COALGEBRAIC BEHAVIORAL METRICS

PAOLO BALDAN, FILIPPO BONCHI, HENNING KERSTAN, AND BARBARA KÖNIG

Dipartimento di Matematica, Università di Padova, Italy
e-mail address: baldan@math.unipd.it

Dipartimento di Informatica, Università di Pisa, Italy
e-mail address: bonchi@di.unipi.it

Universität Duisburg-Essen, Germany
e-mail address: mail@henningkerstan.org

Universität Duisburg-Essen, Germany
e-mail address: barbara_koenig@uni-due.de

ABSTRACT. We study different behavioral metrics, such as those arising from both branching and linear-time semantics, in a coalgebraic setting. Given a coalgebra $\alpha\colon X \to HX$ for a functor $H\colon Set \to Set$, we define a framework for deriving pseudometrics on X which measure the behavioral distance of states.

A crucial step is the lifting of the functor H on Set to a functor $\overline{H}$ on the category $PMet$ of pseudometric spaces. We present two different approaches which can be viewed as generalizations of the Kantorovich and Wasserstein pseudometrics for probability measures. We show that the pseudometrics provided by the two approaches coincide on several natural examples, but in general they differ.

If H has a final coalgebra, every lifting $\overline{H}$ yields in a canonical way a behavioral distance which is usually branching-time, i.e., it generalizes bisimilarity. In order to model linear-time metrics (generalizing trace equivalences), we show sufficient conditions for lifting distributive laws and monads. These results enable us to employ the generalized powerset construction.

1. INTRODUCTION

When considering the behavior of state-based system models embodying quantitative information, such as probabilities, time or cost, the interest normally shifts from behavioral equivalences to behavioral distances. In fact, in a quantitative setting, it is often quite unnatural to ask that two systems exhibit exactly the same behavior, while it can be more reasonable to require that the distance between their behaviors is sufficiently small [GJS90, DGJP04, vBW05, BBLM15, dAFS04, dAFS09, FLT11a].

Coalgebras [Rut00] are a well-established abstract framework where a canonical notion of behavioral equivalence can be uniformly derived. In a nutshell, a coalgebra for a functor $H\colon \mathbb{C} \to \mathbb{C}$ is an arrow $\alpha\colon X \to HX$ in some category $\mathbb{C}$. When $\mathbb{C}$ is Set, the category of

sets and functions, coalgebras represent state machines: X is a set of states, α the transition function, and H describes the type of the transitions performed. For instance, deterministic automata are coalgebras $\langle o, t \rangle \colon X \to 2 \times X^A$: for any state $x \in X$, o specifies whether it is final ($o(x) = 1$) or not ($o(x) = 0$), and t the successor state for any given input in A. Similarly, nondeterministic automata can be seen as coalgebras $\langle o, t \rangle \colon X \to 2 \times \mathcal{P}(X)^A$ where now t assigns, for every state and input, a set of possible successors. Under certain conditions on H, there exists a final H-coalgebra, namely a coalgebra $\zeta \colon Z \to HZ$ such that for every coalgebra $\alpha \colon X \to HX$ there is a unique coalgebra homomorphism $[\![\cdot]\!] \colon X \to Z$. When the base category is Set, Z can be thought of as the set of all H-behaviors and $[\![\cdot]\!]$ as the function mapping every state in X into its behavior. Then two states $x, y \in X$ are said to be behaviorally equivalent when $[\![x]\!] = [\![y]\!]$. For deterministic automata this equivalence coincides with language equivalence while, for nondeterministic automata, it coincides with bisimilarity.

In this paper, we investigate how to exploit coalgebras to derive canonical notions of behavioral distances. The key step in our approach is to lift a functor H on Set to a functor $\overline{H}$ on $PMet$, the category of pseudometric spaces and nonexpansive maps. Given a pseudometric space (X, d_X), the goal is to define a suitable pseudometric on HX. For the (discrete) probability distribution functor $\mathcal{D}$, there are two liftings, known as Wasserstein and Kantorovich liftings, that have been extensively studied in transportation theory [Vil09]. The Kantorovich-Rubinstein duality states that these two liftings coincide.

It is our observation that these notions of liftings can analogously be defined for arbitrary functors H, leading to a rich general theory (Section 5). As concrete examples, besides $\mathcal{D}$, we study the (finite) powerset functor (resulting in the Hausdorff metric), the input functor, the coproduct and product bifunctors. The Kantorovich-Rubinstein duality holds for these, but it does not hold in general (we provide a counterexample).

Once a lifting $\overline{H}$ has been defined, it is easy to derive a behavioral distance on the states of an H-coalgebra $\alpha \colon X \to HX$. First, we turn α into a coalgebra for $\overline{H}$ by endowing X with the trivial discrete distance. Then, for any two states $x, y \in X$, their behavioral distance can be defined as the distance of their images in a final $\overline{H}$-coalgebra $\zeta \colon (Z, d_Z) \to \overline{H}(Z, d_Z)$, namely as $d_Z([\![x]\!], [\![y]\!])$. This notion of distance is well behaved with respect to the underlying behavioral equivalence: we show (Theorem 6.2) that a final H-coalgebra can be enriched with a pseudometric d_Z on Z yielding the final $\overline{H}$-coalgebra. This immediately implies that if two states are behaviorally equivalent, then they are at distance 0 (Lemma 6.6). Moreover, we show how to compute distances on the final coalgebra as well as on arbitrary coalgebras via fixed-point iteration and we prove that the pseudometric obtained on the final coalgebra is indeed a metric (Theorems 6.7 and 6.10). We recover behavioral metrics in the setting of probabilistic systems [DGJP04, vBW06] and of metric transition systems [dAFS09].

The canonical notion of equivalence for coalgebras, in a sense, fully captures the behavior of the system as expressed by the functor H. Sometimes one is interested in coarser equivalences, ignoring some aspects of a computation, a notable example being language equivalence for nondeterministic automata. In the coalgebraic setting this can be achieved by means of the generalized powerset construction [SBBR13, JSS12, JSS15]. The starting observation is that the distinction between the behavior to be observed and the computational effects that are intended to be hidden, is sometimes formally captured by splitting the functor H in two components, a functor F for the observable behavior and a monad T describing the computational effects, e.g., $1 + \_$, $\mathcal{P}$ or $\mathcal{D}$ provides partial, nondeterministic or probabilistic computations, respectively. For instance, the functor

for nondeterministic automata $2 \times \mathcal{P}(X)^A$ can be seen as the composition of the functor $FX = 2 \times X^A$ of deterministic automata, with the powerset monad $T = \mathcal{P}$, capturing nondeterminism. Trace semantics can be derived by viewing a coalgebra $X \to 2 \times \mathcal{P}(X)^A$ as a coalgebra $\mathcal{P}(X) \to 2 \times \mathcal{P}(X)^A$, via a determinization construction. Similarly probabilistic automata can be seen as coalgebras of the form $X \to [0,1] \times \mathcal{D}(X)^A$, yielding coalgebras $\mathcal{D}(X) \to [0,1] \times \mathcal{D}(X)^A$ via determinization. In general terms, the determinization of a coalgebra $X \to FTX$ leads to an F-coalgebra $TX \to FTX$. Formally this can be done whenever there is a distributive law between F and T. In this case, F lifts to a functor $\widehat{F}$ in $\mathcal{EM}(T)$, the Eilenberg-Moore category of T, and the determinized coalgebra can be regarded as an $\widehat{F}$-coalgebra.

In the second part of this paper, we exploit the aforementioned approach [JSS15] for systematically deriving metric trace semantics for coalgebras. The situation is summarized by the diagram in Figure 4 (on page 50). As a first step, building on our technique for lifting functors from Set to $PMet$, we identify conditions under which also natural transformations, monads and distributive laws can be lifted (Theorem 8.1). In this way we obtain an adjunction between $PMet$ and $\mathcal{EM}(\overline{T})$, where $\overline{T}$ is the lifted monad. Via the lifted distributive law we can transfer a functor $\overline{F} \colon PMet \to PMet$ to an endofunctor $\widehat{\overline{F}}$ on $\mathcal{EM}(\overline{T})$. By using the discrete distance, coalgebras of the form $TX \to FTX$ can now live in $\mathcal{EM}(\overline{T})$ and can be equipped with a trace distance via a map into a final coalgebra.

We show that this notion covers known or meaningful trace distances such as a metric on languages for nondeterministic automata or a variant of the total variation distance on distributions for probabilistic automata (Examples 9.4 and 9.6).

Synopsis. Motivated by an example of probabilistic transition systems, we will explain the Wasserstein and Kantorovich liftings for $\mathcal{D}$ in terms of transportation theory (Section 2). After some preliminaries on pseudometrics (Section 3) and a quick look at further motivating examples (Section 4) we will generalize the Kantorovich/Wasserstein liftings to arbitrary endofunctors on Set (Section 5). This leads to the definition of behavioral distance and the proof that it enjoys several desirable properties (Section 6). For trace pseudometrics, we need to study first compositionality of our liftings (Section 7) and then the lifting of natural transformations and monads (Section 8). Using these results, we can then employ the generalized powerset construction [JSS15] to obtain trace pseudometrics and show how this applies to nondeterministic and probabilistic automata (Section 9). Finally, we will discuss related and future work (Section 10).

Notation. We assume that the reader is familiar with the basic notions of coalgebras and category theory but will present some of the definitions below to introduce our notation. For a function $f \colon X \to Y$ and sets $A \subseteq X$, $B \subseteq Y$ we write $f[A] := \{f(a) \mid a \in A\}$ for the *image* of A and $f^{-1}[B] = \{x \in X \mid f(x) \in B\}$ for the *preimage* of B. If $Y \subseteq [0, \infty]$ and $f, g \colon X \to Y$ are functions we write $f \leq g$ if $f(x) \leq g(x)$ for all $x \in X$. Given a natural number $n \in \mathbb{N}$ and a family $(X_i)_{i=1}^n$ of sets X_i we denote the projections of the (cartesian) product of the X_i by $\pi_i^n \colon \prod_{i=1}^n X_i \to X_i$, or just by π_i if n is clear from the context. For a source $(f_i \colon X \to X_i)_{i=1}^n$ we denote the unique mediating arrow to the product by $\langle f_1, \ldots, f_n \rangle \colon X \to \prod_{i=1}^n X_i$. Similarly, given a family of arrows $(f_i \colon X_i \to Y_i)_{i=1}^n$, we write $f_1 \times \cdots \times f_n = \langle f_1 \circ \pi_1, \ldots, f_n \circ \pi_n \rangle \colon \prod_{i=1}^n X_i \to \prod_{i=1}^n Y_i$.

We quickly recap the basic ideas of coalgebras. Let F be an endofunctor on the category Set of sets and functions. An F-coalgebra is just a function $\alpha \colon X \to FX$. Given another

F-coalgebra $\beta\colon Y \to FY$ a coalgebra homomorphism from α to β is a function $f\colon A \to B$ such that $\beta \circ f = Ff \circ \alpha$. We call an F-coalgebra $\kappa\colon \Omega \to F\Omega$ *final* if for any other coalgebra $\alpha\colon X \to FX$ there is a unique coalgebra homomorphism $[\![\cdot]\!]_\alpha\colon X \to \Omega$. The final coalgebra need not exist but if it does it is unique up to isomorphism. It can be considered as the universe of all possible behaviors. If we have an endofunctor F such that a final coalgebra $\kappa\colon \Omega \to F\Omega$ exists then for any coalgebra $\alpha\colon X \to FX$ two states $x_1, x_2 \in X$ are said to be *behaviorally equivalent* if and only if $[\![x_1]\!]_\alpha = [\![x_2]\!]_\alpha$.

A *probability distribution* on a given set X is a function $P\colon X \to [0,1]$ satisfying $\sum_{x \in X} P(x) = 1$. For any set $B \subseteq X$ we define $P(B) = \sum_{x \in B} P(x)$. The *support* of P is the set $\mathrm{supp}(P) := \{x \in X \mid P(x) > 0\}$. We write $\mathcal{D}(X)$ for the set of all probability distributions on a set X and $\mathcal{D}_f(X)$ for the subset containing the distributions on X with finite support.

2. Motivation

The transition system in Figure 1 (taken from [vBW06]) is a purely probabilistic transition system (PTS) without labels with state space $X = \{u, x, y, z\}$ and an arbitrary $\epsilon \in \,]0, 1/2[$. An intuitive understanding of such a system is that in each state the system chooses a transition (indicated by the arrows) to another state using the probabilistic information which is given by the numbers on the arrows.

The state z on the right hand side of Figure 1 is a *final state* so the system terminates with probability one (indicated by the double circle) when reaching that state. Contrary to that, state u on the left hand side can be interpreted as a *fail state* which – once reached – can never be left and the system loops indefinitely in u. Thus the behavior of these states is entirely different.

Figure 1: A PTS

Comparing the behavior of x and y is a bit more complicated – they both have probabilistic transitions to u and z but in state x there is always a bias towards the final state z which is controlled only by the value of ϵ. Due to this x and y are certainly not behaviorally equivalent but similar in the sense that their behavioral distance is ϵ.
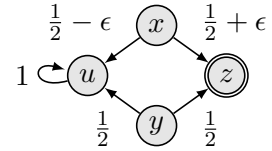
Let us now analyze how we can formally draw that conclusion. First note that we can define a distance between the states u and z based solely on the fact that z is final while u is not. Though we do not yet give it an explicit numerical value, we consider them to be *maximally apart*. Then, in order to compare x and y we need to compare their transitions which (in this example) are probability distributions[1] $P_x, P_y\colon \{u, z\} \to [0, 1]$. In particular $P_x(u) = \frac{1}{2} - \epsilon$, $P_x(z) = \frac{1}{2} + \epsilon$, $P_y(u) = \frac{1}{2}$, $P_y(z) = \frac{1}{2}$, $P_x(x) = P_x(y) = P_y(x) = P_y(y) = 0$. Thus the underlying task is to *define a distance between probability distributions based on a distance on their common domain of definition*.

Let us now tackle this question with a separate, more illustrative example. We will come back to our probabilistic transition system afterwards (in fact, we will even discuss it more thoroughly throughout the whole paper).

---

[1]They are actually probability distributions on X with support $\{u, z\}$. We will discuss this issue later.

2.1. **Wasserstein and Kantorovich Distance.** Suppose we are given the three element set $X = \{A, B, C\}$ along with a distance function (here: a metric) $d\colon X \times X \to \mathbb{R}_0^+$ where

$$d(A, A) = d(B, B) = d(C, C) = 0, \qquad d(A, B) = d(B, A) = 3,$$
$$d(A, C) = d(C, A) = 5, \qquad \text{and} \qquad d(B, C) = d(C, B) = 4.$$

Based on this function we now want to define a distance on probability distributions on X, i.e., a function $d^{\mathcal{D}}\colon \mathcal{D}X \times \mathcal{D}X \to [0, 1]$, so we need to define its value $d^{\mathcal{D}}(P, Q)$ for all probability distributions $P, Q\colon X \to [0, 1]$. As a concrete example, let us take the distributions

$$P(A) = 0.7, \qquad\qquad P(B) = 0.1, \qquad\qquad P(C) = 0.2,$$
$$Q(A) = 0.2, \qquad\qquad Q(B) = 0.3, \text{ and} \qquad\qquad Q(C) = 0.5.$$

In order to define their distance, we can interpret the three elements A, B, C as places where a certain product is produced and consumed (imagine the places are coffee roasting plants, each with an adjacent café where one can taste the coffee). For the sake of simplicity, we assume that the total amount of production equals the amount of consumption[2]. The places are geographically distributed across an area and the distance function just describes their physical distance. Moreover, the above distributions model the supply (P) and demand (Q) of the product in proportion of the total supply or demand. We have illustrated this situation graphically in Figure 2. The numbers on the edges indicate the distance between the places A, B, C whereas the numbers on the nodes indicate supply P (upper value) and demand Q (lower value).

The interpretation given above allows us to find two economically motivated views of defining a distance between P and Q based on the notion of *transportation* which is studied extensively in *transportation theory* [Vil09]. The leading idea is that the product needs to be transported so as to avoid excess supply and meet all demands. As an owner of the three facilities we have two choices: do the transport on our own or find a logistics firm which will do it for us.

Figure 2: Lifting example

If we are organizing the transport ourselves, transporting one unit of our product from one place to another costs one monetary unit per unit of distance traveled. As an example, transporting ten units from A to B will cost $10 \cdot 3 = 30$ monetary units. Formally, we will have to define a function $t\colon X \times X \to [0, 1]$ where $t(x, y)$ describes (in %) the amount of goods to be transported from place x to y such that

▷ supplies are used up: for all $x \in X$ we must have $\sum_{y \in X} t(x, y) = P(x)$, and
▷ demand is satisfied: for all $y \in X$ we must have $\sum_{x \in X} t(x, y) = Q(y)$.

In probability theory, such a function is as a joint probability distribution on $X \times X$ with marginals P and Q and is therefore called a *coupling* of P and Q. In our economic perspective we will just call it a *transportation plan* for supply P and demand Q and write $T(P, Q)$ for the set of all such transportation plans.

If $M \in \mathbb{R}_+$ denotes the total supply (= total demand), the total transportation cost for any transportation plan $t \in T(P, Q)$ is given by the function c_d which is parametrized by
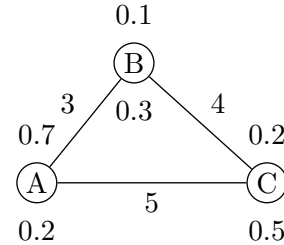
---

[2]If this was not the case we could introduce a dummy place representing either a warehouse (overproduction) or the amount of the product that needs to be bought on the market (underproduction).

the distance function d and defined as

$$c_d(t) := \sum_{x,y \in X} \big(M \cdot t(x,y)\big) \cdot d(x,y) = M \cdot \sum_{x,y \in X} t(x,y) \cdot d(x,y).$$

Since we want to maximize our profits, we will see to it that the total transportation cost is minimized, i.e., we will look for a transportation plan $t^* \in T(P,Q)$ minimizing the value of c_d (it can be shown that such a t^* always exists, by compactness). Using this, we can now define the distance of P and Q to be the relative costs $c_d(t^*)/M$ of such an optimal transportation plan, i.e.,

$$d^{\downarrow \mathcal{D}}(P,Q) := \min \left\{ \sum_{x,y \in X} t(x,y) \cdot d(x,y) \ \middle| \ t \in T(P,Q) \right\} = c_d(t^*)/M.$$

This distance between probability distributions is called the *Wasserstein distance* and it can be shown that this distance is a pseudometric if d is a pseudometric (and we will recover this result later). In our concrete example of Figure 2, the best solution is apparently to first use the local production (zero costs) at each facility and then transport the remaining excess supply of 50% in A to the remaining demands in B (20%) and in C (30%). Thus we obtain

$$d^{\downarrow \mathcal{D}}(P,Q) = 0.2 \cdot d(A,B) + 0.3 \cdot d(A,C) = 0.2 \cdot 3 + 0.3 \cdot 5 = 0.6 + 1.5 = 2.1$$

yielding an optimal (absolute) transportation cost of $2.1 \cdot M$ monetary units.

    If we decide to let a logistics firm do the transportation instead of doing it on our own, we assume that for each place it sets up a price for which it will buy a unit of our product (at a place with overproduction) or sell it (at a place with excess demand). Formally it will do so by giving us a price function $f \colon X \to \mathbb{R}_+$. We will only accept *competitive price functions* which satisfy $|f(x) - f(y)| \leq d(x,y)$ for all places $x,y \in X$. This amounts to saying that if we sell one unit of our product at place x to the logistics firm and buy one at y it does not cost more than transporting it ourselves from x to y. If d is a pseudometric, we will later call this requirement *nonexpansiveness* of the function f. Here, we will denote the set of all these functions by $C(d)$.

    The logistics firm is interested in its own profits which are given by the function g_d which is again parametrized by d and defined as

$$g_d(f) := \sum_{x \in X} f(x) \cdot \Big(\big(M \cdot Q(x)\big) - \big(M \cdot P(x)\big) \Big) = M \cdot \sum_{x \in X} f(x) \cdot \big(Q(x) - P(x)\big)$$

for all competitive price functions $f \in C(d)$. If in this formula the value $Q(x) - P(x)$ is greater than 0, there is underproduction so the logistics firm can sell goods whereas if $Q(x) - P(x) < 0$ it will have to buy them. Naturally, the logistics firm will want to maximize its profits so it will look for a competitive price function $f^* \in C(d)$ maximizing the value of g_d. Based on this we can now define another distance between P and Q to be the relative profit $g_d(f^*)/M$, i.e.,

$$d^{\uparrow \mathcal{D}}(P,Q) := \max \left\{ \sum_{x \in X} f(x) \cdot \big(Q(x) - P(x)\big) \ \middle| \ f \in C(d) \right\} = g_d(f^*)/M.$$

This distance is known as the *Kantorovich distance* and is also a pseudometric if d is. One can show that for our example it will be best if we give our product to the logistics firm

for free[3] in A, i.e., the logistic firm defines the price $f^*(A) = 0$. Moreover, we need to buy it back at B for three monetary units ($f^*(B) = 3$) and for five monetary units at C ($f^*(C) = 5$). This yields as distance

$$d^{\uparrow \mathcal{D}}(P, Q) = \sum_{x \in X} f^*(x) \cdot \big(Q(x) - P(x)\big)$$
$$= 0 \cdot (0.2 - 0.7) + 3 \cdot (0.3 - 0.1) + 5 \cdot (0.5 - 0.2) = 2.1$$

which is exactly the same as the one obtained earlier. In fact, one can prove that $d^{\uparrow \mathcal{D}}(P, Q) \le d^{\downarrow \mathcal{D}}(P, Q)$ so whenever we have a transportation plan t^* and a competitive price function f^* so that $c_d(t^*) = g_d(f^*)$ we know that both are optimal yielding $d^{\uparrow \mathcal{D}}(P, Q) = d^{\downarrow \mathcal{D}}(P, Q) = c_d(t^*)$. As final remark we note that if X is a finite set such a pair (t^*, f^*) will always exist and can be found e.g. using the simplex algorithm from linear programming.

Note that the Kantorovich-Rubinstein duality [Vil09] states that in the probabilistic setting the Wasserstein and the Kantorovich distance coincide. However, this is not necessarily the case when we lift other functors.

2.2. **Behavioral Distance as Fixed Point.** Now that we have finished our little excursion to transportation theory, let us come back to the original example where we wanted to define a distance between the two states x and y of the probabilistic transition system in Figure 1. Since we consider u and z to be maximally apart, we could formally set $d(u, z) = d(z, u) = 1$ and $d(u, u) = d(z, z) = 0$ so we obtain as distance function $d \colon \{u, z\} \times \{u, z\} \to [0, 1]$ the discrete 1-bounded metric on the set $\{u, z\}$. Using this, we could then define the distance of x and y to be the distance of their transition distributions $P_x, P_y \colon \{u, z\} \to [0, 1]$ yielding indeed a distance $d'(x, y) = d^{\downarrow \mathcal{D}}(P_x, P_y) = d^{\uparrow \mathcal{D}}(P_x, P_y) = \epsilon$ as claimed in the beginning.

However, the remaining question we need to answer is how the above procedure gives rise to a proper behavioral distance in the sense that we obtain a sound and complete definition of a distance function on *the whole set* X, i.e., a pseudometric $d \colon X \times X \to [0, 1]$. Notice that the distance between distributions depends on the distances in X, and vice-versa. Thus, we are led to a fixed-point definition, in particular in the presence of cycles in the transition system. We just need to observe that the definition of $d'(x, y)$ immediately yields the following recursive formula $d(x_1, x_2) = d^{\mathcal{D}}\big(P_{x_1}, P_{x_2}\big)$ for *all* $x_1, x_2 \in X$ where $d^{\mathcal{D}}$ is one of the equivalent distances (Wasserstein or Kantorovich) defined above. A known approach for probabilistic systems as the one above is to define its behavioral distance to be a fixed point $d^* \colon X \times X \to [0, 1]$ of the above equation [vBW06, vBSW08]. It is not difficult to see that due to the special structure of the above system, one obtains $d^*(u, z) = d^*(z, u) = 1$ and $d^*(x, y) = d'(x, y) = \epsilon$ which finally validates our initial claim that the distance of x and y is indeed ϵ.

3. PSEUDOMETRIC SPACES

Contrary to the usual definitions, where distances assume values in the half open real interval $[0, \infty[$, our distances assume values in a closed interval $[0, \top]$, where $\top \in]0, \infty]$ is a fixed maximal element (for our examples we will use $\top = 1$ or $\top = \infty$).

---

[3]Apparently, this is only reasonable in presence of a contract that prohibits the logistics firm to use our product or sell it to anyone else.

Definition 3.1 (Pseudometric, Pseudometric Space)**.** Let $\top \in \,]0, \infty]$ be a fixed maximal distance and X be a set. We call a function $d\colon X \times X \to [0, \top]$ a $\top$-*pseudometric* on X (or a *pseudometric* if $\top$ is clear from the context) if it satisfies $d(x, x) = 0$ (*reflexivity*), $d(x, y) = d(y, x)$ (*symmetry*), and $d(x, z) \leq d(x, y) + d(y, z)$ (*triangle inequality*) for all $x, y, z \in X$. If additionally $d(x, y) = 0$ implies $x = y$, d is called a $\top$-*metric* (or a *metric*). A *(pseudo)metric space* is a pair (X, d) where X is a set and d is a (pseudo)metric on X.

A trivial example of a pseudometric is the constant 0-function on any set whereas a simple example of a metric is the so-called *discrete metric* which can be defined on any set X as $d(x, x) = 0$ for all $x \in X$ and $d(x, y) = \top$ for all $x, y \in X$ with $x \neq y$.

The set of (pseudo)metrics over a fixed set X is a complete lattice (since $[0, \top]$ is) with respect to the pointwise order, i.e., for $d_1, d_2 \colon X \times X \to [0, \top]$ we define $d_1 \leq d_2$ if and only if $d_1(x, x') \leq d_2(x, x')$ holds for all $x, x' \in X$.

Lemma 3.2 (Lattice of Pseudometrics)**.** *Let X be a set. Then the set of pseudometrics over X, i.e., $D_X = \{d \mid d\colon X \times X \to [0, \top] \,\wedge\, d \text{ pseudometric}\}$, endowed with the pointwise order, is a complete lattice. The join of a set of pseudometrics $D \subseteq D_X$ is $(\sup D)(x, y) = \sup\{d(x, y) \mid d \in D\}$ for all $x, y \in X$. The meet of D is $\inf D = \sup\{d \mid d \in D_X \,\wedge\, \forall d' \in D. \, d \leq d'\}$.*

Proof. Let $D \subseteq D_X$. We first show that $d'\colon X \times X \to [0, \top]$ defined by $d'(x, y) = \sup\{d(x, y) \mid d \in D\}$ for all $x, y \in X$ is a pseudometric. Reflexivity and symmetry follows immediately from the fact that all $d \in D$ are pseudometrics. Concerning the triangle inequality, for all $x, y, z \in X$:

$$
\begin{aligned}
d'(x, y) + d'(y, z) &= \sup_{d \in D} d(x, y) + \sup_{d \in D} d(y, z) \\
&\geq \sup_{d \in D} \big(d(x, y) + d(y, z)\big) \geq \sup_{d \in D} d(x, z) = d'(x, z)\,.
\end{aligned}
$$

Since d' is a pseudometric, it immediately follows that $\sup D = d$, as desired.

The assertion about the meet is a completely general fact in complete lattices.   □

Observe that while the join of a set of pseudometrics is given by the pointwise supremum, in general, the meet can be smaller than the pointwise infimum. Hereafter, joins and meets of sets of pseudometrics will be implicitly taken in the corresponding lattice.

We will see later (Theorem 3.8) that the completeness of the lattice of pseudometrics ensures that the category whose objects are pseudometric spaces is complete and cocomplete.

3.1. **Calculating with (Extended) Real Numbers.** By $d_e\colon [0, \top]^2 \to [0, \top]$ we denote the ordinary Euclidean metric on $[0, \top]$, i.e., $d_e(x, y) = |x - y|$ for $x, y \in [0, \top] \setminus \{\infty\}$, and – where appropriate – $d_e(x, \infty) = \infty$ if $x \neq \infty$ and $d_e(\infty, \infty) = 0$. Addition is defined in the usual way, in particular $x + \infty = \infty$ for $x \in [0, \infty]$.

In the following lemma we rephrase the well-known fact that for $a, b, c \in [0, \infty)$ we have $|a - b| \leq c \iff a - b \leq c \,\wedge\, b - a \leq c$ to include the cases where a, b, c might be ∞. The proof is a simple case distinction and hence omitted.

Lemma 3.3. *For $a, b, c \in [0, \infty]$ we have $d_e(a, b) \leq c$ if and only if $(a \leq b + c)$ and $(b \leq a + c)$.*

We continue with another simple calculation involving this extended Euclidean distance which will turn out to be useful. Again, we omit the straightforward proof.

Lemma 3.4. *For a finite set A and functions $f, g \colon A \to [0, \infty]$ we have*

1. $d_e\big(\max_{a \in A} f(a), \max_{a \in A} g(a)\big) \leq \max_{a \in A} d_e\big(f(a), g(a)\big)$, *and*
2. $d_e\big(\sum_{a \in A} f(a), \sum_{a \in A} g(a)\big) \leq \sum_{a \in A} d_e\big(f(a), g(a)\big)$.

3.2. Pseudometrics Categorically.

Having established these two intermediary results, we recall that we want to work in a category whose objects are pseudometric spaces. In order to do so we need to define the arrows between them. While there are other, topologically motivated possibilities (taking the *continuous* or *uniformly continuous* functions with respect to the pseudometric topology), we just require that our functions do not increase distances.

Definition 3.5 (Nonexpansive Function, Isometry)**.** Let $\top \in \,]0, \infty]$ be an extended real number and (X, d_X) and (Y, d_Y) be $\top$-pseudometric spaces. We call a function $f \colon X \to Y$ *nonexpansive* if $d_Y \circ (f \times f) \leq d_X$. In this case we write $f \colon (X, d_X) \to (Y, d_Y)$. If equality holds f is called an *isometry*.

Observe that that isometries are not required to be surjective. Moreover, since we work with pseudometrics, where distinct elements can have distance 0, they are not necessarily injective. Note also that in this definition we have used a category theoretic mind-set and written the inequality in an "element-free" version as it will be easier to use in some of the subsequent results. Of course, this inequality is equivalent to requiring $d_Y\big(f(x), f(x')\big) \leq d_X(x, x')$ for all $x, x' \in X$. Simple examples of nonexpansive functions – even isometries – are of course the identity functions on a pseudometric space.

Apparently, if these functions shall be the arrows of a category, we will have to check that nonexpansiveness is preserved by function composition.

Lemma 3.6 (Composition of Nonexpansive Functions)**.** *Let $\top \in \,]0, \infty]$, and (X, d_X), (Y, d_Y) and (Z, d_Z) be $\top$-pseudometric spaces. If $f \colon (X, d_X) \to (Y, d_Y)$ and $g \colon (Y, d_Y) \to (Z, d_Z)$ are nonexpansive, then so is $g \circ f \colon (X, d_X) \to (Z, d_Z)$.*

Proof. Using nonexpansiveness of g and f we immediately conclude that $d_Z \circ \big((g \circ f) \times (g \circ f)\big) = d_Z \circ (g \times g) \circ (f \times f) \leq d_Y \circ (f \times f) \leq d_X$ which is the desired nonexpansiveness of $g \circ f$. $\square$

With this result at hand we now give our category a name. Please note that the definition below actually defines a whole family of categories, parametrized by the chosen maximal element. Despite this, we will just speak of *the* category of pseudometric spaces and keep in mind that there are (uncountably) many with the same properties.

Definition 3.7 (Category of Pseudometric Spaces)**.** Let $\top \in \,]0, \infty]$ be a fixed maximal element. The category $PMet$ has as objects all pseudometric spaces whose pseudometrics have codomain $[0, \top]$. The arrows are the nonexpansive functions between these spaces. The identities are the (isometric) identity functions and composition of arrows is function composition.

This category is complete and cocomplete which in particular implies that it has products and coproducts. We will later see that the respective product and coproduct pseudometrics also arise as special instances of our lifting framework (see Lemmas 5.45 and 5.48).

Theorem 3.8. *$PMet$ is bicomplete, i.e., it is complete and cocomplete.*

Proof. Let $U\colon PMet \to Set$ be the forgetful functor which maps every pseudometric space to its underlying set and each nonexpansive function to the underlying function. Moreover, let $D\colon I \to PMet$ be a small diagram, and define $(X_i, d_i) := D(i)$ for each object $i \in I$. Obviously $UD\colon I \to Set$ is also a small diagram. We show completeness and cocompleteness separately.

Completeness: Let $(f_i\colon X \to X_i)_{i\in I}$ be the limit cone to UD.

Observe that for any $i \in I$, $d_i \circ (f_i \times f_i)\colon X^2 \to [0, \top]$ is a pseudometric and define the pseudometric $d := \sup_{i\in I} d_i \circ (f_i \times f_i)\colon X^2 \to [0, \top]$ as provided by Lemma 3.2.

With this pseudometric all f_j are nonexpansive functions $(X, d) \to (X_j, d_j)$ because for all $j \in I$ and all $x, y \in X$ we have $d_j\big(f_j(x), f_j(y)\big) \leq \sup_{i\in I} d_i\big(f_i(x), f_i(y)\big) = d(x,y)$.

