way to do this is to show that all the structure maps are monotone between the respective orders, so that the commutativity of the necessary diagrams is inherited from CMet. This will be done in 3.4.1. In the rest of the section, we will study the algebras and their morphisms, prove some of their general properties, and show that P is a bimonoidal monad as in the unordered case (proven in Section 2.5).

3.4.1. Monad structure

First of all, by Corollary 3.3.7, if $X \in \mathsf{L\text{-}COMet}$, then $PX \in \mathsf{L\text{-}COMet}$ too.

We will now lift the Kantorovich monad to $\mathsf{L\text{-}COMet}$. To do this, we have to:

(a) Show that if $f : X \to Y$ is monotone, then also $Pf : PX \to PY$ is monotone.

(b) Show that the structure transformations have components $\delta : X \to PX$ and $E : PPX \to PX$ which are monotone.

The commutativity of all relevant diagrams involved is obvious, since the forgetful functor $\mathsf{L\text{-}COMet} \to \mathsf{CMet}$ is faithful.

Proposition 3.4.1. *Let $f : X \to Y$ (short, monotone). Then $Pf : PX \to PY$ is also monotone.*

Proof. Let $C \subseteq Y$ be a closed upper set. We have to prove that

$$(f_*p)(C) \le (f_*q)(C),$$

which means

$$p(f^{-1}(C)) \le q(f^{-1}(C)).$$

Now since f is continuous, $f^{-1}(C)$ is closed. Since f is monotone, $f^{-1}(C)$ is an upper set. By definition of the order on PX, $p(C') \le q(C')$ for all upper closed sets C'. Therefore $(f_*p)(C) \le (f_*q)(C)$. □

Hence P is indeed an endofunctor of $\mathsf{L\text{-}COMet}$.

To prove the monotonicity of the structure maps, we will use the dual characterization of the order in terms of monotone short maps.

Proposition 3.4.2. *Let $X \in \mathsf{L\text{-}COMet}$. Then;*

(a) $\delta : X \to PX$ is an order embedding;

(b) $E : PPX \to PX$ is monotone.

Proof. (a) Let $x \leq y \in X$, and let $f : X \to \mathbb{R}$ (short, monotone). Then

$$\int_X f \, d\delta(x) = f(x) \leq f(y) = \int_X f \, d\delta(y).$$

Therefore $\delta(x) \leq \delta(y)$.

(b) Let $\mu \leq \nu \in PPX$, and again $f : X \to \mathbb{R}$ (short, monotone). By definition, the assignment

$$p \mapsto \int_X f \, dp$$

is monotone as a function $PX \to \mathbb{R}$. Therefore we can write

$$\int_X f \, d(E\mu) = \int_{PX} \left(\int_X f \, dp \right) d\mu(p) \leq \int_{PX} \left(\int_X f \, dp \right) d\nu(p) = \int_X f \, d(E\nu).$$

We conclude that again $E\mu \leq E\nu$. In conclusion, E is monotone. □

We therefore obtain:

Corollary 3.4.3. *(P, δ, E) is a monad on* $\mathsf{L\text{-}COMet}$ *lifting the Kantorovich monad on* CMet.

We will call this monad with the same name whenever this does not cause confusion.

3.4.2. Monoidal structure

We have seen in Section 2.5 that the Kantorovich monad on CMet has a bimonoidal structure which we can interpret in terms of forming joints and marginals. We now extend this structure to $\mathsf{L\text{-}COMet}$.

Just like for the monad structure, it suffices to show that its structure maps $\nabla : PX \otimes PY \to P(X \otimes Y)$ and $\Delta : P(X \otimes Y) \to PX \otimes PY$ are monotone.

Lemma 3.4.4. *Let $X, Y \in$ $\mathsf{L\text{-}COMet}$. Then $\nabla : PX \otimes PY \to P(X \otimes Y)$ is monotone.*

Proof. First of all, let $f : X \otimes Y \to \mathbb{R}$ be monotone, and let $p \in PX$. Then the function

$$\left(\int_X f(x, -) \, dp(x) \right) : Y \to \mathbb{R} \tag{3.4.1}$$

is monotone as well.

Suppose now that $p \leq p'$ and $q \leq q'$. Let $f : X \otimes Y \to \mathbb{R}$ be monotone. Then using the remark above,

$$\begin{aligned}
\int_{X\otimes Y} f(x,y)\, d(p\otimes q)(x,y) &= \int_X \left(\int_Y f(x,y)\, dp(x) \right) dq(y) \\
&\leq \int_X \left(\int_Y f(x,y)\, dp(x) \right) dq'(y) \\
&= \int_Y \left(\int_X f(x,y)\, dq'(y) \right) dp(x) \\
&\leq \int_Y \left(\int_X f(x,y)\, dq'(y) \right) dp'(x) \\
&= \int_{X\otimes Y} f(x,y)\, d(p'\otimes q')(x,y).
\end{aligned}$$

□

Lemma 3.4.5. *Let $X, Y \in \mathsf{L\text{-}COMet}$. Then $\Delta : P(X \otimes Y) \to PX \otimes PY$ is monotone.*

Proof. First of all, notice that if $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$ are monotone, then $(f+g) : X \otimes Y \to \mathbb{R}$ given by $(x,y) \mapsto f(x) + g(y)$ is monotone.

Suppose now that $p \leq q$. Let $f : X \to \mathbb{R}$ be monotone. So it is also monotone as a function $X \otimes Y \to \mathbb{R}$. This means that

$$\int_{X\otimes Y} f(x)\, dp(x,y) \leq \int_{X\otimes Y} f(x)\, dq(x,y),$$

but we can replace both terms as

$$\int_X f(x)\, dp_X(x) \leq \int_X f(x)\, dq_X(x),$$

so $p_X \leq q_X$. The same is true for Y, so $\Delta(p) \leq \Delta(q)$. □

Together with the results of Section 2.5, we get as a corollary:

Theorem 3.4.6. *P is a symmetric bimonoidal monad on $\mathsf{L\text{-}COMet}$.*

This in particular implies that $\Delta \circ \nabla = \mathrm{id}$ (see Proposition 1.2.2 and Corollary 2.5.18) is an order embedding in addition to being a metric embedding.

3.4.3. Order lifting

The map E admits a lifting criterion for the partial order, analogous to the one for the metric that we saw in 2.6.2.

We can prove it using the same technique as in 2.6.2, by first starting with finite sequences.

Proposition 3.4.7. *Let* $\bar{\mu} = ((\bar{\mu}_{m,n})_{m\in M})_{n\in N} \in (X^M)^N$ *and* $\bar{q} = (\bar{q}_{m,n})_{m\in M, n\in N} \in X^{MN}$*. Suppose that*

$$i \circ E^{M,N}(\bar{\mu}) \leq i(\bar{q}) \tag{3.4.2}$$

for the order in PX*, where* $i : X^{MN} \to PX$ *denotes the empirical distribution. Then there exists* $\bar{\nu} \in (X^M)^N$ *such that* $E^{M,N}(\bar{\nu}) = (\bar{q}) \circ \sigma$ *for some permutation* $\sigma \in S_{MN}$*, and* $\bar{\mu} \leq \bar{\nu}$ *for the order in* $(X^M)^N$*.*

Proof. By Corollary 3.2.4, condition (3.4.2) is equivalent to say that there exists a $\bar{\sigma} \in S_{MN}$ such that for all $(m, n) \in MN$,

$$\mu_{m,n} \leq q_{\bar{\sigma}(m,n)}. \tag{3.4.3}$$

Let now

$$\bar{\nu} := ((\bar{\nu}_{m,n})) := ((q_{\bar{\sigma}(m,n)})).$$

Then (3.4.3) implies that $\bar{\mu} \leq \bar{\nu}$. ☐

To go from finite sequences to PX the order density result given by Proposition 3.2.5.

Proposition 3.4.8 (Order lifting)**.** *Let* $X \in \mathsf{COMet}$*. Let* $\mu \in PPX$ *and* $q \in PX$*, such that* $E\mu \leq q$*. Then there exists* $\nu \in PPX$ *such that* $E\nu = q$ *and* $\mu \leq \nu$*.*

Proof. By Proposition 3.2.5, there exist sequences $M_j, N_j \in \mathsf{FinUnif}$ and $\{\bar{p}_j\}, \{\bar{q}_j\}$, such that:

- $\bar{p}_j, \bar{q}_j \in X^{M_j N_j}$ for all j;
- $i(\bar{p}_j) \to E\mu$ and $i(\bar{q}_j) \to q$;
- $\bar{p}_j \leq \bar{q}_j$ in the order of X^{N_j} for all j.

Now by Proposition 2.6.6, we can find a sequence $\{\bar{\mu}_j\}$ such that $\bar{\mu}_j \in (X^{M_j})^{N_j}$ for all j, with $E^{M_j,N_j}(\bar{\mu}_j) = \bar{p}_j$, and such that $i(\bar{\mu}_j) \to \mu$. By the third condition above together with Proposition 3.4.7, we have a sequence $\{\bar{\nu}_j\}$ such that $\bar{\nu}_j \in (X^{M_j})^{N_j}$, $E^{M,N}(\bar{\nu}) = (\bar{q}) \circ \sigma$ for some permutation $\sigma \in S_{MN}$, and $\bar{\mu}_j \leq \bar{\nu}_j$ for all j. By Corollary 2.6.8, we know that $\{i(\bar{\nu}_j)\}$ admits an accumulation point $\nu \in PPX$, so we can find a subsequence $\{\bar{\nu}_{j_k}\}$ converging to ν. Now by continuity of the order,

$$\bar{\mu}_{j_k} \leq \bar{\nu}_{j_k} \quad \Rightarrow \quad \mu \leq \nu,$$

and by continuity of E:

$$\begin{aligned} E\nu &= \lim_{k\to\infty} E \circ i(\bar{\nu}_{j_k}) \\ &= \lim_{k\to\infty} i \circ E^{M_{j_k},N_{j_k}}(\bar{\nu}_{j_k}) \\ &= \lim_{k\to\infty} i(\bar{q}_{j_k}) = q. \end{aligned}$$

□

3.5. Ordered algebras

We have seen in 2.4.3 that the algebras of the Kantorovich monad on CMet are exactly closed convex subsets of Banach spaces, where the algebra map $e : PA \to A$ maps every probability measure to its barycenter. We show that this implies that the algebras of the ordered Kantorovich monad P on $\mathsf{L\text{-}COMet}$ can be identified with *ordered* closed convex subsets of Banach spaces, for which the algebra map $e : PA \to A$ is monotone; and the morphisms of algebras are then just the *monotone* short affine maps.

Lemma 3.5.1. *Let $A \in \mathsf{L\text{-}COMet}$ be an algebra of the unordered Kantorovich monad via an algebra map $e : PA \to A$. Then e is monotone if and only if for all $a, b, c \in A$ and $\lambda \in [0, 1]$,*

$$a \leq b \quad \Rightarrow \quad e(\lambda\delta_a + (1-\lambda)\delta_c) \ \leq \ e(\lambda\delta_b + (1-\lambda)\delta_c).$$

This result is the ordered analogue of the equivalence of (a) and (d) in Theorem 2.4.2. The condition is the defining property of an *ordered barycentric algebra* [Kei08].

Proof. The assumption $a \leq b$ implies $\lambda\delta_a + (1-\lambda)\delta_c \leq \lambda\delta_b + (1-\lambda)\delta_c$ in PA. Therefore if e is monotone, the conclusion follows.

Conversely, suppose that the above implication holds. In order to prove that e is monotone, the density result of Proposition 3.2.5 shows that it is enough to prove $e(i_N(x_n)) \leq e(i_N(y_n))$ for $(x_n), (y_n) \in X^N$ with $i(x_n) \leq i(y_n)$. By Corollary 3.2.4, we can relabel (y_n) such that $x_n \leq y_n$ for every $n \in N$. Writing $N = \{1, \dots, |N|\}$, we therefore have

$$e\left(\frac{1}{|N|}\sum_{i=1}^{k}\delta_{x_i} + \frac{1}{|N|}\sum_{i=k+1}^{|N|}\delta_{y_i}\right) \leq e\left(\frac{1}{|N|}\sum_{i=1}^{k-1}\delta_{x_i} + \frac{1}{|N|}\sum_{i=k}^{|N|}\delta_{y_i}\right)$$

as an instance of the assumption, for every $k = 1, \dots, |N|$. Chaining all these inequalities results in the claimed $e(i_N(x_n)) \leq e(i_N(y_n))$. □

So if we represent A as a closed convex subset of a Banach space, then e is monotone if and only if

$$a \leq b \quad \Rightarrow \quad \lambda\, a + (1-\lambda)\, c \;\leq\; \lambda\, b + (1-\lambda)\, c \tag{3.5.1}$$

holds for all $a, b, c \in A$ and $\lambda \in [0,1]$, since the right-hand side is exactly $e(\lambda\delta_a + (1-\lambda)\delta_b) \leq e(\lambda\delta_b + (1-\lambda)\delta_c)$.

We will prove in 3.5.1 that when the map e is monotone, then it is even *strictly* monotone.

Monotonicity of e turns the algebra A, which is a subset of a Banach space, into a subset of an *ordered* Banach space, in analogy to what happens with ordered algebras of the Radon monad [Kei08].

Definition 3.5.2. *An* ordered Banach space *is a Banach space equipped with a closed positive cone.*

We already know that every convex subset A of a Banach space is a P-algebra in CMet, with the structure map given by integration, and we know that integration is monotone. In order for A to be P-algebra in $\mathsf{L\text{-}COMet}$, what remains to be checked is that A is indeed an object of $\mathsf{L\text{-}COMet}$, i.e. it is L-ordered. This is guaranteed by the Hahn-Banach theorem, which even shows that we can test the order using only *affine* short monotone maps:

Proposition 3.5.3. *Let B be an ordered Banach space. Let $a, b \in B$. Then $a \leq b$ if and only if for every short monotone* linear *functional $h : B \to \mathbb{R}$, $h(a) \leq h(b)$.*

Proof. Let B be a Banach space equipped with a closed positive cone B^+, let $a, b \in B$, and suppose that $a \not\leq b$. This means that the point $v := b - a$ does *not* lie in the cone B^+. Since $\{v\} \subseteq B$ is in particular compact and $B^+ \subseteq B$ is closed and convex, by the Hahn-Banach separation theorem there exists a bounded linear functional $h : B \to \mathbb{R}$ such that

(a) $h(c) \geq 0$ for all $c \in B^+$, and

(b) $h(v) < 0$ strictly.

Without loss of generality, we can assume that h has norm one, so that it is short. Property (a) means exactly that h is monotone. By linearity, property (b) means exactly that $h(a) > h(b)$. We have found a short, monotone, affine map $h : B \to \mathbb{R}$ such that $h(a) > h(b)$. □

Corollary 3.5.4. *Every ordered Banach space is L-ordered.*

Corollary 3.5.5. *Every closed convex subset of an ordered Banach space is a P-algebra in* $\mathsf{L\text{-}COMet}$.

Here is the converse statement:

Theorem 3.5.6. *Every P-algebra in* $\mathsf{L\text{-}COMet}$ *is i to a closed convex subset of an ordered Banach space.*

The proof follows that of the analogous result for ordered barycentric algebras [Kei08, Proposition 3.3]. We also need the following technical result about the L-distance on a P-algebra.

Lemma 3.5.7. *Let A be a P-algebra. Let $x, y, z \in A$ and $\alpha \in [0,1]$. Then*

$$d_L\big(\alpha\, x + (1-\alpha)\, z, \alpha\, y + (1-\alpha)\, z\big) = \alpha\, d_L(x,y). \tag{3.5.2}$$

Proof of Lemma 3.5.7. The proof works along the lines of [CF13, Lemma 8]. We know by Remark 3.3.5 that since e is short and monotone we have

$$d_L\big(\alpha\, x + (1-\alpha)\, z, \alpha\, y + (1-\alpha)\, z\big) \le \alpha\, d_L(x,y).$$

Now by setting $y = z$ we get that

$$d_L\big(\alpha\, x + (1-\alpha)\, y, y\big) \le \alpha\, d_L(x,y),$$

and by setting instead $x = z$, we get

$$d_L\big(x, \alpha\, y + (1-\alpha)\, x\big) \le \alpha\, d_L(x,y).$$

By the triangle inequality,

$$\begin{aligned} d_L(x,y) &\le d_L\big(x, \alpha\, x + (1-\alpha)\, y\big) + d_L\big(\alpha\, x + (1-\alpha)\, y, y\big) \\ &\le d_L\big(x, \alpha\, x + (1-\alpha)\, y\big) + \alpha\, d_L\big(x, y\big) \end{aligned}$$

$$\leq (1-\alpha)\, d_L\big(x,y\big) + \alpha\, d_L\big(x,y\big) = d_L(x,y),$$

so all three inequalities are actually equalities. In particular,

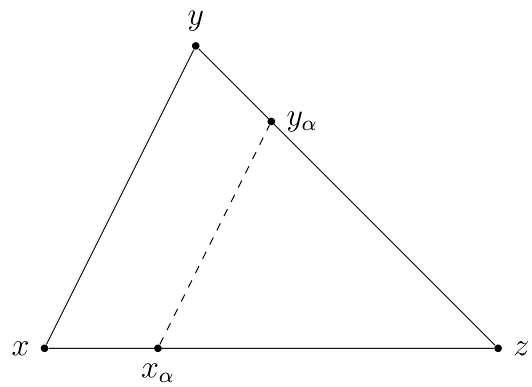
$$d_L\big(x, \alpha\, x + (1-\alpha)\, y\big) + d_L\big(\alpha\, x + (1-\alpha)\, y, y\big) = d_L\big(x, \alpha\, x + (1-\alpha)\, y\big) + \alpha\, d_L\big(x,y\big)$$

implies $d_L\big(\alpha\, x + (1-\alpha)\, y, y\big) = \alpha\, d_L\big(x,y\big)$, and

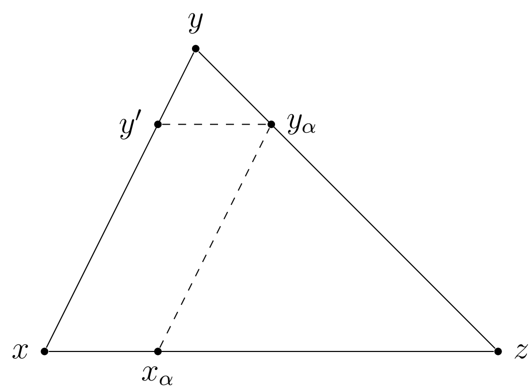
$$d_L\big(x, \alpha\, x + (1-\alpha)\, y\big) + \alpha\, d_L\big(x,y\big) = (1-\alpha)\, d_L\big(x,y\big) + \alpha\, d_L\big(x,y\big)$$

implies $d_L\big(x, \alpha\, x + (1-\alpha)\, y\big) = (1-\alpha)\, d_L\big(x,y\big)$.

d_L is then "affine on lines", or "longitudinally translation-invariant". Let's draw a picture to illustrate. Denote $\alpha\, x + (1-\alpha)\, z$ by x_α and $\alpha\, y + (1-\alpha)\, z$ by y_α. We can represent the situation as:



where by what we have proven above, $d_L(x_\alpha, z) = \alpha\, d_L(x,z)$, and $d_L(y_\alpha, z) = \alpha\, d_L(y,z)$. We have to prove that $d_L(x_\alpha, y_\alpha) = \alpha\, d_L(x,y)$. Consider now the point $y' := \alpha\, y + (1-\alpha)\, x$, which forms a parallelogram with x, x_α and y_α:



If we proved that d_L is translation invariant, then we would conclude that $d(x_\alpha, y_\alpha) = d(x, y') = \alpha\, d(x,y)$, which is the assert of the theorem.

