Compact Quantitative Theories of Convex Algebras

Matteo Mio[a,1]

[a] *CNRS & ENS de Lyon, France*

Abstract

We introduce the concept of compact quantitative equational theory. A quantitative equational theory is defined to be compact if all its consequences are derivable by means of finite proofs. We prove that the theory of interpolative barycentric (also known as convex) quantitative algebras of Mardare et. al. is compact. This serves as a paradigmatic example, used to obtain other compact quantitative equational theories of convex algebras, each axiomatizing some distance on finitely supported probability distributions.

Keywords: Quantitative Algebra, probability distributions, convex algebras, finite proofs.

1 Introduction

At the core of universal algebra is a sound and complete deductive system, due to Birkhoff (see, for example, [33, §3.2]), which allows to derive judgments of the form $E \vdash s = t$, where E is a set of equations, with the meaning 'every algebra that satisfy all equations in E also satisfies the equation $s = t$'. Since Birkhoff's proof system can be seen as a fragment of the deductive apparatus of classical first order logic, every proof is a finite (both in width and depth) tree. This is a key property, and it is fundamental when it comes to applications of universal algebra in computer science: proofs, being finite objects, can be represented and manipulated by machines, they can be enumerated and there are effective procedures to check if a given derivation tree is a valid proof or not.

Quantitative algebra, recently introduced in [22], is an active area of investigation [1,2,5,6,7,8,15,16,23,29,27,26,25,28,21,20,24,34,35,17,11,13] aimed at extending the methods of universal algebra to the study of so-called *quantitative algebras*: algebraic structures $\mathbb{A} = (A, \{op^{\mathbb{A}}\}_{op\in\Sigma})$ further endowed with an extended metric distance $d_A : A \times A \to [0, \infty]$ compatible with the interpretations $op^{\mathbb{A}}$ of the operations $op \in \Sigma$. The key novel concept is that of quantitative equation $s =_\epsilon t$ between terms, for $\epsilon \in [0, \infty)$, which asserts that the distance between s and t is bounded above by ϵ ($d_A(s, t) \leq \epsilon$).

In [22], a sound and complete proof system, analogous to that of Birkhoff, has been investigated for deriving judgments [2] of the form $E \vdash s =_\epsilon t$ that involve quantitative equations *in lieu* of equations. However, unlike Birkhoff's system, the deductive system for quantitative algebras is based on the infinitary first order logic $L_{\omega_1,\omega}$ [18] which, crucially, allows countable conjuctions. As a consequence, a proof in

[1] Email: matteo.mio@ens-lyon.fr

[2] This is a simplification sufficient for the purpose of this introduction. More precisely, the judgments of the proof system are called *quantitative inferences* [22] and are Horn implications involving quantitative equations. Precise definitions are given in Section 2.2.

quantitative algebra is, in general, not a finite object: it is a well founded tree with countable width. More specifically, the key source of infinite width in proofs comes from the presence of the following rule:

$$\frac{E \vdash s =_{\epsilon_0} t \qquad E \vdash s =_{\epsilon_1} t \qquad \ldots \qquad E \vdash s =_{\epsilon_n} t \qquad \ldots}{E \vdash s =_{\epsilon} t} \; \epsilon = \inf\{\epsilon_i\}_{i \in \mathbb{N}}$$

which states that if the distance between s and t is bounded above by ϵ_i (i.e. $d_A(s,t) \leq \epsilon_i$), for all $i \in \mathbb{N}$, then necessarily the distance is also bounded by the infimum of all bounds ($d_A(s,t) \leq \epsilon$).

The non-finiteness of proofs in quantitative equational logic is an obstacle to mechanization, of course. This is unavoidable: it can be shown that the presence of an infinitary rule is required in any sound and complete proof system for quantitative algebra.

Compact Quantitative Theories

On the other hand, it is still possible for some well-behaved quantitative theories E, to satisfy the following property (which is stated precisely in Section 3):

If $E \vdash s =_{\epsilon} t$ is derivable, then it is derivable by a finite proof not using the infinitary rule.

We refer to such quantitative equational theories E as *compact*.

Perhaps surprisingly, several of the main examples of quantitative equational theories that have appeared in the literature turn out to be compact although, to the best of our knowledge, this has never been explicitly observed. One such example, which we fully investigate and generalize in this work, is given by the quantitative theory of *interpolative barycentric algebras* of [22, §10] which axiomatizes the Kantorovich (also known as 1-Wasserstein [32, Ch 6]) metric on finitely supported probability distributions on a metric space. Other examples include the theory of quantitative semilattices [22, §10], axiomatizing the Hausdorff distance on finite subsets of a metric space, and the theory of quantitative convex semilattices [28], axiomatizing the composition of the Kantorovich and Hausdorff distances, on finitely generated convex sets of finitely supported probability distributions on a metric space.

Given the presence of such examples, important in the theory of programming languages and semantics, and the relevance of finitary proofs in practical applications of quantitative algebra, we argue that the systematic study of quantitative compact theories is an interesting endeavour. Natural types of questions which we will explore in forthcoming work, include: what are conditions on theories that guarantee compactness? For example syntactical conditions on the shape of the quantitative equations, or categorical conditions on the corresponding term monads. What is common between the examples mentioned above? In the other direction, assuming that a theory is compact, what can generally be said about its class of models, the corresponding term monad, *et cetera*?

Contribution and Organization of this Work

The goal of this paper is to set the foundation for this line of research by formally introducing the notion of compact quantitative theory and proving, in full details, that the theory of *interpolative barycentric algebras* is compact. By generalizing this example, we also derive a family of compact quantitative theories, each axiomatising some distance (lifting) on probability distributions, such as the k-Wasserstein distance for $k \in [1, +\infty]$, (also studied in [22] for $k < \infty$) and examples involving log-probabilities.

Rather than working with the original theory of quantitative algebra of [22], we formulate our results in the recently introduced generalization of [27] (see §9.1 of [27] for a detailed comparison with [22]). One axis of generalisation consists in allowing quantitative algebras $\mathbb{A} = (A, \{op^{\mathbb{A}}\}_{op \in \Sigma}, d_A)$ whose distances d_A are not required to be a metric and can be arbitrary fuzzy relations. This, beside allowing non-metric distances to be modelled (as in [26]), has the advantage of decoupling the technicalities related to compactness from those related to metric reasoning. For example, by casting the quantitative theory of *interpolative barycentric algebras* of [22] in the context of [27], as we do in Section 4, we obtain a generalized compact axiomatization of the Kantorovich lifting of distances that are not necessarily metrics.

In Section 2 we introduce the necessary technical background. In Section 3 we introduce the notion of compact quantitative theory and provide some examples. In Section 4 we prove the compactness of a generalization of the quantitative theory of *interpolative barycentric algebras* of [22]. In Section 5 we generalize this result and obtain other examples of compact quantitative theories of convex algebras.

2 Technical Background

2.1 Convex Algebras, Probability Distributions and Couplings

Convex algebras, also known under serveral other names including convex spaces [14] and barycentric algebras [31] (see [30,14,19] for overviews), are algebraic structures over the uncountable signature of operations $\Sigma_{\mathrm{CA}} = \{+_p \mid 0 < p < 1\}$ where, for each $p \in (0,1)$, the operation $+_p$ is binary. Therefore, a Σ_{CA}-algebra is a relational structure $\mathbb{A} = (A, \{+^{\mathbb{A}}_p\}_{p\in(0,1)})$ with $+^{\mathbb{A}}_p : A \times A \to A$, for all p.

Definition 2.1 A *convex algebra* is a Σ_{CA}-algebra $\mathbb{A}$ such that:

Idempotency:	$a +^{\mathbb{A}}_p a = a$	$\forall a \in A, \forall p \in (0,1),$
Skew commutativity.:	$a +^{\mathbb{A}}_p b = b +^{\mathbb{A}}_{1-p} a$	$\forall a, b \in A, \forall p \in (0,1),$
Skew associativity:	$(a +^{\mathbb{A}}_p b) +^{\mathbb{A}}_q c = a +^{\mathbb{A}}_{pq} (b +^{\mathbb{A}}_{\frac{(1-p)q}{1-pq}} c)$	$\forall a, b, c \in A, \forall p, q \in (0,1).$

If $\mathbb{A}$ is clear from the context, we just write $a +_p a'$ in place of $a +^{\mathbb{A}}_p a'$. Any convex subset of a real vector space, such as $[0,1]$, is a convex algebra with operations defined as $x +_p y = px + (1-p)y$. But not all convex algebras are of this form. For example, any semilattice $(X, \vee)$ becomes a convex algebra by defining $x +_p y = x \vee y$, for all $p \in (0,1)$.

Alternatively, it is possible to present convex algebras using a signature of n-ary operations $\sum_{i=1}^n p_i x_i$, where $p_i \in [0,1]$ (thus including 0 and 1) and $\sum_{i=1}^n p_i = 1$. In this version, a convex algebra $\mathbb{A}$ is a set A with interpretations of type $A^n \to A$, for all n-ary convex combination expressions, subject to axioms similar to those given above for binary operations. The two approaches are equivalent [30, Prop. 2.3], and one can always rewrite n-ary convex combinations to Σ_{CA}-terms, recursively, as follows:

$$(\text{case } p_1 = 1)\ \sum_{i=1}^n p_i x_i = x_1 \qquad (\text{case } p_1 = 0)\ \sum_{i=1}^n p_i x_i = \sum_{i=2}^n p_i x_i$$

$$(\text{case } 0 < p_1 < 1)\ \sum_{i=1}^n p_i x_i = x_1 +_{p_1} \Big(\sum_{i=2}^n \frac{p_i}{1-p_1} x_i\Big).$$

In what follows we will switch between binary and n-ary operations as most convenient.

Finitely supported distributions on a set play an important role in the theory of convex algebra.

Definition 2.2 Given a set X, a probability distribution on X is a function $\mu\colon X \to [0,1]$ such that $\sum_{x\in X} \mu(x) = 1$. The support of μ is the set $supp(\mu) = \{x \mid \mu(x) > 0\}$. We say that μ is *finitely supported* if $supp(\mu)$ is finite. We denote with $D(X)$ the set of finitely supported probability distributions on X. For $x \in X$, we denote with δ_x the Dirac probability distribution defined as $\delta_x(y) = 1$ if $y = x$ and 0 otherwise.

The set $D(X)$ with operations $+_p$ defined, for all $\mu, \nu \in D(X)$, as $(\mu +_p \nu)(x) = p\mu(x) + (1-p)\nu(x)$, is a convex algebra which, with some abuse of notation, we also denote by $D(X)$. It is well known (see, for example, [30]) that $D(X)$ is the free convex algebra generated by X. This means that for any convex algebra $\mathbb{A}$, there is a one to one correspondence between functions $f : X \to A$ and convex algebra homomorphisms $\hat{f} : D(X) \to \mathbb{A}$, such that $f = \hat{f} \circ \eta_X$, where $\eta_X : X \to D(X)$ is defined as $\eta_X(x) = \delta_x$.

It also means that $D(X)$ can be seen as the quotient of $\mathrm{Terms}_{\Sigma_{\mathrm{CA}}}(X)$, the absolutely free algebra of Σ_{CA}-terms over X, by the axioms of convex algebras (Definition 2.1). For this reason, given $s \in \mathrm{Terms}_{\Sigma_{\mathrm{CA}}}(X)$, we denote with $[s] \in D(X)$ the corresponding probability distribution. Formally, if s is of the form (using n-ary notation) $s = \sum_{i=1}^n p_i x_i$ then $[s] = \sum_{i=1}^n p_i \delta_{x_i}$.

We now turn attention to couplings of finitely supported probability distributions.

Definition 2.3 Let X be a set and $\mu, \nu \in D(X)$. A *coupling* γ of μ and ν is a probability distribution γ on $X \times X$ whose marginals are μ and ν: for all $x \in X$, $\sum_{x,y\in X} \gamma(x,y) = \mu(x)$, and for all $y \in X$, $\sum_{x,y\in X} \gamma(x,y) = \nu(y)$. We denote with $\Gamma(\mu,\nu) \subseteq D(X \times X)$ the set of all couplings on μ and ν.

Note that $\Gamma(\mu,\nu)$ is always nonempty because the independent product $\mu \times \nu$ of μ and ν, defined as $(\mu \times \nu)(\langle x, x'\rangle) = \mu(x) \cdot \nu(x')$, is a coupling. Furthermore, viewing $\Gamma(\mu,\nu)$ as a (closed and bounded) subset of $[0,1]^{n\times n}$ ($n = |supp(\mu) \cup supp(\nu)|$) we observe that $\Gamma(\mu,\nu)$ is compact, by the Heine-Borel theorem.

Lemma 2.4 *Given any set X and $\mu, \nu \in D(X)$, the set of couplings $\Gamma(\mu,\nu)$ is nonempty and compact.*

The following simple[3] property of couplings is important: a convex combination of couplings is a coupling of the convex combinations.