Moreover, if $\big(f_i'\colon (X', d') \to (X_i, d_i)\big)_{i\in I}$ is a cone to D, $(f_i'\colon X' \to X_i)_{i\in I}$ is a cone to UD and hence there is a unique function $g\colon X' \to X$ in Set satisfying $f_i \circ g = f_i'$ for all $i \in I$. We finish our proof by showing that this is a nonexpansive function $(X', d') \to (X, d)$. By nonexpansiveness of the f_i' we have for all $i \in I$ and all $x, y \in X'$ that $d_i(f_i'(x), f_i'(y)) \leq d'(x,y)$ and thus also

$$d\big(g(x), g(y)\big) = \sup_{i\in I} d_i\Big(f_i\big(g(x)\big), f_i\big(g(y)\big)\Big) = \sup_{i\in I} d_i\big(f_i'(x), f_i'(y)\big) \leq \sup_{i\in I} d'(x,y) = d'(x,y).$$

We conclude that $\big(f_i\colon (X, d) \to (X_i, d_i)\big)_{i\in I}$ is a limit cone to D.

Cocompleteness: Let $(f_i\colon X_i \to X)_{i\in I}$ be the colimit co-cone from UD and M_X be the set of all pseudometrics $d_X\colon X^2 \to [0, \top]$ on X such that the f_i are nonexpansive functions $(X_i, d_i) \to (X, d_X)$. We define the pseudometric $d := \sup_{d_X \in M_X} d_X$, as given by Lemma 3.2. With this pseudometric all f_j are nonexpansive functions $(X_j, d_j) \to (X, d)$ because for all $j \in I$ and all $x, y \in X_j$ we have $d\big(f_j(x), f_j(y)\big) = \sup_{d_X \in M_X} d_X\big(f_j(x), f_j(y)\big) \leq \sup_{d_X \in M_X} d_j(x,y) = d_j(x,y)$.

Moreover, if $\big(f_i'\colon (X_i, d_i) \to (X', d')\big)_{i\in I}$ is a co-cone from D, $(f_i'\colon X_i \to X')_{i\in I}$ is a co-cone from UD and hence there is a unique function $g\colon X \to X'$ in Set satisfying $g \circ f_i = f_i'$ for all $i \in I$. We finish our proof by showing that this is a nonexpansive function $(X, d) \to (X', d')$. Let $d_g := d' \circ (g \times g)\colon X^2 \to [0, \top]$, then it is easy to see that d_g is a pseudometric on X. Moreover, for all $i \in I$ and all $x, y \in X_i$ we have

$$d_g\big(f_i(x), f_i(y)\big) = d'\Big(g\big(f_i(x)\big), g\big(f_i(y)\big)\Big) = d'\big(f_i'(x), f_i'(y)\big) \leq d_i(x,y)$$

due to nonexpansiveness of $f_i'\colon (X_i, d_i) \to (X', d')$. Thus all f_i are nonexpansive if seen as functions $(X_i, d_i) \to (X, d_g)$ and we have $d_g \in M_X$. Using this we observe that for all $x, y \in X$ we have $d'\big(g(x), g(y)\big) = d_g(x,y) \leq \sup_{d_X \in M_X} d_X(x,y) = d(x,y)$ which shows that g is a nonexpansive function $(X', d') \to (X, d)$. We conclude that indeed $\big(f_i\colon (X_i, d_i) \to (X, d)\big)_{i\in I}$ is a colimit co-cone from D. □

For our purposes it will turn out to be useful to consider the following alternative characterization of the triangle inequality using the concept of nonexpansive functions.

Lemma 3.9. *A reflexive and symmetric function (see Definition 3.1) $d\colon X^2 \to [0, \top]$ satisfies the triangle inequality if and only if for all $x \in X$ the function $d(x, \_)\colon (X, d) \to ([0, \top], d_e)$ is nonexpansive.*

Proof. Let $d\colon X^2 \to [0, 1]$ be a reflexive and symmetric function. We show both implications separately for all $x, y, z \in X$.

▷ Using the triangle inequality and symmetry we know that $d(x, y) \leq d(x, z) + d(y, z)$ and $d(x, z) \leq d(x, y) + d(y, z)$. With Lemma 3.3 we conclude that $d_e(d(x, y), d(x, z)) \leq d(y, z)$.

▷ Using reflexivity of d, the triangle inequality for d_e and nonexpansiveness of $d(x, \_)$ we get $d(x, z) = d_e(d(x, x), d(x, z)) \leq d_e(d(x, x), d(x, y)) + d_e(d(x, y), d(x, z)) \leq d(x, y) + d(y, z)$. $\qquad\square$

4. Examples of Behavioral Distances

Equipped with this basic knowledge about our pseudometrics, let us now take a look at a few examples which we will use throughout the paper to demonstrate our theory. All the claims we make in these examples will be justified by our results. Our first example are the purely probabilistic systems like the system in Figure 1 in the beginning.

Example 4.1 (Probabilistic Systems and Behavioral Distance [vBW06]). The probability distribution functor $\mathcal{D}\colon Set \to Set$ maps each set X to the set of all probability distributions on it, i.e., $\mathcal{D}X = \left\{p\colon X \to [0, 1] \mid \sum_{x \in X} p(x) = 1\right\}$ and each function $f\colon X \to Y$ to the function $\mathcal{D}f\colon \mathcal{D}X \to \mathcal{D}Y$ with $\mathcal{D}f(p)(y) = \sum_{x \in f^{-1}[\{y\}]} p(x)$. We consider purely probabilistic transition systems without labels as coalgebras of the form $\alpha\colon X \to \mathcal{D}(X + \mathbb{1})$. Thus $\alpha(x)(y)$, for $x, y \in X$, is the probability of a transition from a state x to y and $\alpha(x)(\checkmark)$ gives the probability of terminating in x.

Van Breugel and Worrell [vBW06] introduced a metric for a continuous version of these systems by considering a discount factor $c \in \,]0, 1[$. Instantiating their framework to the discrete case we obtain the behavioral distance $d\colon X^2 \to [0, 1]$, defined as the least solution (with respect to the order $d_1 \leq d_2 \iff \forall x, y \in X. d_1(x, y) \leq d_2(x, y)$) of the equation $d(x, y) = \overline{d}\big(\alpha(x), \alpha(y)\big)$ for all $x, y \in X$. The lifted pseudometric $\overline{d}\colon (\mathcal{D}(X + \mathbb{1}))^2 \to [0, 1]$ is defined in two steps:

▷ First, $\widehat{d}\colon (X + \mathbb{1})^2 \to [0, 1]$ is defined as $\widehat{d}(x, y) = c \cdot d(x, y)$ if $x, y \in X$, $\widehat{d}(\checkmark, \checkmark) = 0$ and $\widehat{d}(x, y) = 1$ otherwise.

▷ Then, for all $P_1, P_2 \in \mathcal{D}(X + \mathbb{1})$, $\overline{d}(P_1, P_2)$ is defined as the supremum of all values $\sum_{x \in X + \mathbb{1}} f(x) \cdot (P_1(x) - P_2(x))$, with $f\colon (X + \mathbb{1}, \widehat{d}) \to ([0, 1], d_e)$ being an arbitrary nonexpansive function.

Our concrete example from Figure 1 is an instance of such a system and if we employ the aforementioned approach the behavioral distance of u and z is $d(u, z) = 1$ and hence $d(x, y) = c \cdot \epsilon$. We will see in Example 6.11 that this example can be captured by our framework. Moreover, we will also see that it is possible to set $c = 1$ resulting in $d(x, y) = \epsilon$.

It is easy to see that also the state space of a deterministic automaton can be equipped with a pseudometric which arises as a solution of a fixed point equation.

Example 4.2 (Bisimilarity Pseudometric for Deterministic Automata). We consider deterministic automata as coalgebras $\langle o, s \rangle\colon X \to 2 \times X^A$ in Set. For each state $x \in X$ the output value $o(x)$ determines whether x is final $(o(x) = 1)$ or not $(o(x) = 0)$ and the successor function $s(x)\colon A \to X$ determines, for each label $a \in A$, the unique a-successor $s(x)(a) \in X$. Given a pseudometric $d\colon X^2 \to [0, 1]$, we obtain a new pseudometric d^F on $2 \times X^A$ by defining, for every $(o_1, s_1), (o_2, s_2) \in 2 \times X^A$,

$$d^F\big((o_1, s_1), (o_2, s_2)\big) = \max\left\{d_2(o_1, o_2), c \cdot \max_{a \in A} d\big(s_1(a), s_2(a)\big)\right\},$$

where $c \in\ ]0,1]$ a discount factor and d_2 is the discrete 1-bounded metric on 2. Using our coalgebra $\langle o, s \rangle \colon X \to 2 \times X^A$ we get a fixed point equation on the complete lattice of pseudometrics on X by requiring, for all $x_1, x_2 \in X$,

$$d(x_1, x_2) = \max \left\{ d_2\big(o(x_1), o(x_2)\big), c \cdot \max_{a \in A} d\big(s(x_1)(a), s(x_2)(a)\big) \right\}.$$

If we take the least fixed point of this equation, the distance of two states x_1 and x_2 will be 1 if one state is final and the other is not. Otherwise their distance is the c-discounted maximal distance of their successors. Note that for $c = 1$ we obtain a discrete metric, where the distance of two states is 0 if they accept the same language and 1 otherwise, hence discount factors $c < 1$ yield more interesting metrics.

Let us finally consider so-called *metric transition systems* [dAFS09]. Each state of such a system comes equipped with a function which maps elements of a set of so-called *propositions* to a kind of non-discrete truth value in a pseudometric space.

Example 4.3 (Branching Distance for Metric Transition Systems [dAFS09]). Let $\Sigma = \{r_1, \ldots, r_n\}$ be a finite set of *propositions* where each proposition $r \in \Sigma$ is associated with a pseudometric space (M_r, d_r) which is bounded, i.e., we must have a finite $\top \in\ ]0, \infty[$ such that $d_r \colon M_r^2 \to [0, \top]$. A *valuation* of Σ is a function $u \colon \Sigma \to \cup_{r \in \Sigma} M_r$ that assigns to each $r \in \Sigma$ an element of M_r, i.e., we require $u(r) \in M_r$. We denote the set of all these valuations by $\mathcal{U}[\Sigma]$ and remark that it is apparently isomorphic to the set $M_1 \times \cdots \times M_n$ by means of the bijective function which maps a valuation u to the tuple $\big(u(r_1), \ldots, u(r_n)\big)$.

A *metric transition system* [dAFS09, Def. 6] is a tuple $(S, \tau, \Sigma, [\cdot])$ with a set S of states, a transition relation $\tau \subseteq S \times S$, a finite set Σ of propositions and a function $[\cdot] \colon S \to \mathcal{U}[\Sigma]$ assigning a valuation $[s]$ to each state $s \in S$. We define $\tau(s) := \{s' \in S \mid (s, s') \in \tau\}$ and require that $\tau(s)$ is finite.

The (directed) propositional distance between two valuations $u, v \in \mathcal{U}[\Sigma]$ is given by [dAFS09, Def. 10] $\overline{pd}(u, v) = \max_{r \in \Sigma} d_r\big(u(r), v(r)\big)$. The (undirected) branching distance $d \colon S \times S \to \mathbb{R}_0^+$ is defined as [dAFS09, Def. 13] the smallest fixed-point of the following equation, where $s, t \in S$:

$$d(s, t) = \max \left\{ \overline{pd}([s], [t]), \max_{s' \in \tau(s)} \min_{t' \in \tau(t)} d(s', t'), \max_{t' \in \tau(t)} \min_{s' \in \tau(s)} d(s', t') \right\} \qquad (4.1)$$

Note that, apart from the first argument, this coincides with the Hausdorff distance between the successors of s and t.
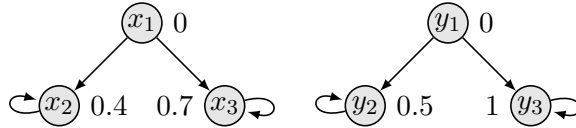

Figure 3: A metric transition system

We consider the concrete example system in Figure 3 ([dAFS09, Fig. 1]) with a single proposition $r \in \Sigma$, where $M_r = [0, 1]$ is equipped with the Euclidean distance d_e. Since the states x_2, x_3, y_2, y_3 only have themselves as successors, computing their distance according to (4.1) is easy: we just have to take the propositional distances of the valuations. This results in $d(x_2, y_2) = |0.4 - 0.5| = 0.1$, $d(x_2, y_3) = 0.6$, $d(x_3, y_2) = 0.2$, $d(x_3, y_3) = 0.3$.

Moreover, $\overline{pd}([x_1], [y_1]) = 0$ and thus $d(x_1, y_1)$ equals the Hausdorff distance of the reals associated with the sets of successors which is 0.3 (the maximal distance of any successor to the closest successor in the other set of successors, here: the distance from y_3 to x_3).

In order to model such transition systems as coalgebras we consider the product multifunctor $P: Set^n \to Set$ where $P(X_1, \ldots, X_n) = X_1 \times \cdots \times X_n$. Then coalgebras are of the form $c: S \to P(M_{r_1}, \ldots, M_{r_n}) \times \mathcal{P}_f(S)$, where $\mathcal{P}_f$ is the finite powerset functor and $c(s) = \big([s][r_1], \ldots, [s][r_n], \tau(s)\big)$. As we will see later in Example 6.12, the right-hand side of (4.1) can be seen as lifting of a pseudometric d on X to a pseudometric on $P(M_{r_1}, \ldots, M_{r_n}) \times \mathcal{P}_f(X)$.

We will later see that in all the examples above, we obtain a coalgebraic *bisimilarity pseudometric*: For any coalgebra $c: X \to FX$ let us denote the respective least fixed point by $bd_c: X^2 \to [0, \top]$. If a final F-coalgebra $z: Z \to FZ$ exists and some additional conditions hold (which is the case for all our examples) we have

$$bd_c(x, y) = 0 \quad \Longleftrightarrow \quad [\![x]\!]_c = [\![y]\!]_c$$

for all $x, y \in X$ where $[\![\cdot]\!]_c: X \to Z$ is the unique map into the final coalgebra.

5. Lifting Functors to Pseudometric Spaces

Generalizing from our examples, we now establish a general framework for deriving behavioral distances. The crucial step is to find, for an endofunctor F on Set, a way to transform a pseudometric on X to a pseudometric on FX. This induces a lifting of the functor F in the following sense.

Definition 5.1 (Lifting to Pseudometric Spaces)**.** Let $U: PMet \to Set$ be the forgetful functor which maps every pseudometric space to its underlying set. A functor $\overline{F}: PMet \to PMet$ is called a *lifting* of a functor $F: Set \to Set$ if the diagram on the right commutes. In this case, for any pseudometric space (X, d), we denote by d^F the pseudometric on FX which we obtain by applying $\overline{F}$ to (X, d).

$$\begin{array}{ccc} PMet & \xrightarrow{\overline{F}} & PMet \\ \downarrow{\scriptstyle U} & & \downarrow{\scriptstyle U} \\ Set & \xrightarrow{F} & Set \end{array}$$

Such a lifting is always monotone on pseudometrics in the following sense.

Theorem 5.2 (Monotonicity of Lifting)**.** *Let $\overline{F}: PMet \to PMet$ be a lifting of $F: Set \to Set$ and $d_1, d_2: X \times X \to [0, \top]$ be pseudometrics on X. Then $d_1 \le d_2$ implies $d_1^F \le d_2^F$.*

Proof. Observe first that for two pseudometrics d_1 and d_2 on the same set X, we have $d_1 \le d_2$ iff there is some $f: (X, d_2) \to (X, d_1)$ in $PMet$ such that $Uf = \mathrm{id}_X$ in Set.

Using the above fact, we can prove our statement. Let $f: (X, d_2) \to (X, d_1)$ in $PMet$ be such that $Uf = \mathrm{id}_X$. Then $\overline{F}f: \overline{F}(X, d_2) \to \overline{F}(X, d_1)$ is in $PMet$ as well by functoriality. By definition of a lifting we have $\overline{F}(X, d_i) = (FX, d_i^F)$, hence $U\overline{F}(X, d_i) = U(FX, d_i^F) = FX$. Moreover, $U\overline{F}f = FUf = F\mathrm{id}_X = \mathrm{id}_{FX}$. Therefore $\overline{F}f$ shows that $d_1^F \le d_2^F$. $\qquad\square$

In order to define a lifting to $PMet$ we will just use one simple tool, an evaluation function which describes how to transform an element of $F[0, \top]$ to a real number.

Definition 5.3 (Evaluation Function, Evaluation Functor)**.** Let F be an endofunctor on Set. An *evaluation function* for F is a function $ev_F: F[0, \top] \to [0, \top]$. Given such a function,

we define the *evaluation functor* to be the endofunctor $\widetilde{F}$ on $Set/[0, \top]$, the slice category[4] over $[0, \top]$, via $\widetilde{F}(g) = ev_F \circ Fg$ for all $g \in Set/[0, \top]$. On arrows $\widetilde{F}$ is defined as F.

We quickly remark that by definition of $\widetilde{F}$ on arrows, it is immediately clear that one indeed obtains a functor so the name is justified.

5.1. **The Kantorovich Lifting.** Let us now consider an endofunctor F on Set with an evaluation function ev_F. Given a pseudometric space (X, d), our first approach to lift d to FX will be to take the smallest possible pseudometric d^F on FX such that, for all nonexpansive functions $f: (X, d) \to ([0, \top], d_e)$, also $\widetilde{F}f: (FX, d^F) \to ([0, \top], d_e)$ is nonexpansive again, i.e., we want to ensure that for all $t_1, t_2 \in FX$ we have $d_e\big(\widetilde{F}f(t_1), \widetilde{F}f(t_2)\big) \leq d^F(t_1, t_2)$. This idea immediately leads us to the following definition which corresponds to the maximization of the logistic firm's prices in the introductory example.

Definition 5.4 (Kantorovich Distance). Let $F: Set \to Set$ be a functor with an evaluation function ev_F. For every pseudometric space (X, d) the *Kantorovich distance* on FX is the function $d^{\uparrow F}: FX \times FX \to [0, \top]$, where

$$d^{\uparrow F}(t_1, t_2) := \sup \left\{ d_e\left(\widetilde{F}f(t_1), \widetilde{F}f(t_2)\right) \mid f: (X, d) \to ([0, \top], d_e) \right\}$$

for all $t_1, t_2 \in FX$.

Note that a nonexpansive function $f: (X, d) \to ([0, \top], d_e)$ always exists. If $X = \emptyset$ it is the unique empty function and for $X \neq \emptyset$ every constant function is nonexpansive. Moreover, it is easy to show that $d^{\uparrow F}$ is a pseudometric.

Theorem 5.5 (Kantorovich Pseudometric). *For every pseudometric space (X, d) the Kantorovich distance $d^{\uparrow F}$ is a pseudometric on FX.*

Proof. Reflexivity and symmetry are an immediate consequence of the fact that d_e is a metric. We now show the triangle inequality. Let $t_1, t_2, t_3 \in FX$, then

$$d^{\uparrow F}(t_1, t_2) + d^{\uparrow F}(t_2, t_3)$$

$$= \sup_{f: (X,d) \to ([0,\top],d_e)} d_e\left(\widetilde{F}f(t_1), \widetilde{F}f(t_2)\right) + \sup_{f: (X,d) \to ([0,\top],d_e)} d_e\left(\widetilde{F}f(t_2), \widetilde{F}f(t_3)\right)$$

$$\geq \sup_{f: (X,d) \to ([0,\top],d_e)} \left(d_e\left(\widetilde{F}f(t_1), \widetilde{F}f(t_2)\right) + d_e\left(\widetilde{F}f(t_2), \widetilde{F}f(t_3)\right)\right)$$

$$\geq \sup_{f: (X,d) \to ([0,\top],d_e)} d_e\left(\widetilde{F}f(t_1), \widetilde{F}f(t_3)\right) = d^{\uparrow F}(t_1, t_3)$$

where the first inequality is a simple property of the supremum and the second inequality follows again from the fact that d_e is a metric. □

Using this pseudometric we can now immediately define our first lifting.

Definition 5.6 (Kantorovich Lifting). Let $F: Set \to Set$ be a functor with an evaluation function ev_F. We define the *Kantorovich lifting* of F to be the functor $\overline{F}: PMet \to PMet$, $\overline{F}(X, d) = (FX, d^{\uparrow F})$, $\overline{F}f = Ff$.

---

[4]The slice category $Set/[0, \top]$ has as objects all functions $g: X \to [0, \top]$ where X is an arbitrary set. Given g as before and $h: Y \to [0, \top]$, an arrow from g to h is a function $f: X \to Y$ satisfying $h \circ f = g$.

Since $\overline{F}$ inherits the preservation of identities and composition of morphisms from F we just need to prove that nonexpansive functions are mapped to nonexpansive functions to obtain functoriality of $\overline{F}$.

Theorem 5.7. *The Kantorovich lifting $\overline{F}$ is a functor on pseudometric spaces.*

Proof. $\overline{F}$ preserves identities and composition of arrows because F does. Moreover, it preserves nonexpansive functions: Let $f\colon (X,d_X) \to (Y,d_Y)$ be nonexpansive and $t_1, t_2 \in FX$, then

$$d_Y^{\uparrow F}\big(Ff(t_1), Ff(t_2)\big) = \sup_{g\colon (Y,d_Y)\to([0,\top],d_e)} d_e\Big(\widetilde{F}(g\circ f)(t_1), \widetilde{F}(g\circ f)(t_2)\Big)$$

$$\leq \sup_{h\colon (X,d_X)\to([0,\top],d_e)} d_e\Big(\widetilde{F}(h)(t_1), \widetilde{F}(h)(t_2)\Big) = d_X^{\uparrow F}(t_1, t_2)$$

due to the fact that since both f and g are nonexpansive also their composition $(g\circ f)\colon (X,d_X) \to ([0,\top],d_e)$ is nonexpansive. □

With this result at hand we are almost done: The only remaining task is to show that $\overline{F}$ is a lifting of F in the sense of Definition 5.1 but this is indeed obvious by definition of $\overline{F}$.

An important property of this lifting is that it preserves isometries, which is a bit tricky to show. While one might be tempted to think that this is immediately true because functors preserve isomorphisms, it is easy to see that the isomorphisms of $PMet$ are the *bijective* isometries. However, there are of course also isometries which are not bijective and thus not isomorphisms. Simple examples for this arise by taking the unique discrete metric space $(\mathbb{1}, d_{\mathbb{1}})$ and mapping it into any other pseudometric space (X,d) with $|X| > 1$. Any function $\mathbb{1} \to X$ is necessarily isometric but certainly not bijective.

Theorem 5.8. *The Kantorovich lifting $\overline{F}$ of a functor F preserves isometries.*

Proof. Let $f\colon (X,d_X) \to (Y,d_Y)$ be an isometry, i.e., f satisfies $d_Y \circ (f \times f) = d_X$. Since the Kantorovich lifting $\overline{F}$ is a functor on pseudometric spaces, we already know that $\overline{F}f$ is nonexpansive, i.e., we know that $d_Y^{\uparrow F} \circ (Ff \times Ff) \leq d_X^F$ thus we only have to show the opposite inequality. We will do that by constructing for every nonexpansive function $g\colon (X,d_X) \to ([0,\top],d_e)$ a nonexpansive function $h\colon (Y,d_Y) \to ([0,\top],d_e)$ such that $h\circ f = g$ which implies that for every $t_1, t_2 \in FX$ we have equality $d_e(\widetilde{F}h(Ff(t_1)), \widetilde{F}h(Ff(t_2))) = d_e(\widetilde{F}g(t_1), \widetilde{F}g(t_2))$. Then we have

$$d_Y^{\uparrow F} \circ (Ff \times Ff)(t_1, t_2) = \sup\Big\{d_e\Big(\widetilde{F}h(Ff(t_1)), \widetilde{F}h(Ff(t_2))\Big) \mid h\colon (Y,d_Y)\to([0,\top],d_e)\Big\}$$

$$\geq \sup\Big\{d_e\Big(\widetilde{F}g(t_1), \widetilde{F}g(t_2)\Big) \mid g\colon (X,d_X)\to([0,\top],d_e)\Big\} = d_X^{\uparrow F}(t_1, t_2)\,.$$

We define h via $h(y) := \inf_{x\in X}\{g(x) + d_Y(f(x), y)\}$. This function is nonexpansive, in fact, for all $y_1, y_2 \in Y$ we have

$$h(y_1) = \inf_{x\in X}\{g(x) + d_Y(f(x), y_1)\} \leq \inf_{x\in X}\{g(x) + d_Y(f(x), y_2) + d_Y(y_1, y_2)\}$$

$$= \inf_{x\in X}\{g(x) + d_Y(f(x), y_2)\} + d_Y(y_1, y_2) = h(y_2) + d_Y(y_1, y_2)$$

where the inequality follows by the fact that d_Y satisfies the triangle inequality. The same calculation also yields $h(y_2) \leq h(y_1) + d_Y(y_1, y_2)$. We use Lemma 3.3 with $a = h(y_1)$, $b = h(y_2)$ and $c = d_Y(y_1, y_2)$ to conclude that $d_e(h(y_1), h(y_2)) = d_e(a,b) \leq c = d_Y(y_1, y_2)$.

We can apply Lemma 3.3 with $a = g(x)$, $b = g(x')$ and $c = d(x, x')$ in the other direction using nonexpansiveness of g to obtain $g(x') + d_X(x', x) = b + c \geq a = g(x)$. Thus $h(f(x)) = \inf_{x' \in X}\{g(x') + d_Y(f(x'), f(x))\} = \inf_{x' \in X}\{g(x') + d_X(x', x)\} \geq g(x)$ and with $x' = x$ we obtain the equality $h(f(x)) = g(x)$ for all $x \in X$. $\square$

With this result in place, let us quickly discuss the name of our lifting. We chose the name *Kantorovich* because our definition is reminiscent of the Kantorovich pseudometric in probability theory. If we take the proper combination of functor and evaluation function, we can recover that pseudometric (in the discrete case) as the first instance of our framework.

Example 5.9 (Kantorovich Lifting of the Distribution Functor)**.** We take $\top = 1$ and the probability distribution functor $\mathcal{D}$ (or the sub-probability distribution functor). As evaluation $ev_{\mathcal{D}}: \mathcal{D}[0, 1] \to [0, 1]$ we define, for each $P \in \mathcal{D}[0, 1]$, $ev(P)$ to be the expected value of the identity function on $[0, 1]$, i.e., $ev_{\mathcal{D}}(P) := \sum_{x \in [0,1]} x \cdot P(x)$. Then for any function $g: X \to [0, 1]$ and any (sub)distribution $P \in \mathcal{D}X$ we can easily see that $\widetilde{\mathcal{D}}g(P) = ev_{\mathcal{D}} \circ \mathcal{D}g(P) = \sum_{x \in X} g(x) \cdot P(x)$. Using this, we can see that for every pseudometric space (X, d) we obtain the usual Kantorovich pseudometric $d^{\uparrow \mathcal{D}}: (\mathcal{D}X)^2 \to [0, 1]$, where

$$d^{\uparrow \mathcal{D}}(P_1, P_2) = \sup\left\{ \left| \sum_{x \in X} f(x) \cdot (P_1(x) - P_2(x)) \right| \;\middle|\; f: (X, d) \to ([0, 1], d_e) \right\}$$

for all (sub)probability distributions $P_1, P_2: X \to [0, 1]$.

Let us now consider the question whether the Kantorovich lifting preserves metrics, i.e., we want to check whether the Kantorovich pseudometric $d^{\uparrow F}$ is a metric for a metric space (X, d). The next example shows that this is not necessarily the case.

Example 5.10 (Kantorovich Lifting of the Squaring Functor)**.** The *squaring functor* on *Set* is the functor $S: Set \to Set$ where $SX = X \times X$ for each set X and $Sf = f \times f$ for each function $f: X \to Y$ (see e.g. [AHS90, Ex. 3.20 (10)]).

We take $\top = \infty$ and the evaluation function $ev_S: [0, \infty] \times [0, \infty] \to [0, \infty]$, $ev_S(r_1, r_2) = r_1 + r_2$. For a metric space (X, d) with $|X| \geq 2$ take $t_1 = (x_1, x_2) \in SX$ with $x_1 \neq x_2$ and define $t_2 := (x_2, x_1)$. Clearly $t_1 \neq t_2$ but for every nonexpansive function $f: (X, d) \to ([0, \top], d_e)$ we have $\widetilde{S}f(t_1) = f(x_1) + f(x_2) = f(x_2) + f(x_1) = \widetilde{S}f(t_2)$ and thus $d^{\uparrow S}(t_1, t_2) = 0$.

5.2. **The Wasserstein Lifting.** We have seen that our first lifting approach bears close resemblance to the original Kantorovich pseudometric on probability measures. We will now also define a generalized version of the Wasserstein pseudometric and compare it with our generalized Kantorovich pseudometric. To do that we first need to define generalized couplings, which can be understood as a generalization of joint probability measures.

Definition 5.11 (Coupling)**.** Let $F: Set \to Set$ be a functor and $n \in \mathbb{N}$. Given a set X and $t_i \in FX$ for $1 \leq i \leq n$ we call an element $t \in F(X^n)$ such that $F\pi_i(t) = t_i$ a *coupling* of the t_i (with respect to F). We write $\Gamma_F(t_1, \ldots, t_n)$ for the set of all these couplings.

Using these couplings we now proceed to define an alternative pseudometric on FX. As was the case for the Kantorovich distance, we only need to choose an evaluation function ev_F for our functor and then use the corresponding evaluation functor $\widetilde{F}$ (see Definition 5.3)

Definition 5.12 (Wasserstein Distance). Let $F\colon Set \to Set$ be a functor with evaluation function ev_F. For every pseudometric space (X, d) the *Wasserstein distance* on FX is the function $d^{\downarrow F}\colon FX \times FX \to [0, \top]$ given by

$$d^{\downarrow F}(t_1, t_2) := \inf \left\{ \widetilde{F}d(t) \,\middle|\, t \in \Gamma_F(t_1, t_2) \right\} . \tag{5.1}$$

for all $t_1, t_2 \in FX$.

Given $t_1, t_2 \in FX$, we say that a coupling $t \in \Gamma_F(t_1, t_2)$ is optimal whenever $d^{\downarrow F}(t_1, t_2) = \widetilde{F}d(t)$. Note that such an optimal coupling need not always exist.

In contrast to the Kantorovich distance, where the respective set cannot be empty because nonexpansive functions always exist, here it might be the case that no coupling exists and thus $d^{\downarrow F}(t_1, t_2) = \inf \emptyset = \top$. Without any additional conditions we cannot even prove that $\Gamma_F(t_1, t_1) \neq \emptyset$ which we would certainly need for reflexivity and thus we do not automatically obtain a pseudometric. The only property we get for free is symmetry.

Lemma 5.13 (Symmetry of the Wasserstein Distance). *For all pseudometric spaces (X, d) the Wasserstein distance $d^{\downarrow F}$ is symmetric.*

Proof. Let $t_1, t_2 \in FX$ and let $\sigma := \langle \pi_2, \pi_1 \rangle$ be the swap map on $X \times X$, i.e., $\sigma\colon X \times X \to X \times X$, $\sigma(x_1, x_2) = (x_2, x_1)$. If there is a coupling $t_{12} \in \Gamma_F(t_1, t_2)$ we define $t_{21} := F\sigma(t_{12}) \in F(X \times X)$ and observe that it satisfies $F\pi_1(t_{21}) = F\pi_1(F\sigma(t_{12})) = F(\pi_1 \circ \sigma)(t_{12}) = F\pi_2(t_{12}) = t_2$ and analogously $F\pi_2(t_{21}) = t_1$, thus $t_{21} \in \Gamma_F(t_2, t_1)$. Moreover, due to symmetry of d (i.e., $d \circ \sigma = d$), we obtain $\widetilde{F}d(t_{21}) = ev_F\big(Fd(t_{21})\big) = ev_F\big(Fd(F\sigma(t_{12}))\big) = ev_F\big(F(d \circ \sigma)(t_{12})\big) = ev_F\big(Fd(t_{12})\big) = \widetilde{F}d(t_{12})$ which yields the desired symmetry. If no coupling of t_1 and t_2 exists, there is also no coupling of t_2 and t_1 because otherwise we would get a coupling of t_1 and t_2 using the above method. Thus $d^{\downarrow F}(t_1, t_2) = \top = d^{\downarrow F}(t_2, t_1)$. $\square$

In order to be able to guarantee the other two properties of a pseudometric we will restrict our attention to well-behaved evaluation functions.

Definition 5.14 (Well-Behaved Evaluation Function). Let ev_F be an evaluation function for a functor $F\colon Set \to Set$. We call ev_F *well-behaved* if it satisfies the following conditions:

W1. $\widetilde{F}$ is monotone, i.e., for $f, g\colon X \to [0, \top]$ with $f \leq g$, we have $\widetilde{F}f \leq \widetilde{F}g$.

W2. For any $t \in F([0, \top]^2)$ we have $d_e\big(ev_F(t_1), ev_F(t_2)\big) \leq \widetilde{F}d_e(t)$ for $t_i := F\pi_i(t)$.

W3. $ev_F^{-1}[\{0\}] = Fi[F\{0\}]$ where $i\colon \{0\} \hookrightarrow [0, \top]$ is the inclusion map.

