University of Reading

Isometric study of Wasserstein spaces - the real line

Article

Accepted Version

Gehér, G. P., Titkos, T. and Virosztek, D. (2020) Isometric study of Wasserstein spaces - the real line. Transactions of the American Mathematical Society, 373 (8). pp. 5855-5883. ISSN 1088-6850 doi: 10.1090/tran/8113 Available at https://centaur.reading.ac.uk/85689/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1090/tran/8113

Publisher: American Mathematical Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

University of Reading

Reading's research outputs online

ISOMETRIC STUDY OF WASSERSTEIN SPACES – THE REAL LINE

GYÖRGY PÁL GEHÉR, TAMÁS TITKOS, AND DÁNIEL VIROSZTEK

Abstract. Recently Kloeckner described the structure of the isometry group of the quadratic Wasserstein space $\mathcal{W}_2(\mathbb{R}^n)$. It turned out that the case of the real line is exceptional in the sense that there exists an exotic isometry flow. Following this line of investigation, we compute $\mathrm{Isom}\,(\mathcal{W}_p(\mathbb{R}))$, the isometry group of the Wasserstein space $\mathcal{W}_p(\mathbb{R})$ for all $p \in [1,\infty) \setminus \{2\}$. We show that $\mathcal{W}_2(\mathbb{R})$ is also exceptional regarding the parameter p: $\mathcal{W}_p(\mathbb{R})$ is isometrically rigid if and only if $p \neq 2$. Regarding the underlying space, we prove that the exceptionality of $p = 2$ disappears if we replace $\mathbb{R}$ by the compact interval $[0,1]$. Surprisingly, in that case, $\mathcal{W}_p([0,1])$ is isometrically rigid if and only if $p \neq 1$. Moreover, $\mathcal{W}_1([0,1])$ admits isometries that split mass, and $\mathrm{Isom}\,(\mathcal{W}_1([0,1]))$ cannot be embedded into $\mathrm{Isom}\,(\mathcal{W}_1(\mathbb{R}))$.

Contents

1. Introduction 1
1.1. Motivation and State of the Art 1
1.2. Main results, content of the paper 2
1.3. Technical preliminaries 3
2. Isometric study of $\mathcal{W}_p([0,1])$ 4
2.1. $p = 1$ – Isometries splitting mass 5
2.2. $p > 1$ – Isometric rigidity 8
3. Isometric study of $\mathcal{W}_p(\mathbb{R})$ 11
3.1. $p = 1$ – Isometric rigidity 11
3.2. $p > 1$, $p \neq 2$ – Characterization of isometric embeddings 16
3.3. $p = 2$ – A functional analytic description of the exotic flow 20
References 22

1. Introduction

1.1. **Motivation and State of the Art.** Given a complete and separable metric space X, one defines its Wasserstein space as the collection of sufficiently concentrated Borel probability measures endowed with a metric which is calculated by means of optimal transport (see the precise definitions in Subsection 1.3). This notion has strong connections to many flourishing areas in pure and applied mathematics including probability theory [4, 5], theory of (stochastic) partial differential equations [10, 11], geometry of metric spaces [16, 19, 21], machine learning [1, 20], and many more. Besides of these connections, the p-Wasserstein space itself is an interesting object, being a measure theoretic analogue of L^p spaces [13].

1991 *Mathematics Subject Classification.* Primary: 54E40; 46E27 Secondary: 60A10; 60B05.

Key words and phrases. Wasserstein space, isometric embeddings, isometric rigidity, exotic isometry flow.

Gehér was supported by the Leverhulme Trust Early Career Fellowship (ECF-2018-125), and also by the Hungarian National Research, Development and Innovation Office (Grant no. K115383); Titkos was supported by the Hungarian National Research, Development and Innovation Office - NKFIH (grant no. PD128374 and grant no. K115383), by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, and by the ÚNKP-18-4-BGE-3 New National Excellence Program of the Ministry of Human Capacities; Virosztek was supported by the ISTFELLOW program of the Institute of Science and Technology Austria (project code IC1027FELL01) and partially supported by the Hungarian National Research, Development and Innovation Office – NKFIH (grant no. K124152 and grant no. KH129601).

When working in a metric setting, a natural question arises: *can we compute the group of isometries?* The answer is known for various concrete metric spaces. However, the problem about the isometric embedding semigroup is usually incomparably more difficult, hence an answer is known only for a few cases. Classical examples from functional analysis include the Banach–Lamperti theorem [15] which describes the semigroup of all linear isometric embeddings of L^p spaces, or the Banach–Stone theorem which describes the group of all linear isometries of the Banach space of all continuous functions over a compact Hausdorff space. We now recall some more recent examples, concentrating on those where the metric space consists of Borel probability measures. The common theoretical importance of all the forthcoming metrics is that they metrise the weak convergence of measures. Molnár proved in [18] that the space of all Borel probability measures over $\mathbb{R}$ endowed with the Lévy metric is isometrically rigid, that is, each isometry is a push-forward map induced by an isometry of the underlying space $\mathbb{R}$. This result has been generalised for separable real Banach spaces in [8]. For a more detailed overview of similar results we refer the reader to the survey [25], where the case of the Kolmogorov-Smirnov [6] and the Kuiper distances [7] are also discussed.

Bertrand and Kloeckner wrote a series of papers [2, 3, 12, 13, 14] about the isometry groups of quadratic Wasserstein spaces over various metric spaces. Here we only recall one of Kloeckner's results [12, Theorem 1.1] in which he described the isometry group of the quadratic Wasserstein space over $\mathbb{R}$, and showed the surprising fact that this space admits so-called exotic isometries whose action is wild in a sense.

We would like to point out that in all of the above results about measures the assumption of bijectivity of the distance preserving maps was crucial in order to obtain these descriptions. In the present paper we continue our study on the (not necessarily bijective) isometric embeddings of Wasserstein spaces, which we started in the recent paper [9], where we provided a complete description for the case of the discrete metric space.

1.2. **Main results, content of the paper.** Kloeckner proved in [12] that the isometric structure of $\mathcal{W}_2(\mathbb{R})$ is exceptional among $\mathcal{W}_2(\mathbb{R}^n)$ spaces. Namely, there exists an exotic isometry flow of $\mathcal{W}_2(\mathbb{R})$. Our aim here is to show that $\mathcal{W}_2(\mathbb{R})$ is exceptional regarding the parameter p as well. It turns out that exceptionality of $p = 2$ disappears if we replace $\mathbb{R}$ by $[0, 1]$. Surprisingly, in that case $p = 1$ is exceptional. The main result of this paper is to get the full picture in the case of the real line. That is, we compute $\mathrm{Isom}\,(\mathcal{W}_p(\mathbb{R}))$, the isometry group of the Wasserstein space $\mathcal{W}_p(\mathbb{R})$ for all $p \in [1, \infty) \setminus \{2\}$, see the table below, where C_2 denotes the two-element group.

	$\mathrm{Isom}(\mathcal{W}_p([0,1]))$ (Isometric rigidity)	$\mathrm{Isom}(\mathcal{W}_p(\mathbb{R}))$ (Isometric rigidity)
$p = 1$	$C_2 \times C_2$ (not rigid)	$\mathrm{Isom}(\mathbb{R})$ (rigid)
$p > 1$, $p \neq 2$	C_2 (rigid)	$\mathrm{Isom}(\mathbb{R})$ (rigid)
$p = 2$	C_2 (rigid)	$\mathrm{Isom}(\mathbb{R}) \ltimes \mathrm{Isom}(\mathbb{R})$ (not rigid)

In fact, besides describing the isometry group, we are able to answer more challenging questions regarding the isometric structure. Below we summarize our results, and briefly sketch the method.

Section 2 is devoted to handle the case of the interval. In Subsection 2.1 we characterize all isometric embeddings of $\mathcal{W}_1([0, 1])$. Using the Harris inequality, we find an extremal metric property which is satisfied exactly for those measures that are either Dirac masses, or convex combinations of two Dirac masses concentrated on $\{0, 1\}$. Thus we are able to obtain that isometric embeddings are automatically surjective, and that the isometry group is the Klein group. Moreover, $\mathcal{W}_1([0, 1])$ admits isometries that split mass, which is quite unusual for Wasserstein spaces (see the aforementioned papers of Bertrand and Kloeckner). The case $p > 1$ is investigated in Subsection 2.2. Using induction and finding some extremal metric properties, we show that the set of all measures supported on 2^N points with equi-distributed weights are left invariant ($N \in \mathbb{N}$). As a consequence, we again have that every isometric embedding is necessarily surjective, and the

isometry group consists of the push-forward maps induced by the two isometries of the interval.

The case of the real line is investigated in Section 3. The main result of Subsection 3.1 is that every isometry of $\mathcal{W}_1(\mathbb{R})$ is implemented by an isometry of $\mathbb{R}$. The main issue here is to find a metric characterization of Dirac masses which we do by examining when the diameter of the metric midpoint set of two measures is minimal. This process naturally leads to the notions of vertical and horizontal bisecting measures, and that of adjacent measures. As a consequence of our description, we conclude that there are two isometries of $\mathcal{W}_1([0,1])$ that cannot be extended to an isometry of $\mathcal{W}_1(\mathbb{R})$, even though $\mathcal{W}_1([0,1])$ embeds naturally into $\mathcal{W}_1(\mathbb{R})$. Furthermore, the isometry group of $\mathcal{W}_1(\mathbb{R})$ does not contain an isomorphic copy of the isometry group of $\mathcal{W}_1([0,1])$.

In Subsection 3.2 we describe the general form of (not necessarily surjective) isometric embeddings of $\mathcal{W}_p(\mathbb{R})$ for $p > 1$ and $p \neq 2$. The key ingredients here are the Banach–Lamperti theorem and an abstract Mankiewicz-type extension lemma. We show that every isometric embedding is a composition of a push-forward of an isometry of $\mathbb{R}$ and a map which acts as a translation on quantile functions. In Subsection 3.3 we take a closer look at Kloeckner's result [12, Theorem 1.1] on $\mathrm{Isom}\,(\mathcal{W}_2(\mathbb{R}))$, in particular, at its exotic isometry flow. In [12], these exotic isometries were defined explicitly on measures supported on at most two points, and it was proved that there exists a unique extension to the whole space. However, the action of the exotic isometry flow was not given explicitly on general measures. Our contribution here is to provide a functional analytic description of this action on the whole $\mathcal{W}_2(\mathbb{R})$ in terms of quantile functions, which involves the well-known Volterra- and a composition operator.

1.3. **Technical preliminaries.** The aim of this subsection is to fix the terminology.

Definition 1.1 (Isometric embedding, isometry)**.** Let $\big(X, \varrho\big)$ be a metric space. A self-map $f\colon X \to X$ is called an *isometric embedding* if it preserves the distance, that is,

$$\varrho(f(x), f(y)) = \varrho(x, y) \qquad (x, y \in X).$$

Surjective isometric embeddings are termed *isometries.*

Note that isometric embeddings acting on X form a unital semigroup which we denote by $\mathrm{IsEmb}(X)$. The symbol $\mathrm{Isom}(X)$ stands for the group of all isometries. We denote by $\mathcal{P}(X)$ the set of all Borel probability measures on X.

Definition 1.2 (Coupling measure)**.** Let X be a complete and separable metric space and let $\mu, \nu \in \mathcal{P}(X)$. A Borel probability measure π on $X \times X$ is said to be a *coupling* of μ and ν if the marginals of π are μ and ν, that is, $\pi\,(A \times X) = \mu(A)$ and $\pi\,(X \times B) = \nu(B)$ for all Borel sets $A, B \subseteq X$. The set of all couplings is denoted by $\Pi(\mu, \nu)$.

By means of couplings, in other words *transport plans*, we can define the p-Wasserstein space and the corresponding p-Wasserstein distance. For more details we refer the reader to the fundamental works of Villani [23, 24].

Definition 1.3 (p-Wasserstein space)**.** Let $\big(X, \varrho\big)$ be a complete and separable metric space, and $p \geq 1$ be a parameter. The *p-Wasserstein space* $\mathcal{W}_p(X)$ is the set of all $\mu \in \mathcal{P}(X)$ that satisfy

$$\int_X \varrho(x, \hat{x})^p \,\mathrm{d}\mu(x) < \infty$$

for some (hence all) $\hat{x} \in X$, endowed with the *p-Wasserstein distance*

$$d_{\mathcal{W}_p}(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} \varrho(x, y)^p \,\mathrm{d}\pi(x, y) \right)^{\frac{1}{p}}.$$

In words, the p-Wasserstein space $\mathcal{W}_p(X)$ is the set of all probability distributions that have finite moment of order p, endowed with the p-Wasserstein metric $d_{\mathcal{W}_p}$. We remark that $d_{\mathcal{W}_p}$ metrizes the weak convergence and is sensitive to large distances in X. Clearly, the embedding of X as Dirac masses

$$\iota\colon X \to \mathcal{W}_p(X), \qquad \iota(x) := \delta_x$$

is distance preserving. Moreover, isometries of the underlying space appear in $\mathrm{Isom}(\mathcal{W}_p(X))$ by means of a natural group homomorphism given in (1.1) below. Throughout this paper the set of all Dirac masses is denoted by $\Delta(X)$.

Definition 1.4 (Push-forward)**.** For a measurable map $g\colon X \to X$ the induced *push-forward map* $g_\#\colon \mathcal{P}(X) \to \mathcal{P}(X)$ is defined by

$$\big(g_\#(\mu)\big)(A) = \mu(g^{-1}[A]) \qquad (A \subseteq X \text{ Borel set},\ \mu \in \mathcal{P}(X))$$

where $g^{-1}[A] = \{x \in X \,|\, g(x) \in A\}$. We call $g_\#(\mu)$ the *push-forward* of μ with g. If $\psi \in \mathrm{Isom}(\mathbb{R})$, then the push-forward map $\psi_\#$ is an isometry of $\mathcal{W}_p(X)$, and the embedding

$$\#\colon \mathrm{Isom}(X) \to \mathrm{Isom}\left(\mathcal{W}_p(X)\right), \qquad \psi \mapsto \psi_\# \tag{1.1}$$

is a group homomorphism. Isometries of the form $\psi_\#$ are called *trivial isometries*.

A special feature of Wasserstein spaces on the real line is that the Wasserstein distance of measures can be calculated by means of their cumulative distribution and quantile functions.

Definition 1.5 (Cumulative distribution and quantile functions)**.** In case of $\big(X, \varrho\big) = \big(\mathbb{R}, |\cdot|\big)$, the *cumulative distribution function* of a measure $\mu \in \mathcal{P}(\mathbb{R})$ is defined as

$$F_\mu\colon \mathbb{R} \to [0,1], \quad x \mapsto F_\mu(x) := \mu\left((-\infty, x]\right).$$

We use the shorthand notation $F_\mu(x-) := \lim_{t \nearrow x} F_\mu(t)$ for the limit from the left. The *quantile function* of μ (or the right-continuous generalized inverse of F_μ) is denoted by F_μ^{-1} and is defined as

$$F_\mu^{-1}\colon (0,1) \to \mathbb{R}, \quad y \mapsto F_\mu^{-1}(y) := \sup\left\{x \in \mathbb{R} \,|\, F_\mu(x) \le y\right\}.$$

In case of $\big(X, \varrho\big) = \big([0,1], |\cdot|\big)$ we shall handle the cumulative distribution and the quantile functions of a $\mu \in \mathcal{P}([0,1])$ as $[0,1] \to [0,1]$ functions. The quantile function in this case is defined by right-continuity at 0 and it takes the value 1 at 1.

