J. Korean Math. Soc. **55** (2018), No. 2, pp. 447–469
https://doi.org/10.4134/JKMS.j170297
pISSN: 0304-9914 / eISSN: 2234-3008

TYPE SPACES AND WASSERSTEIN SPACES

Shichang Song

Abstract. Types (over parameters) in the theory of atomless random variable structures correspond precisely to (conditional) distributions in probability theory. Moreover, the logic (resp. metric) topology on the type space corresponds to the topology of weak (resp. strong) convergence of distributions. In this paper, we study metrics between types. We show that type spaces under d^*-metric are isometric to Wasserstein spaces. Using optimal transport theory, two formulas for the metrics between types are given. Then, we give a new proof of an integral formula for the Wasserstein distance, and generalize some results in optimal transport theory.

1. Introduction

Continuous logic is a continuous version of first order logic that has been developed recently; see [3] and [5] for reference. The set of truth values in continuous logic is the interval $[0, 1]$ instead of the truth values $\{\text{True, False}\}$ in classical logic. Continuous logic is better suited for applications to metric structures than classical first order logic.

In the setting of continuous logic, Ben Yaacov [1] studied the theories of ($[0, 1]$-valued) random variable structures and their atomless counterparts. He axiomatized the theory of random variable structures by RV and its atomless counterpart by ARV. For the theory ARV, Ben Yaacov studied separable categoricity, type spaces, quantifier elimination, ω-stability, *etc* in [1]. On type spaces of ARV, motivated by the work of Berkes and Rosenthal [7], Ben Yaacov, Berenstein, and Henson [2] studied three different topologies on the space of types: the logic topology, the metric topology and the canonical base topology, in which the last two agree. They showed that types in the theory ARV correspond to conditional distributions. More specifically, in [2, Theorem 3.3] (resp. [2, Theorem 6.1]), they showed that the logic (resp. metric) topology on

Received May 3, 2017; Accepted June 26, 2017.
2010 *Mathematics Subject Classification.* 03C90, 60B10.
Key words and phrases. random variables, type spaces, Wasserstein distances.
This work was partially supported by "National Natural Science Fund of China" with Grant No. 11401020.

the type space corresponds to the topology of weak (resp. strong) convergence of distributions.

In this paper, following the work of [2], we study type spaces of ARV. In [2], they studied topologies on type spaces, while we consider metrics on type spaces. We realize that type spaces of ARV and Wasserstein spaces from optimal transport theory are closely related. We define a metric d^* between types, which is equivalent to the usual d-metric between types in continuous logic. Under d^*, type spaces are isometric to Wasserstein spaces. Our main results are two formulas for the d-metric and the d^*-metric between types. Theorem 4.5 is for a special case and Theorem 4.10 is for the most general case. Theorem 4.5 generalizes a classical result Theorem 2.6, which gives an integral formula for the Wasserstein distance between probability measures on the real line. To prove Theorem 4.10, we use the Kantorovich-Rubinstein duality formula from optimal transport theory. Then, using model theoretic results and rearrangements of measurable functions, we present a new proof of Theorem 2.6. Finally, we borrow ideas from model theory to define conditional Wasserstein distances. Our definition of conditional Wasserstein distances generalizes the definitions of Wasserstein distances. We apply model theory to prove some results for conditional Wasserstein spaces.

We assume the reader is familiar with basics of continuous logic. This paper is organized as follows. Section 2 gives the background and notations from analysis. We introduce rearrangements and Wasserstein distances in optimal transport. In Section 3, we present basics of the theory ARV. In Section 4, we study metrics on type spaces of ARV. Two formulas for the d-metric and the d^*-metric between types in ARV are given; see Theorems 4.5 and 4.10. In Section 5, a new proof of an integral formula for the Wasserstein distance between probability measures on the real line is given. In Section 6, we study conditional Wasserstein distances.

2. Background from analysis

In this section, we present the background and notations from analysis such as rearrangements and Wasserstein distances. They are used to prove Theorems 4.5 and 4.10.

2.1. Rearrangements

In this section, we discuss the notion of symmetric rearrangement of a Borel set and symmetric-decreasing rearrangement of a Borel measurable function. Then we introduce Lemma 2.2, which will be used to prove Theorem 4.5.

For a Borel subset $A \subseteq \mathbb{R}$ of finite Lebesgue measure, A^* is called the *symmetric rearrangement of the set A* if it is the open interval centered at the origin whose length is the measure of A. Hence, $A^* = \left\{x \mid |x| < \frac{\lambda(A)}{2}\right\}$, where λ is the standard Lebesgue measure on $\mathbb{R}$. For a Borel measurable function

$f\colon \mathbb{R} \to \mathbb{R}$, vanishing at infinity, f^* is the called the *symmetric-decreasing rearrangement of* f if $f^*(x) = \int_0^\infty \chi_{\{|f|>t\}^*}(x)dt$.

Fact 2.1 ([11, Pages 80–81])**.** (i) *$f^*(x)$ is nonnegative.*

(ii) *$f^*(x)$ is radially symmetric and nonincreasing; i.e., $f^*(x) = f^*(y)$ if $|x| = |y|$ and $f^*(x) \geq f^*(y)$ if $|x| \leq |y|$.*

(iii) *$f^*(x)$ is lower semi-continuous.*

(iv) *For every $t > 0$, $\{x \mid f^*(x) > t\} = \{x \mid |f(x)| > t\}^*$. An easy, but important, consequence of this fact is that $\operatorname{dist}(f^*(x)) = \operatorname{dist}(|f(x)|)$.*

Lemma 2.2 (Nonexpansivity of rearrangement, [11, Theorem 3.5])**.** *Let $J\colon \mathbb{R} \to \mathbb{R}$ be a nonnegative convex function such that $J(0) = 0$. Let f and g be nonnegative functions on $\mathbb{R}$, vanishing at infinity. Then*

$$\int_{\mathbb{R}} J\big(f^*(x) - g^*(x)\big)dx \leq \int_{\mathbb{R}} J\big(f(x) - g(x)\big)dx. \tag{1}$$

If we also assume that J is strictly convex, $f = f^*$, and f is strictly decreasing, then equality in (1) *implies that $g = g^*$.*

Now we define the rearrangements of Borel functions from $[0,1]$ to $[0,1]$, which will be used to give a new proof of Theorem 2.6 in Section 5.

For a Borel measurable function $f\colon [0,1] \to [0,1]$, we define $\tilde{f}\colon \mathbb{R} \to \mathbb{R}$ as follows:

$$\tilde{f}(x) = \begin{cases} f(x) & x \in [0,1], \\ f(-x) & x \in [-1,0], \\ 0 & \text{otherwise.} \end{cases}$$

Then the symmetric-decreasing rearrangement $\tilde{f}^*$ of $\tilde{f}$ is as follows:

$$\tilde{f}^*(x) = \int_0^\infty \chi_{\{|\tilde{f}|>t\}^*}(x)dt = \int_0^1 \chi_{\{\tilde{f}>t\}^*}(x)dt.$$

Let $f^*(x)$ be the restriction of $\tilde{f}^*$ on $[0,1]$. Then $f^*(x)$ is from $[0,1]$ to $[0,1]$. We call f^* the *decreasing rearrangement* of f. By Fact 2.1, we get

$$\tilde{f}^*(x) = \begin{cases} f^*(x) & x \in [0,1], \\ f^*(-x) & x \in [-1,0], \\ 0 & \text{otherwise.} \end{cases}$$

Following Fact 2.1, we have the following.

Fact 2.3. *Let f be a Borel measurable function from $[0,1]$ to $[0,1]$. Let f^* be the decreasing rearrangement of f. Then*

(i) *$f^*(x)$ is nonincreasing and lower semi-continuous.*

(ii) $\operatorname{dist}\big(f^*(x)\big) = \operatorname{dist}\big(f(x)\big)$.

Corollary 2.4 (Nonexpansivity of rearrangement)**.** *Let $J\colon [0,1] \to [0,1]$ be a nonnegative convex function such that $J(0) = 0$. Let $f, g\colon [0,1] \to [0,1]$ be*

Borel. Then

$$\int_0^1 J\big(f^*(x) - g^*(x)\big)dx \le \int_0^1 J\big(f(x) - g(x)\big)dx.$$

Proof. This follows directly from Lemma 2.2. □

2.2. Wasserstein distances

In this section, we present the necessary background on Wasserstein distances, which will be used in Section 4. In the rest of this subsection, we will mostly follow the notions and results in Villani's book [16, Chapter 6].

Let $(\mathcal{X}, d)$ be a Polish metric space, and let $p \in [1, \infty)$. For probability measures μ, ν on $\mathcal{X}$, the *Wasserstein distance of order* p between μ and ν is defined as follows:

$$\begin{aligned} W_p(\mu, \nu) &= \inf_{\pi \in \Pi(\mu,\nu)} \Big(\int_{\mathcal{X}} d(x,y)^p d\pi(x,y) \Big)^{1/p} \\ &= \inf \Big\{ \big[\mathbb{E} d(X,Y)^p\big]^{\frac{1}{p}} \mid \mathrm{dist}(X) = \mu, \mathrm{dist}(Y) = \nu \Big\}, \end{aligned}$$

where $\Pi(\mu, \nu)$ denote the collection of all probability measures on $\mathcal{X} \times \mathcal{X}$ with marginals μ and ν on the first and second factors respectively.

Let $P(\mathcal{X})$ be the space of Borel probability measures on $\mathcal{X}$. The *Wasserstein space of order* p is defined as

$$P_p(\mathcal{X}) := \big\{\mu \mid \mu \in P(\mathcal{X}) \text{ and } \int_{\mathcal{X}} d(x_0, x)^p \mu(dx) < +\infty \text{ for some } x_0 \in \mathcal{X}\big\}.$$

Throughout this paper, a Wasserstein space is always equipped with a distance W_p for corresponding p. For $p = 1$, we have the following formula:

Theorem 2.5 (Kantorovich-Rubinstein duality formula, [16, (6.3)]). *Let $(\mathcal{X}, d)$ be a Polish metric space. Then for all $\mu, \nu \in P_1(\mathcal{X})$,*

$$W_1(\mu, \nu) = \sup_{\|\psi\|_{Lip} \le 1} \Big\{ \int_{\mathcal{X}} \psi d\mu - \int_{\mathcal{X}} \psi d\nu \Big\}, \tag{2}$$

where the Lipschitz norm of a real valued function ψ on $\mathcal{X}$ is defined as

$$\|\psi\|_{Lip} := \sup_{x \ne y} \frac{|\psi(x) - \psi(y)|}{d(x,y)}.$$

The following theorem gives an integral formula for the Wasserstein distance between probability distributions on the line, which is used in the proof of Theorem 4.5. Meanwhile, Theorem 4.5 generalizes Theorem 2.6.

Theorem 2.6. *For all $\mu, \nu \in P_1(\mathbb{R})$, one has*

$$W_1(\mu, \nu) = \int_{-\infty}^{+\infty} |F(x) - G(x)| dx,$$

where F and G are distribution functions for μ and ν respectively.

Proof. For a proof using the Kantorovich-Rubinstein duality formula, see Proposition 20.11 in Dudley's book [9]. □

Remark 2.7. Theorem 2.6 was discovered independently by the following people at least: Dall'Aglio [8], Vallander [15], Szulga [14]. In Section 5, the author will revisit this theorem and will give his proof using rearrangements and model theoretic results; see Theorem 5.2.