While condition W1 is quite natural, for conditions W2 and W3 some explanations are in order. Condition W2 ensures that $\widetilde{F}\mathrm{id}_{[0,\top]} = ev_F\colon F[0, \top] \to [0, \top]$ is nonexpansive once d_e is lifted to $F[0, \top]$ (recall that for the Kantorovich lifting we require $\widetilde{F}f$ to be nonexpansive for *any* nonexpansive f). By definition of the evaluation functor $\widetilde{F}$ and the t_i, we have $d_e\big(\widetilde{F}\pi_1(t), \widetilde{F}\pi_2(t)\big) = d_e\big(ev_F \circ F\pi_1(t), ev_F \circ F\pi_2(t)\big) = d_e\big(ev_F(t_1), ev_F(t_2)\big)$ so Condition W2 can equivalently be stated as $d_e\big(\widetilde{F}\pi_1(t), \widetilde{F}\pi_2(t)\big) \leq \widetilde{F}d_e(t)$.

Condition W3 requires that exactly the elements of $F\{0\}$ are mapped to 0 via ev_F. This ensures the reflexivity of the Wasserstein pseudometric.

Lemma 5.15 (Reflexivity of the Wasserstein Function). *Let F be an endofunctor on Set with evaluation function ev_F. If ev_F satisfies Condition W3 of Definition 5.14 then for any pseudometric space (X, d) the Wasserstein distance $d^{\downarrow F}$ is reflexive.*

Proof. Let $t_1 \in FX$. To show reflexivity we will construct a coupling $t \in \Gamma_F(t_1, t_1)$ such that $\widetilde{F}d(t) = 0$. In order to do that, let $\delta \colon X \to X^2, \delta(x) = (x, x)$ and define $t := F\delta(t_1)$. Then $F\pi_i(t) = F(\pi_i \circ \delta)(t_1) = F(\mathrm{id}_X)(t_1) = t_1$ and thus $t \in \Gamma_F(t_1, t_1)$. Since d is reflexive, $d \circ \delta \colon X \to [0, \top]$ is the constant zero function. Let $i \colon \{0\} \hookrightarrow [0, \top], i(0) = 0$ and for any set X let $!_X \colon X \to \{0\}, !_X(x) = 0$. Then also $i \circ !_X \colon X \to [0, \top]$ is the constant zero function and thus $d \circ \delta = i \circ !_X$. Using this we can conclude that

$$\widetilde{F}d(t) = \widetilde{F}d\big(F\delta(t_1)\big) = \widetilde{F}(d \circ \delta)(t_1) = \widetilde{F}(i \circ !_X)(t_1) = ev_F\Big(Fi\big((F!_X)(t_1)\big)\Big) = 0$$

where the last equality follows from the fact that $F!_X(t_1) \in F\{0\}$ and Condition W3 of Definition 5.14. $\qquad\square$

Before we continue our efforts to obtain a Wasserstein pseudometric, we check that well-behaved evaluation functions exist but not every evaluation function is well-behaved.

Example 5.16 (Evaluation Function for the Powerset Functor). We take $\top = \infty$ and consider the powerset functor $\mathcal{P}$. First, we show that the evaluation function $\sup \colon \mathcal{P}[0, \infty] \to [0, \infty]$ where $\sup \emptyset := 0$ is well-behaved.

W1. Let $f, g \colon X \to [0, \infty]$ with $f \leq g$. Let $S \in \mathcal{P}X$, i.e., $S \subseteq X$. Then we have $\widetilde{\mathcal{P}}f(S) = \sup f[S] = \sup \{f(x) \mid x \in S\} \leq \sup \{g(x) \mid x \in S\} = \sup g[S] = \widetilde{\mathcal{P}}g(S)$.

W2. For any subset $S \subseteq [0, \infty]^2$ we have to show the inequality

$$d_e\big(\sup \pi_1[S], \sup \pi_2[S]\big) \leq \sup d_e[S]. \tag{5.2}$$

For $S = \emptyset$ this is true because $\sup \emptyset = 0$ and thus both sides of the inequality are 0. Otherwise we define $s_i := \sup \pi_i[S]$. Clearly, if $s_1 = s_2$ then the left hand side of (5.2) is 0 and thus the inequality holds. Without loss of generality we now assume $s_1 < s_2$ and distinguish two cases.

If $s_2 < \infty$ then for any $\epsilon > 0$ we can find a pair $(t_1, t_2) \in S$ such that $s_2 - \epsilon < t_2$ because s_2 is the supremum. Moreover, $t_1 \leq s_1$ and if $\epsilon < s_2 - s_1$ also $t_1 \leq t_2$. By combining these inequalities, we conclude that for every $\epsilon \in]0, s_2 - s_1[$ we have a pair $(t_1, t_2) \in S$ such that $d_e(s_1, s_2) - \epsilon = s_2 - s_1 - \epsilon \leq s_2 - t_1 - \epsilon = s_2 - \epsilon - t_1 < t_2 - t_1 = d_e(t_1, t_2)$. Since ϵ can be arbitrarily small, we thus must have (5.2).

If $s_2 = \infty$ then $d_e(s_1, s_2) = \infty$. However, $s_2 = \infty$ also implies that for every non-negative real number $r \in \mathbb{R}_+$ we can find an element $(t_1, t_2) \in S$ such that $t_2 > r$. Especially, for $r > s_1$ we have $t_2 > r > s_1 \geq t_1$ and thus $d_e(t_1, t_2) = t_2 - t_1 \geq t_2 - s_1 > r - s_1$. Since r can be arbitrarily large, we thus must have $\sup d_e[S] = \infty$ and thus (5.2) is an equality.

W3. We have $\mathcal{P}i[\mathcal{P}\{0\}] = \mathcal{P}i[\{\emptyset, \{0\}\}] = \{i[\emptyset], i[\{0\}]\} = \{\emptyset, \{0\}\} = \sup^{-1}[\{0\}]$.

Whenever we work with the finite powerset functor $\mathcal{P}_f$ we can of course use max instead of sup with the convention $\max \emptyset = 0$.

In contrast to the above, $\inf \colon \mathcal{P}([0, \infty]) \to [0, \infty]$ is not well-behaved. It neither satisfies Condition W2, nor Condition W3: $\inf d_e[S] \geq d_e\big(\inf \pi_1[S], \inf \pi_2[S]\big)$ fails for $S = \{(0, 1), (1, 1)\}$ and $\{0, 1\} \in \inf^{-1}[\{0\}]$.

Staying with Condition W3 for a while we remark that it can be expressed as a weak pullback diagram, thus fitting nicely into a coalgebraic framework.

Lemma 5.17 (Weak Pullback Characterization of Condition W3). *Let F be an endofunctor on Set with evaluation function ev_F and $i\colon \{0\} \hookrightarrow [0, \top]$ be the inclusion function. For any set X we denote the unique arrow into $\{0\}$ by $!_X\colon X \to \{0\}$. Then ev_F satisfies Condition W3 of Definition 5.14, i.e., $ev_F^{-1}[\{0\}] = F[F\{0\}]$ if and only if the diagram on the right is a weak pullback.*

$$\begin{array}{ccc} F\{0\} & \xrightarrow{\ !_{F\{0\}}\ } & \{0\} \\ {\scriptstyle Fi}\downarrow & & \downarrow{\scriptstyle i} \\ F[0, \top] & \xrightarrow[ev_F]{} & [0, \top] \end{array}$$

Proof. We consider the extended diagram on the right hand side below. Commutativity of the diagram is equivalent to $ev_F^{-1}[\{0\}] \supseteq Fi[F\{0\}]$.

Given a set X and a function $f\colon X \to F[0, \top]$ as depicted below, we conclude again by commutativity ($i \circ !_X = ev_F \circ f$) that $f(x) \in ev_F^{-1}[\{0\}]$ for all $x \in X$.

Now we show that the weak universality is equivalent to the other inclusion. First suppose that $ev_F^{-1}[\{0\}] \subseteq Fi[F\{0\}]$ then for $f(x) \in ev_F^{-1}[\{0\}]$ we can choose a (not necessarily unique) $x_0 \in F\{0\}$ such that $f(x) = Fi(x_0)$. If we define $\varphi\colon X \to F\{0\}$ by $\varphi(x) = x_0$ then clearly φ makes the above diagram commute and thus we have a weak pullback.

Conversely if the diagram is a weak pullback we consider the set $X = ev_F^{-1}[\{0\}]$ and the function $f\colon ev_F^{-1}[\{0\}] \hookrightarrow F[0, \top], f(x) = x$. Now for any $x \in ev_F^{-1}[\{0\}]$ we have $Fi(\varphi(x)) = (Fi \circ \varphi)(x) = f(x) = x$, hence – since $\varphi(x) \in F\{0\}$ – we have $x \in Fi[F\{0\}]$. This shows that indeed $ev_F^{-1}[\{0\}] \subseteq Fi[F\{0\}]$ holds. $\qquad\square$

The only missing step towards the Wasserstein pseudometric is the observation that if F preserves weak pullbacks we can define new couplings based on given ones.

Lemma 5.18 (Gluing Lemma). *Let F be an endofunctor on Set, X a set, $t_1, t_2, t_3 \in FX$, $t_{12} \in \Gamma_F(t_1, t_2)$, and $t_{23} \in \Gamma_F(t_2, t_3)$ be couplings. If F preserves weak pullbacks then there is a coupling $t_{123} \in \Gamma_F(t_1, t_2, t_3)$ such that*

$$F\langle \pi_1, \pi_2 \rangle(t_{123}) = t_{12} \quad and \quad F\langle \pi_2, \pi_3 \rangle(t_{123}) = t_{23}$$

where $\pi_i\colon X^3 \to X$ are the projections of the ternary product. Moreover, $t_{13} := F\langle \pi_1, \pi_3 \rangle(t_{123})$ is a coupling of t_1 and t_3, i.e., we have $t_{13} \in \Gamma_F(t_1, t_3)$.

Proof. Let $\tau_i\colon X \times X$ be the projections of the binary product. We first observe that the following diagram is a pullback square.

$$\begin{array}{ccccc} & & X \times X \times X & & \\ & {\scriptstyle \langle \pi_1, \pi_2 \rangle}\swarrow & & \searrow{\scriptstyle \langle \pi_2, \pi_3 \rangle} & \\ X \times X & & & & X \times X \\ & {\scriptstyle \tau_2}\searrow & & \swarrow{\scriptstyle \tau_1} & \\ & & X & & \end{array}$$

Given any set P along with functions $p_1, p_2\colon P \to X \times X$ satisfying the condition $\tau_2 \circ p_1 = \tau_1 \circ p_2$ the unique mediating arrow $u\colon P \to X \times X \times X$ is given by $u = \langle \tau_1 \circ p_1, \tau_2 \circ p_1, \tau_2 \circ p_2 \rangle = \langle \tau_1 \circ p_1, \tau_1 \circ p_2, \tau_2 \circ p_2 \rangle$. Now let us look at the following diagram.

$$F\langle\pi_1,\pi_2\rangle \quad F(X\times X\times X) \quad F\langle\pi_2,\pi_3\rangle$$

$$F\tau_1 \qquad F(X\times X) \qquad\qquad\qquad F(X\times X) \qquad F\tau_2$$

$$FX \qquad\qquad F\tau_2 \qquad FX \qquad F\tau_1 \qquad\qquad\qquad FX$$

Since F preserves weak pullbacks, the inner part of this diagram is a weak pullback. We recall that $t_{12} \in \Gamma_F(t_1,t_2)$ and $t_{23} \in \Gamma_F(t_2,t_3)$ so we have $F\tau_2(t_{12}) = t_2 = F\tau_1(t_{23})$. Using this, we can use the (weak) universality of the pullback to obtain[5] an element $t_{123} \in F(X\times X\times X)$ which satisfies the two equations of the lemma and moreover $F\pi_1(t_{123}) = F(\tau_1\circ\langle\pi_1,\pi_2\rangle)(t_{123}) = F\tau_1\circ F\langle\pi_1,\pi_2\rangle(t_{123}) = F\tau_1(t_{12}) = t_1$ and analogously $F\pi_2(t_{123}) = t_2$, $F\pi_3(t_{123}) = t_3$ yielding $t_{123} \in \Gamma_F(t_1,t_2,t_3)$.

For $t_{13} := F\langle\pi_1,\pi_3\rangle(t_{123})$ we observe that $F\tau_1(t_{13}) = F\tau_1\big(F\langle\pi_1,\pi_3\rangle(t_{123})\big) = F\big(\tau_1\circ\langle\pi_1,\pi_3\rangle\big)(t_{123}) = F\pi_1(t_{123}) = t_1$ and similarly $F\tau_2(t_{13}) = t_3$ so $t_{13} \in \Gamma_F(t_1,t_3)$ as claimed.
□

With the help of this lemma we can now finally give sufficient conditions to guarantee that the Wasserstein distance satisfies the triangle inequality. Apparently, since we use the above lemma, this will work only for weak pullback preserving functors and we will also need Conditions W1 and W2 of Definition 5.14 for the proof.

Lemma 5.19 (Triangle Inequality for the Wasserstein Function). *Let F be an endofunctor on Set with evaluation function ev_F. If F preserves weak pullbacks and ev_F satisfies Conditions W1 and W2 of Definition 5.14 then for any pseudometric space (X,d) the Wasserstein distance $d^{\downarrow F}$ satisfies the triangle inequality.*

Proof. We will use the characterization of the triangle inequality given by Lemma 3.9 (page 10). Hence, given any pseudometric space (X,d) we just have shown that for every $t_1 \in FX$ the function $d^{\downarrow F}(t_1,\_)\colon (FX,d^{\downarrow F}) \to ([0,\top],d_e)$ is nonexpansive, i.e., that the inequality

$$d_e\big(d^{\downarrow F}(t_1,t_2),d^{\downarrow F}(t_1,t_3)\big) \leq d^{\downarrow F}(t_2,t_3) \tag{5.3}$$

holds for all $t_2,t_3 \in FX$. We will show this in several steps.

First of all we consider the case where no coupling of t_2 and t_3 exists. In this case the right hand side of (5.3) is $\top$ and it is easy to see that the the left hand side can never exceed that value because $d^{\downarrow F}$ is non-negative. Thus in the remainder of the proof we only consider the case where $\Gamma_F(t_2,t_3) \neq \emptyset$.

As next step we observe that if $\Gamma_F(t_1,t_2) = \Gamma_F(t_1,t_3) = \emptyset$ the left hand side of (5.3) is 0 and the right hand side is non-negative. Thus we are left with the cases where $\Gamma_F(t_1,t_2) \neq \emptyset$ or $\Gamma_F(t_1,t_3) \neq \emptyset$.

Let us first assume that $\Gamma_F(t_1,t_2) \neq \emptyset$ and recall that we required $\Gamma_F(t_2,t_3) \neq \emptyset$. With the Gluing Lemma (Lemma 5.18) we can then conclude that also $\Gamma_F(t_1,t_3) \neq \emptyset$. Similarly, if $\Gamma_F(t_1,t_3) \neq \emptyset$ we can use the swap map as in the proof of Lemma 5.13 to see that $\Gamma_F(t_3,t_1) \neq \emptyset$. As above the Gluing Lemma yields $\Gamma_F(t_2,t_1) \neq \emptyset$ and again using the swap

_____

[5]Explicitly: Consider $\{t_2\}$ with functions $p_1,p_2\colon \{t_2\} \to F(X\times X)$ where $p_1(t_2) = t_{12}$ and $p_2(t_2) = t_{23}$, then by the weak pullback property there is a (not necessarily unique) function $u\colon \{t_2\} \to F(X\times X\times X)$ satisfying $F(\langle\pi_1,\pi_2\rangle)\circ u = p_1$ and $F(\langle\pi_2,\pi_3\rangle)\circ u = p_2$. We simply define $t_{123} := u(t_2)$.

map we conclude that $\Gamma_F(t_1, t_2) \neq \emptyset$. Thus the sole remaining case is the case where all couplings exists, i.e., we have $\Gamma_F(t_1, t_2) \neq \emptyset$, $\Gamma_F(t_2, t_3) \neq \emptyset$ and $\Gamma_F(t_1, t_3) \neq \emptyset$.

As intermediate step we recall that for all $x \in X$ the function $d(x, \_)$ is nonexpansive (see Lemma 3.9, page 10). Using the projections $\pi_i \colon X^3 \to X$ of the product this can be formulated as the inequality $d_e \circ (d \times d) \circ \langle \langle \pi_1, \pi_2 \rangle, \langle \pi_1, \pi_3 \rangle \rangle \leq d \circ \langle \pi_2, \pi_3 \rangle$ and the monotonicity of $\widetilde{F}$ (Condition W1) implies that also the inequality

$$\widetilde{F}\big(d_e \circ (d \times d) \circ \langle \langle \pi_1, \pi_2 \rangle, \langle \pi_1, \pi_3 \rangle \rangle\big) \leq \widetilde{F}(d \circ \langle \pi_2, \pi_3 \rangle) \tag{5.4}$$

holds. We will now use this inequality to prove (5.3) for the remaining case in which all couplings exists.

As already pointed out before, for any $t_{12} \in \Gamma_F(t_1, t_2)$ and $t_{23} \in \Gamma_F(t_2, t_3)$ the Gluing Lemma (Lemma 5.18) yields a $t_{123} \in \Gamma_F(t_1, t_2, t_3)$ and a coupling $t_{13} := F(\langle \pi_1, \pi_3 \rangle)(t_{123}) \in \Gamma_F(t_1, t_3)$. Plugging in t_{123} in the inequality (5.4) above yields $\widetilde{F}d_e\left(F(d \times d)(t_{12}, t_{13})\right) \leq \widetilde{F}d(t_{23})$. Using well-behavedness (Condition W2) of ev_F on the left hand side we obtain the following, intermediary result:

$$d_e\left(\widetilde{F}d(t_{12}), \widetilde{F}d(t_{13})\right) \leq \widetilde{F}d(t_{23}). \tag{5.5}$$

If we define $d_{ij} := d^{\downarrow F}(t_i, t_j)$ we can express (5.3) as $d_e(d_{12}, d_{13}) \leq d_{23}$. This is obviously true for $d_{12} = d_{13}$ so without loss of generality we assume $d_{12} < d_{13}$ and claim that for all $\epsilon > 0$ there is a coupling $t_{12} \in \Gamma_F(t_1, t_2)$ such that for all couplings $t_{13} \in \Gamma_F(t_1, t_3)$ we have

$$d_e(d_{12}, d_{13}) \leq \epsilon + d_e\left(\widetilde{F}d(t_{12}), \widetilde{F}d(t_{13})\right). \tag{5.6}$$

To prove this claim, we recall that the Wasserstein distance is defined as an infimum, so we have $d_{13} \leq \widetilde{F}d(t_{13})$ for all couplings t_{13}. Moreover, for the same reason we can pick, for every $\epsilon > 0$, a coupling $t_{12} \in \Gamma_F(t_1, t_2)$, such that $\widetilde{F}d(t_{12}) - d_{12} \leq \epsilon$ which can equivalently be stated as $\widetilde{F}d(t_{12}) \leq d_{12} + \epsilon$. With this fixed coupling we now proceed to establish (5.4) for all $t_{13} \in \Gamma_F(t_1, t_3)$.

If $d_{13} = \infty$ we have $d_e(d_{12}, d_{13}) = \infty$ but also $\widetilde{F}d(t_{13}) = \infty$ and $\widetilde{F}d(t_{12}) \leq d_{12} + \epsilon < \infty + \epsilon = \infty$ and therefore $d_e\big(\widetilde{F}d(t_{12}), \widetilde{F}d(t_{13})\big) = \infty$ and thus (5.6) is valid. For $d_{13} < \infty$ we have

$$d_e(d_{12}, d_{13}) = d_{13} - d_{12} \leq \widetilde{F}d(t_{13}) - \left(\widetilde{F}d(t_{12}) - \epsilon\right) = \epsilon + \left(\widetilde{F}d(t_{13}) - \widetilde{F}d(t_{12})\right)$$

$$\leq \epsilon + \big|\widetilde{F}d(t_{13}) - \widetilde{F}d(t_{12})\big| \leq \epsilon + d_e\left(\widetilde{F}d(t_{12}), \widetilde{F}d(t_{13})\right)$$

where the last inequality is due to the fact that $\widetilde{F}d(t_{12}) < \infty$. Hence we have established our claimed validity of (5.6). Using this, (5.5) and the fact that – as above – given $\epsilon > 0$ we have a coupling t_{23} such that $\widetilde{F}d(t_{23}) \leq d_{23} + \epsilon$ we obtain the inequality

$$d_e(d_{12}, d_{13}) \leq \epsilon + d_e\left(\widetilde{F}d(t_{12}), \widetilde{F}d(t_{13})\right) \leq \epsilon + \widetilde{F}d(t_{23}) \leq 2\epsilon + d_{23}$$

which also proves $d_e(d_{12}, d_{13}) \leq d_{23}$. Indeed if $d_e(d_{12}, d_{13}) > d_{23}$ then we would have $d_e(d_{12}, d_{13}) = d_{23} + \epsilon'$ and we just take $\epsilon < \epsilon'/2$ which yields the contradiction $d_e(d_{12}, d_{13}) \leq 2\epsilon + d_{23} < \epsilon' + d_{23} = d_e(d_{12}, d_{13})$. $\qquad\square$

Combining this result with our previous considerations we finally obtain the desired result which guarantees that the Wasserstein distance is indeed a pseudometric.

Theorem 5.20 (Wasserstein Pseudometric)**.** *Let F be an endofunctor on Set with evaluation function ev_F. If F preserves weak pullbacks and ev_F is well-behaved then for any pseudometric space (X, d) the Wasserstein distance $d^{\downarrow F}$ is a pseudometric.*

Proof. Reflexivity is given by Lemma 5.15, symmetry by Lemma 5.13 and the triangle inequality by Lemma 5.19. □

With this result in place we can now finally study the Wasserstein lifting of a functor. Of course, our requirements on F and ev_F are just sufficient conditions to prove that the Wasserstein distance is a pseudometric so it might be possible to give a more general definition. However, we will always work with weak pullback preserving functors and well-behaved evaluation functions so the following definition suffices.

Definition 5.21 (Wasserstein Lifting)**.** Let F be a weak pullback preserving endofunctor on *Set* with well-behaved evaluation function ev_F. We define the *Wasserstein lifting* of F to be the functor $\overline{F} \colon PMet \to PMet$, $\overline{F}(X, d) = (FX, d^{\downarrow F})$, $\overline{F}f = Ff$.

Of course, we will have to check the functoriality. Its proof relies on Condition W1 of Definition 5.14, the monotonicity of $\widetilde{F}$.

Theorem 5.22. *The Wasserstein lifting $\overline{F}$ is a functor on pseudometric spaces.*

Proof. $\overline{F}$ preserves identities and composition of arrows because F does. Moreover, it preserves nonexpansive functions: Let $f \colon (X, d_X) \to (Y, d_Y)$ be nonexpansive and $t_1, t_2 \in FX$. Every $t \in \Gamma_F(t_1, t_2)$ satisfies $Ff(t_i) = Ff\big(F\pi_i(t)\big) = F(f \circ \pi_i)(t) = F\big(\pi_i \circ (f \times f)\big)(t) = F\pi_i\big(F(f \times f)(t)\big)$. Hence we can calculate

$$d_X^{\downarrow F}(t_1, t_2) = \inf\left\{\widetilde{F}d_X(t) \mid t \in \Gamma_F(t_1, t_2)\right\}$$

$$\geq \inf\left\{\widetilde{F}d_X(t) \mid t \in F(X \times X), \ F\pi_i\big(F(f \times f)(t)\big) = Ff(t_i)\right\} \qquad (5.7)$$

$$\geq \inf\left\{\widetilde{F}d_Y\big(F(f \times f)(t)\big) \mid t \in F(X \times X), \ F\pi_i\big(F(f \times f)(t)\big) = Ff(t_i)\right\} \quad (5.8)$$

$$\geq \inf\left\{\widetilde{F}d_Y(t') \mid t' \in \Gamma_F\big(Ff(t_1), Ff(t_2)\big)\right\} = d_Y^{\downarrow F}\big(Ff(t_1), Ff(t_2)\big). \qquad (5.9)$$

In this calculation the inequality (5.7) is due to our initial observation. Furthermore, (5.8) holds because f is nonexpansive, i.e., $d_X \geq d_Y \circ (f \times f)$ and applying the monotonicity (Condition W1 of Definition 5.14) of $\widetilde{F}$ yields $\widetilde{F}d_X \geq \widetilde{F}(d_Y \circ (f \times f)) = \widetilde{F}d_Y \circ F(f \times f)$. The last inequality, (5.9), is due to the fact that there might be more couplings t' than those obtained via $F(f \times f)$. □

Let us now study the properties of the Wasserstein lifting. As was the case for the Kantorovich lifting, also the Wasserstein lifting preserves isometries.

Theorem 5.23. *The Wasserstein lifting $\overline{F}$ of a functor F preserves isometries.*

Proof. Let $f \colon (X, d_X) \to (Y, d_Y)$ be an isometry. Since $\overline{F}$ is a functor, $\overline{F}f$ is nonexpansive, i.e., for all $t_1, t_2 \in FX$ we have $d_X^{\downarrow F}(t_1, t_2) \geq d_Y^{\downarrow F}\big(Ff(t_1), Ff(t_2)\big)$. Now we show the opposite direction, i.e., that for all $t_1, t_2 \in FX$ we have $d_X^{\downarrow F}(t_1, t_2) \leq d_Y^{\downarrow F}\big(Ff(t_1), Ff(t_2)\big)$.

If $\Gamma_F\big(Ff(t_1), Ff(t_2)\big) = \emptyset$ we have $d_Y^{\downarrow F}\big((Ff(t_1), Ff(t_2)\big) = \top \geq d_X^{\downarrow F}(t_1, t_2)$. Otherwise we will construct for each coupling $t \in \Gamma_F\big(Ff(t_1), Ff(t_2)\big)$ a coupling $\gamma(t) \in \Gamma_F(t_1, t_2)$ such

that $\widetilde{F}d_X\big(\gamma(t)\big) = \widetilde{F}d_Y(t)$ because then we have

$$d_X^{\downarrow F}(t_1, t_2) = \inf_{t' \in \Gamma_F(t_1, t_2)} \widetilde{F}d_X(t') \leq \inf_{t \in \Gamma_F\big(Ff(t_1), Ff(t_2)\big)} \widetilde{F}d_X\big(\gamma(t)\big)$$

$$= \inf_{t \in \Gamma_F\big(Ff(t_1), Ff(t_2)\big)} \widetilde{F}d_Y(t) = d_Y^{\downarrow F}\big(Ff(t_1), Ff(t_2)\big)$$

as desired. In this calculation the inequality is due to the fact that $\gamma(t) \in \Gamma_F(t_1, t_2)$ is a coupling and there might be other couplings which are not in the image of γ.

In order to construct $\gamma \colon \Gamma_F\big(Ff(t_1), Ff(t_2)\big) \to \Gamma_F(t_1, t_2)$, we consider the diagram below where $\pi_1 \colon X \times Y \to X$, $\pi_2 \colon Y \times X \to X$, and $\tau_i \colon Y \times Y \to Y$ are the respective projections of the products.

This diagram consists of pullbacks: it is easy to check that the diagram commutes. The unique mediating arrows are constructed as follows: in every item we assume that P is another object – with a commuting square – and we construct the mediating morphism to the pullback object.

▷ For the lower left part let P be a set with $p_1 \colon P \to X$, $p_2 \colon P \to Y \times Y$ such that $f \circ p_1 = \tau_1 \circ p_2$, then define $u \colon P \to X \times Y$ as $u := \langle p_1, \tau_2 \circ p_2 \rangle$.

▷ Analogously, for the lower right part let P be a set with $p_1 \colon P \to Y \times Y$, $p_2 \colon P \to X$ such that $\tau_2 \circ p_1 = f \circ p_2$, then define $u \colon P \to Y \times X$ as $u := \langle \tau_1 \circ p_1, p_2 \rangle$.

▷ Finally, for the upper part let P be a set with $p_1 \colon P \to X \times Y$, $p_2 \colon P \to Y \times X$ such that $(f \times \mathrm{id}_Y) \circ p_1 = (\mathrm{id}_Y \times f) \circ p_2$, then define $u \colon P \to X \times X$ as $u := \langle \pi_1 \circ p_1, \pi_2 \circ p_2 \rangle$.

We apply the weak pullback preserving functor F to the diagram and obtain the following diagram which hence consists of three weak pullbacks.

Given a coupling $t \in \Gamma_F\big(Ff(t_1), Ff(t_2)\big) \subseteq F(Y \times Y)$ we know $F\tau_i(t) = Ff(t_i) \in FY$.

Since the lower left square in the diagram is a weak pullback, we obtain an element $s_1 \in F(X \times Y)$ with $F\pi_1(s_1) = t_1$ and $F(f \times \mathrm{id}_Y)(s_1) = t$. Similarly, from the lower right square, we obtain $s_2 \in F(Y \times X)$ with $F\pi_2(s_2) = t_2$ and $F(\mathrm{id}_Y \times f)(s_2) = t$. Again by

the weak pullback property we obtain our $\gamma(t) \in F(X \times X)$ with $F(\mathrm{id}_X \times f)\big(\gamma(t)\big) = s_1$, $F(f \times \mathrm{id}_X)\big(\gamma(t)\big) = s_2$.

We convince ourselves that $\gamma(t)$ is indeed a coupling of t_1 and t_2: Let $\pi'_i \colon X \times X \to X$ be the missing projections. Using $\pi_1 \circ (\mathrm{id}_X \times f) = \pi'_1$ we have $F\pi'_1\big(\gamma(t)\big) = F\big(\pi_1 \circ (\mathrm{id}_X \times f)\big)\big(\gamma(t)\big) = F\pi_1 \circ F(\mathrm{id}_X \times f)\big(\gamma(t)\big) = F\pi_1(s_1) = t_1$ and analogously $F\pi'_2\big(\gamma(t)\big) = F\big(\pi_2 \circ (f \times \mathrm{id}_X)\big)\big(\gamma(t)\big) = F\pi_2 \circ F(f \times \mathrm{id}_X)\big(\gamma(t)\big) = F\pi_2(s_2) = t_2$. Moreover, using $f \times f = (f \times \mathrm{id}_Y) \circ (\mathrm{id}_X \times f)$ and functoriality we have

$$F(f \times f)\big(\gamma(t)\big) = F\big((f \times \mathrm{id}_Y) \circ (\mathrm{id}_X \times f)\big)\big(\gamma(t)\big)$$
$$= F(f \times \mathrm{id}_Y) \circ F(\mathrm{id}_X \times f)\big(\gamma(t)\big) = F(f \times \mathrm{id}_Y)(s_1) = t$$

and thus $\widetilde{F}d_Y(t) = \widetilde{F}d_Y(F(f \times f)(\gamma(t))) = \widetilde{F}(d_Y \circ (f \times f))(\gamma(t)) = \widetilde{F}(d_X(\gamma(t)))$ as desired. Note that the last equality is due to the fact that f is an isometry. $\qquad\square$

In contrast to the Kantorovich lifting, we can prove that metrics are preserved by the Wasserstein lifting in certain situations.

Theorem 5.24 (Preservation of Metrics)**.** *Let F be a weak pullback preserving endofunctor on Set with well-behaved evaluation function ev_F and (X, d) be a metric space. If for all $t_1, t_2 \in FX$ where $d^{\downarrow F}(t_1, t_2) = 0$ there is an optimal F-coupling $\gamma(t_1, t_2) \in \Gamma_F(t_1, t_2)$ such that $0 = d^{\downarrow F}(t_1, t_2) = \widetilde{F}d\big(\gamma(t_1, t_2)\big)$ then $d^{\downarrow F}$ is a metric and thus $\overline{F}(X, d) = (FX, d^{\downarrow F})$ is a metric space.*

Proof. Let (X, d) be a metric space. By Theorem 5.20 $d^{\downarrow F}$ is a pseudometric. Thus we just have to show that for any $t_1, t_2 \in FX$ the fact that $d^{\downarrow F}(t_1, t_2) = 0$ implies $t_1 = t_2$.