Note that the cumulative distribution function of a $\mu \in \mathcal{P}([0,1])$ is monotone increasing, continuous from the right and take the value 1 at the point 1. Conversely, any function $F\colon [0,1] \to [0,1]$ satisfying the above three conditions is the cumulative distribution function of some Borel probability measure on $[0,1]$. Consequently, for any measure $\mu \in \mathcal{P}([0,1])$, the function F_μ^{-1} is a cumulative distribution function of some measure $\nu \in \mathcal{P}([0,1])$, that is, $F_\nu = F_\mu^{-1}$.

As was mentioned above, in our setting the p-Wasserstein distance can be expressed in terms of the cumulative distribution and quantile functions. Namely, Vallender proved in [22] that

$$d_{\mathcal{W}_1}(\mu,\nu) = \int_{-\infty}^{\infty} |F_\mu(x) - F_\nu(x)| \ \mathrm{d}x = \int_0^1 \left|F_\mu^{-1}(x) - F_\nu^{-1}(x)\right| \ \mathrm{d}x \tag{1.2}$$

for all $\mu, \nu \in \mathcal{W}_1(\mathbb{R})$. Moreover, Vallender's formula can be generalized in the following way:

$$d_{\mathcal{W}_p}(\mu,\nu) = \left(\int_0^1 \left|F_\mu^{-1} - F_\nu^{-1}\right|^p \mathrm{d}t\right)^{\frac{1}{p}} \qquad (p > 1,\ \mu,\nu \in \mathcal{W}_p(\mathbb{R})), \tag{1.3}$$

see for instance [24, Remarks 2.19]. These two formulae will play an important role in the sequel.

2. Isometric study of $\mathcal{W}_p([0,1])$

Knowing Kloeckner's result on the exotic isometry flow in $\mathcal{W}_2(\mathbb{R})$, a natural question arises: how does the isometry group look like when one replaces $\mathbb{R}$ by the compact interval $[0,1]$. Investigating this problem we found on the one hand that exceptionality of $p = 2$ disappears, and instead the case $p = 1$ becomes exceptional. On the other hand, it turned out that some observations regarding $\mathcal{W}_1([0,1])$ can be used when dealing with $\mathcal{W}_1(\mathbb{R})$, thus we decided to start with the case of $\mathcal{W}_p([0,1])$ spaces. The aim of this section is to describe the structure of all (not necessarily surjective) distance preserving self-maps of $\mathcal{W}_p([0,1])$. In fact, we will prove that every isometric embedding is automatically surjective, thus $\mathrm{Isom}\left(\mathcal{W}_p([0,1])\right) = \mathrm{IsEmb}\left(\mathcal{W}_p([0,1])\right)$.

2.1. $p = 1$ − **Isometries splitting mass.** We start by naming two maps that arise naturally.

Definition 2.1 (Reflection in $\mathcal{W}_1([0,1])$)**.** The isometry group of $[0,1]$ is isomorphic to C_2, and is generated by $r\colon [0,1] \to [0,1]$, $r(x) := 1 - x$. According to (1.1), the push-forward map $r_\#$ is an isometry of $\mathcal{W}_1([0,1])$ which we call *reflection*.

Definition 2.2 (Flip operation)**.** The map

$$j\colon \mathcal{W}_1([0,1]) \to \mathcal{W}_1([0,1]), \quad \mu \mapsto j(\mu), \quad F_{j(\mu)} = F_\mu^{-1}$$

is called the *flip operation*. The map j is surjective, and thus we see from (1.2) that $j \in \mathrm{Isom}(\mathcal{W}_1([0,1]))$.

We remark that the flip operation does not send Dirac masses to Dirac masses in general, which is quite unusual among isometries of Wasserstein spaces. Indeed,

$$j(\delta_t) = t\delta_0 + (1-t)\delta_1 \qquad (0 \le t \le 1).$$

The essential part of our argument will be to show that the flip operation and reflection generate the semigroup $\mathrm{IsEmb}(\mathcal{W}_1([0,1]))$. As these two maps are bijective, it will follow that $\mathrm{IsEmb}(\mathcal{W}_1([0,1])) = \mathrm{Isom}(\mathcal{W}_1([0,1])) = C_2 \times C_2$, the Klein group.

When describing the form of isometric embeddings, it is a natural idea to identify those subsets of the space in question that can be characterized by means of certain extremal metric properties. Our first observation is that

$$\mathrm{diam}(\mathcal{W}_1([0,1])) = \sup\left\{ \int_0^1 |F_\mu(t) - F_\nu(t)|\, dt \,\middle|\, \mu,\nu \in \mathcal{W}_1([0,1]) \right\} = 1,$$

and this supremum is attained if and only if $\{\mu,\nu\} = \{\delta_0,\delta_1\}$. Therefore,

$$\{\delta_0,\delta_1\} = \{\varphi(\delta_0),\varphi(\delta_1)\} \qquad (\varphi \in \mathrm{IsEmb}(\mathcal{W}_1([0,1]))).$$

Note also that the triangle inequality $d_{\mathcal{W}_1}(\delta_0,\delta_1) \le d_{\mathcal{W}_1}(\delta_0,\mu) + d_{\mathcal{W}_1}(\mu,\delta_1)$ is saturated for every measure $\mu \in \mathcal{W}_1([0,1])$. Indeed,

$$d_{\mathcal{W}_1}(\delta_0,\mu) + d_{\mathcal{W}_1}(\mu,\delta_1) = \int_{[0,1]} |0-y|\ \mathrm{d}\mu(y) + \int_{[0,1]} |x-1|\ \mathrm{d}\mu(x) = 1$$

holds for all $\mu \in \mathcal{W}_1([0,1])$. Let us define the set

$$S_t := \left\{ \mu \in \mathcal{W}_1([0,1]) \,\middle|\, d_{\mathcal{W}_1}(\delta_0,\mu) = t \right\} \qquad (t \in [0,1]),$$

which we call the *t-slice*. Clearly, if $\varphi \in \mathrm{IsEmb}(\mathcal{W}_1([0,1]))$ with $\varphi(\delta_0) = \delta_0$, then $\varphi(S_t) \subseteq S_t$.

In the next Claim we characterize those elements in S_t that have maximal distance. The Harris inequality plays an important role in our argument. We introduce the following notations for functions F and G:

$$(F \wedge G)(x) := \min\{F(x),G(x)\} \quad \text{and} \quad (F \vee G)(x) := \max\{F(x),G(x)\}.$$

Claim 2.3. Let $0 \le t \le 1$. The t-slice S_t has diameter $2t(1-t)$. That is,

$$d_{\mathcal{W}_1}(\rho,\sigma) \le 2t(1-t) \qquad (\rho,\sigma \in S_t)$$

and

$$d_{\mathcal{W}_1}(\rho,\sigma) = 2t(1-t) \quad \Longleftrightarrow \quad \{\rho,\sigma\} = \{(1-t)\delta_0 + t\delta_1, \delta_t\}.$$

Proof. The statement is trivial if $t = 0$ or $t = 1$. Let $t \in (0,1)$ and $\rho,\sigma \in S_t$ be arbitrary but fixed. By (1.2) we have

$$\int_0^1 F_\rho(x)\ \mathrm{d}x = \int_0^1 F_\sigma(x)\ \mathrm{d}x = 1 - t. \tag{2.1}$$

First, we show that $\int_0^1 (F_\rho \wedge F_\sigma)(x)\ \mathrm{d}x \ge (1-t)^2$ holds. As both F_ρ and F_σ are monotone increasing, we have

$$(F_\rho(x) - F_\rho(y))(F_\sigma(x) - F_\sigma(y)) \ge 0 \qquad (x,y \in [0,1]).$$

Consequently,

$$\int_0^1 \int_0^1 \left(F_\rho(x) - F_\rho(y)\right)\left(F_\sigma(x) - F_\sigma(y)\right)\ \mathrm{d}y\ \mathrm{d}x \geq 0, \tag{2.2}$$

which is equivalent to

$$\int_0^1 F_\rho(x)F_\sigma(x)\ \mathrm{d}x \geq \int_0^1 F_\rho(x)\ \mathrm{d}x \cdot \int_0^1 F_\sigma(x)\ \mathrm{d}x. \tag{2.3}$$

As $0 \leq F_\rho(x) \leq 1$ and $0 \leq F_\sigma(x) \leq 1$, we have $\left(F_\rho \wedge F_\sigma\right)(x) \geq F_\rho(x)F_\sigma(x)$ for all $x \in [0,1]$. Therefore, combining this with (2.1) and (2.3) we obtain

$$(1-t)^2 \leq \int_0^1 F_\rho(x)F_\sigma(x)\ \mathrm{d}x \leq \int_0^1 \left(F_\rho \wedge F_\sigma\right)(x)\ \mathrm{d}x. \tag{2.4}$$

Second, we prove that inequalities in (2.4) are equalities if and only if

$$\{\rho, \sigma\} = \{(1-t)\delta_0 + t\delta_1, \delta_t\}.$$

Since both F_ρ and F_σ are continuous from the right, so are the functions $F_\rho \cdot F_\sigma$ and $F_\rho \wedge F_\sigma$. Therefore, by the equivalence of (2.2) and (2.3), the first inequality in (2.4) is saturated if and only if we have

$$F_\rho(x) = F_\rho(y) \quad \text{or} \quad F_\sigma(x) = F_\sigma(y) \quad \left((x,y) \in [0,1) \times [0,1)\right). \tag{2.5}$$

Moreover, the second inequality in (2.4) is saturated if and only if

$$F_\rho(x)F_\sigma(x) = \left(F_\rho \wedge F_\sigma\right)(x) \quad \left(x \in [0,1)\right),$$

which means that we have

$$F_\rho(x) \in \{0,1\} \quad \text{or} \quad F_\sigma(x) \in \{0,1\} \quad \left(x \in [0,1)\right). \tag{2.6}$$

Notice that if any of the distribution functions F_σ and F_ρ is constant on $[0,1)$, then by (2.1) its value must be $1-t$ on $[0,1)$. Observe also that by (2.6) at most one of them is constant on $[0,1)$, say F_σ is not. This means that we have $F_\sigma(x) < F_\sigma(y)$ for some $0 \leq x < y < 1$. However by (2.5), this implies $F_\rho(\tilde{x}) = F_\rho(\tilde{y})$ for all $0 \leq \tilde{x} \leq x$ and $y \leq \tilde{y} < 1$, and thus F_ρ must be constant $1-t$ on $[0,1)$, or equivalently, $\rho = (1-t)\delta_0 + t\delta_1$. It follows from (2.6) that $F_\sigma(x) \in \{0,1\}$ for all $x \in [0,1)$ which means that σ must be a Dirac measure. By (2.1) we conclude $\sigma = \delta_t$.

Now, using (1.2), we get on the one hand that

$$d_{\mathcal{W}_1}(\rho,\sigma) = \int_0^1 |F_\rho(x) - F_\sigma(x)|\ \mathrm{d}x = \int_0^1 \left(F_\rho \vee F_\sigma\right)(x) - \left(F_\rho \wedge F_\sigma\right)(x)\ \mathrm{d}x. \tag{2.7}$$

On the other hand, we have

$$\int_0^1 \left(F_\rho \vee F_\sigma\right)(x) + \left(F_\rho \wedge F_\sigma\right)(x)\ \mathrm{d}x = \int_0^1 F_\rho(x) + F_\sigma(x)\ \mathrm{d}x = 2(1-t). \tag{2.8}$$

Finally, combining (2.7) and (2.8) with inequality (2.4), we conclude that

$$d_{\mathcal{W}_1}(\rho,\sigma) = 2(1-t) - 2\int_0^1 \left(F_\rho \wedge F_\sigma\right)(x)\ \mathrm{d}x \leq 2(1-t) - 2(1-t)^2 = 2t(1-t)$$

with equality if and only if $\{\rho,\sigma\} = \{(1-t)\delta_0 + t\delta_1, \delta_t\}$. The proof is complete. □

We remark here that Claim 2.3 roughly speaking describes the shape of $\mathcal{W}_1([0,1])$, suggesting that the action of a $\varphi \in \mathrm{IsEmb}(\mathcal{W}_1([0,1]))$ on $\left\{\delta_0, \delta_{\frac{1}{2}}\right\}$ determines φ completely. This is indeed the case and we make this precise as follows.

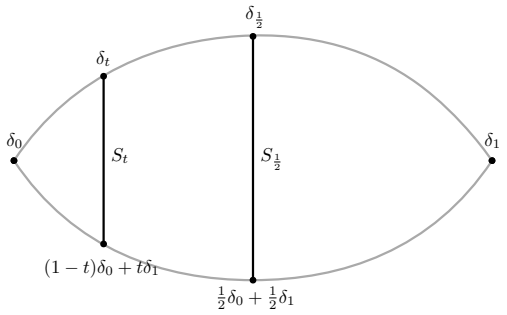

FIGURE 1. The shape of the Wasserstein space $\mathcal{W}_1([0,1])$.

Claim 2.4. Let $\varphi\colon \mathcal{W}_1([0,1]) \to \mathcal{W}_1([0,1])$ be an isometric embedding such that $\varphi(\delta_0) = \delta_0$ and $\varphi\left(\delta_{\frac{1}{2}}\right) = \delta_{\frac{1}{2}}$. Then $\varphi(\mu) = \mu$ for all $\mu \in \mathcal{W}_1([0,1])$.

Proof. Using Claim 2.3, we obtain

$$\{\varphi\left((1-t)\delta_0 + t\delta_1\right), \varphi\left(\delta_t\right)\} = \{(1-t)\delta_0 + t\delta_1, \delta_t\} \quad (t \in [0,1]).$$

In particular, $\varphi\left(\delta_1\right) = \delta_1$. As for all $0 < t < 1$ we have

$$d_{\mathcal{W}_1}\left(\delta_t, \delta_{\frac{1}{2}}\right) = \left|t - \tfrac{1}{2}\right| < \tfrac{1}{2} \quad \text{and} \quad d_{\mathcal{W}_1}\left((1-t)\delta_0 + t\delta_1, \delta_{\frac{1}{2}}\right) = \tfrac{1}{2},$$

we get that $\varphi\left(\delta_t\right) = \delta_t$. Therefore it is enough to show that any measure $\mu \in \mathcal{W}_1([0,1])$ is completely determined by its distances from Dirac masses. This can be seen in the following way: by (1.2) we have

$$d_{\mathcal{W}_1}\left(\mu, \delta_t\right) = \int_0^t F_\mu(x)\ \mathrm{d}x + \int_t^1 \left(1 - F_\mu(x)\right)\ \mathrm{d}x \quad (t \in [0,1]), \tag{2.9}$$

hence

$$\lim_{h \searrow 0} \frac{d_{\mathcal{W}_1}\left(\mu, \delta_{t+h}\right) - d_{\mathcal{W}_1}\left(\mu, \delta_t\right)}{h} = \lim_{h \searrow 0} \frac{1}{h} \int_t^{t+h} \left(2F_\mu(x) - 1\right)\ \mathrm{d}x = 2F_\mu(t) - 1 \tag{2.10}$$

holds for all $t \in [0,1)$. The proof is done. □

Now we are in the position to present the main result of this subsection.