Lemma 2.5 *Let X be a set, and $\mu_1, \mu_2, \nu_1, \nu_2 \in D(X)$. Let $\gamma_1 \in \Gamma(\mu_1, \nu_1)$ and $\gamma_2 \in \Gamma(\mu_2, \nu_2)$. Then, for all $p \in (0, 1)$, it holds that $\gamma_1 +_p \gamma_2 \in \Gamma(\mu_1 +_p \mu_2, \nu_1 +_p \nu_2)$.*

2.2 Quantitative Algebra

Quantitative algebra, as originally introduced in [22], deals with algebraic structures $\mathbb{A} = (A, \{op^{\mathbb{A}}\}_{op\in\Sigma})$, for some signature of function symbols Σ, further endowed with an *extended metric* $d_A \colon A\times A \to \mathbb{R}_{\geq 0}\cup\{\infty\}$ which makes all operations 1-Lipschitz: $d_A\big(op^{\mathbb{A}}(a_1, \dots, a_n), op^{\mathbb{A}}(a'_1, \dots, a'_n)\big) \leq \max_{i=1\dots n} d_A(a_i, a'_i)$.

Here we instead follow a generalisation of the entire apparatus, recently proposed in [27]. Quantitative algebras, in the sense of [27], are algebraic structures $\mathbb{A} = (A, \{op^{\mathbb{A}}\}_{op\in\Sigma})$, for some some signature of functions symbols Σ, further endowed with an arbitrary[4] fuzzy relation $d_A \colon A \times A \to [0, 1]$ not subject to any additional constraints such as being a metric or making all operations 1-Lipschitz. Such constraints can be expressed by quantitative equations (cf. Example 2.17) in the same way that commutativity of a binary operation can be expressed by equations in universal algebra. We refer to [27] for a detailed exposition and, in particular, to [27, §9.1] for a comparison with [22]. Here we restrict attention only to the key definitions needed, later in Section 3, to formalize the concept of *compact quantitative theory*.

Definition 2.6 Given a set A, a function of type $d : A \times A \to [0, 1]$ is called a *fuzzy relation* on A. We also refer to the pair (A, d) as a fuzzy relation. Given two fuzzy relations (A, d_A) and (B, d_B), a function $f : A \to B$ is *1-Lipschitz* (or also *nonexpansive*) if $d_B\big(f(a), f(a')\big) \leq d_A(a, a')$, for all $a, a' \in A$.

Definition 2.7 Given a set A we denote with d_1^A, or just d_1 if A is clear from the context, the *discrete fuzzy relation* $d_1^A : A \times A \to [0, 1]$ defined as $d_1^A(a, a') = 1$, for all $a, a' \in A$.

Note that, for all fuzzy relations (B, d_B), any function $f : A \to B$ is 1-Lipschitz as a map $f : (A, d_A^1) \to (B, d_B)$. Also note that fuzzy relations are not required to be metrics nor pseudometrics. For example, the discrete fuzzy relation is neither a metric nor a pseudometric because $d(a, a) \neq 0$.

Fuzzy relations on finite sets can be represented by finite matrices. For example:

$$\text{the fuzzy relation } \left(A = \{a_1, a_2\} \;,\; d_A = \begin{cases} (a_1, a_1) \mapsto 0.5 \\ (a_1, a_2) \mapsto 1 \\ (a_2, a_1) \mapsto 0.3 \\ (a_2, a_2) \mapsto 0 \end{cases} \right), \text{ is represented by: } \begin{array}{cc} & \begin{array}{cc} a_1 & a_2 \end{array} \\ \begin{array}{c} a_1 \\ a_2 \end{array} & \begin{bmatrix} 0.5 & 1 \\ 0.3 & 0 \end{bmatrix} \end{array}.$$

Definition 2.8 [Quantitative Algebra] Let Σ be a possibly infinite signature of function symbols $op \in \Sigma$, each having finite arity $ar(op) \in \mathbb{N}$. A *quantitative algebra* is a structure $\mathbb{A} = (A, \{op^{\mathbb{A}}\}_{op\in\Sigma}, d_A)$, where:

(i) $(A, \{op^{\mathbb{A}}\}_{op\in\Sigma})$ is an algebra, in the usual sense of universal algebra: a set A with interpretations of the operations: $op^{\mathbb{A}} : A^{ar(op)} \to A$.

(ii) (A, d_A) is a fuzzy relation.

Given quantitative algebras $\mathbb{A}$ and $\mathbb{B}$, a homomorphism of quantitative algebras is a map $f : A \to B$ which is both 1-Lipschitz as $f : (A, d_A) \to (B, d_B)$ and a homomorphism of the underlying algebras.

Whereas equations in ordinary universal algebra must be satisfied by all interpretations of the variables, in quantitative equations the set of variables, say B, is specified in a context, namely a fuzzy relation $d_B \colon B \times B \to [0, 1]$, and crucially the quantitative equation must only be satisfied by 1-Lipschitz interpretations. This is made formal by the following definition.

Definition 2.9 [Quantitative Equation] A *quantitative equation* is an expression of one of the two following forms:

[3] The proof is straightforward and is omitted but is available in the Appendix.

[4] The choice of valuing the distance in $[0, 1]$ is purely motivated by simplicity reasons. All the results of [27] hold when $[0, 1]$ is replaced by any complete lattice, such as $\mathbb{R}_{\geq 0} \cup \{\infty\}$ or any quantale. See [29].

$$\forall(B, d_B).s =_\epsilon t \qquad\qquad \forall(B, d_B).s = t$$

where (B, d_B) is a fuzzy relation on a (possibly infinite) set B, $s, t \in \mathrm{Terms}_\Sigma(B)$ are Σ-terms built from the set of generators B and ϵ is a real number in $[0, 1]$. When (B, d_B) is finite, we often write the quantitative equation using matrix notation as, for example:

$$\forall\left(\begin{array}{cc} & \begin{array}{cc} b_1 & b_2 \end{array} \\ \begin{array}{c} b_1 \\ b_2 \end{array} & \begin{bmatrix} 1 & \epsilon \\ 1 & 1 \end{bmatrix} \end{array}\right).\ s =_\epsilon t\ .$$

Definition 2.10 [Satisfiability Relation] Let $\mathbb{A}$ be a quantitative algebra. We define the *satisfiability relation* as follows:

$$\mathbb{A} \models \forall(B, d_B).s =_\epsilon t \quad \text{iff} \quad d_A\big(\iota(s), \iota(t)\big) \leq \epsilon \quad \text{for all 1-Lipschitz } \iota : (B, d_B) \to (A, d_A)$$

$$\mathbb{A} \models \forall(B, d_B).s = t \quad \text{iff} \quad \iota(s) = \iota(t) \quad \text{for all 1-Lipschitz } \iota : (B, d_B) \to (A, d_A)$$

where $\iota(s), \iota(t) \in A$ denote the intepretations of the terms s, t, using the extension $\iota\colon \mathrm{Terms}_\Sigma(B) \to \mathbb{A}$ of ι to terms defined, as usual in universal algebra, by structural induction on terms.

Remark 2.11 Following the universal algebra textbook [33] and [27], the "$\forall$" symbol in quantitative equations is adopted to remind the universal quantification involved in the satisfiability relation. In [29] (see also [12]) the notation $(B, d_B) \vdash s =_\epsilon t$ is used *in lieu* of $\forall(B, d_B).s =_\epsilon t$. We reserve the usage of the "$\vdash$" symbol, later on, for the syntactical consequence relation of quantitative algebra.

Remark 2.12 The reader familiar with the original apparatus of quantitative algebra of [22] will recognize that quantitative equations $\forall(A, d).s =_\epsilon t$, in the sense of Definition 2.9, coincide with *basic inferences* of [22] of the form: $\{a =_{d(a,a')} a' \mid a, a' \in A\} \vdash s =_\epsilon t$. Also note that, unlike [22], we defined two types of quantitative equations: with equality $(=)$ and with quantitative equality $(=_\epsilon)$. This is because the models are not necessarily metric spaces, and so equality $(=)$ cannot be equivalently be expressed as distance zero $(=_0)$. We refer to [27, §9.1] for a detailed discussion.

Remark 2.13 The 1-Lipschitz condition on interpretations is on the whole (potentially infinite) set B and not on the finite subset $B' \subseteq B$ of variables appearing in s, t. To appreciate this point, consider $(B, d_B) = ([0, 1], d_{[0,1]})$ with the standard Euclidean metric. All 1-Lipschitz maps $f\colon (B, d_B) \to (2, d^2_1)$, from (B, d_B) to the two-element set with the discrete fuzzy relation, are constant. However, there are non-constant 1-Lipschitz maps $f\colon (\{0, 1\}, d_{B'}) \to (2, d^2_1)$ where $d_{B'}$ is the restriction of d_B to $\{0, 1\} \subsetneq [0, 1]$.

Note that if no interpretation $\iota\colon (B, d_B) \to (A, d_A)$ is 1-Lipschitz, the quantitative equation is trivially satisfied. Also (see comment after Definition 2.7) if d_B is the discrete fuzzy relation on B, every interpretation $\iota : B \to A$ is 1-Lipschitz. Hence the familiar type of equations, quantified over all possible interpretations, can be expressed by taking $d_B = d^B_1$ to be the discrete fuzzy relation.

Example 2.14 The following quantitative equation expresses that $op \in \Sigma$ is commutative.

$$\forall\left(\begin{array}{cc} & \begin{array}{cc} b_1 & b_2 \end{array} \\ \begin{array}{c} b_1 \\ b_2 \end{array} & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{array}\right).\ op(b_1, b_2) = op(b_2, b_1)\ .$$

When d_B is not discrete, the 1-Lipschitz restriction on the intepretations becomes meaningful.

Example 2.15 The following quantitative equation expresses that any two points having distance smaller or equal than 0 (and thus, 0) must be equal:

$$\forall\left(\begin{array}{cc} & \begin{array}{cc} b_1 & b_2 \end{array} \\ \begin{array}{c} b_1 \\ b_2 \end{array} & \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \end{array}\right).\ b_1 = b_2\ .$$

By collecting infinitely many quantitative equations, with varying values defining d_B, it is possible to express common properties such as the symmetry of the distance (for all ϵ, $d(x,y) \leq \epsilon \Rightarrow d(y,x) \leq \epsilon$), or the 1-Lipschitz property of an operation $op \in \Sigma$ (for all ϵ, $d(x,y) \leq \epsilon \Rightarrow d(op(x), op(y)) \leq \epsilon$).

Example 2.16 Let $E = \{\phi_\epsilon\}$ be the set of quantitative equations, indexed by $\epsilon \in [0,1]$, of the form:

$$\phi_\epsilon = \qquad \forall(\begin{array}{c c} & \begin{array}{cc} b_1 & b_2 \end{array} \\ \begin{array}{c} b_1 \\ b_2 \end{array} & \begin{bmatrix} 1 & \epsilon \\ 1 & 1 \end{bmatrix} \end{array}).\ b_2 =_\epsilon b_1.$$

A quantitative algebra $\mathbb{A}$ satisfies E (i.e., $\forall \phi_\epsilon \in E, \mathbb{A} \models \phi_\epsilon$) if and only if d_A is symmetric.

Example 2.17 Let $op \in \Sigma$ be unary and let $E = \{\phi_\epsilon\}$ be the set of quantitative equations, indexed by $\epsilon \in [0,1]$, of the form:

$$\phi_\epsilon = \qquad \forall(\begin{array}{c c} & \begin{array}{cc} b_1 & b_2 \end{array} \\ \begin{array}{c} b_1 \\ b_2 \end{array} & \begin{bmatrix} 1 & \epsilon \\ 1 & 1 \end{bmatrix} \end{array}).\ op(b_1) =_\epsilon op(b_2).$$

A quantitative algebra $\mathbb{A}$ satisfies E if and only if $op^{\mathbb{A}} : (A, d_A) \to (A, d_A)$ is 1-Lipschitz.

With these definitions in place, one defines as expected the *consequence relation* between a set E of quantitative equations and a quantitative equation ψ.