Since d is a metric the preimage $d^{-1}[\{0\}]$ is the set $\Delta_X = \{(x, x) \mid x \in X\}$. Hence the square on the left below is a pullback and adding the projections yields $\pi_1 \circ e = \pi_2 \circ e$ where $e \colon \Delta_X \hookrightarrow X \times X$ is the inclusion. Furthermore, by Lemma 5.17 we know that due to Condition W3 of Definition 5.14 the square on the right is a weak pullback.

$$
\begin{array}{ccc}
\Delta_X & \xrightarrow{\ !_{\Delta_X}\ } & \{0\} \\
{\scriptstyle e}\uparrow & & \uparrow{\scriptstyle i} \\
X \underset{\pi_2}{\overset{\pi_1}{\rightleftarrows}} X \times X & \xrightarrow{\ d\ } & [0, \top]
\end{array}
\qquad\qquad
\begin{array}{ccc}
F\{0\} & \xrightarrow{\ !_{F\{0\}}\ } & \{0\} \\
{\scriptstyle Fi}\downarrow & & \uparrow{\scriptstyle i} \\
F[0, \top] & \xrightarrow{\ ev_F\ } & [0, \top]
\end{array}
$$

Since F preserves weak pullbacks, applying it to the first diagram yields a weak pullback. By combining this diagram with the right diagram we obtain the diagram below where the outer rectangle is again a weak pullback.

$$
\begin{array}{ccccccc}
& & F\Delta_X & \xrightarrow{\ F!_{\Delta_X}\ } & F\{0\} & \xrightarrow{\ !_{F\{0\}}\ } & \{0\} \\
& {\scriptstyle F\pi_1} & {\scriptstyle Fe}\downarrow & & \downarrow{\scriptstyle Fi} & {\scriptstyle ev_F} & \uparrow{\scriptstyle i} \\
FX & \underset{F\pi_2}{\overset{}{\rightleftarrows}} & F(X \times X) & \xrightarrow{\ Fd\ } & F[0, \top] & \longrightarrow & [0, \top] \\
& & & \underset{\widetilde{F}d}{\searrow} & & &
\end{array}
$$

Let $t := \gamma(t_1, t_2) \in F(X \times X)$, i.e., $d^{\downarrow F}(t_1, t_2) = \widetilde{F}d(t) = 0$. Since we have a weak pullback, there exists $t' \in F\Delta_X$ with $Fe(t') = t$. (Since Fe is an embedding, t' and t actually coincide.) This implies that $t_1 = F\pi_1(t) = F\pi_1\big(Fe(t')\big) = F\pi_2\big(Fe(t')\big) = F\pi_2(t) = t_2$. $\qquad\square$

Apparently, Theorem 5.24 admits the following simple corollary.

Corollary 5.25. *Let F be a weak pullback preserving endofunctor on Set with well-behaved evaluation function ev_F and (X, d) be a metric space. If the infimum in (5.1) is always a minimum then $d^{\downarrow F}$ is a metric and thus $\overline{F}(X, d) = (FX, d^{\downarrow F})$ is a metric space.*

Please note that a similar restriction for the Kantorovich lifting (i.e., requiring that the supremum in Definition 5.4 is a maximum) does *not* yield preservation of metrics: In Example 5.10 the supremum is always a maximum but we do not get a metric. Let us now compare both lifting approaches.

Lemma 5.26. *Let F be an endofunctor on Set with evaluation function ev_F and (X, d) be a pseudometric space. If ev_F satisfies Conditions W1 and W2 of Definition 5.14 then for all $t_1, t_2 \in FX$, all $t \in \Gamma_F(t_1, t_2)$ and all nonexpansive functions $f: (X, d) \rightarrow ([0, \top], d_e)$ we have $d_e\big(\widetilde{F}f(t_1), \widetilde{F}f(t_2)\big) \leq \widetilde{F}d(t)$.*

Proof. We have $d_e \circ (f \times f) \leq d$ since f is nonexpansive. Due to monotonicity of the evaluation functor (Condition W1) we obtain $\widetilde{F}d_e \circ F(f \times f) = \widetilde{F}\big(d_e \circ (f \times f)\big) \leq \widetilde{F}d$. Furthermore:

$$d_e\big(\widetilde{F}f(t_1), \widetilde{F}f(t_2)\big) = d_e\Big(\widetilde{F}f\big(F\pi_1(t)\big), \widetilde{F}f\big(F\pi_2(t)\big)\Big) = d_e\Big(\widetilde{F}(f \circ \pi_1)(t), \widetilde{F}(f \circ \pi_2)(t)\Big)$$

$$= d_e\Big(\widetilde{F}\big(\pi_1 \circ (f \times f)\big)(t), \widetilde{F}\big(\pi_2 \circ (f \times f)\big)(t)\Big) = d_e\Big(\widetilde{F}\pi_1\big(F(f \times f)(t)\big), \widetilde{F}\pi_2\big(F(f \times f)(t)\big)\Big)$$

$$\leq \widetilde{F}d_e\big(F(f \times f)(t)\big) = \widetilde{F}\big(d_e \circ (f \times f)\big)(t) \leq \widetilde{F}d(t)$$

where the first inequality is due to Condition W2, the second due to the above observation which was based on Condition W1. $\square$

Using this result we can see that under certain conditions the Wasserstein distance is an upper bound for the Kantorovich pseudometric.

Theorem 5.27 (Comparison of the two Liftings). *Let F be an endofunctor on Set. If ev_F satisfies Conditions W1 and W2 of Definition 5.14 then for all pseudometric spaces (X, d) we have $d^{\uparrow F} \leq d^{\downarrow F}$.*

Proof. Let $t_1, t_2 \in FX$. We know (see the discussion after Definition 5.4) that a nonexpansive function $f: (X, d) \rightarrow ([0, \top], d_e)$ always exists but (see the discussion after Definition 5.12) couplings do not have to exist, so we have to distinguish two cases. If $\Gamma_F(t_1, t_2) = \emptyset$ we have $d^{\downarrow F}(t_1, t_2) = \top$ and clearly $d^{\uparrow F} \leq \top$. Otherwise we can apply Lemma 5.26 to obtain the desired inequality. $\square$

This inequality can be strict as the following example shows.

Example 5.28 (Wasserstein Lifting of the Squaring Functor). It is easy to see that the squaring functor $S: Set \rightarrow Set$, $SX = X \times X$, $Sf = f \times f$ preserves weak pullbacks. Some simple calculations show that the evaluation function $ev_S: S[0, \infty] \rightarrow [0, \infty]$, $ev_S(r_1, r_2) = r_1 + r_2$ given in Example 5.10 is well-behaved [Ker16, Ex. 5.4.28]. We now continue Example 5.10 where we considered a metric space (X, d) with at least two elements, chose an element $t_1 = (x_1, x_2) \in SX = X \times X$ with $x_1 \neq x_2$ and defined $t_2 = (x_2, x_1)$. The unique coupling $t \in \Gamma_S(t_1, t_2)$ is $t = \big((x_1, x_2), (x_2, x_1)\big)$. Using that d is a metric we conclude that $d^{\downarrow S}(t_1, t_2) = \widetilde{S}d(t) = d(x_1, x_2) + d(x_2, x_1) = 2d(x_1, x_2) > 0$. However, in Example 5.10 we calculated $d^{\uparrow S}(t_1, t_2) = 0$.

Whenever the inequality in Theorem 5.27 can be replaced by an equality we will in the following say that the *Kantorovich-Rubinstein duality* or simply *duality* holds. In this case we obtain a canonical notion of distance on FX for any given pseudometric space (X, d).

To prove that the duality holds and simultaneously to calculate the distance of $t_1, t_2 \in FX$ it is sufficient to find a nonexpansive function $f \colon (X, d) \to ([0, \top], d_e)$ and a coupling $t \in \Gamma_F(t_1, t_2)$ such that $d_e\big(\widetilde{F}f(t_1), \widetilde{F}f(t_2)\big) = \widetilde{F}d_e(t)$. Due to Theorem 5.27 this value equals $d^{\uparrow F}(t_1, t_2) = d^{\downarrow F}(t_1, t_2)$. We will often employ this technique for the upcoming examples.

Example 5.29 (Duality for the Identity Functor). We consider the identity functor Id with the identity function as evaluation function, i.e., $ev_{\mathrm{Id}} = \mathrm{id}_{[0, \top]}$. For any $t_1, t_2 \in X$, $t := (t_1, t_2)$ is the unique coupling of t_1, t_2. Hence, $d^{\downarrow F}(t_1, t_2) = d(t_1, t_2)$. With the function $d(t_1, \_) \colon (X, d) \to ([0, \top], d_e)$ which is nonexpansive due to Lemma 3.9 we obtain duality because we have $d(t_1, t_2) = d_e\big(d(t_1, t_1), d(t_1, t_2)\big) \le d^{\uparrow F}(t_1, t_2) \le d^{\downarrow F}(t_1, t_2) = d(t_1, t_2)$ and thus equality. Similarly, if we define $ev_{\mathrm{Id}}(r) = c \cdot r$ for $r \in [0, \top]$, $0 < c \le 1$, the Kantorovich and Wasserstein liftings coincide and we obtain the discounted distance $d^{\uparrow F}(t_1, t_2) = d^{\downarrow F}(t_1, t_2) = c \cdot d(t_1, t_2)$.

Example 5.30 (Duality for the Distribution Functors). It is known that the probability distribution functor $\mathcal{D}$ and its variants preserve weak pullbacks [Sok11, Proposition 3.3]. It is easy show that the evaluation function $ev_{\mathcal{D}} \colon \mathcal{D}[0, 1] \to [0, 1]$, $ev_{\mathcal{D}}(P) = \sum_{x \in [0,1]} x \cdot P(x)$ which we have defined in Example 5.9 is well-behaved for all variants of the distribution functor (i.e., distributions or subdistributions with countable or finite supports). In order to do so, one just has to use the definition of the evaluation functor and the triangle inequality for the absolute value. We omit this proof [Ker16, Ex. 5.4.30] here and just observe that we recover the usual Wasserstein pseudometric, i.e., for any (sub)probability distributions $P_1, P_2 \colon X \to [0, 1]$ we have

$$d^{\downarrow \mathcal{D}}(P_1, P_2) = \inf \left\{ \sum_{x_1, x_2 \in X} d(x_1, x_2) \cdot P(x_1, x_2) \mid P \in \Gamma_{\mathcal{D}}(P_1, P_2) \right\}$$

and – for proper distributions only – the Kantorovich-Rubinstein duality [Vil09] from transportation theory for the discrete case. Moreover, in this case it is known (and easy to see using a linear program formulation) that for finite supports the above infimum is always a minimum. In the case of subdistributions we do not have duality: Let $P, Q \colon \mathbb{1} \to [0, 1]$ be subdistributions on the singleton set $\mathbb{1}$, i.e., $P(\checkmark) = p$ and $Q(\checkmark) = q$ with $p, q \in [0, 1]$. The only pseudometric on $\mathbb{1}$ is the discrete metric d so any function $f \colon \mathbb{1} \to [0, 1]$ is nonexpansive and we have $\widetilde{\mathcal{D}}f(P) = f(\checkmark) \cdot P(\checkmark)$. Hence the Kantorovich distance of P and Q is achieved for the function f where $f(\checkmark) = 1$ and equals $d^{\uparrow F}(P, Q) = |p - q|$. However, if $p \ne q$ it is easy to see that there are no couplings of P and Q so $d^{\downarrow F}(P, Q) = 1$. Thus for any p, q where $|p - q| < 1$ we do not have equality.

Example 5.31 (The Hausdorff Pseudometric for Finite Sets). Similar to Example 5.16 we assume $\top = \infty$ but here we just consider the finite powerset functor $\mathcal{P}_f$ with evaluation function $\max \colon \mathcal{P}_f([0, \infty]) \to [0, \infty]$ with $\max \emptyset = 0$ and $\min \emptyset = \infty$. We claim that in this setting we obtain duality and both pseudometrics are equal to the *Hausdorff pseudometric* d_H on $\mathcal{P}_f(X)$ which is defined as, for all $X_1, X_2 \in \mathcal{P}_f X$,

$$d_H(X_1, X_2) = \max \left\{ \max_{x_1 \in X_1} \min_{x_2 \in X_2} d(x_1, x_2), \max_{x_2 \in X_2} \min_{x_1 \in X_1} d(x_1, x_2) \right\} .$$

Note that this distance is ∞, if either X_1 or X_2 is empty.

We show our claim by proving that if X_1, X_2 are both non-empty there exists a coupling and a nonexpansive function that both witness the Hausdorff distance. Assume that the first value $\max_{x_1 \in X_1} \min_{x_2 \in X_2} d(x_1, x_2)$ is maximal and assume that $y_1 \in X_1$ is the element of X_1 for which the maximum is reached. Furthermore let $y_2 \in X_2$ be the closest element in X_2, i.e., the element for which $d(y_1, y_2)$ is minimal. We know that for all $x_1 \in X_1$ there exists $x_2^{x_1}$ such that $d(x_1, x_2^{x_1}) \leq d(y_1, y_2)$ and for all $x_2 \in X_2$ there exists $x_1^{x_2}$ such that $d(x_1^{x_2}, x_2) \leq d(y_1, y_2)$. Specifically, $x_2^{y_1} = y_2$. We use the coupling $T \subseteq X \times X$ with

$$T = \{(x_1, x_2^{x_1}) \mid x_1 \in X_1\} \cup \{(x_1^{x_2}, x_2) \mid x_2 \in X_2\} \ .$$

Indeed, we obviously have $\mathcal{P}_f \pi_i(T) = X_i$ and $\mathcal{P}_f d(T)$ contains all distances between the elements above, of which the distance $d(y_1, y_2) = d^H(X_1, X_2)$ is maximal. We define a nonexpansive function $f \colon (X, d) \to ([0, \top], d_e)$ as $f(x) = \min_{x_2 \in X_2} d(x, x_2)$. Then we have

$$\max \mathcal{P}_f f(X_1) = \max f[X_1] = \max_{x_1 \in X_1} \min_{x_2 \in X_2} d(x_1, x_2) = d^H(X_1, X_2)$$

and $\max \mathcal{P}_f f(X_2) = \max f[X_2] = 0$. Hence, the difference of both values is $d^H(X_1, X_2)$. It remains to show that f is nonexpansive. Let $x, y \in X$ and let $x_2, y_2 \in X_2$ be elements for which the distances $d(x, x_2), d(y, y_2)$ are minimal. Hence $d(x, x_2) \leq d(x, y_2) \leq d(x, y) + d(y, y_2)$ and $d(y, y_2) \leq d(y, x_2) \leq d(y, x) + d(x, x_2)$. Lemma 3.3 implies that $d(x, y) \geq d_e\big(d(x, x_2), d(y, y_2)\big) = d_e\big(f(x), f(y)\big)$.

If $X_1 = X_2 = \emptyset$, we can use the coupling $T = \emptyset = \emptyset \times \emptyset$ and any function f. If, instead $X_1 = \emptyset$, $X_2 \neq \emptyset$, no coupling exists thus $d^{\downarrow F} = \infty$ and we can take the constant ∞-function to show that also $d^{\uparrow F} = \infty$ is attained.

It would also be interesting to consider the general or countable powerset functor, use the supremum as (well-behaved) evaluation function and consider the resulting liftings. By generalizing the argument and finding suitable approximations for the infimum and supremum, one can again show that the Wasserstein as well as the Kantorovich lifting coincide with the Hausdorff metric (with supremum/infimum replacing maximum/minimum).

On the other hand, we can argue that in this case optimal couplings do not necessarily exist, because the Hausdorff pseudometric for countable sets does not preserve metrics. If we take the Euclidean metric and consider the sets $X_1 = \{0\} \cup \{1/n \mid n \in \mathbb{N}\}$ and $X_2 = \{1/n \mid n \in \mathbb{N}\}$ then their Hausdorff distance is 0 although X_1 and X_2 are different. Thus, due to Theorem 5.24, there cannot be an optimal coupling for X_1 and X_2.

As another example of our lifting approaches, we consider the input functor $\_^A$ on Set where A is an arbitrary set. It maps each set X to the set X^A of functions with domain A and codomain X and each function $f \colon X \to Y$ to the function $f^A \colon X^A \to Y^A$ where $f^A(g \colon A \to X) = f \circ g$. Rutten showed in [Rut00] that functors on Set which preserve pullbacks also preserve weak pullbacks, which applies to the input functor. We will now present different well-behaved evaluation functions for this functor and the resulting Wasserstein pseudometrics.

Example 5.32 (Wasserstein Lifting for the Input Functor)**.** We consider the input functor $\_^A$ with finite input set A and claim that the evaluation functions $ev_F \colon [0, \top]^A \to [0, \top]$ which are listed in the table below are well-behaved and yield the given Wasserstein pseudometric on X^A for any pseudometric space (X, d). Note that in the third case it is necessary to normalize with $|A|^{-1}$ so that the distance does not exceed $\top$.

maximal distance $\top$	$ev_F(s)$	$d^{\downarrow F}(s_1, s_2)$
$\top \in]0, \infty]$	$\max\limits_{a \in A} s(a)$	$\max\limits_{a \in A} d\big(s_1(a), s_2(a)\big)$
$\top = \infty$	$\sum\limits_{a \in A} s(a)$	$\sum\limits_{a \in A} d\big(s_1(a), s_2(a)\big)$
$\top \in]0, \infty[$	$|A|^{-1} \sum\limits_{a \in A} s(a)$	$|A|^{-1} \sum\limits_{a \in A} d\big(s_1(a), s_2(a)\big)$

In order to show this we first observe that for any $f \colon X \to [0, \top]$ we have $\widetilde{F}f = ev_F \circ f^A$ so applying it to $s \in X^A$ yields either $\max_{a \in A} f\big(s(a)\big)$ or $\sum_{a \in A} f\big(s(a)\big)$ or $|A|^{-1} \sum_{a \in A} f\big(s(a)\big)$. With this we proceed to show well-behavedness.

W1. For $f_1, f_2 \colon X \to [0, \top]$ with $f_1 \leq f_2$ we obviously also have $\widetilde{F}f_1 \leq \widetilde{F}f_2$.

W2. Let $s \in ([0, \top]^2)^A$ and $s_i := \pi_i^A(t)$, i.e., necessarily $s = \langle s_1, s_2 \rangle$. We have to show the inequality $d_e\big(ev_F(s_1), ev_F(s_2)\big) \leq \widetilde{F}d_e(s)$ where the right hand side evaluates to $ev_F\big(d_e^A(s)\big) = ev_F(d_e \circ s) = ev_F(d_e \circ \langle s_1, s_2 \rangle)$. Using this we can see that the inequality is an immediate consequence of Lemma 3.4 (page 9) by taking $f = s_1$, $g = s_2$ or, in the last case, $f = |A|^{-1}s_1$ and $g = |A|^{-1}s_2$.

W3. We have $ev_F^{-1}[\{0\}] = \{s \colon A \to [0, \top] \mid ev_F(s) = 0\}$. Clearly for all functions this is the case only if s is the constant 0-function. Since $\{0\}$ is a final object in Set, there is a unique function $z \colon A \to \{0\}$. Thus $Fi[F\{0\}] = i^A[\{0\}^A] = \{i^A(z)\} = \{i \circ z\}$ and clearly $i \circ z \colon A \to [0, \top]$ is also the constant 0-function.

Now if we have $s_1, s_2 \in X^A$ their unique coupling is $s := \langle s_1, s_2 \rangle \colon A \to X \times X$. Moreover $\widetilde{F}d(s) = ev_F\big(d^A(s)\big) = ev_F\big(d \circ \langle s_1, s_2 \rangle\big)$ and using the different evaluation functions we obtain the pseudometrics given in the table above.

We conclude our list of examples with the Wasserstein lifting of the machine functor which we will use several times in the remainder of this paper.

Example 5.33 (Wasserstein Lifting of the Machine Functor)**.** We equip the machine functor $M_B = B \times \_^A$ with the evaluation function $ev_{M_B} \colon B \times [0, \top]^A \to [0, \top]$, $(o, s) \mapsto c \cdot ev_I(s)$ where $c \in]0, 1]$ is a discount factor and and ev_I is one of the evaluation functions for the input functor from Example 5.32. For any pseudometric space (X, d) we can easily see that for two elements $(o_1, s_1), (o_2, s_2) \in B \times X^A$ we have a unique coupling if and only if $o_1 = o_2$, namely $(o_1, \langle s_1, s_2 \rangle)$ (for $o_1 \neq o_2$ no coupling exists at all). Thus the Wasserstein distance of any two elements as above is 1 if $o_1 \neq o_2$ and $c \cdot ev_I(d \circ \langle s_1, s_2 \rangle)$ otherwise.

5.3. **Lifting Multifunctors.** While the functors we considered so far can be nicely lifted using our theory, there are other functors that require a more general treatment. For instance, consider the output functor $F = B \times \_$ for some fixed set B. As in Example 5.33 we have a coupling for $t_1, t_2 \in FX = B \times X$ with $t_i = (b_i, x_i)$ if and only if $b_1 = b_2$. Consequently, if $b_1 \neq b_2$ then irrespective of the evaluation function we choose and of the distance between x_1 and x_2 in (X, d), the lifted Wasserstein pseudometric will always result in $d^{\downarrow F}(t_1, t_2) = \top$. This can be counterintuitive, e.g., taking $B = [0, 1]$, $X \neq \emptyset$ and $t_1 = (0, x)$ and $t_2 = (\epsilon, x)$ for a small $\epsilon > 0$ and an $x \in X$. The reason is that we think of $B = [0, 1]$ as if it were endowed with a non-discrete pseudometric, like e.g. the Euclidean metric d_e, plugged into the product *after* the lifting. This intuition can be formalized by considering the lifting of the product seen as a functor from $Set \times Set$ into Set. More generally, it can be seen that the definitions and results introduced so far for endofunctors in Set extend to multifunctors on Set, i.e., to functors $F \colon Set^n \to Set$ on the product category Set^n for any natural number $n \in \mathbb{N}$. The

only difference is that we start with n pseudometric spaces instead of one. Due to this, the definitions and results are technically more complicated than in the endofunctor setting but they capture exactly the same ideas as before.

For clarity we provide some of the multifunctor results here but it is safe to skip the results at a first read, continue with studying the product and coproduct bifunctors in Section 5.4 (page 31) – they will play an important role for the later development of our theory – and only look at the exact multifunctor definitions when necessary.

The formal definition of multifunctor lifting is a straightforward extension of Definition 5.1 with only a little bit of added technical complexity.

Definition 5.34 (Lifting of a Multifunctor). Let $U\colon PMet \to Set$ be the forgetful functor which maps every pseudometric space to its underlying set and denote by $U^n\colon PMet^n \to Set^n$ the n-fold product of U with itself. A functor $\overline{F}\colon PMet^n \to PMet$ is called a *lifting* of a functor $F\colon Set^n \to Set$ if it satisfies $U \circ \overline{F} = F \circ U^n$.

These multifunctor liftings can be used to obtain endofunctor liftings by fixing all but one parameter [Ker16, Lem. 5.4.37]. As in the endofunctor case any multifunctor lifting is monotone in the following sense: If we have pseudometrics $d_i \le e_i$ on common sets X_i we also have $(d_1, \ldots, d_n)^F \le (e_1, \ldots, e_n)^F$ where $(d_1, \ldots, d_n)^F$ and $(e_1, \ldots, e_n)^F$ denote the pseudometrics on $F(X_1, \ldots, X_n)$ which we obtain by applying $\overline{F}$ to $(X_1, d_1), \ldots, (X_n, d_n)$ or $(X_1, e_1), \ldots, (X_n, e_n)$ respectively. We omit the proof of this property since the line of argument is exactly the same as used in the proof of Theorem 5.2 (page 13). One just has to take proper care of putting the universal quantification (for all $1 \le i \le n$) in the right place. Also for the following results we will just provide a reference to the corresponding endofunctor result and omit the simple (but admittedly tedious) calculations.

The really useful feature of considering multifunctor liftings is based on the fact that we have a slightly different domain of definition for evaluation functions which will also help us to solve the problems we described initially.

Definition 5.35 (Multifunctor Evaluation Function and Evaluation Multifunctor). Let $F\colon Set^n \to Set$ be a multifunctor. We call any function $ev_F\colon F([0, \top], \ldots, [0, \top]) \to [0, \top]$ an *evaluation function* for F. Given such an evaluation function, the *evaluation multifunctor* is the functor $\widetilde{F}\colon (Set/[0, \top])^n \to Set/[0, \top]$ where $\widetilde{F}(g_1, \ldots, g_n) := ev_F \circ F(g_1, \ldots, g_n)$ for all $g_i \in Set/[0, \top]$ and on arrows $\widetilde{F}$ coincides with F.

Using this function, we immediately get the Kantorovich pseudometric and the corresponding lifting.

Definition 5.36 (Kantorovich Distance for Multifunctors). Let $F\colon Set^n \to Set$ be a functor with evaluation function $ev_F\colon F([0, \top], \ldots, [0, \top]) \to [0, \top]$ and let $(X_1, d_1), \ldots, (X_n, d_n)$ be arbitrary pseudometric spaces. The *Kantorovich distance* is the function $(d_1, \ldots, d_n)^{\uparrow F}\colon \big(F(X_1, \ldots, X_n)\big)^2 \to [0, \top]$, where

$$(d_1, \ldots, d_n)^{\uparrow F}(t_1, t_2) := \sup_{\substack{(f_1, \ldots, f_n) \\ f_i\colon (X_i, d_i) \to ([0, \top], d_e)}} d_e\Big(\widetilde{F}(f_1, \ldots, f_n)(t_1), \widetilde{F}(f_1, \ldots, f_n)(t_2)\Big)$$

for all $t_1, t_2 \in F(X_1, \ldots, X_n)$.

By adapting the proof of Theorem 5.5 (page 14) we can prove that this function is indeed a pseudometric on $F(X_1, \ldots, X_n)$ and can thus define the Kantorovich lifting of

F as $\overline{F}\colon PMet^n \to PMet$, $\overline{F}((X_1, d_1), \ldots, (X_n, d_n)) = (F(X_1, \ldots, X_n), (d_1, \ldots, d_n)^{\uparrow F})$ and $\overline{F}f = Ff$. The soundness of this definition can be shown using the same line of argument as in Theorem 5.7 (page 15).

Not only the Kantorovich but also the Wasserstein lifting can be transferred to the multifunctor setting. For this we first need to define couplings.

Definition 5.37 (Coupling). Let $F\colon Set^n \to Set$ be a functor and $m \in \mathbb{N}$. Given sets $X_1, \ldots, X_n$ and elements $t_j \in F(X_1, \ldots, X_n)$ for $1 \leq j \leq m$ we call an element $t \in F(X_1^m, \ldots, X_n^m)$ such that $F(\pi_{1,j}, \ldots, \pi_{n,j})(t) = t_j$ a *coupling* of the t_j (with respect to F) where $\pi_{i,j}$ are the projections $\pi_{i,j}\colon X_i^m \to X_i$. We write $\Gamma_F(t_1, t_2, \ldots, t_m)$ for the set of all these couplings.

Using these couplings we can then again define a Wasserstein distance.

Definition 5.38 (Wasserstein Distance for Multifunctors). Let $F\colon Set^n \to Set$ be a functor with evaluation function $ev_F\colon F([0, \top], \ldots, [0, \top]) \to [0, \top]$ and let (X_1, d_1), $\ldots$, (X_n, d_n) be arbitrary pseudometric spaces. The *Wasserstein distance* is the function $(d_1, \ldots, d_n)^{\downarrow F}\colon \big(F(X_1, \ldots, X_n)\big)^2 \to [0, \top]$, where

$$(d_1, \ldots, d_n)^{\downarrow F}(t_1, t_2) := \inf_{t \in \Gamma_F(t_1, t_2)} \widetilde{F}(d_1, \ldots, d_n)(t).$$

for all $t_1, t_2 \in F(X_1, \ldots, X_n)$.

As before, we will use well-behaved evaluation functions along with pullback preserving functors to obtain a Wasserstein pseudometric.

Definition 5.39 (Well-Behaved Multifunctor Evaluation Function). We call a multifunctor evaluation function $ev_F\colon F([0, \top], \ldots, [0, \top]) \to [0, \top]$ *well-behaved* if it satisfies the following three properties.

W1. $\widetilde{F}$ is monotone, i.e., given $f_i, g_i\colon X_i \to [0, \top]$ with $f_i \leq g_i$ for all $1 \leq i \leq n$, we also have $\widetilde{F}(f_1, \ldots, f_n) \leq \widetilde{F}(g_1, \ldots, g_n)$.

W2. Let $\pi_i\colon [0, \top]^2 \to [0, \top]$ be the projections. For all couplings $t \in F([0, \top]^2, \ldots, [0, \top]^2)$ we require $d_e\big(\widetilde{F}(\pi_1, \ldots, \pi_1)(t), \widetilde{F}(\pi_2, \ldots, \pi_2)(t)\big) \leq \widetilde{F}(d_e, \ldots, d_e)(t)$.

W3. $ev_F^{-1}[\{0\}] = F(i, \ldots, i)[F(\{0\}, \ldots, \{0\})]$ where $i\colon \{0\} \hookrightarrow [0, \top]$ is the inclusion map.

By using a generalization [Ker16, Lem. 5.4.47] of the Gluing Lemma (Lemma 5.18) and well-behavedness we can prove sufficient conditions for the Wasserstein distance to be a pseudometric just as we did in Theorem 5.20.

Theorem 5.40. *Let $F\colon Set^n \to Set$ be a functor with evaluation function ev_F. If F preserves weak pullbacks and ev_F is well-behaved then the Wasserstein distance is a pseudometric.*

This result gives rise to the Wasserstein lifting: For a weak pullback preserving functor $F\colon Set^n \to Set$ with well-behaved evaluation function $ev_F\colon F([0, \top], \ldots, [0, \top]) \to [0, \top]$ the *Wasserstein lifting* is the functor $\overline{F}\colon PMet^n \to PMet$, where $\overline{F}\big((X_1, d_1), \ldots, (X_n, d_n)\big) = \big(F(X_1, \ldots, X_n), d\big)$ with $d = (d_1, \ldots, d_n)^{\downarrow F}$ and $\overline{F}f = Ff$. This definition is justified by adapting the proof of Theorem 5.22 to obtain functoriality.

With these results at hand we quickly summarize a few of the properties for the multifunctor liftings which arise as natural generalizations of Lemma 5.26 and Theorems 5.8, 5.23, 5.24 and 5.27.

Theorem 5.41. *Let $F\colon Set^n \to Set$ be a functor with evaluation function ev_F.*

1. *Both liftings preserve isometries.*
2. *If ev_F satisfies Conditions W1 and W2 then $(d_1,\ldots,d_n)^{\uparrow F} \leq (d_1,\ldots,d_n)^{\downarrow F}$ holds for all pseudometric spaces (X_i, d_i).*
3. *If F preserves weak pullbacks, ev_F is well-behaved and for all $t_1, t_2 \in F(X_1,\ldots,X_n)$ with $(d_1,\ldots,d_n)^{\downarrow F}(t_1,t_2) = 0$ there is a coupling $\gamma(t_1,t_2) \in \Gamma_F(t_1,t_2)$ such that $0 = (d_1,\ldots,d_n)^{\downarrow F}(t_1,t_2) = \widetilde{F}(d_1,\ldots,d_n)\big(\gamma(t_1,t_2)\big)$ then $(d_1,\ldots,d_n)^{\downarrow F}$ is a metric for all metric spaces (X_i, d_i).*

Of course, whenever the two pseudometrics coincide for a functor and an evaluation function, we say that the *Kantorovich-Rubinstein duality* or short *duality* holds.

5.4. **The Product and Coproduct Bifunctors.**

We conclude our section on multifunctors by considering two important examples at length, the product and the coproduct bifunctor.

Definition 5.42 (Product Bifunctor). The *product bifunctor* is the bifunctor $F\colon Set^2 \to Set$ where $F(X_1, X_2) = X_1 \times X_2$ for all sets X_1, X_2 and $F(f_1, f_2) = f_1 \times f_2$ for all functions $f_i\colon X_i \to Y_i$.

This functor fits nicely into our theory since it preserves pullbacks [Ker16, Lem. 5.4.53]. The proof is simple and hence omitted. Let us now discuss possible evaluation functions for this functor. They are similar to those for the input functor in Example 5.32 but we add some additional parameters as weighting factors to have additional flexibility and to demonstrate that one could choose different evaluation functions (this could be done analogously for the input functor).

Lemma 5.43 (Evaluation Functions for the Product Bifunctor). *Let F be the product bifunctor. The evaluation functions $ev_F\colon [0,\top]^2 \to [0,\top]$ presented below are well-behaved.*

Maximal Distance $\top$	Other Parameters	$ev_F(r_1, r_2)$
$\top \in\ ]0,\infty]$	$c_1, c_2 \in\ ]0,1]$	$\max\{c_1 r_1, c_2 r_2\}$
$\top = \infty$	$c_1, c_2 \in\ ]0,\infty[\,,\ p \in \mathbb{N}$	$(c_1 x_1^p + c_2 x_2^p)^{1/p}$
$\top \in\ ]0,\infty[$	$c_1, c_2 \in\ ]0,1], c_1 + c_2 \leq 1,\ p \in \mathbb{N}$	$(c_1 x_1^p + c_2 x_2^p)^{1/p}$

Proof. Apparently the only difference between the second and the third row is the range of the parameters. It ensures that $ev_F(r_1, r_2) \in [0, \top]$. We proceed by checking all three conditions for well-behavedness:

W1. Let $f_i, g_i\colon X_i \to [0, \top]$ with $f_i \leq g_i$ be given. For the maximum we have $\widetilde{F}(f_1, f_2) = \max\{c_1 f_1, c_2 f_2\} \leq \max\{c_1 g_1, c_2 g_2\} = \widetilde{F}(g_1, g_2)$ and for the second evaluation function, we also obtain $\widetilde{F}(f_1, f_2) = \big(c_1 \cdot f_1^p + c_2 \cdot f_2^p\big)^{1/p} \leq \big(c_1 \cdot g_1^p + c_2 \cdot g_2^p\big)^{1/p} = \widetilde{F}(g_1, g_2)$ due to monotonicity of all involved functions since $c_1, c_2 > 0$.