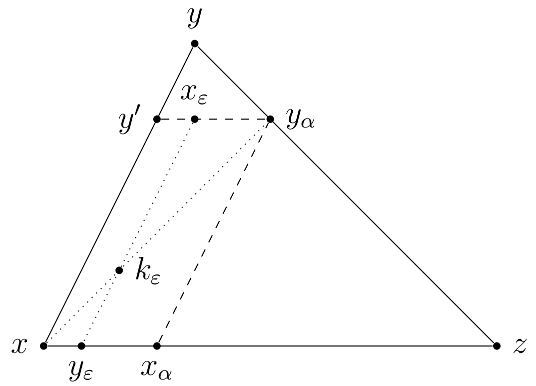
Now for $\varepsilon \in (0,1)$ consider the points

$$x_\varepsilon := \varepsilon\ x_\alpha + (1-\varepsilon)\, x, \quad y_\varepsilon := \varepsilon\ y_\alpha + (1-\varepsilon)\, y,$$

and

$$k_e := (1-\varepsilon)\, x_\varepsilon + \varepsilon\ y_\varepsilon = \varepsilon\ y_\alpha + (1-\varepsilon)\, x,$$

which can represent as:



We have that, by monotonicity of e,

$$d_L(x_\varepsilon, k_\varepsilon) = d_L\big(\varepsilon\ x_\alpha + (1-\varepsilon)\, x, \varepsilon\ y_\alpha + (1-\varepsilon)\, x\big) \leq \varepsilon\ d_L(x_\alpha, y_\alpha).$$

Moreover, since k_ε is on the same line of x_ε and y_ε,

$$d_L(x_\varepsilon, k_\varepsilon) = \varepsilon\ d_L(x_\varepsilon, y_\varepsilon).$$

Therefore,

$$d_L(x_\varepsilon, y_\varepsilon) = \varepsilon^{-1}\ d_L(x_\varepsilon, k_\varepsilon) \leq d_L(x_\alpha, y_\alpha).$$

By taking the limit $\varepsilon \to 0$, we then get

$$d_L(x, y) \leq d_L(x_\alpha, y_\alpha),$$

and by symmetry we also have the opposite inequality. So $d_L(x, y) = d_L(x_\alpha, y_\alpha)$. □

We can now prove the theorem.

Proof of Theorem 3.5.6. By what we already know, it is enough to show that if B is a Banach space and $A \subseteq B$ is a closed convex subset equipped with a closed partial order, then we can equip B itself with a closed partial order that restricts to the given order on A. So let $x \in B$ be considered positive if it is of the form $\lambda(y_+ - y_-)$ for $\lambda \geq 0$ and $y_+ \geq y_-$ in A. Using the fact that taking convex combinations in A is monotone, it is easy to see that this defines a convex cone. Taking $x \geq y$ if and only if $x - y$ is in the cone recovers the original order, since $x - y = \lambda(z_+ - z_-)$ for $z_+ \geq z_-$ in A and $\lambda > 0$ implies $\frac{1}{1+\lambda}x + \frac{\lambda}{1+\lambda}z_- = \frac{1}{1+\lambda}y + \frac{\lambda}{1+\lambda}z_+$. Together with $z_+ \geq z_-$, we hence obtain $x \leq y$ from the general theory of ordered topological barycentric algebras [Kei08, Corollary 4.2].

We cannot assume that the cone in B defined this way is closed, so we take its closure. To check that the resulting embedding is still an order embedding, then we have to show that the order of A already contains all the inequalities that that are added by taking the closure of the cone. In other words, we have to prove that whenever the sequence $\lambda_n(z_{+n} - z_{-n})$ for some $\lambda_n \geq 0$ and $z_{+n} \geq z_{-n} \in A$ and tends to $y - x$, then $x \leq y$ for the order of A. So suppose that

$$d(\lambda_n(z_{+n} - z_{-n}), y - x) \to 0,$$

or, rewriting everything in terms of only convex combinations (elements of A),

$$\frac{1}{\alpha_n} d\big(\alpha_n\, x + (1-\alpha_n)\, z_{+n}, \alpha_n\, y + (1-\alpha_n)\, z_{-n}\big) \to 0, \tag{3.5.3}$$

with $\alpha_n = \frac{1}{1+\lambda_n}$. Now consider the L-distance on A. We have from Lemma 3.5.7 and the triangle inequality for d_L that

$$\begin{aligned}
d_L(x,y) &= \frac{1}{\alpha_n} d_L\big(\alpha_n\, x + (1-\alpha_n)\, z_{+n}, \alpha_n\, y + (1-\alpha_n)\, z_{+n}\big) \\
&\leq \frac{1}{\alpha_n} d_L\big(\alpha_n\, x + (1-\alpha_n)\, z_{+n}, \alpha_n\, y + (1-\alpha_n)\, z_{-n}\big) \\
&\quad + \frac{1}{\alpha_n} d_L\big(\alpha_n\, y + (1-\alpha_n)\, z_{-n}, \alpha_n\, y + (1-\alpha_n)\, z_{+n}\big) \\
&= \frac{1}{\alpha_n} d_L\big(\alpha_n\, x + (1-\alpha_n)\, z_{+n}, \alpha_n\, y + (1-\alpha_n)\, z_{-n}\big) \\
&\qquad + \frac{1}{\alpha_n} (1-\alpha_n)\, d_L\big(z_{-n}, z_{+n}\big) \\
&= \frac{1}{\alpha_n} d_L\big(\alpha_n\, x + (1-\alpha_n)\, z_{+n}, \alpha_n\, y + (1-\alpha_n)\, z_{-n}\big) + 0,
\end{aligned}$$

since $z_{-n} \leq z_{+n}$. Since the L-distance on A is bounded above by the usual distance, the expression above is bounded by the quantity (3.5.3), which by assumption tends to zero, so necessarily $d_L(x,y) = 0$. Since A is L-ordered, then necessarily (Proposition 3.3.6) we have that $x \leq y$. □

In the unordered case, the P-morphisms are the short affine maps, i.e. the short maps which respect convex combinations. In the ordered case, they are additionally required to be monotone. Overall, we therefore have:

Theorem 3.5.8. *For P the ordered Kantorovich monad on* L-COMet*, the category of P-algebras is equivalent to the category of closed convex subsets of ordered Banach spaces with short affine monotone maps.*

We will refer to P-algebras in $\mathsf{L\text{-}COMet}$ as *ordered P-algebras.* These of course include those with trivial order.

Just as for the unordered case of 2.4.3, we have a natural bijection

$$\mathsf{L\text{-}COMet}(X, A) \cong \mathsf{L\text{-}COMet}^P(PX, A) \tag{3.5.4}$$

which we can interpret now as the fact that Choquet theory restricts to *monotone* maps. Without mentioning monads, it means the following: *given an L-ordered metric space X and a an ordered Banach space (or a closed, convex subset thereof) A, there is a bijection between short monotone maps $X \to A$, and affine, monotone maps $PX \to A$.*

Equivalently, it means the following.

Corollary 3.5.9. *Let X be L-ordered, and A be a P-algebra. Let $\tilde{f} : PX \to A$ be short and affine, but not necessarily monotone. Then $\tilde{f}$ is monotone (for the usual stochastic order) if and only if it is the affine extension of a monotone function.*

We can also give an explicit proof of the corollary, which can help the intuition.

Proof. First of all, we know by Proposition 3.4.1 that if f is monotone, then Pf is also monotone. Composing with e, which is monotone (since we are considering ordered algebras), gives a monotone map $e \circ (Pf) = \tilde{f}$.

Conversely, suppose that $\tilde{f}$ is monotone for the usual stochastic order. Then $f \circ \delta$ is monotone as well, since δ is an order embedding. □

We can also interpret the correspondence in terms of the dual system of Definition 2.1.6:

Remark 3.5.10. Let $X \in \mathsf{L\text{-}COMet}$. Short monotone functions $X \to \mathbb{R}$ form a convex cone in $\mathrm{Lip}(X)$, which we denote $C_{\geq}$. The stochastic order on PX induces a cone in $M(X)$ which is the *dual cone* $(C_{\geq})^*$ of $C_{\geq}$. In other words, we have an *ordered* equivalent of the duality:

- If $\int f \, d\mu \geq 0$ for all $f \in C_{\geq}$, then $\mu \in (C_{\geq})^*$;
- If $\int f \, d\mu \geq 0$ for all $\mu \in (C_{\geq})^*$, then $f \in C_{\geq}$.

3.5.1. The integration map is strictly monotone

For real random variables, it is well-known that if $p < q$ strictly, then $e(p) < e(q)$ strictly [Fis80, Theorem 1]. The interpretation is that *if one moves a nonzero amount of mass upwards in the order, then the center of mass will strictly rise.* Here we give a general version of the same statement, which works in any ordered P-algebra (in particular, in any ordered Banach space).

Proposition 3.5.11. *Let A be an ordered P-algebra, and let $p, q \in PA$. Suppose that $p \leq q$ in the usual stochastic order, and $e(p) = e(q)$. Then $p = q$.*

In order to prove the proposition we use the following version of the Hahn-Banach theorem [Lue69, Section 5.12, Theorem 1].

Theorem 3.5.12 (Geometric Hahn-Banach theorem)**.** *Let B be a normed vector space. Let A be an affine subset of B, let U be a convex, open set, and suppose that A and U are disjoint. Then there exists a bounded linear functional $\phi : B \to \mathbb{R}$ and a constant c such that:*

- *$\phi(a) = c$ for all $a \in A$;*
- *$\phi(u) > c$ for all $u \in U$.*

We are ready now to prove the statement.

Proof of Proposition 3.5.11. Suppose that $p \leq q$. By definition of stochastic order, we know that there exists a join $r \in P(A \otimes A)$ whose support lies entirely in the relation $\{\leq\} \subset A \otimes A$. We want to prove that in fact, r must be supported on the diagonal $D := \{(a, a), a \in A\}$, since that would imply that $p = q$.

Consider $A \otimes A$ embedded in a Banach space, as we know from the previous section. Now suppose that r is *not* entirely supported on the diagonal. Then there exists an $(a, b) \in A \otimes A$ with $a < b$ strictly, such that every open neighborhood U of (a, b) has strictly positive measure. Now take a neighborhood U disjoint from D (which we can, since $a \neq b$), and convex (which we can, since we are in a locally convex space). So U is convex and open, D is an affine subspace, and U and D are disjoint. By Theorem 3.5.12, we can find then a short affine map $h : A \otimes A \to \mathbb{R}$ such that:

- $h(d) = 0$ for all $d \in D$;
- $h(u) > 0$ for all $u \in U$.

Now consider the integral of h with respect to the measure r,

$$\int_{A \otimes A} h(a, b)\, dr(a, b). \tag{3.5.5}$$

Since h is short and affine, it commutes with the integrals,

$$\int_{A \otimes A} h(a, b)\, dr(a, b)$$

$$= h\Big(\int_{A\otimes A} a\, dr(a,b), \int_{A\otimes A} b\, dr(a,b)\Big)$$
$$= h\Big(\int_A a\, dp(a), \int_A b\, dq(b)\Big)$$
$$= h(e(p), e(q)).$$

At the same time, we have that (3.5.5) is bounded below by the integration restricted on U, which is strictly positive by assumption:

$$\int_{A\otimes A} h(a,b)\, dr(a,b) \geq \int_U h(a,b)\, dr(a,b) > 0. \tag{3.5.6}$$

Therefore $h(e(p), e(q)) > 0$ strictly. But since h is identically zero on the diagonal D, we must conclude that $(e(p), e(q))$ is *not* in D, i.e. that $e(p) \neq e(q)$. □

3.5.2. Higher structure

We now consider L-COMet as a category enriched in posets, or equivalently as a *locally posetal* 2-category. Concretely, we put $f \leq g$ for $f, g : X \to Y$ if and only if $f(x) \leq g(x)$ for all $x \in X$. This property/2-cell is preserved by P:

Proposition 3.5.13. *Let $f \leq g : X \to Y$. Then $Pf \leq Pg : PX \to PY$.*

Proof. Let $h : Y \to \mathbb{R}$ (monotone). We have that for every $x \in X$, $f(x) \leq g(x)$ in Y, therefore $h \circ f(x) \leq h \circ g(x)$. Since all the measures in PX are positive (or equivalently, positive linear functionals), we get that for every $p \in P$,

$$\int_X h\, d(f_*p) = \int_X h \circ f\, dp \leq \int_X h \circ g\, dp = \int_X h\, d(f_*p),$$

i.e. (since it holds for every such h) $(Pf)(p) \leq (Pg)(p)$. Since this holds for every p, we get finally $Pf \leq Pg$. □

Corollary 3.5.14. *P is a (strict) 2-functor, and so also a strict 2-monad, on* L-COMet *(as a strict 2-category).*

Consider now the adjunction given by the bijection (3.5.4). The operations $f \mapsto e \circ (Pf)$ and $\tilde{f} \mapsto \tilde{f} \circ \delta$ forming the bijection are monotone:

- If $f, g : X \to A$ and $f \leq g$, then $Pf \leq Pg$ by Proposition 3.5.13, and then $e \circ (Pf) \leq e \circ (Pf)$ by monotonicity of e;

- If $\tilde{f}, \tilde{g} : PX \to A$ and $\tilde{f} \le \tilde{g}$, then in particular they preserve the order on the delta measures, so that $\tilde{f} \circ \delta \le \tilde{f} \circ \delta$.

Therefore, the correspondence

$$\mathsf{L\text{-}COMet}(X, A) \cong \mathsf{L\text{-}COMet}^P(PX, A) \tag{3.5.7}$$

is not just a bijection of sets, but also an isomorphism of partial orders. In other words, it is an adjunction in the enriched (locally posetal) sense.

From the abstract point of view, the 2-monad P induces a 2-adjunction, which implies an equivalence of the hom-preorders in (3.5.7). But since all the objects of our categories are partial orders, all the hom-categories are skeletal, and so such equivalence of preorders must be an isomorphism of partial orders.

Let's now give a 2-categorical analogue of the concept of *separating points*. In an L-ordered space, by definition, the morphisms to $\mathbb{R}$ are enough to distinguish points and to determine the order. Here is how we can formalize the statement, by defining an analogue of *coseparators* for locally posetal 2-categories.

Definition 3.5.15. *Let* C *be a locally posetal 2-category. We call a* 2-coseparator *an object* S *of* C *such that the 2-functor*

$$\mathsf{C}(-, S) : \mathsf{C}^{\mathrm{op}} \to \mathsf{Poset}$$

is locally fully faithful.

By definition, $\mathbb{R}$ is a 2-coseparator in the categories $\mathsf{L\text{-}OMet}$ and $\mathsf{L\text{-}COMet}$. Conversely, we can characterize the categories $\mathsf{L\text{-}OMet}$ and $\mathsf{L\text{-}COMet}$ as being *exactly* the full subcategories of OMet and COMet on which $\mathbb{R}$ is a 2-coseparator.

Thanks to the Hahn-Banach theorem (in our case, by Proposition 3.5.3), we know that the order on P-algebras is determined even just by *affine* short monotone maps:

Proposition 3.5.16. *Let* $X \in \mathsf{L\text{-}COMet}$*, and let* A *be a* P*-algebra. Consider two maps* $f, g : X \to A$*. Then* $f \le g$ *in the pointwise order if and only if for every* P-morphism $h : A \to \mathbb{R}$*, we have* $h \circ f \le h \circ g$*.*

Proof. Since h is required to be monotone, one direction is trivial.

Suppose now that $f \not\le g$. Then by definition there exists $x \in X$ such that $f(x) \not\le g(x)$ in A. By Proposition 3.5.3 we know that there exists an affine map $h : A \to \mathbb{R}$ such that $h(f(x)) > h(g(x))$ strictly, So $h \circ f \not\le h \circ g$. $\square$

Corollary 3.5.17. *The real line* $\mathbb{R}$ *is a 2-coseparator in the Eilenberg-Moore category of* P*, i.e. of* P*-algebras and* P*-morphisms (affine maps).*

3.5.3. Convex monotone maps as oplax morphisms

In this subsection we will consider $\mathsf{L\text{-}COMet}$ a strict 2-category, and P a strict 2-monad, in the sense explained in 3.5.2.

This means that for algebras of the ordered Kantorovich monad, the algebra morphisms are not the only interesting maps: there are also *lax* algebra morphisms. A lax P-morphism $f : A \to B$ is a short, monotone map together with a 2-cell (which here is a property rather than a structure),

$$\begin{tikzcd} PA \arrow[r, "Pf"] \arrow[d, "e"'] & PB \arrow[d, "e"] \arrow[dl, phantom, "\swArrow"] \\ A \arrow[r, "f"'] & B \end{tikzcd} \tag{3.5.8}$$

which means that $e(f(p)) \leq f(e(p))$ for all $p \in PX$.

These maps are well known, at least in a special case.

Proposition 3.5.18. *Let A be an unordered P-algebra, and consider $\mathbb{R}$ with its usual order. Let $f : A \to \mathbb{R}$ be short (and automatically monotone). Then f is a lax P-morphism if and only if it is a concave function.*

Proof. Diagram (3.5.8) can be written explicitly as:

$$\int_A f(a)\, dp(a) \leq f\left(\int_A a\, dp(a) \right) \tag{3.5.9}$$

for any $p \in PA$. By the generalized Jensen's inequality, this is equivalent to

$$\lambda\, f(a) + (1-\lambda)\, f(b) \;\leq\; f\big(\lambda\, a + (1-\lambda)\, b\big) \tag{3.5.10}$$

for all $a, b \in A$ and $\lambda \in [0,1]$. This is the usual definition of a concave function. □

More in general, we think of lax P-morphisms as *monotone* concave functions. Dually, *oplax* P-morphisms—which are as in (3.5.8) but with the inequality oriented the opposite way—correspond to monotone convex functions. We therefore have the following categories:

- $P\mathsf{Alg}_s$ the category of P-algebras and *strict* P-morphisms (affine maps);
- $P\mathsf{Alg}_l$ the category of P-algebras and *lax* P-morphisms (concave maps);
- $P\mathsf{Alg}_o$ the category of P-algebras and *oplax* P-morphisms (convex maps).

All these categories are again locally posetal 2-categories, and since they contain all affine maps, they all admit $\mathbb{R}$ as a 2-coseparator.

We have then proven the following:

Theorem 3.5.19. *Consider the monad P on* $\mathsf{L\text{-}COMet}$. *Then:*

- $P\mathsf{Alg}_s$ *is equivalent to the category of closed convex subsets $A \subseteq E$ with E an ordered Banach space, with morphisms given by monotone* affine *short maps;*
- $P\mathsf{Alg}_l$ *is equivalent to the category of closed convex subsets $A \subseteq E$ with E an ordered Banach space, with morphisms given by monotone* concave *short maps;*
- $P\mathsf{Alg}_o$ *is equivalent to the category of closed convex subsets $A \subseteq E$ with E an ordered Banach space, with morphisms given by monotone* convex *short maps.*

Remark 3.5.20. It is a very well-known fact that the composition $f \circ g$ of two convex functions $f, g : \mathbb{R} \to \mathbb{R}$ may not be a convex function; and that if f is in addition monotone, then $f \circ g$ *is* convex. We now explain how this makes perfect sense within our framework. We write $(\mathbb{R}, \leq)$ for the $\mathbb{R} \in \mathsf{CMet}$ equipped with its usual order, and $(\mathbb{R}, =)$ for $\mathbb{R} \in \mathsf{CMet}$ equipped with the discrete order. Technically all our maps are assumed to be short, but the same considerations should apply more generally.

By Proposition 3.5.18, a concave function $\mathbb{R} \to \mathbb{R}$ is the same thing as a lax P-morphism $(\mathbb{R}, =) \to (\mathbb{R}, \leq)$; monotonicity is a trivial requirement. A monotone concave function $\mathbb{R} \to \mathbb{R}$ is the same thing as a lax P-morphism $(\mathbb{R}, \leq) \to (\mathbb{R}, \leq)$. In our formalism, both functions are technically monotone, but with respect to different orders on the domain. Due to the possibility of composing in $P\mathsf{Alg}_l$, we have:

- Two concave monotone functions $(\mathbb{R}, \leq) \to (\mathbb{R}, \leq)$ can be composed, giving again a concave monotone function $(\mathbb{R}, \leq) \to (\mathbb{R}, \leq)$;
- A concave monotone function $(\mathbb{R}, =) \to (\mathbb{R}, \leq)$ can be postcomposed with a concave monotone function $(\mathbb{R}, \leq) \to (\mathbb{R}, \leq)$, giving a concave monotone function $(\mathbb{R}, =) \to (\mathbb{R}, \leq)$;
- Two concave monotone functions $(\mathbb{R}, =) \to (\mathbb{R}, \leq)$ *cannot* be composed, since domain and codomain do not match.

We see that in this framework, the rule for when the composition of concave functions is again concave is just elementary category theory. Of course, the same applies to convex functions as oplax P-morphisms.