Theorem 2.5. Let $\varphi \in \mathrm{IsEmb}(\mathcal{W}_1([0,1]))$, that is,

$$d_{\mathcal{W}_1}\left(\varphi(\mu), \varphi(\nu)\right) = d_{\mathcal{W}_1}\left(\mu, \nu\right) \qquad (\mu, \nu \in \mathcal{W}_1([0,1])).$$

Then $\varphi \in \{\mathrm{id}_{\mathcal{W}_1([0,1])}, r_\#, j, r_\# j\}$, where $r_\# j = j r_\#$. Consequently, every isometric embedding is surjective, that is,

$$\mathrm{IsEmb}(\mathcal{W}_1([0,1])) = \mathrm{Isom}(\mathcal{W}_1([0,1])).$$

Moreover, this isometry group is isomorphic to the Klein group $C_2 \times C_2$.

Proof. Clearly, $\varphi(\delta_0) \in \{\delta_0, \delta_1\}$ and it follows from Claim 2.3 that

$$\varphi\left(\delta_{\frac{1}{2}}\right) \in \left\{\delta_{\frac{1}{2}}, \tfrac{1}{2}\delta_0 + \tfrac{1}{2}\delta_1\right\}.$$

Therefore we have four cases to check. If $\varphi\left(\delta_0\right) = \delta_0$ and $\varphi\left(\delta_{\frac{1}{2}}\right) = \delta_{\frac{1}{2}}$, then by Claim 2.4 $\varphi = \mathrm{id}_{\mathcal{W}_1([0,1])}$. Next, if $\varphi\left(\delta_0\right) = \delta_1$ and $\varphi\left(\delta_{\frac{1}{2}}\right) = \delta_{\frac{1}{2}}$, then the isometric embedding $r_\#\varphi$ sends δ_0 to δ_0 and $\delta_{\frac{1}{2}}$ to $\delta_{\frac{1}{2}}$. Consequently, $r_\#\varphi = \mathrm{id}_{\mathcal{W}_1([0,1])}$ and $\varphi = r_\#$. Similarly, if $\varphi\left(\delta_0\right) = \delta_1$ and $\varphi\left(\delta_{\frac{1}{2}}\right) = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$, then $j\varphi$ leaves δ_0 and $\delta_{\frac{1}{2}}$ invariant, which implies $\varphi = j$.

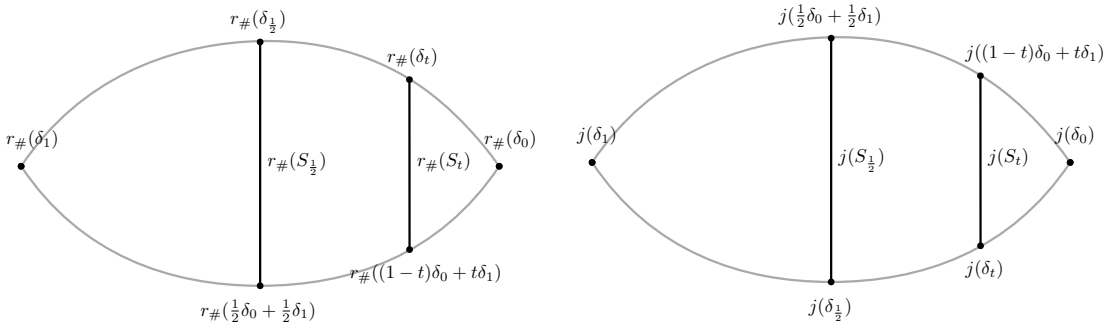

FIGURE 2. The action of $r_\#$ and j on $\mathcal{W}_1([0,1])$, cf. Figure 1.

Finally, if $\varphi(\delta_0)=\delta_0$ and $\varphi\left(\delta_{\frac12}\right)=\frac12\delta_0+\frac12\delta_1$, then $r_\# j\varphi$ and $jr_\#\varphi$ are isometric embeddings leaving both δ_0 and $\delta_{\frac12}$ invariant. Therefore $\varphi=jr_\#=r_\# j$. □

We close this subsection by noting that the metric structure of $\mathcal{W}_1([a,b])$ is similar to that of $\mathcal{W}_1([0,1])$ for all $a<b$. Consequently, our method works for any compact interval. Consider the function $\lambda_{a,b}\colon [0,1]\to[a,b]$, $\lambda_{a,b}(t)=a+(b-a)t$ and the following push-forward bijection: $\xi_{a,b}\colon \mathcal{W}_1([0,1])\to\mathcal{W}_1([a,b])$, $\xi_{a,b}(\mu)=\lambda_{a,b\#}(\mu)$. Notice that $\varphi\in\mathrm{IsEmb}\left(\mathcal{W}_1([a,b])\right)$ holds if and only if $(\xi_{a,b})^{-1}\circ\varphi\circ\xi_{a,b}\in\mathrm{IsEmb}\left(\mathcal{W}_1([0,1])\right)$, so

$$\mathrm{IsEmb}\left(\mathcal{W}_1([a,b])\right)=\mathrm{Isom}\left(\mathcal{W}_1([a,b])\right)=C_2\times C_2.$$

In the next subsection we continue by describing all isometric embeddings of the Wasserstein space $\mathcal{W}_p([0,1])$ for parameters $p>1$.

2.2. $p>1$ − **Isometric rigidity.** Similarly to the case of $p=1$, it turns out that every isometric embedding is surjective. However, in contrast to the case $p=1$, we prove *isometric rigidity*, that is, $\mathrm{Isom}(\mathcal{W}_p([0,1]))$ is isomorphic to $\mathrm{Isom}([0,1])=C_2$ for $p>1$.

Note that (1.3) and the strict convexity of the L^p-norm for $p>1$ implies the following: for any $\mu,\nu\in\mathcal{W}_p([0,1])$ and $s\in[0,1]$ there exists a unique measure $\gamma_{\mu,\nu}(s)\in\mathcal{W}_p([0,1])$ such that

$$d_{\mathcal{W}_p}\left(\mu,\gamma_{\mu,\nu}(s)\right)=s\cdot d_{\mathcal{W}_p}\left(\mu,\nu\right)\quad\text{and}\quad d_{\mathcal{W}_p}\left(\gamma_{\mu,\nu}(s),\nu\right)=(1-s)\cdot d_{\mathcal{W}_p}\left(\mu,\nu\right),$$

moreover, $\gamma_{\mu,\nu}(s)$ is defined by the equation

$$F^{-1}_{\gamma_{\mu,\nu}(s)}=(1-s)F^{-1}_\mu+sF^{-1}_\nu\qquad(s\in[0,1]).$$

Consequently, for any isometric embedding $\varphi\colon \mathcal{W}_p([0,1])\to\mathcal{W}_p([0,1])$ we have the following compatibility equation:

$$\varphi\left(\gamma_{\mu,\nu}(s)\right)=\gamma_{\varphi(\mu),\varphi(\nu)}(s)\qquad(\mu,\nu\in\mathcal{W}_p([0,1]),s\in[0,1]).$$

In particular, if φ leaves μ and ν fixed, then it leaves $\gamma_{\mu,\nu}(s)$ fixed for all $s\in[0,1]$.

Note that $\gamma_{\mu,\nu}$ is the unique geodesic parametrized by the unit interval $[0,1]$ between μ and ν. We define the *convex hull* $\mathbf{co}(S)$ of a set $S\subseteq\mathcal{W}_p([0,1])$ as the *closure* of the set of all measures with quantile functions of the form

$$\sum_{j=1}^N\alpha_jF^{-1}_{\nu_j},\qquad\left(N\in\mathbb{N},\,\{(\nu_j,\alpha_j)\}_{j=1}^N\subseteq S\times(0,1],\,\sum_{j=1}^N\alpha_j=1\right).$$

In other words, $\mathbf{co}(S)$ is the set of those measures whose quantile functions belong to the L^p-closed convex hull of quantile functions of measures in S. Now we can generalize the above remark: if an isometric embedding leaves every element of S invariant, then it leaves every element of $\mathbf{co}(S)$ fixed, as well.

Similarly to the case $p=1$, every $\varphi\in\mathrm{IsEmb}(\mathcal{W}_p([0,1]))$ satisfies $\{\varphi(\delta_0),\varphi(\delta_1)\}=\{\delta_0,\delta_1\}$. Without loss of generality we assume from now on that

$$\varphi(\delta_0)=\delta_0\quad\text{and}\quad\varphi(\delta_1)=\delta_1$$

(otherwise we can work with $r_{\#}\varphi$). Set $M_0 := \{\delta_0, \delta_1\}$ and $M_1 := \mathbf{co}\,(M_0)$. Obviously, we have $M_1 = \Delta([0,1])$, and thus

$$\varphi(\delta_t) = \delta_t \quad (t \in [0,1]). \tag{2.11}$$

Let us introduce the notation for any measure $\mu \in \mathcal{W}_p([0,1])$ and any nonempty set $H \subseteq \mathcal{W}_p([0,1])$

$$\mathrm{dist}_{\mathcal{W}_p}(\mu, H) := \inf\left\{ d_{\mathcal{W}_p}(\mu,\nu) \;\middle|\; \nu \in H \right\}.$$

We are searching for elements of $\mathcal{W}_p([0,1])$ that can be characterized by means of extremal metric properties. More precisely, in the next Claim we determine the maximizer of $\mathrm{dist}_{\mathcal{W}_p}(\mu, M_1)$.

Claim 2.6. For every measure $\mu \in \mathcal{W}_p([0,1])$ we have

$$\mathrm{dist}_{\mathcal{W}_p}(\mu, M_1) \le \tfrac{1}{2}. \tag{2.12}$$

Moreover, $\mathrm{dist}_{\mathcal{W}_p}(\mu, M_1) = \frac{1}{2}$, if and only if $\mu = \frac{1}{2}(\delta_0 + \delta_1)$.

Proof. The first statement of the claim easily follows from the fact that

$$d^p_{\mathcal{W}_p}\left(\mu, \delta_{\frac{1}{2}}\right) = \int_{[0,1]} \left|x - \tfrac{1}{2}\right|^p \mathrm{d}\mu(x) \le \left(\tfrac{1}{2}\right)^p \qquad (\mu \in \mathcal{W}_p([0,1])). \tag{2.13}$$

As for the second statement, note that equality in (2.13) implies $\mu(\{0,1\}) = 1$. Consequently, if $d_{\mathcal{W}_p}\left(\mu, \delta_{\frac{1}{2}}\right) = \frac{1}{2}$, then $\mu = (1-\alpha)\delta_0 + \alpha\delta_1$ for some $\alpha \in [0,1]$. Moreover, notice that

$$d^p_{\mathcal{W}_p}((1-\alpha)\delta_0 + \alpha\delta_1, \delta_0) = \alpha \;\text{ and }\; d^p_{\mathcal{W}_p}((1-\alpha)\delta_0 + \alpha\delta_1, \delta_1) = 1-\alpha,$$

hence equality in (2.12) implies $\mu = \frac{1}{2}(\delta_0 + \delta_1)$. The following easy calculation shows that $\mathrm{dist}_{\mathcal{W}_p}\left(\frac{1}{2}(\delta_0+\delta_1), M_1\right)$ is indeed $\frac{1}{2}$:

$$d^p_{\mathcal{W}_p}(\tfrac{1}{2}(\delta_0+\delta_1), \delta_t) = \tfrac{1}{2}(t^p + (1-t)^p)) \qquad \text{for all} \quad t \in [0,1].$$

This expression is strictly convex in $t \in [0,1]$ and has a unique minimizer $t = \frac{1}{2}$. The proof is complete. □

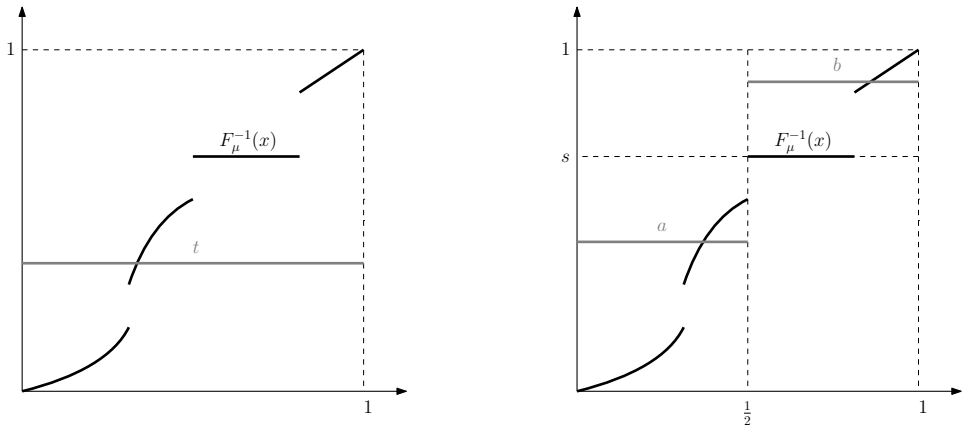

FIGURE 3. Illustration for Claims 2.6 and 2.8.

The following corollary highlights the goal of the previous claim. Namely, if the largest distance from a pointwise invariant set is attained uniquely, then that unique maximizer must be a fixed point as well.

Corollary 2.7. If $\varphi \in \mathrm{IsEmb}\,(\mathcal{W}_p([0,1]))$ with $\varphi(\delta_0) = \delta_0$, then

$$\varphi\left(\tfrac{1}{2}\delta_a + \tfrac{1}{2}\delta_b\right) = \tfrac{1}{2}\delta_a + \tfrac{1}{2}\delta_b \quad (0 \le a \le b \le 1).$$

Proof. It is easy to see by considering the inverse distribution functions that

$$M_2 := \mathbf{co}\Big(M_1 \cup \left\{\tfrac{1}{2}\delta_0 + \tfrac{1}{2}\delta_1\right\}\Big) = \left\{\tfrac{1}{2}\delta_a + \tfrac{1}{2}\delta_b \;\middle|\; 0 \le a \le b \le 1\right\}.$$

Using Claim 2.6 and (2.11) we conclude that $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ is a fixed point of φ. Therefore φ acts identically on M_2. □

As some kind of an induction step, in the following claim we identify the maximizers of the distance from the set M_2.

Claim 2.8. For every measure $\mu \in \mathcal{W}_p([0,1])$ we have

$$d_{\mathcal{W}_p}(\mu, M_2) \leq \left(\tfrac{1}{2}\right)^{1+\frac{1}{p}},$$

and $d_{\mathcal{W}_p}(\mu, M_2) = \left(\tfrac{1}{2}\right)^{1+\frac{1}{p}}$ if and only if $\mu = \frac{3}{4}\delta_0 + \frac{1}{4}\delta_1$ or $\mu = \frac{1}{4}\delta_0 + \frac{3}{4}\delta_1$.