W2. Let $\pi_i\colon [0, \top]^2 \to [0, \top]$ be the product projections and define $t := (x_{11}, x_{21}, x_{12}, x_{22}) \in F([0,\top]^2, [0,\top]^2) = [0,\top]^2 \times [0,\top]^2$. We have to show the inequality

$$d_e\Big(\widetilde{F}(\pi_1, \pi_1)(t), \widetilde{F}(\pi_2, \pi_2)(t)\Big) \leq \widetilde{F}(d_e, d_e)(t). \qquad (5.10)$$

To do this, we first observe that the right hand side of this inequality evaluates to $\widetilde{F}(d_e, d_e)(t) = ev_F\big(d_e(x_{11}, x_{21}), d_e(x_{12}, x_{22})\big)$. Moreover, we have $\widetilde{F}(\pi_i, \pi_i)(t) =$

$ev_F(x_{i1}, x_{i2})$ so if we define $z_i = ev_F(x_{i1}, x_{i2})$ the left hand side of (5.10) can be rewritten as $d_e(z_1, z_2)$. Thus (5.10) is equivalent to

$$d_e(z_1, z_2) \leq ev_F\big(d_e(x_{11}, x_{21}), d_e(x_{12}, x_{22})\big). \qquad (5.11)$$

If $z_1 = z_2$ this is obviously true because $d_e(z_1, z_2) = 0$ and the right hand side of (5.11) is non-negative. We now assume $z_1 > z_2$ (the other case is symmetrical). For $\infty = z_1 > z_2$ inequality (5.11) holds because then $x_{11} = \infty$ or $x_{12} = \infty$ and $x_{21}, x_{22} < \infty$ (otherwise we would have $z_2 = \infty$) so both the left hand side and the right hand side are ∞. Thus we can now restrict our attention to $\infty > z_1 > z_2$ where necessarily also $x_{11}, x_{12}, x_{21}, x_{22} < \infty$ (otherwise we would have $z_1 = \infty$ or $z_2 = \infty$). According to Lemma 3.3 (page 8), the inequality (5.11) is equivalent to showing the two inequalities $z_1 \leq z_2 + ev_F\big(d_e(x_{11}, x_{21}), d_e(x_{12}, x_{22})\big)$ and $z_2 \leq z_1 + ev_F\big(d_e(x_{11}, x_{21}), d_e(x_{12}, x_{22})\big)$. By our assumption ($\infty > z_1 > z_2$) the second of these inequalities is satisfied, so we just have to show the first.

1. For the discounted maximum as evaluation function we have $z_i = \max\{c_1 x_{i1}, c_2 x_{i2}\}$. If for z_1 the maximum is attained for the first element, i.e., if $z_1 = c_1 x_{11}$, we can conclude that $z_2 + \max\{c_1 d_e(x_{11}, x_{21}), c_2 d_e(x_{12}, x_{22})\} \geq z_2 + c_1 d_e(x_{11}, x_{21}) = z_2 + c_1|x_{11} - x_{21}| \geq z_2 + c_1(x_{11} - x_{21}) = z_2 + c_1 x_{11} - c_1 x_{21} = z_2 + z_1 - c_1 x_{21} = z_1 + (z_2 - c_1 x_{21}) \geq z_1$ because $z_2 = \max\{c_1 x_{21}, c_2 x_{22}\} > c_1 x_{21}$ and therefore $(z_2 - c_1 x_{21}) \geq 0$. The same line of argument can be applied if $z_1 = c_2 x_{12}$.

2. For the second evaluation function we can simply use the Minkowski inequality to obtain the result [Ker16, Lem. 5.4.54].

W3. Both functions satisfy Condition W3 of Definition 5.35, because $F(i, i)[F(\{0\}, \{0\})] = (i \times i)[\{0\} \times \{0\}] = \{(0,0)\}$ and for both evaluation functions apparently $ev_F^{-1}[\{0\}] = \{(0,0)\}$. $\qquad \square$

Using these well-behaved evaluation functions we can now lift the product bifunctor using our multifunctor lifting framework.

Lemma 5.44 (Product Pseudometrics). *Let F be the product bifunctor of Definition 5.42. For the evaluation functions presented in Lemma 5.43 the Kantorovich-Rubinstein duality holds and the supremum [infimum] of the Kantorovich [Wasserstein] pseudometric is always a maximum [minimum]. Moreover, for all pseudometric spaces (X_1, d_1), (X_2, d_2) we obtain the lifted pseudometrics $(d_1, d_2)^F : (X_1 \times X_2)^2 \to [0, \top]$ as given in the table below.*

$ev_F(r_1, r_2)$	$(d_1, d_2)^F\big((x_1, x_2), (y_1, y_2)\big)$
$\max\{c_1 r_1, c_2 r_2\}$	$\max\{c_1 d_1(x_1, y_1), c_2 d_2(x_2, y_2)\}$
$\big(c_1 x_1^p + c_2 x_2^p\big)^{1/p}$	$\big(c_1 d_1(x_1, y_1)^p + c_2 d_2(x_2, y_2)^p\big)^{1/p}$

Proof. Let (X_1, d_1), (X_1, d_2) be pseudometric spaces, $\pi_i : X_1^2 \to X_1$ and $\tau_i : X_2^2 \to X_2$ be the projections and let $t_1 = (x_1, x_2), t_2 = (y_1, y_2) \in F(X_1, X_2) = X_1 \times X_2$ be given. We define $t := (x_1, y_1, x_2, y_2) \in F(X_1^2, X_2^2)$ and observe that $F(\pi_1, \tau_1)(t) = t_1$, $F(\pi_2, \tau_2)(t) = t_2$ and thus $t \in \Gamma_F(t_1, t_2)$ is a coupling of t_1 and t_2. In the following we will construct nonexpansive functions $f_i : (X_i, d_i) \to ([0, \top], d_e)$ such that $d_e\big(\widetilde{F}(f_1, f_2)(t_1), \widetilde{F}(f_1, f_2)(t_2)\big) = \widetilde{F}(d_1, d_2)(t)$ holds. Due to Theorem 5.41 (page 31) we can then conclude that duality holds and both supremum and infimum are attained.

1. For the first evaluation function we have $\widetilde{F}(d_1, d_2)(t) = \max\{c_1 d_1(x_1, y_1), c_2 d_2(x_2, y_2)\}$ and assume without loss of generality that $c_1 d_1(x_1, y_1)$ is the maximal element. We define

$f_1 := d_1(x_1, \_)$, which is nonexpansive due to Lemma 3.9 (page 10), and f_2 to be the constant zero-function which is obviously nonexpansive as a constant function. Then we have:

$$d_e\Big(\widetilde{F}(f_1, f_2)(t_1), \widetilde{F}(f_1, f_2)(t_2)\Big)$$

$$= d_e\big(\max\{c_1 f_1(x_1), c_2 f_2(x_2)\}, \max\{c_1 f_1(y_1), c_2 f_2(y_2)\}\big)$$

$$= d_e\big(\max\{c_1 f_1(x_1), 0\}, \max\{c_1 f_1(y_1), 0\}\big) = d_e\big(c_1 f_1(x_1), c_1 f_1(y_1)\big)$$

$$= c_1 d_1(x_1, y_1) = \max\{c_1 d_1(x_1, y_1), c_2 d_2(x_2, y_2)\} = \widetilde{F}(d_1, d_2)(t)$$

The case where $c_2 d_2(x_2, y_2)$ is the maximal element is treated analogously.

2. For the second evaluation function we define $f_1 := d_1(x_1, \_)$ and $f_2 := d_2(x_2, \_)$ which are nonexpansive by Lemma 3.9 and obtain

$$d_e\Big(\widetilde{F}(f_1, f_2)(t_1), \widetilde{F}(f_1, f_2)(t_2)\Big)$$

$$= d_e\left((c_1 f_1^p(x_1) + c_2 f_2^p(x_2))^{1/p}, (c_1 f_1^p(y_1) + c_2 f_2^p(y_2))^{1/p}\right)$$

$$= d_e\left(0, (c_1 d_1^p(x_1, y_1) + c_2 d_2^p(x_2, y_2))^{1/p}\right)$$

$$= (c_1 d_1^p(x_1, y_1) + c_2 d_2^p(x_2, y_2))^{1/p} = \widetilde{F}(d_1, d_2)(t)$$

which completes the proof. $\qquad\square$

While all the product pseudometrics of Lemma 5.44 are well-known, we point out a specifically interesting one, the undiscounted maximum pseudometric.

Lemma 5.45 (Binary Products in $PMet$). *If $c_1 = c_2 = 1$ for the first evaluation function in Lemma 5.44 we obtain for two given pseudometric spaces (X_1, d_1) and (X_2, d_2) as lifted pseudometric the function $d_\infty \colon (X_1 \times X_2)^2 \to [0, \top]$, with $d\big((x_1, x_2), (y_1, y_2)\big) = \max\{d_1(x_1, y_1), d_2(x_2, y_2)\}$. The resulting pseudometric space $(X_1 \times X_2, d_\infty)$ is exactly the categorical product of (X_1, d_1) and (X_2, d_2) in $PMet$.*

Proof. This follows from Theorem 3.8 (page 9) by taking $I = 2$ and $f_i = \pi_i \colon X_1 \times X_2 \to X$. $\qquad\square$

In a completely analogous way as for the product bifunctor, we will now introduce and study the coproduct bifunctor.

Definition 5.46 (Coproduct Bifunctor). The *coproduct bifunctor* is the functor $F \colon Set^2 \to Set$, where $F(X_1, X_2) = X_1 + X_2$ for all sets X_1, X_2 and $F(f_1, f_2) = f_1 + f_2$ for all functions $f_1 \colon X_1 \to Y_1$, $f_2 \colon X_2 \to Y_2$. Explicitly[6], the function $f_1 + f_2 \colon X_1 + X_2 \to Y_1 + Y_2$ is given via the assignment $f_1 + f_2(x, i) = \big(f_i(x), i\big)$.

As for the product bifunctor, one can easily show that this bifunctor preserves pullbacks [Ker16, 5.4.58] and we omit the simple proof. For this functor we will just consider one type of evaluation function, parametrized by the maximal element $\top$ of the pseudometrics.

Lemma 5.47 (Evaluation Function for the Coproduct Bifunctor). *Let F be the coproduct bifunctor. The function $ev_F \colon [0, \top] + [0, \top] \to [0, \top]$, where $ev_F(x, i) = x$, is well-behaved.*

Proof. We show the three properties of a well-behaved evaluation function.

---

[6]We use the representation $X_1 + X_2 \cong X_1 \times \{1\} \cup X_2 \times \{2\}$.

W1. Let $f_1, f_2, g_1, g_2 \colon X \to [0, \top]$ with $f_1 \le g_1$, $f_2 \le g_2$ and $(z, i) \in F(X_1, X_2) = X_1 + X_2$. We have $\widetilde{F}(f_1, f_2)(z, i) = ev_F\big(F(f_1, f_2)(z, i)\big) = f_i(z) \le g_i(z) = ev_F\big(F(g_1, g_2)(z, i)\big) = \widetilde{F}(g_1, g_2)(z, i)$.

W2. Let $t = \big((x, y), i\big) \in F([0, \top]^2, [0, \top]^2) = [0, \top]^2 \times \{1, 2\}$. We obtain $\widetilde{F}(d_e, d_e)(t) = ev_F\big(d_e(x, y), i\big) = d_e(x, y) = d_e\big(ev_F(x, i), ev_F(y, i)\big) = d_e\Big(\widetilde{F}(\pi_1, \pi_1)(t), \widetilde{F}(\pi_2, \pi_2)(t)\Big)$.

W3. Let $i \colon 0 \hookrightarrow [0, \top]$ be the inclusion function. We have $Fi[F(\{0\}, \{0\})] = (i + i)[\{0\} + \{0\}] = \{0\} \times \{1, 2\} = ev_F^{-1}[\{0\}]$.   □

With this evaluation function we can now employ our multifunctor lifting framework to obtain the following coproduct pseudometric.

Lemma 5.48 (Coproduct Pseudometric). *For the coproduct bifunctor of Definition 5.46 and the evaluation function of Lemma 5.47 the Kantorovich-Rubinstein duality holds, the supremum of the Kantorovich pseudometric is always a maximum, the infimum of the Wasserstein pseudometric is a minimum whenever a coupling exists and we obtain the coproduct pseudometric*

$$d_+ \colon (X_1 + X_2)^2 \to [0, \top], \quad d_+\big((x_1, i_1), (x_2, i_2)\big) = \begin{cases} d_i(x_1, x_2), & \text{if } i_1 = i_2 = i \\ \top, & \text{else} \end{cases}.$$

Proof. Let (X_1, d_1), (X_1, d_2) be pseudometric spaces, $\pi_i \colon X_1^2 \to X_1$ and $\tau_i \colon X_2^2 \to X_2$ be the projections and and $t_1, t_2 \in F(X_1, X_2) = X_1 + X_2$, say $t_1 = (z, i)$, $t_2 = (z', i')$. We distinguish two cases.

1. For $i = i'$ we define $t = \big((z, z'), i\big)$ and observe that $F(\pi_1, \tau_1)\big((z, z'), i\big) = t_1$, and $F(\pi_2, \tau_2)\big((z, z'), i\big) = t_2$, thus $t \in \Gamma_F(t_1, t_2)$. Furthermore $\widetilde{F}(d_1, d_2)(t) = d_i(z, z')$. If $i = i' = 1$ we define $f_1 := d_1(z, \_) \colon (X_1, d_1) \to ([0, \top], d_e)$ which is nonexpansive according to Lemma 3.9 and consider an arbitrary nonexpansive function $f_2 \colon (X_2, d_2) \to ([0, \top], d_e)$ (e.g. the constant zero-function). Then we have $d_e\big(\widetilde{F}(f_1, f_2)(t_1), \widetilde{F}(f_1, f_2)(t_2)\big) = d_e\big(\widetilde{F}(f_1, f_2)(z, 1), \widetilde{F}(f_1, f_2)(z', 1)\big) = d_e\big(f_1(z), f_1(z')\big) = d_e\big(0, d_1(z, z')\big) = d_1(z, z') = d_i(z, z')$. The case $i = i' = 2$ can be treated analogously.

2. In the case where $i \ne i'$, there is no coupling that projects to (z, i) and (z', i'), thus $(d_1, d_2)^{\downarrow F}(t_1, t_2) = \top$. We show that also $(d_1, d_2)^{\uparrow F}(t_1, t_2) = \top$. We define f_1 to be the constant 0-function and f_2 the constant $\top$-function. We have $d_e\big(\widetilde{F}(f_1, f_2)(t_1), \widetilde{F}(f_1, f_2)(t_2)\big) = d_e\big(\widetilde{F}(f_1, f_2)(z, i), \widetilde{F}(f_1, f_2)(z', j)\big) = d_e\big(f_i(z), f_j(z')\big) = d_e(0, \top) = \top$.   □

We conclude this section by the observation that this yields the categorical coproduct.

Lemma 5.49 (Binary Coproducts in $PMet$). *The pseudometric space $(X_1 + X_2, d_+)$ where d_+ is the pseudometric given in Lemma 5.48 is exactly the categorical coproduct of (X_1, d_1) and (X_2, d_2) in $PMet$.*

Proof. This follows from Theorem 3.8 by taking $I = 2$ and $f_i = \iota_i \colon X_i \to X_1 + X_2$.   □

6. Bisimilarity Pseudometrics

We now want to use our lifting framework to derive bisimilarity pseudometrics. We assume an arbitrary lifting $\overline{F} \colon PMet \to PMet$ of an endofunctor F on Set and, for any pseudometric space (X, d), we write d^F for the pseudometric obtained by applying $\overline{F}$ to (X, d). Such a

lifting can be obtained as an endofunctor lifting, by taking a lifted multifunctor and fixing all parameters apart from one or by the composition of such functors.

We first observe that any coalgebra $c\colon X \to FX$ for a set functor $F\colon Set \to Set$ can be "lifted" to a coalgebra for the lifting $\overline{F}\colon PMet \to PMet$ by endowing X with a canonical pseudometric, i.e., the least pseudometric making c a nonexpansive function (that actually turns c into an isometry).

Lemma 6.1 (Lifting Coalgebras to $PMet$). *Let $\overline{F}\colon PMet \to PMet$ be a lifting of a functor $F\colon Set \to Set$ and let $c\colon X \to FX$ be a coalgebra. The mapping associating each pseudometric $d\colon X \times X \to [0, \top]$ with $d^F \circ (c \times c)$ is monotonic over the lattice of pseudometrics on X, hence it has a least fixed point*

$$d_c = \inf \left\{ d \mid d\colon X \times X \to [0, \top] \ pseudometric \wedge \ d^F \circ (c \times c) \le d \right\}$$

and $c\colon (X, d_c) \to (FX, d_c^F)$ is an isometry.

Moreover, if $c'\colon Y \to FY$ is another coalgebra and $f\colon X \to Y$ is a coalgebra homomorphism then $f\colon (X, d_c) \to (Y, d_{c'})$ is nonexpansive. It is an isometry if additionally $\overline{F}$ preserves isometries.

Proof. We first observe that the mapping associating each pseudometric $d\colon X \times X \to [0, \top]$ with $d^F \circ (c \times c)$ is monotonic over the lattice of pseudometrics on X. Let $d, d'\colon X \times X \to [0, \top]$ be pseudometrics such that $d \le d'$. Then by Theorem 5.2, $d^F \le d'^F$ and thus $d^F \circ (c \times c) \le d'^F \circ (c \times c)$, as desired. The fact that d_c is a fixed point exactly says that $c\colon (X, d_c) \to (FX, d_c^F)$ is an isometry.

Now, let $c'\colon Y \to FY$ be another coalgebra and $f\colon X \to Y$ be a coalgebra homomorphism. We first show that $f\colon (X, d_c) \to (Y, d_{c'})$ is nonexpansive. Consider the pseudometric $d_Y\colon Y \times Y \to [0, \top]$ defined as

$$d_Y = \sup \left\{ d \mid d\colon Y \times Y \to [0, \top] \ pseudometric \ \wedge \ f\colon (X, d_c) \to (Y, d) \ nonexpansive \right\} .$$

Since the supremum is taken pointwise, $f\colon (X, d_c) \to (Y, d_Y)$ is nonexpansive and thus $Ff\colon (FX, d_c^F) \to (FY, d_Y^F)$ is also nonexpansive. This allows us to deduce that

$$d_Y^F \circ (c' \times c') \le d_Y . \tag{6.1}$$

In fact we have

$$
\begin{aligned}
d_Y^F &\circ (c' \circ f \times c' \circ f) \\
&= d_Y^F \circ (Ff \circ c \times Ff \circ c) && [\text{since } f \text{ is coalgebra hom.: } Ff \circ c = c' \circ f] \\
&\le d_c^F \circ (c \times c) && [\text{since } Ff \text{ is nonexpansive}] \\
&\le d_c . && [\text{since } c \text{ is nonexpansive}]
\end{aligned}
$$

The above means that $f\colon (X, d_c) \to (Y, d_Y^F \circ (c' \times c'))$ is nonexpansive. Since d_Y is defined as the supremum of the pseudometrics making f nonexpansive, we conclude Equation (6.1). Given the characterization of $d_{c'}$ as least of pre-fixed points, it follows that $d_{c'} \le d_Y$. Recalling that $f\colon (X, d_c) \to (Y, d_Y)$ was nonexpansive, we conclude that $d_{c'} \circ (f \times f) \le d_Y \circ (f \times f) \le d_c$ thus $f\colon (X, d_c) \to (Y, d_{c'})$ is nonexpansive, as desired.

We finally prove that, whenever $\overline{F}$ preserves isometries, $f\colon (X, d_c) \to (Y, d_{c'})$ is an isometry. Define a pseudometric $d_X\colon X \times X \to [0, \top]$ as $d_X = d_{c'} \circ (f \times f)$. Observe that $f\colon (X, d_X) \to (Y, d_{c'})$ is isometric by definition. Hence $Ff\colon (FX, d_X^F) \to (FY, d_{c'}^F)$ is an

isometry. Moreover we have

$$
\begin{aligned}
d_X(x, x') \quad &= d_{c'} \circ (f \times f) && [\text{by definition of } d_X] \\
&= d_{c'}^F \circ (c' \circ f \times c' \circ f) && [\text{since } c' \colon (Y, d_{c'}) \to (FY, d_{c'}^F) \text{ is an isometry}] \\
&= d_{c'}^F \circ (Ff \circ c \times Ff \circ c) && [\text{since } f \text{ is coalgebra hom.: } Ff \circ c = c' \circ f] \\
&= d_X^F \circ (c \times c). && [\text{since } Ff \colon (FX, d_X^F) \to (FY, d_{c'}^F) \text{ is an isometry}]
\end{aligned}
$$

The above shows that $d_X^F \circ (c \times c) = d_X$. Therefore, by the characterization of d_c as least of pre-fixed points, it follows that $d_c \leq d_X$. Hence

$$
d_c \leq d_X = d_{c'} \circ (f \times f) \leq d_c
$$

the last inequality being nonexpansiveness of $f \colon (X, d_c) \to (Y, d_{c'})$. Hence we conclude that f is indeed an isometry. $\qquad\square$

For a coalgebra $c \colon X \to FX$, the least fixpoint pseudometric $d_c \colon X \times X \to [0, \top]$ above, given by the Knaster-Tarski theorem [Tar55], can be characterized also via transfinite induction. For every ordinal i we construct a pseudometric $d_i \colon X \times X \to [0, \top]$ as follows:

▷ $d_0 := 0$ is the zero pseudometric,
▷ $d_{i+1} := d_i^F \circ (c \times c)$ for all ordinals i and
▷ $d_j := \sup_{i<j} d_i$ for all limit ordinals j.

This sequence converges to d_θ for some ordinal θ.

Moreover, if $c' \colon X \times X \to [0, \top]$ is another coalgebra and $f \colon X \to Y$ is a coalgebra homomorphism and the fixpoint pseudometric $d_{c'} \colon X \times X \to [0, \top]$ is reached by the construction above at ordinal ζ, then $\theta \leq \zeta$.

We can now easily show that if $z \colon Z \to FZ$ is a final F-coalgebra, then its lifting is a final $\overline{F}$-coalgebra.

Theorem 6.2 (Final Coalgebra Construction for Liftings)**.** *Let* $\overline{F} \colon PMet \to PMet$ *be a lifting of a functor* $F \colon Set \to Set$ *such that* F *has a final coalgebra* $z \colon Z \to FZ$*. Then* $z \colon (Z, d_z) \to (FZ, d_z^F)$*, where* d_z *is defined as in Lemma 6.1 is the final* $\overline{F}$*-coalgebra.*

Proof. By Lemma 6.1, if we let $d_z = \inf\{d \mid d \colon Z \times Z \to [0, \top] \text{ pseudometric} \wedge d^F \circ (z \times z) \leq d\}$ then $z \colon (X, d_z) \to (FZ, d_z^F)$ is an isometry.

Let $c \colon (X, d_X) \to \overline{F}(X, d_X)$ be any $\overline{F}$-coalgebra. Again, Lemma 6.1, provides a pseudometric $d_c = \inf\{d \mid d \colon X \times X \to [0, \top] \text{ pseudometric} \wedge d^F \circ (c \times c) \leq d\}$ such that $c \colon (X, d_c) \to (FZ, d_c^F)$ is nonexpansive. Since c is also nonexpansive with respect to d_X, i.e., $d_X^F \circ (c \times c) \leq d_X$, we must have $d_c \leq d_X$.

Now, consider the underlying F-coalgebra $c \colon X \to FX$ in *Set*. Since z is the final F-coalgebra, there is a unique function $f \colon X \to Z$ such that $z \circ f = Ff \circ c$. By Lemma 6.1, $f \colon (X, d_c) \to (Z, d_z)$ is nonexpansive, thus $d_z \circ (f \times f) \leq d_c$. Recalling that $d_c \leq d_X$ we conclude that also $f \colon (X, d_X) \to (Z, d_z)$ is nonexpansive, as desired. $\qquad\square$

As a first simple example of this construction we consider the machine functor. We will look at more examples at the end of this section.

Example 6.3 (Final Coalgebra for the Lifted Machine Functor)**.** We consider the machine endofunctor $M_2 = 2 \times \_^A$, take $\top = 1$ and as evaluation function we use $ev_{M_2} \colon [0, 1] \times [0, 1]^A$ with $ev_{M_2}(o, s) = c \cdot \max_{a \in A} s(a)$ for $0 < c < 1$ as in Example 5.33 (page 28).

It is well-known that the carrier of the final M_2-coalgebra is 2^{A^*}. Moreover, from Example 5.33 we know that for any pseudometric d on 2^{A^*} we obtain as Wasserstein pseudometric the function $d^{\downarrow F} \colon \left(2 \times (2^{A^*})^A\right)^2 \to [0,1]$ where, for all $(o_1, s_1), (o_2, s_2) \in 2 \times (2^{A^*})^A$,

$$d^{\downarrow F}\big((o_1, s_1), (o_2, s_2)\big) = \max \left\{ d_2(o_1, o_2), c \cdot \max_{a \in A} d\big(s_1(a), s_2(a)\big) \right\}$$

where d_2 is the discrete metric on 2. Thus the fixed-point equation induced by Lemma 6.1 is given by, for $L_1, L_2 \in 2^{A^*}$,

$$d(L_1, L_2) = \max \left\{ d_2\big(L_1(\epsilon), L_2(\epsilon)\big), c \cdot \max_{a \in A} d\big(\lambda w.L_1(aw), \lambda w.L_2(aw)\big) \right\}.$$

Now because d_2 is the discrete metric with $d_2(0,1) = 1$ we can easily see that $d_{2^{A^*}}$ as defined below is indeed the least fixed-point of this equation and thus $(2^{A^*}, d_{2^{A^*}})$ is the carrier of the final $\overline{M_2}$-coalgebra.

$$d_{2^{A^*}} \colon 2^{A^*} \times 2^{A^*} \to [0,1], \quad d_{2^{A^*}}(L_1, L_2) = c^{\inf\{n \in \mathbb{N} \mid \exists w \in A^n. L_1(w) \neq L_2(w)\}}.$$

Thus the distance between two languages $L_1, L_2 \colon A^* \to 2$ can be determined by looking for a word w of minimal length which is contained in one and not in the other. Then, the distance is computed as $c^{|w|}$.

We already noted in the beginning of this paper that for any set X, the set of pseudo-metrics over X, with pointwise order, is a complete lattice. Moreover, by Theorem 5.2 the lifting $\overline{F}$ induces a monotone function $\_^F$ which maps any pseudometric d on X to d^F on FX. If, additionally, such a function is ω-continuous, i.e., if it preserves the supremum of ω-chains, the least fixpoint metric in Lemma 6.1 can be obtained by a construction that converges at an ordinal less than or equal to ω. It is easy to see that this is the case in Example 6.3. With some more effort one can also show that also the liftings induced by the finite powerset functor and the probability distribution functor with finite support are ω-continuous [BBKK14, Prop. P.6.1].

The lifting of the final F-coalgebra to a final $\overline{F}$-coalgebra, which is provided by Theorem 6.2, allows us to move from a qualitative to a quantitative behavior analysis: Instead of just considering equivalences, in $PMet$ we can now measure the distance of behaviors using the final coalgebra.

Definition 6.4 (Bisimilarity Pseudometric). Let $\overline{F} \colon PMet \to PMet$ be a lifting of a Set-endofunctor F for which a final coalgebra $z \colon Z \to FZ$ exists and $d_z \colon X \times X \to [0, \top]$ be the fixpoint pseudometric of Lemma 6.1. For any F-coalgebra $c \colon X \to FX$, the *bisimilarity distance* on X is the function $bd_c \colon X \times X \to [0, \top]$ where

$$bd_c(x, y) := d_z\big(\llbracket x \rrbracket_c, \llbracket y \rrbracket_c\big)$$

for all $x, y \in X$ with $\llbracket \cdot \rrbracket_c \colon X \to Z$ being the unique map into the final coalgebra. Since d_z is a pseudometric also bd_c is a pseudometric.

Let us check how this definition applies to deterministic automata.

Example 6.5 (Bisimilarity Pseudometric for Deterministic Automata). We instantiate the above definition for the machine functor $M_2 = 2 \times \_^A$ with maximal distance $\top = 1$ and evaluation function $ev_{M_2} \colon [0, 1] \times [0, 1]^A$ with $ev_{M_2}(o, s) = c \cdot \max_{a \in A} s(a)$ for $0 < c < 1$ as in Example 6.3. We recall that for any coalgebra $\alpha \colon X \to 2 \times X^A$ the unique map

$[\![\cdot]\!]_\alpha \colon X \to 2^{A^*}$ into the final coalgebra maps each state $x \in X$ to the language $[\![x]\!]_\alpha \colon A^* \to 2$ it accepts. Using the final coalgebra pseudometric from Example 6.3 we have

$$bd_\alpha \colon X \times X \to [0,1], \quad bd_\alpha(x,y) = c^{\inf\{n \in \mathbb{N} \mid \exists w \in A^n. [\![x]\!]_\alpha(w) \neq [\![y]\!]_\alpha(w)\}}.$$

Thus the distance between two states $x, y \in X$ is determined by the shortest word w which is contained in the language of one state and not in the language of the other. Then the distance is computed as $c^{|w|}$.

The name *bisimilarity* pseudometric is motivated by the fact that states which are mapped to the same element in the final coalgebra are at distance 0.

Lemma 6.6. *Let $\overline{F}$, F, $c \colon X \to FX$, $[\![\cdot]\!]_c$ and bd_c be as in Definition 6.4. For all $x, x' \in X$,*
$$[\![x]\!]_c = [\![x']\!]_c \implies bd_c(x, x') = 0\,.$$

Proof. Since d_z is a pseudometric, it is reflexive and thus $bd_c(x, x') = d_z\big([\![x]\!]_c, [\![y]\!]_c\big) = 0.$ $\square$

If d_z is a metric, also the converse of the above implication holds. We will later (Theorem 6.10, page 39) provide sufficient conditions which guarantee that this is the case.

Before doing so, we show that under some mild conditions the bisimilarity pseudometric can be computed without exploring the entire final coalgebra (which might be too large) but, in a sense, restricting to its relevant part.

Theorem 6.7 (Bisimilarity Pseudometric)**.** *Let $\overline{F} \colon PMet \to PMet$ be a lifting of a functor $F \colon Set \to Set$ such that F has a final coalgebra and assume that $\overline{F}$ preserves isometries. Let furthermore $c \colon X \to FX$ be an arbitrary coalgebra. Then the pseudometric $d_c \colon X \times X \to [0, \top]$ defined in Lemma 6.1 is the bisimilarity pseudometric, i.e., we have $bd_c = d_c$.*

Proof. Let $z \colon Z \to FZ$ be the final coalgebra of F in *Set* and let $[\![\cdot]\!]_c \colon X \to Z$ be the unique function such that $z \circ [\![\cdot]\!]_c = F[\![\cdot]\!]_c \circ c$. By Lemma 6.1, if d_z and d_c are the least fixpoint distances on Z and X respectively, $[\![\cdot]\!]_c \colon (X, d_c) \to (Z, d_z)$ is an isometry. Moreover, by Theorem 6.2, $z \colon (Z, d_z) \to \overline{F}(Z, d_z)$ is the final coalgebra for $\overline{F}$. Therefore $bd_c(x, y) = d_z\big([\![x]\!]_c, [\![y]\!]_c\big) = d_c(x, y).$ $\square$

We now want to find conditions which ensure that the behavioral distance is a proper metric. To this aim, we proceed by recalling the final coalgebra construction via the final chain which was first presented in the dual setting (free/initial algebra).

Definition 6.8 (Final Chain Construction [Adá74])**.** Let $\mathbb{C}$ be a category with terminal object $\mathbb{1}$ and limits of ordinal-indexed cochains. For any functor $F \colon \mathbb{C} \to \mathbb{C}$ the *final chain* consists of objects W_i for all ordinals i and *connection morphisms* $p_{i,j} \colon W_j \to W_i$ for all ordinals $i \le j$. The objects are defined as $W_0 := \mathbb{1}$, $W_{i+1} := FW_i$ for all ordinals i, and $W_j := \lim_{i<j} W_i$ for all limit ordinals j. The morphisms are determined by $p_{0,i} := !_{W_i} \colon W_i \to \mathbb{1}$, $p_{i,i} = \mathrm{id}_{W_i}$ for all ordinals i, $p_{i+1,j+1} := Fp_{i,j}$ for all ordinals $i < j$ and if j is a limit ordinal the $p_{i,j}$ are the morphisms of the limit cone. They satisfy $p_{i,k} = p_{i,j} \circ p_{j,k}$ for all ordinals $i \le j \le k$. We say that the chain *converges* in λ steps if $p_{\lambda,\lambda+1} \colon W_{\lambda+1} \to W_\lambda$ is an isomorphism.

This construction does not necessarily converge (e.g. for the unrestricted powerset functor $\mathcal{P}$ on *Set*), but if it does, we always obtain a final coalgebra.

Theorem 6.9 (Final Coalgebra via the Final Chain [Adá74])**.** *Let $\mathbb{C}$ be a category with terminal object $\mathbb{1}$ and limits of ordinal-indexed cochains. If the final chain of a functor $F\colon \mathbb{C} \to \mathbb{C}$ converges in λ steps then $p_{\lambda,\lambda+1}^{-1}\colon W_\lambda \to FW_\lambda$ is the final coalgebra.*

We now show under which circumstances the least fixpoint pseudometric d_z is a metric and how our construction relates to the construction of the final chain.