3.6. The exchange law

There is an even stronger compatibility condition that we can impose between the metric and the order. It is a unidirectional commutation of the relation $x \leq y$ with the relation $d(x, y) < r$:

Definition 3.6.1. $X \in \mathsf{OMet}$ *satisfies the* exchange law *if and only if, for every* $x, y, \bar{y} \in X$ *such that* $x \leq y$ *and* $d(y, \bar{y}) < r$*, there exists* $\bar{x} \in X$ *such that* $\bar{x} \leq \bar{y}$ *and* $d(x, \bar{x}) < r$.

With "$\leq r$" in place of "$< r$", we obtain a slightly stronger condition which has already been used in the context of the stochastic order in [HLL18, Proposition 3.8]. Note that our condition holds if and only if for $x \leq y$ and $d(y, \bar{y}) < r$, we can find $\bar{x}$ such that $d(x, \bar{x}) \leq r$.

In pictures, the exchange law ways that for every configuration of points

$$\begin{tikzcd} & x \arrow[dr, dotted, no head] & \\ x' \arrow[ur, "\leq"] & & y \end{tikzcd}$$

such that the distance between x and y (dotted line) is less than r, we can complete the diagram to

$$\begin{tikzcd} & x \arrow[dr, dotted, no head] & \\ x' \arrow[ur, "\leq"] \arrow[dr, dotted, no head] & & y \\ & y' \arrow[ur, "\leq"'] & \end{tikzcd}$$

such that the distance between x' and y' is also less than r.

Remark 3.6.2. Every ordered Banach space trivially satisfies the exchange law: given x, y, x' with $x' \leq x$, one can always define

$$y' := y - x + x'.$$

This way, $d(x', y') = d(x, y)$, and

$$y - y' = y - y + x - x' = x - x' \geq 0.$$

We denote $\mathsf{X\text{-}OMet}$ and $\mathsf{X\text{-}COMet}$ the full subcategories of OMet and COMet whose objects are spaces satisfying the exchange law of Definition 3.6.1. The categories $\mathsf{X\text{-}OMet}$ and $\mathsf{X\text{-}COMet}$ are full subcategories of $\mathsf{L\text{-}OMet}$ and $\mathsf{L\text{-}COMet}$:

Proposition 3.6.3. *Suppose that $X \in \mathsf{OMet}$ satisfies the exchange law. Then X is L-ordered.*

In order to prove the proposition we use the following remark from [HLL18, Proposition 3.8]

Lemma 3.6.4. *Let $X \in \mathsf{X\text{-}OMet}$. Then for every lower set $L \subset X$, the function*

$$d(-, L) : x \mapsto \inf_{l \in L} d(x, l) \tag{3.6.1}$$

is monotone (and short).

Proof of Lemma 3.6.4. Only monotonicity is nontrivial. So suppose $x \le y$. By the exchange law, for every $l \in L$ and $\varepsilon > 0$, there exists $l' \le l$ such that $d(x, l') \le d(y, l) + \varepsilon$. Since L is lower, necessarily $l' \in L$. Therefore

$$\inf_{l' \in L} d(x, l') \le \inf_{l \in L} d(y, l),$$

as was to be shown. □

Proof of Proposition 3.6.3. Suppose that X satisfies the exchange law, and suppose that $x \not\le y$. Denote by $\downarrow \{y\}$ the down-set of y:

$$\downarrow \{y\} := \{y' \in X \text{ such that } y' \le y\}.$$

Then by assumption $x \notin \downarrow \{y\}$. Since the order is closed, $\downarrow \{y\}$ is closed too. Therefore $d(x, L)$ is nonzero, and $d(y, L) = 0$. By Lemma 3.6.4, since X satisfies the exchange law, $d(-, L)$ is short and monotone. So we have found a short monotone function $d(-, L) : X \to \mathbb{R}$ such that $d(x, L) > d(y, L)$ strictly. □

Since every $X \in \mathsf{X\text{-}COMet}$ is L-ordered, we can apply the Kantorovich monad. The resulting space PX will also satisfy the exchange law:

Proposition 3.6.5. *Suppose $X \in \mathsf{X\text{-}COMet}$. Then $PX \in \mathsf{X\text{-}COMet}$ as well.*

Proof. Let $p, q, \bar{q} \in PX$ with $p \le q$ and $d(q, \bar{q}) < r$ for some $r > 0$. Choose $\varepsilon > 0$ such that $8\,\varepsilon < r - d(q, \bar{q})$. By Proposition 3.2.5, we can find empirical distributions $p_\varepsilon, q_\varepsilon, \bar{q}_\varepsilon$ with $p_\varepsilon \le q_\varepsilon$, which are ε-close to $p, q, \bar{q}$, respectively. Without loss of generality, we can assume that they all come from some power X^N for some large enough $N \in \mathsf{FinUnif}$. This means that, after possibly permuting the components, we have

$$p_\varepsilon = i((x_n)_{n \in N}), \quad q_\varepsilon = i((y_n)_{n \in N}), \quad \bar{q}_\varepsilon = i((\bar{y}_n)_{n \in N})$$

for suitable $(x_n), (y_n), (\bar{y}_n) \in X^N$ such that for all $n \in N$, we have $x_n \leq y_n$ in X. Now since X satisfies the exchange law, we can find $\bar{x}_n \in X$ for every n such that $\bar{x}_n \leq \bar{y}_n$ and such that $d(\bar{x}_n, x_n) \leq d(\bar{y}_n, y_n) + \varepsilon$. Call now $\bar{p}_\varepsilon := ((\bar{x}_n)_{n\in N})$. We have that

$$\begin{aligned} d(\bar{p}_\varepsilon, p_\varepsilon) &= |N|^{-1} \sum_{n\in N} d(\bar{x}_n, x_n) \leq |N|^{-1} \sum_{n\in N} d(\bar{y}_n, y_n) + \varepsilon \\ &= d(\bar{q}_\varepsilon, q_\varepsilon) + \varepsilon \leq d(\bar{q}_\varepsilon, \bar{q}) + d(q, \bar{q}) + d(q, q_\varepsilon) + \varepsilon \\ &\leq d(q, \bar{q}) + 3\,\varepsilon < r. \end{aligned}$$

We can now find Cauchy sequences $\{p_j\}, \{q_j\}, \{\bar{q}_j\}$ tending arbitrarily fast respectively to $p, q, \bar{q}$, with $p_j \leq q_j$ and such that $p_j, q_j, \bar{q}_j$ are empirical distributions coming from X^{N_j}, with $N_j \in \mathsf{FinUnif}$ for all j. We can take as first elements of the three sequences the values obtained above,

$$p_1 := p_\varepsilon, \quad q_1 := q_\varepsilon, \quad \bar{q}_1 := \bar{q}_\varepsilon.$$

Since the sequence $\{\bar{q}_j\}$ can be chosen to tend to $\bar{q}$ arbitrarily fast, suppose $d(\bar{q}_j, \bar{q}) < 2^{1-j}\,\varepsilon$. This way,

$$\begin{aligned} d(\bar{q}_\ell, \bar{q}_{\ell+1}) &\leq \big(d(\bar{q}_\ell, \bar{q}) + d(\bar{q}, \bar{q}_{\ell+1})\big) \leq \big(2^{1-\ell} + 2^{-\ell}\big)\,\varepsilon \\ &= 2^{-\ell}(2+1)\,\varepsilon = 2^{-\ell} \cdot 3\,\varepsilon\,. \end{aligned}$$

We can obtain a sequence $\{\bar{p}_j\}$ in the following way: start with the above $\bar{p}_1 := \bar{p}_\varepsilon$. Now given $\bar{p}_j$ coming from X^{N_j} with $\bar{p}_j \leq \bar{q}_j$, we know that by the argument above we can find an empirical distribution $\bar{p}_{j+1}$ coming from $X^{N_{j+1}}$ such that $\bar{p}_{j+1} \leq \bar{q}_{j+1}$, and such that $d(\bar{p}_{j+1}, \bar{p}_j) < 2^{-\ell} \cdot 4\,\varepsilon$. This way we would get, for every $k \geq j$,

$$\begin{aligned} d(\bar{p}_k, \bar{p}_j) &\leq \sum_{\ell=j}^{k-1} d(\bar{p}_\ell, \bar{p}_{\ell+1}) < \sum_{\ell=j}^{k-1} 2^{-\ell} \cdot 4\,\varepsilon \\ &= 2^{-j} \sum_{\ell=0}^{k-1-j} 2^{-\ell} \cdot 4\,\varepsilon < 2^{-j} \cdot 2 \cdot 4\,\varepsilon = 2^{-j} \cdot 8\,\varepsilon\,. \end{aligned}$$

With such a choice of $\{\bar{q}_j\}$, the sequence $\{\bar{p}_j\}$ is Cauchy. Let $\bar{p}$ be its limit. Then we have that by continuity,

$$\bar{p} = \lim_j \bar{p}_j \leq \lim_j \bar{q}_j = \bar{q},$$

and

$$\begin{aligned} d(p,\bar{p}) = \lim_j d(p,\bar{p}_j) &\leq d(p,p_\varepsilon) + d(p_\varepsilon,\bar{p}_\varepsilon) + \lim_j d(\bar{p}_\varepsilon,\bar{p}_j) \\ &< \varepsilon + d(q,\bar{q}) + 3\,\varepsilon + \lim_j d(\bar{p}_\varepsilon,\bar{p}_j) = d(q,\bar{q}) + 4\,\varepsilon + \lim_j d(\bar{p}_1,\bar{p}_j) \\ &< d(q,\bar{q}) + 4\,\varepsilon + 2^{-1}\cdot 8\,\varepsilon = d(q,\bar{q}) + 8\,\varepsilon < r. \end{aligned}$$

□

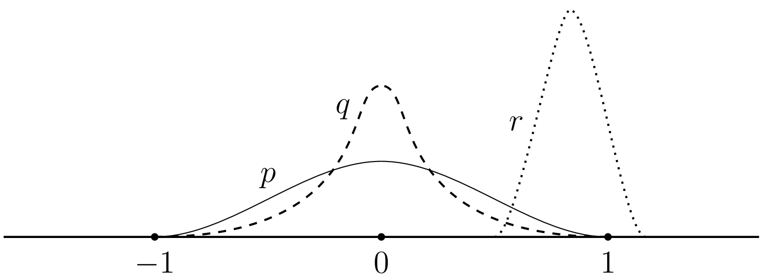
Therefore, P restricts to a monad on $\mathsf{X\text{-}COMet}$.

4. Convex Orders

The stochastic order of Chapter 3 can be thought of as comparing probability measures in terms of *how far up they are* for the order of the underlying space. In this chapter we will study another order, which compares measures in terms of *how spread*, or *how random they are.*

Measuring the amount of "randomness" or "risk" of a probability distribution is something of utmost importance in probability and statistics, and there are several quantities designed to accomplish such a task, like variance and entropy. However, all these quantities necessarily induce a total preorder, which does not in general encode as much information as a partial order. Intuitively, a single number can measure only "how much" the randomness is, but not "where", or "in which way".

Example 4.0.1. Consider for example the probability distributions on $\mathbb{R}$ whose densities are represented in the following picture:



One can say that p is "more random" or "more spread" than q *over the same values.* Instead, while r looks more "peaked" than q, it is so over different elements: it has indeed less randomness *quantitatively*, but over *different regions.* In a partial order, we would say that q and r are incomparable. The same would be true, in higher dimensions, if the two distributions were spread along different directions. This is what we mean by "where the randomness is".

The first partial order on probability distributions formalizing "increasing randomness" was introduced, as far as we know, by Blackwell [Bla51]. In the following years, several researchers from different fields have given similar definitions, from Strassen in probability theory [Str65], to Stiglitz and Rothschild in economics [RS70]. Just as for the stochastic order, this new order, which we will denote by $\preceq_c$, is known in the literature under many names: *risk order* [RS70], *convex order* [KA10], and *Choquet order* [Win85]. Again, there are mainly three more or less equivalent ways to define it:

(a) $p \preceq_c q$ if and only if q can be obtained from p by composition with a mean preserving kernel, or "dilation";

(b) $p \preceq_c q$ if and only if there exist random variables X and Y with laws p and q, respectively, such that X can be written as a conditional expectation of Y;

(c) $p \preceq_c q$ if and only if for every convex function $f : X \to \mathbb{R}$ of a certain class (for example, continuous),
$$\int f\, dp \leq \int f\, dq.$$

A possible interpretation of the first condition is that *q can be obtained by adding noise to p*, or *diffusion without drift*, or *casual, unbiased errors*. A possible interpretation of the second condition is that p can be obtained from q by "partially averaging", or "concentrating" some components of q. A possible interpretation of the third condition is that for any choice of *risk-seeking* utility function, the expected utility with measure p is less than the expected utility with measure q (with risk-averse utilities, reverse the inequality). These conditions are known to be equivalent with some degree of generality. Winkler, in particular, has proven a very general equivalence theorem, valid in any locally convex topological vector space [Win85, Theorem 1.3.6].

When the underlying space is ordered, it is interesting to compare the convex order with the stochastic order. The two orders are in some way *transverse*, meaning that two distributions cannot be comparable for both orders unless they are equal (see Section 4.3). One can also define a new order comprising both orders, called sometimes *increasing convex order* [KA10], or *second order stochastic dominance* [Fis80], which is also of use in applications: for example, in economics, *concave monotone* functions represent increasing risk-averse utilities. This more general order is not as well studied as the Choquet order and, for example, no duality result seems to be known in general (we will prove such a result in Section 4.4).

In this chapter we want to give a categorical definition and treatment of the orders described above, as well as a full duality theorem. We start by studying in detail the idea of *partially averaging a measure*. As we have seen in 1.1.2, a possible interpretation of probability monads and their algebras is that the monad P defines an operation of *average*, or *expectation*, under which the algebras are closed. We introduce a categorical formalism to model *expressions evaluated partially*, which can be defined for arbitrary monads on concrete categories. This allows us to define an order of "partial evaluations" on all algebras of the Kantorovich monad. The resulting order appears in the literature, and it has been studied at least by Winkler [Win85], who proved that such a construction is equivalent to the traditional Choquet order on all bounded spaces. We prove that, in a metric setting, the boundedness assumption is not necessary, so that the equivalence of the two orders always holds.

We know from Chapter 3 that convex functions are characterized categorically as the oplax P-morphisms. As we sketched above, the Choquet order, which is equivalent to the order of partial evaluations, is dual to convex functions. This is not a coincidence: as we will show, the relationship between the partial evaluation order and convex functions has a deep categorical meaning, coming from the ordered Choquet adjunction (3.5.7). This connection permits to characterize the partial evaluation order in terms of a universal property, as an *oplax codescent object* in the sense of [Lac02]. From the universal property we can then easily derive a general duality result valid for all ordered Banach spaces, Corollary 4.4.10, which, as far as we know, is new.

Our theory of partial averages, just like the classical concepts of martingale and conditional expectation, necessarily takes place in a convex space. Therefore, for the whole of this chapter, we will work only with P-algebras. There is a way to generalize convex functions and diffusion to metric spaces that are not necessarily convex, for example graphs or manifolds. This will, however, not be pursued in this work.

Outline.

- In Section 4.1 we give a categorical definition of "partial evaluations" in terms of monads, and explain the intuition behind it.

- In Section 4.2 we instantiate the definition of partial evaluation in the case of the Kantorovich monad. We prove that the resulting relation is a closed partial order (Theorem 4.2.4), and even a P-algebra (Proposition 4.2.10). In 4.2.1 we prove (Theorem 4.2.14) that over every P-algebra, the partial evaluation order is equivalent to the existence of a conditional expectation or of a dilation, extending Winkler's result [Win85, Theorem 1.3.6] to possibly unbounded spaces in a metric setting, and connecting to the known literature on the Choquet order. In 4.2.2 we give a convergence result for bounded monotone nets in this order (Theorem 4.2.18), extending another result by Winkler [Win85, Theorem 2.4.2] to possibly unbounded spaces in a metric setting.

- In Section 4.3 we compare the partial evaluation order with the stochastic order. In 4.3.1 we show that the two orders are always transverse (Corollary 4.3.1). In 4.3.2 we define the *lax partial evaluation order* as the composite of the two orders, and prove that it also defines a P-algebra (Proposition 4.3.14).

- In Section 4.4 we prove (Theorem 4.4.3) that the ordered Choquet adjunction of equation (3.5.7) connects the partial evaluation order with convex functions (oplax P-morphisms). In 4.4.1 we show that this connection characterizes the partial evaluation orders in terms of a universal property, as an *oplax codescent object* (Theorem 4.4.5), whose properties we study in 4.4.2. Finally, in 4.4.3, we show that by its universal property, the lax partial evaluation order is dual to monotone convex functions over every ordered Banach space, fully generalizing all results known to us in the literature (Theorem 4.4.9 and Corollary 4.4.10).

Most of the material in this chapter will be part of a paper which is currently in preparation.

4.1. Partial evaluations

We have seen in 1.1.2 that monads can be interpreted in terms of spaces of *formal expressions*. Suppose now that we have a monad T and a T-algebra (A, e), for example, the free commutative monoid monad of 1.1.2, together with the commutative monoid of natural numbers with addition. Consider now the formal sums

$$2 + 3 + 4 \qquad \text{and} \qquad 5 + 4.$$

These formal sums have the same result, 9. But moreover, the second sum is in some way *closer* to the result: the first term in the second sum is already the (actual) sum of the first two terms in the first sum. In other words, the second formal sum is a *partial evaluation* of the first one: *part* of the formal expression has already been evaluated.

Let's try to make this precise. The idea is that there is a *formal sum of formal sums*, i.e. a formal sum with one level of brackets (see 1.1.2) such that removing the brackets yields the term on the left, and performing the operations in the brackets (and then removing them) yields the term on the right. That is:

$$\begin{tikzcd}
& (2+3)+(4) \arrow[dl, "\text{remove brackets}"'] \arrow[dr, "\text{evaluate brackets}"] & \\
2+3+4 & & 5+4
\end{tikzcd}$$

As we have seen in 1.1.2, the "formal sums of formal sums" live in TTA. The map which can be seen as "removing the brackets" is the composition map $\mu : TTA \to TA$, and the map that evaluates the expressions within the brackets is the image of the evaluation map e under the functor T, i.e. $(Te) : TTA \to TA$. We can then give a precise definition of partial evaluations in terms of monads. Since we are talking about elements, we need the category in question to be concrete (but this approach can be generalized).

Definition 4.1.1. *Let (T, η, μ) be a monad on a concrete category C, and (A, e) a T-algebra. Let $s, t \in TA$. We say that s is a* partial evaluation *of t if and only if there exists a $\sigma \in TTA$ such that $(Te)(\sigma) = s$, and $\mu(\sigma) = t$.*

From the definition we have immediately the following result, which is a sort of consistency check: if s is a partial evaluation of t, then s and t necessarily must have the same result (in the example above, 9).

Proposition 4.1.2 (Law of total evaluation)**.** *Let $s, t \in TA$ like above, and suppose that s is a partial evaluation of t. Then s and t have necessarily the same "total evaluation", i.e. $e(s) = e(t)$.*

Proof. The composition square of the T-algebra (A, e) is a commutative diagram

$$\begin{tikzcd} TTA \arrow[r, "Te"] \arrow[d, "\mu"] & TA \arrow[d, "e"] \\ TA \arrow[r, "e"] & A \end{tikzcd}$$

Now suppose that s is a partial evaluation of t. Then by definition, there exists a $\sigma \in TTA$ such that $(Te)(\sigma) = s$, and $\mu(\sigma) = t$. But then, since the square above commutes,

$$e(s) = e \circ (Te)(\sigma) = e \circ \mu(\sigma) = e(t).$$

□

This may remind the reader of the "law of total expectation" that random variables and conditional expectations satisfy. We will see that this analogy is precise: partial evaluations for a probability monad correspond exactly to conditional expectations, which one can see as "partial expectations". More on this in 4.2.1.

Moreover, any expression has two trivial partial evaluations: itself, and its total result (viewed as a formal expression).

Proposition 4.1.3. *Let A be a T-algebra like above, and $t \in TA$. Then:*

(a) t is a partial evaluation of itself;

(b) $\eta \circ e(t)$ is a partial evaluation of t.

Proof.

(a) Consider $(T\eta)(t) \in TTA$. Then $\mu \circ (T\eta)(t) = t$ by the right unitality of the monad, and $(Te) \circ (T\eta)(t) = T(e \circ \eta)(t) = t$ by functoriality of T together with the unit condition of the algebra. Therefore, t is a partial evaluation of itself.