Definition 2.18 [Consequence Relation] Let E be a set of quantitative equations and ψ be a quantitative equation. We write $E \models \psi$, and say that *ψ is a consequence of E*, if:

$$\forall \mathbb{A}, \quad \text{if } \left(\textstyle\bigwedge_{\phi \in E} \mathbb{A} \models \phi \right) \text{ then } \mathbb{A} \models \psi.$$

A key fact in quantitative algebra is that the consequence relation can be axiomatized: there exists a proof system for deriving judgments of the form $E \vdash \psi$ such that $E \vdash \phi$ is derivable if and only if $E \models \phi$ holds. This proof system, introduced in [27, §4], is the quantitative algebra analogous of the well known deductive system of Birkhoff for equations. It includes a number of *finitary* axioms and rules, some for handling equality judgments (entirely analogous to those of Birkhoff's proof system) like:

$$\frac{}{E \vdash \forall(B, d_B).s = s} \qquad\qquad \frac{E \vdash \forall(B, d_B).s = t}{E \vdash \forall(B, d_B).t = s}$$

and some for handling quantitive equality judgments, like:

$$\frac{}{E \vdash \forall(B, d_B).b =_\epsilon b'}\ d_B(b, b') = \epsilon \qquad\qquad \frac{E \vdash \forall(B, d_B).s =_\epsilon t}{E \vdash \forall(B, d_B).s =_\delta t}\ \delta \geq \epsilon$$

The axiom on the left allows to derive distances (provided by the context d_B) between variables and the "weakening" rule on the right states that if from E the upper bound $\leq \epsilon$ on the distance between s and t is derivable, then also the weaker upper bound $\leq \delta$ is derivable. An important feature of the proof system is the substitution rule (see Rule 4.e in Definition 4.1 of [27]):

$$\frac{E \vdash \forall(B, d_B).s =_\star t \qquad \left\{ E \vdash \forall(C, d_C).\sigma(b) =_{d_B(b,b')} \sigma(b') \mid b, b' \in B,\ d_B(b, b') < 1 \right\}}{E \vdash \forall(C, d_C).\sigma(s) =_\star \sigma(t).}\ \sigma : B \to \mathrm{Terms}_\Sigma(C)$$

where $s =_\star t$ is either $s = t$ or $s =_\epsilon t$ for some $\epsilon \in [0,1]$ and $\sigma(s), \sigma(t)$ denote the application of the substitution σ to the terms s, t, defined as usual. The set of premises on the right-side witnesses that the substitution σ is 1-Lipschitz. For this reason, we say that the substitution rule can only be applied with *provably* 1-Lipschitz substitutions σ. Note that any substitution σ is provably 1-Lipschitz when $d_B = d_1^B$ is the discrete distance on B (in this case the right premise is empty). Also note that the substitution rule is finitary whenever $\{b, b' \mid d_B(b, b') < 1\}$ is finite, as it happens in the special case $d_B = d_1^B$. We refer to [27, §4] for a detailed presentation of the other rules of the proof system. Here we only focus on the fact that, unlike Birkhoff's proof system, the proof system of quantitative algebra includes one infinitary rule:

$$\frac{E \vdash \forall(B,d_B).s =_{\epsilon_0} t \qquad E \vdash \forall(B,d_B).s =_{\epsilon_1} t \qquad \dots \qquad E \vdash \forall(B,d_B).s =_{\epsilon_n} t \qquad \dots}{E \vdash \forall(B,d_B).s =_{\epsilon} t}\ \epsilon = \inf\{\epsilon_i\}_{i\in\mathbb{N}}$$

This rule states that if the bound $\leq \epsilon_i$ (in $\forall(B,d_B).s =_{\epsilon_i} t$) is derivable form E, for each ϵ_i, then also the "limit" $\leq \epsilon$ bound ($E \vdash \forall(B,d_B).s =_{\epsilon} t$) is derivable. This rule has appeared in the literature under several names: it is named *order completeness rule* in [27], *Archimedean rule* in [22] and *continuity rule* in [8]. As a consequence of the presence of this rule, proofs in the proof system of quantitative algebra are, in general, well-founded infinite trees because some nodes, corresponding to the application of the infinitary rule, have countably infinite width. The presence of some kind of infinitary rule is necessary, in the sense that there is no axiomatisation of the consequence relation ($E \models \phi$) given by a set of finitary Horn implications in first order logic. This follows from the fact that the category of pseudometric spaces and 1-Lipschitz maps (which is the category of models of a quantitative theory E, see Example 3.5 below) is not locally finitely presentable (see [3, Ex. 2.2] and [4, §5]).

A main result of quantitative algebra is the existence of free algebras. We refer to [27, §5] for a detailed formulation. Here we only state the fact that, given a quantitative equational theory E and a fuzzy relation (A,d_A), the *free E-quantitative algebra generated by* (A,d_A) is isomorphic to $\big(\mathrm{Terms}_\Sigma(A)/_{\equiv}, \{op^F\}_{op\in\Sigma}, d_F\big)$, where:

(i) the carrier is the set of Σ-terms over A modulo the equivalence relation $\equiv$ of "provable equality" defined as: $s \equiv t \Leftrightarrow E \vdash \forall(A,d_A).s = t$,

(ii) the interpretation of the operations is defined as: $op^F([s_1]_\equiv,\dots,[s_n]_\equiv) = [op(s_1,\dots,s_n)]_\equiv$,

(iii) the distance is defined by "provable ϵ-equality": $d_F([s]_\equiv,[t]_\equiv) \leq \epsilon \Leftrightarrow E \vdash \forall(A,d_A).s =_\epsilon t$.

It can be shown that definitions (ii–iii) are valid for all choices of representatives $s \in [s]_\equiv$, $t \in [t]_\equiv$.

3 Compact Quantitative Theories

We now introduce the concept of a compact quantitative equational theory. In what follows a possibly infinite signature Σ of function symbols, each having finite arity, is fixed.

Definition 3.1 [Compact Quantitative Theory] Let E be a set of quantitative equations. We say that E is *compact* if, for all quantitative equations ϕ, if $E \models \phi$ then there exists a finite proof (i.e., a proof tree never using infinitary rules) of $E \vdash \phi$.

Remark 3.2 We use the term *theory* as synonym of *set* of quantitative equations. Often, the term *theory* is instead used to indicate a set of (quantitative) equations closed under the consequence relation. But, in such setting, every consequence ϕ of E trivially admits a finite proof, because an axiom of the proof system (see "Init", Def. 4.1 of [27]) allows to derive $E \vdash \phi$ whenever $\phi \in E$. Hence, the notion of compactness of Definition 3.1 is meaningful for sets E not closed under deducibility.

The following sequence of propositions illustrate the notion of compact quantitative theory.

Proposition 3.3 *Let $\Sigma = \emptyset$ and $E = \emptyset$. Then E is compact.*

Proof. First, observe that since there are no operations ($\Sigma = \emptyset$) the models of $E = \emptyset$ are all fuzzy relations (A,d_A). Now let ϕ be a quantitative equation. We need to prove that if $\emptyset \models \phi$ then $\emptyset \vdash \phi$ has a finite derivation. The assumption $\emptyset \models \phi$ means that $(A,d_A) \models \phi$, for all fuzzy relations (A,d_A).

Let us first consider the case ϕ is of the form $\forall(B,d_B).b = b'$. We claim that the assumption implies that $b = b'$. Indeed if $b \neq b'$, the fuzzy relation (B,d_B) itself and the identity interpretation (which is always 1-Lipschitz), witness that $(B,d_B) \not\models \phi$. Now that the claim has been established, we observe that $E \vdash \forall(B,d_B).b = b$ can be derived by one of the axioms of the proof system.

Let us now consider the case ϕ is of the form $\forall(B,d_B).b =_\epsilon b'$. We claim that the assumption implies that $\epsilon \geq d_B(b,b')$. Indeed if $\epsilon < d_B(b,b')$, the fuzzy relation (B,d_B) itself and the identity interpretation (which is always 1-Lipschitz), witness that $(B,d_B) \not\models \phi$. Now that the claim has been established, we observe that $E \vdash \forall(B,d_B).b =_{d_B(b,b')} b'$ can be can be derived by one of the axioms, and from it we can apply a weakening rule (because $\epsilon \geq d_B(b,b')$) to derive $E \vdash \forall(B,d_B).b =_\epsilon b'$. $\square$

As we already observed, the semantics of equations $s = t$, in standard universal algebra, can be expressed by quantitative equations of the form $\forall(B, d_1).s = t$, using the discrete fuzzy relation on a set B containing all variables appearing in s and t. All proof rules of Birkhoff's system have corresponding finitary rules in the proof system of quantitative algebra, when dealing with quantitative equations of the form $\forall(B, d_1).s = t$. In particular, the familiar substitution rule of Birkhoff proof system (see [33, Rule EL5]) has a corresponding rule because any substitution σ is, as already remarked, provably 1-Lipschitz when using the discrete distance d_1. This has the following consequence.

Proposition 3.4 *Fix an arbitrary Σ and let $E = \{\phi_i\}_{i\in I}$ be a collection of quantitative equations ϕ_i of the form $\forall(B_i, d_1^{B_i}).s_i = t_i$, where $d_1^{B_i}$ is the discrete fuzzy relation on B_i. Then E is compact.*

Proof. Let $\mathcal{E} = \{s_i = t_i\}_{i\in I}$ be the set of equations (in the sense of universal algebra) corresponding to E. Note that the models of E are quantitative algebras $\mathbb{A} = (A, \{op^{\mathbb{A}}\}_{op\in\Sigma}, d_A)$ whose underlying algebra $(A, \{op^{\mathbb{A}}\}_{op\in\Sigma})$ is a model of $\mathcal{E}$ and d_A is an arbitrary fuzzy relation. Now, consider an arbitrary quantitative equation ψ and assume $E \models \psi$. We need to show that $E \vdash \psi$ has a finite derivation.

Consider first the case that ψ is of the form $\forall(B, d_B).s = t$, for some $s, t \in \mathrm{Terms}_\Sigma(B)$. The assumption $E \models \psi$ means that all models of E satisfy $\forall(B, d_B).s = t$. In particular, for each model $(A, \{op^{\mathbb{A}}\}_{op\in\Sigma})$ of $\mathcal{E}$, the quantitative algebra $\mathbb{A} = (A, \{op^{\mathbb{A}}\}_{op\in\Sigma}, d_0^A)$ is a model of E, with d_0^A the constant zero fuzzy relation ($d_0^A(a, a') = 0$). Hence $\mathbb{A} \models \forall(B, d_B).s = t$. Note that any function $f\colon B \to A$ is 1-Lipschitz as a map $f\colon (B, d_B) \to (A, d_0^A)$. Therefore $\mathbb{A} \models \forall(B, d_B).s = t$ means that $\iota(s) = \iota(t)$ holds for all interpretations ι. Hence $\iota(s) = \iota(t)$ holds, for any interpretation ι, in any model of $\mathcal{E}$. By completness of Birkhoff's proof system, this implies that $s = t$ is derivable from $\mathcal{E}$ by means of a finite derivation. By "copying" this finite derivation, in the proof system of quantitative algebra, we obtain a corresponding finite proof of $E \vdash \forall(B, d_1^B).s = t$, where d_1^B is the discrete fuzzy relation on B. Finally, we can derive $E \vdash \forall(B, d_B).s = t$ from $E \vdash \forall(B, d_1^B).s = t$ by applying a substitution rule of the proof system, with the identity substitution ($\sigma(b) = b$) of type $\sigma : (B, d_1^B) \to \mathrm{Terms}_\Sigma(B, d_B)$, which is trivially provably 1-Lipschitz.

Now consider the case that ψ is of the form $\forall(B, d_B).s =_\epsilon t$. The assumption $E \models \psi$ implies that any model $\mathbb{A}$ of E satisfies $\forall(B, d_B).s =_\epsilon t$. But since, among the models of E, there are models with the discrete fuzzy relation d_1^A, which assigns distance 1 to the intepretations of s and t, we deduce that $\epsilon = 1$. The quantitative equation $E \vdash \forall(B, d_B).s =_1 t$ can be finitely proved by an axiom of the proof system (see axiom (h), in Definition 4.1 of [27]). $\square$

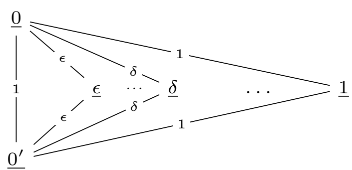
However, as soon as quantitative theories E including some quantitative equations of the form $\forall(B, d_B).s =_\epsilon t$ are considered, it is very easy to incur in examples that are not compact.

Proposition 3.5 *Let $\Sigma = \emptyset$ and let E be the theory of pseudometric spaces defined as the union of the following sets of quantitative equations, for all $\epsilon, \epsilon_1, \epsilon_2 \in [0, 1]$:*

$$\forall\Big(\begin{array}{cc} & b \\ b & \begin{bmatrix}1\end{bmatrix}\end{array}\Big).\ b =_0 b \qquad \forall\Big(\begin{array}{cc} & \begin{array}{cc} b_1 & b_2\end{array} \\ \begin{array}{c} b_1 \\ b_2\end{array} & \begin{bmatrix}1 & \epsilon \\ 1 & 1\end{bmatrix}\end{array}\Big).\ b_2 =_\epsilon b_1 \qquad \forall\Big(\begin{array}{cc} & \begin{array}{ccc} b_1 & b_2 & b_3\end{array} \\ \begin{array}{c} b_1 \\ b_2 \\ b_3\end{array} & \begin{bmatrix}1 & \epsilon_1 & 1 \\ 1 & 1 & \epsilon_2 \\ 1 & 1 & 1\end{bmatrix}\end{array}\Big).\ b_1 =_{\min\{1,\epsilon_1+\epsilon_2\}} b_3\,.$$

Then the quantitative theory E is not compact.

Proof. It is easy to verify (see [27, §2.3]) that the models of E are fuzzy relations (A, d_A) such that d_A is a pseudometric: symmetric ($d_A(a, a') = d_A(a', a)$), self-distance zero ($d_A(a, a) = 0$) and satisfying the triangular inequality ($d_A(a, c) \leq d_A(a, b) + d_A(b, c)$). Consider the fuzzy relation (B, d_B) where $B = \{\underline{0}, \underline{0'}\} \cup \{\underline{\epsilon} \mid 0 < \epsilon \leq 1\}$ (the unit interval with "two zeros") and d_B is a symmetric and self-distance zero fuzzy relation specified by: $d_B(\underline{0}, \underline{0'}) = 1$, $d(\underline{0}, \underline{\epsilon}) = d(\underline{0'}, \underline{\epsilon}) = \epsilon$ and $d_B(\underline{\epsilon}, \underline{\delta}) = |\epsilon - \delta|$, for all $\epsilon, \delta \in (0, 1]$:



Hence d_B does not satisfy the triangular inequality because $d_B(\underline{0},\underline{0'}) = d_B(\underline{0'},\underline{0}) = 1$. Let ϕ be the quantitative equation $\forall(B,d_B).\underline{0} =_0 \underline{0'}$. We show that $E \models \phi$ but all proofs of $E \vdash \phi$ are infinite (i.e., they involve the infinitary rule).