Now we discuss convergence in Wasserstein spaces. Let $p \in [1, \infty)$ and let $(\mu_i)_{i\in\mathbb{N}}$ and μ be in $P_p(\mathcal{X})$. The notation $\mu_i \to \mu$ means that μ_i *converges weakly* to μ; *i.e.*, $\int_{\mathcal{X}} \varphi(x)\mu_i(dx) \to \int_{\mathcal{X}} \varphi(x)\mu(dx)$ for all bounded continuous functions $\varphi\colon \mathcal{X} \to \mathbb{R}$. We say $(\mu_i)_{i\in\mathbb{N}}$ *converges weakly in* $P_p(\mathcal{X})$ if for some (and thus every) $x_0 \in \mathcal{X}$, one has $\mu_i \to \mu$ and

$$\int_{\mathcal{X}} d(x_0, x)^p \mu_i(dx) \to \int_{\mathcal{X}} d(x_0, x)^p \mu(dx).$$

Theorem 2.8 ([16, Theorem 6.9])**.** *Let $(\mathcal{X}, d)$ be a Polish space and let $p \in [1, \infty)$. Then the topology defined by the Wasserstein distance W_p is the same as the topology of weak convergence in $P_p(\mathcal{X})$. In other words, if $(\mu_i)_{i\in\mathbb{N}}$ is a sequence in $P_p(\mathcal{X})$ and μ is in $P_p(\mathcal{X})$, then μ_i converges weakly in $P_p(\mathcal{X})$ to μ if and only if $W_p(\mu_i, \mu) \to 0$.*

Remark 2.9. we will see that Theorem 2.8 is a consequence of the separable categoricity of ARV by the Ryll-Nardzewski Theorem for continuous logic.

3. Basics of the theory ARV

In this section, we summarize basic model theoretic properties of the theory ARV and results about topologies on type spaces of ARV.

3.1. The theory ARV

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. It is *atomless* if for every $A \in \mathcal{F}$ with $\mu(A) > 0$, there exists $B \in \mathcal{F}$ such that $B \subseteq A$ and $0 < \mu(B) < \mu(A)$. Let $L^1\big((\Omega, \mathcal{F}, \mu), [0,1]\big)$, or $L^1(\mu, [0,1])$, denote the L^1-space of classes of $[0,1]$-valued $\mathcal{F}$-measurable functions equipped with L^1-metric. A ($[0,1]$-valued) *random variable structure* is based on a set of the form $M = L^1\big((\Omega, \mathcal{F}, \mu), [0,1]\big)$, where $(\Omega, \mathcal{F}, \mu)$ is a probability space. It is called an *atomless random variable structure*, if its underlying probability space is atomless. We consider the signature $L_{\mathsf{RV}} = \{\mathbf{0}, \neg, \dot{-}, \frac{1}{2}, I\}$, where $\mathbf{0}$ is a constant symbol, $\dot{-}$ is a binary function symbol, $\neg$ and $\frac{1}{2}$ are unary function symbols, and I is a unary predicate symbol. Symbols of L_{RV} in M are interpreted naturally. Then $\mathcal{M} = \Big(L^1\big((\Omega, \mathcal{F}, \mu), [0,1]\big), \mathbf{0}, \neg, \dot{-}, \frac{1}{2}, I, d\Big)$ is an L_{RV}-structure. Ben Yaacov [1] axiomatized the class of all $[0,1]$-valued random variable structures as L_{RV}-structures by the theory RV and its atomless counterpart by the theory ARV.

The following theorem includes basic properties of the theory ARV.

Theorem 3.1 ([1, Theorem 2.17]). *Let $\mathcal{M} = L^1\big((\Omega, \mathcal{F}, \mu), [0,1]\big)$ be a model of* ARV*. Then the theory* ARV *is complete, separably categorical, and admits quantifier elimination. The universal part of* ARV *is* RV*, and* ARV *is the model completion of* RV*. If $A \subseteq M$, then $\mathrm{dcl}(A) = \mathrm{acl}(A) = L^1\big((\Omega, \sigma(A), \mu), [0,1]\big) \subseteq M$, where $\sigma(A)$ is the σ-algebra of measurable sets generated by the random variables in A. Two tuples f and g in M^n have the same type over a set $A \subseteq M$ if and only if they have the same joint conditional distribution over $\sigma(A)$.*

3.2. The topologies on type spaces

Here, we introduce the results from [2] that connect the logic (resp. metric) topology with the topology of weak (resp. strong) convergence. We begin by defining some notions of convergence of distributions from probability theory, *e.g.*, weak convergence and strong convergence. We follow the definitions in [7].

Let $(\Omega, \mathscr{A}, m)$ be a probability space. An *n-dimensional conditional distribution over $\mathscr{A}$*, denoted by μ, is an $L^1\big((\Omega, \mathscr{A}, m), [0,1]\big)$-valued Borel probability measure on $\mathbb{R}^n$. More precisely, it satisfies the following:

- $\mu(B) \geq 0$ a.s. for all Borel sets $B \subseteq \mathbb{R}^n$.
- $\mu(\mathbb{R}^n) = 1$ a.s.
- $\mu(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mu(B_i)$ a.s. for all disjoint Borel sets $B_1, B_2, \ldots \subseteq \mathbb{R}^n$.

Let $\mathfrak{D}_{\mathbb{R}^n}(\mathscr{A})$ denote the space of all n-dimensional conditional distributions over $\mathscr{A}$. For every Borel set $B \subseteq \mathbb{R}^n$, we denote by $\mathfrak{D}_B(\mathscr{A})$ the space of all n-dimensional conditional distributions over $\mathscr{A}$ which, as measures, are supported by B. Note that the n-dimensional conditional distributions over $\emptyset$ are exactly the n-dimensional distributions.

Let f be an n-tuple of real-valued random variables. The (*joint*) *conditional distribution of f over $\mathscr{A}$*, denoted by $\mu = \mathrm{dist}(f \mid \mathscr{A})$, is an n-dimensional conditional distribution over $\mathscr{A}$ given by

$$\mu(B) = \mathbb{P}(f \in B \mid \mathscr{A}) \text{ for all Borel sets } B \subseteq \mathbb{R}^n.$$

A sequence $(f_i)_{i\in\mathbb{N}} \subseteq L^1\big((\Omega, \mathscr{A}, m), [0,1]\big)$ *converges weakly to f* if for every $g \in L^\infty\big((\Omega, \mathscr{A}, m), [0,1]\big)$, we have $\mathbb{E}(f_i g) \to \mathbb{E}(fg)$. We say the sequence $(f_i)_{i\in\mathbb{N}}$ *converges strongly to f* if it converges in L^1. Let μ and $(\mu_i)_{i\in\mathbb{N}}$ be n-dimensional conditional distributions over $\mathscr{A}$. We say the sequence $(\mu_i)_{i\in\mathbb{N}}$ *converges weakly* (*resp. strongly*) to μ if for every x with $\mu(\{x\}) = 0$, we have $\mu_i\big((-\infty, x]\big)$ converges weakly (resp. strongly) to $\mu\big((-\infty, x]\big)$.

Theorem 3.2 ([2, Theorems 3.3 and 6.1]). *Let f be an n-tuple in a model $\mathcal{M}$ of* ARV *and let A be a subset of M. Let $\mathscr{A}$ denote $\sigma(A)$, the σ-algebra of measurable sets generated by the random variables in A. Then the joint conditional distribution* $\mathrm{dist}(f \mid \mathscr{A})$ *only depends on* $\mathrm{tp}(f/A)$*. Moreover,*

(i) *the mapping*

$$\begin{aligned}\zeta\colon S_n(A) &\to \mathfrak{D}_{[0,1]^n}(\mathscr{A})\\ \operatorname{tp}(f/A) &\mapsto \operatorname{dist}(f \mid \mathscr{A})\end{aligned}$$

between $S_n(A)$ equipped with the logic topology and $\mathfrak{D}_{[0,1]^n}(\mathscr{A})$ equipped with the topology of weak convergence is a homeomorphism.

(ii) *We identify $S_n(A)$ with $\mathfrak{D}_{[0,1]^n}(\mathscr{A})$. Then the metric topology and the topology of strong convergence of conditional distributions agree.*

From now on, because of the above theorem, we use ζ to identify $S_n(A)$, the set of n-types over A, with $\mathfrak{D}_{[0,1]^n}\big(\sigma(A)\big)$, the set of n-dimensional conditional distributions over $\sigma(A)$.

4. The metrics on type spaces of ARV

In this section, we study the d-metric and the d^*-metric between types. The main results are two formulas for the d-metric and the d^*-metric between types. Theorem 4.5 is for a special case and Theorem 4.10 is for the most general case. Note that Theorem 4.5 generalizes Theorem 2.6.

First, we show a theorem concerning the relation between type spaces of ARV and Wasserstein spaces. Fix a signature L for metric structures and an consistent L-theory T. Let $\mathcal{M}$ be a model of T and let $A \subseteq M$. Denote the $L(A)$-structure $(\mathcal{M}, a)_{a\in A}$ by $\mathcal{M}_A$, and set T_A to be the $L(A)$-theory of $\mathcal{M}_A$.

Definition 4.1. For each $n \geq 1$, let $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$ be in a metric structure M^n, we define $d(a, b)$ to be $\max_{1\leq i\leq n} d(a_i, b_i)$, and $d^*(a, b)$ to be $\sum_{i=1}^n d(a_i, b_i)$. Let T_A be as above and let $\mathcal{M}_A = (\mathcal{M}, a)_{a\in A}$ be any model of T_A in which every type in $S_n(T_A)$ is realized. For all $p, q \in S_n(T_A)$, we define $d(p, q)$ to be

$$\inf\big\{\max_{1\leq i\leq n} d(b_i, c_i) \mid \mathcal{M}_A \models p[b_1, \dots, b_n] \text{ and } \mathcal{M}_A \models q[c_1, \dots, c_n]\big\},$$

and $d^*(p, q)$ to be

$$\inf\big\{\sum_{i=1}^n d(b_i, c_i) \mid \mathcal{M}_A \models p[b_1, \dots, b_n] \text{ and } \mathcal{M}_A \models q[c_1, \dots, c_n]\big\}.$$

Note that $d(a, b) \leq d^*(a, b) \leq nd(a, b)$ and $d(p, q) \leq d^*(p, q) \leq nd(p, q)$. The metric d^* therefore defines the same topology on type spaces as the usual metric d. Model theoretic results such as the Ryll-Nardzewski Theorem for continuous logic only depend on the metric topology on type spaces, so this d^*-metric is adequate for model theoretic purposes. The reason for us to prefer d^*-metric is the following:

Theorem 4.2. *Let f be an n-tuple in a model $\mathcal{M}$ of* ARV*. Then* $\operatorname{dist}(f)$ *is a probability measure on $[0,1]^n$. Moreover, the mapping*

$$\eta_n\colon S_n(\mathsf{ARV}) \to P_1\big([0,1]^n\big)$$

$$\mathrm{tp}(f) \mapsto \mathrm{dist}(f)$$

is an isometric isomorphism between $\big(S_n(\mathsf{ARV}), d^*\big)$ *and* $\Big(P_1\big([0,1]^n\big), W_1\Big)$, *where the metric on which the definition of* W_1 *depends is defined by*

$$d^*_{[0,1]^n}(a,b) = \sum_{i=1}^{n} |a_i - b_i| \text{ for all } a, b \in [0,1]^n.$$