(b) Consider $\eta(t) \in TTA$. We have a diagram

$$\begin{tikzcd} TA \arrow[r, "\eta"] \arrow[d, "e"] & TTA \arrow[d, "Te"] \\ A \arrow[r, "\eta"] & TA \end{tikzcd}$$

which commutes by naturality of η. Now $\mu \circ \eta(t) = t$ by the left unitality of the monad, and $(Te) \circ \eta(t) = \eta \circ e(t)$ by the commutativity of the diagram above. Therefore $\eta \circ e(t)$ is a partial evaluation of t. $\square$

There is another very tempting property to expect from partial evaluations, namely that if s is a partial evaluation of t and t is a partial evaluation of u, then s is a partial evaluation of u as well. This sort of composition property, or transitivity, as far as we know has not been proven for general monads.[1] However, we will prove (Proposition 4.2.5) that for the Kantorovich monad it always holds.

4.2. The partial evaluation order

Let's now study partial evaluations for algebras of the Kantorovich monad. In this section, we will consider the Kantorovich monad on *unordered* spaces, i.e. on CMet. Let's instantiate Definition 4.1.1 in our setting:

Definition 4.2.1. *Let A be a P-algebra. Let $p, q \in PA$. We say that q is a* partial evaluation *of p, and we write $q \preceq_c p$, if and only if there exists a $\mu \in PPA$ such that $E\mu = p$ and $(Pe)\mu = q$.*

As we have seen in the beginning of the chapter, the intuition is that p is "more concentrated" than q, or "closer to a delta at its center of mass". From the statistical

[1] It is known to be true for several classes of monads, however, like cartesian monads on Set.

point of view, p is better approximated by just looking at its expectation than q, since q is "more spread out".

Note the direction of the relation. This is motivated by the fact that this order is equivalent to the convex order (see the introduction of this chapter), which is conventionally defined in the same direction, so that we avoid confusion in the notation.

Directly from Proposition 4.1.2 we have a law of total evaluation for P:

Corollary 4.2.2. *Let $p, q \in PA$ like above, and suppose that $q \preceq_c p$. Then necessarily $e(p) = e(q)$, i.e. p and q must have the same expectation.*

Just as well, from Proposition 4.1.3 we have the following trivial evaluations:

Corollary 4.2.3. *Let $p \in PA$ like above. Then $p \preceq_c p$, and $\delta_{e(p)} \preceq_c p$.*

We can view the latter as: p is necessarily more spread than its center of mass.

As said in Section 4.1, it is tempting to check whether partial evaluations can be composed. We prove here that, for the Kantorovich monad, this is indeed the case. Moreover, since $p \preceq_c q$ can be seen as p being "less random" than q, it is tempting to check whether $p \preceq_c q$ and $q \preceq_c p$ imply $p = q$: this is also the case. And partial evaluations also respect approximations by sequences. In rigor, we then want to prove the following result:

Theorem 4.2.4. *Let A be a P-algebra. The partial evaluation relation $\preceq$ on PA is a closed partial order.*

In order to prove the theorem, it is convenient to first look at some partial results. First of all, transitivity, or composition of partial evaluations, deserves some particular attention. The result follows in particular from the disintegration result of Theorem 2.6.9.

Lemma 4.2.5. *Let A be a P-algebra. Let $p, q, r \in PA$ and let $\mu, \nu \in PPA$ such that $(Pe)\mu = p$, $E\mu = (Pe)\nu = q$, and $E\nu = r$. Then there exists $\rho \in PPA$ such that $(Pe)\rho = p$ and $E\rho = q$.*

Proof. Consider the commutative diagram:[2]

$$
\begin{tikzcd}
 & PA & \\
PPA \arrow[ur, "E"] \arrow[dd, "Pe"'] & & PPA \arrow[ul, "Pe"'] \arrow[dd, "E"] \\
 & PPPA \arrow[ul, "PPe"] \arrow[ur, "E"'] \arrow[dd, "PE"] & \\
PA & & PA \\
 & PPA \arrow[ul, "Pe"] \arrow[ur, "E"'] &
\end{tikzcd}
\tag{4.2.1}
$$

(which commutes by the composition, associativity, and naturality squares). Then we have that p sits in the bottom left corner, q in the top corner, and r in the bottom right corner, while μ sits in the top left corner, and ν in the top right. By Theorem 2.6.9, setting $f = e$, there exists an $\alpha \in PPPA$ making (4.2.1) commute. Therefore $\phi := (PE)\,\alpha$ is such that $(Pe)\rho = p$ and $E\rho = q$. $\square$

Antisymmetry is also interesting: we first notice that convex functions are sensitive to partial evaluations. This is a very deep connection, which will be explored further in Sections 4.4 and 4.4.

Lemma 4.2.6. *Let $f : A \to \mathbb{R}$ be short and convex. Let $\mu \in PPA$. Then*

$$\int_A f \, d(e_* \mu) \le \int_A f \, d(E\mu). \tag{4.2.2}$$

Proof. Rewriting both sides of (4.2.2), we have to prove that

$$\int_{PA} \left\{ f\left(\int_A a \, dr(a) \right) \right\} d\mu(r) \le \int_{PA} \left\{ \int_A f(a) \, dr(a) \right\} d\mu(r).$$

Now, since f is convex (i.e. a lax P-morphism), for every $r \in PA$,

$$f\left(\int_A a \, dr(a) \right) \le \int_A f(a) \, dr(a).$$

By monotonicity of integration, then (4.2.2) holds. $\square$

[2] This diagram is given by the first three levels of the bar construction [Mac00, Chapter VII, Section 6]. It seems that the bar construction is a sort of higher-level categorification of the partial evaluation relation. However, a detailed higher-categorical analysis of these ideas is beyond the scope of this work.

We now use the following known fact, following from [Win85, Lemma in Section 0.7], which says that convex functions are enough to test Borel probability measures on A, since the σ-algebra they define is exactly the Borel one.

Lemma 4.2.7. *Let A be a P-algebra, and let $p, q \in PA$. Suppose that for every $f : A \to \mathbb{R}$ Lipschitz and convex,*

$$\int_A f\,dp = \int_A f\,dq. \tag{4.2.3}$$

Then $p = q$.

We are now ready to prove the rest of the theorem.

Proof of Theorem 4.2.4. Reflexivity follows from Corollary 4.2.3.

Transitivity is given exactly by Lemma 4.2.5.

Antisymmetry follows from Lemmas 4.2.6 and 4.2.7.

For the closure, let $\{p_i\}, \{q_i\}$ be Cauchy sequences in PA tending to p and q respectively, and such that p_i is a partial evaluation of q_i for all i. This means that there exists a (generic) sequence $\{\mu_i\}$ in PPA such that for all i, $(Pe)\mu_i = p_i$ and $E\mu_i = q_i$. By Corollary 2.6.8, the sequence $\{\mu_i\}$ has at least one accumulation point μ. Now since both E and (Pe) are continuous, $(Pe)\mu = p$ and $E\mu = q$, so p is a partial evaluation of q.

Therefore the partial evaluation relation is a closed partial order. $\square$

So, in particular, PA with the partial evaluation relation is an ordered metric space. In addition, it satisfies the exchange law, thanks to the metric lifting that we saw in 2.6.2.

Proposition 4.2.8. *Let A be a P-algebra. The partial evaluation order on PA satisfies the exchange law of Definition 3.6.1. In other words, given $p, q, \bar{q} \in PA$ with $p \preceq_c q$ and $d(q, \bar{q}) < r$, there exists a $\bar{p} \in PA$ such that $\bar{p} \preceq_c \bar{q}$ and $d(p, \bar{p}) < r$.*

Proof. By definition there exists $\mu \in PPA$ such that $E\mu = q$ and $e_*\mu = p$. Since $d(q, c) < r$, by the metric lifting property given in Proposition 2.6.5, there exists $\bar{\mu} \in PPA$ such that $E\bar{\mu} = \bar{q}$ and $d(\mu, \bar{\mu}) < r$. Define now $\bar{p} := e_*\bar{\mu}$. We have that by construction, $\bar{p} \preceq_c \bar{q}$, and $d(p, \bar{p}) = d(e_*\mu, e_*\bar{\mu}) \leq d(\mu, \bar{\mu}) < r$. $\square$

Corollary 4.2.9. *Therefore, for any P-algebra A, we have (by Proposition 3.6.3) that $(PA, \preceq_c)$ is L-ordered.*

Moreover, it is even a P-algebra.

Proposition 4.2.10. *Let A be a P-algebra in $\mathsf{L\text{-}COMet}$. The map $E : PPA \to PA$ is monotone as a a map $P(PA, \preceq_c) \to (PA, \preceq_c)$. Therefore, $(PA, \preceq_c)$ is itself a P-algebra.*

We know already that PA is a P-algebra, with its usual order. However, not all orders on PA are compatible with the algebra structure: as we saw in Section 3.5, the only order that are allowed are those for which the structure map is monotone, in this case, $E : PPA \to PA$.

Proof. By Lemma 3.5.1, it suffices to show that for every $p, q, r \in PA$ such that $p \preceq_c q$ and every $\lambda \in [0, 1]$,

$$\lambda\, p + (1 - \lambda)\, r \quad \preceq_c \quad \lambda\, q + (1 - \lambda)\, r.$$

Now suppose that $p \preceq_c q$. Then by definition there exists $\mu \in PPA$ such that $e_* \mu = p$ and $E\mu = q$. Consider now the measure

$$\mu' := \lambda\, \mu + (1 - \lambda)\, \delta_*(r) \quad \in PPA.$$

Since e_* is affine, we have that

$$E(\mu') = \lambda\, E(\mu) + (1 - \lambda)\, E(\delta_*(r)) = \lambda\, q + (1 - \lambda)\, r,$$

and since E is affine, we have that

$$e_*(\mu') = \lambda\, e_*(\mu) + (1 - \lambda)\, e_*(\delta_*(r)) = \lambda\, p + (1 - \lambda)\, r.$$

Therefore

$$\lambda\, p + (1 - \lambda)\, r \quad \preceq_c \quad \lambda\, q + (1 - \lambda)\, r,$$

which means that $E : P(PA, \preceq_c) \to (PA, \preceq_c)$ is monotone. □

So the space $(PA, \preceq_c)$ can be embedded into an ordered Banach space.

Remark 4.2.11. Just like the stochastic order (Remark 3.5.10), the partial evaluation order $\preceq_c$ on PA induces a cone in $M(A)$, in the sense of the dual system defined in 2.1.6.

We will see that this cone is the *dual cone to convex functions*. More on that in Section 4.4.

4.2.1. Equivalence with conditional expectations

In probability theory there exists already a concept that intuitively is a "partial expectation", namely, *conditional expectation*. In this subsection we want to prove that the two concepts are in some sense equivalent. First of all, a caveat: elements of PA are *probability distributions, not random variables*. So any statement involving PA has to do with the *law* of random variables. In particular, equality in PA corresponds to equality in distribution.

The material in this subsection is closely related to the work of Winkler and Weizsäcker (see [Win85] and the discussion therein). However, their equivalence theorem relies on the assumption that the space A is bounded, while ours does not.

Since we have to connect our framework with the usual measure-theoretical approach, in this subsection, and only in this subsection, all the functions will be only assumed to be *measurable*, not necessarily short. Moreover, in this subsection, A will always denote a P-algebra, for example $\mathbb{R}$.

Definition 4.2.12. *Consider a probability space $(X, \mathcal{F}, \mu)$, a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, and measurable mappings $f, g : X \to A$ such that $f_*\mu$ and $g_*\mu$ have finite first moment. We say that g is a* conditional expectation of f given $\mathcal{G}$ *if:*

- *The function g is also $\mathcal{G}$-measurable;*
- *For every G in the σ-algebra $\mathcal{G}$, we have*

$$\int_G g \, d\mu = \int_G f \, d\mu.$$

For brevity, we extend the terminology to the image measures themselves:

Definition 4.2.13. *Let $p, q \in PA$. We say that p is a* conditional expectation *of q* in distribution *if there exist a probability space $(X, \mathcal{F}, \mu)$, a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, and mappings $f, g : X \to A$, with f $\mathcal{F}$-measurable and g $\mathcal{G}$-measurable, such that $p = g_*\mu$, $q = f_*\mu$, and g is a conditional expectation of f given $\mathcal{G}$.*

More informally, we say that p is a conditional expectation of q if they can be written as laws of A-valued random variables which are one the conditional expectation of the other.

Here is now the main result that we want to prove:

Theorem 4.2.14. *Let A be a P-algebra, and let $p, q \in PA$. Then p is a partial evaluation of q if and only if it is a conditional expectation of q in distribution.*

Again, this will not mean that whenever p is a partial evaluation of q, their associated *random variables* are in relationship of conditional expectation: we are only looking at the distributions, and not at the correlations between the random variables. The theorem above rather says that whenever p is a partial evaluation of q there *exists* a coupling between the two distributions which exhibits a conditional expectation relation between the associated random variables.

In order to prove the theorem, we need first to talk about another standard notion in probability: a random map which intuitively "only spreads, but does not translate" (think of diffusion without drift, or the kernel of a martingale). In statistics, this corresponds to "adding unbiased noise", or "casual, not systematic errors".

Definition 4.2.15. *Let A be a P-algebra. A* dilation *is a map $k : A \to PA$, which we write $a \mapsto k_a$, such that for all $a \in A$, $e(k_a) = a$. Let now $p \in PA$. A p-dilation is a map $t : A \to PA$ such that for p-almost all $a \in A$, $e(k_a) = a$.*

The most trivial dilation is the delta. Clearly, every dilation is a p-dilation. Here is a (traditional) disintegration result, of which similar versions are known in the literature (for example, [Win85, Theorem A2]).

Lemma 4.2.16. *Let $X, Y \in \mathsf{CMet}$. Let $r \in P(X \otimes Y)$ be a joint probability measure with marginals $p \in PX$ and $q \in PY$, respectively. Then for p-almost all $x \in X$ there exists a probability measure k_x on Y with the following properties:*

(a) For p-almost all x, k_x is tight, and it has finite first moment;

(b) The assignment $x \mapsto k_x$ is measurable;

(c) The measure k_*p is tight and has finite first moment;

(d) The joint defined by p and k is indeed r, i.e. for every measurable subsets $S \subseteq X$ and $T \subseteq Y$,

$$\int_S k_x(T)\, dp(x) = r(S \times T).$$

Proof. The existence of a measurable assignment $x \mapsto k_x$ is guaranteed by the usual theory of conditional expectation, and tightness is implied, for example, by Weizsäcker's disintegration theorem [Win85, Theorem A2]. The only properties that need to be checked are finite first moment of k_x for p-almost all x, and of k_*p.

Let now $f : Y \to \mathbb{R}$ be the 1-Lipschitz function given by $y \mapsto d(y_0, y)$ for some fixed $y_0 \in Y$. Then we have:

$$\int_X \left(\int_Y f\, dk_x \right) dp(x) = \int_{X \times Y} f\, dr < \infty,$$

since r has finite first moment. It follows that the integrand on the left-most side must be finite p-almost everywhere.

Now consider $\delta_{y_0} \in PY$. We have, by formula (2.1.11),

$$\int_{PY} d(\delta_{y_0}, q')\, d(k_*p)(q') = \int_{PY} \left(\int_Y d(y_0, y)\, dq'(y) \right) d(k_*p)(q')$$

$$= \int_X \left(\int_Y d(y_0, y)\, dk_x(y) \right) dp(x)$$
$$= \int_{X \times Y} d(y_0, y)\, dr(x, y) < \infty,$$

again since r has finite first moment. □

We use the previous lemma to prove the following equivalence result. The following is similar to known results, for example, part of [Win85, Theorem 1.3.6].

Lemma 4.2.17. *Let A be a P-algebra, and let $p, q \in PA$. The following conditions are equivalent:*

(a) p is a conditional expectation of q in distribution;

(b) There exists a joint $r \in P(A \otimes A)$, with marginals p and q, respectively, such that for every measurable set $B \subseteq A$,

$$\int_{B \times A} a\, dr(b, a) = \int_B b\, dp(b); \tag{4.2.4}$$

(c) There exists a p-dilation k such that $E \circ k_* p = q$;

Proof of Lemma 4.2.17.

- (a)⇒(b): Suppose that there exist $(X, \mathcal{F}, \mu)$, a sub-σ-algebra $\mathcal{G}$, and $f, g : X \to A$, such that g is a conditional expectation of f given $\mathcal{G}$. Then take the map $(f, g) : \Omega \to A \otimes A$, and the image measure $r := (f, g)_* \mu \in P(A \otimes A)$. The measure r has the prescribed marginals by construction, and

$$\int_{B \times A} a\, dr(b, a) = \int_{B \times A} a\, d\big((f, g)_* \mu\big)(b, a)$$
$$= \int_{f^{-1}(B) \cap g^{-1}(A)} g(x)\, d\mu(x) = \int_{f^{-1}(B) \cap g^{-1}(A)} f(x)\, d\mu(x)$$
$$= \int_{B \times A} b\, d\big((f, g)_* \mu\big)(b, a) = \int_{B \times A} b\, dr(b, a) = \int_B b\, dp(b).$$

- (b)⇒(a): Suppose that such a joint r exists. Then take as probability space $X = A \otimes A$, with the product σ-algebra, and as measure r, and as sub-σ-algebra $\mathcal{G}$ take the one generated by the projection $\pi_1 : A \otimes A \to A$ to the first component. Denote also $\pi_2 : A \otimes A \to A$ the projection to the *second* component. We have that by construction π_1 is $\mathcal{G}$-measurable, and the sets in the σ-algebra $\mathcal{G}$ are precisely those in the form $B \times A$ for measurable $B \subseteq A$. So from (4.2.17) we get that for every measurable $B \subseteq A$:

$$\int_{B\times A} \pi_1(b,a)\, dr(b,a) = \int_B b\, dp(b) = \int_{B\times A} a\, dr(b,a) = \int_{B\times A} \pi_2(b,a)\, dr(b,a),$$

therefore π_1 is the conditional expectation of π_2.

- (b)⇒(c): Suppose that such a joint r exists, and that is has finite first moment. By Lemma 4.2.16, we can find a measurable map $k : A \to PA$ defined for p-almost all $a \in A$, such that it gives the right joint, i.e. for each B, C:

$$r(B \times C) = \int_B k_b(C)\, dp(b),$$

and such that k_*p is in PPA. So in particular,

$$\int_{B\times A} a\, dr(b,a) = \int_B \left(\int_A a\, dk_b(a) \right) dp(b) = \int_B e(k_b)\, dp(b)$$

must be equal to

$$\int_B b\, dp(b),$$

for each measurable $B \subseteq A$, which means that for p-almost all $a \in A$, $e(k_a) = a$, by Radon-Nikodym. Moreover,

$$E \circ k_*p(B) = \int_{PA} s(B)\, d(k_*p)(s) = \int_A k_a(B)\, dp(a) = \int_A \int_B dr(a,b) = q(B).$$

- (c)⇒(b): Given $k : A \to PA$, we can form the joint $r \in P(A \otimes A)$ as usual:

$$r(B \times C) := \int_{B\times C} dk_b(c)\, dp(b).$$

With this construction, r has the prescribed marginals:

$$r(A \times C) = \int_A \int_C dk_a(c)\, dp(a) = \int_{PA} s(C)\, d(k_*p)(s) = E(k_*p)(C) = q(C),$$

and

$$r(B \times A) = \int_B \int_A dk_b(a)\, dp(b) = \int_B dp(b) = p(B).$$

Moreover:

$$\int_{B\times A} a\, dr(b,a) = \int_B \left(\int_A a\, dk_b(a) \right) dp(b) = \int_B e(k_b)\, dp(b) = \int_B b\, dp(b).$$

□

We are now ready to prove the main theorem.

Proof of Theorem 4.2.14. First of all, suppose that p is a conditional expectation of q. By Lemma 4.2.17, there exists a p-dilation $k : A \to PA$ such that $E \circ k_*(p) = q$. Define $\mu := k_* p \in PPA$. Then $e_*\mu = (e \circ k)_*(p) = p$, and $E\mu = E \circ k_*(p) = q$, so p is a partial evaluation of q.