First, we prove that $E \models \phi$. Let (A,d_A) be a model of E, i.e., a pseudometric space, and let $\iota\colon (B,d_B) \to (A,d_A)$ be a 1-Lipschitz interpretation. Note that $d_A(\iota(\underline{0}),\iota(\underline{0'})) = 0$, because for every $\epsilon > 0$, $d_A(\iota(\underline{0}),\iota(\underline{0'})) \leq d_A(\iota(\underline{0}),\iota(\underline{\epsilon})) + d_A(\iota(\underline{\epsilon}),\iota(\underline{0'})) \leq \epsilon + \epsilon$, by triangular inequality of d_A and the 1-Lipschitz property of ι. As this holds for all 1-Lipschitz ι, we conclude that $(A,d_A) \models \phi$. Hence $E \models \phi$.

Secondly, we show that $E \vdash \phi$ does not have finite proofs. Informally, each judgment $\forall(B,d_B).\underline{0} =_{2\epsilon} \underline{0'}$ has a finite derivation from the judgments $\forall(B,d_B).\underline{0} =_\epsilon \underline{\epsilon}$ and $\forall(B,d_B).\underline{\epsilon} =_\epsilon \underline{0'}$, which are finitely derivable using an axiom of the proof system, and by application of the "triangular inequality" quantitative equation from E. But collecting these judgments to obtain $\forall(B,d_B).\underline{0} =_0 \underline{0'}$ requires an application of the infinitary rule. Formally, we need to exhibit a relational structure $(C,\{R_\epsilon\}_{\epsilon\in[0,1]})$, where each $R_\epsilon \subseteq C \times C$ is a binary relation interpreting $=_\epsilon$, such that C models all the deductive rules of the proof system – except the infinitary rule – and all the quantitative equations in E but not ϕ. Let $C = B$ and define $R_0 = \mathrm{Id}_C$ (the identity relation) and R_ϵ, for each $\epsilon > 0$, as the symmetric reflexive relation satisfying:

$$(\underline{0},\underline{0'}) \in R_\epsilon, \quad (z,\underline{\delta}) \in R_\epsilon \text{ (for } z \in \{\underline{0},\underline{0'}\} \text{ and } 0 < \delta \leq \epsilon), \quad (\underline{\delta},\underline{\lambda}) \in R_\epsilon \text{ (for } 0 < \delta,\lambda \leq 1 \text{ s.t. } |\delta-\lambda| \leq \epsilon).$$

In this model, $\underline{0}$ is arbitrarily close to $\underline{0'}$ (i.e., $(\underline{0},\underline{0'}) \in R_\epsilon$) but not at distance 0 (i.e., $(\underline{0},\underline{0'}) \notin R_0$). ☐

4 The Theory of Interpolative Convex algebras is compact

The theory of *Interpolative Barycentric (IB) quantitative algebras* was introduced in [22] as a main example of quantitative theory. In [22] it was proved that the free IB quantitative algebra generated by a metric space $(X, d\colon X \times X \to [0,1])$ is isomorphic to the quantitative algebra $(D(X), \{+_p\}_{p\in(0,1)}, K(d))$, where $K(d)\colon D(X) \times D(X) \to [0,1]$ is the Kantorovich (see Definition 4.2 below) lifting of d. In this section we recast this result in our chosen apparatus of quantitative algebra of [27], introduced in Section 2.2, where distances are not necessarily metrics. To avoid confusion with [22] we will refer to it as the quantitative theory of *Interpolative convex algebras* ($\mathbb{ICA}$).

Definition 4.1 [Theory $\mathbb{ICA}$] Let Σ_{CA} be the signature of convex algebras. The quantitative theory of *interpolative convex algebras*, denoted by $\mathbb{ICA}$, is defined as the union of the following quantitative equations, for all $p,q \in (0,1)$:

$$\text{Idempotency:} \quad \forall\Big(\begin{array}{cc} & x \\ x & \begin{bmatrix}1\end{bmatrix}\end{array}\Big).\ x +_p x = x \qquad \text{Skew comm.:} \quad \forall\Big(\begin{array}{cc} & \begin{array}{cc} x & y \end{array} \\ \begin{array}{c} x \\ y \end{array} & \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\end{array}\Big).\ x +_p y = y +_{1-p} x$$

$$\text{Skew assoc.:} \quad \forall\Big(\begin{array}{cc} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}\end{array}\Big).\ (x +_p y) +_q z = x +_{pq} (y +_{\frac{(1-p)q}{1-pq}} z)$$

and the following quantitative equations, for all $p \in (0,1)$ and $\epsilon, \delta \in [0,1]$:

$$\text{Interpolative:} \quad \forall\Big(\begin{array}{cc} & \begin{array}{cccc} x & y & w & z \end{array} \\ \begin{array}{c} x \\ y \\ w \\ z \end{array} & \begin{bmatrix}1 & 1 & \epsilon & 1\\ 1 & 1 & 1 & \delta\\ 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\end{bmatrix}\end{array}\Big).\ x +_p y =_{p\epsilon+(1-p)\delta} w +_p z.$$

The first set of quantitative equations simply corresponds to the equations of convex algebras (Definition 2.1), which must hold for all possible interpretations, and are therefore expressed with the discrete fuzzy relation on variables. This implies that all models $\mathbb{A} = (A, \{+^{\mathbb{A}}\}_p, d_A)$ satisfying $\mathbb{ICA}$ are such that the underlying algebra $(A, \{+^{\mathbb{A}}\}_p)$ is a convex algebra. The set of "interpolative" quantitative equations, on the other hand, have a nontrivial fuzzy relation on variables, which constrains interpretations of x and w

(respectively, y and z) to have distance bounded by ϵ (respectively, bounded by δ).

Definition 4.2 [Kantorovich distance lifting] Let (A, d) be a fuzzy relation. The *Kantorovich lifting of* d is a fuzzy relation on $D(A)$ (that is $K(d) : D(A) \times D(A) \to [0, 1]$) defined as follows, for all $\mu, \nu \in D(A)$:

$$K(d)(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} V_d(\gamma) \qquad \text{where} \qquad V_d(\gamma) = \sum_{a,b \in A} \gamma(a, b) \cdot d(a, b).$$

Remark 4.3 Since $\Gamma(\mu, \nu)$ is nonempty $K(d)(\mu, \nu) < \infty$. Furthermore, since all values $\gamma(a, b)$ and $d(a, b)$ are in $[0, 1]$, we have that $0 \leq K(d)(\mu, \nu) \leq 1$. In other words, $K(d)$ is indeed a fuzzy relation on $D(A)$.

Remark 4.4 It is well known (see e.g. [32, p. 68]) that when d is a (pseudo)metric on A then also $K(d)$ is a (pseudo)metric on $D(A)$. In general, however, $K(d)$ does not satisfy these properties. For example, if $d(a, a) = 1$ for some $a \in A$, then $K(d)(\delta_a, \delta_a) = 1$ thus not satisfying the properties of metrics.

Proposition 4.5 *Let (A, d) be a fuzzy relation. The map $\eta_A : (A, d) \to (D(A), K(d))$, defined as $\eta_A(a) = \delta_a$, is 1-Lipschitz.*

Proof. We need to show that $K(d)(\delta_a, \delta_b) \leq d(a, b)$, for all $a, b \in A$. This inequality is witnessed by the independent product $\delta_a \times \delta_b$, which is a coupling, and satisfies $V_d(\delta_a \times \delta_b) = d(a, b)$. □

The first result, relating the quantitative theory $\mathbb{ICA}$ and the Kantorovich distance, states that the Kantorovich distance gives a model of $\mathbb{ICA}$, a statement that can be rephrased as a soundness property.

Proposition 4.6 *Let (A, d_A) be a fuzzy relation and $(D(A), \{+_p\}_{p\in(0,1)})$ be the convex algebra of finitely supported probability distributions. The quantitative algebra $(D(A), \{+_p\}_{p\in(0,1)}, K(d))$ is a model of $\mathbb{ICA}$.*

Proof. We know that $D(A)$ satisfies all axioms (idempotency and skew comm./assoc.) of convex algebras, and thus it satisfies the corresponding quantitative equations in $\mathbb{ICA}$. It remains only to show that it satisfies all instances, for $p \in (0, 1)$ and $\epsilon, \delta \in [0, 1]$, of the interpolative quantitative equation. So assume $\iota : \{x, y, w, z\} \to D(A)$ is any 1-Lipschitz interpretation of the variables and denote with $\mu_x = \iota(x)$, $\mu_y = \iota(y)$, $\mu_w = \iota(w)$ and $\mu_z = \iota(z)$. The 1-Lipschitz property of ι amounts to say that $K(d)(\mu_x, \mu_w) \leq \epsilon$ and $K(d)(\mu_y, \mu_z) \leq \delta$. We need to prove that: $K(d)\big(\mu_x +_p \mu_y, \mu_w +_p \mu_z\big) \leq p\epsilon + (1 - p)\delta$.

By definition of $K(d)$ as an infimum, for any $\lambda > 0$, we can find couplings $\gamma \in \Gamma(\mu_x, \mu_w)$ and $\gamma' \in \Gamma(\mu_y, \mu_z)$ such that: $V_d(\gamma) \leq \epsilon + \lambda$ and $V_d(\gamma') \leq \delta + \lambda$. From γ and γ', we obtain by Lemma 2.5 a coupling $\gamma +_p \gamma' \in \Gamma(\mu_x +_p \mu_y, \mu_w +_p \mu_z)$ such that:

$$V_d(\gamma +_p \gamma') \stackrel{def}{=} \sum_{a,b\in A} \big(p\gamma(a, b) + (1 - p)\gamma'(a, b)\big) \cdot d(a, b) = pV_d(\gamma) + (1 - p)V_d(\gamma').$$

Hence $V_d(\gamma +_p \gamma') \leq p(\epsilon + \lambda) + (1 - p)(\delta + \lambda)$. Since $\lambda > 0$ is arbitrary, by taking $\lambda \to 0$ we deduce that $V_d(\gamma +_p \gamma') \leq p\epsilon + (1 - p)\delta$. Hence $K(d)\big(\mu_x +_p \mu_y, \mu_w +_p \mu_z\big) \leq p\epsilon + (1 - p)\delta$. □

Recall that given $s \in \mathrm{Terms}_{\Sigma_{\mathrm{CA}}}(A)$ a convex algebra term with variables in A we write $[s] \in D(A)$ for the corresponding probability distribution, defined as: $[\sum_{i=1}^n p_i a_i] = \sum_{i=1}^n p_i \delta_{a_i}$.

Corollary 4.7 (Soundness) *Let (A, d_A) be a fuzzy relation and $s, t \in \mathrm{Terms}_{\Sigma_{\mathrm{CA}}}(A)$ be two convex algebra terms with variables in A. Then the following implication holds, for all $\epsilon \in [0, 1]$:*

$$\mathbb{ICA} \vdash \forall(A, d_A).s =_\epsilon t \text{ is derivable} \qquad \Longrightarrow \qquad K(d_A)([s], [t]) \leq \epsilon.$$

Proof. By the properties of the proof system of quantitative algebra, $\mathbb{ICA} \vdash \forall(A, d_A).s =_\epsilon t$ is equivalent to $\mathbb{ICA} \models \forall(A, d_A).s =_\epsilon t$. By Proposition 4.6, $(D(A), \{+_p\}_{p\in(0,1)}, K(d))$ is a model of $\mathbb{ICA}$. The assumption therefore implies that $(D(A), \{+_p\}_{p\in(0,1)}, K(d)) \models (A, d_A).s =_\epsilon t$. This means that for all 1-Lipschitz interpretations $\iota \colon A \to D(A)$ it holds that $K(d_A)(\iota(s), \iota(t)) \leq \epsilon$. The interpretation η_A $(a \mapsto \delta_a)$ is 1-Lipschitz by Lemma 4.5 and, by definition, $\eta_A(s) = [s]$ and $\eta_A(t) = [t]$. Hence $K(d_A)([s], [t]) \leq \epsilon$ holds. □

The following useful corollary can be also be deduced from Proposition 4.6 and the fact that $D(A)$ is the free convex algebra on A. It states, using the termilogy established in [27, Def 7.6], that $\mathbb{ICA}$ is a quantitative *extension* of the theory of convex algebras.

