Towards Trace Metrics via Functor Lifting*

Paolo Baldan[1], Filippo Bonchi[2], Henning Kerstan[3], and Barbara König[3]

1 **Dipartimento di Matematica, Università di Padova, Italy**
baldan@math.unipd.it
2 **CNRS, ENS Lyon, Université de Lyon, France**
filippo.bonchi@ens-lyon.fr
3 **Universität Duisburg-Essen, Germany**
henning.kerstan@uni-due.de, barbara_koenig@uni-due.de

Abstract

We investigate the possibility of deriving metric trace semantics in a coalgebraic framework. First, we generalize a technique for systematically lifting functors from the category **Set** of sets to the category **PMet** of pseudometric spaces, by identifying conditions under which also natural transformations, monads and distributive laws can be lifted. By exploiting some recent work on an abstract determinization, these results enable the derivation of trace metrics starting from coalgebras in **Set**. More precisely, for a coalgebra in **Set** we determinize it, thus obtaining a coalgebra in the Eilenberg-Moore category of a monad. When the monad can be lifted to **PMet**, we can equip the final coalgebra with a behavioral distance. The trace distance between two states of the original coalgebra is the distance between their images in the determinized coalgebra through the unit of the monad. We show how our framework applies to nondeterministic automata and probabilistic automata.

1998 ACM Subject Classification F.3.1 Specifying and Verifying and Reasoning about Programs, D.2.4 Software/Program Verification

Keywords and phrases trace metric, monad lifting, pseudometric, coalgebra

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 (see, e.g., [10, 7, 23, 1, 5, 6, 8]).

Coalgebras [18] are a well-established abstract framework where a canonical notion of behavioral equivalence can be uniformly derived. The behavior of a system is represented as a coalgebra, namely a map of the form $X \to HX$, where X is a state space and H is a functor that describes the type of computation performed. For instance nondeterministic automata can be seen as coalgebras $X \to 2 \times \mathcal{P}(X)^A$: for any state we specify whether it is final or not, and the set of successors for any given input in A. Under suitable conditions a final coalgebra exists which can be seen as minimized version of the system, so that two states are deemed equivalent when they correspond to the same state in the final coalgebra.

* Partially supported by the University of Padova project ANCORE and the DFG project BEMEGA. The second author acknowledges the support by project ANR 12IS0 2001 PACE.

In a recent paper [2] we faced the problem of devising a framework where, given a coalgebra for an endofunctor H on **Set**, one can systematically derive pseudometrics which measure the behavioral distance of states. A first crucial step is the lifting of H to a functor $\overline{H}$ on **PMet**, the category of pseudometric spaces. In particular, we presented two different approaches which can be viewed as generalizations of the Kantorovich and Wasserstein pseudometrics for probability measures. One can prove that the final coalgebra in **Set** can be endowed with a metric, arising as a solution of a fixpoint equation, turning it into the final coalgebra for the lifting $\overline{H}$. Since any coalgebra $X \to HX$ can be seen as a coalgebra in **PMet** by endowing X with the discrete metric, the unique mapping into the final coalgebra provides a behavioral distance on X.

The canonical notion of equivalence for coalgebras, in a sense, fully captures the behavior of the system as expressed by the functor H. As such, it naturally corresponds to bisimulation equivalences already defined for various concrete formalisms. Sometimes one is interested in coarser equivalences, ignoring some aspects of a computation, a notable example being trace equivalence where the computational effect which is ignored is branching.

In this paper, relying on recent work on an abstract determinization construction for coalgebras in [20, 13, 14], we extend the above framework in order to systematically derive trace metrics. The mentioned work starts from the observation that the distinction between the behavior to be observed and the computational effects that are intended to be hidden from the observer, is sometimes formally captured by splitting the functor H characterizing system computations in two components, a functor F for the observable behavior and a monad T describing the computational effects, e.g., lifting $1 + -$, the powerset functor $\mathcal{P}$ or the distribution functor $\mathcal{D}$ 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 $FX = 2 \times X^A$, describing the transitions, 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.

On this basis, [14] develops a framework for deriving behavioral equivalences which only considers the visible behavior, ignoring the computational effects. The core idea consists in "incorporating" the effect of the monad also in the set of states X, which thus becomes TX, by means of a construction that can be seen as an abstract form of determinization. For functors of the shape FT, this can be done by lifting F to a functor $\widehat{F}$ in $\mathcal{EM}(T)$, the Eilenberg-Moore category of T, using a distributive law between F and T. In fact, the final F-coalgebra lifts to the final $\widehat{F}$-coalgebra in $\mathcal{EM}(T)$. The technique works, at the price of some complications, also for functors of the shape TF [14].

Here, we exploit the results in [14] for systematically deriving metric trace semantics for **Set**-based coalgebras. The situation is summarized in the diagram at the end of Subsection 5.1. As a first step, building on our technique for lifting functors from the category **Set** of sets to the category **PMet** of pseudometric spaces, we identify conditions under which also natural transformations, monads and distributive laws can be lifted. 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 $\overline{F}\colon \mathbf{PMet} \to \mathbf{PMet}$ to an endofunctor $\widehat{\overline{F}}$ on $\mathcal{EM}(\overline{T})$. By using the trivial 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 the final coalgebra. This final coalgebra is again obtained by lifting the final $\overline{F}$-coalgebra, i.e. a coalgebra equipped with a behavioral distance, to $\mathcal{EM}(\overline{T})$.

The trace distance between two states of the original coalgebra can then be defined as the distance between their images in the determinized coalgebra through the unit of the monad. We illustrate our framework by thoroughly discussing two running examples, namely nondeterministic automata and probabilistic automata. We show that it allows us to recover known or meaningful trace distances such as the standard ultrametric on word languages for nondeterministic automata or the total variation distance on distributions for probabilistic automata.

The paper is structured as follows. In Section 2 we will introduce our notation and quickly recall the basics of our lifting framework from [2]. Then, in Section 3, we tackle the question of compositionality, i.e. we investigate whether based on liftings of two functors we can obtain a lifting of the composed functor. The lifting of natural transformations and monads is treated in Section 4. Equipped with these tools, we show as main result in Section 5 how to obtain trace pseudometrics in the Eilenberg-Moore category of a lifted monad. We conclude our paper with a discussion on related and future work (Section 6). Proofs can be found in the extended version [`arXiv:1505.08105`].

2 Preliminaries

In this section we recap some basic notions and fix the corresponding notation. We also briefly recall the results in [2] which will be exploited in the paper.

We assume that the reader is familiar with the basic notions of category theory, especially with the definitions of functor, product, coproduct and weak pullbacks.

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. Finally, 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 $x \in X$.

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 $\operatorname{supp}(P) := \{x \in X \mid P(x) > 0\}$.