Theorem 6.10 (Final Coalgebra Metric)**.** *Let $\overline{F}\colon PMet \to PMet$ be a lifting of a functor $F\colon Set \to Set$ which has a final coalgebra $z\colon Z \to FZ$. Let $d_z\colon X \times X \to [0, \top]$ be the least fixpoint pseudometric in Lemma 6.1. If $\overline{F}$ preserves metrics, and the final chain for F converges then d_z is a metric, i.e., all for $z, z' \in Z$ we have $d_z(z, z') = 0 \iff z = z'$. This implies that for every coalgebra $c\colon X \to FX$ and all $x, x' \in X$ we have*

$$[\![x]\!]_c = [\![x']\!]_c \iff bd_c(x, x') = 0\,.$$

Proof. Let $W_i, p_{i,j}\colon W_j \to W_i$ be as in Definition 6.8. Define a pseudometric over each element of the chain $e_i\colon W_i \times W_i \to [0, \top]$ as follows: $e_0\colon \mathbb{1} \times \mathbb{1} \to [0, \top]$ is the (unique!) zero metric on $\mathbb{1}$, $e_{i+1} := e_i^F\colon W_{i+1} \times W_{i+1} \to [0, \top]$ for all ordinals i and $e_j := \sup_{i<j} e_i \circ (p_{i,j} \times p_{i,j})$ for all limit ordinals j. Since the functor preserves metrics e_{i+1} is a metric if e_i is. Given a limit ordinal j, it follows from Lemma 3.2 that e_j is a pseudometric provided that all the e_i with $i < j$ are pseudometrics. To see that e_j is also a metric when all e_i with $i < j$ are metrics we proceed as follows: Suppose $e_j(x, y) = 0$ for some $x, y \in W_j$, then we know that for all $i < j$ we must have $e_i\big(p_{i,j}(x), p_{i,j}(y)\big) = 0$ and thus $p_{i,j}(x) = p_{i,j}(y)$ because the e_i are metrics. Since the cone $(W_j \overset{p_{i,j}}{\twoheadrightarrow} W_i)_{i<j}$ is by definition a limit in *Set* we can now conclude that $x = y$. This is due to the universal property of the limit: Let us assume $x \neq y$ then for the cone $(\{x, y\} \overset{f_i}{\to} W_i)_{i<j}$ with $f_i(x) := p_{i,j}(x)$, and also $f_i(y) := p_{i,j}(y) = p_{i,j}(x)$ there would have to be a unique function $u\colon \{x, y\} \to W_j$ satisfying $p_{i,j} \circ u = f_i$. However, for example $u, u'\colon \{x, y\} \to W_j$ where $u(x) = u(y) = x$ and $u'(x) = u'(y) = y$ are distinct functions satisfying this commutativity which is a contradiction to the uniqueness. Thus our assumption ($x \neq y$) is false and e_j is indeed a metric.

Consider now for each ordinal i, a function $\eta_i\colon Z \to W_i$ defined as follows. For $i = 0$ we let $\eta_0\colon Z \to W_0$ be the unique mapping to the terminal object. For each ordinal i we define $\eta_{i+1} = F\eta_i \circ z$, and for a limit ordinal j, we let η_j be the unique mapping into the limit $\lim_{i<j} W_i$. Observe that for each ordinal i, it holds that

$$\eta_i = p_{i,i+1} \circ F\eta_i \circ z \tag{6.2}$$

For $i = 0$, $\eta_0 = p_{0,1} \circ F\eta_1 \circ z$ is the unique map into the final object. For any ordinal i, assuming (6.2) we can prove that the same property holds for $i + 1$. In fact $\eta_{i+1} = F\eta_i \circ z = F(p_{i,i+1} \circ F\eta_i \circ z) \circ z = F(p_{i,i+1} \circ \eta_{i+1}) \circ z = Fp_{i,i+1} \circ F\eta_{i+1} \circ z = p_{i+1,i+2} \circ F\eta_{i+1} \circ z$, as desired. Finally, for a limit ordinal j, assume the property (6.2) for $i < j$. Recall that $\eta_j\colon Z \to W_j$ is the mediating arrow into the limit, i.e., for all ordinals $i < j$ we have $\eta_i = p_{i,j} \circ \eta_j$ and thus $F\eta_i = Fp_{i,j} \circ F\eta_j = p_{i+1,j+1} \circ F\eta_j$. Therefore

$$
\begin{aligned}
p_{i,j} &\circ (p_{j,j+1} \circ F\eta_j \circ z) \\
&= p_{i,j+1} \circ F\eta_j \circ z = p_{i,i+1} \circ p_{i+1,j+1} \circ F\eta_j \circ z && \text{[by definition of } p_i, j] \\
&= p_{i,i+1} \circ F\eta_i \circ z && \text{[by the above observation]} \\
&= \eta_i && \text{[by inductive hypothesis]}
\end{aligned}
$$

Hence $p_{j,j+1} \circ F\eta_j \circ z$ is a mediating morphisms into the limit and thus it must coincide with η_j, namely $\eta_j = p_{j,j+1} \circ F\eta_j \circ z$ as desired.

Also observe that each $\eta_i \colon (Z, d_z) \to (W_i, e_i)$ is nonexpansive. This is trivially true for $i = 0$, since e_0 is the zero pseudometric. For any ordinal i, assuming $\eta_i \colon (Z, d_z) \to (W_i, e_i)$ nonexpansive, we obtain that $F\eta_i \colon (FZ, d_z^F) \to (FW_i, e_i^F) = (W_{i+1}, e_{i+1})$ is nonexpansive, hence $\eta_{i+1} = F\eta_i \circ z$ is nonexpansive. For a limit ordinal j,

$$
\begin{aligned}
& e_j \circ (\eta_j \times \eta_j) \\
&\quad = (\sup_{i<j}(e_i \circ (p_{i,j} \times p_{i,j}))) \circ (\eta_j \times \eta_j) && \text{[definition of } e_j] \\
&\quad = \sup_{i<j}(e_i \circ (p_{i,j} \circ \eta_j \times p_{i,j} \circ \eta_j)) && \text{[sup componentwise]} \\
&\quad = \sup_{i<j}(e_i \circ (\eta_i \times \eta_i)) && \text{[since } \eta_j \text{ is the mediating arrow into the limit]} \\
&\quad = \sup_{i<j} d_Z = d_Z && \text{[nonexpansiveness of } \eta_i \colon (Z, d_Z) \to (W_i, e_i)]
\end{aligned}
$$

which means that $\eta_j \colon (Z, d_z) \to (W_j, e_j)$ is nonexpansive.

Now, let λ be the ordinal for which the final chain converges in Set. We know that $p_{\lambda,\lambda+1}^{-1} \colon W_\lambda \to FW_\lambda$ is the final coalgebra in Set. Since also $z \colon Z \to FZ$ is, and by (6.2) $\eta_\lambda \colon Z \to W_\lambda$ is a coalgebra morphism, we deduce that η_λ is an isomorphism of sets. Moreover, we proved that $\eta_\lambda \colon (Z, d_z) \to (W_\lambda, e_\lambda)$ is nonexpansive. Recalling that e_λ is a metric, we conclude that also d_z is so. In fact, if $w, w' \in Z$ are such that $d_z(w, w') = 0$, then $e_\lambda(\eta_\lambda(w), \eta_\lambda(w')) \le d_z(w, w') = 0$. Therefore $\eta_\lambda(w) = \eta_\lambda(w')$, and thus $w = w'$ since η_λ is an iso in Set and thus bijective. Since we know now that d_z is a metric, we also know that $0 = bd_c(x, x') = d_z(\llbracket x \rrbracket_c, \llbracket x' \rrbracket_c)$ holds if and only if $\llbracket x \rrbracket_c = \llbracket x' \rrbracket_c$. $\qquad\square$

We will now get back to the examples studied at the beginning of this paper (Examples 4.1 and 4.3, pages 11 and 12) and discuss in which sense they are instances of our framework.

Example 6.11 (Bisimilarity Pseudometric for Probabilistic Systems)**.** In order to model the discounted behavioral distance for purely probabilistic systems as given in Example 4.1 (page 11) in our framework, we set $\top = 1$ and proceed to lift the following three functors: we first consider the identity functor Id with evaluation map $ev_{\mathrm{Id}} \colon [0,1] \to [0,1]$, $ev_{\mathrm{Id}}(z) = c \cdot z$ in order to integrate the discount (Example 5.29, page 26). Then, we take the coproduct with the singleton metric space (Definition 5.46 and Lemmas 5.47 and 5.48, pp. 33 ff.). The combination of the two functors yields the discrete version of the refusal functor of [vBW06], namely $\overline{R}(X, d) = (X + \mathbb{1}, \hat{d})$ where $\hat{d}$ is the coproduct pseudometric taken from Example 4.1. Finally, we lift the probability distribution functor $\mathcal{D}$ to obtain $\overline{\mathcal{D}}$ (Example 5.9, page 16). All functors satisfy the Kantorovich-Rubinstein duality and preserve metrics.

It is easy to see that $\overline{\mathcal{D}}(\overline{R}(X, d)) = (\mathcal{D}(X + 1), \overline{d})$, where $\overline{d}$ is defined as in Example 4.1). Then, the least solution of $d(x, y) = \overline{d}\big(c(x), c(y)\big)$ can be computed as in Theorem 6.7.

Example 6.12 (Bisimilarity Pseudometric for Metric Transition Systems)**.** As in Example 4.3 we consider metric transition systems as coalgebras $c \colon S \to M_1 \times \cdots \times M_n \times \mathcal{P}_f(S)$ where S is a finite set of states and (M_i, d_i) are pseudometric spaces. If, for $i \in \{1, \dots, n\}$, $\pi_i \colon M_1 \times \cdots \times M_n \times \mathcal{P}_f(S) \to M_i$ and $\pi_{n+1} \colon M_1 \times \cdots \times M_n \times \mathcal{P}_f(S) \to \mathcal{P}_f(S)$ are the projections of the product, then each state $s \in S$ is assigned a valuation function $[s] \colon \{1, \dots, n\} \to \cup_{r=1}^n M_r$ where, of course, $[s](i) = \pi_i\big(c(S)\big)$, and the set $\pi_{n+1}\big(c(S)\big)$ of successor states.

To obtain behavioral distances for metric transition systems using our framework we set $\top = \infty$. Moreover, analogously to the product bifunctor of Definition 5.42 we can equip the product multifunctor $P\colon Set^{n+1} \to Set$ with the evaluation function $ev_P\colon [0,\infty]^{n+1} \to [0,\infty]$ where $ev_P(r_1, \ldots, r_{n+1}) = \max\{r_1, \ldots, r_{n+1}\}$ which is a natural generalization of the function presented in Lemma 5.43. As in that lemma, this function is well-behaved in the sense of Definition 5.39 and analogously to Lemma 5.44 we can easily see that duality holds and we obtain the categorical product pseudometric, i.e., for given pseudometric spaces (X_i, d_i) the new pseudometric $d^P\colon (X_1 \times \cdots \times X_{n+1})^2 \to [0,\infty]$, $\quad d^P = \max\{d_1, \ldots, d_{n+1}\}$. Let $\overline{P}$ be the corresponding lifted multifunctor. We instantiate the given pseudometric spaces (M_i, d_i) as fixed parameters and obtain the endofunctor $\overline{F}\colon PMet \to PMet$ with

$$\overline{F}(X,d) = \overline{P}\big((M_1, d_1), \ldots, (M_n, d_n) \times \overline{\mathcal{P}_f}(X,d)\big)$$

where $\overline{\mathcal{P}_f}$ is the lifting of the powerset functor using the evaluation function $\max\colon \mathcal{P}_f([0,\infty]) \to [0,\infty]$ with $\max \emptyset = 0$ as presented in Example 5.31. Then, via Theorem 6.7, we obtain exactly the least solution of

$$d(s,t) = \max\left\{\overline{pd}([s],[t]), \max_{s'\in\tau(s)} \min_{t'\in\tau(t)} d(s',t'), \max_{t'\in\tau(t)} \min_{s'\in\tau(s)} d(s',t')\right\}$$

as in (4.1) in Example 4.3. Except for the fact that we allow $+\infty$ and consider only undirected (symmetric) pseudometrics, this is exactly the branching distance bd^{Ss} for metric transition systems [dAFS09, Def. 13].

7. Compositionality of Liftings

In the remainder of this paper we want to turn our attention to trace (or linear-time) pseudometrics. Since we plan to apply the generalized powerset construction we will have to lift not only functors but also monads from Set to $PMet$. As a preparation for that we now study *compositionality* of functor liftings, i.e., we set off to identify some sufficient conditions ensuring $\overline{F}\,\overline{G} = \overline{FG}$. Unfortunately, this seems to be a quite difficult question in this general setting so our main result only deals with the Wasserstein lifting and requires the existence of optimal couplings. However, whenever it can be applied it allows us to reason modularly and, consequently, to simplify the proofs needed for the treatment of our examples.

As further preparation for the trace pseudometric we will also consider two examples involving the finite powerset monad where optimal couplings do not always exist and manually prove that compositionality holds for these specific cases.

7.1. **Compositionality for Endofunctors.** Given evaluation functions ev_F and ev_G, we can easily construct an evaluation function for the composition FG as follows.

Definition 7.1 (Composition of Evaluation Functions). Let F and G be endofunctors on Set with evaluation functions ev_F and ev_G. We define the composition of ev_F and ev_G to be the evaluation function $ev_F * ev_G\colon FG[0,\top] \to [0,\top]$ for the composed functor FG as $ev_F * ev_G := \widetilde{F}ev_G = ev_F \circ Fev_G$.

Whenever F and G preserve weak pullbacks well-behavedness is inherited.

Theorem 7.2 (Well-Behavedness of Composed Evaluation Function)**.** *Let F, G be endofunctors on Set with evaluation functions ev_F, ev_G. If both functors preserve weak pullbacks and both evaluation functions are well-behaved then also $ev_F * ev_G$ is well-behaved.*

We will split the proof into two technical lemmas. The first of these just summarizes some useful rules for calculations.

Lemma 7.3. *Let F, G be endofunctors on Set with evaluation functions ev_F, ev_G and $a := \langle G\pi_1, G\pi_2 \rangle$ (i.e., the unique mediating arrow into the product $GX \times GX$, where $\pi_i \colon X^2 \to X$ are the projections) and (X, d) an arbitrary pseudometric space. Then the following holds.*

1. *We have $\widetilde{G}d \geq d^{\downarrow G} \circ a$ and if ev_G satisfies Conditions W1 and W2 we also have $\widetilde{G}d \geq d^{\uparrow G} \circ a$*
2. *$\forall t_1, t_2 \in FGX: \quad t \in \Gamma_{FG}(t_1, t_2) \implies Fa(t) \in \Gamma_F(t_1, t_2)$.*
3. *If F and G preserve weak pullbacks then so does FG.*
4. *For any $f \in Set/[0, \top]$ we have $\widetilde{FG}f = \widetilde{F}(\widetilde{G}f)$.*

Proof. We first of all observe that a is the unique mediating arrow into the product $GX \times GX$ satisfying $\tau_i \circ a = G\pi_i$ for the projections $\pi_i \colon X^2 \to X$ and $\tau_i \colon (GX)^2 \to GX$.

1. Let $s \in G(X \times X)$ and define $s_i := G\pi_i(s) = \tau_i \circ a(s)$. Then by definition $s \in \Gamma_G(s_1, s_2)$ and we conclude $\widetilde{G}d(s) \geq \inf\{\widetilde{G}d(s') \mid s' \in \Gamma_G(s_1, s_2)\} = d^{\downarrow G}(s_1, s_2) = d^{\downarrow G}(\tau_1 \circ a(s), \tau_2 \circ a(s)) = d^{\downarrow G} \circ a(s)$. Since $d^{\downarrow G} \geq d^{\uparrow G}$ as shown in Theorem 5.27, the statement follows.
2. We compute $F\tau_i(Fa(t)) = F(\tau_i \circ a)(t) = F(G\pi_i)(t) = FG\pi_i = t_i$.
3. This is indeed clear by definition.
4. Let $f \colon X \to [0, \top]$, then $\widetilde{FG}f = ev_F * ev_G \circ FGf = ev_F \circ Fev_G \circ FGf = ev_F \circ F(ev_G \circ Gf) = \widetilde{F}(\widetilde{G}f)$. $\qquad\square$

We can now use these calculations to prove the second lemma which already finishes the proof of the inheritance of well-behavedness.

Lemma 7.4. *Let F, G be functors with evaluation functions ev_F, ev_G.*

1. *If $\widetilde{F}$ and $\widetilde{G}$ are monotone (Condition W1), then so is the evaluation functor $\widetilde{FG}$ with respect to the composed evaluation function $ev_F * ev_G$.*
2. *If G preserves weak pullbacks, ev_G is well-behaved and $\widetilde{F}$ is monotone (Condition W1) then $ev_F * ev_G$ satisfies Condition W2 of Definition 5.14.*
3. *If F preserves weak pullbacks and ev_F, ev_G satisfy Condition W3 of well-behavedness, then also $ev_F * ev_G$ satisfies Condition W3 of Definition 5.14.*

Proof. 1. Let $f, g \colon X \to [0, \top]$ with $f \leq g$, then by monotonicity of ev_G we have $\widetilde{G}f \leq \widetilde{G}g$ and using monotonicity of ev_F we get $\widetilde{FG}f = \widetilde{F}(\widetilde{G}f) \leq \widetilde{F}(\widetilde{G}g) = \widetilde{FG}g$.

2. Let $\pi_i \colon X^2 \to X$ be the projections of the product. For $t \in FG([0, \top]^2)$ we define $t_i := FG\pi_i(t) \in FG[0, \top]$. By definition $t \in \Gamma_{FG}(t_1, t_2)$ so Lemma 7.3 (page 42) tells us $Fa(t) \in \Gamma_F(t_1, t_2)$ for $a := \langle G\pi_1, G\pi_2 \rangle$. Moreover, since the evaluation function $ev_G \colon (G[0, \top], d_e^{\uparrow G}) \to ([0, \top], d_e)$ is nonexpansive (by definition of the Kantorovich pseudometric), we can apply Lemma 5.26 (page 25) to obtain the inequality $d_e((ev_F * ev_G)(t_1), ev_F * ev_G(t_2)) = d_e(\widetilde{F}ev_G(t_1), \widetilde{F}ev_G(t_2)) \leq \widetilde{F}d_e^{\uparrow G}(Fa(t)) = \widetilde{F}(d_e^{\uparrow G} \circ a)(t)$. By Lemma 7.3 we have $d_e^{\uparrow G} \circ a \leq \widetilde{G}d_e$ and using monotonicity of $\widetilde{F}$ we can continue our inequality with $\widetilde{F}(d_e^{\uparrow G} \circ a)(t) \leq \widetilde{F}(\widetilde{G}d_e)(t) = \widetilde{FG}d_e(t)$ which concludes the proof.

3. By Lemma 5.17 (page 19) we have to show that the following diagram is a weak pullback.

$$
\begin{array}{ccccc}
FG\{0\} & \xrightarrow{\ \ F!_{G\{0\}}\ \ } & F\{0\} & \xrightarrow{\ \ !_{F\{0\}}\ \ } & \{0\} \\[2pt]
{\scriptstyle FGi}\downarrow & & \downarrow{\scriptstyle Fi} & & \downarrow{\scriptstyle i} \\[2pt]
FG[0,\top] & \xrightarrow{\ \ Fev_G\ \ } & F[0,\top] & \xrightarrow{\ \ ev_F\ \ } & [0,\top]
\end{array}
$$

with $!_{FG\{0\}}$ across the top and $ev_F * ev_G$ across the bottom.

Lemma 5.17 tells us that the right square is a weak pullback and since F preserves weak pullbacks also the left square is. The outer part is thus necessarily a weak pullback again yielding by Lemma 5.17 that $ev_F * ev_G$ satisfies Condition W3. □

This lemma concludes the proof of Theorem 7.2. In the light of this result we know that, whenever we start with weak pullback preserving functors F, G along with well-behaved evaluation functions ev_F, ev_G, the Wasserstein distance for FG (with respect to $ev_F * ev_G$) is always a pseudometric (see Theorem 5.20, page 22) so we can safely talk about *the* Wasserstein lifting of FG and study compositionality.

We will now show that a sufficient criterion for compositionality of the Wasserstein lifting is the existence of optimal couplings for G. Again we start with a technical lemma which also contains a small statement about the Kantorovich lifting.

Lemma 7.5. *Let F and G be endofunctors on Set together with evaluation functions $ev_F\colon F[0,\top] \to [0,\top]$, $ev_G\colon G[0,\top] \to [0,\top]$. We define $ev_F * ev_G := ev_F \circ Fev_G$. Then the following properties hold for every pseudometric space (X,d).*

1. *$d^{\uparrow FG} \le (d^{\uparrow G})^{\uparrow F}$.*
2. *If ev_F and ev_G satisfy Conditions W1 and W2 then $d^{\downarrow FG} \ge (d^{\downarrow G})^{\downarrow F}$.*
3. *If for all $t_1, t_2 \in FGX$ there is a function $\nabla(t_1, t_2)\colon \Gamma_F(t_1, t_2) \to \Gamma_{FG}(t_1, t_2)$ such that $\widetilde{FG}d \circ \nabla(t_1, t_2) = \widetilde{F}d^{\downarrow G}$ then $d^{\downarrow FG} \le (d^{\downarrow G})^{\downarrow F}$.*

Proof. Let $t_1, t_2 \in FGX$.

1. Recall that $d^{\uparrow G}$ is the smallest pseudometric such that for every nonexpansive function $f\colon (X,d) \to ([0,\top], d_e)$ also $\widetilde{G}f\colon (GX, d^{\uparrow G}) \to ([0,\top], d_e)$ is nonexpansive (see remark in the beginning of Section 5.1 on page 14). Moreover, $\widetilde{FG}f = \widetilde{F}(\widetilde{G}f)$ by Lemma 7.3. Thus

$$
\begin{aligned}
d^{\uparrow FG}(t_1, t_2) &= \sup\left\{ d_e\left(\widetilde{FG}f(t_1), \widetilde{FG}f(t_2)\right) \mid f\colon (X,d) \to ([0,\top], d_e) \right\} \\
&= \sup\left\{ d_e\left(\widetilde{F}(\widetilde{G}f)(t_1), \widetilde{F}(\widetilde{G}f)(t_2)\right) \mid f\colon (X,d) \to ([0,\top], d_e) \right\} \\
&\le \sup\left\{ d_e\left(\widetilde{F}(g)(t_1), \widetilde{F}(g)(t_2)\right) \mid g\colon (GX, d^{\uparrow G}) \to ([0,\top], d_e) \right\} = (d^{\uparrow G})^{\uparrow F}(t_1, t_2)
\end{aligned}
$$

2. Lemma 7.3 tells us $\widetilde{G}d \ge d^{\downarrow G} \circ a$ and for any coupling $t \in \Gamma_{FG}(t_1, t_2)$ we have $Fa(t) \in \Gamma_F(t_1, t_2)$. Using these facts and the monotonicity of $\widetilde{F}$ we obtain:

$$
\begin{aligned}
d^{\downarrow FG}(t_1, t_2) &= \inf\left\{ \widetilde{FG}d(t) \mid t \in \Gamma_{FG}(t_1, t_2) \right\} = \inf\left\{ \widetilde{F}(\widetilde{G}d)(t) \mid t \in \Gamma_{FG}(t_1, t_2) \right\} \\
&\ge \inf\left\{ \widetilde{F}(d^{\downarrow G} \circ a)(t) \mid t \in \Gamma_{FG}(t_1, t_2) \right\} = \inf\left\{ \widetilde{F}d^{\downarrow G}(Fa(t)) \mid t \in \Gamma_{FG}(t_1, t_2) \right\} \\
&\ge \inf\left\{ \widetilde{F}d^{\downarrow G}(t') \mid t' \in \Gamma_F(t_1, t_2) \right\} = (d^{\downarrow G})^{\downarrow F}(t_1, t_2)
\end{aligned}
$$

3. Using $\nabla(t_1, t_2)$ we compute

$$d^{\downarrow FG}(t_1, t_2) = \inf\left\{\widetilde{FG}d(t') \,\big|\, t' \in \Gamma_{FG}(t_1, t_2)\right\} \leq \inf\left\{\widetilde{FG}d\big(\nabla(t_1, t_2)(t)\big) \,\big|\, t \in \Gamma_F(t_1, t_2)\right\}$$

$$= \inf\left\{\widetilde{F}d^{\downarrow G}(t) \,\big|\, t \in \Gamma_F(t_1, t_2)\right\} = (d^{\downarrow G})^{\downarrow F}(t_1, t_2)$$

which concludes the proof. $\square$

With this result at hand we can now prove compositionality for the Wasserstein lifting.

Theorem 7.6 (Compositionality of the Wasserstein Lifting)**.** *Let F, G be weak pullback preserving endofunctors on Set with well-behaved evaluation functions ev_F, ev_G and (X, d) be a pseudometric space. If for all $t_1, t_2 \in GX$ there is an optimal G-coupling $\gamma(t_1, t_2) \in \Gamma_G(t_1, t_2)$ such that $d^{\downarrow G}(t_1, t_2) = \widetilde{G}d\big(\gamma(t_1, t_2)\big)$ then we have the equality $d^{\downarrow FG} = (d^{\downarrow G})^{\downarrow F}$.*

Proof. From Lemma 7.5.2 we know $d^{\downarrow FG} \geq (d^{\downarrow G})^{\downarrow F}$. We just have to show the other inequality. By our requirement we have a function $\gamma \colon GX \times GX \to G(X \times X)$, such that $d^{\downarrow G} = \widetilde{G}d \circ \gamma$. Moreover, let $\pi_i \colon X^2 \to X$ and $\tau_i \colon (GX)^2 \to GX$ be the projections of the product then γ satisfies $G\pi_i \circ \gamma = \tau_i$. Given $t_1, t_2 \in FGX$ and $t \in \Gamma_F(t_1, t_2)$, we define $\nabla(t_1, t_2)(t) = F\gamma(t)$, then this satisfies the conditions of Lemma 7.5.3. First, we have $F\gamma(t) \in \Gamma_{FG}(t_1, t_2)$ because $FG\pi_i\big(F\gamma(t)\big) = F(G\pi_i \circ \gamma)(t) = F\tau_i(t) = t_i$. Moreover $\widetilde{FG}d\big(F\gamma(t)\big) = ev_F * ev_G \circ F\big(Gd \circ \gamma(t)\big) = ev_F \circ Fev_G \circ F(Gd \circ \gamma)(t) = ev_F \circ F(\widetilde{G}d \circ \gamma)(t) = ev_F \circ Fd^{\downarrow G}(t) = \widetilde{F}d^{\downarrow G}(t)$ so Lemma 7.5.3 yields $d^{\downarrow FG} \leq (d^{\downarrow G})^{\downarrow F}$. $\square$

This criterion will sometimes turn out to be useful for our later results. Nevertheless it provides just a sufficient condition for compositionality as the following examples show.

Example 7.7 (Compositionality for the Distribution Functor)**.** We consider the distribution functor (with finite support) $\mathcal{D}_f$ with the evaluation function defined in Example 5.9. For any pseudometric space (X, d) we have $d^{\downarrow \mathcal{D}_f \mathcal{D}_f} = \big(d^{\downarrow \mathcal{D}_f}\big)^{\downarrow \mathcal{D}_f}$ by Theorem 7.6 because optimal couplings always exist.

Example 7.8 (Compositionality for the Finite Powerset Functor)**.** We consider the finite powerset functor $\mathcal{P}_f$ with the evaluation function defined in Example 5.16. We claim that for any pseudometric space (X, d) we have $d^{\downarrow \mathcal{P}_f \mathcal{P}_f} = \big(d^{\downarrow \mathcal{P}_f}\big)^{\downarrow \mathcal{P}_f}$ although $\mathcal{P}_f$-couplings do not always exist. To verify this, we recall from Lemma 7.5.2 that

$$d^{\downarrow \mathcal{P}_f \mathcal{P}_f} \geq \big(d^{\downarrow \mathcal{P}_f}\big)^{\downarrow \mathcal{P}_f} \tag{7.1}$$

holds. We now show that we always have equality. Let (X, d) be a pseudometric space and $T_1, T_2 \in \mathcal{P}_f \mathcal{P}_f X$. We distinguish three cases:

1. If $T_1 = T_2 = \emptyset$ we know by reflexivity that both sides of (7.1) are 0.
2. If $T_1 = \emptyset \neq T_2$ or $T_1 \neq \emptyset = T_2$ we know from Example 5.31 that $\Gamma_{\mathcal{P}_f}(T_1, T_2) = \emptyset$ and therefore $\big(d^{\downarrow \mathcal{P}_f}\big)^{\downarrow \mathcal{P}_f}(T_1, T_2) = \top$ and thus (7.1) is necessarily an equality because the left hand side can never exceed $\top$.
3. Let $T_1, T_2 \neq \emptyset$. We know from Example 5.31 that we have an optimal coupling $T^* \in \Gamma_{\mathcal{P}_f}(T_1, T_2)$, say $T^* = \{(V_{j1}, V_{j2}) \in \mathcal{P}_f X \times \mathcal{P}_f X \mid j \in J\}$ for a suitable index set J. Then, using the projections $\pi_i \colon (\mathcal{P}_f X)^2 \to \mathcal{P}_f X$, we have $T_i = \mathcal{P}_f \pi_i(T^*) = \pi_i[T^*] =$

$\{\pi_i((V_{j1}, V_{j2})) \mid j \in J\} = \{V_{ji} \mid j \in J\}$. By optimality we thus have:

$$\left(d^{\downarrow \mathcal{P}_f}\right)^{\downarrow \mathcal{P}_f} (T_1, T_2) = \widetilde{\mathcal{P}_f} d^{\downarrow \mathcal{P}_f}(T^*) = \max d^{\downarrow \mathcal{P}_f}[T^*] = \max_{j \in J} d^{\downarrow \mathcal{P}_f}(V_{j1}, V_{j2}). \qquad (7.2)$$

We will make another case distinction:
1. If there is an index $j' \in J$ such that $\Gamma_{\mathcal{P}_f}(V_{j'1}, V_{j'2}) = \emptyset$, we have $d^{\downarrow \mathcal{P}_f}(V_{j'1}, V_{j'2}) = \top$
   and by (7.2) also $\left(d^{\downarrow \mathcal{P}_f}\right)^{\downarrow \mathcal{P}_f}(T_1, T_2) = \top$ which again shows that (7.1) is an equality.
2. Otherwise we can take optimal couplings $V_j^* \in \Gamma_{\mathcal{P}_f}(V_{j1}, V_{j2})$ (see Example 5.31, page 26). Continuing (7.2) we have

$$\left(d^{\downarrow \mathcal{P}_f}\right)^{\downarrow \mathcal{P}_f} (T_1, T_2) = \max_{j \in J} \widetilde{\mathcal{P}_f} d(V_j^*) = \max_{j \in J} \max d[V_j^*] \qquad (7.3)$$

We define $T := \left\{V_j^* \mid j \in J\right\} \in \mathcal{P}_f \mathcal{P}_f(X \times X)$ and calculate for the projections $\pi_i \colon X \times X \to X$

$$\mathcal{P}_f \mathcal{P}_f \pi_i(T) = \mathcal{P}_f \pi_i[T] = \left\{\mathcal{P}_f \pi_i(V_j^*) \mid j \in J\right\} = \{V_{ji} \mid j \in J\} = T_i$$

and thus $T \in \Gamma_{\mathcal{P}_f \mathcal{P}_f}(T_1, T_2)$. Moreover we have

$$\begin{aligned}
d^{\downarrow \mathcal{P}_f \mathcal{P}_f}(T_1, T_2) &\leq \widetilde{\mathcal{P}_f \mathcal{P}_f} d(T) = \max \left((\mathcal{P}_f \max)(\mathcal{P}_f \mathcal{P}_f d(T)) \right) \\
&= \max \left(\max \left[(\mathcal{P}_f d)[T] \right] \right) = \max \left(\max \left[\{d[V_j^*] \mid j \in J\} \right] \right) \\
&= \max \left(\{\max d[V_j^*] \mid j \in J\} \right) = \max_{j \in J} \max d[V_j^*]. \qquad (7.4)
\end{aligned}$$

Thus using this, (7.3) and (7.1) we conclude that

$$d^{\downarrow \mathcal{P}_f \mathcal{P}_f}(T_1, T_2) \leq \max_{j \in J} \max d[V_j^*] = \left(d^{\downarrow \mathcal{P}_f}\right)^{\downarrow \mathcal{P}_f} (T_1, T_2) \leq d^{\downarrow \mathcal{P}_f \mathcal{P}_f}(T_1, T_2)$$

which proves equality also in this last case.

We conclude our study of compositionality for endofunctors with another example for which one again has to show compositionality separately. This result will later turn out to be helpful to obtain trace pseudometrics for nondeterministic automata. The proof follows closely the approach used in the previous example, hence it is omitted.

Example 7.9. As in Example 5.33 (page 28) we equip the machine functor with the evaluation function $ev_{M_2} \colon 2 \times [0, 1]^A \to [0, 1]$, $(o, s) \mapsto c \cdot ev_I(s)$ where $c \in\,]0, 1]$ and ev_I is one of the evaluation functions for the input functor from Example 5.32. Moreover, for the powerset functor we use the maximum as evaluation function (see Example 5.16).

   Although couplings for M_2 do not always exist, one simply can adapt [Ker16, Ex. 5.6.9] the approach employed in Example 7.8 to show that we have $d^{\downarrow \mathcal{P}_f M_2} = \left(d^{\downarrow M_2}\right)^{\downarrow \mathcal{P}_f}$.