Proof. By (1.3), we have

$$d^p_{\mathcal{W}_p}\left(\mu, \tfrac{1}{2}\delta_a + \tfrac{1}{2}\delta_b\right) = \int_0^{\frac{1}{2}} \left|F_\mu^{-1}(x) - a\right|^p \, \mathrm{d}x + \int_{\frac{1}{2}}^1 \left|F_\mu^{-1}(x) - b\right|^p \, \mathrm{d}x. \tag{2.14}$$

Let us introduce the set

$$R_s := \left\{\mu \in \mathcal{W}_p([0,1]) \,\middle|\, F_\mu^{-1}\left(\tfrac{1}{2}\right) = s\right\} \qquad (s \in [0,1]), \tag{2.15}$$

A simple rescaling of Claim 2.6 gives that for any $\mu \in R_s$ we have

$$\inf_{a\in[0,1]} \int_0^{\frac{1}{2}} \left|F_\mu^{-1}(x) - a\right|^p \, \mathrm{d}x = \inf_{a\in[0,s]} \int_0^{\frac{1}{2}} \left|F_\mu^{-1}(x) - a\right|^p \, \mathrm{d}x \leq \tfrac{1}{2}\left(\tfrac{s}{2}\right)^p,$$

and equality holds if and only if

$$F_\mu^{-1}(x) = \begin{cases} 0 & \text{for } x \in \left[0, \frac{1}{4}\right) \\ s & \text{for } x \in \left[\frac{1}{4}, \frac{1}{2}\right) \end{cases},$$

in which case the minimizer is $a = \frac{s}{2}$ (see Figure 3). And similarly,

$$\inf_{b\in[0,1]} \int_{\frac{1}{2}}^1 \left|F_\mu^{-1}(x) - b\right|^p \, \mathrm{d}x \leq \tfrac{1}{2}\left(\tfrac{1-s}{2}\right)^p$$

and equality holds if and only if

$$F_\mu^{-1}(x) = \begin{cases} s & \text{for } x \in \left[\frac{1}{2}, \frac{3}{4}\right) \\ 1 & \text{for } x \in \left[\frac{3}{4}, 1\right] \end{cases},$$

in which case the minimizer is $b = \frac{s+1}{2}$. By (2.14), this means that

$$d^p_{\mathcal{W}_p}(\mu, M_2) = \inf_{0\leq a\leq b\leq 1} d^p_{\mathcal{W}_p}\left(\mu, \tfrac{1}{2}\delta_a + \tfrac{1}{2}\delta_b\right) \leq \tfrac{1}{2}\left(\tfrac{s}{2}\right)^p + \tfrac{1}{2}\left(\tfrac{1-s}{2}\right)^p \quad (\mu \in R_s) \tag{2.16}$$

with equality if and only if $\mu = \frac{1}{4}\delta_0 + \frac{1}{2}\delta_s + \frac{1}{4}\delta_1$. By the strict convexity of $x \mapsto x^p$, the right hand side of (2.16) is maximal if and only if $s \in \{0,1\}$, and in both cases, the maximum value is $\left(\frac{1}{2}\right)^{p+1}$. The proof is done. □

Corollary 2.9. Let $\varphi \in \mathrm{IsEmb}(\mathcal{W}_p([0,1]))$ that leaves δ_0 invariant. Then all measures of the form $\mu = \frac{1}{4}\sum_{j=1}^4 \delta_{a_j}$ (where $0 \leq a_1 \leq \cdots \leq a_4 \leq 1$) are invariant under the action of φ.

Proof. With the same argument as in the previous corollary, one can see that we have either $\varphi\left(\frac{3}{4}\delta_0 + \frac{1}{4}\delta_1\right) = \frac{3}{4}\delta_0 + \frac{1}{4}\delta_1$ and $\varphi\left(\frac{1}{4}\delta_0 + \frac{3}{4}\delta_1\right) = \frac{1}{4}\delta_0 + \frac{3}{4}\delta_1$, or the other way around. However, by considering the distances from δ_0, the latter is easily excluded. Now, set

$$M_3 := \mathbf{co}\left(M_2 \cup \left\{\tfrac{3}{4}\delta_0 + \tfrac{1}{4}\delta_1, \tfrac{1}{4}\delta_0 + \tfrac{3}{4}\delta_1\right\}\right),$$

and observe that

$$M_3 = \left\{ \tfrac{1}{4}\sum_{j=1}^4 \delta_{a_j} \,\middle|\, 0 \leq a_1 \leq \cdots \leq a_4 \leq 1 \right\}.$$

Since pointwise invariance of a set forces the pointwise invariance of its convex hull, the proof is complete. □

This procedure can be iterated: following this pattern, it can be shown that measures of the form

$$\mu = \tfrac{1}{2^N} \sum_{j=1}^{2^N} \delta_{a_j} \quad (0 \le a_1 \le \cdots \le a_{2^N})$$

are left invariant under isometric embeddings fixing δ_0. These measures form a dense subset of $\mathcal{W}_p([0,1])$, thus every element of $\mathcal{W}_p([0,1])$ is left invariant by an isometric embedding fixing δ_0. We recall again that if $\varphi(\delta_0) \neq \delta_0$, then $r_\# \varphi(\delta_0) = \delta_0$, which forces $r_\# \varphi = \mathrm{id}_{\mathcal{W}_p([0,1])}$, or equivalently, $\varphi = r_\#$. Now, we summarize the above observations and formulate the main result of this subsection.

Theorem 2.10. Let $p > 1$ and let $\varphi \in \mathrm{IsEmb}(\mathcal{W}_p([0,1]))$, that is,

$$d_{\mathcal{W}_p}(\varphi(\mu), \varphi(\nu)) = d_{\mathcal{W}_p}(\mu, \nu) \qquad (\mu, \nu \in \mathcal{W}_p([0,1])).$$

Then we have the following two possibilities: either $\varphi = \mathrm{id}_{\mathcal{W}_p([0,1])}$, or $\varphi = r_\#$. Consequently, every isometric embedding is surjective, that is,

$$\mathrm{IsEmb}(\mathcal{W}_p([0,1])) = \mathrm{Isom}(\mathcal{W}_p([0,1])) = C_2.$$

As at the end of the previous subsection, one can examine the isometric embeddings of $\mathcal{W}_p([a,b])$ using a map defined very similarly as $\xi_{a,b}$. One then obtains that all isometric embeddings are bijective, and that there are only the two trivial isometries.

3. Isometric study of $\mathcal{W}_p(\mathbb{R})$

We have seen that the structure of $\mathrm{Isom}(\mathcal{W}_p([0,1]))$ can be different for different parameters p, and that $\mathrm{IsEmb}(\mathcal{W}_p([0,1])) = \mathrm{Isom}(\mathcal{W}_p([0,1]))$ for all p. Our next goal is to examine isometries and isometric embeddings of Wasserstein spaces over the real line. In contrast to the interval case, here it will turn out that $\mathrm{Isom}(\mathcal{W}_p(\mathbb{R})) \subsetneq \mathrm{IsEmb}(\mathcal{W}_p(\mathbb{R}))$. However, we will also see that the structure of the isometry group can be again different for different parameters p, and that the same holds for the semigroup $\mathrm{IsEmb}(\mathcal{W}_p(\mathbb{R}))$. During our investigation, the parameters $p = 1$ and $p = 2$ have to be handled separately.

As for the $p = 2$ case, we recall that the structure of $\mathrm{Isom}(\mathcal{W}_2(\mathbb{R}))$ has been described by Kloeckner in [12]. In particular, Kloeckner showed that $\mathcal{W}_2(\mathbb{R})$ admits non-trivial isometries, moreover there exists a so-called exotic flow of isometries that does not even preserve the shape of measures. We will discuss this exotic flow in detail in the last subsection.

3.1. $p = 1$ – Isometric rigidity.

The goal of this subsection is to describe the isometry group of $\mathcal{W}_1(\mathbb{R})$. Namely, we prove that it admits only trivial isometries. One important difference between the $p = 1$ and $p > 1$ cases is that the L^1 norm is not strictly convex. Consequently, in $\mathcal{W}_1(\mathbb{R})$ the optimal transport plan between measures is not unique (let alone the geodesic curve). Thus Kloeckner's idea of characterizing Dirac masses by means of geodesics cannot be adapted for $p = 1$. Actually, here the main difficulty is to find a metric characterization of Dirac masses. We start with a definition.

Definition 3.1 (Metric midpoints)**.** For $\mu, \nu \in \mathcal{W}_1(\mathbb{R})$, $\mu \neq \nu$, the following set is called the *metric midpoint set* of μ and ν:

$$M(\mu,\nu) := \left\{ \xi \in \mathcal{W}_1(\mathbb{R}) \,\middle|\, d_{\mathcal{W}_1}(\mu,\xi) = d_{\mathcal{W}_1}(\xi,\nu) = \tfrac{1}{2} d_{\mathcal{W}_1}(\mu,\nu) \right\}.$$

We continue with proving a metric property of the metric midpoint set, during which we will identify two special elements of $M(\mu,\nu)$.

Claim 3.2. For $\mu, \nu \in \mathcal{W}_1(\mathbb{R})$, $\mu \neq \nu$ we always have

$$\tfrac{1}{2} d_{\mathcal{W}_1}(\mu,\nu) \le \mathrm{diam}(M(\mu,\nu)) \le d_{\mathcal{W}_1}(\mu,\nu). \tag{3.1}$$

Proof. The second inequality is trivial by the triangle inequality, hence we shall only focus on the first inequality. For the sake of brevity, we introduce the notation $D := d_{\mathcal{W}_1}(\mu,\nu) \neq 0$. By a simple geometric consideration it is obvious that D is exactly the Lebesgue measure of the Borel set

$$\mathcal{S} := \left\{ (x,y) \in \mathbb{R} \times [0,1] \,\middle|\, F_\mu(x) \le y \le F_\nu(x) \text{ or } F_\nu(x) \le y \le F_\mu(x) \right\}.$$

Hence, there exist two numbers $h \in (0,1)$ and $v \in \mathbb{R}$ such that the following four sets have Lebesgue measure $\frac{D}{2}$:

$$\mathcal{S} \cap ((-\infty, v) \times [0,1]), \ \ \mathcal{S} \cap ((v, \infty) \times [0,1]), \ \ \mathcal{S} \cap (\mathbb{R} \times [0,h)), \ \ \mathcal{S} \cap (\mathbb{R} \times (h,1]).$$

Let us define the sets

$$\mathcal{S}_1 := \mathcal{S} \cap ((-\infty, v) \times [0,h)), \ \ \mathcal{S}_2 := \mathcal{S} \cap ((v, \infty) \times [0,h)),$$
$$\mathcal{S}_3 := \mathcal{S} \cap ((v, \infty) \times (h,1]), \ \ \mathcal{S}_4 := \mathcal{S} \cap ((-\infty, v) \times (h,1]),$$

and denote the Lebesgue measures of these sets by $\alpha_1, \alpha_2, \alpha_3$ and α_4, respectively. By the choice of v and h we have

$$\alpha_1 = \alpha_3, \ \ \alpha_2 = \alpha_4, \ \ \text{and} \ \ \alpha_1 + \alpha_2 = \alpha_3 + \alpha_4 = \tfrac{D}{2}.$$

From here we consider two cases. First, assume that $\alpha_2 = \alpha_4 = 0$. Suppose for a moment that we have $F_\mu(v) < h$ (the case $F_\nu(v) < h$ is handled similarly). Then by right-continuity we obtain easily that $F_\mu(x) = F_\nu(x)$ for all $v \le x < \sup\{x \in \mathbb{R} \,|\, F_\mu(x) < h\}$. Therefore by monotonicity of F_μ, we can choose h to be $F_\mu(v)$. By doing so, we may assume without loss of generality from now on that

$$F_\mu(v) \ge h \quad \text{and} \quad F_\nu(v) \ge h.$$

Notice that since $\alpha_4 = 0$, we must have

$$(F_\mu \vee F_\nu)(v-) \le (F_\mu \wedge F_\nu)(v).$$

Hence, the measures ξ and η defined in the following way are clearly in $M(\mu,\nu)$ and their distance is obviously D, which proves the inequality for this case:

$$F_\xi(x) := \begin{cases} F_\mu(x) & \text{if } x < v \\ F_\nu(x) & \text{if } x \ge v \end{cases} \quad \text{and} \quad F_\eta(x) := \begin{cases} F_\nu(x) & \text{if } x < v \\ F_\mu(x) & \text{if } x \ge v \end{cases}.$$

Second, assume that $\alpha_2 = \alpha_4 > 0$. Notice that we cannot have both $h \le F_\mu(v)$ and $h \le F_\nu(v)$, since that would imply $\mathcal{S}_2 = \emptyset$. Similarly, having both $F_\mu(v-) \le h$ and $F_\nu(v-) \le h$ would imply $\mathcal{S}_4 = \emptyset$. Hence by symmetry we may assume without loss of generality that

$$F_\mu(v) < h < F_\nu(v-).$$

Now, we define two measures $\xi_v^{\mu,\nu}$ and $\xi_h^{\mu,\nu}$ with their cumulative distribution functions:

$$F_{\xi_v^{\mu,\nu}}(x) := \begin{cases} F_\mu(x) & \text{if } x < v \\ F_\nu(x) & \text{if } x \ge v \end{cases} \tag{3.2}$$

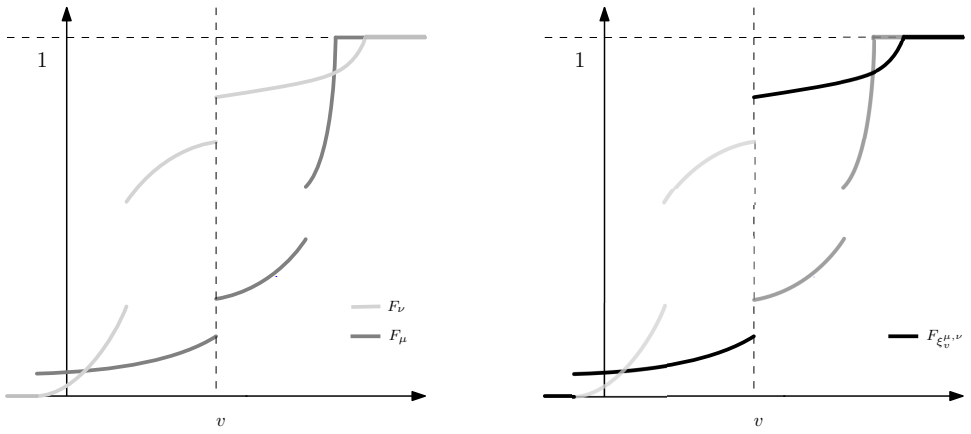

FIGURE 4. $F_{\xi_v^{\mu,\nu}}$ splits the area between F_μ and F_ν vertically.

and

$$F_{\xi_h^{\mu,\nu}}(x) := \begin{cases} F_\nu(x) & \text{if } x < F_\nu^{-1}(h) \\ h & \text{if } F_\nu^{-1}(h) \le x < F_\mu^{-1}(h) \\ F_\mu(x) & \text{if } x \ge F_\mu^{-1}(h) \end{cases}. \tag{3.3}$$

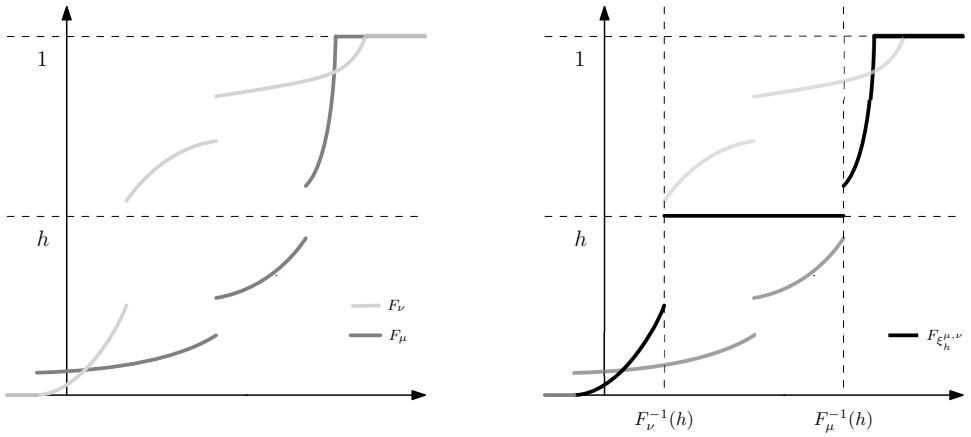

FIGURE 5. $F_{\xi_h^{\mu,\nu}}$ splits the area between F_μ and F_ν horizontally.