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 $\pi_i\colon \prod_{i=1}^n X_i \to X_i$. For a source $(f_i\colon X \to X_i)_{i=1}^n$ we denote the unique mediating arrow to the product by $\langle f_1, \dots, f_n\rangle\colon X \to \prod_{i=1}^n X_i$. Similarly, given a family of arrows $(f_i\colon X_i \to Y_i)_{i=1}^n$, we write $f_1 \times \cdots \times f_n = \langle f_1 \circ \pi_1, \dots, f_n \circ \pi_n\rangle\colon \prod_{i=1}^n X_i \to \prod_{i=1}^n Y_i$.

For $\top \in (0, \infty]$ and a set X we call any function $d\colon X^2 \to [0, \top]$ a ($\top$-)*distance* on X (for our examples we will use $\top = 1$ or $\top = \infty$). Whenever d satisfies, for all $x, y, z \in X$, $d(x, x) = 0$ (reflexivity), $d(x, y) = d(y, x)$ (symmetry) and $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality) we call it a *pseudometric* and if it additionally satisfies $d(x, y) = 0 \implies x = y$ we call it a *metric*. Given such a function d on a set X, we say that (X, d) is a pseudometric/metric space. By $d_e\colon [0, \top]^2 \to [0, \top]$ we denote the ordinary Euclidean distance 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 $x \in [0, \infty]$. We call a function $f\colon X \to Y$ between pseudometric spaces (X, d_X) and (Y, d_Y) *nonexpansive* and write $f\colon (X, d_X) \xrightarrow{1} (Y, d_Y)$ if $d_Y \circ (f \times f) \leq d_X$. If equality holds we call f an *isometry*.

By choosing a fixed maximal element $\top$ in our definition of distances, we ensure that the set of pseudometrics over a fixed set with pointwise order is a complete lattice (since $[0, \top]$ is) and we obtain a complete and cocomplete category of pseudometric spaces and nonexpansive functions, which we denote by $\mathbf{PMet}$. Given a functor F on $\mathbf{Set}$, we aim at constructing a functor $\overline{F}$ on **PMet** which is a lifting of F in the following sense.

▶ **Definition 2.1** (Lifting). Let $U\colon \mathbf{PMet} \to \mathbf{Set}$ be the forgetful functor which maps every pseudometric space to its underlying set. A functor $\overline{F}\colon \mathbf{PMet} \to \mathbf{PMet}$ is called a *lifting* of a functor $F\colon \mathbf{Set} \to \mathbf{Set}$ if it satisfies $U\overline{F} = FU$.

Similarly to predicate lifting of coalgebraic modal logic [19], lifting to **PMet** can be conveniently defined once a suitable (evaluation) function from $F[0,\top]$ to $[0,\top]$ is fixed.

▶ **Definition 2.2** (Evaluation Function & Evaluation Functor). Let F be an endofunctor on **Set**. An *evaluation function* for F is a function $ev_F\colon F[0,\top] \to [0,\top]$. Given such a function, we define the *evaluation functor* to be the endofunctor $\widetilde{F}$ on $\mathbf{Set}/[0,\top]$, the slice category[1] over $[0,\top]$, via $\widetilde{F}(g) = ev_F \circ Fg$ for all $g \in \mathbf{Set}/[0,\top]$. On arrows $\widetilde{F}$ is defined as F.

A first lifting technique leads to what we called the Kantorovich pseudometric, which is the smallest possible pseudometric d^F on FX such that, for all nonexpansive functions $f\colon (X,d) \xrightarrow{1} ([0,\top],d_e)$, also $\widetilde{F}f\colon (FX,d^F) \xrightarrow{1} ([0,\top],d_e)$ is again nonexpansive.

▶ **Definition 2.3** (Kantorovich Pseudometric & Kantorovich Lifting). Let $F\colon \mathbf{Set} \to \mathbf{Set}$ be a functor with an evaluation function ev_F. For every pseudometric space (X,d) the *Kantorovich pseudometric* on FX is the function $d^{\uparrow F}\colon FX \times FX \to [0,\top]$, where for all $t_1,t_2 \in FX$:

$$d^{\uparrow F}(t_1,t_2) := \sup\left\{d_e\left(\widetilde{F}f(t_1),\widetilde{F}f(t_2)\right) \mid f\colon (X,d) \xrightarrow{1} ([0,\top],d_e)\right\} .$$

The *Kantorovich lifting* of the functor F is the functor $\overline{F}\colon \mathbf{PMet} \to \mathbf{PMet}$ defined as $\overline{F}(X,d) = (FX, d^{\uparrow F})$ and $\overline{F}f = Ff$.

This definition is sound, i.e., $d^{\uparrow F}$ is guaranteed to be a pseudometric so that we indeed obtain a lifting of the functor. A dual way for obtaining a pseudometric on FX relies on ideas from probability and transportation theory. It is based on the notion of couplings, which can be understood as a generalization of joint probability measures.

▶ **Definition 2.4** (Coupling). Let $F\colon \mathbf{Set} \to \mathbf{Set}$ be a functor and $n \in \mathbb{N}$. Given a set X and $t_i \in FX$ for $1 \le i \le n$ 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 $\Gamma_F(t_1,t_2,\ldots,t_n)$ for the set of all these couplings.

Based on these couplings we are now able to define an alternative distance on FX.

▶ **Definition 2.5** (Wasserstein Distance & Wasserstein Lifting). Let $F\colon \mathbf{Set} \to \mathbf{Set}$ 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, for all $t_1,t_2 \in FX$,

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

If $d^{\downarrow F}$ is a pseudometric for all pseudometric spaces (X,d), we define the *Wasserstein lifting* of F to be the functor $\overline{F}\colon \mathbf{PMet} \to \mathbf{PMet}$, $\overline{F}(X,d) = (FX, d^{\downarrow F})$, $\overline{F}f = Ff$.

The names Kantorovich and Wasserstein used for the liftings derive from transportation theory [25]. Indeed we obtain a transport problem if we instantiate F with the distribution functor $\mathcal{D}$ (see also Example 2.9 below). In order to measure the distance between two

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

probability distributions $s, t\colon X \to [0,1]$ it is useful to think of the following analogy: assume that X is a collection of cities (with distance function d between them) and s, t represent supply and demand (in units of mass). The distance between s, t can be measured in two ways: the first is to set up an optimal transportation plan with minimal costs (also called coupling) to transport goods from cities with excess supply to cities with excess demand. The cost of transport is determined by the product of mass and distance. In this way we obtain the Wasserstein distance. A different view is to imagine a logistics firm that is commissioned to handle the transport. It sets prices for each city and buys and sells for this price at every location. However, it has to ensure that the price function (here, f) is nonexpansive, i.e., the difference of prices between two cities is smaller than the distance of the cities, otherwise it will not be worthwhile to outsource this task. This firm will attempt to maximize its profit, which can be considered as the Kantorovich distance of s, t. The Kantorovich-Rubinstein duality informs us that these two views lead to the exactly same result.

