ISOMETRIC RIGIDITY OF WASSERSTEIN SPACES: 
THE GRAPH METRIC CASE

GERGELY KISS AND TAMÁS TITKOS

Abstract. The aim of this paper is to prove that the Wasserstein space $W_p(X)$ is isometrically rigid for all $p \geq 1$ whenever $X$ is a countable graph metric space. As a consequence, we obtain that for every countable group $H$ there exists a $p$-Wasserstein space whose isometry group is isomorphic to $H$.

1. Introduction

Due to its deep impact on both pure and applied sciences, one of the most intensively studied metric spaces nowadays is the so-called $p$-Wasserstein space $W_p(X)$: the collection of Borel probability measures on a complete separable metric space $(X, \varrho)$ with finite $p$-th moment, endowed with a transport related metric $d_p$ which is calculated by means of optimal couplings and the $p$-th power of the underlying distance $\varrho$. We mention here only three comprehensive textbooks [1, 22, 24], more references and precise definitions will follow later. In this paper we consider those Wasserstein spaces, whose underlying metric space $(X, \varrho)$ is a graph metric space. This class contains many important metric spaces, just to mention a few: any countable set with the discrete metric; the set of natural numbers and the set of integers with the usual $|\cdot|$-distance; $d$-dimensional lattices endowed with the $l_1$-metric or the $l_\infty$-metric (for the relevance of these metrics in pattern recognition see e.g. [21]), finite strings with the Hamming distance (as it was mentioned in [5] in connection with the quantum 1-Wasserstein distance, the classical 1-Wasserstein distance with respect to the Hamming metric is called Ornstein’s distance, and was first considered in [20]); finite regular trees (see the very recent manuscript [9]).
When working with a structure, the most fundamental and natural task is to explore its transformations and symmetries. In the case of metric spaces, such symmetries are isometries, that is, distance preserving bijections. In the recent past, many authors investigated isometries of various important metric spaces of probability measures \([2, 6, 7, 11–14, 16, 19, 23, 25]\). In [19] Molnár investigated the structure of isometries of the space of distribution functions with respect to the Lévy distance. Later Gehér and Titkos generalised his result to the Lévy-Prokhorov metric in [11]. Namely, it was shown that if the space \(\mathcal{P}(X)\) of all Borel probability measures of a real Banach space \((X, \| \cdot \|)\) is endowed with the Lévy-Prokhorov metric \(d_{\text{LP}}\), then the isometry group of \((\mathcal{P}(X), d_{\text{LP}})\) is isomorphic to the isometry group of the underlying space \(X\). Bertrand and Kloeckner showed that a similar phenomenon occurs when one considers a 2-Wasserstein space built on a negatively curved metric space: all isometries of the space of measures are push-forwards of the isometries of the underlying space. This phenomenon is called isometric rigidity. Finally, we mention a very recent rigidity result, Santos-Rodriguez proved that \(p\)-Wasserstein spaces built on compact rank one symmetric spaces are all isometrically rigid [23] if \(p > 1\).

Our main result is Theorem 1, where we prove that Wasserstein spaces over graph metric spaces are all isometrically rigid. As a consequence, in Corollary 2 we will conclude that for every countable group \(G\) and for every value of \(p \geq 1\) there exists a Wasserstein space whose isometry group is isomorphic to \(G\).

Before going into the details, we make some comments on rigidity. It comes easy to say that these rigidity results are not surprising because of the intimate connection between \(d_p\) and \(\varrho\). It is well known that if \(p \geq 1\) then the distance between any two Dirac measures equals to the distance of their supporting points, and every measure can be approximated by convex combinations of Dirac measures. In other words, \(\mathcal{W}_p(X)\) contains an isometric copy of \(X\), and the convex hull of this copy is dense in \(\mathcal{W}_p(X)\). Moreover, \(\mathcal{W}_p(X)\) inherits many nice properties such as completeness, compactness, existence of geodesics, etc from \(X\). So one may have the impression that although the Wasserstein space \(\mathcal{W}_p(X)\) is much bigger than \(X\) (see e.g. [17] for many interesting results), the strong connection between the metrics does not allow \(\mathcal{W}_p(X)\) to have more symmetries than \(X\) has. A possible sketch of proof is this:

**Step 1.** Prove that an isometry \(\Phi : \mathcal{W}_p(X) \rightarrow \mathcal{W}_p(X)\) leaves the set of Dirac masses invariant. Once it is done, one sees easily that the action on Dirac masses is generated by an isometry \(f : X \rightarrow X\), that is, \(\Phi(\delta_x) = \delta_{f(x)}\) for all \(x \in X\).