Conversely, suppose that there exists $\mu \in PPA$ such that $e_*\mu = p$ and $E\mu = q$. We want to find a joint r satisfying condition (4.2.4). We apply to μ the composition:

$$PPA \xrightarrow{\mathrm{diag}_*} P(PA \otimes PA) \xrightarrow{(\delta\circ e\otimes \mathrm{id})_*} P(PA \otimes PA) \xrightarrow{\nabla_*} PP(A \otimes A) \xrightarrow{E} P(A \otimes A),$$

where diag is the diagonal map $p \mapsto (p, p)$, and ∇ is the monoidal map of Section 2.5 giving product probabilities $(p, q) \mapsto p \otimes q$. We obtain $r := E \circ \nabla_*(\delta \circ e \otimes \mathrm{id})_* \,\mathrm{diag}_*\, \mu$. The pair of marginals of r can be obtained by applying the map Δ of Section 2.5. Using Proposition 2.5.15 together with Corollary 2.5.18, and naturality of Δ, the following diagram commutes:

$$\begin{tikzcd}
PPA \arrow[r, "\mathrm{diag}_*"] \arrow[dr, "\mathrm{diag}"'] & P(PA\otimes PA) \arrow[r, "(\delta\circ e\otimes\mathrm{id})_*"] \arrow[d, "\Delta"] & P(PA\otimes PA) \arrow[r, "\nabla_*"] \arrow[dr, "\mathrm{id}"'] & PP(A\otimes A) \arrow[r, "E"] \arrow[d, "\Delta_*"] & P(A\otimes A) \arrow[dd, "\Delta"] \\
 & PPA\otimes PPA \arrow[drr, "(\delta\circ e)_*\otimes\mathrm{id}"'] & & P(PA\otimes PA) \arrow[d, "\Delta"] & \\
 & & & PPA\otimes PPA \arrow[r, "E\otimes E"] & PA\otimes PA
\end{tikzcd}$$

Therefore $\Delta r = ((E \circ \delta_* \circ e_*) \otimes E)(\mu, \mu)$, which by the right unitality diagram of P (Theorem 2.3.8) is equal to $(e_*\mu, E\mu) = (p, q)$. So r has the right marginals. Moreover,

$$\begin{aligned}
\int_{B\times A} a \, dr(b,a) &= \int_{B\times A} a \, d\big(E \circ \nabla_*(\delta\circ e\otimes \mathrm{id})_* \,\mathrm{diag}_*\, \mu\big)(b,a) \\
&= \int_{P(A\otimes A)} \bigg(\int_{B\otimes A} a \, dr'(b,a) \bigg) d\big(\nabla_*(\delta\circ e\otimes \mathrm{id})_* \,\mathrm{diag}_*\, \mu\big)(r') \\
&= \int_{PA\otimes PA} \bigg(\int_{B\otimes A} a \, dp'(b)\, dq'(a) \bigg) d\big((\delta\circ e\otimes \mathrm{id})_* \,\mathrm{diag}_*\, \mu\big)(p',q') \\
&= \int_{PA\otimes PA} \bigg(\int_{B} dp'(b) \int_A a \, dq'(a) \bigg) d\big((\delta\circ e\otimes \mathrm{id})_* \,\mathrm{diag}_*\, \mu\big)(p',q') \\
&= \int_{PA\otimes PA} p'(B)\, e(q') \, d\big((\delta\circ e\otimes \mathrm{id})_* \,\mathrm{diag}_*\, \mu\big)(p',q') \\
&= \int_{PA\otimes PA} \delta_{e(p')}(B)\, e(q') \, d\big(\,\mathrm{diag}_*\, \mu\big)(p',q')
\end{aligned}$$

$$= \int_{PA} \delta_{e(p')}(B)\, e(p')\, d\mu(p').$$

The integrand is equal to $e(p')$ when $e(p')$ lies inside B, and zero otherwise. Therefore the integral is equal to

$$= \int_{e^{-1}(B)} e(p')\, d\mu(p')$$

$$= \int_B b\, d(e_*\mu)(b) = \int_B b\, dp(b),$$

so equation (4.2.4) holds. By Lemma 4.2.17, then p is a conditional expectation of q. $\square$

So, in particular, the law of total evaluation of Corollary 4.2.2 corresponds indeed to the law of total expectation. Moreover, we have gained an extra interpretation of the partial evaluation order: if $p \preceq_c q$, we can view q as "p plus unbiased noise", or "p after diffusion", or "p plus casual errors".

4.2.2. Convergence properties

Here we will prove that the convex order on an unordered P-algebra satisfies the so-called *Levi property* [AT07, Definition 2.44]: every bounded monotone net converges topologically. The result is similar to Lebesgue's monotone convergence theorem, and it reminds us of Doob's martingale and backward-martingale convergence theorems (see the discussions at the beginning and at the end of [Win85, Section 2.4]). As stated before, elements of PA correspond however to *laws* of random variables, so the convergence results here, from the point of view of random variables, will correspond in general to convergence *in distribution*. It is possible to obtain finer convergence results in this categorical framework, but this will not be pursued in the present work.

The main result, Theorem 4.2.18, is analogous to a result of Winkler [Win85, Theorem 2.4.2]. The theorem there requires the domain to be bounded (since what Winkler calls "measure-convex sets" are necessarily bounded), while ours does not. However, we require the domain to sit in a Banach space, while he only requires a locally convex topological vector space.

Theorem 4.2.18. *Let A be an unordered P-algebra. Let $\{p_\alpha\}$ be a net in PA bounded above by some q.*

(a) If $\{p_\alpha\}$ is monotonically decreasing, then it admits an infimum p, and $p_\alpha \to p$ topologically;

(b) *If $\{p_\alpha\}$ is monotonically increasing, then it admits a supremum p, and $p_\alpha \to p$ topologically.*

First a useful, general lemma, which says that a for a monotone net (or sequence), one accumulation point is enough to have convergence, thanks to monotonicity.

Lemma 4.2.19. *Let K be a compact topological space with a closed partial order $(\leq)$ on it. Let $\{x_\alpha\}$ be a monotone increasing (resp. decreasing) net in X. Then $\{x_\alpha\}$ admits a supremum (resp. infimum), and converges to it topologically.*

Proof. We will prove the statement for increasing nets, the decreasing case is analogous.

First of all, since the space is compact, in order to prove convergence it is enough to prove that $\{x_\alpha\}$ admits a unique accumulation point. So let x, y be accumulation points. Then we can find subnets $\{x_{\alpha_\beta}\}$ and $\{x_{\alpha_\gamma}\}$ converging to x and y, respectively. By finality of subnets, for every β we can find a γ such that $\alpha_\beta \leq \alpha_\gamma$, which by monotonicity implies $x_{\alpha_\beta} \leq x_{\alpha_\gamma}$. Since the relation is closed, this implies that $x \leq y$. In the same way we can conclude that $y \leq x$, which implies $y = x$. So any accumulation point must be unique, and it must be the limit of the net.

Now let x be the limit of $\{x_\alpha\}$. Since the net is monotone, for every $\alpha \leq \beta$, $x_\alpha \leq x_\beta$. We can take the topological limit over β, and closedness gives then $x_\alpha \leq x$ for all α. Therefore x is an upper bound. Now suppose that for some y, $x_\alpha \leq y$ for all α. By closedness again, this implies $x \leq y$, so x is a supremum. □

Thanks to the previous lemma, and to the properness of E proven in 2.6.3, the main result now follows easily:

Proof of Theorem 4.2.18. By hypothesis, $\{p_\alpha\}$ is contained in the down-set $\downarrow \{q\}$. By definition of the order,

$$\downarrow \{q\} = (Pe) \circ E^{-1}(q).$$

Now by Theorem 2.6.7, $E^{-1}(q)$ is compact. Since Pe is continuous, $(Pe) \circ E^{-1}(q)$ must be compact as well. Suppose now $\{p_\alpha\}$ is monotonically increasing (resp. decreasing). By Lemma 4.2.19, it admits a supremum (resp. infimum), and it tends to it topologically. □

4.3. Interaction with the underlying order

In the previous section, we have defined an order on PA, the partial evaluation order, which is different from the usual stochastic order. Suppose now that A is an *ordered* algebra. Here we want to study the interaction between the usual stochastic order on

PA, which we recall is denoted by $\leq$, and the partial evaluation order, which we recall is denoted by $\preceq_c$. From now on A will always assumed to be a P-algebra in L-COMet, i.e. ordered. An example to keep in mind is $\mathbb{R}$ with its usual order. Of course, unordered spaces are included as well as a trivial case.

4.3.1. Transversality

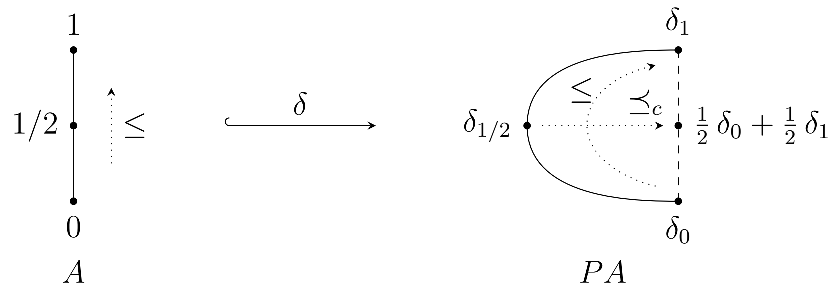
The first result that we have follows quite easily from the law of total expectation, and says that *the only way that p and q are comparable for both orders is if they are equal.* In other words, the two orders are somehow "transverse" to each other:

Corollary 4.3.1. *If p is a partial evaluation of q and $p \leq q$ or $p \geq q$ in the usual stochastic order, then $p = q$.*

Proof. If p is a partial evaluation of q, by Corollary 4.2.2, $e(p) = e(q)$. Then, by Proposition 3.5.11, $p = q$. □

A possible interpretation of this result is the following: we have seen in 4.2.1 that we can interpret $p \preceq_c q$ as "q can be obtained by p by adding unbiased noise". If the noise is really unbiased, then q cannot lie globally "higher" or "lower" than p in the stochastic order. The noise should spread the distribution around the same center of mass. The same reasoning can be done, in the other direction, by thinking of p as a "concentration" of q.

Example 4.3.2. Set A to be the interval $[0, 1]$ with its usual order. We can embed it via δ into PA as its set of extreme points, which is the solid C-shaped line on the right (the picture is intended as a sketch, the real space is infinite-dimensional):



Now the usual stochastic order $\leq$ on PA is directed somewhat "vertically" in the picture (for example, $\delta_0 \leq \delta_1$), while the partial evaluation order is directed "horizontally" (for example, $\delta_{1/2} \preceq_c \frac{1}{2}\,\delta_0 + \frac{1}{2}\,\delta_1$). This is what we mean by "the two orders being transverse to each other".

We can express this property also in terms of the dual system of Definition 2.1.6.

Remark 4.3.3. We know that the orders $\leq$ and $\preceq_c$ induce cones in the space $M(A)$ (Remarks 3.5.10 and 4.2.11). Transversality of $\leq$ and $\preceq_c$, in this setting, means exactly that the two associated cones have trivial intersection.

4.3.2. The lax partial evaluation relation

We have seen that the stochastic and partial evaluation orders are in some way transverse to each other. This makes it difficult to compare distributions which have both different concentration as well as different expected height. We define now a relation that is "diagonal", i.e. which follows both directions one after the other. The resulting order is also known in the literature, and it forms the basis of the so-called *second-order stochastic dominance* [Fis80]. Intuitively, this new order is to the order $\preceq_c$ as supermartingales are to martingales.

Definition 4.3.4. *Let A be an ordered P-algebra. Let $p, q \in PA$. We say that p is a* lax partial evaluation *of q, and we write $p \preceq_l q$, if and only if there exists a $p' \in PA$ such that:*

(a) $p \leq p'$ for the usual stochastic order on PA;

(b) $p' \preceq_c q$ for the partial evaluation order on PA.

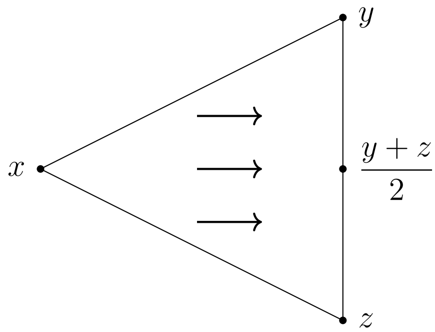
Technically, this is called the *composite relation*. In diagrams, we are defining $\preceq_l$ as the composite arrow:

$$\begin{tikzcd} & p' \arrow[dr, "\preceq_c"] & \\ p \arrow[ur, "\leq"] \arrow[rr, "\preceq_l"'] & & q \end{tikzcd}$$

Equivalently, $p \preceq_l q$ if and only if there exists a $\mu \in PPA$ such that $p \leq (Pe)\mu$ for the usual stochastic order on PA, and $q = E\mu$. In particular, if the order on A is trivial, the orders $\preceq_c$ and $\preceq_l$ coincide.

Intuitively, $p \preceq_l q$ means that in order to obtain q from p one must *first* move the mass upward (stochastic order), and *then* let it spread. Conversely, to obtain p from q, one must first concentrate the mass (partially evaluate), and then move it downward. The order of these operations is important, and it cannot always be interchanged, as the following example shows.

Example 4.3.5. Consider the following triangle in $\mathbb{R}^2$, ordered horizontally from right to left (as the arrows show):



Notice that $x \not\leq y$ and $x \not\leq z$, since the order points exactly to the right. Consider now the measures $p = \delta_x$ and $q = \frac{1}{2}\,\delta_y + \frac{1}{2}\,\delta_z$ in PA. Intuitively, q is "more spread" than p, and also more "to the right" (upwards in the stochastic order). Now, there exist a p' such that $p \leq p' \preceq_c q$, namely a delta at the midpoint of y and z: we can first move p to the right (upward in the order) until the midpoint of y and z, and then spread it vertically to obtain q. But there is no measure q' which is below q in the stochastic order and more spread than x: if we try to spread x nontrivially, or if we move q to the left, we leave the space.

However, when we *can* first spread and then move upwards, we can also first move upwards and then spread:

Proposition 4.3.6. *Suppose that we have $p, q \in PA$ and some $q' \in PA$ such that we have a diagram*

$$\begin{tikzcd} p \arrow[dr, "\preceq_c"'] & & q \\ & q' \arrow[ur, "\leq"'] & \end{tikzcd}$$

Then there exists a $p' \in PA$ completing the diagram to

$$\begin{tikzcd} & p' \arrow[dr, "\preceq_c"] & \\ p \arrow[ur, "\leq"] \arrow[dr, "\preceq_c"'] & & q \\ & q' \arrow[ur, "\leq"'] & \end{tikzcd}$$

which then implies $p \preceq_l q$.

Proof. Suppose that $p \preceq_c q'$. Then by definition there is $\nu \in PPA$ such that $p = e_*\nu$ and $E\nu = q' \leq q$. By the order lifting (Proposition 3.4.8) there exists $\mu \in PPA$ such that $E\mu = q$ and $\nu \leq \mu$. Take now $p' := e_*\mu$. Since $E\mu = q$, $p' \preceq_c q$, and since e_* is monotone, $p = e_*\nu \leq e_*\mu = p'$. $\square$

So, the definition 4.3.4 is the one that "always works". Proposition 4.3.6 is a sort of exchange law for the orders, similar to the exchange law between metric and order in Section 3.6. This suggests a unified treatment in terms of Lawvere metric spaces, which is however beyond the scope of the present work.

Here is an application of Proposition 4.3.6.

Lemma 4.3.7. *The lax partial evaluation relation is transitive.*

Proof. Suppose that we have $p \preceq_l q \preceq_l r$. Then by definition we have p', q' fitting into a diagram

$$\begin{tikzcd} & p' \arrow[dr, "\preceq_c"] & & q' \arrow[dr, "\preceq_c"] & \\ p \arrow[ur, "\leq"] & & q \arrow[ur, "\leq"] & & r \end{tikzcd}$$

Now by Proposition 4.3.6, there exists a $\tilde{q}$ which completes the diagram to

$$\begin{tikzcd} & & \tilde{q} \arrow[dr, "\preceq_c"] & & \\ & p' \arrow[ur, "\leq"] \arrow[dr, "\preceq_c"] & & q' \arrow[dr, "\preceq_c"] & \\ p \arrow[ur, "\leq"] & & q \arrow[ur, "\leq"] & & r \end{tikzcd}$$

Then by transitivity of the two orders,

$$\begin{tikzcd} & \tilde{q} \arrow[dr, "\preceq_c"] & \\ p \arrow[ur, "\leq"] & & r \end{tikzcd}$$

which by definition means that $p \preceq_l r$. $\square$

Reflexivity follows immediately from the reflexivity of $\preceq_c$ and $\leq$.

Remark 4.3.8. A preorder is a monad in the 2-category of relations. Proposition 4.3.6 can be interpreted as a distributive law between the monads $\preceq_c$ and $\leq$, which as it is known makes the composition of the monads a monad itself, i.e. a preorder. This is the abstract reasoning behind the proof of Lemma 4.3.7.

Antisymmetry also follows from antisymmetry of $\preceq_c$ and $\leq$, together with Proposition 4.3.6 and the transversality criterion, Corollary 4.3.1:

Lemma 4.3.9. *The lax partial evaluation relation is antisymmetric.*

Proof. Suppose that $p \preceq_l q$ and $q \preceq_l p$. Then there exist $p', q' \in PA$ such that $p \leq p' \preceq_c q$ and $q \leq q' \preceq_c p$:

$$
\begin{tikzcd}
 & p' \arrow[dr, "\preceq_c"] & \\
p \arrow[ur, "\leq"] & & q \arrow[dl, "\leq"] \\
 & q' \arrow[ul, "\preceq_c"] &
\end{tikzcd}
$$

But then, as the diagram above shows, $p' \preceq_c q \leq q'$. Therefore by Proposition 4.3.6, there exists a $\tilde{q}$ with $p' \leq \tilde{q} \preceq_c q'$, i.e.:

$$
\begin{tikzcd}
 & p' \arrow[dr, "\preceq_c"] \arrow[d, "\leq"] & \\
p \arrow[ur, "\leq"] & \tilde{q} \arrow[d, "\preceq_c"] & q \arrow[dl, "\leq"] \\
 & q' \arrow[ul, "\preceq_c"] &
\end{tikzcd}
$$

Now, as the diagram again shows, $p \leq \tilde{q}$ and $\tilde{q} \preceq_c p$ by transitivity of the two orders. By Corollary 4.3.1, then $p = \tilde{q}$. We are left with

$$
\begin{tikzcd}
 & p' \arrow[dr, "\preceq_c"] \arrow[d, "\leq"] & \\
p \arrow[ur, "\leq"] & p \arrow[d, "\preceq_c"] & q \arrow[dl, "\leq"] \\
 & q' \arrow[ul, "\preceq_c"] &
\end{tikzcd}
$$

where we see that $p \leq p' \leq p$, which implies $p = p'$, and $p \preceq_c q' \preceq_c p$, which implies $p = q'$, by antisymmetry of the two orders. So now we have

$$
\begin{tikzcd}
 & p \arrow[dr, "\preceq_c"] \arrow[d, "\leq"] & \\
p \arrow[ur, "\leq"] & p \arrow[d, "\preceq_c"] & q \arrow[dl, "\leq"] \\
 & p \arrow[ul, "\preceq_c"] &
\end{tikzcd}
$$

where we see that $p \preceq_c q$ and $q \leq p$. Again by Corollary 4.3.1, then $p = q$. $\square$

Closure also holds, using the metric lifting, just like for the partial evaluation case.

Lemma 4.3.10. *The lax partial evaluation relation is closed.*

Proof. Let $\{p_i\}, \{q_i\}$ be Cauchy sequences in PA tending to p and q respectively, and such that $p_i \preceq_l q_i$ for all i. This means that there exists a (generic) sequence $\{\mu_i\}$ in PPA such that for all i, $p_i \le e_* \mu_i$ and $E\mu_i = q_i$. By Corollary 2.6.8, the sequence $\{\mu_i\}$ has at least one accumulation point μ. Now since E is continuous, $E\mu = q$. Moreover, since e_* is continuous, and since the stochastic order is closed, $p \le e_* \mu$. Therefore p is a lax partial evaluation of q. □

We have proven then the analogue of Theorem 4.2.4:

Theorem 4.3.11. *The lax partial evaluation relation is a closed partial order. Therefore,* $(PA, \preceq_l)$ *is an object of* COMet.