In Definition 2.5 we are not guaranteed, in general, that $d^{\downarrow F}$ is a pseudometric. This is the case if we require F to preserve weak-pullbacks and impose the following restrictions on the evaluation function.

▶ **Definition 2.6** (Well-Behaved). Let F be a functor with an evaluation function ev_F. 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 $f \le g$, we have $\widetilde{F}f \le \widetilde{F}g$.
W2. For each $t \in F([0,\top]^2)$ it holds that $d_e(ev_F(t_1), ev_F(t_2)) \le \widetilde{F}d_e(t)$ 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 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). Condition W3 requires that exactly the elements of $F\{0\}$ are mapped to 0 via ev_F. This is necessary for reflexivity of the Wasserstein pseudometric. Indeed, with this definition at hand we were able to prove the desired result.

▶ **Proposition 2.7** ([2]). *If F preserves weak pullbacks and ev_F is well-behaved, then $d^{\downarrow F}$ is a pseudometric for any pseudometric space (X, d).*

From now on, whenever we use the Wasserstein lifting $d^{\downarrow F}$, we implicitly assume to be in the hypotheses of Proposition 2.7. It can be shown that, in general, $d^{\uparrow F} \le d^{\downarrow F}$. Whenever equality holds we say that the functor and the evaluation function satisfy the *Kantorovich-Rubinstein duality*. This is helpful in many situations (e.g., in [24] it allowed to reuse an efficient linear programming algorithm to compute behavioral distance) but it is usually difficult to obtain.

We now recall two examples which will play an important role in this paper. First, we consider the following finitary variant of the powerset functor.

▶ **Example 2.8** (Finite Powerset). The finite powerset functor $\mathcal{P}_{fin}$ assigns to each set X the set $\mathcal{P}_{fin}X = \{S \subseteq X \mid |S| < \infty\}$ and to each function $f\colon X \to Y$ the function $\mathcal{P}_{fin}f\colon \mathcal{P}_{fin}X \to \mathcal{P}_{fin}Y$, $\mathcal{P}_{fin}f(S) := f[S]$. This functor preserves weak pullbacks and the evaluation function $\max\colon \mathcal{P}_{fin}([0,\infty]) \to [0,\infty]$ with $\max\emptyset = 0$ is well-behaved. The Kantorovich-Rubinstein duality holds and the resulting distance is the Hausdorff pseudometric which, for any pseudometric space (X, d) and any $X_1, X_2 \in \mathcal{P}_{fin}X$, is defined as

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

Our second example is the following finite variant of the distribution functor.

▶ **Example 2.9** (Finitely Supported Distributions). The probability distribution functor $\mathcal{D}$ assigns to each set X the set $\mathcal{D}X = \{P\colon X \to [0,1] \mid |\mathrm{supp}(P)| < \infty, P(X) = 1\}$ and to each function $f\colon X \to Y$ the function $\mathcal{D}f\colon \mathcal{D}X \to \mathcal{D}Y$, $\mathcal{D}f(P)(y) = \sum_{x\in f^{-1}[\{y\}]} P(x) = P(f^{-1}[\{y\}])$. It preserves weak pullbacks and the evaluation function $ev_{\mathcal{D}}\colon \mathcal{D}[0,1] \to [0,1]$, $ev_{\mathcal{D}}(P) = \sum_{r\in[0,1]} r \cdot P(r)$ is well-behaved. For any pseudometric space (X,d) we obtain the Wasserstein pseudometric which, for any $P_1, P_2 \in \mathcal{D}X$, is defined as

$$d^{\downarrow\mathcal{D}}(P_1,P_2) = \min\left\{ \sum_{x_1,x_2\in X} d(x_1,x_2)\cdot P(x_1,x_2) \;\middle|\; P \in \Gamma_{\mathcal{D}}(P_1,P_2) \right\}.$$

The Kantorovich-Rubinstein duality [25] holds from classical results in transportation theory.

While these two functors can be nicely lifted using the theory developed so far, there are other functors that require a more general treatment. For instance, consider the endofunctor $F = B \times \_$ (left product with B) for some fixed B. Notice that for $t_1, t_2 \in FX = B \times X$ with $t_i = (b_i, x_i)$ a coupling exists iff $b_1 = b_2$. As a consequence, when $b_1 \neq b_2$, irrespectively 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 $t_2 = (\varepsilon, x)$ for a small $\varepsilon > 0$ and an $x \in X$. The reason is that we think of $B = [0,1]$ as 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 indeed formalized by considering the lifting of the product seen as a functor from $\mathbf{Set} \times \mathbf{Set}$ into $\mathbf{Set}$. More generally, it can be seen that the definitions and results introduced so far for endofunctors on $\mathbf{Set}$ straightforwardly extend to multifunctors on $\mathbf{Set}$, namely functors $F\colon \mathbf{Set}^n \to \mathbf{Set}$ on the product category $\mathbf{Set}^n$ for a natural number $n \in \mathbb{N}$. For ease of presentation we will not spell out the details here (they can be found in [2]), but just provide an important example of a bifunctor (i.e. $n = 2$).

▶ **Example 2.10** (Product Bifunctor). The weak pullback preserving product bifunctor $F\colon \mathbf{Set}^2 \to \mathbf{Set}$ maps two sets X_1, X_2 to $F(X_1,X_2) = X_1 \times X_2$ and two functions $f_i\colon X_i \to Y_i$ to the function $F(f_1,f_2) = f_1 \times f_2$. In this paper we will use the well-behaved evaluation functions $ev_F\colon [0,1]^2 \to [0,1]$ presented in the table below. Therein we also list the pseudometric $(d_1,d_2)^F\colon (X_1 \times X_2)^2 \to [0,\top]$ we obtain for pseudometric spaces (X_1,d_1), (X_2,d_2).

Parameters	$ev_F(r_1,r_2)$	$(d_1,d_2)^F((x_1,x_2),(y_1,y_2))$
$c_1, c_2 \in (0,1]$	$\max\{c_1r_1, c_2r_2\}$	$\max\{c_1d_1(x_1,y_1), c_2d_2(x_2,y_2)\}$
$c_1, c_2 \in (0,1], c_1 + c_2 \leq 1$	$c_1x_1 + c_2x_2$	$c_1d_1(x_1,y_1) + c_2d_2(x_2,y_2)$

For $c_1 = c_2 = 1$, the first evaluation map yields exactly the categorical product in $\mathbf{PMet}$. In both cases the Kantorovich-Rubinstein duality holds and the supremum [infimum] of the Kantorovich [Wasserstein] pseudometric is always a maximum [minimum].