Proof. By Theorem 3.2, we know that η_n is well-defined and bijective. Let f and g be n-tuples in M.

$$\begin{aligned} & W_1\big(\mathrm{dist}(f), \mathrm{dist}(g)\big) \\ &= \inf\Big\{\mathbb{E}\big(\sum_{i=1}^{n} |X_i - Y_i|\big) : \mathrm{dist}(X) = \mathrm{dist}(f), \mathrm{dist}(Y) = \mathrm{dist}(g)\Big\} \\ &= \inf\Big\{\sum_{i=1}^{n} \mathbb{E}\big(|X_i - Y_i|\big) : \mathrm{dist}(X) = \mathrm{dist}(f), \mathrm{dist}(Y) = \mathrm{dist}(g)\Big\} \\ &= d^*\big(\mathrm{tp}(f), \mathrm{tp}(g)\big). \end{aligned}$$

□

Corollary 4.3. *Let f be an n-tuple in a model $\mathcal{M}$ of* ARV. *Then* $\mathrm{dist}(f)$ *is a probability measure on* $[0,1]^n$. *Moreover, the mapping*

$$\eta_n \colon S_n(\mathsf{ARV}) \to P_1\big([0,1]^n\big)$$
$$\mathrm{tp}(f) \mapsto \mathrm{dist}(f)$$

is bijective and bi-Lipschitz between $\big(S_n(\mathsf{ARV}), d\big)$ *and* $\Big(P_1\big([0,1]^n\big), W_1\Big)$, *where the metric on which the definition of* W_1 *depends is defined by*

$$d_{[0,1]^n}(a,b) = \max_{1 \le i \le n} |a_i - b_i| \quad \text{for all } a, b \in [0,1]^n.$$

In particular, when $n = 1$, η_1 is an isometry.

Proof. This follows from Theorem 4.2 and the fact that $d(p,q) \le d^*(p,q) \le nd(p,q)$ for all $p, q \in S_n(\mathsf{ARV})$. □

An explicit formula for a special case

Now, we apply model theory to prove Theorem 4.5, which gives an explicit formula for the d-metric between types for a special case, but still generalizes the formula in Theorem 2.6. In the proof of it, the following result is required.

Proposition 4.4. *Let $(\Omega, \mathcal{F}, \mu)$ be an atomless probability space and let $f, g \colon \Omega \to [0,1]$ be random variables. Suppose $\mu\big(\mathrm{supp}(f)\big) \le \mu\big(\mathrm{supp}(g)\big)$.*

(i) *There are random variables $\tilde{f}, \tilde{g} \colon \Omega \to [0,1]$ such that $\mathrm{dist}(\tilde{f}) = \mathrm{dist}(f)$, $\mathrm{dist}(\tilde{g}) = \mathrm{dist}(g)$, and $d(\tilde{f}, \tilde{g}) = W_1\big(\mathrm{dist}(f), \mathrm{dist}(g)\big)$. Moreover, for any such random variables $\tilde{f}$ and $\tilde{g}$, one has $\mathrm{supp}(\tilde{f}) \subseteq \mathrm{supp}(\tilde{g})$ up to a null set.*

(ii) *Let* $A \subseteq B$ *be in* $\mathcal{F}$ *with* $\mu(A) = \mu\big(\mathrm{supp}(f)\big)$ *and* $\mu(B) = \mu\big(\mathrm{supp}(g)\big)$. *Then there are* $f', g' \colon \Omega \to [0,1]$ *satisfying*:
- $\mathrm{supp}(f') = A$ *and* $\mathrm{supp}(g') = B$;
- $\mathrm{dist}(f) = \mathrm{dist}(f')$ *and* $\mathrm{dist}(g) = \mathrm{dist}(g')$;
- $d(f', g') = W_1\big(\mathrm{dist}(f'), \mathrm{dist}(g')\big)$.

Proof. (i) By Theorem 3.1, the theory ARV is separably categorical. Hence there are $\tilde{f}, \tilde{g} \colon \Omega \to [0,1]$ such that $\mathrm{tp}(f) = \mathrm{tp}(\tilde{f})$, $\mathrm{tp}(g) = \mathrm{tp}(\tilde{g})$, and $d(\tilde{f}, \tilde{g}) = d(\mathrm{tp}(f), \mathrm{tp}(g))$. Then by Theorem 3.1 and Theorem 4.2, one has that $\mathrm{dist}(\tilde{f}) = \mathrm{dist}(f)$, $\mathrm{dist}(\tilde{g}) = \mathrm{dist}(g)$, and $d(\tilde{f}, \tilde{g}) = d^*(\tilde{f}, \tilde{g}) = W_1\big(\mathrm{dist}(f), \mathrm{dist}(g)\big)$.

Suppose $\mathrm{supp}(\tilde{f})$ is not contained in $\mathrm{supp}(\tilde{g})$ up to a null set. Let $D = \mathrm{supp}(\tilde{f}) \backslash \mathrm{supp}(\tilde{g})$, so $\mu(D) > 0$. Since $\mathrm{dist}(f) = \mathrm{dist}(\tilde{f})$ and $\mathrm{dist}(g) = \mathrm{dist}(\tilde{g})$, one has that $\mu(\mathrm{supp}(\tilde{f})) = \mu(\mathrm{supp}(f)) \leq \mu(\mathrm{supp}(\tilde{g})) = \mu(\mathrm{supp}(g))$. Because Ω is atomless, there is a measurable set $C \subseteq \mathrm{supp}(\tilde{g}) \backslash \mathrm{supp}(\tilde{f})$ such that $\mu(C) = \mu(D)$. Let Ω_C and Ω_D denote the probability spaces $(C, \mathcal{F} \restriction C, \tilde{\mu})$ and $(D, \mathcal{F} \restriction D, \tilde{\mu})$ respectively, where $\tilde{\mu} = \frac{\mu}{\mu(C)} = \frac{\mu}{\mu(D)}$. Then $\tilde{f} \restriction D \in L^1(\Omega_D, [0,1])$. By Theorem 3.1, the theory ARV is separably categorical, so all types in $S_n(ARV)$ are realized in any model of ARV. Thus $\mathrm{tp}(\tilde{f} \restriction D)$ is realized in $L^1(\Omega_C, [0,1])$, say by $h \colon C \to [0,1]$. Then by Theorem 3.1, $\mathrm{dist}(h) = \mathrm{dist}(\tilde{f} \restriction D)$. We define $\tilde{f}' \colon \Omega \to [0,1]$ as follows:

$$\tilde{f}'(\omega) = \begin{cases} h(\omega) & \omega \in C, \\ 0 & \omega \in D, \\ \tilde{f}(\omega) & \text{elsewhere.} \end{cases}$$

Then $\mathrm{dist}(\tilde{f}) = \mathrm{dist}(\tilde{f}')$, since $\mu(C) = \mu(D)$ and $\mathrm{dist}(h) = \mathrm{dist}(\tilde{f} \restriction D)$. Furthermore,

$$\begin{aligned} d(\tilde{f}', \tilde{g}) = \int_\Omega |\tilde{f}' - \tilde{g}| d\mu &= \int_{\Omega \backslash (C \cup D)} |\tilde{f}' - \tilde{g}| d\mu + \int_C |\tilde{f}' - \tilde{g}| d\mu + \int_D |\tilde{f}' - \tilde{g}| d\mu \\ &= \int_{\Omega \backslash (C \cup D)} |\tilde{f} - \tilde{g}| d\mu + \int_C |h - \tilde{g}| d\mu + \int_D |0 - \tilde{g}| d\mu \\ &< \int_{\Omega \backslash (C \cup D)} |\tilde{f} - \tilde{g}| d\mu + \int_C (h + \tilde{g}) d\mu + 0 \\ &= \int_{\Omega \backslash (C \cup D)} |\tilde{f} - \tilde{g}| d\mu + \int_C \tilde{g} d\mu + \int_D \tilde{f} d\mu \\ &= \int_{\Omega \backslash (C \cup D)} |\tilde{f} - \tilde{g}| d\mu + \int_C |\tilde{f} - \tilde{g}| d\mu + \int_D |\tilde{f} - \tilde{g}| d\mu \\ &= \int_\Omega |\tilde{f} - \tilde{g}| d\mu = d(\tilde{f}, \tilde{g}) = W_1(\mathrm{dist}(f), \mathrm{dist}(g)). \end{aligned}$$

By Corollary 4.3, this contradicts the fact that $\mathrm{dist}(\tilde{f}') = \mathrm{dist}(f)$ and $\mathrm{dist}(\tilde{g}) = \mathrm{dist}(g)$.

(ii) By (i), there are random variables $\tilde{f}, \tilde{g} \colon \Omega \to [0,1]$ such that $\operatorname{dist}(\tilde{f}) = \operatorname{dist}(f)$, $\operatorname{dist}(\tilde{g}) = \operatorname{dist}(g)$, and $d(\tilde{f}, \tilde{g}) = W_1\big(\operatorname{dist}(f), \operatorname{dist}(g)\big)$. Moreover, we have $\operatorname{supp}(\tilde{f}) \subseteq \operatorname{supp}(\tilde{g})$ up to a null set. Let $C = \operatorname{supp}(\tilde{f})$, $D = \operatorname{supp}(\tilde{g})$, and assume that $C \subseteq D$. Then, one has that $\mu(C) = \mu(A)$ and $\mu(D) = \mu(B)$. Since $\mathcal{F}$ is atomless, there is an atomless and countably σ-generated σ-subalgebra $\mathcal{F}_0$ of $\mathcal{F}$ which contains A, B, C, D. Let $\mathcal{A} = \sigma(C \cup D)$ be the σ-subalgebra σ-generated by C and D in $\mathcal{F}_0$. Then we define $\varphi \colon \mathcal{A} \to \mathcal{F}_0$ by $\varphi(C) = A$ and $\varphi(D) = B$. Since $\mathcal{F}_0$ is atomless and countably σ-generated, $\mathcal{F}_0$ is a homogeneous probability algebra. By [10, Corollary 3.19], φ extends to an automorphism Φ of the probability algebra $\mathcal{F}_0$. Naturally, Φ induces an L_{RV}-automorphism of $L^1\big((\Omega, \mathcal{F}_0, \mu), [0,1]\big)$; we still call it Φ. Let $f' = \Phi(\tilde{f})$ and $g' = \Phi(\tilde{g})$. Then $\operatorname{dist}(f') = \operatorname{dist}(\tilde{f})$, $\operatorname{dist}(g') = \operatorname{dist}(\tilde{g})$, and $d(f', g') = d(\tilde{f}, \tilde{g})$. Moreover, $\operatorname{supp}(f') = A$ and $\operatorname{supp}(g') = B$. □

Next, for a special case, we prove an explicit formula for the d-metric between types.