It is obvious that $F_\nu^{-1}(h) < v < F_\mu^{-1}(h)$, $\xi_v^{\mu,\nu}, \xi_h^{\mu,\nu} \in M(\mu,\nu)$ and $d_{\mathcal{W}_1}(\xi_v^{\mu,\nu}, \xi_h^{\mu,\nu}) = \alpha_1 + \alpha_3$. From here we verify that $\alpha_1 = \alpha_3 \geq \alpha_2 = \alpha_4$ by the following geometric observation. We consider the auxiliary rectangle $\big(F_\nu^{-1}(h), F_\mu^{-1}(h)\big) \times \big(F_\mu(v), F_\nu(v)\big)$, and split it into four parts using the horizontal and vertical lines corresponding to h and v, respectively, see Figure 6. Denoting the area of these pieces by β_i's in accordance with α_i's $(1 \leq i \leq 4)$, we obtain

$$\beta_1 \leq \alpha_1, \;\; \beta_2 \geq \alpha_2, \;\; \beta_3 \leq \alpha_3, \;\; \text{and} \;\; \beta_4 \geq \alpha_4. \tag{3.4}$$

But obviously, depending on h we have either $\beta_4 \leq \beta_1$, or $\beta_2 \leq \beta_3$, which combined with the previous inequalities completes the proof. □

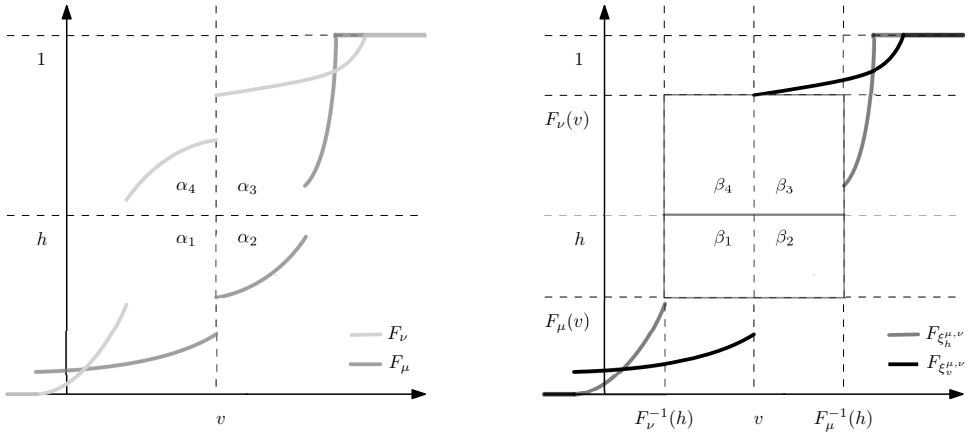

FIGURE 6. Partitioning the area between the graphs with vertical and horizontal lines, and the auxiliary rectangle.

Definition 3.3 (Vertical and horizontal bisecting measures)**.** If $\mu, \nu \in \mathcal{W}_1(\mathbb{R})$ are measures such that $\alpha_2 = \alpha_4 > 0$ with the above defined numbers, then the measures $\xi_v^{\mu,\nu}$ and $\xi_h^{\mu,\nu}$ defined in (3.2)–(3.3) are called the *vertical* and *horizontal bisecting measures* of μ and ν, respectively.

We proceed with examining when the first inequality in (3.1) becomes an equality.

Definition 3.4 (Adjacent measures)**.** Two different elements μ and ν of $\mathcal{W}_1(\mathbb{R})$ are said to be *adjacent*, in notation $\mu \sim \nu$, if there exists an interval $(a,b) \subseteq \mathbb{R}$ such that

(1) $\mu|_{\mathbb{R}\setminus\{a,b\}} = \nu|_{\mathbb{R}\setminus\{a,b\}}$ and
(2) $\mu\big((a,b)\big) = \nu\big((a,b)\big) = 0$.

Or equivalently,

(1') $F_\mu|_{\mathbb{R}\setminus[a,b)} \equiv F_\nu|_{\mathbb{R}\setminus[a,b)}$ and
(2') both $F_\mu|_{[a,b)}$ and $F_\nu|_{[a,b)}$ are constant.

Observe that for adjacent measures we have $\alpha_2 = \alpha_4 > 0$, hence the vertical and horizontal bisecting measures are defined by Definition 3.3.

Claim 3.5. For any $\mu, \nu \in \mathcal{W}_1(\mathbb{R}), \mu \neq \nu$ the following statements are equivalent

(i) $\operatorname{diam}(M(\mu,\nu)) = \frac{1}{2} d_{\mathcal{W}_1}(\mu,\nu)$,
(ii) $\mu \sim \nu$.

Moreover, if $\mu \sim \nu$, then the diameter is attained only for the pair $\left\{\xi_v^{\mu,\nu}, \xi_h^{\mu,\nu}\right\}$.

Proof. We continue to use the notations of Claim 3.2. First, we prove the direction (i)$\Longrightarrow$(ii). As $\frac{D}{2} = \operatorname{diam}(M(\mu,\nu)) \geq \alpha_1 + \alpha_3 \geq \alpha_2 + \alpha_4$, we immediately obtain $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \frac{D}{4}$. Combining this with (3.4) gives

$$\beta_1 \leq \tfrac{D}{4}, \;\; \beta_2 \geq \tfrac{D}{4}, \;\; \beta_3 \leq \tfrac{D}{4} \;\text{ and }\; \beta_4 \geq \tfrac{D}{4},$$

from which, by simple geometric considerations, we conclude $\beta_i = \frac{D}{4} = \alpha_i$ for all $i = 1, 2, 3, 4$. In particular, $\mu \sim \nu$ follows.

As for the reverse direction (ii)$\Longrightarrow$(i), we only need to observe that

$$F_\eta|_{\mathbb{R}\setminus[a,b]} = F_\mu|_{\mathbb{R}\setminus[a,b]} = F_\nu|_{\mathbb{R}\setminus[a,b]} \qquad (\eta \in M(\mu,\nu)).$$

Indeed, elements of $M(\mu,\nu)$ saturate the triangle inequality

$$d_{\mathcal{W}_1}(\mu,\nu) = d_{\mathcal{W}_1}(\mu,\eta) + d_{\mathcal{W}_1}(\eta,\nu).$$

Hence by (1.2) we have $F_\mu \wedge F_\nu \leq F_\eta \leq F_\mu \vee F_\nu$. Now, we basically reduced the problem to the case of the interval, and thus the argument of Claim 2.3 can be applied with a simple rescaling. In such a way one obtains

$$d_{\mathcal{W}_1}(\eta_1,\eta_2) = \int_a^b |F_{\eta_1}(t) - F_{\eta_2}(t)| \mathrm{d}t \leq \tfrac{1}{2} D \qquad (\eta_1, \eta_2 \in M(\mu,\nu))$$

with equality if and only if $\{\eta_1, \eta_2\} = \{\xi_v^{\mu,\nu}, \xi_h^{\mu,\nu}\}$. □

Now, we are in the position to give a metric characterization of Dirac masses.

Claim 3.6. For a measure $\eta \in \mathcal{W}_1(\mathbb{R})$ the following statements are equivalent

(i) $\eta \in \Delta(\mathbb{R})$,
(ii) for all $n \in \mathbb{N}$ there are measures $\mu_n, \nu_n \in \mathcal{W}_1(\mathbb{R})$ such that
  (a) $\mu_n \sim \nu_n$,
  (b) $d_{\mathcal{W}_1}(\mu_n, \nu_n) = n$,
  (c) $\eta \in \{\xi_v^{\mu_n,\nu_n}, \xi_h^{\mu_n,\nu_n}\}$.

Proof. Assume first that $\eta = \delta_t$ for some $t \in \mathbb{R}$. Then the choices $\mu_n := \delta_{t-\frac{1}{2}n}$ and $\nu_n := \delta_{t+\frac{1}{2}n}$ ($n \in \mathbb{N}$) obviously satisfy (ii). Therefore what remained to show is that it is impossible to have $\eta \notin \Delta(\mathbb{R})$ and (ii) at the same time. We shall prove this indirectly, so from now on we assume that $\eta \notin \Delta(\mathbb{R})$ fulfils (ii).

By definition, for all $n \in \mathbb{N}$ there exists an interval $[a_n, b_n)$ such that

$$F_{\mu_n}|_{\mathbb{R}\setminus[a_n,b_n)} = F_{\nu_n}|_{\mathbb{R}\setminus[a_n,b_n)} = F_\eta|_{\mathbb{R}\setminus[a_n,b_n)},$$

and that $F_{\mu_n}|_{[a_n,b_n)}$ and $F_{\nu_n}|_{[a_n,b_n)}$ are both constants. Set

$$\alpha_n := (F_{\mu_n} \wedge F_{\nu_n})(a_n), \quad \text{and} \quad \beta_n := (F_{\mu_n} \vee F_{\nu_n})(a_n).$$

Notice that $(\beta_n - \alpha_n)(b_n - a_n) = n$, hence $b_n - a_n \geq n$.

If $\eta = \xi_h^{\mu_n,\nu_n}$, then $F_\eta|_{[a_n,b_n)}$ is also constant with $F_\eta(a_n) = \frac{\alpha_n+\beta_n}{2}$. A simple geometric consideration shows that in this case we have

$$d_{\mathcal{W}_1}(\eta,\delta_0) \geq \int_{a_n}^{b_n} |F_\eta(t) - F_{\delta_0}(t)| \mathrm{d}t \geq \tfrac{n}{2}.$$

Therefore there exists a number $N \in \mathbb{N}$ such that

$$\eta = \xi_v^{\mu_n,\nu_n} \quad (n \geq N).$$

Again by definition, for all $n \geq N$ we have that F_η is constant α_n on $\left[a_n, \frac{1}{2}(a_n + b_n)\right)$, and constant β_n on $\left[\frac{1}{2}(a_n + b_n), b_n\right)$. As η is not a Dirac mass, we get that there is a maximal positive

number $s_n \geq \frac{b_n - a_n}{2}$ such that F_η is constant on both intervals $\left[\frac{1}{2}(a_n+b_n) - s_n, \frac{1}{2}(a_n+b_n)\right)$ and $\left[\frac{1}{2}(a_n+b_n), \frac{1}{2}(a_n+b_n) + s_n\right)$. Therefore there exists an infinite subset $\mathcal{N}$ of positive integers such that for all $j,k \in \mathcal{N}, j \neq k$ we have

$$\left[a_j, \tfrac{1}{2}(a_j+b_j)\right) \cap \left[\tfrac{1}{2}(a_k+b_k), b_k\right) = \emptyset$$

or

$$\left[a_k, \tfrac{1}{2}(a_k+b_k)\right) \cap \left[\tfrac{1}{2}(a_j+b_j), b_j\right) = \emptyset.$$

Keeping in mind that $b_n - a_n \geq n$ $(n \in \mathbb{N})$ gives that $\left\{\frac{1}{2}(a_j+b_j) \,|\, j \in \mathcal{N}\right\}$ is a set that clusters at $+\infty$ or $-\infty$. Suppose it clusters at least at $+\infty$. Then for large enough numbers $j \in \mathcal{N}$ one easily concludes that

$$d_{\mathcal{W}_1}(\eta, \delta_0) \geq \int_{a_j}^{\frac{1}{2}(a_j+b_j)} |F_\eta(t) - F_{\delta_0}(t)| \,\mathrm{d}t \geq \tfrac{j}{2},$$

which is a contradiction. If $\left\{\frac{1}{2}(a_j+b_j) \,\middle|\, j \in \mathcal{N}\right\}$ clusters only at $-\infty$, then with a similar method we conclude a contradiction. □

Now we are in the position to prove the main result of this subsection, which says that $\mathcal{W}_1(\mathbb{R})$ is isometrically rigid.

Theorem 3.7. Let $\varphi\colon \mathcal{W}_1(\mathbb{R}) \to \mathcal{W}_1(\mathbb{R})$ be an isometry, that is, a bijection satisfying

$$d_{\mathcal{W}_1}(\varphi(\mu), \varphi(\nu)) = d_{\mathcal{W}_1}(\mu, \nu) \qquad (\mu, \nu \in \mathcal{W}_1(\mathbb{R})).$$

Then $\varphi = \psi_\#$ for some $\psi \in \mathrm{Isom}(\mathbb{R})$. Therefore, we also have

$$\mathrm{Isom}(\mathcal{W}_1(\mathbb{R})) = \mathrm{Isom}(\mathbb{R}).$$

Proof. Since φ is an isometry, for every $\mu, \nu, \eta \in \mathcal{W}_1(\mathbb{R}), \mu \neq \nu$ we have

$$\eta \in M(\mu,\nu) \quad \Longleftrightarrow \quad \varphi(\eta) \in M(\varphi(\mu), \varphi(\nu)),$$

and hence also $\mathrm{diam}\,(M(\mu,\nu)) = \mathrm{diam}\,(M(\varphi(\mu),\varphi(\nu)))$. By the above claims this implies that φ preserves adjacency in both directions, and thus φ leaves $\Delta(\mathbb{R})$ invariant. Since we have $d_{\mathcal{W}_1}(\delta_x, \delta_y) = |x-y|$ $(x, y \in \mathbb{R})$, we easily obtain an isometry $\psi\colon \mathbb{R} \to \mathbb{R}$ such that

$$\varphi(\delta_x) = \delta_{\psi(x)} \qquad x \in \mathbb{R}.$$

If ψ is the identity map on $\mathbb{R}$, then by an argument similar to the one in (2.9)–(2.10) in Claim 2.4 we can conclude that φ is the identity on $\mathcal{W}_1(\mathbb{R})$. If ψ is not the identity, then we can replace φ with $\varphi\psi_\#^{-1}$, and in this case we obtain $\varphi = \psi_\#$. □

We finish this subsection with two short remarks. First, we would like to point out a somewhat surprising consequence of Theorems 2.5 and 3.7.