3 Compositionality for the Wasserstein Lifting

Our first step is to study compositionality of functor liftings, i.e., we identify some sufficient conditions ensuring $\overline{F}\,\overline{G} = \overline{FG}$. This technical result will be often very useful since it allows us to reason modularly and, consequently, to simplify the proofs needed in the treatment of our examples. We will explicitly only consider the Wasserstein approach which is the one employed in all the examples of this paper.

Given evaluation functions ev_F and ev_G, we can easily construct an evaluation function for the composition FG by defining $ev_{FG} := \widetilde{F}ev_G = ev_F \circ Fev_G$. Our first observation is that, whenever F and G preserve weak pullbacks, well-behavedness is inherited.

▶ **Proposition 3.1** (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 $ev_{FG} = ev_F \circ Fev_G$ is well-behaved.*

In the light of this result and the fact that FG certainly preserves weak pullbacks if F and G do, we can safely use the Wasserstein lifting for FG. A sufficient criterion for compositionality is the existence of optimal couplings for G.

▶ **Proposition 3.2** (Compositionality). *Let F, G be weak pullback preserving endofunctors on* **Set** *with well-behaved evaluation functions ev_F, ev_G and let (X, d) be a pseudometric space. Then $d^{\downarrow FG} \geq (d^{\downarrow G})^{\downarrow F}$. Moreover, if for all $t_1, t_2 \in GX$ there is an optimal G-coupling, i.e. $\gamma(t_1, t_2) \in \Gamma_G(t_1, t_2)$ such that $d^{\downarrow G}(t_1, t_2) = \widetilde{G}d(\gamma(t_1, t_2))$, then $d^{\downarrow FG} = (d^{\downarrow G})^{\downarrow F}$.*

This criterion will turn out to be very useful for our later results. Nevertheless it provides just a sufficient condition for compositionality as the next example shows.

▶ **Example 3.3.** We consider the finite powerset functor $\mathcal{P}_{\mathit{fin}}$ of Example 2.8 and the distribution functor $\mathcal{D}$ of Example 2.9 with their evaluation functions. Let (X, d) be a pseudometric space.

1. We have $d^{\downarrow \mathcal{D}\mathcal{D}} = \left(d^{\downarrow \mathcal{D}}\right)^{\downarrow \mathcal{D}}$, by Proposition 3.2, because optimal couplings always exist.
2. We have $d^{\downarrow \mathcal{P}_{\mathit{fin}}\mathcal{P}_{\mathit{fin}}} = \left(d^{\downarrow \mathcal{P}_{\mathit{fin}}}\right)^{\downarrow \mathcal{P}_{\mathit{fin}}}$ although $\mathcal{P}_{\mathit{fin}}$-couplings do not always exist.

Note that when we lift the functor $\mathcal{P}_{\mathit{fin}}$ we do not have couplings in the case when we determine the distance between an empty set $\emptyset$ and a non-empty set $Y \subseteq X$, since there exists no subset of $X \times X$ that projects to both.

Compositionality can be defined analogously for multifunctors. Again, we will not spell this out completely but we will use it to obtain the machine bifunctor. Before we can do that, we first need to define another endofunctor.

▶ **Example 3.4** (Input Functor). Let A be a fixed finite set of inputs. The input functor $F = \_^A \colon \mathbf{Set} \to \mathbf{Set}$ maps a set X to the exponential X^A and a function $f \colon X \to Y$ to $f^A \colon X^A \to Y^A$, $f^A(g) = f \circ g$. This functor preserves weak pullbacks. The two evaluation functions listed below are well-behaved and yield the given Wasserstein pseudometric on X^A for any pseudometric space (X, d).

$ev_F(s)$	$d^{\downarrow F}(s_1, s_2)$
$\max_{a \in A} s(a)$	$\max_{a \in A} d\big(s_1(a), s_2(a)\big)$
$\sum_{a \in A} s(a)$	$\sum_{a \in A} d\big(s_1(a), s_2(a)\big)$

By composing this functor with the product bifunctor we obtain the machine bifunctor which we will use to obtain trace semantics.

▶ **Example 3.5** (Machine Bifunctor). Let A be a finite set of inputs, $I = \_^A$ the input functor of Example 3.4, Id the identity endofunctor on **Set** and P be the product bifunctor of Example 2.10. The *machine bifunctor* is the composition $M := P \circ (\mathrm{Id} \times I)$ i.e. the bifunctor $M \colon \mathbf{Set}^2 \to \mathbf{Set}$ with $M(B, X) := B \times X^A$. Since for Id and I there are unique (thus optimal) couplings we have compositionality. Depending on the choices of evaluation function for P and I (for Id we always take $\mathrm{id}_{[0,1]}$) we obtain the following well-behaved evaluation functions $ev_M \colon [0, 1] \times [0, 1]^A \to [0, 1]$.

Parameters	$ev_P(r_1, r_2)$	$ev_I(s)$	$ev_M(o, s)$
$c_1, c_2 \in (0, 1]$	$\max\{c_1 r_1, c_2 r_2\}$	$\max_{a\in A} s(a)$	$\max\left\{c_1 o, c_2 \max_{a\in A} s(a)\right\}$
$c_1, c_2 \in (0, 1], c_1 + c_2 \leq 1$	$c_1 x_1 + c_2 x_2$	$\lvert A\rvert^{-1} \sum_{a\in A} s(a)$	$c_1 o + c_2 \lvert A\rvert^{-1} \sum_{a\in A} s(a)$.

Let (B, d_B), (X, d) be pseudometric spaces. For any $t_1, t_2 \in M(B, X)$ with $t_i = (b_i, s_i) \in B \times X^A$ there is a unique and therefore necessarily optimal coupling $t := (b_1, b_2, \langle s_1, s_2\rangle)$. Depending on the evaluation function, we obtain for the first case $(d_B, d)^{\downarrow M}(t_1, t_2) = \max\{c_1 d_B(b_1, b_2), c_2 \cdot \max_{a\in A} d(s_1(a), s_2(a))\}$ and for the second case $(d_B, d)^{\downarrow M}(t_1, t_2) = c_1 d_B(b_1, b_2) + c_2 |A|^{-1} \sum_{a\in A} d(s_1(a), s_2(a))$.

Usually we will fix the first argument (the set of outputs) of the machine bifunctor and consider the obtained machine endofunctor $M_B := M(B, \_)$. However, for the same reasons as explained above for the product bifunctor, we need to consider it as bifunctor. One notable exception is the case where $B = 2$, endowed with the discrete metric. Then we have the following result.