Corollary 4.8 *Let (A, d_A) be a (possibly infinite) fuzzy relation and $s, t \in \mathrm{Terms}_{\Sigma_{\mathrm{CA}}}(A)$ be two convex algebra terms with variables in A. The following are equivalent:*

(i) *$s = t$ is provable in Birkhoff's proof system from the axioms of convex algebras (Definition 2.1).*

(ii) *$\mathbb{ICA} \vdash \forall(A, d_A).s = t$ is derivable in the proof system of quantitative algebra by a finite proof.*

(iii) *$\mathbb{ICA} \vdash \forall(A, d_A).s = t$ is derivable in the proof system of quantitative algebra.*

Proof. Direction (1) $\Rightarrow$ (2). Following the same argument of Proposition 3.4, the finite proof in Birkhoff's proof system of $s = t$ can be translated to a finite proof in the proof system of quantitative algebra of $\mathbb{ICA} \vdash \forall(A, d_1).s = t$, where d_1 is the discrete fuzzy relation. The identity substitution $\sigma(a) = a$, of type $\sigma : (A, d_1) \to (A, d_A)$ is trivially provably 1-Lipschitz, because d_1 is discrete. Hence we can apply the substitution rule and obtain the desired judgment $\mathbb{ICA} \vdash \forall(A, d_A).s = t$.

Direction (2) $\Rightarrow$ (3) is trivial.

Direction (3) $\Rightarrow$ (1). Assume $\mathbb{ICA} \vdash \forall(A, d_A).s = t$ is derivable. By the soundness of the proof system, this means that, $\mathbb{ICA} \models \forall(A, d_A).s = t$, i.e. for any quantitative algebra $\mathbb{A}$ that is a model of $\mathbb{ICA}$, it holds that $\mathbb{A} \models \forall(A, d_A).s = t$. Consider as $\mathbb{A}$ the quantitative algebra $(D(A), \{+_p\}_{p\in(0,1)}, K(d_A))$. Hence $\iota(s) = \iota(t)$ holds in $\mathbb{A}$ for all 1-Lipschitz interpretations $\iota : (A, d_A) \to (D(A), K(d_A))$. The intepretation $\eta_A(a) = \delta_a$ is 1-Lipschitz by Proposition 4.5. Hence $\eta_A(s) = \eta_A(t)$. Since $(D(A), \{+_p\}_{p\in(0,1)})$ is the free convex algebra on A, this implies that $s = t$ is derivable in Birkhoff's proof system from the axioms of convex algebra. $\square$

Note that the above Corollary (direction $(iii) \Rightarrow (ii)$) establishes half of what is required to conclude that $\mathbb{ICA}$ is compact. We now proceed showing that the quantitative theory $\mathbb{ICA}$ and the Kantorovich distance are also related by a completeness property (Proposition 4.10 below). Before stating it, we prove the following useful, purely syntactic, lemma.

Lemma 4.9 *Let (A, d) be a (possibly infinite) fuzzy relation. Let $e(\boldsymbol{x}) \in \mathrm{Terms}_{\Sigma_{\mathrm{CA}}}(\{x_1, \dots, x_n\})$ be a convex algebra expression on n variables of the form (using n-ary notation) $e(\boldsymbol{x}) = \sum_{i=1}^{n} p_i x_i$. Let $\boldsymbol{a} = (a_1, \dots, a_n)$ and $\boldsymbol{b} = (b_1, \dots, b_n)$ be tuples in A. Denote with $e(\boldsymbol{a})$ and $e(\boldsymbol{b})$ the expressions (with variables in A) obtained by substituting $\boldsymbol{x}$ with $\boldsymbol{a}$ and $\boldsymbol{b}$. Then the following judgment is derivable, by a finite proof, in the proof system of quantitative algebra:*

$$\mathbb{ICA} \vdash \forall(A, d).e(\boldsymbol{a}) =_\epsilon e(\boldsymbol{b}) \qquad\qquad \text{where } \epsilon = \textstyle\sum_{i=1}^{n} p_i \cdot d(a_i, b_i)\ .$$

Proof. By induction on n.

Base case ($n = 1$): in this case, $e(\boldsymbol{x}) = x_1$ and we need to show a finite proof of the judgment: $\mathbb{ICA} \vdash \forall(A, d).a_1 =_{d(a_1,b_1)} b_1$. And indeed it is simply derivable as an instance of one of the axioms of the proof system which uses the information provided by the distance d on A.

Inductive Case ($n > 1$): in this case $e(\boldsymbol{x}) = x_1 +_{p_1} e'(\boldsymbol{x})$ where the expression $e'(\boldsymbol{x}) = \sum_{i=2}^{n} \frac{p_i}{1-p_1} x_i$ only involves variables $\{x_2, \dots, x_n\}$. We need to exhibit a finite proof of the judgment

$$\mathbb{ICA} \vdash \forall(A, d).a_1 +_{p_1} e'(\boldsymbol{a}) =_\epsilon b_1 +_{p_1} e'(\boldsymbol{b}).$$

We derive this judgment by taking the following instance of the Interpolation axiom (in $\mathbb{ICA}$):

$$\forall(\ \begin{array}{cc} & \begin{array}{cccc} x & y & w & z \end{array} \\ \begin{array}{c} x \\ y \\ w \\ z \end{array} & \begin{bmatrix} 1 & 1 & d(a_1,b_1) & 1 \\ 1 & 1 & 1 & \epsilon' \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array}).\ x +_{p_1} y \ =_{p_1 d(a_1,b_1) + (1-p_1)\epsilon'}\ w +_{p_1} z$$

where $\epsilon' = \sum_{i=2}^{n} \frac{p_i}{(1-p_1)} d(a_i, b_i)$, and applying the substitution $\tau : \{x, y, w, z\} \to \mathrm{Terms}_{\Sigma_{\mathrm{CA}}}(A)$ defined as:

$$\tau(x) = a_1 \qquad \tau(y) = e'(\boldsymbol{a}) \qquad \tau(w) = b_1 \qquad \tau(z) = e'(\boldsymbol{b}).$$

This will yield the desired finite proof, since $\epsilon = p_1 d(a_1, b_1) + (1-p_1)\epsilon'$. The application of the substitution rule requires showing that τ is provably 1-Lipschitz, i.e., we need to provide finite sub-derivations of the two judgments:

$$\mathbb{ICA} \vdash \forall(A, d_A).\tau(x) =_{d_X(x,w)} \tau(w) \qquad\qquad \mathbb{ICA} \vdash \forall(A, d_A).\tau(y) =_{d_X(y,z)} \tau(z)$$

where d_X is the fuzzy relation on $\{x, y, w, z\}$ used in the instance of the Interpolation rule above.

For the left-judgment we need to derive $\forall(A, d).a_1 =_{d(a_1,b_1)} b_1$. This, like in the base case, is derivable by instantiating one of the axioms of the proof system. For the right-judgment we need a finite derivation $\forall(A, d).e'(\boldsymbol{a}) =_{\epsilon'} e'(\boldsymbol{b})$, and this exists by inductive hypothesis. Hence τ is provably 1-Lipschitz and the substitution rule can be applied. □

Proposition 4.10 (Completeness) *Let (A, d_A) be a fuzzy relation and $s, t \in \mathrm{Terms}_{\Sigma_{\mathrm{CA}}}(A)$ be two convex algebra terms with variables in A. Then the following implication holds, for all $\epsilon \in [0, 1]$:*

$$K(d)([s], [t]) \leq \epsilon \qquad \Longrightarrow \qquad \mathbb{ICA} \vdash \forall(A, d).\ s =_\epsilon t \quad \textit{is derivable.}$$

Proof.

Let us denote $\mu = [s]$ and $\nu = [t]$. The proof involves two steps:

(i) We first show that for every coupling $\gamma \in \Gamma(\mu, \nu)$ there exists a finite derivation of the judgmet $\mathbb{ICA} \vdash \forall(A, d).\ s =_\lambda t$, where $\lambda = V_d(\gamma)$ specified as in Definition 4.2.

(ii) Since $K(d)(\mu, \nu) \leq \epsilon$ by assumption, and $K(d)(\mu, \nu) = \inf_\gamma V_d(\gamma)$ by definition, we have that $\inf_\gamma V(\gamma) \leq \epsilon$. We can collect countably many derivations of (Step i) and, by applying the infinitary rule, obtain a derivation of:

$$\mathbb{ICA} \vdash \forall(A, d).\ s =_\delta t \qquad\qquad \delta = \inf_\gamma V_d(\gamma)$$

and from this the desired derivation of ($\mathbb{ICA} \vdash \forall(A, d).\ s =_\epsilon t$) is obtained by applying a weakening rule of the proof system, since $\epsilon \geq \delta$.

So it remains to outline the details of (Step i). Fix an arbitrary coupling $\gamma \in \Gamma(\mu, \nu)$. By the defining properties of couplings, we have that, for all $a, b \in A$:

$$\mu(a) = \textstyle\sum_{b \in A} \gamma(a, b) \qquad\qquad \nu(b) = \textstyle\sum_{a \in A} \gamma(a, b).$$

This implies that the convex algebra terms s and t can be proven equal (in the theory of convex algebras, using Birkhoff's proof system) to the following convex algebra (n-ary) expressions:

$$s = \sum_{a \in A} (\sum_{b \in A} \gamma(a, b))\, a = \sum_{a,b \in A} \gamma(a, b)\, a \qquad\qquad t = \sum_{b \in A} (\sum_{a \in A} \gamma(a, b))\, b = \sum_{a,b \in A} \gamma(a, b)\, b.$$

In other words, there is a convex algebra expression $e(\boldsymbol{x}) = \sum \gamma(a, b)\, x_{a,b}$, in the finite set of variables variables $\{x_{a,b} \mid a, b \in \mathit{supp}(\mu) \cup \mathit{supp}(\nu)\}$, such that the following equalities are provable from the axioms of convex algebras: $s = e(\boldsymbol{a})$ and $t = e(\boldsymbol{b})$, where $e(\boldsymbol{a})$ and $e(\boldsymbol{b})$ denote the substitutions of each of the variables $x_{a,b}$ in $e(\boldsymbol{x})$ with $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. By Corollary 4.8, these equalities have corresponding finite derivations of: $\mathbb{ICA} \vdash \forall(A, d).s = e(\boldsymbol{a})$ and $\mathbb{ICA} \vdash \forall(A, d).t = e(\boldsymbol{b})$. From lemma 4.9 we obtain a finite derivation of the following judgment:

$$\mathbb{ICA} \vdash \forall(A, d).\ e(\boldsymbol{a}) =_\lambda e(\boldsymbol{b}) \qquad\qquad \lambda = \sum_{a,b} \gamma(a, b) d(a, b) = V_d(\gamma)\ .$$

We can combine these three derivations, using a deductive rule of the proof system (Rule 4.j in [27, Def 4.1] called "congruence"), to obtain the desired finite derivation of $\mathbb{ICA} \vdash \forall(A, d).\ s =_\lambda t$. □

Corollary 4.11 *The free $\mathbb{ICA}$-quantitative algebra generated by the fuzzy relation (A, d_A) is isomorphic to $(D(A), \{+_p\}_{p \in (0,1)}, K(d_A))$.*

Proof. As discussed at the end of Section 2.2, the carrier of the free quantitative algebra is $\mathrm{Terms}_{\Sigma_{\mathrm{CA}}}(A)/_\equiv$ by the relation $\equiv$ of provable equality, and this coincides (by Corollary 4.8), with provable equality in Birkhoff's proof system from the axioms of convex algebras. Hence the carrier is $(D(A), \{+_p\}_{p \in (0,1)})$, the free convex algebra on A. The distance d_F of the free quantitative algebra is defined as provable distance and, by the soundness (4.6) and completeness (4.10) results, it coincides with $K(d_A)$. □

We are finally ready to show that the quantitative theory $\mathbb{ICA}$ is compact. The statement of completeness (Proposition 4.10) does not assert the existence of a finite proof. By inspection of its proof (specifically Step (ii)), this is due to one single application of the infinitary rule. But the following well known Proposition 4.12, stating that there always exists an *optimal* coupling γ achieving the infimium in the definition of $K(d)$, ensures that Step (ii) is unnecessary. Hence the statement of the completeness Proposition 4.10 can be strengthen by stating that a *finite* proof exists (Corollary 4.13 below). We have delayed this simple observation, and its consequence, to highlight that the soundness and completeness results do not depend on the assumption that optimal couplings exist.

Proposition 4.12 *Let (A, d) be a fuzzy relation. Then, for every $\mu, \nu \in D(A)$ there exists an optimal coupling $\gamma_\star \in \Gamma(\mu, \nu)$ achieving the infimum: $V_d(\gamma_\star) = K(d)(\mu, \nu)$.*

Proof. By Proposition 2.4, $\Gamma(\mu, \nu)$ is a compact topological space and $V_d : \Gamma(\mu, \nu) \to [0, 1]$ is a continuous function and, as such, it attains a minimum value (see [9, Thm 4, p. 362]). □

Corollary 4.13 (Completeness wrt finite proofs) *Let (A, d) be a fuzzy relation and $s, t \in \mathrm{Terms}_{\Sigma_{\mathrm{CA}}}(A)$ be two convex algebra terms with variables in A. Then, for all $\epsilon \in [0, 1]$:*

$$K(d)([s], [t]) \leq \epsilon \qquad \Longrightarrow \qquad \mathbb{ICA} \vdash \forall(A, d).\ s =_\epsilon t \ \ \textit{is derivable with a finite proof.}$$

Theorem 4.14 *The quantitative equational theory $\mathbb{ICA}$ is compact.*

Proof. Let ϕ be a quantitative equation and assume $\mathbb{ICA} \models \phi$ or, equivalently, that $\mathbb{ICA} \vdash \phi$ is derivable by a possibly infinite proof. First, if ϕ is of the form $\forall(A, d).s = t$, by Proposition 4.8 (iii)⇒(ii), $\mathbb{ICA} \vdash \phi$ has a finite proof. Second, if ϕ is of the form $\forall(A, d).s =_\epsilon t$, By Proposition 4.6 we have that $K(d)([s], [t]) \leq \epsilon$ and from Proposition 4.13 we have a finite proof of $\mathbb{ICA} \vdash \phi$. □

5 Analysis of the Proof and Generalizations

In this section we analyze the proof of compactness of $\mathbb{ICA}$ (Theorem 4.14) to get general insights and generalzations. As a result we obtain a family of compact quantitative theories of convex algebras (Theorem 5.3 below) including some interesting examples.