Corollary 3.8. Even though we have $\mathcal{W}_1([0,1]) \subset \mathcal{W}_1(\mathbb{R})$, not every isometry of $\mathcal{W}_1([0,1])$ can be extended into an isometry of $\mathcal{W}_1(\mathbb{R})$. Moreover, no subgroup of $\mathrm{Isom}\,(\mathcal{W}_1(\mathbb{R}))$ is isomorphic to $\mathrm{Isom}\,(\mathcal{W}_1([0,1]))$.

Proof. For the flip operation j we have $j(\Delta([0,1])) \not\subseteq \Delta([0,1])$, however, every isometry of $\mathcal{W}_1(\mathbb{R})$ leaves the set of all Dirac masses invariant. As for the second statement, we note that in $\mathrm{Isom}\,(\mathcal{W}_1(\mathbb{R}))$ the product of any two different elements of order two is never an element of order two, in contrast with the Klein group. □

Second, our proof of the above theorem strongly relies on the assumption that φ is bijective, since for a general $\varphi \in \mathrm{IsEmb}\,(\mathcal{W}_1(\mathbb{R}))$ we usually have $\varphi\,(M(\mu,\nu)) \subsetneq M\,(\varphi(\mu),\varphi(\nu))$. In fact, one can easily construct non-surjective examples with essentially different properties. We continue with two such examples. The first one is the translation on the space of quantile functions. These maps will be crucial in the next section, so we start by a definition for $p \geq 1$.

Definition 3.9 (Translation in $\mathcal{W}_p(\mathbb{R})$)**.** Let $p \geq 1$ and $\nu \in \mathcal{W}_p(\mathbb{R})$ be arbitrary. Then the map defined by

$$\varphi\colon \mathcal{W}_p(\mathbb{R}) \to \mathcal{W}_p(\mathbb{R}), \quad F^{-1}_{\varphi(\mu)} = F^{-1}_\mu + F^{-1}_\nu \quad (\mu \in \mathcal{W}_p(\mathbb{R}))$$

is called a *translation* by the measure ν. By (1.2), this defines an isometric embedding. Clearly, a translation is bijective if and only if $\nu \in \Delta(\mathbb{R})$.

Let us now point out that if φ is the translation by $\frac{1}{2}\delta_{-1}+\frac{1}{2}\delta_1$, then the range of φ contains only such measures whose support is never the whole $\mathbb{R}$, as the quantile function of each $\varphi(\mu)$ jumps at $\frac{1}{2}$. If φ is the translation by the uniform measure on $[0,1]$, then all the slopes of each $F^{-1}_{\varphi(\mu)}$ are at least 1. Therefore all the slopes of each $F_{\varphi(\mu)}$ must be at most 1, hence φ only contains absolutely continuous measures in its range. One could say that such a translation "smoothens out" measures.

Our second example is special in the sense that its range contains only measures which coincide on the open interval $(-1,1)$. Let $E\colon [-1,1)\to[\frac{1}{3},\frac{2}{3}]$ be an arbitrary right-continuous and monotone increasing function. We define $\varphi\colon \mathcal{W}_1(\mathbb{R})\to\mathcal{P}(\mathbb{R})$ by

$$F_{\varphi(\mu)}(x)=\begin{cases}\frac{1}{3}F_\mu(\frac{x+1}{3}) & \text{if } x<-1,\\ E(x) & \text{if } -1\le x<1,\\ \frac{2}{3}+\frac{1}{3}F_\mu(\frac{x-1}{3}) & \text{if } 1\le x.\end{cases}.$$

It is easy to see that indeed φ maps $\mathcal{W}_1(\mathbb{R})$ into $\mathcal{P}(\mathbb{R})$. Next, notice that the following holds for all $\mu,\nu\in\mathcal{W}_1(\mathbb{R})$:

$$\begin{aligned}\int_{-\infty}^{\infty}|F_{\varphi(\mu)}(x)-F_{\varphi(\nu)}(x)|\,\mathrm{d}x&=\int_{-\infty}^{-1}\tfrac{1}{3}|F_\mu(\tfrac{x+1}{3})-F_\mu(\tfrac{x+1}{3})|\,\mathrm{d}x+\int_{1}^{\infty}\tfrac{1}{3}|F_\mu(\tfrac{x-1}{3})-F_\mu(\tfrac{x-1}{3})|\,\mathrm{d}x\\&=\int_{-\infty}^{\infty}|F_\mu(x)-F_\nu(x)|\,\mathrm{d}x=d_{\mathcal{W}_1}(\mu,\nu).\end{aligned}$$

Therefore, substituting $\nu=\delta_0$ and noticing that $\varphi(\delta_0)$ is supported on $[-1,1]$ shows that φ maps $\mathcal{W}_1(\mathbb{R})$ into itself. Hence it is an isometric embedding of $\mathcal{W}_1(\mathbb{R})$.

In contrast to the above examples, one may observe the following fact which shows at least some kind of a rigidity of the Wasserstein space $\mathcal{W}_1(\mathbb{R})$.

Proposition 3.10. Let φ be an isometric embedding of $\mathcal{W}_1(\mathbb{R})$ such that $\varphi(\Delta(\mathbb{R}))\subseteq\Delta(\mathbb{R})$. Then φ is an isometry of $\mathcal{W}_1(\mathbb{R})$.

We omit the proof, as one can easily do it using the ideas of Theorem 3.7.

We have seen that $\mathrm{Isom}\,(\mathcal{W}_1(\mathbb{R}))$ and $\mathrm{Isom}\,(\mathcal{W}_2(\mathbb{R}))$ are essentially different. To get the full picture, we continue by investigating $\mathrm{Isom}\,(\mathcal{W}_p(\mathbb{R}))$ in the case of $p>1$, $p\neq 2$.

3.2. $p>1$, $p\neq 2$ – **Characterization of isometric embeddings.** Similarly to Section 2, here we are able to handle the more general case of isometric embeddings. This time, however, $\mathrm{Isom}(\mathcal{W}_p(\mathbb{R}))$ and $\mathrm{IsEmb}(\mathcal{W}_p(\mathbb{R}))$ are different. We will show that isometric embeddings are compositions of trivial isometries and translations (see Definition 3.9). In particular, it will turn out that every isometric embedding that leaves the set of all Dirac masses $\Delta(\mathbb{R})$ invariant is a trivial isometry. In this subsection it is more convenient to consider $\mathcal{W}_p(\mathbb{R})$ as a space of quantile functions, hence, as a subset of $L^p\big((0,1)\big)$. In order to achieve our goal first, we prove an abstract Mankiewicz-type lemma which ensures that every isometric embedding of $\mathcal{W}_p(\mathbb{R})$ can be extended to an isometric embedding of $L^p((0,1))$. Then we apply the Banach–Lamperti theorem which describes all linear isometric embeddings of $L^p((0,1))$ for $p>1$, $p\neq 2$, see [15, Theorem 3.1].

We call a convex subset of a real Banach space with non-empty interior a convex body. Also, when we talk about bijective distance preserving maps between *two different* metric spaces, then we will call them simply isometries. When bijectivity is not assumed, then we shall call them isometric embeddings.

Now, we state Mankiewicz's theorem.

Theorem 3.11 (Mankiewicz, [17])**.** Let X and Y be two real Banach spaces and $K\subset X$, $M\subset Y$ be convex bodies. Then every isometry $\phi\colon K\to M$ can be extended to an (affine) isometry $\Phi\colon X\to Y$.

We note that every isometric embedding of a Banach space X into a strictly convex Banach space Y is automatically affine (linear up to translation). Indeed, in this case the strict triangle inequality holds in Y, hence the midpoint y of any two points $y_1,y_2\in Y$ is characterized by

$$\|y-y_1\|=\|y-y_2\|=\tfrac{1}{2}\|y_1-y_2\|.$$

Also, the Mazur–Ulam theorem ensures that all isometries between two Banach spaces are affine, although note that this statement fails for isometric embeddings in general.

The reason why we cannot apply Mankiewicz's theorem directly is that although $\mathcal{W}_p(\mathbb{R})$ is a convex and closed subset of $L^p((0,1))$, its interior is empty. However, since $L^p((0,1))$ is a strictly convex Banach space, we can overcome this obstacle with the forthcoming lemma. The linear span of a set $K \subseteq X$ will be denoted by $\operatorname{linspan}(K)$, and its closure by $(\operatorname{linspan}(K))^-$.

Lemma 3.12. Let X be a real, strictly convex Banach space and $K \subset X$ be a convex set (with possibly empty interior) such that $0 \in K$ and $(\operatorname{linspan}(K))^- = X$. Then every isometric embedding $\varphi\colon K \to X$ with $\varphi(0) = 0$ can be uniquely extended to a (linear) isometric embedding $L\colon X \to X$.

Proof. We only need to extend φ to the dense subspace $\operatorname{linspan} K$, as from there extending to the whole space is straightforward by a simple continuity and completeness argument. Let us define a set of finite dimensional subspaces, where $\operatorname{Lat}(X)$ denotes the lattice of all subspaces of X:

$$\mathcal{S} := \{M \in \operatorname{Lat}(X) \mid \dim M < \infty, M \cap K \text{ is a convex body in } M\}.$$

Now, we prove that

$$\operatorname{linspan} K = \cup\{M \mid M \in \mathcal{S}\}.$$

Indeed if $x = \sum_{j=1}^m a_j x_j,\ a_j \in \mathbb{R}, x_j \in K$, then

$$x \in M := \operatorname{linspan}\{x_1, \dots x_m\} = \operatorname{linspan}\{x_{i_1}, \dots x_{i_k}\}$$

where $1 \le i_1 < i_2 < \cdots < i_k \le m$ and the system $x_{i_1}, \dots x_{i_k} \in K$ is a base in M. Obviously, the simplex spanned by 0 and this system is a convex body in M, hence so is $K \cap M$. Therefore we obtain $\operatorname{linspan} K \subseteq \cup\{M \mid M \in \mathcal{S}\}$. For the reverse, let $M \in \mathcal{S}$ with $\dim M = m$, then by definition there must exist $m+1$ affine independent points $x_0, x_1, \dots, x_m$ in $K \cap M$. Clearly, then $\{x_i - x_0\}_{i=1}^m$ is a base in M, and hence $M \subseteq \operatorname{linspan} K$.

Next, we show that for every $M \in \mathcal{S}$ there exists a unique linear extension of $\varphi|_{K\cap M}$ to M, which also happens to be an isometric embedding. By strict convexity, φ is an affine map, which also fixes 0. Therefore the restriction $\varphi|_{K\cap M}$ can be extended to a unique injective linear map

$$L_M\colon M \to \operatorname{linspan}(\varphi(K \cap M)).$$

Clearly, $\varphi(K \cap M)$ is a convex body in $\operatorname{linspan}(\varphi(K \cap M))$, thus by Mankiewicz's theorem L_M must be an isometry too. Also, note that by construction we have

$$M, N \in \mathcal{S}, M \subseteq N \implies L_N|_M = L_M. \tag{3.5}$$

Now, we have an extension for every $M \in \mathcal{S}$, and our goal is to show that if $M, N \in \mathcal{S}$, then $L_M|_{M\cap N} = L_N|_{M\cap N}$. (However, we point out that $M, N \in \mathcal{S}$ does not imply $M \cap N \in \mathcal{S}$ in general.) This will show that the following map is a well-defined, linear, distance-preserving extension of φ:

$$L\colon \operatorname{linspan} K \to X,\ Lx = L_M x \text{ where } x \in M \in \mathcal{S},$$

and thus the proof will be complete. For this observe that

$$M, N \in \mathcal{S} \implies M + N \in \mathcal{S}.$$

Indeed, assume indirectly that $M + N \notin \mathcal{S}$, thus $(M + N) \cap K$ is not a convex body in $M + N$. This also means that $(M + N) \cap K$ spans an affine subspace E of $M + N$ with co-dimension at least 1. As $0 \in K$, the affine subspace E is a linear subspace. But, as both $M \cap K$ and $N \cap K \subset (M + N) \cap K \subset E$ and they are convex bodies in M and N, respectively, we get that $M, N \subset E$ and hence $M + N \subseteq E$, a contradiction. Therefore by (3.5), for every $M, N \in \mathcal{S}$ we obtain

$$L_M|_{M\cap N} = L_{M+N}|_{M\cap N} = L_N|_{M\cap N},$$

which completes the proof. □

As we mentioned earlier, $\mathcal{W}_p(\mathbb{R})$ is a convex and closed subset of $L^p((0,1))$, where we regard elements of $\mathcal{W}_p(\mathbb{R})$ as quantile functions. Let us point out that $\operatorname{linspan}(\mathcal{W}_p(\mathbb{R}))$ is dense in $L^p((0,1))$, since the functions $t \mapsto t^n$, $n \in \mathbb{N}$, are elements of $\mathcal{W}_p(\mathbb{R})$. Therefore, by the above lemma, if $\varphi \in \operatorname{IsEmb}(\mathcal{W}_p(\mathbb{R}))$ and φ fixes δ_0, then φ can be extended to a linear isometric embedding of $L^p((0,1))$. The latter have been characterized in [15, Theorem 3.1], which we recall now.

Definition 3.13 (Regular set-isomorphism)**.** Let $((0,1),\mathcal{L}_{(0,1)},\lambda)$ be the measure space where $\mathcal{L}_{(0,1)}$ stands for the σ-algebra of all Lebesgue sets of $(0,1)$ and λ is the normalized Lebesgue measure. We call a map $T\colon \mathcal{L}_{(0,1)} \to \mathcal{L}_{(0,1)}$, defined modulo sets of measure zero, a *regular set-isomorphism* if the following conditions hold:

(a) $T((0,1)\setminus A) = T((0,1)) \setminus T(A)$ for all Lebesgue sets $A \subseteq (0,1)$,
(b) $T\left(\cup_{n=1}^{\infty} A_n\right) = \cup_{n=1}^{\infty} T(A_n)$ for disjoint Lebesgue sets $A_n \subseteq (0,1)$,
(c) for each Lebesgue set $A \subseteq (0,1)$, we have $\lambda(T(A)) = 0$ if and only if $\lambda(A) = 0$.

A regular set-isomorphism induces a linear transformation on the set of all Lebesgue-measurable functions, which is also denoted by T, and which is characterized by $T\chi_A = \chi_{T(A)}$ where χ_A denotes the characteristic function of a Lebesgue set A (see [15] for more details).

We note that a regular set-isomorphism does not need to be bijective, for instance the map $T(A) := \frac{1}{2}A = \left\{\frac{1}{2}x \mid x \in A\right\}$ defines a non-bijective one. In particular, as can be seen from the Banach–Lamperti theorem below, a typical linear isometric embedding of $L^p((0,1))$ is in fact not bijective. From now on, we will denote the constant function with value 1 by $\mathbf{1}$.

Theorem 3.14 (Banach–Lamperti)**.** Let $1 \le p < \infty$, $p \neq 2$ be a fixed parameter, and assume that $U\colon L^p((0,1)) \to L^p((0,1))$ is a linear isometric embedding. Then there exists a regular set-isomorphism T of the measure space $((0,1),\mathcal{L}_{(0,1)},\lambda)$ such that

$$(Uf)(x) = h(x)\cdot(Tf)(x) \qquad (\text{a.e. } x \in (0,1)),$$

where $h = U\chi_{(0,1)} = U\mathbf{1} = UF_{\delta_1}^{-1} \in L^p((0,1))$.