Theorem 4.5. *Let κ be an uncountable cardinal. Let $\mathcal{M} \models \mathsf{ARV}$ be a κ-universal domain of the form $\big(L^1(\mu, [0,1]), \neg, \frac{1}{2}, \cdot, \dot{-}, I\big)$, where $(\Omega, \mathcal{F}, \mu)$ is an atomless probability space. Suppose $C \subseteq M = L^1(\mu, [0,1])$ is a small subset; let $\mathcal{C}$ be the σ-algebra of measurable sets generated by the random variables in C. Let $a = (a_1, \ldots, a_n) \in M^n$ and $b = (b_1, \ldots, b_n) \in M^n$ be disjointly supported; i.e., they satisfy $a_i \cdot a_j = b_i \cdot b_j = 0$ whenever $i \neq j$ and $1 \leq i, j \leq n$. Then*

$$d\big(\operatorname{tp}(a/C), \operatorname{tp}(b/C)\big) = \max_{1 \leq i \leq n} \int_0^1 \big\|\mathbb{P}(a_i > t \mid \mathcal{C}) - \mathbb{P}(b_i > t \mid \mathcal{C})\big\|_1 dt, \tag{3}$$

where $\|\cdot\|_1$ is the L^1-norm.

Proof. **Case 1**: We first consider the case in which $\mathcal{C} = \emptyset$. By definition,

$$d\big(\operatorname{tp}(a), \operatorname{tp}(b)\big) = \inf\big\{\max_{1 \leq i \leq n} I\big(|x_i - y_i|\big) : x, y \in M, x \models \operatorname{tp}(a), \text{ and } y \models \operatorname{tp}(b)\big\}.$$

Then

$$\begin{aligned} d\big(\operatorname{tp}(a), \operatorname{tp}(b)\big) &= \inf\big\{\max_{1 \leq i \leq n} I\big(|x_i - y_i|\big) : x, y \in M, x \models \operatorname{tp}(a), \text{ and } y \models \operatorname{tp}(b)\big\} \\ &\geq \max_{1 \leq i \leq n} \inf\big\{I\big(|x_i - y_i|\big) : x, y \in M, x \models \operatorname{tp}(a), \text{ and } y \models \operatorname{tp}(b)\big\} \\ &= \max_{1 \leq i \leq n} \int_0^1 \big|\mu(a_i > t) - \mu(b_i > t)\big| dt. \end{aligned}$$

So we need only show that

$$d\big(\operatorname{tp}(a), \operatorname{tp}(b)\big) \leq \max_{1 \leq i \leq n} \int_0^1 \big|\mu(a_i > t) - \mu(b_i > t)\big| dt.$$

Let $\big([0,1], \mathscr{B}, \lambda\big)$ be the standard Lebesgue space. Let $f := (f_1, \ldots, f_n)$ and $g := (g_1, \ldots, g_n)$ whose coordinates are elements of $L^1\big(\lambda, [0,1]\big)$, such that $\operatorname{dist}(f) = \operatorname{dist}(a)$ and $\operatorname{dist}(g) = \operatorname{dist}(b)$. Let A_i be the support set for f_i and

B_i for g_i, for every $i \le n$. Reordering the indices if necessary, we may assume that there is $l \le n$ such that $\mu(A_i) \ge \mu(B_i)$ if and only if $i \le l$. Let A_i' be the closed interval $\left[\sum_{j<i} \mu(A_j), \sum_{j \le i} \mu(A_j)\right]$ for $i \le n$. For $i \le l$, let $B_i' := \left[\sum_{j<i} \mu(A_j), \sum_{j<i} \mu(A_j) + \mu(B_i)\right]$. Thus, $B_i' \subseteq A_i'$ for $i \le l$. Since $a_1, \dots, a_n$ are disjointly supported and the same for $b_1, \dots, b_n$, we know that $f_1, \dots, f_n$ are disjointly supported and the same for $g_1, \dots, g_n$. For each $l < i \le n$, let B_i' be a union of closed intervals satisfying, $B_i' \supseteq A_i'$, $\lambda(B_i') = \mu(B_i)$, and B_i' is disjoint from B_j' for all $j \ne i$. For every $i \le n$, by Proposition 4.4 there are $f_i', g_i' \colon [0,1] \to [0,1]$ satisfying the following:

- $\mathrm{dist}(f_i') = \mathrm{dist}(f_i) = \mathrm{dist}(a_i)$, $\mathrm{dist}(g_i') = \mathrm{dist}(g_i) = \mathrm{dist}(b_i)$;
- The support set of f_i' and g_i' are A_i' and B_i', respectively;
- $d(f_i', g_i') = d\big(\mathrm{tp}(a_i), \mathrm{tp}(b_i)\big)$.

By Corollary 4.3 and Theorem 2.6, we have:

$$d(f_i', g_i') = d\big(\mathrm{tp}(a_i), \mathrm{tp}(b_i)\big) = \int_0^1 \big|\mu(a_i > t) - \mu(b_i > t)\big| dt.$$

Let $f' := (f_1', \dots, f_n')$ and $g' := (g_1', \dots, g_n')$. Note that $\mathrm{dist}(f') = \mathrm{dist}(f) = \mathrm{dist}(a)$ and $\mathrm{dist}(g') = \mathrm{dist}(g) = \mathrm{dist}(b)$ by the fact that they are disjointly supported. By Theorem 3.1, we have $\mathrm{tp}(f') = \mathrm{tp}(f) = \mathrm{tp}(a)$ and $\mathrm{tp}(g') = \mathrm{tp}(g) = \mathrm{tp}(b)$. Then,

$$d\big(\mathrm{tp}(a), \mathrm{tp}(b)\big) \le \max_{1 \le i \le n} d(f_i', g_i') = \max_{1 \le i \le n} \int_0^1 \big|\mu(a_i > t) - \mu(b_i > t)\big| dt.$$

Case 2: We next consider the case in which $\mathcal{C}$ is finite with atoms $C_1, \dots, C_m$. Let $a_{ij} = a_i \cdot \chi_{C_j}$ and $b_{ij} = b_i \cdot \chi_{C_j}$ for every $1 \le j \le m$ and every $1 \le i \le n$. Also for every $1 \le j \le m$, let $\bar{a}_j = (a_{1j}, \dots, a_{nj})$ and $\bar{b}_j = (b_{1j}, \dots, b_{nj})$. For each $1 \le j \le m$, by Proposition 4.4 there are $\bar{a}_j' = (a_{1j}', \dots, a_{nj}')$ and $\bar{b}_j' = (b_{1j}', \dots, b_{nj}')$, whose coordinates are supported in C_j such that $\mathrm{tp}(\bar{a}_j') = \mathrm{tp}(\bar{a}_j)$, $\mathrm{tp}(\bar{b}_j') = \mathrm{tp}(\bar{b}_j)$, and $d(a_{ij}', b_{ij}') = W_1\big(\mathrm{dist}(a_{ij}'), \mathrm{dist}(b_{ij}')\big)$ for all $1 \le i \le n$. Then by Corollary 4.3,

$$d(a_{ij}', b_{ij}') = d\big(\mathrm{tp}(a_{ij}'), \mathrm{tp}(b_{ij}')\big) = \int_0^1 |\mu(a_{ij}' > t) - \mu(b_{ij}' > t)| dt \text{ for all } i \le n.$$

Let $a_i' = \sum_{j=1}^m a_{ij}'$ and $b_i' = \sum_{j=1}^m b_{ij}'$. Let $a' = (a_1', \dots, a_n')$ and $b' = (b_1', \dots, b_n')$. For each $1 \le j \le m$,

$$\mathbb{P}(a_i > t \mid \mathcal{C}) = \sum_{j=1}^m \frac{\mu\big(\{a_i > t\} \cap C_j\big)\chi_{C_j}}{\mu(C_j)} = \sum_{j=1}^m \frac{\mu(a_{ij} > t)\chi_{C_j}}{\mu(C_j)}.$$

Thus we have $\mathrm{dist}\big(a_i \mid \sigma(C)\big) = \mathrm{dist}\big(a_i' \mid \sigma(C)\big)$ and $\mathrm{dist}\big(b_i \mid \sigma(C)\big) = \mathrm{dist}\big(b_i' \mid \sigma(C)\big)$. Then by Theorem 3.1, we have $\mathrm{tp}(a_i/C) = \mathrm{tp}(a_i'/C)$ and $\mathrm{tp}(b_i/C) = \mathrm{tp}(b_i'/C)$. By the fact that the coordinates of a, a', b', and b are disjointly supported, we have $\mathrm{tp}(a/C) = \mathrm{tp}(a'/C)$ and $\mathrm{tp}(b/C) = \mathrm{tp}(b'/C)$, since $\mathrm{dist}\big(a \mid$

$\sigma(C)\big) = \operatorname{dist}\big(a' \mid \sigma(C)\big)$ and $\operatorname{dist}\big(b \mid \sigma(C)\big) = \operatorname{dist}\big(b' \mid \sigma(C)\big)$. For all $1 \le i \le n$, we have

$$\begin{aligned} d(a'_i, b'_i) &= \int_0^1 |a'_i - b'_i| dt = \sum_{j=1}^m \int_{C_j} |a'_i - b'_i| dt = \sum_{j=1}^m \int_{C_j} |a'_{ij} - b'_{ij}| dt \\ &= \sum_{j=1}^m \int_0^1 |a'_{ij} - b'_{ij}| dt = \sum_{j=1}^m d(a'_{ij}, b'_{ij}). \end{aligned}$$

By the fact that $d(a'_{ij}, b'_{ij}) = \int_0^1 |\mu(a'_{ij} > t) - \mu(b'_{ij} > t)| dt$, $\operatorname{dist}\big(a \mid \sigma(C)\big) = \operatorname{dist}\big(a' \mid \sigma(C)\big)$ and $\operatorname{dist}\big(b \mid \sigma(C)\big) = \operatorname{dist}\big(b' \mid \sigma(C)\big)$, we have

$$d(a'_i, b'_i) = \sum_{j=1}^m d(a'_{ij}, b'_{ij}) = \sum_{j=1}^m \int_0^1 \big|\mu(a'_{ij} > t) - \mu(b'_{ij} > t)\big| dt.$$

Thus,

$$d(a'_i, b'_i) = \sum_{j=1}^m \int_0^1 \big|\mu(a_{ij} > t) - \mu(b_{ij} > t)\big| dt. \tag{4}$$

For all $f, g \in M$ with $\operatorname{tp}(f/C) = \operatorname{tp}(a/C)$ and $\operatorname{tp}(g/C) = \operatorname{tp}(b/C)$, we have that for each $1 \le i \le n$,

$$\begin{aligned} d(f_i, g_i) &= \sum_{j=1}^m d(f_{ij}, g_{ij}) \ge \sum_{j=1}^m d\big(\operatorname{tp}(f_{ij}), \operatorname{tp}(g_{ij})\big) \\ &= \sum_{j=1}^m \int_0^1 \big|\mu(f_{ij} > t) - \mu(g_{ij} > t)\big| dt \\ &= \sum_{j=1}^m \int_0^1 \big|\mu(a_{ij} > t) - \mu(b_{ij} > t)\big| dt, \end{aligned}$$

where $f_{ij} = f_i \cdot \chi_{C_j}$ and $g_{ij} = g_i \cdot \chi_{C_j}$. By Equation (4), we have $d(f, g) \ge d(a', b')$, whereby $d(a', b') = d\big(\operatorname{tp}(a/C), \operatorname{tp}(b/C)\big)$. Then by

$$\mathbb{P}(a_i > t \mid \mathcal{C}) = \sum_{j=1}^m \frac{\mu(a_{ij} > t)\chi_{C_j}}{\mu(C_j)},$$

we have

$$\begin{aligned} \big\|\mathbb{P}(a_i > t \mid \mathcal{C}) - \mathbb{P}(b_i > t \mid \mathcal{C})\big\|_1 &= \int_\Omega \Big|\sum_{j=1}^m \big(\mu(a_{ij} > t) - \mu(b_{ij} > t)\big) \frac{\chi_{C_j}}{\mu(C_j)}\Big| d\mu \\ &= \sum_{j=1}^m \int_{C_j} \frac{\big|\mu(a_{ij} > t) - \mu(b_{ij} > t)\big|}{\mu(C_j)} d\mu. \end{aligned}$$