First, by Definition 4.1, the theory $\mathbb{ICA}$ includes quantitative equations corresponding to the equations of convex algebras ("Idempotency", "Skew-comm./assoc.") and quantitative equations of the form:

$$\text{Interpolative:} \quad \forall(\begin{array}{c|c} & \begin{array}{cccc} x & y & w & z \end{array} \\ \begin{array}{c} x \\ y \\ w \\ z \end{array} & \begin{bmatrix} 1 & 1 & \epsilon & 1 \\ 1 & 1 & 1 & \delta \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array}).\ x +_p y \ =_{\epsilon \oplus_p \delta} \ w +_p z.$$

where $\oplus_p : [0, 1]^2 \to [0, 1]$ is the standard convex algebra operation on $[0, 1]$, defined as $x \oplus_p y = px + (1-p)y$. We highlight with $\oplus_p$ the standard operation because, as we discuss later, it is by varying $\oplus_p$ that we will obtain other examples of compact theories.

Second, a lifting method (the Kantorovich distance) is found to map a fuzzy relation (A, d) to a fuzzy relation on $D(A)$, which is the free convex algebra on A. Crucially, the lifting method is expressed as an infimum ($\inf_\gamma V_d(\gamma)$) over evaluations (by the function $V_d : \Gamma(\mu, \nu) \to [0, 1]$) of a family of "witness objects" (the set $\Gamma(\mu, \nu)$ of couplings γ).

Third, the soundness of this lifting method is established (Proposition 4.6 and Corollary 4.7). The key facts used in the proof of soundness are:

(i) the witness objects (couplings) carry a convex algebra structure, as in Lemma 2.5.

(ii) the evaluation V_d satisfies the property: $V_d(\delta_{\langle a,b\rangle}) \leq d(a, b)$, to validate Proposition 4.5.

(iii) the proof of Proposition 4.6 requires that V_d satisfies: $V_d(\gamma +_p \gamma') \leq V_d(\gamma) \oplus_p V_d(\gamma')$.

(iv) the proof of Proposition 4.6 requires that the convex algebra operation $\oplus_p : [0, 1]^2 \to [0, 1]$ is monotone (with respect to the standard order on $[0, 1]$) in both arguments. Monotonicity is required to derive, from the assumptions $V_d(\gamma) \leq \epsilon + \lambda$ and $V_d(\gamma') \leq \delta + \lambda$, that $V_d(\gamma) \oplus_p V_d(\gamma') \leq (\epsilon + \lambda) \oplus_p (\delta + \lambda)$.

(v) the proof of Proposition 4.6 requires that the convex algebra operation $\oplus_p : [0,1]^2 \to [0,1]$ satisfies: $\lim_{\lambda\to 0}(\epsilon+\lambda) \oplus_p (\delta+\lambda) \leq \epsilon \oplus_p \delta$. This property is used to derive $V_d(\gamma) \oplus_p V_d(\gamma') \leq \epsilon \oplus_p \delta$.

Fourth, the completeness of the lifting method is established (Proposition 4.10). The key facts used in the proof of completeness are:

(i) The proof is based on Lemma 4.9. This lamma is purely syntactic and it states that, given two convex algebra terms $e(\boldsymbol{a})$ and $e(\boldsymbol{b})$ of the same shape, it is possible to iterate the "Interpolative" axioms to derive a bound ϵ on their distances. Specifically, if $e(\boldsymbol{x}) = \sum_i p_i x_i$ then $\epsilon = \bigoplus_i p_i d(a_i, b_i)$.

(ii) The proof of completeness is based on the observation that a coupling $\gamma \in \Gamma(\mu,\nu)$ provides a "receipe" for rewriting, using the axioms of convex algebras, the two distributions μ and ν in the same shape (the convex algebra term $e(\boldsymbol{x})$) which allows for the application of Lemma 4.9.

Fifth, and last point, the proof of compactness of Theorem 4.14 uses the finitary completeness Corollary 4.13 which is a consequence of the fact that $\inf_\gamma V_d(\gamma)$ achieves a minimum. This fact is, in turn, proved using the topological compactness of $\Gamma(\mu,\nu)$, from Lemma 2.4, and topological properties of V_d (continuity) ensuring that a minimum is always achived on a compact domain.

The analysis suggests a generalization of the results of Section 4 parametric on choices of $\oplus_p$ different than the standard one. In the rest of this section we develop this idea. The main result, stated as Theorem 5.3, gives a family of examples of compact quantitative theories, including the axiomatisation of the k-Wasserstein distance, for $k \geq 1$, which was also studied in [22], and the ∞-Wasserstein distance.

First, we generalize the notion of Kantorovich distance (Definition 4.2) and of $\mathbb{ICA}$ (Definition 4.1) to arbitrary convex algebras $([0,1], \{\oplus_p\}_{p\in(0,1)})$ on $[0,1]$.

Definition 5.1 [Generalized lifting $K^{\oplus_p}(d)$] Let (A,d) be a fuzzy relation, i.e., $d : A\times A \to [0,1]$. Denote with $V_d^{\oplus_p}$ the unique (recall that $D(A\times A)$ is free on $A \times A$) homomorphic extension of d, defined as:

$$V_d^{\oplus_p} : D(A\times A) \to [0,1] \qquad\qquad \sum_{i=i}^{n} p_i \delta_{\langle a_i, b_i\rangle} \mapsto \bigoplus_{i=1}^{n} p_i d(a_i,b_i) \ .$$

The $\oplus_p$*-lifting of* d is the fuzzy relation $K^{\oplus_p}(d)$ on $D(A)$ defined as follows, for all $\mu,\nu \in D(A)$:

$$K^{\oplus_p}(d)(\mu,\nu) = \inf_{\gamma\in\Gamma(\mu,\nu)} V_d^{\oplus_p}(\gamma) \ .$$

Definition 5.2 [Generalized quantitative theory $\mathbb{ICA}^{\oplus_p}$] The quantitative equational theory $\mathbb{ICA}^{\oplus_p}$ is defined by the union of the quantitative equations "Idempotency", "Skew-comm" and "Skew-assoc" of Definition 4.1 and by the "$\oplus_p$-Interpolative" quantitative equations, for all $p\in(0,1)$ and $\epsilon,\delta\in[0,1]$:

$$\oplus_p\text{–Interpolative:} \quad \forall(\begin{array}{c|c} & \begin{matrix} x & y & w & z \end{matrix} \\ \begin{matrix} x \\ y \\ w \\ z \end{matrix} & \begin{bmatrix} 1 & 1 & \epsilon & 1 \\ 1 & 1 & 1 & \delta \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array}). \ x +_p y \ =_{\epsilon\oplus_p\delta} \ w +_p z \ .$$

Hence, $K(d)$ and $\mathbb{ICA}$ are instances of $K(d)^{\oplus_p}$ and $\mathbb{ICA}^{\oplus_p}$ for the standard convex algebra $([0,1], \{\oplus_p\}_{p\in(0,1)})$ on $[0,1]$, where $x \oplus_p y = px + (1-p)y$.

Theorem 5.3 *Let* $([0,1], \{\oplus_p\}_{p\in(0,1)})$ *be a convex algebra on* $[0,1]$ *such that, for all* $p \in (0,1)$*:*

(i) $\oplus_p$ *is monotone:* $\forall x, x', y \in [0,1]$*, if* $x \leq x'$ *then* $x \oplus_p y \leq x' \oplus_p y$ *and* $y \oplus_p x \leq y \oplus_p x'$*,*

(ii) *for all* $x, y \in [0,1]$*,* $\lim_{\lambda\to 0} \big((x+\lambda)\oplus_p (y+\lambda)\big) \leq x \oplus_p y$*,*

then $\mathbb{ICA}^{\oplus_p}$ *is a sound and complete axiomatization of the distance lifting* $K^{\oplus_p}$*. Furthermore, if:*

(iii) *for all* $d : A\times A \to [0,1]$ *and* $\mu,\nu\in D(A)$*, there exists an optimal coupling:*

$$\exists \gamma_\star \in \Gamma(\mu,\nu). \qquad V_d^{\oplus_p}(\gamma_\star) \ = \inf_{\gamma\in\Gamma(\mu,\nu)} V_d^{\oplus_p}(\gamma)$$

then $\mathbb{ICA}^{\oplus_p}$ *is a compact quantitative equational theory.*

Proof. By definition of $V_d^{\oplus_p}$ as a homomorphism of convex algebras, it holds that $V_d^{\oplus_p}(\delta_{\langle a,b\rangle}) = d(a,b)$ and $V_d^{\oplus_p}(\gamma +_q \gamma') = V_d^{\oplus_p}(\gamma) \oplus_q V_d^{\oplus_p}(\gamma')$. This, together with the assumptions (i) and (ii), ensures that the proof of Soundness can be carried out. Similarly, using the definition $\mathbb{ICA}^{\oplus_p}$, the analogous of Lemma 4.9 can be proved and with it the proof of completeness. Finally, the existence of optimal couplings allows for the strenthening of the completeness theorem to finite proofs, ensuring compactness. □

For a given convex algebra $([0,1], \{\oplus_p\}_{p\in(0,1)})$ on $[0,1]$, conditions (i) and (ii) are usually straightforwards to verify. For (iii), it is typically necessary to establish some topological property (like continuity or, more generally, lower semicontinuity [9, §6]) of the evaluation map $V_d^{\oplus_p} : D(A \times A) \to [0,1]$ that guarantees that the infimum over the compact set $\Gamma(\mu,\nu)$ is always achieved (see [9, Thm 4, p. 362]).

Proposition 5.4 (Examples) *The following operations* $\oplus_p : [0,1]^2 \to [0,1]$*, for* $p \in (0,1)$*, satisfy the axioms of convex algebras (Definition 2.1) and conditions (i–iii) of Theorem 5.3:*

binary version	*n-ary version*	
$x \oplus_p y = px + (1-p)y$	$\bigoplus_i p_i x_i = \sum_i p_i x_i$	*the standard operation*
$x \oplus_p y = \max\{x,y\}$	$\bigoplus_i p_i x_i = \max\{x_i\}_i$	$\vee$*-semilattice*
$x \oplus_p y = (px^k + (1-p)y^k)^{\frac{1}{k}}$	$\bigoplus_i p_i x_i = (\sum_i p_i x_i^k)^{\frac{1}{k}}$	*for some* $k \geq 1$
$x \oplus_p y = x^p \cdot y^{1-p}$	$\bigoplus_i p_i x_i = \prod_i {x_i}^{p_i}$	*log-probabilities*

Hence, by Theorem 5.3, the quantitative theory $\mathbb{ICA}^{\oplus_p}$ is compact for all these examples. As already noted, the Kantorovich distance $K(d)$ (Definition 4.2) is the special case of $K^{\oplus_p}(d)$ for $x \oplus_p y = px + (1-p)y$, the standard convex algebra operation on $[0,1]$. The case $x \oplus_p y = (px^k + (1-p)y^k)^{\frac{1}{k}}$, for some $k \geq 1$, gives as $K^{\oplus_p}(d)$ the Wasserstein k-distance (see [32, Ch. 6]). The case $x \oplus_p y = \max\{x,y\}$ gives as $K^{\oplus_p}(d)$ the Wasserstein ∞-distance (see [10]). Note that, in this case, the evaluation map $V_d^{\oplus_p}$ is not continuous but it is lower semicontinuous by [9, Thm. 4, p. 362], since it is the supremum of all k-Wasserstein evaluations, for $k \geq 1$, which are continuous. Finally, the case $x \oplus_p y = x^p y^{1-p}$ is a distance related to log-probabilities, since $x^p y^{1-p} = \exp(p \log(x) + (1-p)\log(y))$, using the convention $\log(0) = -\infty$ and $\exp(-\infty) = 0$.

6 Conclusions and Future Work

We have introduced the concept of *compact quantitative equational theory* in the quantitative algebra apparatus of [27] and proved that the theory $\mathbb{ICA}$ of *interpolative convex algebras* is compact. This theory was first studied in [22], under the name *interpolative barycentric algebras*, in the less general context of metric spaces and wihtout explictly observing the compactness of the theory. In Section 5 we have obtained an algebraic generalization: to each convex algebra structure on $[0,1]$, satisfying certain regularity conditions, corresponds a compact variant of $\mathbb{ICA}$ axiomatizing a variant of the Kantorovich lifting on probability distributions. We have discussed how some of these variants coincide with well-known distance liftings, such as k-Wasserstein (also studied in [22]), ∞-Wasserstein and involving log-probabilities.