In [15] the reader can find a more general statement that holds for all σ-finite measure spaces, and which includes a converse statement too. However, we shall only need the above very special version. Note that even though $h = U\mathbf{1}$ is not explicitly stated in [15, Theorem 3.1], it can be found in its proof in case when the measure space is finite. Before we can apply the Banach–Lamperti theorem, we have to show that the study of a general isometric embeddings of $\mathcal{W}_p(\mathbb{R})$, $1 < p < \infty$, can be reduced to the study of those isometric embeddings of $\mathcal{W}_p(\mathbb{R})$ that fix δ_0. This is what we do in the next lemma.

Lemma 3.15. Assume that $1 < p < \infty$ and $\varphi\colon \mathcal{W}_p(\mathbb{R}) \to \mathcal{W}_p(\mathbb{R})$ is an isometric embedding, that is,

$$\left\|F_{\varphi(\mu)}^{-1} - F_{\varphi(\nu)}^{-1}\right\|_p = \left\|F_{\mu}^{-1} - F_{\nu}^{-1}\right\|_p \quad (\mu,\nu \in \mathcal{W}_p(\mathbb{R})).$$

Then either

$$F_{\varphi(\delta_t)}^{-1} = F_{\varphi(\delta_0)}^{-1} + t\cdot\mathbf{1} \qquad (t \in \mathbb{R}),$$

or

$$F_{\varphi(\delta_t)}^{-1} = F_{\varphi(\delta_0)}^{-1} - t\cdot\mathbf{1} \qquad (t \in \mathbb{R}).$$

Moreover, we have

$$F_{\varphi(\mu)}^{-1} - F_{\varphi(\delta_0)}^{-1} \in \mathcal{W}_p(\mathbb{R}) \qquad (\mu \in \mathcal{W}_p(\mathbb{R})). \tag{3.6}$$

In particular, the mapping $\tilde{\varphi}\colon \mathcal{W}_p(\mathbb{R}) \to \mathcal{W}_p(\mathbb{R})$ defined by

$$F_{\tilde{\varphi}(\mu)}^{-1} := F_{\varphi(\mu)}^{-1} - F_{\varphi(\delta_0)}^{-1} \qquad (\mu \in \mathcal{W}_p(\mathbb{R})) \tag{3.7}$$

is a well-defined isometric embedding such that either $\tilde{\varphi}(\delta_t) = \delta_t$ for all $t \in \mathbb{R}$, or $\tilde{\varphi}(\delta_t) = \delta_{-t}$ for all $t \in \mathbb{R}$.

Proof. First, we show that $F_{\varphi(\delta_1)}^{-1} - F_{\varphi(\delta_0)}^{-1} \in \{-\mathbf{1},\mathbf{1}\}$. By strict convexity of the norm, we have $F_{\varphi(\delta_x)}^{-1}(t) = (1-x)\cdot F_{\varphi(\delta_0)}^{-1}(t) + x\cdot F_{\varphi(\delta_1)}^{-1}(t)$ for all $x \in \mathbb{R}$ and $0 < t < 1$. Thus for all $x \in \mathbb{R}$ and $0 < t_1 < t_2 < 1$ we obtain

$$0 \le \frac{F_{\varphi(\delta_x)}^{-1}(t_1) - F_{\varphi(\delta_x)}^{-1}(t_2)}{t_1 - t_2} = (1-x)\cdot\frac{F_{\varphi(\delta_0)}^{-1}(t_1) - F_{\varphi(\delta_0)}^{-1}(t_2)}{t_1 - t_2} + x\cdot\frac{F_{\varphi(\delta_1)}^{-1}(t_1) - F_{\varphi(\delta_1)}^{-1}(t_2)}{t_1 - t_2}.$$

But notice that this happens if and only if all the slopes $\frac{F_{\varphi(\delta_0)}^{-1}(t_1) - F_{\varphi(\delta_0)}^{-1}(t_2)}{t_1 - t_2}$ and $\frac{F_{\varphi(\delta_1)}^{-1}(t_1) - F_{\varphi(\delta_1)}^{-1}(t_2)}{t_1 - t_2}$ coincide, or in other words, $F_{\varphi(\delta_1)}^{-1} - F_{\varphi(\delta_0)}^{-1}$ is constant on $(0,1)$. Taking into account the distances,

$F^{-1}_{\varphi(\delta_1)} - F^{-1}_{\varphi(\delta_0)} \in \{\mathbf{1}, -\mathbf{1}\}$ follows. By strict convexity we then easily get the first statement of the lemma.

Next, let $\mu, \nu \in \mathcal{W}_p(\mathbb{R})$ be arbitrary, and let us define the set:

$$\begin{aligned} I_{\mu,\nu} :&= \left\{ s \in \mathbb{R} \mid (1-s)\cdot F_\mu^{-1} + s \cdot F_\nu^{-1} \text{ is monotone increasing} \right\} \\ &= \left\{ s \in \mathbb{R} \mid \exists\, \eta \in \mathcal{W}_p(\mathbb{R})\colon (1-s)\cdot F_\mu^{-1} + s \cdot F_\nu^{-1} = F_\eta^{-1} \right\}. \end{aligned}$$

Clearly, the set $I_{\mu,\nu}$ is always a closed interval that contains $[0,1]$. Now, we observe that for any $\mu, \nu \in \mathcal{W}_p(\mathbb{R})$ we have $[0,\infty) \subseteq I_{\mu,\nu}$ if and only if

$$\frac{F_\mu^{-1}(t_1) - F_\mu^{-1}(t_2)}{t_1 - t_2} \leq \frac{F_\nu^{-1}(t_1) - F_\nu^{-1}(t_2)}{t_1 - t_2} \qquad (0 < t_1 < t_2 < 1). \tag{3.8}$$

In particular, we always have $[0,\infty) \subseteq I_{\delta_0,\mu}$, thus also $[0,\infty) \subseteq I_{\varphi(\delta_0),\varphi(\mu)}$ for all $\mu \in \mathcal{W}_p(\mathbb{R})$. Therefore applying (3.8) with $\varphi(\delta_0)$ and $\varphi(\mu)$, we obtain (3.6) and the rest of the statement follows easily. □

Now, we are in the position to prove our theorem on $\mathrm{IsEmb}(\mathcal{W}_p(\mathbb{R}))$ for $p > 1, p \neq 2$, using the Banach–Lamperti theorem and our Mankiewicz-type extension lemma.

Theorem 3.16. Let $1 < p < \infty, p \neq 2$ and $\varphi \in \mathrm{IsEmb}(\mathcal{W}_p(\mathbb{R}))$, that is

$$d_{\mathcal{W}_p}(\varphi(\mu), \varphi(\nu)) = d_{\mathcal{W}_p}(\mu, \nu) \qquad (\mu, \nu \in \mathcal{W}_p(\mathbb{R})).$$

Then φ is a composition of a trivial isometry and a translation, that is, there exists a $\psi \in \mathrm{Isom}(\mathbb{R})$ and $\nu \in \mathcal{W}_p(\mathbb{R})$ such that

$$F^{-1}_{\varphi(\mu)} = F^{-1}_{\psi_\#(\mu)} + F^{-1}_\nu \qquad (\mu \in \mathcal{W}_p(\mathbb{R})). \tag{3.9}$$

In particular, if φ is also bijective, then it is a trivial isometry, therefore we have

$$\mathrm{Isom}(\mathbb{R}) = \mathrm{Isom}(\mathcal{W}_p(\mathbb{R})) \subsetneq \mathrm{IsEmb}(\mathcal{W}_p(\mathbb{R})).$$

Proof. Consider the mapping $\tilde{\varphi} \in \mathrm{IsEmb}(\mathcal{W}_p(\mathbb{R}))$ defined in (3.7), which either fixes all Dirac measures, or $\tilde{\varphi}(\delta_x) = \delta_{-x}$ for all $x \in \mathbb{R}$. Therefore, it is enough to show that if an isometric embedding $\tilde{\varphi}$ fixes all Dirac measures, then it fixes every measure in $\mathcal{W}_p(\mathbb{R})$.

By Lemma 3.12, we conclude that the map $\tilde{\varphi}$ is a restriction of a linear isometric embedding $U\colon L^p((0,1)) \to L^p((0,1))$, which fixes all constant functions. Hence, in the Banach–Lamperti theorem we have $h(x) = 1$ for a.e. x, therefore

$$Uf(x) = Tf(x) \qquad (\text{a.e. } x \in (0,1)) \tag{3.10}$$

where T is the linear operator generated by a regular set-isomorphism T (which is defined modulo null-sets). Substituting

$$f = \chi_{[a,1)} = F^{-1}_{a\delta_0 + (1-a)\delta_1} \quad \text{and} \quad f = -\chi_{(0,a)} = F^{-1}_{a\delta_{-1} + (1-a)\delta_0}$$

into (3.10) gives

$$\chi_{T([a,1))}, -\chi_{T((0,a))} \in \mathcal{W}_p(\mathbb{R}) \subset L^p((0,1)) \qquad (0 < a < 1).$$

Therefore, by the properties of the regular set-isomorphism T, for every $0 < a < 1$ there exists a $0 < t_a < 1$ such that $T([a,1)) = [t_a, 1)$ and $T((0,a)) = (0, t_a)$. Hence we obtain

$$U\chi_{[a,1)} = \chi_{[t_a,1)} \text{ and } U\chi_{(0,a)} = \chi_{(0,t_a)} \qquad (0 < a < 1),$$

but since U preserves the p-norm, the number t_a must coincide with a for all $0 < a < 1$. Thus we get that U is the identity operator, and the proof of (3.9) is complete.

Finally, note that if φ is also assumed to be bijective, then F_ν^{-1} must be a constant function. Indeed, otherwise there would exist two points $0 < t_1 < t_2 < 1$ such that

$$\frac{F^{-1}_{\varphi(\mu)}(t_1) - F^{-1}_{\varphi(\mu)}(t_2)}{t_1 - t_2} \geq \frac{F_\nu^{-1}(t_1) - F_\nu^{-1}(t_2)}{t_1 - t_2} > 0 \quad (\mu \in \mathcal{W}_p(\mathbb{R})),$$

and therefore Dirac masses would not be in the range of φ. □

3.3. $p = 2$ – **A functional analytic description of the exotic flow.** The aim of this subsection is to take a closer look at Kloeckner's surprising result on $\mathrm{Isom}\,(\mathcal{W}_2(\mathbb{R}))$. We introduce the notation $m(\mu)$ for the center of mass of a $\mu \in \mathcal{W}_1(\mathbb{R})$:

$$m(\mu) = \int_0^1 F_\mu^{-1}(x)\ \mathrm{d}x.$$

The map $r_c\colon \mathbb{R} \to \mathbb{R}, x \mapsto 2c - x$ is called the reflection through $c \in \mathbb{R}$.

Kloeckner showed in [12, Theorem 1.1] that the group $\mathrm{Isom}(\mathcal{W}_2(\mathbb{R}))$ is the semidirect product $\mathrm{Isom}(\mathbb{R}) \ltimes \mathrm{Isom}(\mathbb{R})$. Namely, he showed that every isometry of $\mathcal{W}_2(\mathbb{R})$ is a composition of some of the of the following maps:

(1) a trivial isometry, that is, $\psi_\#$ for some $\psi \in \mathrm{Isom}(\mathbb{R})$;
(2) the map $\mu \mapsto (r_{m(\mu)})_\#(\mu)$, that is, the isometry that reflects every measure through its center of mass; and
(3) a so-called exotic isometry Φ^q for some $q \in \mathbb{R}$, which we focus on in this subsection, see the definition in (3.11) and (3.12).

Note that all of the above types of isometries leave the set of all Dirac measures invariant, moreover, (2)–(3) fix every Dirac measures. Also, in cases (1)–(2) the support of each measure μ is isometrically congruent to the support of its image. However, this is not the case for exotic isometries.

The set of all measures which are supported on at most two points will be denoted by $\Delta_2(\mathbb{R})$. As in [12], we parametrize $\Delta_2(\mathbb{R})$ by $x, p \in \mathbb{R}$, $\sigma \geq 0$ as

$$\mu(x, \sigma, p) := \frac{e^{-p}}{e^p + e^{-p}} \cdot \delta_{x - \sigma e^p} + \frac{e^p}{e^p + e^{-p}} \cdot \delta_{x + \sigma e^{-p}}. \tag{3.11}$$

Let $q \in \mathbb{R}$ be fixed. Using the above parametrization, Kloeckner defined the exotic isometry Φ^q on $\Delta_2(\mathbb{R})$ in the following way:

$$\Phi^q\left(\mu(x, \sigma, p)\right) := \mu(x, \sigma, p + q) \quad (x, \sigma, p \in \mathbb{R}, \sigma \geq 0). \tag{3.12}$$

He proved that this indeed defines an isometry on $\Delta_2(\mathbb{R})$ and that it extends uniquely to an isometry of $\mathcal{W}_2(\mathbb{R})$. He also pointed out that even though the above definition is constructive, it is not very explicit outside $\Delta_2(\mathbb{R})$. Moreover, he illustrated that an explicit formula for general measures supported on three points already seems to be very complicated.

The goal of this subsection is to provide a general explicit formula for the action of exotic isometries, which we shall do by using functional analytic techniques rather than geometric ones. Namely, we prove that if we regard elements of $\mathcal{W}_2(\mathbb{R})$ as quantile functions, then Φ^q extends to a real unitary operator $U_q : L^2((0,1)) \to L^2((0,1))$ which can be written in terms of a composition operator, the Volterra operator, a multiplication operator, and a rank-one projection. First we state a well-known lemma (see for instance [26, Theorem 11.4]) that will be helpful in our considerations.

Lemma 3.17. Let H be a real Hilbert space and S be a subset such that $0 \in S$ and $\mathrm{linspan}\, S$ is dense in H. If $\varphi\colon S \to H$ is an isometric embedding such that $\varphi(0) = 0$, then it can be uniquely extended to a (linear) isometric embedding $L\colon H \to H$.

Now our theorem introduced above reads as follows.