Thus,

$$\big\|\mathbb{P}(a_i > t \mid \mathcal{C}) - \mathbb{P}(b_i > t \mid \mathcal{C})\big\|_1 = \sum_{j=1}^{m} \big|\mu(a_{ij} > t) - \mu(b_{ij} > t)\big|. \tag{5}$$

By Equations (4) and (5), we have

$$d(a_i', b_i') = \int_0^1 \big\|\mathbb{P}(a_i > t \mid \mathcal{C}) - \mathbb{P}(b_i > t \mid \mathcal{C})\big\|_1 dt,$$

whereby

$$d\big(\mathrm{tp}(a/C), \mathrm{tp}(b/C)\big) = d(a', b') = \max_{1\le i\le n} \int_0^1 \big\|\mathbb{P}(a_i > t \mid \mathcal{C}) - \mathbb{P}(b_i > t \mid \mathcal{C})\big\|_1 dt.$$

Case 3: Finally, suppose $\mathcal{C}$ is an infinite σ-algebra. Let $\{\mathcal{C}_j \mid j \in J\}$ be the family of all finite subalgebras of $\mathcal{C}$. Then $\mathcal{C} = \cup_{j\in J}\mathcal{C}_j$. Since all measurable functions are approximated by simple functions, $\mathbb{P}(a_i > t \mid \mathcal{C}) = \sup_{j\in J} \mathbb{P}(a_i > t \mid \mathcal{C}_j)$ *a.s.* and $\mathbb{P}(b_i > t \mid \mathcal{C}) = \sup_{j\in J} \mathbb{P}(b_i > t \mid \mathcal{C}_j)$ *a.s.* for each $1 \le i \le n$. Further,

$$\big\|\mathbb{P}(a_i > t \mid \mathcal{C}) - \mathbb{P}(b_i > t \mid \mathcal{C})\big\|_1 = \sup_{j\in J} \big\|\mathbb{P}(a_i > t \mid \mathcal{C}_j) - \mathbb{P}(b_i > t \mid \mathcal{C}_j)\big\|_1.$$

By the compactness theorem, we know that

$$d\big(\mathrm{tp}(a/C), \mathrm{tp}(b/C)\big) = \sup_{j\in J} d\big(\mathrm{tp}(a/\mathcal{C}_j), \mathrm{tp}(b/\mathcal{C}_j)\big).$$

Thus, by **Case 2** we have

$$d\big(\mathrm{tp}(a/C), \mathrm{tp}(b/C)\big) = \max_{1\le i\le n} \int_0^1 \big\|\mathbb{P}(a_i > t \mid \mathcal{C}) - \mathbb{P}(b_i > t \mid \mathcal{C})\big\|_1 dt. \qquad \square$$

Remark 4.6. (i) When a and b are just elements in M, Theorem 4.5 yields the following formula, which generalizes the formula in Theorem 2.6.

$$d\big(\mathrm{tp}(a/C), \mathrm{tp}(b/C)\big) = \int_0^1 \big\|\mathbb{P}(a > t|\mathcal{C}) - \mathbb{P}(b > t|\mathcal{C})\big\|_1 dt.$$

(ii) In Case 3, one has

$$d\big(\mathrm{tp}(a/C), \mathrm{tp}(b/C)\big) = \sup_{j\in J} d\big(\mathrm{tp}(a/\mathcal{C}_j), \mathrm{tp}(b/\mathcal{C}_j)\big)$$

holds in general and does not need assumptions on a and b in the lemma.

(iii) It follows from the proof of Theorem 4.5 that

$$d^*\big(\mathrm{tp}(a/C), \mathrm{tp}(b/C)\big) = \sum_{i=1}^{n} \int_0^1 \big\|\mathbb{P}(a_i > t|\mathcal{C}) - \mathbb{P}(b_i > t|\mathcal{C})\big\|_1 dt$$

for all disjointly supported $a = (a_1, \dots, a_n), b = (b_1, \dots, b_n) \in M^n$. Indeed, every occurrence of $\max_{1\le i\le n}$ in the proof can be replaced by $\sum_{i=1}^n$.

(iv) In an early draft of [6] with the title *Model theory of probability spaces with an automorphism*, Berenstein and Henson gave a similar explicit formula for the d-metric between types in the theory of atomless probability spaces. See more about their result at [3, Lemma 16.4].

Question 4.7. We see that by Theorem 4.5, for $a, b \in M^n$, being disjointly supported is a sufficient condition for Equation (3). Does Equation (3) hold for a more general class of $a, b \in M^n$? We know that Equation (3) does not hold for all $a, b \in M^n$, since in general the joint distributions of a and b are not determined by the individual distributions of their coordinates, while the right hand side of Equation (3) only depends on the distributions of a and b's coordinates.

The most general case

Now, we are about to prove Theorem 4.10, which considers arbitrary n-types over parameters in ARV. First, we present some results that will be used in the proof.

Lemma 4.8. *Let $\mathcal{M} \models \mathsf{ARV}$ be of the form $M = L^1\big(\mu, [0,1]\big)$ where $(\Omega, \mathcal{F}, \mu)$ is an atomless probability space, and let C be a subset of M. Let $\mathcal{C}$ denote the σ-algebra of measurable sets generated by the random variables in C. Suppose $\mathcal{C}$ is finite with atoms $C_1, \dots, C_m$. Let $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$ be tuples in M. Then*

$$\operatorname{tp}(a/C) = \operatorname{tp}(b/C) \Leftrightarrow \operatorname{tp}\Big(\big(\{a_{ij}\}_{1 \le i \le n, 1 \le j \le m}\big)\Big) = \operatorname{tp}\Big(\big(\{b_{ij}\}_{1 \le i \le n, 1 \le j \le m}\big)\Big),$$

where $a_{ij} = a_i \cdot \chi_{C_j}$ and $b_{ij} = b_i \cdot \chi_{C_j}$ for all $1 \le i \le n$ and $1 \le j \le m$.

Proof. We need only show from right to left. For all Borel $B \subseteq [0,1]^n$, we have

$$\begin{aligned}
&\mathbb{P}\big((a_1, \dots, a_n) \in B \mid \mathcal{C}\big)\\
&= \frac{\mu\big(C_1 \cap \{(a_1, \dots, a_n) \in B\}\big)}{\mu(C_1)} \chi_{C_1} + \cdots + \frac{\mu\big(C_n \cap \{(a_1, \dots, a_n) \in B\}\big)}{\mu(C_n)} \chi_{C_n}\\
&= \frac{\mu((a_{11}, \dots, a_{n1}) \in B)}{\mu(C_1)} \chi_{C_1} + \cdots + \frac{\mu((a_{1m}, \dots, a_{nm}) \in B)}{\mu(C_m)} \chi_{C_m}\\
&= \frac{\mu((b_{11}, \dots, b_{n1}) \in B)}{\mu(C_1)} \chi_{C_1} + \cdots + \frac{\mu((b_{1m}, \dots, b_{nm}) \in B)}{\mu(C_m)} \chi_{C_m}\\
&= \frac{\mu\big(C_1 \cap \{(b_1, \dots, b_n) \in B\}\big)}{\mu(C_1)} \chi_{C_1} + \cdots + \frac{\mu\big(C_n \cap \{(b_1, \dots, b_n) \in B\}\big)}{\mu(C_n)} \chi_{C_n}\\
&= \mathbb{P}\big((b_1, \dots, b_n) \in B \mid \mathcal{C}\big).
\end{aligned}$$

Then by Theorem 3.2, we know $\operatorname{tp}\big((a_1, \dots, a_n)/C\big) = \operatorname{tp}\big((b_1, \dots, b_n)/C\big)$. $\square$

Proposition 4.9. *Let $\mathcal{M} \models \mathsf{ARV}$ be an $\aleph_0$-universal domain and let C be a subset of M. Suppose that $M = L^1\big(\mu, [0,1]\big)$, where $(\Omega, \mathcal{F}, \mu)$ is an atomless probability space. Let $\mathcal{C}$ denote the σ-algebra of measurable sets generated by* *all the random variables in C. Suppose that $\mathcal{C}$ is finite with atoms $C_1, \ldots, C_m$. For all n-tuples $a = (a_1, \ldots, a_n)$ in M, define $a_{ij} := a_i \cdot \chi_{C_j}$ for all $1 \leq i \leq n$ and $1 \leq j \leq m$. The mn-tuple $(a_{11}, \ldots, a_{1m}, \ldots, a_{n1}, \ldots, a_{nm})$ is denoted by $\big(\{a_{ij}\}_{1\leq i\leq n, 1\leq j\leq m}\big)$. Then the mapping*

$$\begin{aligned}\theta_{\mathcal{C}} \colon S_n^{\mathsf{ARV}}(C) &\to S_{mn}(\mathsf{ARV})\\ \operatorname{tp}(a/C) &\mapsto \operatorname{tp}\Big(\big(\{a_{ij}\}_{1\leq i\leq n, 1\leq j\leq m}\big)\Big)\end{aligned}$$

is isometric from $(S_n^{\mathsf{ARV}}(C), d^*)$ into $(S_{mn}(\mathsf{ARV}), d^*)$.

Proof. By Lemma 4.8, $\theta_{\mathcal{C}}$ is well-defined. We will show that $\theta_{\mathcal{C}}$ is isometric. Let $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ be n-tuples in M. We define $a_{ij} = a_i \cdot \chi_{C_j}$ and $b_{ij} = b_i \cdot \chi_{C_j}$ for all $1 \leq i \leq n$ and $1 \leq j \leq m$. Since $C_1, \ldots, C_m$ are the atoms of $\mathcal{C}$, for all $1 \leq i \leq n$ we have $a_i = a_{i1} + \cdots + a_{im}$, $b_i = b_{i1} + \cdots + b_{im}$, and $d^*(a_i, b_i) = \sum_{j=1}^m d^*(a_{ij}, b_{ij})$. Then

$$d^*(a,b) = \sum_{i=1}^n \sum_{j=1}^m d^*(a_{ij}, b_{ij}) = d^*\Big(\big(\{a_{ij}\}_{1\leq i\leq n, 1\leq j\leq m}\big), \big(\{b_{ij}\}_{1\leq i\leq n, 1\leq j\leq m}\big)\Big).$$

By Lemma 4.8, we get

$$d^*\big(\operatorname{tp}(a/C), \operatorname{tp}(b/C)\big) = d^*\Big(\operatorname{tp}\big(\{a_{ij}\}_{1\leq i\leq n, 1\leq j\leq m}\big), \operatorname{tp}\big(\{b_{ij}\}_{1\leq i\leq n, 1\leq j\leq m}\big)\Big),$$

whereby

$$d^*(p,q) = d^*\big(\theta_{\mathcal{C}}(p), \theta_{\mathcal{C}}(q)\big) \text{ for all } p, q \in S_n^{\mathsf{ARV}}(C). \qquad \square$$

Now, we use Theorem 2.5 (the Kantorovich-Rubinstein duality formula) to give a formula for the d^*-metric between n-types over parameters in the most general situation.