▶ **Example 3.6.** Consider the machine endofunctor $M_2 := M(2, \_) = 2 \times \_^A$ with evaluation function $ev_{M_2} \colon 2 \times [0, 1]^A, (o, s) \mapsto c \cdot ev_I(s)$ where $c \in (0, 1]$ and ev_I is one of the evaluation functions for the input functor from Example 3.4. If d_2 is the discrete metric on 2 and $c = c_2$ (where c_2 is the parameter for the evaluation function of the machine bifunctor as in Example 3.5) then the pseudometric obtained via the bifunctor lifting coincides with the one obtained by endofunctor lifting i.e. for all pseudometric spaces (X, d) we have $(d_2, d)^{\downarrow M} = d^{\downarrow M_2}$. Moreover, although couplings for M_2 do not always exist we have $d^{\downarrow \mathcal{P}_{fin} M_2} = \left(d^{\downarrow M_2}\right)^{\downarrow \mathcal{P}_{fin}}$.

4 Lifting of Natural Transformations and Monads

Recall that a monad on an arbitrary category $\mathbf{C}$ is a triple (T, η, μ) where $T \colon \mathbf{C} \to \mathbf{C}$ is an endofunctor and $\eta \colon \mathrm{Id} \Rightarrow T$, $\mu \colon T^2 \Rightarrow T$ are natural transformations called *unit* (η) and *multiplication* (μ) such that the two diagrams below commute.

$$\begin{tikzcd} T \arrow[r, "\eta T"] \arrow[dr, equal] & T^2 \arrow[d, "\mu"] & T \arrow[l, "T\eta"'] \arrow[dl, equal] \\ & T & \end{tikzcd} \qquad \begin{tikzcd} T^3 \arrow[r, "\mu T"] \arrow[d, "T\mu"'] & T^2 \arrow[d, "\mu"] \\ T^2 \arrow[r, "\mu"] & T \end{tikzcd}$$

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.

▶ **Proposition 4.1** (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. Then the following holds for all pseudometric spaces (X, d). For the Kantorovich lifting:*

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

while for the Wasserstein lifting:

3. *If $ev_G \circ \lambda_{[0,\top]} \le ev_F$ then $d^{\downarrow G} \circ (\lambda_X \times \lambda_X) \le d^{\downarrow F}$, i.e. λ_X is nonexpansive.*
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 $d^{\downarrow G} \circ (\lambda_X \times \lambda_X) = d^{\downarrow F}$, i.e. λ_X is an isometry.*

In the rest of the 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 4.2** (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 $ev_T \circ \eta_{[0,\top]} \le \mathrm{id}_{[0,\top]}$ 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 $d^T \in \{d^{\uparrow T}, d^{\downarrow T}\}$. If $ev_T \circ \mu_{[0,\top]} \le ev_T \circ Tev_T$ and compositionality holds for TT, i.e. $(d^T)^T = d^{TT}$, then μ is nonexpansive, i.e. $d^T \circ (\mu_X \times \mu_X) \le (d^T)^T$. This yields the multiplication $\overline{\mu}\colon \overline{T}\,\overline{T} \Rightarrow \overline{T}$ in **PMet**.*

We conclude this section with two examples of liftable monads.

▶ **Example 4.3** (Finite Powerset Monad). The finite powerset functor $\mathcal{P}_{\mathit{fin}}$ of Example 2.8 can be seen as a monad, with unit η consisting of the functions $\eta_X\colon X \to \mathcal{P}_{\mathit{fin}}X$, $\eta_X(x) = \{x\}$ and multiplication given by $\mu_X\colon \mathcal{P}_{\mathit{fin}}\mathcal{P}_{\mathit{fin}}X \to \mathcal{P}_{\mathit{fin}}X$, $\mu_X(S) = \cup S$. We show that 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 $\mathcal{S} \in \mathcal{P}_{\mathit{fin}}(\mathcal{P}_{\mathit{fin}}[0,\top])$ we have $ev_T \circ \mu_{[0,1]}(\mathcal{S}) = \max \cup \mathcal{S} = \max \cup_{S \in \mathcal{S}} S$ and $ev_T \circ Tev_T(\mathcal{S}) = \max(ev_T[\mathcal{S}]) = \max\{\max S \mid S \in \mathcal{S}\}$ and thus both values coincide. Moreover, we recall from Example 3.3.2 that we have compositionality for $\mathcal{P}_{\mathit{fin}}\mathcal{P}_{\mathit{fin}}$. Therefore, by Corollary 4.2 η is an isometry and μ nonexpansive.

▶ **Example 4.4** (Distribution Monad). The probability distribution functor $\mathcal{D}$ of Example 2.9 can be seen as a monad: the unit η consists of the functions $\eta_X\colon X \to \mathcal{D}X$, $\eta_X(x) = \delta_x^X$ where δ_x^X is the Dirac distribution and the multiplication is given by $\mu_X\colon \mathcal{D}\mathcal{D}X \to \mathcal{D}X$, $\mu_X(P) = \lambda x. \sum_{q \in \mathcal{D}X} P(q) \cdot q(x)$. We consider its Wasserstein lifting. Since $[0,1] = \mathcal{D}2$ we can see that $ev_{\mathcal{D}} = \mu_2$. Using this fact and the monad laws we have $ev_{\mathcal{D}} \circ \eta_{[0,1]} = \mu_2 \circ \eta_{\mathcal{D}2} = \mathrm{id}_{\mathcal{D}X} = \mathrm{id}_{[0,1]}$ and also $ev_{\mathcal{D}} \circ \mu_{[0,1]} = \mu_2 \circ \mu_{\mathcal{D}2} = \mu_2 \circ \mathcal{D}\mu_2 = ev_{\mathcal{D}} \circ \mathcal{D}ev_{\mathcal{D}}$. Moreover, since we always have optimal couplings, we have compositionality for $\mathcal{D}\mathcal{D}$ by Proposition 3.2. Thus by Corollary 4.2 η is an isometry and μ nonexpansive.

5 Trace Metrics in Eilenberg-Moore

As mentioned in the introduction, trace semantics can be characterized by means of coalgebras either over Kleisli [17, 11] or over Eilenberg-Moore [20, 14] categories. We focus on the latter approach. We first recall the basic notions of Eilenberg-Moore algebras and distributive laws, and discuss how the results in the paper can be used to "lift" the associated determinization construction. This is then applied to derive trace metrics for nondeterministic automata and probabilistic automata, by relying on suitable liftings of the machine functor.