Theorem 3.18. Let q be a real number. Then the action of the exotic isometry Φ^q is given by the following formula:

$$\begin{aligned} F^{-1}_{\Phi^q(\mu)}(x) = (1 - e^q) \cdot m(\mu) + \left\{e^q + (e^{-q} - e^q) h_q(x)\right\} \cdot F_\mu^{-1}(h_q(x)) \\ + (e^q - e^{-q}) \cdot \int_0^{h_q(x)} F_\mu^{-1}(s)\ \mathrm{d}s \qquad (\mu \in \mathcal{W}_2(\mathbb{R}), 0 < x < 1), \end{aligned} \tag{3.13}$$

where

$$h_q(x) = \frac{x e^{2q}}{1 + (e^{2q} - 1)x} \qquad (x \in (0,1)). \tag{3.14}$$

Proof. Again, we will regard the Wasserstein space $\mathcal{W}_2(\mathbb{R})$ as a convex and closed subset of $L^2((0,1))$ whose linear span is dense in $\mathcal{W}_2(\mathbb{R})$. Therefore by Lemma 3.17 and (3.12) the exotic isometry Φ^q can be extended to a unique linear isometric embedding which we denote by U_q. Let us point out that since $\Delta_2(\mathbb{R})$ is such a subset of $L^2((0,1))$ whose linear span is dense, therefore U_q is the unique bounded linear operator on $L^2((0,1))$ which satisfies

$$U_q\left(F^{-1}_{\mu(x,\sigma,p)}\right) = F^{-1}_{\mu(x,\sigma,p+q)} \qquad (x,\sigma,p\in\mathbb{R},\sigma\geq 0). \tag{3.15}$$

Therefore, it is enough to find a bounded linear operator (without proving its isometric property) that satisfies (3.15). Also, observe that (3.15) is equivalent to $U_q\mathbf{1}=\mathbf{1}$, and $U_q\left(F^{-1}_{\mu_p}\right) = F^{-1}_{\mu_{p+q}}$ $(p\in\mathbb{R})$, where we used the shorthand $\mu_p := \mu(0,1,p)$. Define the operator

$$T_q\colon L^2((0,1))\to L^2((0,1)),\quad T_q = U_q - \mathbf{1}\otimes\mathbf{1},$$

where $(\mathbf{1}\otimes\mathbf{1})f = \int_0^1 f(s)\,\mathrm{d}s\cdot\mathbf{1}$ is the rank-one projection onto the subspace of all constant functions. Next, we describe how T_q acts on certain characteristic functions. Since for all $p\in\mathbb{R}$

$$F^{-1}_{\mu_p} = -e^p\cdot\chi_{\left(0,\frac{e^{-p}}{e^{-p}+e^p}\right)} + e^{-p}\cdot\chi_{\left[\frac{e^{-p}}{e^{-p}+e^p},1\right)}$$

holds, we calculate

$$\begin{aligned} T_q\chi_{\left(0,\frac{e^{-p}}{e^{-p}+e^p}\right)} &= \tfrac{1}{e^{-p}+e^p}\cdot U_q\left(e^{-p}\cdot\mathbf{1}-F^{-1}_{\mu_p}\right) - \tfrac{e^{-p}}{e^{-p}+e^p}\cdot\mathbf{1} = \tfrac{-1}{e^{-p}+e^p}\cdot F^{-1}_{\mu_{p+q}} \\ &= \tfrac{e^{p+q}}{e^{-p}+e^p}\cdot\chi_{\left(0,\frac{e^{-p-q}}{e^{-p-q}+e^{p+q}}\right)} - \tfrac{e^{-p-q}}{e^{-p}+e^p}\cdot\chi_{\left[\frac{e^{-p-q}}{e^{-p-q}+e^{p+q}},1\right)}. \end{aligned}$$

Notice that since $T_q\mathbf{1} = 0\cdot\mathbf{1}$, we have

$$T_q\chi_{\left[\frac{e^{-p}}{e^{-p}+e^p},1\right)} = -\tfrac{e^{p+q}}{e^{-p}+e^p}\cdot\chi_{\left(0,\frac{e^{-p-q}}{e^{-p-q}+e^{p+q}}\right)} + \tfrac{e^{-p-q}}{e^{-p}+e^p}\cdot\chi_{\left[\frac{e^{-p-q}}{e^{-p-q}+e^{p+q}},1\right)}.$$

Now, we define a set and two transformations on it. Let

$$\mathcal{I} := \left\{\chi_{(0,t)}\,\middle|\,0<t<1\right\}\cup\left\{\chi_{[t,1)}\,\middle|\,0<t<1\right\},$$

$$\begin{aligned} T^{(0)}_q\colon\mathcal{I}\to L^2((0,1)),\quad T^{(0)}_q\left(\chi_{\left(0,\frac{e^{-p}}{e^{-p}+e^p}\right)}\right) &= \tfrac{e^p}{e^{-p}+e^p}\cdot\chi_{\left(0,\frac{e^{-p-q}}{e^{-p-q}+e^{p+q}}\right)} \\ T^{(0)}_q\left(\chi_{[t,1)}\right) &= -T^{(0)}_q\left(\chi_{(0,t)}\right)\quad (p\in\mathbb{R},0<t<1) \end{aligned}$$

and

$$\begin{aligned} T^{(1)}_q\colon\mathcal{I}\to L^2((0,1)),\quad T^{(1)}_q\left(\chi_{\left[\frac{e^{-p}}{e^{-p}+e^p},1\right)}\right) &= \tfrac{e^{-p}}{e^{-p}+e^p}\cdot\chi_{\left[\frac{e^{-p-q}}{e^{-p-q}+e^{p+q}},1\right)} \\ T^{(1)}_q\left(\chi_{(0,t)}\right) &= -T^{(1)}_q\left(\chi_{[t,1)}\right)\quad (p\in\mathbb{R},0<t<1). \end{aligned}$$

Observe that $T_q|_{\mathcal{I}} = e^q\cdot T^{(0)}_q + e^{-q}\cdot T^{(1)}_q$. However, at this point we do not know if $T^{(0)}_q$ or $T^{(1)}_q$ can be extended linearly and continuously to the whole space, thus we cannot treat them as operators. A calculation gives

$$T^{(0)}_q\left(\chi_{(0,t)}\right) = C_q\circ T^{(0)}_0\left(\chi_{(0,t)}\right)\quad\text{and}\quad T^{(0)}_q\left(\chi_{[t,1)}\right) = C_q\circ T^{(0)}_0\left(\chi_{[t,1)}\right)$$

for all $0<t<1$ where

$$C_q\colon L^2((0,1))\to L^2((0,1)),\quad (C_qf)(x) = f(h_q(x))\quad (x\in(0,1))$$

is a composition operator with symbol h_q, see (3.14). Notice that C_q is a bounded operator, as h_q maps $[0,1]$ bijectively onto itself, it is a smooth function on a neighbourhood of $[0,1]$, and its derivative is bounded from below by $e^{-2|q|}$ on $[0,1]$.

Next, let $M_{\mathbf{1}-\mathbf{x}}$ stand for the multiplication operator by the function $\mathbf{1}-\mathbf{x}$ where $\mathbf{x}(t)=t$, and V for the Volterra operator: $(Vf)(t)=\int_0^t f(s)\,\mathrm{d}s$ $(t\in(0,1))$. We notice that

$$T_0^{(0)}\left(\chi_{(0,t)}\right)=(1-t)\cdot\chi_{(0,t)}=(M_{\mathbf{1}-\mathbf{x}}-\mathbf{1}\otimes\mathbf{1}+V)\chi_{(0,t)}\qquad(0<t<1).$$

Furthermore, since

$$(M_{\mathbf{1}-\mathbf{x}}-\mathbf{1}\otimes\mathbf{1}+V)\left(\chi_{(0,t)}+\chi_{[t,1)}\right)=(M_{\mathbf{1}-\mathbf{x}}-\mathbf{1}\otimes\mathbf{1}+V)\mathbf{1}=0\cdot\mathbf{1},$$

we also have $T_0^{(0)}\left(\chi_{[t,1)}\right)=(M_{\mathbf{1}-\mathbf{x}}-\mathbf{1}\otimes\mathbf{1}+V)\chi_{[t,1)}$ for all $0<t<1$. Therefore, we obtain that

$$T_0^{(0)}=(M_{\mathbf{1}-\mathbf{x}}-\mathbf{1}\otimes\mathbf{1}+V)|_{\mathcal{I}}.$$

As for $T_q^{(1)}$, we calculate and notice the following for all $0<t<1$:

$$T_q^{(1)}\left(\chi_{(0,t)}\right)=C_q\circ T_0^{(1)}\left(\chi_{(0,t)}\right),\quad T_q^{(1)}\left(\chi_{[t,1)}\right)=C_q\circ T_0^{(1)}\left(\chi_{[t,1)}\right)$$

and

$$\begin{aligned}T_0^{(1)}\left(\chi_{(0,t)}\right)&=-t\cdot\chi_{[t,1)}=-t\cdot\mathbf{1}+t\cdot\chi_{(0,t)}\\&=\left(-\mathbf{1}\otimes\mathbf{1}+I-T_0^{(0)}\right)\chi_{(0,t)}=(M_{\mathbf{x}}-V)\,\chi_{(0,t)}.\end{aligned}$$

As $(M_{\mathbf{x}}-V)\,\mathbf{1}=0\cdot\mathbf{1}$, we conclude that

$$T_0^{(1)}=(M_{\mathbf{x}}-V)\,|_{\mathcal{I}}.$$

Therefore, the above observations together imply

$$\begin{aligned}U_q|_{\mathcal{I}}&=(T_q+\mathbf{1}\otimes\mathbf{1})|_{\mathcal{I}}=e^q\cdot T_q^{(0)}+e^{-q}\cdot T_q^{(1)}+(\mathbf{1}\otimes\mathbf{1})|_{\mathcal{I}}\\&=e^q\cdot C_q\circ T_0^{(0)}+e^{-q}\cdot C_q\circ T_0^{(1)}+(\mathbf{1}\otimes\mathbf{1})|_{\mathcal{I}}\\&=\left(e^q\cdot C_q\cdot(M_{\mathbf{1}-\mathbf{x}}-\mathbf{1}\otimes\mathbf{1}+V)+e^{-q}\cdot C_q\cdot(M_{\mathbf{x}}-V)+\mathbf{1}\otimes\mathbf{1}\right)\big|_{\mathcal{I}}.\end{aligned}$$

Thus we eventually conclude

$$U_q=C_q\cdot\left[(1-e^q)\cdot(\mathbf{1}\otimes\mathbf{1})+e^q\cdot I+(e^{-q}-e^q)\cdot M_{\mathbf{x}}+(e^q-e^{-q})\cdot V\right],$$

which implies (3.13), at least for almost every $x\in(0,1)$. However, since two right-continuous functions are equal almost everywhere on $(0,1)$ if and only if they coincide on $(0,1)$, we easily conclude (3.13). □

Acknowledgements

This paper is based on discussions made during research visits at the Institute of Science and Technology (IST) Austria, Klosterneuburg. We thank the warm hospitality of the Erdős group. We are also grateful to Lajos Molnár for his comments on earlier versions of the manuscript, and to László Erdős for his suggestions on the structure and highlights of this paper.

References

[1] M. Arjovsky, S. Chintala, L. Bottou, *Wasserstein Generative Adversarial Networks*, Proccedings of Machine Learning Research (2017), 214–223.
[2] J. Bertrand, and B. Kloeckner, *A geometric study of Wasserstein spaces: Hadamard spaces* J. Topol. Anal. 4(4) (2012), 515–542.
[3] J. Bertrand, and B. Kloeckner, *A geometric study of Wasserstein spaces: isometric rigidity in negative curvature,* Int. Math. Res. Notices 2016 (5), 1368–1386.
[4] E. Boissard, T. Le Gouic, J.-M. Loubes, *Distribution's template estimate with Wasserstein*, Bernoulli 21(2) 2015, 740–759.
[5] O. Butkovsky, *Subgeometric rates of convergence of Markov processes in the Wasserstein metric,* Annals of Appl. Prob 24(2) (2014), 526–552.
[6] G. Dolinar, and L. Molnár, *Isometries of the space of distribution functions with respect to the Kolmogorov–Smirnov metric,* J. Math. Anal. Appl. **348** (2008), 494–498.
[7] Gy. P. Gehér, *Surjective Kuiper isometries*, Houston J. Math. 44 (2018), 263–281.
[8] Gy. P. Gehér, and T. Titkos, *A characterisation of isometries with respect to the Lévy-Prokhorov metric,* Annali della Scuola Normale Superiore di Pisa - Classe di Scienze, Vol. XIX (2019), 655–677.

[9] Gy. P. Gehér, T. Titkos, D. Virosztek, *On isometric embeddings of Wasserstein spaces – the discrete case*, J. Math. Anal. Appl., to appear, arXiv: 1809.01101
[10] M. Hairer, J.C. Mattingly, M Scheutzow, *Asymptotic coupling and a general form of Harris'theorem with applications to stochastic delay equations* Probab. Theory Related Fields 149 (2011), 223–259.
[11] M. Hairer, J.C. Mattingly, *Spectral gaps in Wasserstein distances and the 2D stochastic Navier–Stokes equations* Ann. Probab. 36(6) (2008), 2050–2091.
[12] B. Kloeckner, *A geometric study of Wasserstein spaces: Euclidean spaces,* Annali della Scuola Normale Superiore di Pisa - Classe di Scienze IX, 2 (2010), 297–323.
[13] B. Kloeckner, *A generalization of Hausdorff dimension applied to Hilbert cubes and Wasserstein spaces,* J. Topol. Anal. 4(2) (2012), 203–235.
[14] B. Kloeckner, *A geometric study of Wasserstein spaces: Ultrametrics* Mathematika 61 (2015), 162–178.
[15] J. Lamperti, *On the isometries of certain function-spaces* Pacific J. Math. 8 (1958) 459–466.
[16] J. Lott, C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. 169 (2009), 903–991.
[17] P. Mankiewicz, *On extension of isometries in normed linear spaces*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. 20 (1972), 367–371.
[18] L. Molnár, *Lévy isometries of the space of probability distribution functions,* J. Math. Anal. Appl. 380 (2011), 847–852.
[19] M.-K. von Renesse and K.-T. Sturm, Transport inequalities, gradient estimates, entropy, and Ricci curvature, Comm. Pure Appl. Math. 58 (2005), no. 7, p. 923–940.
[20] S. Srivastava, C. Li, D.B. Dunson, *Scalable Bayes via Barycenter in Wasserstein Space*, Proccedings of Machine Learning Research 19(8) 2018, 1–35. 2018.
[21] K.-T. Sturm, On the geometry of metric measure spaces. I, II, Acta Math. 196 (2006), no. 1, 65–131 and 133–177.
[22] S. S. Vallender, *Calculation of the Wasserstein distance between probability distributions on the line,* Theory Probab. Appl. 18 (1973), 784–786.
[23] C. Villani, *Optimal Transport, Old and New,* Springer, 2009.
[24] C. Villani, *Topics in optimal transportation,* Graduate studies in Mathematics vol. 58, American Mathemtical Society, Providence, RI, 2003.
[25] D. Virosztek, *Maps on probability measures preserving certain distances — a survey and some new results,* Acta Sci. Math. (Szeged) 84 (2018), 65–80.
[26] J.H. Wells, L.R. Williams, Embeddings and extensions in analysis, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 84. Springer-Verlag, New York-Heidelberg, 1975.

György Pál Gehér, Department of Mathematics and Statistics, University of Reading, Whiteknights, P.O. Box 220, Reading RG6 6AX, United Kingdom
Email address: G.P.Geher@reading.ac.uk or gehergyuri@gmail.com
http://www.math.u-szeged.hu/~gehergy

Tamás Titkos, Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Reáltanoda u. 13-15., Budapest H-1053, Hungary, and BBS University of Applied Sciences, Alkotmány u. 9., Budapest H-1054, Hungary
Email address: titkos.tamas@renyi.mta.hu
http://renyi.hu/~titkos

Dániel Virosztek, Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria
Email address: daniel.virosztek@ist.ac.at
http://pub.ist.ac.at/~dviroszt