Theorem 4.10. *Let κ be an uncountable cardinal. Let $\mathcal{M} \models \mathsf{ARV}$ be a κ-universal domain of the form $\big(L^1(\mu, [0,1]), \neg, \frac{1}{2}, \cdot, \dot{-}, I\big)$, where $(\Omega, \mathcal{F}, \mu)$ is an atomless probability space. Suppose $C \subseteq M = L^1(\mu, [0,1])$ is a small subset; let $\mathcal{C}$ be the σ-algebra of measurable sets generated by the random variables in C. Let $\{\mathcal{C}_j \mid j \in J\}$ be the family of all finite subalgebras of $\mathcal{C}$. Then $\mathcal{C} = \cup_{j\in J} \mathcal{C}_j$. For all $p, q \in S_n(ARV_C)$ and for all $j \in J$, let p_j and q_j be the restriction of p and q in $S_n(ARV_{\mathcal{C}_j})$ respectively. Then let $\nu_j = \eta_{mn} \circ \theta_{\mathcal{C}_j}(p_j)$ and $\xi_j = \eta_{mn} \circ \theta_{\mathcal{C}_j}(q_j)$, where η_{mn} and $\theta_{\mathcal{C}}$ are from Theorem 4.2 and Proposition 4.9. Then*

$$d^*(p,q) = \sup_{j\in J} \sup_{\|\psi\|_{Lip}\leq 1} \Big\{\int_{[0,1]^{mn}} \psi d\nu_j - \int_{[0,1]^{mn}} \psi d\xi_j\Big\},$$

where $\|\psi\|_{Lip} \leq 1$ means

$$|\psi(x) - \psi(y)| \leq \sum_{i=1}^{mn} |x_i - y_i|, \forall x = (x_1, \ldots, x_{mn}), y = (y_1, \ldots, y_{mn}) \in [0,1]^{mn}.$$

Proof. Take $j \in J$. By Proposition 4.9, $d^*(p_j, q_j) = d^*\big(\theta_{\mathcal{C}_j}(p_j), \theta_{\mathcal{C}_j}(q_j)\big)$. By Theorem 4.2, we have $d^*\big(\theta_{\mathcal{C}_j}(p_j), \theta_{\mathcal{C}_j}(q_j)\big) = W_1\big(\eta_{mn} \circ \theta_{\mathcal{C}_j}(p_j), \eta_{mn} \circ \theta_{\mathcal{C}_j}(q_j)\big) = W_1(\nu_j, \xi_j)$. By Theorem 2.5, we have

$$W_1(\nu_j, \xi_j) = \sup_{\|\psi\|_{Lip} \le 1} \Big\{ \int_{[0,1]^{mn}} \psi d\nu_j - \int_{[0,1]^{mn}} \psi d\xi_j \Big\},$$

where $\|\psi\|_{Lip} \le 1$ means $|\psi(x) - \psi(y)| \le \sum_{i=1}^{mn} |x_i - y_i|$ for all $x = (x_1, \dots, x_{mn})$ and $y = (y_1, \dots, y_{mn}) \in [0,1]^{mn}$.

Then, by Remark 4.6(ii) we have

$$d^*(p, q) = \sup_{j \in J} d^*(p_j, q_j) = \sup_{j \in J} \sup_{\|\psi\|_{Lip} \le 1} \Big\{ \int_{[0,1]^{mn}} \psi d\nu_j - \int_{[0,1]^{mn}} \psi d\xi_j \Big\}. \qquad \square$$

5. A new proof of an integral formula for the Wasserstein distance

In this section, using rearrangements and model theoretic results, we give a new proof of Theorem 2.6. First, we need the following lemma, which is interesting in itself.

Lemma 5.1. *Let f, g be simple, nonincreasing, right-continuous functions from the standard Lebesgue space $\big([0,1], \mathscr{B}, \lambda\big)$ to $[0,1]$. Then*

$$\int_0^1 \big|f(x) - g(x)\big| dx = \int_0^1 \big|\lambda(f > t) - \lambda(g > t)\big| dt.$$

Proof. Let D be the set

$$\big\{(x, y) \mid x, y \in [0,1], f(x) \le y \le g(x) \text{ or } g(x) \le y \le f(x)\big\},$$

and let $I(D)$ be the set

$$I(D) = \big\{(x, y) \mid x, y \in [0,1], f(x) < y < g(x) \text{ or } g(x) < y < f(x)\big\}.$$

Since f, g are simple, nonincreasing, right-continuous functions, we know that $I(D) \subseteq D$, and there is a set S, which is a finite union of rectangles, such that $I(D)$ and D are equal to S up to null sets. Therefore, $(\lambda \times \lambda)(D) = (\lambda \times \lambda)\big(I(D)\big)$. Hence,

$$\int_0^1 \big|f(x) - g(x)\big| dx = (\lambda \times \lambda)(D) = (\lambda \times \lambda)\big(I(D)\big).$$

Note that $\lambda\ (f > s)$ and $\lambda\ (g > s)$ are also simple, nonincreasing, right-continuous functions. Let D' be the set

$$\big\{(s, t) \mid s, t \in [0,1], \lambda(f > s) \le t \le \lambda(g > s) \text{ or } \lambda(g > s) \le t \le \lambda(f > s)\big\},$$

and let $I(D')$ be the set

$$\big\{(s, t) \mid s, t \in [0,1], \lambda(f > s) < t < \lambda(g > s) \text{ or } \lambda(g > s) < t < \lambda(f > s)\big\}.$$

Since $\lambda(f > s)$ and $\lambda(g > s)$ are simple, nonincreasing, right-continuous functions, we know that $I(D') \subseteq D'$, and there is a set S', which is a finite union of rectangles, such that $I(D')$ and D' are equal to S' up to null sets. Therefore, $(\lambda \times \lambda)(D') = (\lambda \times \lambda)\big(I(D')\big)$. Hence

$$\int_0^1 \big|\lambda(f > t) - \lambda(g > t)\big| dt = (\lambda \times \lambda)(D') = (\lambda \times \lambda)\big(I(D')\big).$$

We will show that if $(x, y) \in I(D)$, then $(y, x) \in D'$, and if $(s, t) \in I(D')$, then $(t, s) \in D$.

If $(x, y) \in I(D)$, then $f(x) < y < g(x)$ or $g(x) < y < f(x)$. For all $u, v \in [0,1]$ and all simple, nonincreasing, right-continuous functions $h\colon [0,1] \to [0,1]$, elementary arguments yield the following facts:

$$\lambda(h > v) \leq u \text{ if and only if } h(u) \leq v; \tag{6}$$

$$\lambda(h > v) > u \text{ if and only if } h(u) > v. \tag{7}$$

Suppose $f(x) < y < g(x)$. Then, by (6), $\lambda(f > y) \leq x$; by (7), $\lambda(g > y) > x$. Hence $\lambda(f > y) \leq x < \lambda(g > y)$, whereby $(y, x) \in D'$. If $g(x) < y < f(x)$, then by (6), $\lambda(g > y) \leq x$; by (7), $\lambda(f > y) > x$. Hence, $\lambda(g > y) \leq x < \lambda(f > y)$, whereby $(y, x) \in D'$. Therefore, $I(D) \subseteq D'$.

If $(s, t) \in I(D')$, then $\lambda(f > s) < t < \lambda(g > s)$ or $\lambda(g > s) < t < \lambda(f > s)$. Suppose $\lambda(f > s) < t < \lambda(g > s)$. By (6), $f(t) \leq s$ and by (7), $g(t) > s$. Hence, $f(t) \leq s < g(t)$, whereby $(t, s) \in D$. If $\lambda(g > s) < t < \lambda(f > s)$, then by (6), $g(t) \leq s$ and by (7), $f(t) > s$. Hence, $g(t) \leq s < f(t)$, whereby $(t, s) \in D$. Therefore, $I(D') \subseteq D$.

Hence $(\lambda \times \lambda)\big(I(D)\big) \leq (\lambda \times \lambda)(D')$ and $(\lambda \times \lambda)\big(I(D')\big) \leq (\lambda \times \lambda)(D)$. Since $(\lambda \times \lambda)(D) = (\lambda \times \lambda)\big(I(D)\big)$ and $(\lambda \times \lambda)(D') = (\lambda \times \lambda)\big(I(D')\big)$, we have $(\lambda \times \lambda)(D) = (\lambda \times \lambda)(D')$, whereby

$$\int_0^1 \big|f(x) - g(x)\big| dx = \int_0^1 \big|\lambda(f > t) - \lambda(g > t)\big| dt. \qquad \square$$

Now we are ready to prove:

Theorem 5.2 (Theorem 2.6 revisited)**.** *For all $\mu, \nu \in P_1(\mathbb{R})$, one has*

$$W_1(\mu, \nu) = \int_{-\infty}^{+\infty} |F(x) - G(x)| dx, \tag{8}$$

where F and G are distribution functions for μ and ν respectively.

Proof. **Case 1**: We consider the case in which $\mu, \nu \in P_1([0,1]) \subseteq P_1(\mathbb{R})$. Let $([0,1], \mathscr{B}, \lambda)$ be the standard Lebesgue space. Let $f, g\colon [0,1] \to [0,1]$ be measurable functions such that $\operatorname{dist}(f) = \mu$ and $\operatorname{dist}(g) = \nu$. Then by Theorem 4.2, $W_1(\mu, \nu) = d(\operatorname{tp}(f), \operatorname{tp}(g))$. By Proposition 4.4, we may further assume that $d(f, g) = d(\operatorname{tp}(f), \operatorname{tp}(g)) = I\big(|f - g|\big) = \int_0^1 |f(x) - g(x)| dx$. Let $F(t) = \lambda(f \leq t)$ and $G(t) = \lambda(g \leq t)$.

First, suppose f and g are simple functions. We consider the decreasing rearrangement f^* of f and the decreasing rearrangement g^* of g from $[0,1]$

to $[0,1]$; for the definition, see Section 2.1. By Fact 2.3, $\mathrm{dist}(f^*) = \mathrm{dist}(f)$, $\mathrm{dist}(g^*) = \mathrm{dist}(g)$, and f^* and g^* are nonincreasing and lower semi-continuous. Because f and g are simple and $\mathrm{dist}(f) = \mathrm{dist}(f^*), \mathrm{dist}(g) = \mathrm{dist}(g^*)$, we have f^* and g^* are also simple. Thus, by Fact 2.3 we have that f^* and g^* are simple, nonincreasing and lower semi-continuous, whereby they are right-continuous. Since $J(x) = |x|$ is a nonnegative convex function with $J(0) = 0$, by Corollary 2.4 we have

$$\int_0^1 \big|f^*(x) - g^*(x)\big|dx \leq \int_0^1 \big|f(x) - g(x)\big|dx = d\big(\mathrm{tp}(f), \mathrm{tp}(g)\big).$$

Since $\mathrm{dist}(f^*) = \mathrm{dist}(f) = \mu$ and $\mathrm{dist}(g^*) = \mathrm{dist}(g) = \nu$, we have $W_1(\mu,\nu) \leq I\big(|f^* - g^*|\big)$, whereby $I\big(|f^* - g^*|\big) = d\big(\mathrm{tp}(f), \mathrm{tp}(g)\big) = W_1(\mu,\nu)$. Therefore, Equation (8) holds by Lemma 5.1.