5.1 Generalized Powerset Construction

An *Eilenberg-Moore algebra* for a monad (T, η, μ) is a $\mathbf{C}$-arrow $a\colon TA \to A$ making the left and middle diagram below commute. Given two such algebras $a\colon TA \to A$ and $b\colon TB \to B$, a morphism from a to b is a $\mathbf{C}$ arrow $f\colon A \to B$ making the right diagram below commute.

$$\begin{tikzcd} A \arrow[r, "\eta_A"] \arrow[dr, equal] & TA \arrow[d, "a"] \\ & A \end{tikzcd} \qquad \begin{tikzcd} T^2A \arrow[r, "\mu_A"] \arrow[d, "Tc"'] & TA \arrow[d, "a"] \\ TA \arrow[r, "a"'] & A \end{tikzcd} \qquad \begin{tikzcd} TA \arrow[r, "Tf"] \arrow[d, "a"'] & TB \arrow[d, "b"] \\ A \arrow[r, "f"'] & B \end{tikzcd}$$

Eilenberg-Moore algebras and their morphisms form a category denoted by $\mathcal{EM}(T)$. A functor $\widehat{F}\colon \mathcal{EM}(T) \to \mathcal{EM}(T)$ is called a *lifting* of $F\colon \mathbf{C} \to \mathbf{C}$ to $\mathcal{EM}(T)$ if $U^T\widehat{F} = FU^T$, with $U^T\colon \mathcal{EM}(T) \to \mathbf{C}$ the forgetful functor. A natural transformation $\lambda\colon TF \Rightarrow FT$ is an $\mathcal{EM}$*-law* (also called *distributive law*) if it satisfies:

$$\begin{tikzcd} F \arrow[r, equal] \arrow[d, "\eta F"'] & F \arrow[d, "F\eta"] \\ TF \arrow[r, "\lambda"'] & FT \end{tikzcd} \qquad \begin{tikzcd} T^2F \arrow[r, "T\lambda"] \arrow[d, "\mu F"'] & TFT \arrow[r, "\lambda T"] & FT^2 \arrow[d, "F\mu"] \\ TF \arrow[rr, "\lambda"'] & & FT \end{tikzcd}$$

Liftings and $\mathcal{EM}$-laws are related by the following folklore result (see e.g. [13]).

▶ **Proposition 5.1.** *There is a bijective correspondence between $\mathcal{EM}$-laws and liftings to $\mathcal{EM}$-categories.*

$\mathcal{EM}$-laws and liftings are crucial to characterize trace semantics via coalgebras. Given a coalgebra $c\colon X \to FTX$, for a functor F and a monad (T, η, μ) such that there is a distributive law $\lambda\colon TF \Rightarrow FT$, one can build an F-coalgebra as

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

If there exists a final F-coalgebra $\omega\colon \Omega \to F\Omega$, one can define a semantic map for the FT-coalgebra c into Ω. First let $[\![-]\!]\colon TX \to \Omega$ be the unique coalgebra morphism from $c^\sharp$. Then take the map $[\![-]\!] \circ \eta\colon X \to \Omega$.

$$\begin{tikzcd} X \arrow[r, "\eta"] \arrow[d, "c"'] & TX \arrow[r, "{[\![-]\!]}"] \arrow[dl, "c^\sharp"] & \Omega \arrow[d, "\omega"] \\ FTX \arrow[rr, "{F[\![-]\!]}"'] & & T\Omega \end{tikzcd}$$

One can readily check that $c^\sharp$ is an algebra map from the T-algebra μ_X to $\widehat{F}\mu_X$, namely it is an $\widehat{F}$-coalgebra or, equivalently, a λ-*bialgebra* [21, 15]. Similarly for ω, Ω carries a T-algebra structure obtained by finality and hence the final F-coalgebra ω can be lifted in order to obtain the final $\widehat{F}$-coalgebra (see [13, Prop. 4]).

This result holds for arbitrary categories and, in particular, we can reuse it for our setting: we only need an $\mathcal{EM}$-law on $\mathbf{PMet}$. Note that Proposition 4.1 not only provides sufficient conditions for monad liftings but also can be exploited to lift $\mathcal{EM}$-laws. Indeed the additional commutativity requirements for $\mathcal{EM}$-laws trivially hold when all components are nonexpansive.

▶ **Corollary 5.2** (Lifting of an $\mathcal{EM}$-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 an $\mathcal{EM}$-law. If the evaluation functions satisfy $ev_G \circ Gev_F \circ \lambda_{[0,\top]} \leq ev_F \circ Fev_G$ and compositionality holds for FG, then λ is nonexpansive and hence $\overline{\lambda}\colon \overline{F}\,\overline{G} \Rightarrow \overline{G}\,\overline{F}$ is also an $\mathcal{EM}$-law.*

We will now consider $\mathcal{EM}$-laws for nondeterministic and probabilistic automata. In the first case, T is the powerset monad $\mathcal{P}_{\mathit{fin}}$ and F is the machine functor $M_2 = 2 \times \_^A$, while in the second case T is the distribution monad $\mathcal{D}$ and F is the machine functor $M_{[0,1]} = [0,1] \times \_^A$. Note however that while in the first case Corollary 5.2 is directly applicable, this is not true in the second case, since we need to deal with multifunctors.

▶ **Example 5.3** ($\mathcal{EM}$-law for Nondeterministic Automata)**.** Let $(\mathcal{P}_{\mathit{fin}}, \eta, \mu)$ be the finite powerset monad from Example 4.3. The $\mathcal{EM}$-law $\lambda\colon \mathcal{P}_{\mathit{fin}}(2 \times \_^A) \Rightarrow 2 \times \mathcal{P}_{\mathit{fin}}(\_)^A$ is defined, for any set X, as

$$\lambda_X(S) = \big(o, \lambda a \in A.\, \{s'(a) \mid (o', s') \in S\}\big), \quad \text{where} \quad o = \begin{cases} 1 & \exists s' \in X^A.(1, s') \in S \\ 0 & \text{else} \end{cases}.$$

This is exactly the one exploited for the standard powerset construction from automata theory [20]. Indeed, for a nondeterministic automaton $c\colon X \to 2 \times \mathcal{P}_{\mathit{fin}}(X)^A$, the map $[\![-]\!] \circ \eta_X$ assigns to each state its accepted language. Corollary 5.2 ensures that it is nonexpansive (see the extended version [`arXiv:1505.08105`] for a detailed proof).

▶ **Example 5.4** ($\mathcal{EM}$-law for Probabilistic Automata)**.** Let $(\mathcal{D}, \eta, \mu)$ be the distribution monad from Example 4.4 and M be the machine bifunctor from Example 3.5. There is a known [20] $\mathcal{EM}$-law $\lambda\colon \mathcal{D}([0,1] \times \_^A) \Rightarrow [0,1] \times \mathcal{D}^A$ given by the assignment

$$\lambda_X(P) = \left(\sum_{r \in [0,1]} r \cdot P(r, X^A), \lambda a \in A. \lambda x \in X. \sum_{s \in X^A,\, s(a) = x} P([0,1], s) \right)$$

Also this $\mathcal{EM}$-law is nonexpansive, as shown in the extended version [`arXiv:1505.08105`].