**Step 2.** Prove that this action extends to the set of finitely supported measures, that is, \(\Phi(\sum_{j \in J} \lambda_j \delta_{x_j}) = \sum_{j \in J} \lambda_j \delta_{f(x_j)}\) holds for all finite index set \(J\), \(x_j \in X\) \(\lambda_j \geq 0\) and \(\sum_{j \in J} \lambda_j = 1\).
Step 3. Since the set of finitely supported measures is dense in $\mathcal{W}_p(X)$, and $\Phi$ is continuous, $\Phi$ must be the push-forward of $f^{-1}$, where $f$ is the above defined isometry.

The problem with this sketch is that it does not work in general. And even if it works, these seemingly easy steps can be nontrivial. For example, Step 1 fails if $p = 1$ and $X = [0, 1]$. In that case, there exists an isometry $j$ (called flip) which is mass splitting, i.e. which does not leave the set of Dirac masses invariant: $j(\delta_t) = t\delta_0 + (1 - t)\delta_1$ for all $t \in [0, 1]$. For more details see [13, Section 2.1]. Step 2 can fail (even if the first step can be done) as it was shown by Kloeckner in [16]: if $p = 2$ and $X = \mathbb{R}$, then there exist a flow of strangely behaving isometries. These isometries leave all Dirac masses fixed, but they differ from the identity of $\mathcal{W}_2(\mathbb{R})$, for more details see [16, Section 5.1]. We mention that all these strange isometries disappear once we modify the value of $p$: it was proved in [13] that if $p \neq 1$ then $\mathcal{W}_p([0, 1])$ is isometrically rigid, and similarly, if $p \neq 2$ then $\mathcal{W}_p(\mathbb{R})$ is isometrically rigid. Furthermore, Gehér et al. showed in [14, Section 2] that for every $p \geq 1$ there exists a compact metric space $X$ such that $\mathcal{W}_p(X)$ admits mass splitting isometries.

Summarising the above results, we can say that isometric rigidity of Wasserstein spaces is a quite regular phenomenon (only a few non-rigid example is known), but there is no general recipe which helps to decide whether a space is rigid or not.

2. THE MAIN RESULT

First we fix the terminology. We say that a countable metric space $(X, \varrho)$ is a graph metric space if there exists a connected graph $G := G(X, E)$ with vertex set $X$ and edge set $E$, such that the edge-length of the shortest path between any two vertices $a, b \in X$ equals to $\varrho(a, b)$. Since existence of loops and multiple edges do not change the length of the shortest path, one can assume that the graph in question is simple.\footnote{The following characterization of graph metric spaces was proved in [3]: a countable metric space $(X, \varrho)$ is a graph metric space if and only if the distance between every two points of $X$ is an integer, and if $a, b \in X$ and $\varrho(a, b) \geq 2$ then there exists a point $x \in M$ such that $\varrho(a, x) > 0$, $\varrho(x, b) > 0$, and $x$ saturates the triangle inequality: $\varrho(a, b) = \varrho(a, x) + \varrho(x, b)$. It was assumed in [3] that the graph is finite, but obviously, the proof works in the countable case as well.} In this paper $p$ will always denote a fixed real number such that $p \geq 1$. The symbol $\mathcal{M}_+(X)$ stands for the set of nonnegative Borel measures on $X$. In our setting each measure $\mu \in \mathcal{M}_+(X)$ is uniquely determined by its value on singletons ($\mu(A) = \sum_{x \in A} \mu(\{x\})$ for all $A \subseteq X$) and therefore $\mu$ can be handled as a one-variable function on $X$. For the sake of simplicity we will write shortly $\mu(x)$ instead of $\mu(\{x\})$. With $\mathcal{W}_p(X)$ we denote the set of all probability measures such that

$$\sum_{x \in X} \varrho^p(x, \hat{x}) \cdot \mu(x) < \infty$$
for some (hence all) \( \hat{x} \in X \). The support \( \mu \) of a \( \mu \in \mathcal{W}_p(X) \) in this setting equals to the set \( \{ x \in X \mid \mu(x) \neq 0 \} \). A Borel probability measure \( \pi \) on \( X \times X \) is said to be a \textit{coupling} for \( \mu \) and \( \nu \) if the marginals of \( \pi \) are \( \mu \) and \( \nu \), that is,

\[
\sum_{x' \in X} \pi(x, x') = \mu(x) \quad \text{and} \quad \sum_{x' \in X} \pi(x', x) = \nu(x).
\]