Second, we consider general $[0,1]$-valued random variables f and g. For all $h_1, h_2 \in L^1\big([0,1],[0,1]\big)$,

$$W_1(\mathrm{dist}(h_1), \mathrm{dist}(h_2)) \leq d(h_1,h_2) = \int_0^1 |h_1(x) - h_2(x)|dx.$$

Since the set of simple functions is dense in $L^1\big(\lambda,[0,1]\big)$, one has that the set $S = \{\mathrm{dist}(h) \mid h\colon [0,1] \to [0,1] \text{ is a simple function}\}$ is dense in $(P_1([0,1]), W_1)$. In the product space $P_1([0,1]) \times P_1([0,1])$, we consider the maximum metric induced by W_1. Then the set of pairs (μ,ν) satisfying Equation (8) is a closed subset of $P_1([0,1]) \times P_1([0,1])$. We already know that all pairs in $S \times S$ satisfy Equation (8). Then by the fact that S is dense in $(P_1([0,1]), W_1)$, we know that for all μ,ν in $P_1([0,1])$, Equation (8) holds.

Case 2: We consider the case in which $\mu,\nu \in P_1([-n,n]) \subseteq P_1(\mathbb{R})$ for some $n \in \mathbb{N}$. We define a function $\varphi\colon \mathbb{R} \to \mathbb{R}$ by $\varphi(x) = \frac{x+n}{2n}$. This function φ induces $\Phi\colon P_1([-n,n]) \to P_1([0,1])$ by defining $\Phi(\mu)(A) = \mu(\varphi^{-1}(A))$, where μ is in $P_1([-n,n])$ and A is a Lebesgue measurable subset of $[0,1]$. Let f and g be any two random variables with $\mathrm{dist}(f) = \mu$ and $\mathrm{dist}(g) = \nu$. Then $\mathrm{dist}(\varphi \circ f) = \mathrm{dist}(\Phi(\mu))$ and $\mathrm{dist}(\varphi \circ g) = \mathrm{dist}(\Phi(\nu))$. Also $\|f - g\|_1 = 2n\|\varphi \circ f - \varphi \circ g\|_1$. Hence $2nW_1(\Phi(\mu), \Phi(\nu)) \leq W_1(\mu,\nu)$. Similarly, one has $2nW_1(\Phi(\mu), \Phi(\nu)) \geq W_1(\mu,\nu)$, whereby $2nW_1(\Phi(\mu), \Phi(\nu)) = W_1(\mu,\nu)$. Let $\Phi(F)$ and $\Phi(G)$ denote the distribution functions for $\Phi(\mu)$ and $\Phi(\nu)$ respectively. Then $\Phi(F)(x) = \Phi(\mu)((-\infty,x]) = \mu(\varphi^{-1}(-\infty,x]) = \mu((-\infty,2nx-n]) = F(2nx-n)$ for $x \in \mathbb{R}$. Similarly, $\Phi(G)(x) = G(2nx-n)$. Then by substitution, one has

$$\int_{\mathbb{R}} |F - G|dx = 2n\int_{\mathbb{R}} |\Phi(F) - \Phi(G)|dx.$$

By Case 1, one has

$$W_1(\Phi(\mu), \Phi(\nu)) = \int_{\mathbb{R}} |\Phi(F) - \Phi(G)|dx.$$

Therefore, Equation (8) holds for μ and ν.

Case 3: The most general case. Let T denote $\bigcup_{n\in\mathbb{N}} P_1([-n,n])$. Then T is dense in $(P_1(\mathbb{R}), W_1)$. In the product space $P_1(\mathbb{R}) \times P_1(\mathbb{R})$, we consider the maximum metric induced by W_1. Then the set of pairs (μ,ν) satisfying Equation (8) is a closed subset of $P_1(\mathbb{R}) \times P_1(\mathbb{R})$. By Case 2, we know that all pairs in $T \times T$ satisfy Equation (8). Then by the fact that T is dense in $(P_1(\mathbb{R}), W_1)$, we know that for all μ,ν in $P_1(\mathbb{R})$, Equation (8) holds. □

6. Conditional Wasserstein distances

By Theorem 3.2, the mapping

$$\begin{aligned}\zeta\colon S_n^{\mathsf{ARV}}(A) &\to \mathfrak{D}_{[0,1]^n}(\mathscr{A})\\ \operatorname{tp}(f/A) &\mapsto \operatorname{dist}(f \mid \mathscr{A})\end{aligned}$$

is a homeomorphism between the type space $S_n^{\mathsf{ARV}}(A)$ equipped with the logic topology and the space $\mathfrak{D}_{[0,1]^n}(\mathscr{A})$ of conditional distributions over $\mathscr{A}$ equipped with the topology of weak convergence. If we consider the d^*-metric on the type spaces over $\emptyset$ in ARV, then by Theorem 4.2, the mapping

$$\begin{aligned}\eta_n\colon S_n(\mathsf{ARV}) &\to \mathfrak{D}_{[0,1]^n}\\ \operatorname{tp}(f) &\mapsto \operatorname{dist}(f)\end{aligned}$$

is an isometric isomorphism between the type space $\big(S_n(\mathsf{ARV}), d^*\big)$ and the Wasserstein space $(\mathfrak{D}_{[0,1]^n}, W_1)$. In model theory, types over parameters are as important as types over $\emptyset$, so it is natural and potentially valuable to define a new distance on $\mathfrak{D}_{[0,1]^n}(\mathscr{A})$ which generalizes the Wasserstein distance between distributions, so that ζ becomes an isometric isomorphism. In this subsection, we will discuss this new metric, called conditional Wasserstein distance, and define conditional Wasserstein spaces.

Let $(\Omega,\mathcal{F},m)$ be a κ-saturated probability space, where κ is an uncountable cardinal. Let $\mathscr{A}$ be a σ-subalgebra of $\mathcal{F}$ that is small with respect to κ. For a Polish metric space $\mathcal{X}$, an *n-dimensional conditional distribution over $\mathscr{A}$ on $\mathcal{X}$*, say μ, is an $L^1\big((\Omega,\mathscr{A},m),[0,1]\big)$-valued Borel probability measure on $\mathcal{X}$. More precisely, it satisfies the following:

- $\mu(B) \geq 0$ a.s. for all Borel sets $B \subseteq \mathcal{X}$.
- $\mu(\mathcal{X}) = 1$ a.s.
- $\mu(\bigcup_{i=1}^\infty B_i) = \sum_{i=1}^\infty \mu(B_i)$ a.s. for all disjoint Borel sets $B_1, B_2, \ldots \subseteq \mathcal{X}$.

Every random variable f from Ω to $\mathcal{X}$ determines a conditional distribution over $\mathscr{A}$, denoted by $\operatorname{dist}(f \mid \mathscr{A})$. It is defined by $\operatorname{dist}(f \mid \mathscr{A})(B) := \mathbb{P}(f \in B \mid \mathscr{A})$ for every Borel set $B \subseteq \mathcal{X}$, where $\mathbb{P}(\cdot \mid \mathscr{A})$ is the conditional probability. Let $\mathfrak{D}_{\mathcal{X}}(\mathscr{A})$ denote the set of all conditional distributions over $\mathscr{A}$ on $\mathcal{X}$. For every Borel set $B \subseteq \mathcal{X}$, we denote by $\mathfrak{D}_B(\mathscr{A})$ the set of all conditional distributions over $\mathscr{A}$ which, as measures, are supported by B.

Definition 6.1 (Conditional Wasserstein distance)**.** Let $(\Omega,\mathcal{F},m)$ denote a κ-saturated probability space for an uncountable cardinal κ, and let $\mathscr{A}$ denote a small σ-subalgebra of $\mathcal{F}$. Let $(\mathcal{X}, d)$ be a Polish metric space and let $p \in [1, \infty)$. For conditional distributions μ and ν over $\mathscr{A}$ on $\mathcal{X}$, we define the *conditional Wasserstein distance of order p over* $\mathscr{A}$ as follows:

$$W_p(\mu, \nu) = \inf\Big\{ \big[\mathbb{E}d(f,g)^p\big]^{\frac{1}{p}} \mid f, g \colon \Omega \to \mathcal{X}, \operatorname{dist}(f \mid \mathscr{A}) = \mu, \operatorname{dist}(g \mid \mathscr{A}) = \nu \Big\}.$$

In the definition, we need Ω to be κ-saturated with $\kappa > |\mathscr{A}|$. Otherwise, for some $\mu \in \mathfrak{D}_{\mathcal{X}}(\mathscr{A})$, it is possible that no $f \colon \Omega \to \mathcal{X}$ satisfies $\operatorname{dist}(f \mid \mathscr{A}) = \mu$. For example, suppose that $(\Omega, \mathcal{F}, m)$ is the standard Lebesgue space $([0,1], \mathcal{L}, \lambda)$. By [13, Proposition 4.1], the separable structure $M = L^1(\lambda, [0,1])$ is not $\aleph_0$-saturated, and there exist $f \in M$ and a type $p(y)$ in $S_1(f)$ such that it is not realized in M. By Theorem 3.1, the type $p(y)$ corresponds $\mu \in \mathfrak{D}_{[0,1]^n}\big(\sigma(f)\big)$ such that no $g \in M$ satisfying $\operatorname{dist}(g \mid \sigma(f)) = \mu$.

Nonetheless, there is an alternative approach to the definition as follows:

Definition 6.2. Let $(\mathcal{X}, d)$ be a Polish metric space and let $\mathscr{A}$ be a probability algebra. Let $p \in [1, \infty)$. For $\mu, \nu \in \mathfrak{D}_{\mathcal{X}}(\mathscr{A})$, the *conditional Wasserstein distance of order p over* $\mathscr{A}$ is defined as

$$W_p(\mu, \nu) := \inf_{\Omega, f, g} \big\{ [\mathbb{E}d(f,g)^p]^{\frac{1}{p}} \big\},$$

where the infimum is over all probability spaces $(\Omega, \mathcal{F}, \mu)$ with $\mathscr{A} \subseteq \mathcal{F}$ and all random variables $f, g \colon \Omega \to \mathcal{X}$ such that $\operatorname{dist}(f \mid \mathscr{A}) = \mu$ and $\operatorname{dist}(g \mid \mathscr{A}) = \nu$.

Remark 6.3. These two definitions are equivalent to each other. Let W_p^{Ω} denote the distance defined in Definition 6.1 and let W_p denote the one defined in Definition 6.2. Clearly, for all $\mu, \nu \in \mathfrak{D}_{\mathcal{X}}(\mathscr{A})$, one has $W_p(\mu, \nu) \le W_p^{\Omega}(\mu, \nu)$. For all $\epsilon > 0$, by definition there are f, g such that $W_p(\mu, \nu) + \epsilon \ge [\mathbb{E}d(f,g)^p]^{\frac{1}{p}}$. Since Ω in Definition 6.1 is κ-saturated for some large cardinal κ and $\mathscr{A}$ is small, there are $f', g' \colon \Omega \to \mathcal{X}$ such that $\operatorname{dist}\big((f', g') \mid \mathscr{A}\big) = \operatorname{dist}\big((f, g) \mid \mathscr{A}\big)$. Therefore, $\operatorname{dist}(f' \mid \mathscr{A}) = \operatorname{dist}(f \mid \mathscr{A})$, $\operatorname{dist}(g' \mid \mathscr{A}) = \operatorname{dist}(g \mid \mathscr{A})$, and $[\mathbb{E}d(f,g)^p]^{\frac{1}{p}} = [\mathbb{E}d(f', g')^p]^{\frac{1}{p}}$. Consequently, $W_p(\mu, \nu) + \epsilon \ge [\mathbb{E}d(f', g')^p]^{\frac{1}{p}} \ge W_p^{\Omega}(\mu, \nu)$. Hence $W_p^{\Omega}(\mu, \nu) = W_p(\mu, \nu)$.