Methods similar to those adopted in this paper can be used to prove the compactness of the quantitative theories of semilattices [22] and of convex semilattices [28]. Both these theories, just like $\mathbb{ICA}$, turn out to be manageble due to topological compactness reasons. For example, finitely generated convex sets of finitely supported probability distributions are compact. An interesting direction of future research is to explore the connection between compactness of quantitative theories and topological compactness.

Our definitions and results have been formulated in the context of the quantitative algebra apparatus of [27] based on fuzzy relations, but they can be adapted to the same apparatus where *generalized metric spaces* (such as metric spaces) take the place of fuzzy relations, as described in [27, §8], or in the quantitative algebra apparatus of [12] based on arbitrary relational Horn theories.

A proof of Proposition 5.4 is available in the Appendix

References

[1] Adámek, J., *Varieties of quantitative algebras and their monads*, in: *37th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS'22)*, ACM (2022).
https://doi.org/10.1145/3531130.3532405

[2] Adámek, J., M. Dostál and J. Velebil, *Strongly finitary monads for varieties of quantitative algebras*, in: *10th Conference on Algebra and Coalgebra in Computer Science (CALCO'23)*, volume 270 of *LIPIcs*, pages 10:1–10:14, Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2023).
https://doi.org/10.4230/LIPICS.CALCO.2023.10

[3] Adámek, J., S. Milius, L. S. Moss and H. Urbat, *On finitary functors and their presentations*, J. Comput. Syst. Sci. **81**, pages 813–833 (2015).
https://doi.org/10.1016/J.JCSS.2014.12.002

[4] Adamek, J. and J. Rosicky, *Locally Presentable and Accessible Categories*, London Mathematical Society Lecture Note Series, Cambridge University Press (1994).

[5] Bacci, G., G. Bacci, K. G. Larsen and R. Mardare, *A complete quantitative deduction system for the bisimilarity distance on markov chains*, Log. Methods Comput. Sci. **14** (2018).
https://doi.org/10.23638/LMCS-14(4:15)2018

[6] Bacci, G., R. Mardare, P. Panangaden and G. Plotkin, *An algebraic theory of markov processes*, in: *33rd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS'18)*, pages 679–688, ACM (2018).
https://doi.org/10.1145/3209108.3209177

[7] Bacci, G., R. Mardare, P. Panangaden and G. D. Plotkin, *Propositional logics for the lawvere quantale*, in: M. Kerjean and P. B. Levy, editors, *Proceedings of the 39th Conference on the Mathematical Foundations of Programming Semantics, MFPS XXXIX, Indiana University, Bloomington, IN, USA, June 21-23, 2023*, volume 3 of *EPTICS*, EpiSciences (2023).
https://doi.org/10.46298/ENTICS.12292

[8] Bacci, G., R. Mardare, P. Panangaden and G. D. Plotkin, *Sum and tensor of quantitative effects*, Log. Methods Comput. Sci. **20** (2024).
https://doi.org/10.46298/LMCS-20(4:9)2024

[9] Bourbaki, N., *Elements of General Topology*, Addison-Wesley Publishing Company (1966).

[10] Champion, T., L. De Pascale and P. Juutinen, *The ∞-wasserstein distance: Local solutions and existence of optimal transport maps*, SIAM Journal on Mathematical Analysis **40**, pages 1–20 (2008). https://doi.org/10.1137/07069938X.
https://doi.org/10.1137/07069938X

[11] D'Angelo, K., S. Gurke, J. M. Kirss, B. König, M. Najafi, W. Rozowski and P. Wild, *Behavioural metrics: Compositionality of the kantorovich lifting and an application to up-to techniques*, in: R. Majumdar and A. Silva, editors, *35th International Conference on Concurrency Theory, CONCUR 2024, September 9-13, 2024, Calgary, Canada*, volume 311 of *LIPIcs*, pages 20:1–20:19, Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2024).
https://doi.org/10.4230/LIPICS.CONCUR.2024.20

[12] Ford, C., S. Milius and L. Schröder, *Monads on categories of relational structures*, in: F. Gadducci and A. Silva, editors, *9th Conference on Algebra and Coalgebra in Computer Science, CALCO 2021, August 31 to September 3, 2021, Salzburg, Austria*, volume 211 of *LIPIcs*, pages 14:1–14:17, Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2021).
https://doi.org/10.4230/LIPICS.CALCO.2021.14

[13] Forster, J., L. Schröder, P. Wild, H. Beohar, S. Gurke, B. König and K. Messing, *Quantitative graded semantics and spectra of behavioural metrics*, in: J. Endrullis and S. Schmitz, editors, *33rd EACSL Annual Conference on Computer Science Logic, CSL 2025, February 10-14, 2025, Amsterdam, Netherlands*, volume 326 of *LIPIcs*, pages 33:1–33:21, Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2025).
https://doi.org/10.4230/LIPICS.CSL.2025.33

[14] Fritz, T., *Convex spaces i: Definition and examples* (2015). 0903.5522.
https://arxiv.org/abs/0903.5522

[15] Gavazzo, F., *Quantitative behavioural reasoning for higher-order effectful programs: Applicative distances*, in: *Logic in Computer Science, LICS 2018*, pages 452–461, ACM (2018).
https://doi.org/10.1145/3209108.3209149

[16] Gavazzo, F. and C. D. Florio, *Elements of quantitative rewriting*, Proc. ACM Program. Lang. **7**, pages 1832–1863 (2023).
https://doi.org/10.1145/3571256

[17] Goncharov, S., D. Hofmann, P. Nora, L. Schröder and P. Wild, *Kantorovich functors and characteristic logics for behavioural distances*, in: O. Kupferman and P. Sobocinski, editors, *Foundations of Software Science and Computation Structures – 26th International Conference, FoSSaCS 2023, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2023, Paris, France, April 22-27, 2023, Proceedings*, volume 13992 of *LNCS*, pages 46–67, Springer (2023).
https://doi.org/10.1007/978-3-031-30829-1_3

[18] Hodges, W., *Model Theory*, volume 42 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press (1993).

[19] Jacobs, B., *Convexity, duality and effects*, in: C. S. Calude and V. Sassone, editors, *Theoretical Computer Science*, pages 1–19, Springer Berlin Heidelberg, Berlin, Heidelberg (2010), ISBN 978-3-642-15240-5.

[20] Jurka, J., S. Milius and H. Urbat, *Algebraic reasoning over relational structures*, in: V. De Paiva and A. Simpson, editors, *Proc. 40th Conference on Mathematical Foundations of Programming Semantics (MFPS)*, volume 4 of *ENTICS*, pages 13:1–13:20 (2024).

[21] Lago, U. D., F. Honsell, M. Lenisa and P. Pistone, *On quantitative algebraic higher-order theories*, in: A. P. Felty, editor, *7th International Conference on Formal Structures for Computation and Deduction, FSCD 2022, August 2-5, 2022, Haifa, Israel*, volume 228 of *LIPIcs*, pages 4:1–4:18, Schloss Dagstuhl – Leibniz-Zentrum für Informatik (2022).
https://doi.org/10.4230/LIPICS.FSCD.2022.4

[22] Mardare, R., P. Panangaden and G. D. Plotkin, *Quantitative algebraic reasoning*, in: M. Grohe, E. Koskinen and N. Shankar, editors, *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16, New York, NY, USA, July 5-8, 2016*, pages 700–709, ACM (2016).
https://doi.org/10.1145/2933575.2934518

[23] Mardare, R., P. Panangaden and G. D. Plotkin, *On the axiomatizability of quantitative algebras*, in: *32nd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017, Reykjavik, Iceland, June 20-23, 2017*, pages 1–12, IEEE Computer Society (2017).
https://doi.org/10.1109/LICS.2017.8005102

[24] Milius, S. and H. Urbat, *Equational axiomatization of algebras with structure*, in: M. Bojańczyk and A. Simpson, editors, *Proc. Foundations of Software Science and Computation Structures (FoSSaCS)*, volume 11425 of *Lecture Notes Comput. Sci. (ARCoSS)*, pages 400–417, Springer (2019).

[25] Mio, M., R. Sarkis and V. Vignudelli, *Combining nondeterminism, probability, and termination: Equational and metric reasoning*, in: *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*, pages 1–14, IEEE (2021).
https://doi.org/10.1109/LICS52264.2021.9470717

[26] Mio, M., R. Sarkis and V. Vignudelli, *Beyond nonexpansive operations in quantitative algebraic reasoning*, in: *Proc. 37th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS'22)*, pages 52:1–52:13, ACM (2022).
https://doi.org/10.1145/3531130.3533366

[27] Mio, M., R. Sarkis and V. Vignudelli, *Universal quantitative algebra for fuzzy relations and generalised metric spaces*, Log. Methods Comput. Sci. **20**, pages 19:1–19:56 (2024).

[28] Mio, M. and V. Vignudelli, *Monads and quantitative equational theories for nondeterminism and probability*, in: I. Konnov and L. Kovács, editors, *31st International Conference on Concurrency Theory, CONCUR 2020, September 1-4, 2020, Vienna, Austria (Virtual Conference)*, volume 171 of *LIPIcs*, pages 28:1–28:18, Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2020).
https://doi.org/10.4230/LIPICS.CONCUR.2020.28

[29] Sarkis, R., *Lifting Algebraic Reasoning to Generalized Metric Spaces*, Phd thesis, ENS de Lyon (2024). Available at https://doi.org/10.5281/zenodo.14001076.

[30] Sokolova, A. and H. Woracek, *Termination in convex sets of distributions*, Logical Methods in Computer Science **Volume 14, Issue 4**, 15 (2018), ISSN 1860-5974.
https://doi.org/10.23638/LMCS-14(4:17)2018

[31] Stone, M. H., *Postulates for the barycentric calculus*, Ann. Math. **29**, pages 25–30 (1949).

[32] Villani, C., *Optimal Transport: Old and New*, Springer (2009).

[33] Wechler, W., *Universal Algebra for Computer Scientists*, volume 25 of *EATCS Monographs on Theoretical Computer Science*, Springer (1992), ISBN 978-3-642-76773-9.
https://doi.org/10.1007/978-3-642-76771-5

[34] Wild, P. and L. Schröder, *Characteristic Logics for Behavioural Metrics via Fuzzy Lax Extensions*, in: I. Konnov and L. Kovács, editors, *31st International Conference on Concurrency Theory, CONCUR 2020, September 1-4, 2020, Vienna, Austria (Virtual Conference)*, volume 171 of *LIPIcs*, pages 27:1–27:23, Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2020). 2007.01033.
https://doi.org/10.4230/LIPIcs.CONCUR.2020.27

[35] Wild, P. and L. Schröder, *A quantified coalgebraic van benthem theorem*, in: S. Kiefer and C. Tasson, editors, *Foundations of Software Science and Computation Structures - 24th International Conference, FOSSACS 2021, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2021, Luxembourg City, Luxembourg, March 27 - April 1, 2021, Proceedings*, volume 12650 of *LNCS*, pages 551–571, Springer (2021).
https://doi.org/10.1007/978-3-030-71995-1_28

A Proofs of Statements of Section 2

Lemma A.1 *Given any set X and $\mu, \nu \in D(X)$, the set of couplings $\Gamma(\mu,\nu)$ is nonempty and compact.*

Proof. The set $\Gamma(\mu,\nu)$ is nonempty since it always contains the independent product of μ and ν. Let $n = |supp(\mu) \cup supp(\nu)|$. Note that $\Gamma(\mu,\nu)$ can be seen as a subset of the Euclidean space $\mathbb{R}^{n\times n}$ and that μ and ν can be seen as points in $\mathbb{R}^n$. Let $\pi_L, \pi_R : \mathbb{R}^{n\times n} \to \mathbb{R}^n$ be the two projection maps, which are continuous. By definition, $\Gamma(\mu,\nu) = \pi_L^{-1}(\{\mu\}) \cap \pi_R^{-1}(\{\nu\})$. Hence $\Gamma(\mu,\nu)$ is closed, being the intersection of two closed sets. Also, $\Gamma(\mu,\nu)$ is bounded since it is a subset of $[0,1]^{n\times n}$. Hence $\Gamma(\mu,\nu)$ is compact by the Heine-Borel theorem. $\square$

Lemma A.2 *Let X be a set, and $\mu_1, \mu_2, \nu_1, \nu_2 \in D(X)$. Let $\gamma_1 \in \Gamma(\mu_1,\nu_1)$ and $\gamma_2 \in \Gamma(\mu_2,\nu_2)$. Then, for all $p \in (0,1)$, it holds that $\gamma_1 +_p \gamma_2 \in \Gamma(\mu_1 +_p \mu_2, \nu_1 +_p \nu_2)$.*

Proof. The assumptions on γ_1 and γ_2 amounts to:

$$\begin{array}{ccc} \forall x \in X & \qquad & \forall y \in X \\ \begin{aligned} \textstyle\sum_{x,y} \gamma_1(x,y) &= \mu_1(x) \\ \textstyle\sum_{x,y} \gamma_2(x,y) &= \mu_2(x) \end{aligned} & & \begin{aligned} \textstyle\sum_{x,y} \gamma_1(x,y) &= \nu_1(y) \\ \textstyle\sum_{x,y} \gamma_2(x,y) &= \nu_2(y) \end{aligned} \end{array}$$

To simplify notation, let us introduce $G = p\gamma_1 + (1-p)\gamma_2$. We need to show that:

- $\sum_{x,y} G(x,y) = p\mu_1(x) + (1-p)\mu_2(x) \qquad \forall x$,
  and this holds because the left hand side is equivalent to:
  $$p\big(\sum_{x,y} \gamma_1(x,y)\big) + (1-p)\big(\sum_{x,y} \gamma_2(x,y)\big) \qquad \forall x \in X.$$
- $\sum_{x,y} G(x,y) = p\nu_1(y) + (1-p)\nu_2(y) \qquad \forall y$,
  and this holds because the left hand side is equivalent to:
  $$p\big(\sum_{x,y} \gamma_1(x,y)\big) + (1-p)\big(\sum_{x,y} \gamma_2(x,y)\big) \qquad \forall y \in X.$$

$\square$

B Proofs of Statements in Section 5

In this section we expand the statement of Proposition 5.4 The following numerical identity is useful:

Lemma B.1 (Numerical Fact) *Let $p, q \in (0,1)$. Then*

$$(1-pq)(1-\tfrac{(1-p)q}{1-pq}) = 1-q.$$

Proof. $(1-pq)(1-\frac{(1-p)q}{1-pq}) = 1-pq-(1-p)q = 1-pq-q+pq = 1-q$. $\square$

First, we prove that all $\oplus_p$ operations of Proposition 5.4 indeed satisfy the axioms of convex algebras, and that the stated n-ary versions indeed correspond to the binary definitions.