The set of all couplings (which is never empty because the product measure is a coupling) is denoted by \( \Pi(\mu, \nu) \). Sometimes couplings are called transport plans, as \( \pi(x, x') \) is the weight of mass that is transported from \( x \) to \( x' \) while \( \mu \) is transported to \( \nu \) along \( \pi \). Assuming that the cost function is \( \varrho^p \), the total cost of transforming \( \mu \) into \( \nu \) along the plan \( \pi \) is

\[
C_{\varrho, p}(\pi) := \sum_{(x, y) \in X \times X} \varrho^p(x, x') \cdot \pi(x, x').
\]

The \( p \)-Wasserstein distance \( d_p \) of \( \mu, \nu \in \mathcal{W}_p(X) \) is defined as

\[
d_p(\mu, \nu) := \left( \inf_{\pi \in \Pi(\mu, \nu)} C_{\varrho, p}(\pi) \right)^{1/p}.
\]

We refer to the metric space \( (\mathcal{W}_p(X), d_p) \) as the \( p \)-Wasserstein space \( \mathcal{W}_p(X) \). It is known (see e.g. Theorem 1.5 in [1] with \( c = \varrho^p \)) that the infimum in (4) is in fact a minimum. Those transport plans that minimise the transport cost are called \textit{optimal transport plans}. Let us denote the set of all Dirac measures (or point masses) by \( \Delta(X) \), and the set of all finitely supported measures by \( \mathcal{F}(X) \). It is easy to see that \( X \) embeds into \( \mathcal{W}_p(X) \) isometrically \( (d_p(\delta_x, \delta_y) = \varrho(x, y) \text{ for all } x, y \in X) \) and that \( \mathcal{F}(X) \) is dense in \( \mathcal{W}_p(X) \). Furthermore, every isometry \( \psi : X \to X \) induces an isometry \( \psi_\# : \mathcal{W}_p(X) \to \mathcal{W}_p(X) \) by push forward

\[
(\psi_\#(\mu))(x) := \mu(\psi^{-1}(x)) \quad (x \in X)
\]

Isometries of the form \( \psi_\# \) are called \textit{trivial isometries}. We say that \( \mathcal{W}_p(X) \) is \textit{isometrically rigid} if every isometry of \( \mathcal{W}_p(X) \) is trivial.

Now we are ready to state and prove the main result of the paper.

**Theorem 1.** Let \( (X, \varrho) \) be a countable graph metric space and let \( p \geq 1 \) be a fixed real number. Then for any isometry \( \Phi : \mathcal{W}_p(X) \to \mathcal{W}_p(X) \) there exists an isometry \( \psi : X \to X \) of the underlying space such that \( \Phi = \psi_\# \). In other words, \( \mathcal{W}_p(X) \) is isometrically rigid for all \( p \geq 1 \).

**Proof.** The strategy of proof is similar to the sketch mentioned in the introduction.
Step 1. First we prove that $\Phi$ maps the set of Dirac masses onto itself. The key step is the characterization of the property that two measures differ only in one pair of neighbouring points, meaning that there exists $u, v \in X$ such that $\varrho(u, v) = 1$ and $\{z \in X \mid \mu(z) \neq \nu(z)\} = \{u, v\}$. For $\mu, \nu \in W_p(X)$ let us introduce the set

$$(6) \quad B_s(\mu, \nu) := \{\xi \in W_p(X) \mid d_p(\mu, \xi) \leq \sqrt{s}d_p(\mu, \nu), \quad d_p(\xi, \nu) \leq \sqrt{(1-s)d_p(\mu, \nu)}\}.$$ 

It is obvious that $B_s(\mu, \nu)$ is non-empty, because $\xi_s = (1-s)\mu + s\nu$ belongs to $B_s(\mu, \nu)$. Let us make an other simple observation: if $\Phi(\mu) = \mu$ and $\Phi(\nu) = \nu$ hold, then $\Phi(\xi) \in B_s(\mu, \nu)$ whenever $\xi \in B_s(\mu, \nu)$. Indeed,

$$(7) \quad d_p(\mu, \Phi(\xi)) = d_p(\Phi(\mu), \Phi(\xi)) = d_p(\mu, \xi) \leq \sqrt{s}d_p(\mu, \nu),$$

and similarly,

$$(8) \quad d_p(\Phi(\xi), \nu) = d_p(\Phi(\xi), \Phi(\nu)) = d_p(\xi, \nu) \leq \sqrt{(1-s)d_p(\mu, \nu)}.$$ 

This is one of the reasons why it would be useful to know whether $\xi_s$ is the only element of $B_s(\mu, \nu)$. We show that conditions (A) and (B) below are equivalent.