Any FT-coalgebra $c\colon X \to FTX$ can always be regarded as an $\overline{F}\,\overline{T}$-coalgebra by equipping X with the discrete metric assigning $\top$ to non equal states (in this way, c is trivially nonexpansive). The consequence of the nonexpansiveness of the $\mathcal{EM}$-law λ is the following: the "generalized determinization" procedure for nondeterministic and probabilistic automata can now be lifted to pass from $\overline{F}\,\overline{T}$-coalgebras to $\widehat{\overline{F}}$-coalgebras in $\mathcal{EM}(\overline{T})$ by using the upper adjunction in the diagram below (analogously to [13, 14]).

$$\begin{tikzcd}
\mathbf{PMet} \arrow[loop left, "\overline{F}"] \arrow[rr, bend left, "L^{\overline{T}}"] \arrow[d, "U"'] & & \mathcal{EM}(\overline{T}) \arrow[loop right, "\widehat{\overline{F}}"] \arrow[ll, bend left, "U^{\overline{T}}"] \arrow[d, "V"] \\
\mathbf{Set} \arrow[loop left, "F"] \arrow[rr, bend left, "L^{T}"] & & \mathcal{EM}(T) \arrow[loop right, "\widehat{F}"] \arrow[ll, bend left, "U^{T}"]
\end{tikzcd}$$

Since we can also lift the final $\overline{F}$-coalgebra to $\mathcal{EM}(\overline{T})$, we can use it to define trace distance. This procedure is detailed in the next section.

5.2 Final Coalgebra for the Lifted Machine Functor

If we fix the first component of the machine bifunctor M on **Set** we obtain an endofunctor $M_B\colon \mathbf{Set} \to \mathbf{Set}$, $M_B(X) = B \times \_^A$. It is known [16] that the final coalgebra for this functor is $\kappa\colon B^{A^*} \to B \times (B^{A^*})^A$ with $\kappa(t) = (t(\varepsilon), \lambda a \in A.\lambda w \in A^*.t(aw))$. We employ an analogous construction with our lifted machine bifunctor $\overline{M}$ on **PMet**, i.e. we fix a pseudometric space (B, d_B) of outputs and consider coalgebras of the functor $\overline{M}_{(B,d_B)} := \overline{M}((B, d_B), \_)$. To obtain the final coalgebra for this functor in **PMet**, we use the following result from [2].

▶ **Proposition 5.5** ([2, Thm. 6.1]). *Let $\overline{F}\colon \mathbf{PMet} \to \mathbf{PMet}$ be a lifting of a functor $F\colon \mathbf{Set} \to \mathbf{Set}$ which has a final coalgebra $\kappa\colon \Omega \to F\Omega$. For every ordinal i we construct a pseudometric $d_i\colon \Omega \times \Omega \to [0, \top]$ as follows: $d_0 := 0$ is the zero pseudometric, $d_{i+1} := d_i^F \circ (\kappa \times \kappa)$ for all ordinals i and $d_j = \sup_{i<j} d_i$ for all limit ordinals j. This sequence converges for some ordinal θ, i.e., $d_\theta = d_\theta^F \circ (\kappa \times \kappa)$. Moreover $\kappa\colon (\Omega, d_\theta) \xrightarrow{1} (F\Omega, d_\theta^F)$ is the final $\overline{F}$-coalgebra.*

It is hence enough to do fixed-point iteration for the functor F on the determinized state set TX in order to obtain trace distance. The lifted monad is ignored at this stage, but its lifting is of course necessary to establish the Eilenberg-Moore category and its adjunction.

We now consider our two examples, where in both cases F is the machine functor M_B (for two different choices of B):

▶ **Example 5.6** (Final Coalgebra Pseudometric). Let M be the machine bifunctor.

1. We start with nondeterministic automata where the output set is $B = 2$ and we use the discrete metric d_2 as distance on 2 as in Example 3.6. As maximal distance we take $\top = 1$ and as evaluation function we use $ev_M(o, s) = c \cdot \max_{a\in A} s(a)$ for $0 < c < 1$.
   For any pseudometric d on 2^{A^*} – the carrier of the final M_2-coalgebra – we know that for elements $(o_1, s_1), (o_2, s_2) \in 2 \times (2^{A^*})^A$ we have the Wasserstein pseudometric $d^{\downarrow F}\big((o_1, s_1), (o_2, s_2)\big) = \max\big\{d_2(o_1, o_2), c \cdot \max_{a\in A} d\big(s_1(a), s_2(a)\big)\big\}$. Thus the fixed-point equation from Proposition 5.5 is, for $L_1, L_2 \in 2^{A^*}$,

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

   Now because d_2 is the discrete metric with $d_2(0, 1) = 1$ we 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.

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

   A determinized coalgebra has as carrier set sets of states $\mathcal{P}(X)$. Each of these sets is mapped to the language that it accepts and 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 $c^{|w|}$. This corresponds to the standard ultrametric on words.
2. Next we consider probabilistic automata where $B = [0, 1]$ equipped with the standard Euclidean metric d_e.
   Furthermore the remaining parameters are set as follows: let $\top = 1$ and the evaluation function is $ev_M(o, s) = c_1 o + c_2|A|^{-1} \sum_{a\in A} s(a)$ for $c_1, c_2 \in (0, 1)$ such that $c_1 + c_2 \leq 1$ as in Example 3.5. This time, the machine functor must be lifted as a bifunctor in order to obtain the appropriate distance (cf. the discussion before Example 2.10).

For any pseudometric d on $[0,1]^{A^*}$ we know that for $(r_1,s_1),(r_2,s_2) \in [0,1]\times([0,1]^{A^*})^A$ we have $d^{\downarrow F}((r_1,s_1),(r_2,s_2)) = c_1|r_1-r_2| + \frac{c_2}{|A|}\cdot\sum_{a\in A} d(s_1(a),s_2(a))$. Thus the fixed-point equation from Proposition 5.5 is, for $p_1,p_2 \in [0,1]^{A^*}$:

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

It is again easy to see that $d_{[0,1]^{A^*}} \colon [0,1]^{A^*}\times[0,1]^{A^*} \to [0,1]$ as presented below is the least fixed-point of this equation and therefore $([0,1]^{A^*}, d_{[0,1]^{A^*}})$ the carrier of the final $\overline{M}_{([0,1],d_e)}$-coalgebra.