7.2. **Compositionality for Multifunctors.** We conclude the analysis of compositionality with a short explanation on how our theory extends to multifunctors.

   For $n \in \mathbb{N}$ we denote by $[n] := \{1, \ldots, n\} \subseteq \mathbb{N}$ the set of all positive natural numbers less than or equal to n. Now let $n_i \in \mathbb{N}$ for all $i \in [n]$ and $F \colon Set^n \to Set$ and $G_i \colon Set^{n_i} \to Set$

(for $i \in [n]$) be multifunctors with evaluation functions $ev_F\colon F([0, \top]^n) \to [0, \top]$ and $ev_{G_i}\colon G_i([0, \top]^{n_i}) \to [0, \top]$. We define $N := \sum_{i=1}^n n_i$ and define the functor

$$H := F \circ \prod_{i=1}^n G_i = F \circ (G_1 \times G_2 \times \cdots \times G_n)\colon Set^N \to Set$$

Then we can define the evaluation function $ev_H\colon H([0, \top]^N) \to [0, \top]$ by

$$ev_H := ev_F \circ F(ev_{G_1}, ev_{G_2}, \ldots, ev_{G_n}).$$

In this setting, compositionality of the lifting means that whenever we have N pseudometric spaces (X_i, d_i) the pseudometric $(d_1, \ldots, d_N)^H$ is equal to

$$\left((d_1, \ldots, d_{n_1})^{G_1}, (d_{n_1+1}, \ldots, d_{n_1+n_2})^{G_2} \ldots, (d_{N-n_n+1}, \ldots, d_N)^{G_n} \right)^F.$$

In the upcoming examples we will just use the Wasserstein lifting and we only have the following two cases:

1. $n = 1$, $n_1 = 2$ so that $F\colon Set \to Set$ is an endofunctor with evaluation function $ev_F\colon F[0, \top] \to [0, \top]$ and $G\colon Set^2 \to Set$ is a bifunctor with evaluation function $ev_G\colon G([0, \top], [0, \top]) \to [0, \top]$. Then we have $N = n_1 = 2$ and obtain the bifunctor $H = F \circ G\colon Set^2 \to Set$ with evaluation $ev_H = ev_F \circ Fev_G\colon FG([0, 1], [0, 1]) \to [0, 1]$. Compositionality means that for an two pseudometric spaces (X_1, d_1), (X_2, d_2) we have $(d_1, d_2)^{\downarrow H} = ((d_1, d_2)^{\downarrow G})^{\downarrow F}$.

2. $n = 2$, $n_1 = n_2 = 1$ so that $F\colon Set^2 \to Set$ is a bifunctor with evaluation function $ev_F\colon F([0, \top], [0, \top]) \to [0, \top]$ and $G_1, G_2\colon Set \to Set$ are endofunctors with evaluations $ev_{G_i}\colon G_i[0, \top] \to [0, \top]$. Then we have $N = n_1 + n_2 = 1 + 1 = 2$ and obtain the bifunctor $H = F \circ (G_1 \times G_2)\colon Set^2 \to Set$ with evaluation $ev_H = ev_F \circ F(ev_{G_1}, ev_{G_2})\colon F(G_1[0, \top], G_2[0, \top]) \to [0, \top]$. Compositionality means that for an two pseudometric spaces (X_1, d_1), (X_2, d_2) we have $(d_1, d_2)^{\downarrow H} = (d_1^{\downarrow G}, d_2^{\downarrow G_2})^{\downarrow F}$.

The results presented for endofunctors work analogously in the multifunctor case (the proofs can be transferred almost verbatim), so we do not explicitly present them here. Instead, we will use compositionality to obtain the machine bifunctor.

Example 7.10 (Machine Bifunctor). Let $I = \_^A$ be the input functor, Id the identity endofunctor on Set and P be the product bifunctor of Definition 5.42. The *machine bifunctor* is the composition $M := P \circ (\text{Id} \times I)$, i.e., the bifunctor $M\colon Set^2 \to Set$ with $M(B, X) := B \times X^A$. We compute the composed evaluation function which, of course, depends on the evaluation functions for P and I (for Id we always take $\text{id}_{[0, \top]}$). Let $(o, s) \in [0, \top] \times [0, \top]^A$, then $ev_M(o, s) = ev_P \circ P(\text{id}_{[0, \top]}, ev_I)(o, s) = ev_P \circ (\text{id}_{[0, \top]} \times ev_I)(o, s) = ev_P(o, ev_I(s))$. By instantiating ev_P and ev_I as in the table below (see also Example 5.32 and Lemma 5.44), we obtain the corresponding evaluation functions $ev_M\colon [0, \top] \times [0, \top]^A \to [0, \top]$. They are well-behaved since all involved functors preserve weak pullbacks and for Id and I there are unique (thus optimal) couplings so we have compositionality by a multifunctor equivalent to Theorem 7.2. Again we introduce constants c_1, c_2 as weighting factors and we impose certain requirements to ensure well-behavedness. In particular we have to require that the distance does not exceed the upper bound $\top$.

Parameters	$ev_P(r_1, r_2)$	$ev_I(s)$	$ev_M(o, s)$
$c_1, c_2 \in {]0, 1]}$	$\max\{c_1 r_1, c_2 r_2\}$	$\max_{a \in A} s(a)$	$\max\{c_1 o, c_2 \max_{a \in A} s(a)\}$
$c_1, c_2 \in {]0, 1]}$, $c_1 + c_2 \leq 1$	$c_1 x_1 + c_2 x_2$	$\|A\|^{-1} \sum_{a \in A} s(a)$	$c_1 o + c_2 \|A\|^{-1} \sum_{a \in A} s(a)$
$c_1, c_2 \in {]0, \infty[}$, $\top = \infty$	$c_1 x_1 + c_2 x_2$	$\sum_{a \in A} s(a)$	$c_1 o + c_2 \sum_{a \in A} s(a)$

Now let $(B, d_B), (X, d)$ be pseudometric spaces. For any $t_1, t_2 \in M(B, X)$ with $t_i = (b_i, s_i) \in B \times X^A$ the unique and therefore necessarily optimal coupling is $t := (b_1, b_2, \langle s_1, s_2 \rangle)$. We compute the Wasserstein distance

$$(d_B, d)^{\downarrow M}(t_1, t_2) = \widetilde{M}(d_B, d)(t) = ev_M \circ M(d_B, d)(t)$$
$$= ev_M \circ (d_B \times d^A)(b_1, b_2, \langle s_1, s_2 \rangle) = ev_M(d_B(b_1, b_2), d \circ \langle s_1, s_2 \rangle).$$

We obtain in the first case $(d_B, d)^{\downarrow M}(t_1, t_2) = \max\{c_1 d_B(b_1, b_2), c_2 \cdot \max_{a \in A} d(s_1(a), s_2(a))\}$, in the second case $(d_B, d)^{\downarrow M}(t_1, t_2) = c_1 d_B(b_1, b_2) + c_2 |A|^{-1} \sum_{a \in A} d(s_1(a), s_2(a))$ and in the third case $(d_B, d)^{\downarrow M}(t_1, t_2) = c_1 d_B(b_1, b_2) + c_2 \sum_{a \in A} d(s_1(a), s_2(a))$. Of course one has to choose which of these pseudometrics fits into the respective context. While the first one selects either the distance of the output values or the maximal distance of the successors and neglects the other one, the latter two accumulate the distances. Depending on our maximal element, we have to make sure that we stay within the selected measuring interval $[0, \top]$ by proper scaling of the values.

Usually we will fix the first argument of the machine bifunctor (the set of outputs) of the machine bifunctor and just consider the machine endofunctor $M_B := M(B, \_)$ as in Examples 7.9 and 5.33 (pages 28 and 45). However, for the same reasons as explained before for the product bifunctor, we often need to lift it as a bifunctor and then fix the first component of the lifted bifunctor. One notable exception is the case where B is endowed with the discrete metric. Then it is easy to show that the endofunctor lifting and the bifunctor lifting coincide, i.e., for all pseudometric spaces (X, d) we have $(d_B, d)^{\downarrow M} = d^{\downarrow M_B}$ [Ker16, Ex. 5.6.11].

Let us finish our short excursion to the theory of multifunctor compositionality with another example which shows how the machine bifunctor lifting helps to obtain suitable bisimilarity pseudometrics.

Example 7.11 (Bisimilarity Pseudometric for Automata with Real Outputs)**.** We consider the machine endofunctor with output set $[0, 1]$, i.e., the functor $M_{[0,1]} = [0, 1] \times \_^A$ which arises out of the machine bifunctor M by fixing the first component to $[0, 1]$. As maximal distance we set $\top = 1$ and equip the machine bifunctor M with the evaluation function $ev_M \colon [0, 1] \times [0, 1]^A \to [0, 1]$ where $ev_M(o, s) = c_1 o + c_2 |A|^{-1} \sum_{a \in A} s(a)$ for $c_1, c_2 \in {]0, 1[}$ such that $c_1 + c_2 \leq 1$ as in Example 7.10. Moreover, we recall that the carrier of the final $M_{[0,1]}$-coalgebra is $[0, 1]^{A^*}$ [JSS15, Lem. 3].

If we equip $[0, 1]$ with the Euclidean metric d_e and use our knowledge from Example 7.10 we know that for any pseudometric d on $[0, 1]^{A^*}$ we obtain as Wasserstein pseudometric the function $(d_e, d)^{\downarrow F} \colon ([0, 1] \times ([0, 1]^{A^*})^A)^2 \to [0, 1]$ with

$$(d_e, d)^{\downarrow F}((r_1, s_1), (r_2, s_2)) = c_1 |r_1 - r_2| + \frac{c_2}{|A|} \cdot \sum_{a \in A} d(s_1(a), s_2(a)).$$

We now want to obtain the final coalgebra for the endofunctor $\overline{M}_{([0,1],d_e)} = \overline{M}\big(([0,1],d_e),\_\big)$ on $PMet$ which is a lifting of $M_{[0,1]}$. For this we use the fixed-point equation induced by Lemma 6.1 (page 35). It is given by, for $p_1, p_2 \in [0,1]^{A^*}$, the equation

$$d(p_1, p_2) = c_1 |p_1(\epsilon) - p_2(\epsilon)| + \frac{c_2}{|A|} \cdot \sum_{a \in A} d\big(\lambda w.p_1(aw), \lambda w.p_2(aw)\big).$$

As in the previous example a simple calculation shows that the function

$$d_{[0,1]^{A^*}} \colon [0,1]^{A^*} \times [0,1]^{A^*} \to [0,1], \quad d_{[0,1]^{A^*}}(p_1, p_2) = c_1 \cdot \sum_{w \in A^*} \left(\frac{c_2}{|A|}\right)^{|w|} |p_1(w) - p_2(w)|$$

is the least fixed-point of this equation so if we equip $[0,1]^{A^*}$ with this pseudometric we obtain the final $\overline{M}_{([0,1],d_e)}$-coalgebra. Thus for a probabilistic automaton $\alpha \colon X \to [0,1] \times X^A$ the bisimilarity pseudometric as given in Definition 6.4 (page 37) is the function

$$bd_\alpha \colon X \times X \to [0,1], \quad bd_\alpha(x,y) = c_1 \cdot \sum_{w \in A^*} \left(\frac{c_2}{|A|}\right)^{|w|} |[\![x]\!]_\alpha(w) - [\![y]\!]_\alpha(w)|$$

where the unique map into the final coalgebra $[\![\cdot]\!]_\alpha \colon X \to [0,1]^{A^*}$ maps each state to the function describing the output value of the automaton for each finite word when starting from the respective state.

8. Lifting Natural Transformations and Monads

If we have a monad on Set, we can of course use our framework to lift the endofunctor T to a functor $\overline{T}$ on pseudometric spaces. A natural question that arises is, whether we also obtain a monad on pseudometric spaces, i.e., if the components of the unit and the multiplication are nonexpansive with respect to the lifted pseudometrics. In order to answer this question, we first take a closer look at sufficient conditions for lifting natural transformations.

Theorem 8.1 (Lifting of a Natural Transformation). *Let F, G be endofunctors on Set with evaluation functions ev_F, ev_G and $\lambda \colon F \Rightarrow G$ be a natural transformation. The following two properties hold for the Kantorovich lifting.*

1. *If $ev_G \circ \lambda_{[0,\top]} \leq ev_F$ then λ_X is nonexpansive for all pseudometric spaces (X,d), i.e., $d^{\uparrow G} \circ (\lambda_X \times \lambda_X) \leq d^{\uparrow F}$.*
2. *If $ev_G \circ \lambda_{[0,\top]} = ev_F$ then λ_X is an isometry for all pseudometric spaces (X,d), i.e., $d^{\uparrow G} \circ (\lambda_X \times \lambda_X) = d^{\uparrow F}$.*

Moreover, similar properties hold for the Wasserstein lifting.

3. *If $ev_G \circ \lambda_{[0,\top]} \leq ev_F$ then λ_X is nonexpansive for all pseudometric space (X,d), i.e., $d^{\downarrow G} \circ (\lambda_X \times \lambda_X) \leq d^{\downarrow F}$.*
4. *If $ev_G \circ \lambda_{[0,\top]} = ev_F$ and the Kantorovich-Rubinstein duality holds for F, i.e., $d^{\uparrow F} = d^{\downarrow F}$, then λ_X is an isometry for all pseudometric spaces, i.e., $d^{\downarrow G} \circ (\lambda_X \times \lambda_X) = d^{\downarrow F}$.*

Proof. Let $t_1, t_2 \in FX$.

1. By naturality of λ and $ev_G \circ \lambda_{[0,\top]} \leq ev_F$ we obtain for every $f\colon X \to [0,\top]$ the equality $\widetilde{G}f \circ \lambda_X = ev_G \circ Gf \circ \lambda_X = ev_G \circ \lambda_{[0,\top]} \circ Ff \leq ev_F \circ Ff = \widetilde{F}f$. Using this we compute $d^{\uparrow G}\big(\lambda_X(t_1), \lambda_X(t_2)\big) = \sup\{d_e\big(\widetilde{G}f\big(\lambda_X(t_1)\big), \widetilde{G}f\big(\lambda_X(t_2)\big)\big) \mid f\colon (X,d) \to ([0,\top],d_e)\} \leq \sup\{d_e\big(\widetilde{F}f(t_1), \widetilde{F}f(t_2)\big) \mid f\colon (X,d) \to ([0,\top],d_e)\} = d^{\uparrow F}(t_1,t_2)$.
2. We just have to replace the inequalities in the previous part by equalities.
3. Naturality of λ yields the equations $\lambda_X \circ F\pi_i = G\pi_i \circ \lambda_{X\times X}$ and $\lambda_{[0,\top]} \circ Fd = Gd \circ \lambda_{X\times X}$ where $\pi_i\colon X \times X \to X$ are the projections of the product and $d\colon X \times X \to [0,\top]$ is a pseudometric on X. Using the first equality we can see that $\lambda_{X\times X}$ maps every coupling $t \in \Gamma_F(t_1,t_2)$ to a coupling $\lambda_{X\times X}(t) \in \Gamma_G\big(\lambda_X(t_1), \lambda_X(t_2)\big)$ because $G\pi_i\big(\lambda_{X\times X}(t)\big) = \lambda_X(F\pi_i(t)) = \lambda_X(t_i)$. Moreover, we can use our requirement $(ev_G \circ \lambda_{[0,\top]} \leq ev_F)$ and the second equality to obtain $\widetilde{G}d\big(\lambda_{X\times X}(t)\big) = ev_G \circ Gd \circ \lambda_{X\times X}(t) = ev_G \circ \lambda_{[0,\top]} \circ Fd(t) \leq ev_F \circ Fd(t) = \widetilde{F}d(t)$. With these preparations at hand we can conclude that $d^{\downarrow G}\big(\lambda_X(t_1), \lambda_X(t_2)\big) = \inf\{\widetilde{G}d(t') \mid t' \in \Gamma_G\big(\lambda_X(t_1), \lambda_X(t_2)\big)\} \leq \inf\{\widetilde{G}d\big(\lambda_{X\times X}(t)\big) \mid t \in \Gamma_F(t_1,t_2)\} \leq \inf\{\widetilde{F}d(t) \mid t \in \Gamma_F(t_1,t_2)\} = d^{\downarrow F}(t_1,t_2)$.
4. Using the previous two results and the fact that Wasserstein is an upper bound yields $d^{\uparrow F} = d^{\uparrow G} \circ (\lambda_X \times \lambda_X) \leq d^{\downarrow G} \circ (\lambda_X \times \lambda_X) \leq d^{\downarrow F}$ and since $d^{\uparrow F} = d^{\downarrow F}$ all these inequalities are equalities. $\qquad\square$

In the remainder of this paper we will call a natural transformation λ nonexpansive [an isometry] if (and only if) each of its components are nonexpansive [isometries] and write $\overline{\lambda}$ for the resulting natural transformation from $\overline{F}$ to $\overline{G}$. Instead of checking nonexpansiveness separately for each component of a natural transformation, we can just check the above (in-)equalities involving the two evaluation functions.

By applying these conditions on the unit and multiplication of a given monad, we can now provide sufficient criteria for a monad lifting.

Corollary 8.2 (Lifting of a Monad)**.** *Let (T, η, μ) be a Set-monad and ev_T an evaluation function for T. Then the following holds.*

1. *If $ev_T \circ \eta_{[0,\top]} \leq \mathrm{id}_{[0,\top]}$ then η is nonexpansive for both liftings. Hence we obtain the unit $\overline{\eta}\colon \overline{\mathrm{Id}} \Rightarrow \overline{T}$ in $PMet$.*
2. *If $ev_T \circ \eta_{[0,\top]} = \mathrm{id}_{[0,\top]}$ then η is an isometry for both liftings.*
3. *Let $d^T \in \big\{d^{\uparrow T}, d^{\downarrow T}\big\}$. If $ev_T \circ \mu_{[0,\top]} \leq ev_T \circ Tev_T$ and compositionality holds for TT, i.e., $(d^T)^T = d^{TT}$, then μ is nonexpansive, i.e., $d^T \circ (\mu_X \times \mu_X) \leq (d^T)^T$. This yields the multiplication $\overline{\mu}\colon \overline{T}\,\overline{T} \Rightarrow \overline{T}$ in $PMet$.*

Proof. This is an immediate consequence of Theorem 8.1. For the unit take $F = \mathrm{Id}$ with evaluation function $ev_F = \mathrm{id}_{[0,\top]}$, hence $d^{\uparrow F} = d^{\downarrow F} = d$ and $G = T$, $ev_G = ev_T$, $\lambda = \eta\colon \mathrm{Id} \Rightarrow T$. For the multiplication take $F = TT$, $G = T$, $ev_F = ev_{TT} = ev_T \circ Tev_T$, $ev_G = ev_T$ and $\lambda = \mu$. $\qquad\square$

We conclude this section with two examples of liftable monads.

Example 8.3 (Lifting of the Finite Powerset Monad)**.** We recall that the finite powerset functor $\mathcal{P}_f$ is part of a monad with unit η consisting of the functions $\eta_X\colon X \to \mathcal{P}_f X$, $\eta_X(x) = \{x\}$ and multiplication given by $\mu_X\colon \mathcal{P}_f \mathcal{P}_f X \to \mathcal{P}_f X$, $\mu_X(S) = \cup S$. We check if our conditions for the Wasserstein lifting are satisfied. Given $r \in [0,\infty]$ we have $ev_T \circ \eta_{[0,\infty]}(r) = \max\{r\} = r$ and for $\mathcal{S} \in \mathcal{P}_f(\mathcal{P}_f[0,\top])$ we have $ev_T \circ \mu_{[0,1]}(\mathcal{S}) = \max \cup \mathcal{S} = \max \cup_{S \in \mathcal{S}} S$ and

$ev_T \circ Tev_T(\mathcal{S}) = \max\left(ev_T[\mathcal{S}]\right) = \max\left\{\max S \mid S \in \mathcal{S}\right\}$ and thus it is easy to see that both values coincide. Moreover, we recall from Example 7.8 that we have compositionality for $\mathcal{P}_f\mathcal{P}_f$. Therefore, by Corollary 8.2 η is an isometry and μ nonexpansive.

Example 8.4 (Lifting of the Distribution Monad With Finite Support)**.** It is known that the probability distribution functor $\mathcal{D}_f$ is part of a monad: the unit η consists of the functions $\eta_X\colon X \to \mathcal{D}_f X$, $\eta_X(x) = \delta_x^X$ where δ_x^X is the Dirac distribution and the multiplication is given by $\mu_X\colon \mathcal{D}_f\mathcal{D}_f X \to \mathcal{D}_f X$, $\mu_X(P) = \lambda x.\sum_{q\in\mathcal{D}_f X} P(q)\cdot q(x)$. We consider its Wasserstein lifting. Since $[0,1] = \mathcal{D}_f 2$ we can see that $ev_{\mathcal{D}_f} = \mu_2$. Using this and the monad laws we have $ev_{\mathcal{D}_f} \circ \eta_{[0,1]} = \mu_2 \circ \eta_{\mathcal{D}_f 2} = \mathrm{id}_{\mathcal{D}_f X} = \mathrm{id}_{[0,1]}$ and also $ev_{\mathcal{D}_f} \circ \mu_{[0,1]} = \mu_2 \circ \mu_{\mathcal{D}_f 2} = \mu_2 \circ \mathcal{D}_f \mu_2 = ev_{\mathcal{D}_f} \circ \mathcal{D}_f ev_{\mathcal{D}_f}$. Moreover, since we always have optimal couplings, we have compositionality for $\mathcal{D}_f\mathcal{D}_f$ by Theorem 7.6. Thus by Corollary 8.2 η is an isometry and μ nonexpansive.

9. Trace Pseudometrics

Combining all our previous results we now want to use the generalized powerset construction [SBBR13] on $PMet$ instead of Set to obtain trace pseudometrics. The basic setup is summarized in the (non-commutative) diagram in Figure 4. We quickly recall that in the usual, qualitative setting (bottom part of Figure 4) we have to start with a coalgebra $c\colon X \to FTX$ where F is an endofunctor with final coalgebra $z\colon Z \to FZ$ and (T,η,μ) is a monad on Set. Using a distributive law $\lambda\colon TF \Rightarrow FT$ we can then consider the determinization $c^\sharp$ of c which is defined as

$$c^\sharp := \left(TX \xrightarrow{Tc} TFTX \xrightarrow{\lambda_{TX}} FTTX \xrightarrow{F\mu_X} FTX\right).$$

Figure 4: Trace pseudometrics via the generalized powerset construction

With this determinized coalgebra we define two states $x,y \in X$ of the original coalgebra c to be trace equivalent if and only if $[\![\eta_X(x)]\!]_c = [\![\eta_X(y)]\!]_c$ holds.

The underlying reason why this technique works is that the distributive law λ yields a unique lifting (with respect to the canonical Eilenberg-Moore adjunction $L^T \dashv U^T$ of the functor F to a functor $\widehat{F}$ on the Eilenberg-Moore category $\mathcal{EM}(T)$. The determinization of a coalgebra is nothing but the application of another lifting

$$L\colon CoAlg\left(FT\right) \to CoAlg\left(\widehat{F}\right)$$

of the free algebra functor L^T, i.e., $c^\sharp = U^T L(c)\colon TX \to FTX$ is the F-coalgebra underlying the $\widehat{F}$-coalgebra $L(c)\colon \mu_X \to \widehat{F}\mu_X$.

In order to move to a quantitative setting (upper part of Figure 4) we need to require that both the functor F and the monad (T,η,μ) can be lifted. Then clearly any FT-coalgebra $c\colon X \to FTX$ can be regarded as an $\overline{F}\,\overline{T}$-coalgebra $\overline{c}$ by equipping the state space X with the discrete metric d assigning $\top$ to non equal states (in this way, $\overline{c}$ is trivially nonexpansive). Moreover, if we can ensure that the $\mathcal{EM}$ law λ is nonexpansive, thus yielding a distributive law $\overline{\lambda}\colon \overline{T}\,\overline{F} \Rightarrow \overline{F}\,\overline{T}$, we can use exactly the same ideas as before. In particular, we can lift
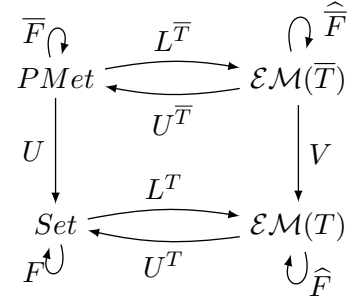
the lifted functor $\overline{F}$ to a functor $\widehat{\overline{F}}\colon \mathcal{EM}(\overline{T}) \to \mathcal{EM}(\overline{T})$ on the Eilenberg-Moore category of the lifted monad. With this, we can lift the free algebra functor $L^{\overline{T}}$ to a functor

$$L'\colon CoAlg\left(\overline{F}\,\overline{T}\right) \to CoAlg\left(\widehat{\overline{F}}\right)$$

which allows us to determinize $\overline{c}$ (as $\overline{F}\,\overline{T}$-coalgebra in $PMet$) to the $\widehat{\overline{F}}$-coalgebra $L'\overline{c}\colon \overline{\mu}_X \to \widehat{\overline{F}}\overline{\mu}_X$ which is given by the underlying $\overline{F}$-coalgebra

$$\overline{c}^{\sharp} := U^T L'\overline{c} = \overline{F}\,\overline{\mu_X} \circ \overline{\lambda_{TX}} \circ \overline{T}\,\overline{c}\colon \overline{T}(X,d) \to \overline{F}\,\overline{T}(X,d)\,. \tag{9.1}$$

If we now equip TX with the behavioral pseudometric $bd_{c^{\sharp}}\colon (TX)^2 \to [0,\top]$ as in Definition 6.4, we can define the trace pseudometric on X via the unit η as follows.

Definition 9.1 (Trace Pseudometric)**.** Let F be an endofunctor and (T,η,μ) be a monad on Set. If F has a final coalgebra $z\colon Z \to FZ$ in Set, F has a lifting $\overline{F}\colon PMet \to PMet$, (T,η,μ) has a lifting $(\overline{T},\overline{\eta},\overline{\mu})$, and there is an $\mathcal{EM}$-law $\lambda\colon TF \to FT$ which can be lifted to an $\mathcal{EM}$-law $\overline{\lambda}\colon \overline{T}\,\overline{F} \to \overline{F}\,\overline{T}$ then for any coalgebra $c\colon X \to FTX$ we define the *trace pseudometric* to be

$$td_c := bd_{c^{\sharp}} \circ (\eta_X \times \eta_X)\colon X \times X \to [0,\top]$$

where $c^{\sharp} = F\mu_X \circ \lambda_{TX} \circ Tc\colon TX \to FTX$ is the determinization of $c\colon X \to FTX$ and $bd_{c^{\sharp}}\colon (TX)^2 \to [0,\top]$ is the corresponding bisimilarity pseudometric (Definition 6.4, page 37).

In order to apply this definition to our two main examples (nondeterministic and probabilistic automata) the only missing thing is the lifting of the $\mathcal{EM}$-law to $PMet$. We note that Theorem 8.1 (page 48) not only provides sufficient conditions for monad liftings but also can be exploited to lift distributive laws. The additional commutativity requirements for $\mathcal{K}\ell$-laws or $\mathcal{EM}$-laws trivially hold when all components are nonexpansive. For the Wasserstein lifting it suffices to require compositionality on the left hand side of the law and to check one inequality.

Corollary 9.2 (Lifting of a Distributive Law)**.** *Let F,G be weak pullback preserving endofunctors on Set with well-behaved evaluation functions ev_F, ev_G and $\lambda\colon FG \Rightarrow GF$ be a distributive law. If the evaluation functions satisfy $(ev_G * ev_F) \circ \lambda_{[0,\top]} \leq ev_F * ev_G$ and compositionality holds for FG, then λ is nonexpansive for the Wasserstein lifting and hence $\overline{\lambda}\colon \overline{F}\,\overline{G} \Rightarrow \overline{G}\,\overline{F}$ is also a distributive law.*

Proof. We use the evaluation function $ev_F * ev_G$ for FG and for GF the evaluation function $ev_G * ev_F$. By Theorem 8.1 (page 48) we know that $d^{\downarrow GF} \circ (\lambda_X \times \lambda_X) \leq d^{\downarrow FG}$ and by Lemma 7.5.2 (page 43) we have $(d^{\downarrow F})^{\downarrow G} \leq d^{\downarrow GF}$. Plugging everything together we conclude that for every pseudometric space (X,d) we have

$$(d^{\downarrow F})^{\downarrow G} \circ (\lambda_X \times \lambda_X) \leq d^{\downarrow GF} \circ (\lambda_X \times \lambda_X) \leq d^{\downarrow FG} = (d^{\downarrow G})^{\downarrow F}$$

which is the desired nonexpansiveness of λ_X. $\qquad\square$

In the remainder of this section we will consider two examples where in both cases G is the machine endofunctor $M_B = B \times \_^A$ (for $B = 2$ and $B = [0,1]$). It is well-known [JSS15, Lem. 3] that for every output set B the final coalgebra for M_B is

$$z = \langle o_z, s_z \rangle \colon B^{A^*} \to B \times \left(B^{A^*}\right)^A$$

which maps any function $f\colon A^* \to B$ to the tuple $z(f) = \big(o_z(f), s_z(f)\big)$. The output value $o_z(f)$ is the value of f on the empty word, i.e., $o_z(f) = f(\epsilon)$ and the successor function $s_z(f)\colon A \to B^{A^*}$ assigns to each letter $a \in A$ the function $s_z(f)(a)\colon A^* \to B$. Its value on a word $w \in A^*$ is equal to the value of f on the word aw, formally $s_z(f)(a)(w) = f(aw)$. In order to lift the machine functor we have two possibilities:

1. We can lift it as an endofunctor obtaining an endofunctor $\overline{M_B}$ on $PMet$.
2. We lift the machine bifunctor M of Example 7.10 to obtain a lifted bifunctor $\overline{M}\colon PMet^2 \to PMet$. Then we fix a pseudometric d_B on the outputs B and consider the induced endofunctor $\overline{M}_{(B,d_B)} := \overline{M}\big((B, d_B), \_\big)$.

In the first case we can of course simply apply Corollary 9.2 from above but in the second case we have to prove nonexpansiveness of λ separately. We will employ the first approach for nondeterministic automata (where $B = 2$) and the second one for probabilistic automata (where $B = [0, 1]$).

9.1. Trace Pseudometric for Nondeterministic Automata.

We will now consider a known $\mathcal{EM}$-law $\lambda\colon \mathcal{P}_f M_2 \Rightarrow M_2 \mathcal{P}_f$ for finitely branching nondeterministic automata [JSS15, p. 867]. Here, Corollary 9.2 is directly applicable using $F = \mathcal{P}_f$ and $G = M_2 = 2 \times \_^A$.

Lemma 9.3 (Distributive Law for Nondeterministic Automata). *Let $(\mathcal{P}_f, \eta, \mu)$ be the finite powerset monad from Example 8.3 with the maximum as evaluation function and $M_2 = 2 \times \_^A$ be the deterministic automaton functor equipped with the evaluation function $ev_{M_2}\colon 2 \times [0, \top]^A \to [0, \top]$, $ev_{M_2}(o, s) = c \cdot \max_{a \in A} s(a)$ with $c \in {]0, 1]}$. We consider the $\mathcal{EM}$-law $\lambda\colon \mathcal{P}_f(2 \times \_^A) \Rightarrow 2 \times \mathcal{P}_f(\_)^A$ on Set which is defined, for any set X, as $\lambda_X(S) = \langle o, s \rangle$ with $o(S) = 1$ if there is an $s' \in X^A$ such that $(1, s') \in S$ else $o(S) = 0$ and the successor functions $s(S)\colon A \to \mathcal{P}X$, where $s(A) = \{s'(a) \mid (o', s') \in S\}$ for every $S \in \mathcal{P}\big(2 \times X^A\big)$. This law is nonexpansive.*

Proof. In the notation of Corollary 9.2 we have $F = 2 \times \_^A = M_2$, and $G = \mathcal{P}_f$. The composed evaluation functions are $ev_{\mathcal{P}_f} * ev_{M_2}\colon \mathcal{P}_f(2 \times [0, 1]^A) \to [0, 1]$ where for $S \in \mathcal{P}_f(2 \times [0, 1]^A)$

$$ev_{\mathcal{P}_f} * ev_{M_2}(S) = ev_{\mathcal{P}_f} \circ \mathcal{P}_f ev_{M_2}(S) = \max\left\{ ev_{M_2}(o, s) \mid (o, s) \in S \right\}$$

$$= \max\left\{ c \cdot \max_{a \in A} s(a) \mid (o, s) \in S \right\} = c \cdot \max_{(o,s) \in S} \max_{a \in A} s(a)$$

and $ev_{M_2} * ev_{\mathcal{P}_f}\colon 2 \times (\mathcal{P}_f[0, 1])^A \to [0, 1]$ where for $(o, s) \in 2 \times (\mathcal{P}_f X)^A$

$$ev_{M_2} * ev_{\mathcal{P}_f}(S) = ev_{M_2} \circ M_2 ev_{\mathcal{P}_f}(o, s) = ev_{M_2}\big(o, \max \circ s\big) = c \cdot \max_{a \in A} \max s(a)$$

As we have seen in Example 7.9 (page 45) we have compositionality for $\mathcal{P}_f M_2$ and the Wasserstein lifting. We want to apply Corollary 9.2 to show nonexpansiveness. For this we just have to check that $(ev_{\mathcal{P}_f} * ev_{M_2}) \circ \lambda_{[0,1]} \le ev_{M_2} * ev_{\mathcal{P}_f}$ holds. For $S \in \mathcal{P}_f(2 \times [0, 1]^A)$ we have $(ev_{\mathcal{P}_f} * ev_{M_2}) \circ \lambda_{[0,1]}(S) = c \cdot \max_{a \in A} \max\{s(a) \mid (o, s) \in S\} = c \cdot \max_{a \in A} \max_{(o,s) \in S} s(a) = ev_{M_2} * ev_{\mathcal{P}_f}(S)$ which concludes the proof. $\square$

With this result at hand we can now define the trace pseudometric for finitely branching nondeterministic automata.