Again, the order satisfies the exchange law with the metric, provided that the underlying space does:

Proposition 4.3.12. *Suppose that A satisfies the exchange law. Then $(PA, \preceq_l)$ satisfies the exchange law too.*

Proof. Suppose that A satisfies the exchange law. First of all, by Proposition 3.6.5, $(PA, \le)$ also satisfies the exchange law. Now suppose that $p \preceq_l q$ and $d(q, \bar{q}) < r$. This means that there exists a $p' \in PA$ such that $p \le p' \preceq_c q$. By the exchange law of $\preceq_c$, we know that there exists a $\bar{p'} \in PA$ such that $\bar{p'} \preceq_c \bar{q}$ and $d(p', \bar{p'}) < r$. So we have $p \le p'$ and $d(p', \bar{p'}) < r$. By the exchange law for the stochastic order, then there exists $\bar{p}$ such that $d(p, \bar{p}) < r$ and $\bar{p} \le \bar{p'} \preceq_c \bar{q}$, so that $\bar{p} \preceq_l \bar{q}$. □

Proposition 4.3.13. *For every order P-algebra A, $(PA, \preceq_l)$ is L-ordered.*

Proof. We know that A can be embedded into an ordered Banach space B (Theorem 3.5.6). The space B is itself a P-algebra, and it satisfies the exchange law (Remark 3.6.2). Therefore by Proposition 4.3.12, $(PB, \preceq_l)$ satisfies the exchange law, which implies that it is L-ordered (Proposition 3.6.3). The space $(PA, \preceq_l)$ can be embedded into $(PB, \preceq_l)$ isometrically. We want to prove that the order $(PB, \preceq_l)$ restricts on PA to the order $(PA, \preceq_l)$ (we know this is true for the stochastic order, but a priori not for the order $\preceq_l$). Now suppose that $p, q \in PA$ are such that there exists $p' \in PB$ with $p \le p' \preceq_c q$. Since A is a convex subset of B, PA is closed under taking partial evaluations, so necessarily $p' \in PA$. Therefore $(PA, \preceq_l)$ is L-ordered too. □

Just like for $\preceq_c$, also $\preceq_l$ has a P-algebra structure.

Proposition 4.3.14. *Let A be a P-algebra in* $\mathsf{L\text{-}COMet}$*. The map $E : PPA \to PA$ is also monotone as a map $P(PA, \preceq_l) \to (PA, \preceq_l)$. Therefore, $(PA, \preceq_l)$ is as well a P-algebra.*

Proof. Just like for $\preceq_c$, we can use Lemma 3.5.1. It suffices to show that for every $p, q, r \in PA$ such that $p \preceq_l q$ and every $\lambda \in [0, 1]$,

$$\lambda\, p + (1-\lambda)\, r \quad \preceq_l \quad \lambda\, q + (1-\lambda)\, r.$$

Now suppose that $p \preceq_l q$. Then by definition there exists $p' \in PA$ such that $p \leq p' \preceq_c q$. Take now the probability measure

$$\lambda\, p' + (1-\lambda)\, r.$$

Since E is monotone for the stochastic order, by Lemma 3.5.1 we have that

$$\lambda\, p + (1-\lambda)\, r \quad \leq \quad \lambda\, p' + (1-\lambda)\, r.$$

Jut as well, since E is monotone for the partial evaluation order (Proposition 4.2.10), again by Lemma 3.5.1 we have that

$$\lambda\, p' + (1-\lambda)\, r \quad \preceq_c \quad \lambda\, q + (1-\lambda)\, r.$$

Therefore

$$\lambda\, p + (1-\lambda)\, r \quad \leq \quad \lambda\, p' + (1-\lambda)\, r \quad \preceq_c \quad \lambda\, q + (1-\lambda)\, r,$$

which by definition means

$$\lambda\, p + (1-\lambda)\, r \quad \preceq_l \quad \lambda\, q + (1-\lambda)\, r,$$

$\square$

Therefore, also $(PA, \preceq_l)$ can be embedded in an ordered Banach space.

Remark 4.3.15. In terms of the dual systems 2.1.6, also the order $\preceq_l$ induces a cone in $M(X)$, just like the stochastic order (Remark 3.5.10) and the partial evaluation order (Remark 4.2.11). The cone induced by $\preceq_l$, in particular, is the *Minkowski sum* of the cones of the orders $\leq$ and $\preceq_c$.

We will see that this cone is the *dual cone to monotone convex functions.* More on that in Section 4.4.

Corollary 4.3.16. *The maps $(PA, \leq) \to (PA, \preceq_l)$ and $(PA, \preceq_c) \to (PA, \preceq_l)$ induced by the identity on the underlying spaces are (monotone and) affine.*

We will not label such maps, by a map $(PA, \preceq_c) \to (PA, \preceq_l)$ we will always mean the one given above, unless otherwise specified.

4.4. Universal property and duality

In this section we explore the deep link that there is between the orders $\preceq_c$ and $\preceq_l$ and convex functions. We will show that the orders are uniquely characterized by a universal property, as *oplax codescent objects* [Lac02], by means of a refinement of the ordered Choquet adjunction (3.5.7). We will then study the consequences of this universal property, which establishes a dual characterization of the orders.

Part of the material in this section works for arbitrary 2-monads on locally posetal 2-categories, and can be thought of as an instance of the general theory of codescent objects given by Lack [Lac02]. We will try to keep both notations, categorical and analytic, whenever possible. The duality to real-valued functions, however, is characteristic of the Kantorovich monad.

4.4.1. Universal property

Intuitively, a concave function is a function that is "larger in the middle". Alternatively, integrals of concave maps assign a larger value to more "concentrated" measures. In economics, for example, they correspond for example to risk-averse utility functions. This property is true in general, and has a deep categorical meaning, which we will now try to show.

Let A and B be (ordered) P-algebras. We know that the ordered Choquet adjunction (3.5.7) gives an isomorphism of partial orders between short monotone maps (not necessarily affine) maps $f : A \to B$ and their affine extensions $\tilde{f} : PA \to B$, by the assignments:

$$f \mapsto \tilde{f} := e \circ (Pf),$$

$$\tilde{f} \mapsto f := \tilde{f} \circ \delta.$$

There are two convex structure involved here: the one of A, and the one of PA. The map $\tilde{f}$ is affine for the mixtures in PA, i.e. of *measures*. This does *not* mean that its restriction $f : A \to B$ is affine on A.

Now suppose that $f : A \to B$ is a *concave* map, i.e. a lax P-morphism. This is reflected by the affine extension $\tilde{f}$ in the following way.

Lemma 4.4.1. *Let A and B be P-algebras in* L-COMet. *Let $f : A \to B$ be a morphism of* L-COMet, *i.e. a short monotone map, not necessarily affine. Then $f : A \to B$ is a* *lax P-morphism (concave function) if and only if we have a 2-cell*

$$
\begin{tikzcd}
PPA \arrow[r, "E"] \arrow[d, "Pe"'] \arrow[dr, phantom, "\swarrow"] & PA \arrow[d, "\tilde{f}"] \\
PA \arrow[r, "\tilde{f}"'] & B
\end{tikzcd}
\tag{4.4.1}
$$

or, in terms of traditional inequalities, if and only if:

$$
\int_A f\, d(E\mu) \leq \int_A f\, d(e_*\mu) \tag{4.4.2}
$$

for all $\mu \in PPA$.

This proposition works for any monad on a locally posetal 2-category, and we will give a diagrammatic proof that works in general.

Proof. First suppose (3.5.8). Using $\tilde{f} = e \circ (Pf)$, we can decompose the diagram (4.4.1) as

$$
\begin{tikzcd}
PPA \arrow[rr, "E"] \arrow[dr, "PPf"] \arrow[dd, "Pe"'] & & PA \arrow[dr, "Pf"] & \\
& PPB \arrow[rr, "E"] \arrow[dd, "Pe"] \arrow[dl, phantom, "\swarrow"] & & PB \arrow[dd, "e"] \\
PA \arrow[dr, "Pf"'] & & & \\
& PB \arrow[rr, "e"'] & & B
\end{tikzcd}
$$

where:

- The upper parallelogram commutes by naturality of E;
- The bottom right square is the composition square of B;
- The left parallelogram is exactly the image of (3.5.8) under P.

Vice versa, suppose (4.4.1). Using $f = \tilde{f} \circ \delta$, we can decompose the diagram (3.5.8) as

$$
\begin{tikzcd}
& PA \arrow[rrr, "P\delta"] \arrow[dl, "e"'] \arrow[dr, "\delta"] \arrow[drr, "\mathrm{id}"] & & & PPA \arrow[dl, "E"'] \arrow[dr, "P\tilde{f}"] & \\
A \arrow[dr, "\delta"'] & & PPA \arrow[r, "E"'] \arrow[dl, "Pe"'] \arrow[d, phantom, "\Longleftarrow"] & PA \arrow[dr, "\tilde{f}"] & & PB \arrow[dl, "e"] \\
& PA \arrow[rrr, "\tilde{f}"'] & {} & & B &
\end{tikzcd}
$$

where now:

- The right diamond commutes since $\tilde{f}$ is a P-morphism;
- The two upper triangles are the unit triangles for P;
- The lower trapezoid is exactly (4.4.1);
- The left diamond commutes by the naturality of the unit.

□

Lemma 4.4.1 can be interpreted in terms of the partial evaluation order, as diagram (4.4.1) (equivalently, inequality (4.4.2)) easily shows. In particular, by reversing the 2-cell:[3]

Corollary 4.4.2. *Let $\tilde{f} : PA \to B$ be the affine extension of $f : A \to B$. Lemma 4.4.1 says precisely that $\tilde{f} : PA \to B$ is monotone as a map $(PA, \preceq_c) \to B$ if and only if this f is an oplax P-morphism.*

In other words, any affine map $PA \to B$ preserves the partial evaluation order if and only if it is the affine extension of a convex function.

We know that any affine map $\tilde{f} : PA \to B$ is the affine extension of *some* $f : A \to B$ (actually, a unique f). Lemma 4.4.1 says that if $\tilde{f}$ preserves the partial evaluation order, then this f *must* be a convex function. For $B = \mathbb{R}$, this corresponds to a stronger version of Lemma 4.2.6.

Almost as a corollary, we have the following duality theorem.

Theorem 4.4.3. *Let A, B be P-algebras in* $\mathsf{L\text{-}COMet}$*, and $\tilde{f} : PA \to B$ be short and affine, but not necessarily monotone. Then:*

(a) $\tilde{f}$ is monotone as a map $(PA, \leq) \to B$ if and only if it is the affine extension of a monotone *map $f : A \to B$;*

(b) $\tilde{f}$ is monotone as a map $(PA, \preceq_c) \to B$ if and only if it is the affine extension of a convex *map $f : A \to B$.*

(c) $\tilde{f}$ is monotone as a map $(PA, \preceq_l) \to B$ if and only if it is the affine extension of a convex, monotone *map $f : A \to B$.*

[3] Again, the reversal of the order is purely conventional, since, as we will shortly prove, the partial evaluation order is equivalent to the convex order, which is usually directed from the *more concentrated* to the *less concentrated*.

To prove the theorem we proceed in the following way. We know that the lax partial evaluation order $\preceq_l$ is the composite order of the orders $\leq$ and $\preceq_c$. We then show that a function preserve the composite order if and only if it preserves both orders separately.

Lemma 4.4.4. *Let A be a P-algebra, X be any ordered space, and suppose that let $\tilde{f} : PA \to X$ be a function, not necessarily monotone. Then $\tilde{f}$ is monotone as a map $(PA, \preceq_l) \to X$ if and only if it is monotone as a map $(PA, \preceq_c) \to X$ and as a map $(PA, \leq) \to X$.*

Proof. Let $p, q \in PA$. First of all, $p \leq q$ implies $p \preceq_l q$, and $p \preceq_c q$ also implies $p \preceq_l q$. Therefore, if $\tilde{f}$ is monotone for the order $\preceq_l$, it is necessarily monotone for the orders $\leq$ and $\preceq_c$ separately.

Conversely, suppose that $\tilde{f}$ is monotone for the orders $\leq$ and $\preceq_c$ separately. Suppose that $p \preceq_l q$. Then by definition there exists a $p \in PA$ such that $p \leq p'$ and $p' \preceq_c q$. This implies that $\tilde{f}(p) \leq \tilde{f}(p')$, and $\tilde{f}(p') \preceq_c \tilde{f}(q)$. Again by definition of the order $\preceq_l$, then $\tilde{f}(p) \preceq_l \tilde{f}(q)$. □

The proof of the theorem follows now straightforwardly.

Proof of Theorem 4.4.3.

(a) This is exactly Corollary 3.5.9, following from the Choquet adjunction (3.5.4).

(b) By Corollary 4.4.2, and setting the order on A to be trivial, $\tilde{f}$ preserves the order $\preceq_c$ if and only if it is the affine extension of a convex map.

(c) By Lemma 4.4.4, $\tilde{f}$ is monotone for the order $\preceq_l$ if and only if it is monotone for the orders $\leq$ and $\preceq_c$. By the two previous conditions, $\tilde{f}$ is monotone for the order $\preceq_l$ if and only if it is the affine extension of a short, monotone map. □

We can restate Theorem 4.4.3 in the following equivalent way, which we can think of as a refinement of the Choquet adjunction for the case of algebras.

Theorem 4.4.5. *Let A and B be a P-algebras in* $\mathsf{L\text{-}COMet}$. *The ordered Choquet adjunction* (3.5.7) *restricts to a natural isomorphism of partial orders*

$$P\mathsf{Alg}_o\big(A, B\big) \cong P\mathsf{Alg}_s\big((PA, \preceq_l), B\big) \tag{4.4.3}$$

between convex monotone *maps $A \to B$ and* affine monotone *maps $(PA, \preceq_l) \to B$.*

By taking A trivially ordered, we also obtain that there is a natural isomorphism of partial orders

$$P\mathsf{Alg}_o\big(A,B\big) \cong P\mathsf{Alg}_s\big((PA,\preceq_c),B\big) \tag{4.4.4}$$

between *convex* maps $A \to B$ and *affine monotone* maps $(PA,\preceq_c) \to B$.

Theorem 4.4.5 means precisely that the (lax) partial evaluation order satisfies a *2-dimensional universal property* in $P\mathsf{Alg}_s$: for every B and every (monotone) convex map $f : A \to B$, there exists a unique monotone affine map $\tilde{f}$ making this diagram commute:

$$\begin{tikzcd}
A \arrow[d, "\delta"'] \arrow[dr, "f"] & \\
(PA,\preceq_l) \arrow[r, dashed, "\tilde{f}"'] & B
\end{tikzcd}$$

This particular universal property was first studied by Lack in [Lac02], and given the following name: $(PA,\preceq_l)$ is the *oplax codescent object* of the P-algebra A.

The 2-dimensional nature of this colimit is visible in the following equivalent form: for every affine monotone map $PA \to B$ with a 2-cell

$$\begin{tikzcd}
& PA \arrow[dr, "\tilde{f}"] \arrow[dd, phantom, "\Uparrow"] & \\
PPA \arrow[ur, "E"] \arrow[dr, "Pe"'] & & B \\
& PA \arrow[ur, "\tilde{f}"'] &
\end{tikzcd}$$

there exists a unique affine monotone map $(PA,\preceq_l) \to B$ making this diagram commute:

$$\begin{tikzcd}
& PA \arrow[dr] \arrow[drr, "\tilde{f}"] \arrow[dd, phantom, "\Uparrow"] & & \\
PPA \arrow[ur, "E"] \arrow[dr, "Pe"'] & & (PA,\preceq_l) \arrow[r, dashed] & B \\
& PA \arrow[ur] \arrow[urr, "\tilde{f}"'] & &
\end{tikzcd}$$

We see that this colimit is similar to a coequalizer, but in an *oplax* way: the coequalizer of E and Pe (which is exactly $e : PA \to A$) *identifies* any two measures p, q such that $p = e_*\mu$ and $q = E\mu$. In the order $(PA,\preceq_l)$, instead, p and q are not identified, there is merely an arrow between them, an arrow of $\preceq_c$. This oplax version of a coequalizer is known in the literature as *op-coinserter* (see for example [Lac02]). Just as it happens in our case, in any locally posetal 2-category the oplax codescent object is simply given by an op-coinserter.

The order $\preceq_l$ is now uniquely characterized by a universal property: it is in some sense inevitable, as it arises naturally from the 2-dimensional theory of monads and algebras. By choosing A trivially ordered, the same is true for $\preceq_c$. The concept of partial evaluation, which was motivated only by some intuitions about formal expressions, now has a precise categorical characterization.

Without reference to monads, Theorem 4.4.5 implies the following statement:

Corollary 4.4.6. *Let A and B be closed convex subsets of ordered Banach spaces. There is a bijective correspondence inducing an isomorphism of partial orders between* convex monotone *maps $A \to B$ and* affine *maps $PA \to B$* which are monotone for the order $\preceq_l$.

4.4.2. Applications of the universal property

Theorem 4.4.5 has a number of consequences of interest.

Corollary 4.4.7. *The assignment $A \mapsto (PA, \preceq_l)$ gives a left adjoint to the inclusion functor $P\mathsf{Alg}_s \hookrightarrow P\mathsf{Alg}_l$.*

It is interesting to look at the unit and counit of this adjunction, which are induced from the unit and counit δ, e of the ordered Choquet adjunction 3.5.7: For each algebra A,

- The unit is given by the lax P-morphism $A \to (PA, \preceq_l)$ induced by the unit $\delta : A \to PA$;
- The counit is given by the strict P-morphism $(PA, \preceq_l) \to A$ induced by the counit of the adjunction, which is the algebra map $e : PA \to A$.

Whenever it does not lead to confusion, we will call the maps δ and e in the same way (always specifying their domain and codomain).

The first condition, somewhat counterintuitive at first, is that $\delta : A \to (PA, \preceq_l)$, which we know is monotone, is also a *convex map*. Let's see why. We have a diagram:

$$\begin{tikzcd} PA \arrow[r, "P\delta"] \arrow[d, "e"'] \arrow[dr, "\mathrm{id}"] & P(PA, \preceq_c) \arrow[d, "E"] \\ A \arrow[r, "\delta"'] \arrow[ur, phantom, "\Nearrow" near start] & (PA, \preceq_c) \end{tikzcd}$$

where the upper triangle commutes by the right unitality diagram of P, and the 2-cell $\delta \circ e \Rightarrow \mathrm{id}$ comes from the fact that $\delta \circ e(p) \preceq_c p$ trivially (Corollary 4.2.3). Therefore $\delta : A \to (PA \preceq_c)$ is an oplax P-morphism. The same can be said about $\delta : A \to (PA \preceq_l)$.

The second condition says that $e : (PA, \preceq_l) \to A$, which we know is affine, is also *monotone*. We know it is monotone for the stochastic order $\leq$, since A is an ordered algebra, but we need to show that it is also monotone for the partial evaluation order $\preceq_c$. So suppose $p \preceq_c q$ in $(PA, \preceq_c)$. Then by the "law of total evaluation", $e(p) = e(q)$, so in particular, $e(p) \leq e(q)$. Therefore $e : (PA, \preceq_l) \to A$ is monotone.

Here is a second important order-theoretical consequence.

Proposition 4.4.8. *For any algebra A, the maps δ and e establish a Galois connection between A and $(PA, \preceq_l)$: for every $a \in A$ and $p \in PA$,*

$$\delta_a \preceq_l p \quad \text{if and only if} \quad a \leq e(p).$$

Proof. Let A be a P-algebra. We have the following diagrams

$$\begin{tikzcd} A \arrow[r, "\delta"] \arrow[dr, equal] & (PA, \preceq_l) \arrow[d, "e"] \arrow[dl, phantom, "=" near start] & (PA, \preceq_l) \arrow[r, "e"] \arrow[dr, equal] & A \arrow[d, "\delta"] \arrow[dl, phantom, "\Swarrow" near start] \\ & A & & (PA, \preceq_l) \end{tikzcd} \tag{4.4.5}$$

where the first diagram commutes by the unit condition of A, and the second one has a 2-cell, as we saw before, by Corollary 4.2.3. The conditions $\mathrm{id} \leq e \circ \delta$ (implied by $\mathrm{id} = e \circ \delta$) and $\delta \circ e \preceq_l \mathrm{id}$ give then a Galois connection. $\square$

The closure operator on $(PA, \preceq_l)$ associated to the Galois connection is $\delta \circ e$, which we can view as "center of mass", or "total evaluation". Therefore we can also view A as the set of invariant elements of $(PA, \preceq_l)$ under this closure operator.