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

Here, a determinized coalgebra has as carrier distributions on states $\mathcal{D}(X)$. Each such distribution is mapped to a function $p\colon A^* \to [0,1]$ assigning numerical values to words. Then the distance, which can be thought of as a form of total variation distance with discount, is computed by the above formula.
If instead of working in the interval $[0,1]$ we use $[0,\infty]$ with $\top=\infty$, we can drop the conditions $c_1,c_2<1$ and $c_1+c_2\le 1$. In this case we may set $c_2 := |A|$ and $c_1 := 1/2$ and then the above distance is equal to the total variation distance, i.e.,

$$d_{[0,\infty]^{A^*}}(p_1,p_2) = \frac{1}{2}\cdot\sum_{w\in A^*}|p_1(w)-p_2(w)|\ .$$

6 Conclusion, Related and Future Work

In the last years, an impressive amount of papers has studied behavioral distances for both probabilistic and nondeterministic systems (see, e.g., [10, 7, 23, 1, 5, 6, 8]). The necessity of a general understanding of such metrics is not a mere intellectual whim but it is perceived also by researchers exploiting distances for differential privacy and quantitative information flow (see for instance [4]). As far as we know, the first use of coalgebras for this purpose dates back to [23], where the authors consider systems and distance for a fixed endofunctor on **PMet**. In [2], we introduced the Kantorovich and Wasserstein approaches as a general way to define "canonical liftings" to **PMet** and behavioral distances by finality. These are usually branching-time, while many properties of interest for applications (see again [4]) are usually expressed by means of distances on set of traces. In this paper, we have shown that the work developed in [2] can be fruitfully combined with [14] to obtain various trace distances.

Among the several trace distances introduced in literature, it is worth to mention [1, 5, 6, 8]. Similar to the trace distance we obtain in Example 5.6 for probabilistic automata is the one introduced in [1] for Semi-Markov chains with residence time. In [5, 6], both branching-time and linear-time distances are introduced for *metric transition systems*, namely Kripke structures where states are associated with elements of a fixed (pseudo-)metric space M, that would correspond to coalgebras of the form $X \to M\times\mathcal{P}(X)$. In [2], we have shown an example capturing branching-time distance for metric transition systems, but for linear distances we require a distributive law of the form $\mathcal{P}(M\times\_) \Rightarrow M\times\mathcal{P}(\_)$, for which we would need at least M carrying an algebra for the monad $\mathcal{P}$. We also plan to investigate trace metrics in a Kleisli setting [11], where it might be easier to incorporate such examples.

There are two other direct consequences of our work that we did not explain in the main text, but that are important properties of the distances that we obtain (and, indeed, are mentioned in [4] amongst the desiderata for "good" metrics). First, the behavioral branching-distance for $\overline{F}\,\overline{T}$ provides an upper bound to the linear-distance $\overline{F}$, analogously to the well-known fact that bisimilarity implies trace equivalence. To see this, it is enough to observe that there is a functor from the category of $\overline{F}\,\overline{T}$-coalgebras to the one of $\overline{F}$-coalgebras mapping $c\colon X \to \overline{F}\,\overline{T}X$ into $c^\sharp\colon \overline{T}X \to \overline{F}\,\overline{T}X$.

Second, since the final map $[\![-]\!]$ is a morphism in $\mathcal{EM}(\overline{T})$, the behavioral distance for $\overline{F}$ is nonexpansive w.r.t. the operators of the monad $\overline{T}$. Nonexpansiveness with respect to some operators is a desirable property which has been studied, for instance in [7], as a generalization of the notion of being a congruence for behavioral equivalence. Several researchers are now studying syntactic rule formats ensuring this and other sorts of compositionality (see e.g. [9] and the references therein) and we believe that our Corollary 5.2 may provide some helpful insights.

In this perspective, however, our results are still unsatisfactory if compared to what happens in the case of behavioral equivalences. From a fibrational point of view, one has a *canonical lifting* to **Rel** (the category of relations and relation preserving morphisms) such that compositionality holds on the nose and distributive laws always lift [12, Exercise 4.4.6]. The forgetful functor $U\colon \mathbf{PMet} \to \mathbf{Set}$ is also a fibration [2], but Kantorovich and Wasserstein liftings are not always so well-behaved. Fibrations might be useful also to guarantee soundness of up-to techniques [3] for behavioral distances that, hopefully, will lead to more efficient proofs and algorithms.

Another interesting future work would be to show that Kantorovich and Wasserstein liftings arise from some universal properties, i.e., that they are the smallest and largest metric in some continuum of metrics with certain properties. Here we would like to draw inspiration from [22] which characterizes the Giry monad via a universal property on monad morphisms.

Finally, we would like to have an abstract understanding of the Kantorovich-Rubinstein duality. Preliminary attempts suggest that this is very difficult: indeed the proof for the probabilistic case relies on specific properties of distributions.

References

1 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.

2 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.

3 Filippo Bonchi, Daniela Petrisan, Damien Pous, and Jurriaan Rot. Coinduction up to in a fibrational setting. In *Proc. of CSL-LICS '14*, 2014.

4 Konstantinos Chatzikokolakis, Daniel Gebler, Catuscia Palamidessi, and Lili Xu. Generalized bisimulation metrics. In *CONCUR 2014 – Concurrency Theory*, Lecture Notes in Computer Science, pages 32–46, 2014.

5 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.

6 Luca de Alfaro, Marco Faella, and Mariëlle Stoelinga. Linear and branching system metrics. *IEEE Transactions on Software Engineering*, 25(2), 2009.

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

8 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.

9 Daniel Gebler and Simone Tini. Compositionality of approximate bisimulation for probabilistic systems. In *Proc. of EXPRESS/SOS '13*, pages 32–46, 2013.

10 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.

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

12 Bart Jacobs. Introduction to coalgebra. Towards mathematics of states and observations, 2012. Draft.

13 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.

14 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).

15 Bartek Klin. Bialgebras for structural operational semantics: An introduction. *Theoretical Computer Science*, 412(38):5043–5069, 2011.

16 E.G. Manes and M.A. Arbib. Algebraic approaches to program semantics. *Texts and Monographs in Computer Science*, 1986.

17 John Power and Daniele Turi. A coalgebraic foundation for linear time semantics. *Electronic Notes in Theoretical Computer Science*, 29:259–274, 1999.

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

19 Lutz Schröder. Expressivity of coalgebraic modal logic: The limits and beyond. *Theoretical Computer Science*, 390(2–3):230–247, 2008.

20 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.

21 Daniele Turi and Gordon D. Plotkin. Towards a mathematical operational semantics. In *Proc. of LICS '97*, pages 280–291, 1997.

22 Franck van Breugel. The metric monad for probabilistic nondeterminism, April 2005.

23 Franck van Breugel and James Worrell. A behavioural pseudometric for probabilistic transition systems. *Theoretical Computer Science*, 331:115–142, 2005.

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

25 Cédric Villani. *Optimal Transport – Old and New*, volume 338 of *A Series of Comprehensive Studies in Mathematics.* Springer, 2009.