Definition 6.4 (Conditional Wasserstein space)**.** Let $(\Omega, \mathcal{F}, m)$, $\mathscr{A}$, $(\mathcal{X}, d)$, p be as above. For every $x \in \mathcal{X}$, let c_x denote the constant function from Ω to $\mathcal{X}$ whose value is x. The *conditional Wasserstein space of order p over* $\mathscr{A}$ is defined as

$$P_p(\mathcal{X}, \mathscr{A}) := \big\{ \mu \in \mathfrak{D}_{\mathcal{X}}(\mathscr{A}) \mid W_p\big(\mu, \operatorname{dist}(c_{x_0} \mid \mathscr{A})\big) < \infty \text{ for some } x_0 \in \mathcal{X} \big\}.$$

Suppose $\mathcal{X} = [0,1]^n$ with the distance defined by

$$d((a_1, \ldots, a_n), (b_1, \ldots, b_n)) = \sum_{i=1}^{n} d(a_i, b_i)$$

for all $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in [0,1]^n$. Then $P_p([0,1]^n, \mathscr{A}) = \mathfrak{D}_{[0,1]^n}(\mathscr{A})$. Now we connect conditional Wasserstein spaces to types spaces over set of parameters.

Theorem 6.5. *Let $\mathcal{M}$ be a model of* ARV *of the form $L^1(\mu,[0,1])$, where $(\Omega,\mathcal{F},\mu)$ is an atomless probability space. Suppose $\mathcal{M}$ is a κ-universal domain for an uncountable cardinal κ. Let A be a small subset of M and let $\mathscr{A}=\sigma(A)$ be the σ-algebra of measurable sets generated by the random variables in A. Let f be an n-tuple in M. Then $\operatorname{dist}(f\mid\mathscr{A})$ is a conditional distribution over $\mathscr{A}$ on $[0,1]^n$. Moreover, the mapping*

$$\begin{aligned}\eta_n^A\colon S_n(A)&\to P_1\big([0,1]^n,\mathscr{A}\big)\\ \operatorname{tp}(f/A)&\mapsto \operatorname{dist}(f\mid\mathscr{A})\end{aligned}$$

is an isometric isomorphism between $\big(S_n(A),d^*\big)$ and $\big(P_1([0,1]^n,\mathscr{A}),W_1\big)$, where the metric on which the definition of W_1 depends is defined by

$$d^*_{[0,1]^n}(a,b)=\sum_{i=1}^n|a_i-b_i| \quad \textit{for all } a,b\in[0,1]^n.$$

Proof. This follows from the definition of conditional Wasserstein distances and Theorem 3.2. □

If $\mathscr{B}\subseteq\mathscr{A}$, then we define $\pi_{\mathscr{B}}\colon \mathfrak{D}_{\mathcal{X}}(\mathscr{A})\to\mathfrak{D}_{\mathcal{X}}(\mathscr{B})$ as follows: for all Borel $B\subseteq\mathcal{X}$, define $\pi_{\mathscr{B}}(\mu)(B)=\mathbb{E}\big(\mu(B)\mid\mathscr{B}\big)$ for all $\mu\in\mathfrak{D}_{\mathcal{X}}(\mathscr{A})$.

Proposition 6.6. *Let $\mathscr{A}$ be a probability algebra and let $\mu,\nu\in\mathfrak{D}_{[0,1]^n}(\mathscr{A})$. Then*

$$W_1(\mu,\nu)=\sup_{\mathscr{B}\subseteq\mathscr{A}}\big\{W_1\big(\pi_{\mathscr{B}}(\mu),\pi_{\mathscr{B}}(\nu)\big)\mid \mathscr{B} \textit{ is a finite subalgebra of } \mathscr{A}\big\}.$$

Proof. Let $\mathcal{M}$ be a κ-universal domain for ARV with $\kappa>|\mathscr{A}|$. We assume that M is of the form $L^1\big((\Omega,\mathcal{F},m),[0,1]\big)$ where $\mathscr{A}\subseteq\mathcal{F}$. By Remark 4.6(ii), for all $f,g\in M^n$, we have

$$\begin{aligned}&d\big(\operatorname{tp}(f/A),\operatorname{tp}(g/A)\big)\\ &=\sup_{\mathscr{B}\subseteq\mathscr{A}}\big\{d\big(\operatorname{tp}(f/\mathscr{B}),\operatorname{tp}(g/\mathscr{B})\big)\mid \mathscr{B}\text{ is a finite subalgebra of }\mathscr{A}\big\}.\end{aligned}$$

By Theorem 6.5, we have

$$\begin{aligned}&W_1\big(\operatorname{dist}(f\mid\mathscr{A}),\operatorname{dist}(g\mid\mathscr{A})\big)\\ &=\sup_{\mathscr{B}\subseteq\mathscr{A}}\big\{W_1\big(\operatorname{dist}(f\mid\mathscr{B}),\operatorname{dist}(g\mid\mathscr{B})\mid \mathscr{B}\text{ is a finite subalgebra of }\mathscr{A}\big\}.\end{aligned}$$

Hence,

$$\begin{aligned}&W_1\big(\operatorname{dist}(f\mid\mathscr{A}),\operatorname{dist}(g\mid\mathscr{A})\big)\\ &=\sup_{\mathscr{B}\subseteq\mathscr{A}}\big\{W_1\big(\pi_{\mathscr{B}}(\operatorname{dist}(f\mid\mathscr{A})),\pi_{\mathscr{B}}(\operatorname{dist}(g\mid\mathscr{A}))\big)\mid \mathscr{B}\text{ is finite subalgebra}\big\}.\end{aligned}$$

Since $(\Omega,\mathcal{F},\mu)$ is κ-saturated and $\mathscr{A}$ is small, for all $\mu,\nu\in\mathfrak{D}_{[0,1]^n}(\mathscr{A})$, there are $f,g\in M$ such that $\operatorname{dist}(f\mid\mathscr{A})=\mu$ and $\operatorname{dist}(g\mid\mathscr{A})=\nu$. The rest follows. □

Theorem 6.7. *Let $(\Omega, \mathscr{A}, m)$ be a probability space, where $\mathscr{A}$ is countably σ-generated. Then the topology induced by the conditional Wasserstein distance W_1 on $\mathfrak{D}_{[0,1]^n}(\mathscr{A})$ is the same as the topology of weak convergence if and only if $(\Omega, \mathscr{A}, m)$ is a discrete probability space.*

Proof. Suppose $\mathscr{A} = \sigma(f)$ for some random variable $f\colon \Omega \to [0,1]$. Then by [13, Theorem 1.1] and [13, Proposition 3.2], $\mathsf{ARV(f)}$ is separably categorical if and only if f is discrete. Then by the Ryll-Nardzewski Theorem for continuous logic ([4, Fact 1.14]), we get that on $S_n\big(\mathsf{ARV(f)}\big)$ the metric topology and the logic topology coincide if and only if f is discrete. The rest follows from Theorem 3.2 and Theorem 6.5. □

Remark 6.8. Comparing the preceding theorem with Theorem 2.8, we see the difference between conditional Wasserstein spaces over non-atomic $\mathscr{A}$ and Wasserstein spaces.

Theorem 6.9. *Let $(\Omega, \mathscr{A}, m)$ be a probability space. Suppose that $\mathscr{A}$ is σ-generated by a subset A of $L^1\big((\Omega, \mathscr{A}, m), [0,1]\big)$. Then the conditional Wasserstein space $\big(\mathfrak{D}_{[0,1]^n}(\mathscr{A}), W_1\big)$ has density character $\leq |A| + \aleph_0$.*

Proof. By Theorem 6.5, there is an isometric isomorphism between $\big(S_n(A), d^*\big)$ and $\big(\mathfrak{D}_{[0,1]^n}(\mathscr{A}), W_1\big)$. Since ARV is ω-stable, we have that $(S_n(A), d^*)$ has density character $\leq |A| + \aleph_0$. Hence, $\big(\mathfrak{D}_{[0,1]^n}(\mathscr{A}), W_1\big)$ has density character $\leq |A| + \aleph_0$. □

Acknowledgement. This work is based on the author's PhD Thesis [12]. The author is grateful to his thesis advisor C. Ward Henson for his kind guidance and support. Also, the author would like to thank Chen-Yu Huang.

References

[1] I. Ben Yaacov, *On theories of random variables*, Israel J. Math. **194** (2013), no. 2, 957–1012.

[2] I. Ben Yaacov, A. Berenstein, and C. W. Henson, *Almost indiscernible sequences and convergence of canonical bases*, J. Symb. Log. **79** (2014), no. 2, 460–484.

[3] I. Ben Yaacov, A. Berenstein, C. W. Henson, and A. Usvyatsov, *Model theory for metric structures*, in: Model theory with applications to algebra and analysis. Vol. 2, 315–427, London Math. Soc. Lecture Note Ser., **350**, Cambridge Univ. Press, Cambridge, 2008.

[4] I. Ben Yaacov and A. Usvyatsov, *On d-finiteness in continuous structures*, Fund. Math. **194** (2007), no. 1, 67–88.

[5] ———, *Continuous first order logic and local stability*, Trans. Amer. Math. Soc. **362** (2010), no. 10, 5213–5259.

[6] A. Berenstein and C. W. Henson, *Model theory of probability spaces*, submitted.

[7] I. Berkes and H. P. Rosenthal, *Almost exchangeable sequences of random variables*, Z. Wahrsch. Verw. Gebiete **70** (1985), no. 4, 473–507.

[8] G. Dall'Aglio, *Sugli estremi dei momenti delle funzioni di ripartizione doppia*, Ann. Scuoloa Norm. Sup. Pisa (3) **10** (1956), 35–74.

[9] R. Dudley, *Probabilities and Metrics*, Convergence of laws on metric spaces, with a view to statistical testing, Lecture Notes Series, No. 45, Matematisk Institut, Aarhus Universitet, Aarhus, 1976.

[10] D. H. Fremlin, *Measure algebras*, in: Handbook of Boolean Algebras, Volume 3, North-Holland, 1989.

[11] E. H. Lieb and M. Loss, *Analysis*, Second Edition, Graduate Studies in Mathematics, Volume 14, American Mathematical Society, 2001.

[12] S. Song, *Model theory and probability*, Thesis, University of Illinois at Urbana-Champaign, 2011.

[13] ———, *On d-finite tuples in random variable structures*, Fund. Math. **221** (2013), no. 3, 221–230.

[14] A. Szulga, *On the Wasserstein metric*, Transactions of the Eighth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Prague, 1978), Vol. B, 267–273, Reidel, 1978.

[15] S. Vallander, *Calculations of the Vasseršteĭn distance between probability distributions on the line*, Teor. Verojatnost. i Primenen. **18** (1973), 824–827.

[16] C. Villani, *Optimal Transport*, Old and New, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 338. Springer-Verlag, 2009.

Shichang Song
Department of Mathematics
Beijing Jiaotong University
Beijing, P. R. China
Email address: ssong@bjtu.edu.cn