4.4.3. Duality

Let A and B be P-algebras in $\mathsf{L\text{-}COMet}$, and $f : A \to B$ a short, monotone, convex map. We have seen (Theorem 4.4.3) that whenever $p \preceq_l q$, then

$$\int f \, dp \leq \int f \, dq,$$

and analogous statements hold for the orders $\leq$ and $\preceq_c$ as well (Corollary 3.5.9 and Lemma 4.4.1). It is now natural to ask the dual question: given $p \leq q \in PA$, is it true that $p \preceq_l q$ if and only if for all convex monotone functions, $\int f \, dp \leq \int f \, dq$? One can ask similar questions for the orders $\leq$ and $\preceq_c$.

First of all, the answer to these questions depends on whether we *fix* the space B (for example $B = \mathbb{R}$) and we look just at maps into B, or we allow maps into all possible

P-algebras. For the second case, the answer is always positive, in a somewhat trivial way (by the Yoneda lemma), and it works for arbitrary locally posetal 2-categories. A more interesting question, in our case, is by fixing $B = \mathbb{R}$.

In this case, the statement still holds true, thanks to the Hahn-Banach theorem (or categorically, thanks to the fact that $\mathbb{R}$ is a 2-coseparator, see 3.5.2).

Theorem 4.4.9. *Let A be a P-algebra, and let $p, q \in PA$. Then:*

(a) $p \leq q$ if and only if for every monotone *$f : A \to \mathbb{R}$, its affine extension $\tilde{f}$ satisfies $\tilde{f}(p) \leq \tilde{f}(q)$.*

(b) $p \preceq_c q$ if and only if for every convex *$f : A \to \mathbb{R}$, its affine extension $\tilde{f}$ satisfies $\tilde{f}(p) \leq \tilde{f}(q)$.*

(c) $p \preceq_l q$ if and only if for every convex monotone *$f : A \to \mathbb{R}$, its affine extension $\tilde{f}$ satisfies $\tilde{f}(p) \leq \tilde{f}(q)$.*

Proof. We know that all three orders $\leq$, $\preceq_c$ and $\preceq_l$ equip PA with the structure of a P-algebra. By Corollary 3.5.17, we can determine the orders just by looking at affine, monotone functions into $\mathbb{R}$. Now by Theorem 4.4.3,

(a) Affine monotone functions $(PA, \leq) \to \mathbb{R}$ are exactly the affine extensions of monotone functions $f : A \to \mathbb{R}$;

(b) Affine monotone functions $(PA, \preceq_c) \to \mathbb{R}$ are exactly the affine extensions of convex functions $f : A \to \mathbb{R}$;

(c) Affine monotone functions $(PA, \preceq_l) \to \mathbb{R}$ are exactly the affine extensions of convex monotone functions $f : A \to \mathbb{R}$; $\square$

All functions are assumed short, but by linearity the same holds equivalently for Lipschitz functions. Without reference to monads, Theorem 4.4.9 reads this way:

Corollary 4.4.10. *Let A be a closed convex subset of an ordered Banach space, and let $p, q \in PA$. Consider the following inequality:*

$$\int f \, dp \leq \int f \, dq. \tag{4.4.6}$$

Then:

(a) $p \leq q$ if and only if (4.4.6) holds for every Lipschitz monotone *$f : A \to \mathbb{R}$.*

(b) $p \preceq_c q$ if and only if (4.4.6) holds for every Lipschitz convex *$f : A \to \mathbb{R}$.*

(c) $p \preceq_l q$ if and only if (4.4.6) *holds for every Lipschitz* convex monotone $f : A \to \mathbb{R}$.

Remark 4.4.11. In terms of the dual systems of Definition 2.1.6, we know (Remarks 3.5.10, 4.2.11, and 4.3.15) that the three orders induce cones in $M(A)$. We also know (Remark 3.5.10) that the cone associated to the stochastic order $\leq$ is the dual cone to monotone functions $C_{\leq}$. Corollary 4.4.10 implies analogous statements for the other two orders:

(a) The cone associated to the partial evaluation order $\preceq_c$ is the dual cone to convex functions C_c ;

(b) The cone associated to the lax partial evaluation order $\preceq_l$ is the dual cone to convex monotone functions C_l.

We know moreover (Remark 4.3.15) that the cone of $\preceq_l$ is the Minkowski sum of the cones of $\leq$ and $\preceq_c$. Therefore, the statements above imply that

$$(C_{\leq} \cap C_c)^* = (C_{\leq})^* + (C_c)^*.$$

Remark 4.4.12. If one interprets functions $f : A \to \mathbb{R}$ as utility functions, as in economics, we then have the following very appealing interpretations:

(a) $p \leq q$ if and only if for every utility function *compatible with the order*, the expected utility with measure p is less or equal than the expected utility with measure q;

(b) $p \preceq_c q$ if and only if for every *risk-seeking* utility function, the expected utility with measure p is less or equal than the expected utility with measure q;

(c) $p \preceq_l q$ if and only if for every *risk-seeking* utility function *compatible with the order*, the expected utility with measure p is less or equal than the expected utility with measure q.

(Equivalently, the same statement for risk-averse utilities can be obtained by reversing the order $\preceq_c$, considering "larger" the more concentrated measures.)

Results of this kind have been proven many times in the literature in different contexts, for example for the case of $A = \mathbb{R}$ [Bla51, Str65, RS70], and for the case of unordered A (as in [Win85, Theorem 1.3.6], where however the full equivalence holds only for a bounded region). As far as we know, however, this statement had never been proven for general ordered Banach spaces. Moreover, in our setting, it is enough to restrict to *Lipschitz* maps $A \to \mathbb{R}$.

A. Additional category theory material

This appendix contains some material of purely categorical nature, which is used in the main text of the work. In particular:

- Section A.1 contains the rigorous definitions of bimonoidal monads, which we use in Sections 1.2, 2.5, and 3.4.2 in order to talk about joints and marginals;

- Section A.2 contains a result about Kan extensions of lax monoidal functors, used in Section 2.3 to prove the monad structure of P from its universal property.

Additional context for both sections is given in the papers [FP17] and [FP18]. For all the details we refer to dedicated texts in category theory, for example [Mac00] for a general treatment, and [AM10] for monoidal categories and functors.

A.1. Monoidal, opmonoidal and bimonoidal monads

We recall the definition of the different monoidal structures for a functor, for the case of braided (including symmetric) monoidal categories. For more results and more general definitions, we refer to [AM10].

Let $(\mathsf{C}, \otimes)$ and $(\mathsf{D}, \otimes)$ be braided monoidal categories.

Definition A.1.1. *A* lax monoidal functor $(\mathsf{C}, \otimes) \to (\mathsf{D}, \otimes)$ *is a triple* (F, η, ∇)*, such that:*

(a) $F : C \to D$ *is a functor;*

(b) The "unit" $\eta : 1_{\mathsf{D}} \to F(1_{\mathsf{C}})$ *is a morphism of* D*;*

(c) The "composition" $\nabla : F(-) \otimes F(-) \Rightarrow F(- \otimes -)$ *is a natural transformation of functors* $\mathsf{C} \times \mathsf{C} \to \mathsf{D}$*;*

(d) *The following "associativity" diagram commutes for every* X, Y, Z *in* C*:*

$$\begin{tikzcd}
(FX \otimes FY) \otimes FZ \arrow[r, "\cong"] \arrow[d, "\nabla_{X,Y} \otimes \mathrm{id}"] & FX \otimes (FY \otimes FZ) \arrow[d, "\mathrm{id} \otimes \nabla_{Y,Z}"] \\
F(X \otimes Y) \otimes FZ \arrow[d, "\nabla_{X \otimes Y, Z}"] & FX \otimes F(Y \otimes Z) \arrow[d, "\nabla_{X, Y \otimes Z}"] \\
F((X \otimes Y) \otimes Z) \arrow[r, "\cong"] & F(X \otimes (Y \otimes Z))
\end{tikzcd}$$

(e) *The following "unitality" diagrams commute for every* X *in* C*:*

$$\begin{tikzcd}
1_{\mathsf{D}} \otimes FX \arrow[r, "\eta \otimes \mathrm{id}"] \arrow[d, "\cong"] & F(1_{\mathsf{C}}) \otimes FX \arrow[d, "\nabla_{1_{\mathsf{C}}, X}"] \\
FX & F(1_{\mathsf{C}} \otimes X) \arrow[l, "\cong"]
\end{tikzcd}
\qquad
\begin{tikzcd}
FX \otimes 1_{\mathsf{D}} \arrow[r, "\mathrm{id} \otimes \eta"] \arrow[d, "\cong"] & FX \otimes F(1_{\mathsf{C}}) \arrow[d, "\nabla_{X, 1_{\mathsf{C}}}"] \\
FX & F(X \otimes 1_{\mathsf{C}}) \arrow[l, "\cong"]
\end{tikzcd}$$

We say that (F, η, ∇) *is also* braided*, or* symmetric *if* C *is symmetric, if in addition the multiplication commutes with the braiding:*

$$\begin{tikzcd}
FX \otimes FY \arrow[r, "\cong"] \arrow[d, "\nabla"] & FY \otimes FX \arrow[d, "\nabla"] \\
F(X \otimes Y) \arrow[r, "\cong"] & F(Y \otimes X)
\end{tikzcd}$$

Definition A.1.2. *Let* (F, η_F, ∇_F) *and* (G, η_G, ∇_G) *be lax monoidal functors* $(\mathsf{C}, \otimes) \to (\mathsf{D}, \otimes)$*. A* lax monoidal natural transformation*, or just* monoidal natural transformation *when it's clear from the context, is a natural transformation* $\alpha : F \Rightarrow G$ *which is compatible with the unit and multiplication map. In particular, the following diagrams must commute (for all* $X, Y \in \mathsf{C}$*):*

$$\begin{tikzcd}
1_{\mathsf{D}} \arrow[r, "\eta_F"] \arrow[dr, "\eta_G"'] & F(1_{\mathsf{C}}) \arrow[d, "\alpha_{1_{\mathsf{C}}}"] \\
& G(1_{\mathsf{C}})
\end{tikzcd}
\qquad
\begin{tikzcd}
FX \otimes FY \arrow[r, "\nabla_F"] \arrow[d, "\alpha_X \otimes \alpha_Y"] & F(X \otimes Y) \arrow[d, "\alpha_{X \otimes Y}"] \\
GX \otimes GY \arrow[r, "\nabla_G"] & G(X \otimes Y)
\end{tikzcd}$$

Definition A.1.3. *An* oplax monoidal functor $(\mathsf{C}, \otimes) \to (\mathsf{D}, \otimes)$ *is a triple* (F, ϵ, Δ)*, such that:*

(a) $F : C \to D$ *is a functor;*

(b) *The "counit"* $\epsilon : F(1_{\mathsf{C}}) \to 1_{\mathsf{D}}$ *is a morphism of* D*;*

(c) *The "comultiplication"* $\Delta : F(-\otimes -) \Rightarrow F(-)\otimes F(-)$ *is a natural transformation of functors* $\mathsf{C}\times\mathsf{C}\to\mathsf{D}$*;*

(d) *The following "coassociativity" diagram commutes for every* X, Y, Z *in* C*:*

$$\begin{tikzcd}
F((X\otimes Y)\otimes Z) \arrow[r, "\cong"] \arrow[d, "\Delta_{X\otimes Y,Z}"] & F(X\otimes(Y\otimes Z)) \arrow[d, "\Delta_{X,Y\otimes Z}"] \\
F(X\otimes Y)\otimes FZ \arrow[d, "\Delta_{X,Y}\otimes\mathrm{id}"] & FX\otimes F(Y\otimes Z) \arrow[d, "\mathrm{id}\otimes\Delta_{Y,Z}"] \\
(FX\otimes FY)\otimes FZ \arrow[r, "\cong"] & FX\otimes(FY\otimes FZ)
\end{tikzcd}$$

(e) *The following "counitality" diagrams commute for every* X *in* C*:*

$$\begin{tikzcd}
F(1_{\mathsf{C}}\otimes X) \arrow[r, "\Delta_{1_{\mathsf{C}},X}"] \arrow[d, "\cong"] & F(1_{\mathsf{C}})\otimes FX \arrow[d, "\epsilon\otimes\mathrm{id}"] \\
FX & 1_{\mathsf{D}}\otimes FX \arrow[l, "\cong"]
\end{tikzcd}
\qquad
\begin{tikzcd}
F(X\otimes 1_{\mathsf{C}}) \arrow[r, "\Delta_{X,1_{\mathsf{C}}}"] \arrow[d, "\cong"] & FX\otimes F(1_{\mathsf{C}}) \arrow[d, "\mathrm{id}\otimes\epsilon"] \\
FX & FX\otimes 1_{\mathsf{D}} \arrow[l, "\cong"]
\end{tikzcd}$$

We say that (F,ϵ,Δ) *is also* braided*, or* symmetric *if* C *is symmetric, if in addition the comultiplication commutes with the braiding:*

$$\begin{tikzcd}
F(X\otimes Y) \arrow[r, "\cong"] \arrow[d, "\Delta"] & F(Y\otimes X) \arrow[d, "\Delta"] \\
FX\otimes FY \arrow[r, "\cong"] & FY\otimes FX
\end{tikzcd}$$

Definition A.1.4. *Let* (F,ϵ_F,Δ_F) *and* (G,ϵ_G,Δ_G) *be oplax monoidal functors* $(\mathsf{C},\otimes)\to(\mathsf{D},\otimes)$*. An* oplax monoidal natural transformation*, or just* monoidal natural transformation *when it's clear from the context, is a natural transformation* $\alpha : F\Rightarrow G$ *which is compatible with the counit and comultiplication map. In particular, the following diagrams must commute (for all* $X, Y\in\mathsf{C}$*):*

$$\begin{tikzcd}
1_{\mathsf{D}} & F(1_{\mathsf{C}}) \arrow[l, "\epsilon_F"'] \arrow[d, "\alpha_{1_{\mathsf{C}}}"] \\
& G(1_{\mathsf{C}}) \arrow[ul, "\epsilon_G"]
\end{tikzcd}
\qquad
\begin{tikzcd}
FX\otimes FY \arrow[d, "\alpha_X\otimes\alpha_Y"] & F(X\otimes Y) \arrow[l, "\Delta_F"'] \arrow[d, "\alpha_{X\otimes Y}"] \\
GX\otimes GY & G(X\otimes Y) \arrow[l, "\Delta_G"']
\end{tikzcd}$$

Definition A.1.5. *A* bilax monoidal functor $(\mathsf{C},\otimes)\to(\mathsf{D},\otimes)$ *is a "quintuplet"* $(F,\eta,\nabla,\epsilon,\Delta)$ *such that:*

(a) $(F,\eta,\nabla) : (\mathsf{C},\otimes)\to(\mathsf{D},\otimes)$ *is a lax monoidal functor;*

(b) $(F, \epsilon, \Delta) : (\mathsf{C}, \otimes) \to (\mathsf{D}, \otimes)$ *is an oplax monoidal functor;*

(c) *The following "bimonoidality" diagram commutes:*

$$
\begin{tikzcd}
 & F(W \otimes X) \otimes F(Y \otimes Z) \arrow[dl, "\nabla_{W\otimes X, Y \otimes Z}"'] \arrow[dr, "\Delta_{W,X} \otimes \Delta_{Y,Z}"] & \\
F(W \otimes X \otimes Y \otimes Z) \arrow[d, "\cong"'] & & FW \otimes FX \otimes FY \otimes FZ \arrow[d, "\cong"] \\
F(W \otimes Y \otimes X \otimes Z) \arrow[dr, "\Delta_{W\otimes Y, X \otimes Z}"'] & & FW \otimes FY \otimes FX \otimes FZ \arrow[dl, "\nabla_{W,Y} \otimes \nabla_{X,Z}"] \\
 & F(W \otimes Y) \otimes F(X \otimes Z) &
\end{tikzcd}
$$

(d) *The following three "unit/counit" diagrams commute:*

$$
\begin{tikzcd}
1 \arrow[r, "\eta"] \arrow[dr, equal] & F(1) \arrow[d, "\epsilon"] \\
 & 1
\end{tikzcd}
\qquad
\begin{tikzcd}
1 \arrow[r, "\eta"] \arrow[d, "\cong"'] & F(1) \arrow[r, "\cong"] & F(1 \otimes 1) \arrow[d, "\Delta_{1,1}"] \\
1 \otimes 1 \arrow[rr, "\eta \otimes \eta"'] & & F(1) \otimes F(1)
\end{tikzcd}
$$

$$
\begin{tikzcd}
1 & F(1) \arrow[l, "\epsilon"'] & F(1 \otimes 1) \arrow[l, "\cong"'] \\
1 \otimes 1 \arrow[u, "\cong"] & & F(1) \otimes F(1) \arrow[u, "\nabla_{1,1}"'] \arrow[ll, "\epsilon \otimes \epsilon"]
\end{tikzcd}
$$

Definition A.1.6. *Let* $(F, \epsilon_F, \Delta_F)$ *and* $(G, \epsilon_G, \Delta_G)$ *be bilax monoidal functors* $(\mathsf{C}, \otimes) \to (\mathsf{D}, \otimes)$. *A* bilax monoidal natural transformation, *or just* monoidal natural transformation *when it's clear from the context, is a natural transformation* $\alpha : F \Rightarrow G$ *which is a lax and oplax natural transformation.*

Definition A.1.7. *Now, we define:*

- *A* monoidal monad *is a monad in the bicategory of monoidal categories, lax monoidal functors, and monoidal natural transformations;*
- *An* opmonoidal monad *is a monad in the bicategory of monoidal categories, oplax monoidal functors, and monoidal natural transformations;*
- *A* bimonoidal monad *is a monad in the bicategory of braided monoidal categories, bilax monoidal functors, and monoidal natural transformations.*

In the third definition, we need the symmetry (or at least a braiding) in order to express the bimonoid equation that is part of the definition of bilax monoidal functor [AM10], even if the functor itself if not braided. If the functor is braided, we can define in addition:

- *A* braided *(resp.* symmetric*)* monoidal monad *is a monad in the bicategory of braided (resp. symmetric) monoidal categories, braided lax monoidal functors, and monoidal natural transformations;*
- *A* braided *(resp.* symmetric*)* opmonoidal monad *is a monad in the bicategory of braided (resp. symmetric) monoidal categories, braided oplax monoidal functors, and monoidal natural transformations;*
- *A* braided *(resp.* symmetric*)* bimonoidal monad *is a monad in the bicategory of braided (resp. symmetric) monoidal categories, braided bilax monoidal functors, and monoidal natural transformations.*

A.2. Kan extensions of lax monoidal functors

There are some results on when a left Kan extension of lax or strong monoidal functors is again monoidal [MT08, Theorem 1], [Pat12, Proposition 4] in such a way that the Kan extension also holds in MonCat, which is the bicategory of monoidal categories, lax monoidal functors, and monoidal transformations. 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 [Kou15, Theorem 1.1b], [Web16, Theorem 2.4.4], which can be applied to the monoidal category 2-monad. Since neither of these results applies verbatim to our situation, we derive a result of this type tailored to our needs.

For a monoidal category C, we denote its unit $e : 1 \to \mathsf{C}$ and multiplication $\otimes : \mathsf{C} \times \mathsf{C} \to \mathsf{C}$ 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.

Theorem A.2.1. *Let the following hypotheses be satisfied:*

- *In* MonCat*, we have a diagram*

$$\begin{tikzcd} \mathsf{C} \arrow[rr, "F"] \arrow[dr, "G"'] & \arrow[d, phantom, "\Downarrow\lambda"] & \mathsf{D} \\ & \mathsf{C}' \arrow[ur, "L"'] & \end{tikzcd} \tag{A.2.1}$$

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

$$\begin{tikzcd} \mathsf{C}\times\mathsf{C} \arrow[rr, "F\times F"] \arrow[dr, "G\times G"'] & & \mathsf{D}\times\mathsf{D} \arrow[r, "\otimes"] & \mathsf{D} \\ & \mathsf{C}'\times\mathsf{C}' \arrow[ur, "L\times L"'] \arrow[u, phantom, "\Downarrow\lambda\times\lambda"] & & \end{tikzcd} \tag{A.2.2}$$

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

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 2.3.9, but not for the proof of Theorem 2.3.3, for which we really need Theorem A.2.1.