Lemma B.2 *The following operations satisfy the axioms of convex algebras.*

binary version	*n-ary version*	
$x \oplus_p y = px + (1-p)y$	$\bigoplus_i p_i x_i = \sum_i p_i x_i$	*the standard operation*
$x \oplus_p y = \max\{x,y\}$	$\bigoplus_i p_i x_i = \max\{x_i\}_i$	*$\vee$-semilattice*
$x \oplus_p y = (px^k + (1-p)y^k)^{\frac{1}{k}}$	$\bigoplus_i p_i x_i = (\sum_i p_i x^k)^{\frac{1}{k}}$	*for some $k \geq 1$*
$x \oplus_p y = x^p \cdot y^{1-p}$	$\bigoplus_i p_i x_i = \prod_i {x_i}^{p_i}$	*log-probabilities*

Proof.

Case $x \oplus_p y = px + (1-p)x$. This is the standard convex algebra on $[0,1]$.

Case $x \oplus_p y = \max\{x, y\}$. This is a $\vee$-semilattice and thus a convex algebra structure.

Case $x \oplus_p y = (px^k + (1-p)y^k)^{\frac{1}{k}}$. Idempotency and skew-commutativity are straightforward. For skew-associativity we need to show that $(x \oplus_p y) \oplus_q z = x \oplus_{pq} (y \oplus_{\frac{(1-p)q}{1-pq}} z)$:

$$
\begin{aligned}
(x \oplus_p y) \oplus_q z &= \sqrt[k]{q(\sqrt[k]{px^k + (1-p)y^k})^k + (1-q)z^k} \\
&= \sqrt[k]{q(px^k + (1-p)y^k) + (1-q)z^k} \\
&= \sqrt[k]{pqx^k + ((1-p)q)y^k + (1-q)z^k} \\
&= \sqrt[k]{pqx^k + (1-pq)\tfrac{(1-p)q}{1-pq}y^k + (1-q)z^k} \\
&\quad \text{by Lemma B.1} \\
&= \sqrt[k]{pqx^k + (1-pq)\tfrac{(1-p)q}{1-pq}y^k + (1-pq)(1-\tfrac{(1-p)q}{1-pq})z^k} \\
&= \sqrt[k]{pqx^k + (1-pq)\Big(\tfrac{(1-p)q}{1-pq}y^k + (1-\tfrac{(1-p)q}{1-pq})z^k\Big)} \\
&= \sqrt[k]{pqx^k + (1-pq)\Big(\sqrt[k]{\tfrac{(1-p)q}{1-pq}y^k + (1-\tfrac{(1-p)q}{1-pq})z^k}\Big)^k} \\
&= x \oplus_{pq} (y \oplus_{\frac{(1-p)q}{1-pq}} z)
\end{aligned}
$$

To prove that the n-ary expression corresponds to the binary definition, we need to verify the recursive equations:

$$(\text{case } p_1 = 1)\ \bigoplus_{i=1}^n p_i x_i = x_1 \qquad\qquad (\text{case } p_1 = 0)\ \bigoplus_{i=1}^n p_i x_i = \bigoplus_{i=2}^n p_i x_i$$

which clearly hold and:

$$(\text{case } 0 < p_1 < 1)\ \bigoplus_{i=1}^n p_i x_i = x_1 +_{p_1} \big(\bigoplus_{i=2}^n \tfrac{p_i}{1-p_1} x_i\big).$$

and this is proven as follows:

$$
\begin{aligned}
\bigoplus_{i=1}^n p_i x_i &= \sqrt[k]{\textstyle\sum_i p_i x^k} \\
&= \sqrt[k]{p_1 x^k + (1-p_1)\textstyle\sum_{i=2}^n \frac{p_i}{1-p_1} x^k} \\
&= \sqrt[k]{p_1 x^k + (1-p_1)\big(\sqrt[k]{\textstyle\sum_{i=2}^n \frac{p_i}{1-p_1} x^k}\big)^k} \\
&= x_1 +_{p_1} \big(\textstyle\bigoplus_{i=2}^n \frac{p_i}{1-p_1} x_i\big)
\end{aligned}
$$

Case $x \oplus_p y = x^p y^{1-p}$. This case is entirely analogous to the previous one. Idempotency and skew-commutativity are straightforward. For skew-associativity we need to show that $(x \oplus_p y) \oplus_q z = x \oplus_{pq} (y \oplus_{\frac{(1-p)q}{1-pq}} z)$.

$$\begin{array}{rcl}
(x \oplus_p y) \oplus_q z & = & \left(x^p y^{1-p}\right)^q z^{1-q} \\
& = & x^{pq} y^{(1-p)q} z^{1-q} \\
& = & x^{pq} y^{(1-pq)\frac{(1-p)q}{1-pq}} z^{1-q} \\
& & \text{by Lemma B.1} \\
& = & x^{pq} y^{(1-pq)\frac{(1-p)q}{1-pq}} z^{(1-pq)(1-\frac{(1-p)q}{1-pq})} \\
& = & x^{pq} \left(y^{\frac{(1-p)q}{1-pq}} z^{(1-\frac{(1-p)q}{1-pq})}\right)^{1-pq} \\
& = & x \oplus_{pq} (y \oplus_{\frac{(1-p)q}{1-pq}} z).
\end{array}$$

To prove that the n-ary expression corresponds to the binary definition, we need to verify the recursive equations:

$$(\text{case } p_1 = 1) \ \bigoplus_{i=1}^n p_i x_i = x_1 \qquad\qquad (\text{case } p_1 = 0) \ \bigoplus_{i=1}^n p_i x_i = \bigoplus_{i=2}^n p_i x_i$$

which clearly hold and:

$$(\text{case } 0 < p_1 < 1) \ \bigoplus_{i=1}^n p_i x_i = x_1 +_{p_1} \left(\bigoplus_{i=2}^n \frac{p_i}{1-p_1} x_i\right).$$

and this is proven as follows:

$$\begin{array}{rcl}
\bigoplus_{i=1}^n p_i x_i & = & \prod_{i=1}^n {x_i}^{p_i} \\
& = & x_1^{p_1} \cdot \left(\prod_{i=2}^n {x_i}^{\frac{p_i}{1-p_1}}\right)^{1-p_1} \\
& = & x_1 +_{p_1} \left(\bigoplus_{i=2}^n \frac{p_i}{1-p_1} x_i\right)
\end{array}$$

□

Next, it remains to verify that that all operations $\oplus_p$ of Theorem 5.3 satisfy points (i–iii). We use the following standard result ([9, Thm. 3, p. 361]).

Lemma B.3 *Let X be a topological space, $Y \subseteq X$ be a compact subset and $f : X \to [0,1]$ be lower semicontinuous. Then there exists $y_\star \in Y$ such that $f(y_\star) = \inf_{y \in Y} f(y)$, i.e., f attains its infimum on a Y.*

Lemma B.4 *The following operations satisfy points (i–iii) of Theorem 5.3.*

binary version	*n-ary version*	
$x \oplus_p y = px + (1-p)y$	$\bigoplus_i p_i x_i = \sum_i p_i x_i$	*the standard operation*
$x \oplus_p y = \max\{x, y\}$	$\bigoplus_i p_i x_i = \max\{x_i\}_i$	$\vee$*-semilattice*
$x \oplus_p y = (px^k + (1-p)y^k)^{\frac{1}{k}}$	$\bigoplus_i p_i x_i = (\sum_i p_i x^k)^{\frac{1}{k}}$	*for some* $k \geq 1$
$x \oplus_p y = x^p \cdot y^{1-p}$	$\bigoplus_i p_i x_i = \prod_i {x_i}^{p_i}$	*log-probabilities*

Proof. Points (i) and (ii) are straightforward to verify. It remains to verify (iii), that is: for all $d : A \times A \to [0,1]$ and $\mu, \nu \in D(A)$, there exists an optimal coupling $\gamma_\star \in \Gamma(\mu, \nu)$:

$$V_d^{\oplus}(\gamma_\star) = \inf_{\gamma \in \Gamma(\mu,\nu)} V_d^{\oplus}(\gamma)$$

We establish point (iii) by proving that all $V_d^{\oplus} : D(A \times A) \to [0,1]$ are lower semicontinuous, which guarantees the result by Lemma B.3 and compactness of $\Gamma(\mu, \nu)$ (Proposition 2.4). The three cases:

$$\text{Case } x \oplus_p y = px + (1-p)y \qquad V_d^{\oplus}(\gamma) = \sum_{\langle a,b\rangle \in supp(\gamma)} \gamma(\langle a,b\rangle) d(a,b)$$

$$\text{Case } x \oplus_p y = (px^k + (1-p)x^k)^{\frac{1}{k}} \qquad V_d^{\oplus}(\gamma) = \sqrt[k]{\sum_{\langle a,b\rangle \in supp(\gamma)} \gamma(\langle a,b\rangle) d(a,b)^k}$$

$$\text{Case } x \oplus_p y = x^p \cdot y^{1-p} \qquad V_d^{\oplus}(\gamma) = \prod_{\langle a,b\rangle \in supp(\gamma)} d(a,b)^{\gamma(\langle a,b\rangle)}$$

are all continuous, hence lower semicontinuous. The last case:

$$\text{Case } x \oplus_p y = \max\{x,y\} \qquad V_d^{\oplus}(\gamma) = \max_{\langle a,b\rangle \in supp(\gamma)} d(a,b)$$

is not continuous. To see this, take $A = \{a,b\}$ and d the discrete metric on A: $d(a,a) = d(b,b) = 0$ and $d(a,b) = d(b,a) = 1$. Then $V_d^{\oplus}(\delta_{\langle a,a\rangle}) = 0$ and, for all $p > 0$, $V_d^{\oplus}(\delta_{\langle a,a\rangle} +_p \delta_{\langle a,b\rangle}) = 1$. Therefore:

$$0 = V_d^{\oplus}\Big(\lim_{p\to 0}(\delta_{\langle a,a\rangle} +_p \delta_{\langle a,b\rangle})\Big) \neq \lim_{p\to 0} V_d^{\oplus}\Big(\delta_{\langle a,a\rangle} +_p \delta_{\langle a,b\rangle}\Big) = 1.$$

However we now show that $V_d^{\oplus}$ is lower semicontinuous. First, we use the well-known fact that the ∞-Wasserstein is the limit of all k-Wasserstein. This follows from the following equality, for all $x, y \in [0,1]$:

$$\max\{x,y\} = \sup_{k\geq 1} \sqrt[k]{px^k + (1-p)y^k} \tag{B.1}$$

which implies (passing to n-ary version) that, for all $\gamma \in D(A \times A)$, we have:

$$\max_{\langle a,b\rangle \in supp(\gamma)} d(a,b) = \sup_{k\geq 1} \Big(\sqrt[k]{\sum_{\langle a,b\rangle \in supp(\gamma)} \gamma(\langle a,b\rangle) d(a,b)^k} \Big) \tag{B.2}$$

Secondly, as already observed, for every $k \geq 1$ the right hand side expressions (inside the supremum) is continuous in γ. Hence $V_d^{\oplus}$ is the pointwise supremum of a family of continuous functions, and as such is lower semicontinuous (see [9, Thm 4, p. 362]).

A standard proof of B.1 goes as follows. Assume, without loss of generality that $x \geq y$, i.e., $x = \max\{x,y\}$. For large k, y^k is much smaller than x^k. Hence $px^k + (1-p)y^k \approx px^k$ and therefore $\sqrt[k]{x^k + (1-p)y^k} \approx \sqrt[k]{px^k} = p^{\frac{1}{k}} x$. Since $\lim_{k\to\infty} p^{\frac{1}{k}} = 1$ we thus have $\lim_{k\to\infty} \sqrt[k]{px^k + (1-p)y^k} = x$. □