Example 9.4 (Trace Pseudometric for Nondeterministic Automata). We consider the machine endofunctor $M_2 = 2 \times \_^A$. As maximal distance we take $\top = 1$ and as evaluation

function we use $ev_{M_2} \colon [0,1] \times [0,1]^A$ with $ev_{M_2}(o,s) = c \cdot \max_{a \in A} s(a)$ for $0 < c < 1$ as in Example 6.3 (page 36) and lift the functor using the Wasserstein lifting.

We now take a finitely branching nondeterministic automaton which is a coalgebra $\alpha \colon X \to 2 \times (\mathcal{P}_f X)^A$. Its determinization is the powerset automaton $\alpha^\sharp \colon \mathcal{P}_f X \to 2 \times (\mathcal{P}_f X)^A$ whose states are sets of states of the original automaton. We recall from Example 6.5 (page 37) that the bisimilarity pseudometric is the function

$$bd_{\alpha^\sharp} \colon \mathcal{P}_f X \times \mathcal{P}_f X \to [0,1], \quad bd_{\alpha^\sharp}(S,T) = c^{\inf\{n \in \mathbb{N} | \exists w \in A^n . [\![S]\!]_{\alpha^\sharp}(w) \neq [\![T]\!]_{\alpha^\sharp}(w)\}} .$$

If we apply the construction of Definition 9.1 using the unit $\eta_X(x) = \{x\}$ of the powerset monad we obtain the trace pseudometric

$$td_\alpha \colon X \times X \to [0,1], \quad td_\alpha(x,y) = c^{\inf\{n \in \mathbb{N} | \exists w \in A^n . [\![\{x\}]\!]_{\alpha^\sharp}(w) \neq [\![\{y\}]\!]_{\alpha^\sharp}(w)\}} .$$

Thus the trace distance of states x and y of α is given by a word w of minimal length which is contained in the language of the state $\{x\}$ of the determinization $\alpha^\sharp$ and *not* contained in the language of the state $\{y\}$. Then, the distance is computed as $c^{|w|}$.

9.2. **Trace Pseudometric for Probabilistic Automata.** Our next example of an $\mathcal{EM}$-law will be of the shape $\lambda \colon \mathcal{D}_f M_{[0,1]} \to M_{[0,1]} \mathcal{D}_f$. This is much more complicated because we need to consider multifunctors to obtain the correct lifting.

Lemma 9.5 (Distributive Law for Probabilistic Automata)**.** *Let $(\mathcal{D}_f, \eta, \mu)$ be the distribution monad with finite support from Example 8.4 (page 50) and M be the machine bifunctor from Example 7.10 (page 46). There is a known*[7] *$\mathcal{EM}$-law $\lambda \colon \mathcal{D}_f([0,1] \times \_^A) \Rightarrow [0,1] \times \mathcal{D}_f^A$ in Set where $\lambda_X = \langle o_X, s_X \rangle$ with*

$$o_X(P) = \sum_{r \in [0,1]} r \cdot P(r, X^A), \quad and \quad s_X(P) \colon A \to \mathcal{D}_f X, s_X(P)(a)(x) = \sum_{s' \in X^A, s'(a) = x} P([0,1], s')$$

for all sets X and all distributions $P \colon [0,1] \times X^A \to [0,1]$. The endofunctors on both sides of this law can be seen a bifunctors F, G for which one component is fixed. They arise by composition of the distribution functor, the identity functor and the machine bifunctor as follows. The bifunctor on the left hand side (the domain of λ) is $F = \mathcal{D}_f \circ M$ and the bifunctor on the right hand side (the codomain of λ) is $G = M \circ (\mathrm{Id} \times \mathcal{D}_f)$ and to get our law we need to fix the respective first parameter to be $[0,1]$.

If we use the usual distribution functor evaluation function as given in Examples 5.9 and 5.30 (pages 16 and 26), the identity function as evaluation function for Id and the discounted sum for the machine bifunctor as in Example 7.10 (page 46), lift both bifunctors and then fix their first component to the metric space $([0,1], d_e)$ then the above $\mathcal{EM}$-law is nonexpansive.

Proof. Since all of the involved (bi)functors have optimal couplings, we have compositionality and the evaluation functions for the composed functors are

$$ev_F := ev_{\mathcal{D}_f} \circ \mathcal{D}_f ev_M \colon \mathcal{D}_f([0,1] \times [0,1]^A) \to [0,1] \quad \text{and}$$

$$ev_G := ev_M \circ M(\mathrm{id}_{[0,1]}, ev_{\mathcal{D}_f}) \colon [0,1] \times (\mathcal{D}_f X)^A \to [0,1] .$$

---

[7]This law arises out of the so-called *strength map* of the monad [JSS15, Lem. 4]. It is also straightforward to show this just using the properties of an $\mathcal{EM}$-law [Ker16].

For any set X we define the function

$$\Lambda_X := \langle \mathbf{o}_1, \mathbf{o}_2, \mathbf{s} \rangle \colon \mathcal{D}_f\big([0,1]^2 \times (X \times X)^A\big) \to [0,1]^2 \times \big(\mathcal{D}_f(X \times X)\big)^A$$

where for any $P \in \mathcal{D}_f([0,1] \times [0,1] \times (X \times X)^A)$

$$\mathbf{o}_1(P) = \sum_{r \in [0,1]} r \cdot P\big(r, [0,1], (X \times X)^A\big), \quad \mathbf{o}_2(P) = \sum_{r \in [0,1]} r \cdot P\big([0,1], r, (X \times X)^A\big), \text{ and}$$

$$\mathbf{s}(P) \colon A \to \mathcal{D}_f(X \times X), \quad \mathbf{s}(P)(a)(x,y) = \sum_{s' \in (X \times X)^A, \ s'(a)=(x,y)} P([0,1]^2, s')$$

completely analogous to the definition of the components λ_X of the distributive law in Lemma 9.5. A lengthy but straightforward computation[8] shows that for any $P_1, P_2 \in \mathcal{D}_f\big([0,1] \times X^A\big)$ and any $P \in \Gamma_F(P_1, P_2) \subseteq \mathcal{D}_f\big([0,1] \times [0,1] \times (X \times X)^A\big)$ this function satisfies the requirements $\Lambda_X(P) \in \Gamma_G\big(\lambda_X(P_1), \lambda_X(P_2)\big)$ and $\widetilde{G}(d_B, d)\big(\Lambda_X(P)\big) \leq \widetilde{F}(d_B, d)(P)$. Using these properties we can conclude that

$$
\begin{aligned}
(d_B, d)^{\downarrow G}\big(\lambda_X(P_1), \lambda_X(P_2)\big) &= \inf\Big\{ \widetilde{G}(d_B, d)(P') \mid P' \in \Gamma_G\big(\lambda_X(P_1), \lambda_X(P_2)\big) \Big\} \\
&\leq \inf\Big\{ \widetilde{G}(d_B, d)\big(\Lambda_X(P)\big) \mid P \in \Gamma_F(P_1, P_2) \Big\} \\
&\leq \inf\Big\{ \widetilde{F}(d_B, d)(P) \mid t \in \Gamma_F(P_1, P_2) \Big\} = (d_B, d)^{\downarrow F}(P_1, P_2)
\end{aligned}
$$

which, due to compositionality, proves the desired nonexpansiveness of λ_X. $\qquad\square$

Using this law, we can now define the trace pseudometric for probabilistic automata.

Example 9.6 (Trace Pseudometric for Probabilistic Automata)**.** As in Example 7.11 (page 47) we consider the machine functor $M_{[0,1]} = [0,1] \times \_^A$ which arises out of the machine bifunctor M by fixing the first component to $[0,1]$. As maximal distance we set $\top = 1$ and equip the machine bifunctor M with the evaluation function $ev_M \colon [0,1] \times [0,1]^A \to [0,1]$ where $ev_M(o,s) = c_1 o + c_2 |A|^{-1} \sum_{a \in A} s(a)$ for $c_1, c_2 \in {]0,1[}$ such that $c_1 + c_2 \leq 1$ as in Example 7.10. We lift this bifunctor using the Wasserstein lifting and then fix its first component to $([0,1], d_e)$.

For a probabilistic automaton $\alpha \colon X \to [0,1] \to (\mathcal{D}_f X)^A$ its determinization is the $M_{[0,1]}$-coalgebra $\alpha^\sharp \colon \mathcal{D}_f X \to [0,1] \to (\mathcal{D}_f X)^A$ whose state space are distributions on the states of the original automaton. From Example 7.11 we know that we obtain the following bisimilarity pseudometric

$$bd_{\alpha^\sharp} \colon \mathcal{D}_f X \times \mathcal{D}_f X \to [0,1], \quad bd_{\alpha^\sharp}(p,q) = c_1 \cdot \sum_{w \in A^*} \left(\frac{c_2}{|A|}\right)^{|w|} \big|[\![p]\!]_{\alpha^\sharp}(w) - [\![q]\!]_{\alpha^\sharp}(w)\big| .$$

If we apply the construction of Definition 9.1 using the unit $\eta_X(x) = \delta_x^X$ of the finite distribution monad we obtain the trace pseudometric

$$td_\alpha \colon X \times X \to [0,1], \quad td_\alpha(x,y) = c_1 \cdot \sum_{w \in A^*} \left(\frac{c_2}{|A|}\right)^{|w|} \big|[\![\delta_x^X]\!]_{\alpha^\sharp}(w) - [\![\delta_y^X]\!]_{\alpha^\sharp}(w)\big| .$$

---

[8]It is presented in the appendix of the extended version of our CALCO '15 paper [BBKK15].

Thus the trace distance of states x and y of α is given by the distance of their Dirac distributions in the determinization. We obtain a metric which is a weighted version of the total variation distance.

10. Conclusion, Related and Future Work

In this paper we have demonstrated how a substantial part of the coalgebraic machinery for modeling and analyzing labeled transition systems can be extended from a qualitative to a quantitative setting. The crucial idea for this is the idea to lift a functor from the category Set of sets and functions to the category $PMet$ of pseudometric spaces and nonexpansive functions. While all the remaining results require a bit of effort, they arise naturally once such lifting has been defined. The big advantage of our approach is that we try

▷ to keep it as general as possible (by using coalgebra and not restricting to a specific class of transition systems) and

▷ to minimize the amount of additional information needed.

Instead of assuming that a transition system already comes equipped with some distance function on the state space, we give canonical definitions of bisimilarity and trace pseudometrics in the sense that they arise automatically out of the coalgebraic model. The only information we have to provide is the evaluation function which explains how we can evaluate the effect of applying the branching functor to real numbers as a single real number.

Whenever someone is interested in defining a new type of transition system, he or she can now automatically derive canonical notions of both behavioral equivalences and pseudometrics.

Since we suggested two possible liftings (Kantorovich and Wasserstein), we should discuss their respective advantages and disadvantages.

The Wasserstein lifting is somewhat easier to understand due to the transport plan analogy (Section 2) and behaves better with respect to compositionality (Section 7). Furthermore, under some mild conditions, the lifting preserves metrics. Also, there is an interesting fibrational theory underlying the Wasserstein lifting that we are currently working out.

The Kantorovich lifting has other advantages: first it can be characterized via a universal property. Namely, it is the smallest lifting that makes the evaluation function $ev_F \colon (F[0, \top], d_e^F) \to ([0, \top], d_e)$ nonexpansive. Furthermore, the Kantorovich (pseudo)metric seems to be useful for logics and could simplify the proof of a Hennessy-Milner theorem for quantitative coalgebraic logics. If we assume that modal logic formulas are denoted by φ and, for a given coalgebra $c \colon X \to FX$, are associated with a function $[\![\varphi]\!]_c \colon X \to [0, \top]$, a quantitative version of the Hennessy-Milner theorem reads as follows:

$$d_c(x, y) = \sup_{\varphi} d_e([\![\varphi]\!]_c(x), [\![\varphi]\!]_c(y)) \,.$$

If we furthermore observe that bisimulation-invariant modal formulas generalize to nonexpansive quantitative formulas, this formula is similar in nature to the definition of the Kantorovich lifting.

Note that in several cases both liftings coincide, making the decision easy. In all other cases the Wasserstein (pseudo)metric is larger than the Kantorovich (pseudo)metric.

10.1. **Related Work.** The ideas for our framework are not only heavily influenced by transportation theory [Vil09] but also by work on quantitative variants of (bisimulation) equivalence of probabilistic systems. In that context Giacalone, Jou and Smolka observed in the nineties that probabilistic Larsen-Skou bisimulation [LS89] is too strong and therefore introduced a metric based on the notion of ϵ-bisimulations [GJS90]. Such a bisimulation is a relaxation of the usual probabilistic bisimulation relation which allows matching the steps of another state with a probability that is not exactly same but can be at most ϵ apart with respect to the Euclidean metric on $[0, 1]$. Based on this, two states are exactly ϵ apart if this is the smallest value such that the two states are ϵ-bisimilar.

A second approach to behavioral distances is based on logics. Labeled Markov processes (LMP) are generalizations of reactive probabilistic transition system to fairly arbitrary (namely analytic) state spaces which involve some measure theoretic results. Surprisingly, probabilistic bisimilarity for these systems can be expressed via a simple modal logic without negation [DEP98] in the sense that two states of an LMP are bisimilar if and only if they satisfy the same formulae. Using this logical framework Desharnais, Gupta, Jagadeesan and Panangaden defined a family of metrics between LMPs [DGJP04] via functional expressions, which can be understood as quantitative generalization of the logical formulae. If evaluated on a state of an LMP, such a functional expression measures the extent (as real number between 0 and 1) to which a formula is satisfied in that state. Then, for any set of such functional expressions, the distance of two LMPs is the supremum of all differences (with respect to the Euclidean distance on $[0, 1]$) of these functional expressions.

A third, coalgebraic approach, which inspired us to develop our framework, is used by van Breugel and Worrell [vBW05, vBW06]. As we have presented in the examples in this paper, they define both a discounted and an undiscounted pseudometric on probabilistic systems via a fixed point approach using the usual Kantorovich pseudometric for probability measures. Moreover, they show that this metric is related to the logical pseudometric by Desharnais et al. [vBW05]. We quickly point out that metrics in a coalgebraic setting appeared already earlier in a paper by de Vink and Rutten. They used ultrametrics[9] and the category of ultrametric spaces in order to define coalgebraic bisimulation for continuous probabilistic transition systems [dVR97, dVR99]. However, they use it mainly as a technical tool to get a final coalgebra and not in order to study bisimilarity distances.

Not only the definition of distances for probabilistic systems and the study of their theoretical importance but also their efficient approximation or exact computation has been the focus of several recent research papers [vBW01, FPP04, TDZ11, CPP12]. In particular, Chen, van Breugel and Worrell proved that both the discounted and the undiscounted bisimilarity pseudometric for probabilistic systems can be computed exactly in polynomial time exploiting algorithms to solve linear programs [CvBW12]. Taking some inspiration from this work, one year later Bacci, Bacci, Larsen and Mardare proposed an on-the-fly approach for the exact computation of bisimilarity distances [BBLM13] which they proved to be practically much more efficient than the earlier approximation algorithms.

Behavioral distances have not only been studied for probabilistic systems but also for other types of transition systems. An example which also appears in the main text is the branching distance for metric transition systems [dAFS09]. Moreover, a thorough comparison of various different behavioral distances on labeled transition systems has recently been

---

[9]An ultrametric is a reflexive and symmetric function $d\colon X^2 \to [0, 1]$ satisfying the implication $d(x, y) \implies x = y$ and the strong triangle inequality $d(x, z) \leq \max\{d(x, y), d(y, z)\}$.

carried out by Fahrenberg, Legay and Thrane. They transfer van Glabbeek's quantitative linear-time–branching time spectrum [vG90] to a quantitative setting [FLT11b, FL14].

The lifting of a monad to the bicategory of $\mathcal{V}$-matrices (where $\mathcal{V}$ is a quantale), similar to our Wasserstein lifting, has earlier been studied in [Hof07]. After our original publication [BBKK14] on lifting functors to $PMet$, there have been other suggestions to categorically characterize such liftings in even more general settings [KS15, BKV15]. Both approaches are based on Kan extensions, where the work by Katsumata and Sato generalizes the Kantorovich lifting and the method by Balan, Kurz and Velebil is reminiscent of the Wasserstein lifting.

10.2. **Future Work.** This paper proposes a paradigmatic shift from qualitative to quantitative behavior analysis and although many basic results are in place there is still a lot of work ahead. We will first discuss a few open questions whose answers (if they exist) might yield improvements of our current framework. Then we discuss further possible generalizations.

In light of the fact that we propose two different lifting approaches, the first apparent question is whether there are conditions that guarantee that these liftings coincide, i.e., such that the Kantorovich-Rubinstein duality holds. However, some preliminary attempts suggest that this is very difficult. The proof for the duality in the (arbitrary) probabilistic setting is domain specific and cannot easily be generalized.

Another valid question concerning the two different liftings is whether they can be captured by some *universal properties*. Although we use a coalgebraic and thus category theoretic framework, our intuition comes from transportation and probability theory. It would be interesting to figure out whether there is some general category theoretic construction such that our liftings are two ends of this construction in a similar way as the Kleisli and Eilenberg-Moore categories of a monad are initial and final objects in a category of adjunctions. A possible source of inspiration for this line of work could be van Breugel's draft paper on the metric monad [vB05] which describes a generalization of the (continuous) Giry monad in terms of universal properties involving monad morphisms.

In order to obtain trace pseudometrics we employed the generalized powerset construction by Silva, Bonchi, Bonsangue and Rutten [SBBR13], one of the two most prominent coalgebraic approaches to traces. Another well-known approach, suggested by Hasuo, Jacobs and Sokolova [HJS07], is to work in the Kleisli category. Whether our framework can be modified or extended to work in that setting remains an interesting open question. A possible basis for answering this could be the recent comparison between these two approaches in the qualitative setting which was carried out by Jacobs, Silva and Sokolova [JSS15]. However, this is already quite complicated in the qualitative case, and thus very likely to be even more involved in the quantitative case. In particular, for this comparison one needs – in addition to the distributive law ($\mathcal{EM}$-law) – a $\mathcal{K}\ell$-law and a so-called extension natural transformation which connects the two distributive laws.

While our theory is already at a quite general level, there are several further possible generalizations. Among these, one could be to drop symmetry and study so-called *directed pseudometrics* as is done in the case of metric transition systems [dAFS09]. This would lay the foundation to study simulation distances from a coalgebraic perspective.

An even more general idea is to study certain reflexive functions $d\colon X^2 \to L$ as generalized metrics, where L is a (complete) lattice with possibly some additional structure. This could result in a theory in which one can model distances for conditional transition systems (CTS). These systems were proposed by Adámek, Bonchi, Hülsbusch, König, Milius and Silva [ABH+12] and are similar to featured transition systems [CHS+10]. Formally, a CTS is a

labeled transition system with the following semantics. Once the environment chooses a label (which represents a condition or a feature), all transitions with this label remain (but the label is dropped) and all the other transitions vanish. Then one is interested in the behavior of the resulting unlabeled transition system. If A is the set of conditions, we could take the complete lattice (with respect to the subset ordering) $L = \mathcal{P}A$ as codomain of our generalized metrics. A natural distance of two states x, y could be the set of conditions for which they are not bisimilar. This is a further generalization of a distance recently proposed by Atlee, Fahrenberg and Legay for featured transition systems: their (simulation) distance only counts how many features prevent simulation [AFL15].

A less drastic, yet interesting generalization could result from replacing the Euclidean metric with a different metric which has recently been suggested by Chatzikokolakis, Gebler, Palamidessi and Xu for probabilistic systems [CGPX14]. Our Kantorovich lifting is obviously based on the Euclidean metric and we could simply replace $d_e \colon [0, \top]^2 \to [0, \top]$ with a different metric in this definition and study the resulting liftings. However, it is unclear how to proceed for the Wasserstein lifting since the Euclidean metric only plays a role in Condition W2 of the well-behavedness of the respective evaluation function.

The use of so-called up-to techniques [San98] can significantly reduce both memory consumption and running time for equivalence checks between two specific states of a labeled transition system. This was demonstrated recently by Bonchi and Pous for the equivalence check of nondeterministic finite automata [BP13]. Together with Petrişan and Rot, they discovered a fibrational basis for these techniques [BPPR14, Rot15]. Based on some preliminary research, we know that our entire framework can be seen from a fibrational perspective as well, leading to yet another generalization. We hope that this fibrational view will lead to more general proofs and efficiently computable algorithms, possibly by using the aforementioned up-to techniques.

Another valid question is, if (and how) our techniques can be generalized to other categories than Set. In order to talk about distances between states it is likely that this will require at least concrete categories.

As we have discussed above, for probabilistic systems it is possible to efficiently compute the lifted distance by using ideas from linear programming. While our framework provides a solid theoretical basis for reasoning about several behavioral distances, its algorithmic applicability as yet is somewhat limited. For this we would need

▷ an efficient method to compute the lifted distances, and possibly also

▷ an efficient method to automatically derive the fixed point of Lemma 6.1.

However, it is most likely that the efficiency of these methods will to a great extent rely on the specific system under consideration.

Last but not least, the coalgebraic theory in Set has benefited a lot from the huge amount of examples. While we have already looked at several examples throughout this paper, there is yet a lot more examples that should be worked out explicitly. In particular, it would be interesting to see if (and how) we can recover other behavioral distances from the literature.

Acknowledgement

References

[ABH+12] Jiří Adámek, Filippo Bonchi, Mathias Hülsbusch, Barbara König, Stefan Milius, and Alexandra Silva. A coalgebraic perspective on minimization and determinization. In Lars Birkedal, editor, *Foundations of Software Science and Computational Structures*, volume 7213 of *Lecture Notes in Computer Science*, pages 58–73. Springer Berlin Heidelberg, 2012.

[Adá74] Jiří Adámek. Free algebras and automata realizations in the language of categories. *Commentationes Mathematicae Universitatis Carolinae*, 015(4):589–602, 1974.

[AFL15] Joanne M. Atlee, Uli Fahrenberg, and Axel Legay. Measuring behaviour interactions between product-line features. In *Proceedings of the Third FME Workshop on Formal Methods in Software Engineering*, Formalise '15, pages 20–25, Piscataway, NJ, USA, 2015. IEEE Press.

[AHS90] Jiří Adámek, Horst Herrlich, and George E. Strecker. *Abstract and Concrete Categories. The Joy of Cats*, volume 17. [John Wiley and Sons, New York], 2006 [1990].

[BBKK14] Paolo Baldan, Filippo Bonchi, Henning Kerstan, and Barbara König. Behavioral Metrics via Functor Lifting. In Venkatesh Raman and S. P. Suresh, editors, *34th International Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2014)*, volume 29 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 403–415. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2014.

[BBKK15] Paolo Baldan, Filippo Bonchi, Henning Kerstan, and Barbara König. Towards trace metrics via functor lifting. In Lawrence S. Moss and Paweł Sobociński, editors, *6th Conference on Algebra and Coalgebra in Computer Science (CALCO'15)*, volume 35 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 35–49. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2015.

[BBLM13] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. On-the-fly exact computation of bisimilarity distances. In Nir Piterman and Scott A. Smolka, editors, *Tools and Algorithms for the Construction and Analysis of Systems*, volume 7795 of *Lecture Notes in Computer Science*, pages 1–15. Springer Berlin Heidelberg, 2013.

[BBLM15] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. On the total variation distance of semi-markov chains. In *Foundations of Software Science and Computation Structures (FOSSACS 2015)*, volume 9034 of *Lecture Notes in Computer Science*. Springer, 2015.

[BKV15] Adriana Balan, Alexander Kurz, and Jiří Velebil. Extensions of functors from Set to $\mathcal{V}$-cat. In *Proc. of CALCO '15*, volume 35 of *LIPIcs*, pages 17–34. Schloss Dagstuhl – Leibniz Center for Informatics, 2015.

[BP13] Filippo Bonchi and Damien Pous. Checking NFA equivalence with bisimulations up to congruence. In *Proceedings of the 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, POPL '13, pages 457–468, New York, NY, USA, 2013. ACM.

[BPPR14] Filippo Bonchi, Daniela Petrişan, Damien Pous, and Jurriaan Rot. Coinduction up-to in a fibrational setting. In *Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, CSL-LICS '14. ACM, 2014.

[CGPX14] Konstantinos Chatzikokolakis, Daniel Gebler, Catuscia Palamidessi, and Lili Xu. Generalized bisimulation metrics. In Paolo Baldan and Daniele Gorla, editors, *CONCUR 2014 – Concurrency Theory*, volume 8704 of *Lecture Notes in Computer Science*, pages 32–46. Springer Berlin Heidelberg, 2014.

[CHS+10] Andreas Classen, Patrick Heymans, Pierre-Yves Schobbens, Axel Legay, and Jean-François Raskin. Model checking lots of systems: Efficient verification of temporal properties in software product lines. In *Proceedings of the 32Nd ACM/IEEE International Conference on Software Engineering - Volume 1*, ICSE '10, pages 335–344, New York, NY, USA, 2010. ACM.

[CPP12] Gheorghe Comanici, Prakash Panangaden, and Doina Precup. On-the-fly algorithms for bisimulation metrics. In *2012 Ninth International Conference on Quantitative Evaluation of Systems (QEST)*, pages 94–103. IEEE Computer Society, 2012.

[CvBW12] Di Chen, Franck van Breugel, and James Worrell. On the complexity of computing probabilistic bisimilarity. In Lars Birkedal, editor, *Foundations of Software Science and Computational Structures*, volume 7213 of *Lecture Notes in Computer Science*, pages 437–451. Springer, 2012.

[dAFS04] Luca de Alfaro, Marco Faella, and Mariëlle Stoelinga. Linear and branching metrics for quantitative transition systems. In *Proc. of ICALP '04*, volume 3142 of *Lecture Notes in Computer Science*, pages 97–109. Springer, 2004.

[dAFS09] Luca de Alfaro, Marco Faella, and Mariëlle Stoelinga. Linear and branching system metrics. *IEEE Trans. Software Eng.*, 35(2):258–273, 2009.

[DEP98] Josée Desharnais, Abbas Edalat, and Prakash Panangaden. A logical characterization of bisimulation for labeled Markov processes. In *Proceedings of the 13th IEEE Symposium on Logic in Computer Science, Indianapolis*, volume 179(2), pages 478–489. IEEE Press, New York, 1998.

[DGJP04] Josée Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labelled markov processes. *Theoretical Computer Science*, 318(3):323 – 354, 2004.

[dVR97] Erik P. de Vink and Jan J.M.M. Rutten. Bisimulation for probabilistic transition systems: a coalgebraic approach. In Pierpaolo Degano, Roberto Gorrieri, and Alberto Marchetti-Spaccamela, editors, *Automata, Languages and Programming*, volume 1256 of *Lecture Notes in Computer Science*, pages 460–470. Springer Berlin Heidelberg, 1997.

[dVR99] Erik P. de Vink and Jan J.M.M Rutten. Bisimulation for probabilistic transition systems: a coalgebraic approach. *Theoretical Computer Science*, 221(1–2):271 – 293, 1999.

[FL14] Uli Fahrenberg and Axel Legay. The quantitative linear-time–branching-time spectrum. *Theoretical Computer Science*, 538(0):54 – 69, 2014. Quantitative Aspects of Programming Languages and Systems (2011-12).

[FLT11a] Uli Fahrenberg, Axel Legay, and Claus Thrane. The quantitative linear-time–branching-time spectrum. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2011)*, volume 13 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 103–114. Schloss Dagstuhl – Leibniz Center for Informatics, 2011.

[FLT11b] Uli Fahrenberg, Axel Legay, and Claus Thrane. The Quantitative Linear-Time–Branching-Time Spectrum. In Supratik Chakraborty and Amit Kumar, editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2011)*, volume 13 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 103–114, Dagstuhl, Germany, 2011. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[FPP04] Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite markov decision processes. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, UAI '04, pages 162–169, Arlington, Virginia, United States, 2004. AUAI Press.

[GJS90] Alessandro Giacalone, Chi-Chang Jou, and Scott A. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Proc. IFIP TC2 Working Conference on Programming Concepts and Methods*, pages 443–458. North-Holland, 1990.

[HJS07] Ichiro Hasuo, Bart Jacobs, and Ana Sokolova. Generic trace semantics via coinduction. *Logical Methods in Computer Science*, 3 (4:11):1–36, 2007.

[Hof07] Dirk Hofmann. Topological theories and closed objects. *Advances in Mathematics*, 215(2):789–824, 2007.

[JSS12] Bart Jacobs, Alexandra Silva, and Ana Sokolova. Trace semantics via determinization. In Dirk Pattinson and Lutz Schröder, editors, *Coalgebraic Methods in Computer Science*, volume 7399 of *Lecture Notes in Computer Science*, pages 109–129. Springer Berlin Heidelberg, 2012.

[JSS15] Bart Jacobs, Alexandra Silva, and Ana Sokolova. Trace semantics via determinization. *Journal of Computer and System Sciences*, 81(5):859 – 879, 2015. 11th International Workshop on Coalgebraic Methods in Computer Science, CMCS 2012 (Selected Papers).

[Ker16] Henning Kerstan. *Coalgebraic Behavior Analysis – From Qualitative to Quantitative Analyses*. PhD thesis, Universität Duisburg-Essen, 2016.

[KS15] Shin-ya Katsumata and Tetsuya Sato. Codensity liftings of monads. In *Proc. of CALCO '15*, volume 35 of *LIPIcs*, pages 156–170. Schloss Dagstuhl – Leibniz Center for Informatics, 2015.

[LS89] Kim G. Larsen and Arne Skou. Bisimulation through probabilistic testing (preliminary report). In *Conference Record of the 16th ACM Symposium on Principles of Programming Languages (POPL)*, pages 344–352, 1989.

[Rot15] Jurriaan Cornelis Rot. *Enhanced Coinduction*. PhD thesis, Universiteit Leiden, 2015.

[Rut00]   Jan J.M.M. Rutten. Universal coalgebra: a theory of systems. *Theoretical Computer Science*, 249:3–80, 2000.

[San98]   Davide Sangiorgi. On the bisimulation proof method. *Mathematical Structures in Computer Science*, 8(5):447–479, 1998.

[SBBR13]   Alexandra Silva, Filippo Bonchi, Marcello M. Bonsangue, and Jan J. M. M. Rutten. Generalizing determinization from automata to coalgebras. *Logical Methods in Computer Science*, 9(1), 2013.

[Sok11]   Ana Sokolova. Probabilistic systems coalgebraically: A survey. *Theoretical Computer Science*, 412(38):5095–5110, 2011. CMCS Tenth Anniversary Meeting.

[Tar55]   Alfred Tarski. A lattice-theoretical theorem and its applications. *Pacific Journal of Mathematics*, 5:285–309, 1955.

[TDZ11]   Mathieu Tracol, Josée Desharnais, and Abir Zhioua. Computing distances between probabilistic automata. In Mieke Massink and Gethin Norman, editors, *9th Workshop on Quantitative Aspects of Programming Languages (QAPL 2011)*, volume 57 of *EPTCS*, pages 148–162, 2011.

[vB05]   Franck van Breugel. The metric monad for probabilistic nondeterminism. `http://www.cse.yorku.ca/~franck/research/drafts/monad.pdf`, 2005.

[vBSW08]   Franck van Breugel, Babita Sharma, and James Worrell. Approximating a behavioural pseudometric without discount for probabilistic systems. *Logical Methods in Computer Science*, 4 [2:2]:1 – 23, 2008.

[vBW01]   Franck van Breugel and James Worrell. An algorithm for quantitative verification of probabilistic transition systems. In Kim G. Larsen and Mogens Nielsen, editors, *CONCUR 2001 — Concurrency Theory*, volume 2154 of *Lecture Notes in Computer Science*, pages 336–350. Springer Berlin Heidelberg, 2001.

[vBW05]   Franck van Breugel and James Worrell. A behavioural pseudometric for probabilistic transition systems. *Theoretical Computer Science*, 331(1):115–142, 2005.

[vBW06]   Franck van Breugel and James Worrell. Approximating and computing behavioural distances in probabilistic transition systems. *Theoretical Computer Science*, 360(1):373–385, 2006.

[vG90]   Robert J. van Glabbeek. The linear time - branching time spectrum. In J. C. M. Baeten and J. W. Klop, editors, *CONCUR '90 Theories of Concurrency: Unification and Extension*, volume 458 of *Lecture Notes in Computer Science*, pages 278–297. Springer Berlin Heidelberg, 1990.

[Vil09]   Cédric Villani. *Optimal Transport – Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin Heidelberg, 2009.