Proof. Given a lax monoidal functor $X : \mathsf{C}' \to \mathsf{D}$ 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

$$\begin{tikzcd} \mathsf{C} \arrow[rr, "F"] \arrow[dr, "G"'] & \arrow[d, phantom, "\Downarrow\lambda"] & \mathsf{D} \\ & \mathsf{C}' \arrow[ur, bend left, "L"] \arrow[ur, bend right, "X"'] \arrow[ur, phantom, "\Rightarrow u"] & \end{tikzcd} \equiv \begin{tikzcd} \mathsf{C} \arrow[rr, "F"] \arrow[dr, "G"'] & \arrow[d, phantom, "\Downarrow\chi"] & \mathsf{D} \\ & \mathsf{C}' \arrow[ur, bend right, "X"'] & \end{tikzcd} \tag{A.2.3}$$

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

$$\begin{tikzcd} & \mathsf{C}' \arrow[dd, bend right, "L"'] \arrow[dd, bend left, "X"] \arrow[dd, phantom, "\Rightarrow u"] \\ 1 \arrow[ur, "e"] \arrow[dr, "e"'] \arrow[r, phantom, "\Uparrow\eta_L"] & \\ & \mathsf{D} \end{tikzcd} \equiv \begin{tikzcd} & \mathsf{C}' \arrow[dd, bend left, "X"] \\ 1 \arrow[ur, "e"] \arrow[dr, "e"'] \arrow[r, phantom, "\Uparrow\eta_X"] & \\ & \mathsf{D} \end{tikzcd} \tag{A.2.4}$$

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

$$
\begin{tikzcd}
 & \mathsf{C} \arrow[dr, bend left, "G"] & \\
1 \arrow[ur, "e"] \arrow[rr, "e"] \arrow[dr, "e"'] \arrow[r, phantom, "\Uparrow\eta_G"] \arrow[r, phantom, shift right=3, "\Nearrow\eta_L"'] & {} & \mathsf{C}' \arrow[dl, bend left, "L"] \\
 & \mathsf{D} &
\end{tikzcd}
\equiv
\begin{tikzcd}
 & \mathsf{C} \arrow[dr, bend left, "G"] \arrow[dd, bend right, "F"'] & \\
1 \arrow[ur, "e"] \arrow[dr, "e"'] \arrow[r, phantom, "\Nearrow\eta_F"] & {} \arrow[r, phantom, "\overset{\lambda}{\Rightarrow}"] & \mathsf{C}' \arrow[dl, bend left, "L"] \\
 & \mathsf{D} &
\end{tikzcd}
\tag{A.2.5}
$$

and similarly for χ. Since η_G is an isomorphism, (A.2.4) follows if we can prove it after postcomposing with η_G,

$$
\begin{tikzcd}
 & \mathsf{C} \arrow[dr, "G"] & \\
1 \arrow[ur, "e"] \arrow[rr, "e"] \arrow[dr, "e"'] \arrow[r, phantom, "\Uparrow\eta_G \cong"] \arrow[r, phantom, shift right=3, "\Nearrow\eta_L"'] & {} & \mathsf{C}' \arrow[dl, bend right, "L"'] \arrow[dl, bend left, "X"] \arrow[dl, phantom, "\Searrow u"] \\
 & \mathsf{D} &
\end{tikzcd}
\equiv
\begin{tikzcd}
 & \mathsf{C} \arrow[dr, "G"] \arrow[dd, "F"] & \\
1 \arrow[ur, "e"] \arrow[dr, "e"'] \arrow[r, phantom, "\Nearrow\eta_F"] & {} \arrow[r, phantom, "\overset{\lambda}{\Rightarrow}"] & \mathsf{C}' \arrow[dl, bend right, "L"'] \arrow[dl, bend left, "X"] \arrow[dl, phantom, "\Searrow u"] \\
 & \mathsf{D} &
\end{tikzcd}
$$

$$
\equiv
\begin{tikzcd}
 & \mathsf{C} \arrow[dr, "G"] \arrow[dd, "F"] & \\
1 \arrow[ur, "e"] \arrow[dr, "e"'] \arrow[r, phantom, "\Nearrow\eta_F"] & {} \arrow[r, phantom, "\overset{\chi}{\Rightarrow}"] & \mathsf{C}' \arrow[dl, bend left, "X"] \\
 & \mathsf{D} &
\end{tikzcd}
\equiv
\begin{tikzcd}
 & \mathsf{C} \arrow[dr, "G"] & \\
1 \arrow[ur, "e"] \arrow[rr, "e"] \arrow[dr, "e"'] \arrow[r, phantom, "\Uparrow\eta_G \cong"] \arrow[r, phantom, shift right=3, "\Nearrow\eta_X"'] & {} & \mathsf{C}' \arrow[dl, bend left, "X"] \\
 & \mathsf{D} &
\end{tikzcd}
$$

which proves the claim.

Proving compatibility with the multiplication

$$
\begin{tikzcd}
\mathsf{C}' \times \mathsf{C}' \arrow[r, "\otimes"] \arrow[d, bend right=50, "L\times L"'] \arrow[d, bend left=50, "X\times X"] \arrow[d, phantom, "\overset{u\times u}{\Rightarrow}"] & \mathsf{C}' \arrow[d, bend left=50, "X"] \arrow[dl, phantom, "\Nearrow\mu_X"] \\
\mathsf{D} \times \mathsf{D} \arrow[r, "\otimes"'] & \mathsf{D}
\end{tikzcd}
\equiv
\begin{tikzcd}
\mathsf{C}' \times \mathsf{C}' \arrow[r, "\otimes"] \arrow[d, bend right=50, "L\times L"'] \arrow[dr, phantom, "\Nearrow\mu_L"] & \mathsf{C}' \arrow[d, bend right=50, "L"'] \arrow[d, bend left=50, "X"] \arrow[d, phantom, "\overset{u}{\Rightarrow}"] \\
\mathsf{D} \times \mathsf{D} \arrow[r, "\otimes"'] & \mathsf{D}
\end{tikzcd}
\tag{A.2.6}
$$

works similarly, but is a bit trickier. We use compatibility of λ with the multiplication

$$
\begin{tikzcd}
\mathsf{C}\times\mathsf{C} \arrow[rr, "\otimes"] \arrow[dr, bend left, "G\times G"] \arrow[dd, bend right=60, "F\times F"'] & & \mathsf{C} \arrow[dr, bend left, "G"] & \\
\arrow[r, phantom, "\overset{\lambda\times\lambda}{\Longrightarrow}"] & \mathsf{C}'\times\mathsf{C}' \arrow[rr, "\otimes"] \arrow[dl, bend left, "L\times L"] \arrow[ur, phantom, "\Uparrow\mu_G"] \arrow[dr, phantom, "\Uparrow\mu_L"] & & \mathsf{C}' \arrow[dl, bend left, "L"] \\
\mathsf{D}\times\mathsf{D} \arrow[rr, "\otimes"'] & & \mathsf{D} &
\end{tikzcd}
\equiv
\begin{tikzcd}
\mathsf{C}\times\mathsf{C} \arrow[rr, "\otimes"] \arrow[dd, bend right=60, "F\times F"'] \arrow[ddrr, phantom, "\Nearrow\mu_F"] & & \mathsf{C} \arrow[dr, bend left, "G"] \arrow[dd, bend right, "F"'] & \\
& & \arrow[r, phantom, "\overset{\lambda}{\Longrightarrow}"] & \mathsf{C}' \arrow[dl, bend left, "L"] \\
\mathsf{D}\times\mathsf{D} \arrow[rr, "\otimes"'] & & \mathsf{D} &
\end{tikzcd}
\tag{A.2.7}
$$

and similarly for χ, in order to compute

$$
\begin{tikzcd}
\mathsf{C}\times\mathsf{C} \arrow[rr, "\otimes"] \arrow[dr, bend left, "G\times G"] \arrow[dd, bend right=60, "F\times F"'] & & \mathsf{C} \arrow[dr, bend left, "G"] & \\
\arrow[r, phantom, "\overset{\lambda\times\lambda}{\Longrightarrow}"] & \mathsf{C}'\times\mathsf{C}' \arrow[rr, "\otimes"] \arrow[dl, bend right, "L\times L"'] \arrow[dl, bend left, "X\times X"] \arrow[ur, phantom, "\Uparrow\mu_G\,\cong"] \arrow[dr, phantom, "\Nearrow\mu_X"] \arrow[dl, phantom, "\Searrow u\times u"] & & \mathsf{C}' \arrow[dl, bend left, "X"] \\
\mathsf{D}\times\mathsf{D} \arrow[rr, "\otimes"'] & & \mathsf{D} &
\end{tikzcd}
$$

$$
\equiv
\begin{tikzcd}
\mathsf{C}\times\mathsf{C} \arrow[rr, "\otimes"] \arrow[dr, bend left, "G\times G"] \arrow[dd, bend right=60, "F\times F"'] & & \mathsf{C} \arrow[dr, bend left, "G"] & \\
\arrow[r, phantom, "\overset{\chi\times\chi}{\Longrightarrow}"] & \mathsf{C}'\times\mathsf{C}' \arrow[rr, "\otimes"] \arrow[dl, bend left, "X\times X"] \arrow[ur, phantom, "\Uparrow\mu_G\,\cong"] \arrow[dr, phantom, "\Nearrow\mu_X"] & & \mathsf{C}' \arrow[dl, bend left, "X"] \\
\mathsf{D}\times\mathsf{D} \arrow[rr, "\otimes"'] & & \mathsf{D} &
\end{tikzcd}
$$

$$
\equiv
\begin{tikzcd}
\mathsf{C}\times\mathsf{C} \arrow[rr, "\otimes"] \arrow[dd, bend right=60, "F\times F"'] \arrow[ddrr, phantom, "\Nearrow\mu_F"] & & \mathsf{C} \arrow[dr, bend left, "G"] \arrow[dd, bend right, "F"'] & \\
& & \arrow[r, phantom, "\overset{\chi}{\Longrightarrow}"] & \mathsf{C}' \arrow[dl, bend left, "X"] \\
\mathsf{D}\times\mathsf{D} \arrow[rr, "\otimes"'] & & \mathsf{D} &
\end{tikzcd}
$$

$$\equiv \quad \begin{tikzcd}
\mathsf{C}\times\mathsf{C} \arrow[r, "\otimes"] \arrow[dd, bend right, "F\times F"'] \arrow[ddr, phantom, "\Nearrow \mu_F"] & \mathsf{C} \arrow[dr, bend left, "G"] \arrow[dd, bend right, "F"'] \arrow[dr, phantom, "\Rightarrow \lambda"] & \\
 & & \mathsf{C}' \arrow[dl, bend right, "L"'] \arrow[dl, bend left, "X"] \arrow[dl, phantom, "\Searrow u"] \\
\mathsf{D}\times\mathsf{D} \arrow[r, "\otimes"'] & \mathsf{D} &
\end{tikzcd}$$

$$\equiv \quad \begin{tikzcd}
\mathsf{C}\times\mathsf{C} \arrow[rr, "\otimes"] \arrow[dr, bend left, "G\times G"] \arrow[dd, bend right, "F\times F"'] \arrow[dr, phantom, "\Rightarrow \lambda\times\lambda"'] & & \mathsf{C} \arrow[dr, bend left, "G"] \arrow[dl, phantom, "\Nearrow \mu_G \cong"] & \\
 & \mathsf{C}'\times\mathsf{C}' \arrow[rr, "\otimes"] \arrow[dl, bend right, "L\times L"'] \arrow[dr, phantom, "\Nearrow \mu_L"] & & \mathsf{C}' \arrow[dl, bend right, "L"'] \arrow[dl, bend left, "X"] \arrow[dl, phantom, "\Searrow u"] \\
\mathsf{D}\times\mathsf{D} \arrow[rr, "\otimes"'] & & \mathsf{D} &
\end{tikzcd}$$

Now the natural transformation (A.2.2) is epic, so that $\lambda \times \lambda$, whiskered by $\mathsf{D}\times\mathsf{D} \to \mathsf{D}$, can be cancelled. μ_G is an isomorphism, so that it can be cancelled as well. Finally $G \times G$ is essentially surjective, and therefore pre-whiskering by it can also be cancelled. We are then left with (A.2.6).

Now suppose that η'_L and μ'_L give another monoidal structure on L. For λ to be monoidal, the equations (A.2.5) and (A.2.7) need to be satisfied. But now by (A.2.5) and the invertibility of η_G, we get $\eta'_L = \eta_L$. Similarly, by (A.2.7) together with the fact that like above, μ_G is an isomorphism, $\lambda \otimes \lambda$ is epic, and $G \times G$ is essentially surjective, we conclude that $\mu'_L = \mu_L$. $\square$

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. Like this, (A.2.7) becomes a full cylinder, with the two caps formed by $\lambda\times\lambda$ and λ, and with the three multiplications wrapping around. The equation (A.2.5), but with the $\lambda \times \lambda$ cap collapsed to a single point, so that one obtains a cone with λ on the base.

Bibliography

[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.

[AM10] M. Aguiar and S. Mahajan. *Monoidal Functors, Species and Hopf Algebras.* AMS, 2010.

[AT07] Charalambos D. Aliprantis and Rabee Tourky. *Cones and duality*, volume 84 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2007.

[Bas15] Giuliano Basso. A Hitchhiker's guide to Wasserstein distances. Available at: http://n.ethz.ch, 2015.

[Bla51] D. Blackwell. Comparison of experiments. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102, 1951.

[CF13] Valerio Capraro and Tobias Fritz. On the axiomatization of convex subsets of Banach spaces. *Proc. Amer. Math. Soc.*, 141, 2013. arXiv:1105.1270.

[Dob06] Ernst-Erich Doberkat. Eilenberg-Moore algebras for stochastic relations. *Inform. and Comput.*, 204(12):1756–1781, 2006.

[Edg98] Gerald A. Edgar. *Integral, Probability, and Fractal Measures.* Springer, 1998.

[Edw78] David A. Edwards. On the existence of probability measures with given marginals. *Ann. Inst. Fourier (Grenoble)*, 28(4):53–78, v, 1978.

[Fis80] Peter C. Fishburn. Stochastic dominance and moments of distributions. *Math. Oper. Res.*, 5(1):94–100, 1980.

[FKM16] Soichiro Fujii, Shin-ya Katsumata, and Paul-André Melliès. Towards a formal theory of graded monads. In *Foundations of software science and computation structures*, volume 9634 of *Lecture Notes in Comput. Sci.*, pages 513–530. Springer, 2016.

[FP17] Tobias Fritz and Paolo Perrone. A Probability Monad as the Colimit of Finite Powers, 2017. Submitted. arXiv:1712.05363.

[FP18] Tobias Fritz and Paolo Perrone. Bimonoidal Structure of Probability Monads. *Proceedings of MFPS 34, ENTCS*, 2018. arXiv:1804.03527.

[Fra01] U. Franz. What is Stochastic Independence? *Non-commutativity, infinite-dimensionality and probability at the crossroads. Proceedings of the RIMS workshop on infinite-dimensional analysis and quantum probability.*, 2001.

[Fri09] Tobias Fritz. Convex Spaces I: Definition and Examples, 2009. arXiv:0903.5522.

[GHK+03] G. Gierz, K. H. Hoffman, K. Keimel, J. D. Lawson, M. W. Mislove, and D. Scott. *Continuous Lattices and Domains*. Cambridge University Press, 2003.

[Gir82] Michèle Giry. A Categorical Approach to Probability Theory. In *Categorical aspects of topology and analysis*, volume 915 of *Lecture Notes in Mathematics*. 1982.

[GL17] Jean Goubault-Larrecq. Complete Quasi-Metrics for Hyperspaces, Continuous Valuations, and Previsions. arXiv:1707.03784, 2017.

[Gud73] Stan Gudder. Convex structures and operational quantum mechanics. *Comm. Math. Phys.*, 29:249–264, 1973.

[HLL18] Fumio Hiai, Jimmie Lawson, and Yongdo Lim. The stochastic order of probability measures on ordered metric spaces. *Journal of Mathematical Analysis and Applications*, 464(1):707–724, 2018. arXiv:1709.04187.

[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.

[JP89] C. Jones and J. D. Plotkin. A Probabilistic Powerdomain of Evaluations. *Proceedings of the Fourth Annual Symposium of Logics in Computer Science*, 1989.

[KA10] Arjun K. Gupta and Mohammad A. S. Aziz. Convex Ordering of Random Variables and its Applications in Econometrics and Actuarial Science . *European Journal of Pure and Applied Mathematics*, 3(5):779–785, 2010.

[Kei08] Klaus Keimel. The monad of probability measures over compact ordered spaces and its Eilenberg-Moore algebras. *Topology Appl.*, 156(2):227–239, 2008.

[Kel84] Hans G. Kellerer. Duality Theorems for Marginal Problems. *Zeitschrift für Warscheinlichkeitstheorie und verwandte Gebiete*, 67:399–432, 1984.

[Kou15] Seerp R. Koudenburg. Algebraic Kan extensions in double categories. *Theory and Applications of Categories*, 30(5), 2015. arXiv:1406.6994.

[Lac02] Stephen Lack. Codescent objects and coherence. *Journal of Pure and Applied Algebra*, 175(1-3), 2002.

[Law62] William Lawvere. The category of probabilistic mappings. Available at https://ncatlab.org/nlab/files/lawvereprobability1962.pdf, 1962.

[Law73] William Lawvere. Metric spaces, generalized logic and closed categories. *Rendiconti del seminario matematico e fisico di Milano*, 43, 1973.

[Law86] William Lawvere. Taking Categories Seriously. *Revista Colombiana de Matematicas*, 20, 1986.

[Law17] Jimmie Lawson. Ordered probability spaces. *J. Math. Anal. Appl.*, 455(1):167–179, 2017. arXiv:1612.03213.

[Leh55] E. L. Lehmann. Ordered Families of Distributions. *Annals of Mathematical Statistics*, 26(3):399–419, 1955.

[Lei16] Tom Leinster. Monoidal Categories with Projections, 2016. n-Category Cafë blog post, https://golem.ph.utexas.edu/category/2011/05/an_operadic_introduction_to_en.html.

[Lue69] David G. Luenberger. *Optimization by vector space methods*. John Wiley & Sons, 1969.

[Mac00] Saunders Mac Lane. *Categories for the Working Mathematician.* Springer, 2000.

[MPP16] Radu Mardare, Prakash Panangaden, and Gordon D. Plotkin. Quantitative Algebraic Reasoning. In *Proceedings of the Thirty first Annual IEEE Symposium on Logic in Computer Science (LICS 2016)*, pages 700–709. IEEE Computer Society Press, 2016.

[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.

[MW10] F. Marmolejo and R. J. Wood. Monads as extension systems. *TAC*, 24(4):84–113, 2010.

[Nac65] Leopold Nachbin. *Topology and Order.* Van Nostrand, 1965.

[Pat12] Ross Paterson. Constructing applicative functors. In Jeremy Gibbons and Pablo Nogueira, editors, *Mathematics of Program Construction*, pages 300–323. Springer, 2012.

[PP85] J. Pearl and A. Paz. Graphoids: a graph-based logic for reasoning about relevance relations. *UCLA Computer Science Dept., Technical Report 850038*, 1985.

[PP02] Gordon D. Plotkin and John Power. Notions of Computation Determine Monads. 2002.

[RS70] M. Rothschild and J. E. Stiglitz. Increasing risk: I. A definition. *Journal of Economic Theory*, 2:225–243, 1970.

[Sim18] Alex Simpson. Category-theoretic Structure for Independence and Conditional Independence. *Electronic Notes in Theoretical Computer Science*, 336:281–297, 2018.

[SS07] Moshe Shaked and George Shanthikumar. *Stochastic Orders.* Springer, 2007.

[Sto49] Marshall H. Stone. Postulates for the barycentric calculus. *Ann. Mat. Pura Appl. (4)*, 29:25–30, 1949.

[Str65] Volker Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36(2423–39), 1965.

[Świ74] Tadeusz Świrszcz. Monadic functors and convexity. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 22, 1974.

[vB05] Franck van Breugel. The Metric Monad for Probabilistic Nondeterminism. Available at http://www.cse.yorku.ca, 2005.

[vBHMW05] Franck van Breugel, Claudio Hermida, Michael Makkai, and James Worrell. An Accessible Approach to Behavioural Pseudometrics. In *Proceedings of the 32nd International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 3580 of *Lecture Notes in Computer Science*, pages 1018–1030. Springer, 2005.

[Vil09] Cédric Villani. *Optimal transport: old and new*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

[Web16] Mark Weber. Algebraic Kan extensions along morphisms of internal algebra classifiers. *Tbilisi Mathematical Journal*, 9(1), 2016. arXiv:1511.04911.

[Win85] Gerhard Winkler. *Choquet order and simplices with applications in probabilistic models*. Lecture Notes in Mathematics. Springer, 1985.