(A) There exist an $\eta \in M_+(X)$ and $u, v \in X$ with $\varrho(u, v) = 1$ such that

$$(9) \quad \mu = \eta + \alpha \cdot \delta_u \quad \text{and} \quad \nu = \eta + \alpha \cdot \delta_v.$$ 

(B) $d_p(\mu, \nu) = \sqrt{\alpha}$ and $B_s(\mu, \nu) = \{(1-s)\mu + s\nu\}$ holds for any $s \in (0, 1)$.

First we make a trivial observation: $d_p^{\varrho}(\mu, \nu) \geq |\mu(x) - \nu(x)|$ holds for all $x, y \in X$. Indeed, since $x \neq y$ implies $\varrho(x, y) \geq 1$, we have to move at least $|\mu(x) - \nu(x)|$ mass to $x$ or from $x$ with at least 1 cost. According to this, if we assume (A) then it is obvious that $d_p(\mu, \nu) \geq \sqrt{\alpha}$, and $d_p(\mu, \nu) \leq \sqrt{\alpha}$ also holds because moving $\alpha$ weight from $u$ to $v$ and leaving everything else fixed is a good transport plan.

Now assume that $\xi \in B_s(\mu, \nu)$. We have to show that $\xi = \xi_s = \eta + (1-s)\alpha \delta_u + s\alpha \delta_v$. Let us denote by $\pi^*$ that transport plan between $\mu$ and $\xi$ which realizes the Wasserstein distance, that is,

$$(10) \quad \alpha s \geq d_p^{\varrho}(\mu, \xi) = \sum_{(x, y) \in X \times X} \varrho(x, y)\pi^*(x, y).$$
The sum on the right hand side can be written as

\[
\sum_{x \in X} \sum_{y \in X} \rho(x, y)(x, y) \pi^*(x, y) + \left[ \sum_{y \in X} \rho(u, y)\pi^*(u, y) + \pi^*(u, u) \right] - \pi^*(u, u)
\]

which gives

\[
\alpha s \geq \mu(u) - \pi^*(u, u)
\]

because

\[
\sum_{y \in X} \rho(u, y)\pi^*(u, y) + \pi^*(u, u) \geq \sum_{y \in X} \pi^*(u, y) = \mu(u)
\]

and

\[
\sum_{x \in X} \sum_{y \in X} \rho(x, y)\pi^*(x, y) \geq 0.
\]

Since \( \pi \in \Pi(\mu, \xi) \), we have \( \xi(u) \geq \pi(u, u) \), and thus (A) implies (after some re-arrangement) that

\[
\xi(u) \geq \pi^*(u, u) \geq \eta(u) + (1 - s)\alpha \delta u(u).
\]

Combination of (10) and (15) asserts now that

\[
s\alpha \geq d_p^p(\mu, \xi) \geq |\xi(u) - \mu(u)| \geq |\eta(u) + (1 - s)\alpha(\eta(u) + \alpha \delta u(u))| = s\alpha.
\]

Furthermore, a very similar calculation with \( \xi \) and \( \nu \) gives

\[
(1 - s)\alpha \geq d_p^p(\nu, \xi) \geq |\xi(v) - \nu(v)| \geq |\eta(v) + s\alpha - (\eta(v) + \alpha \delta v(v))| = (1 - s)\alpha.
\]

Since every inequality in (16) and (17) is actually an equality, we get

\[
\xi(u) = \eta(u) + (1 - s)\alpha \delta u(u) \quad \text{and} \quad \xi(v) = \eta(v) + s\alpha \delta v(v)
\]

which imply with \( d_p^p(\mu, \xi) = |\mu(u) - \xi(u)| \) and \( d_p^p(\xi, \nu) = |\xi(v) - \nu(v)| \) that

\[
\xi = \eta + (1 - s)\alpha \delta u + s\alpha \delta v = \xi_s.
\]

To prove (B)\( \Rightarrow \) (A) assume that \( \xi_s \) is the only element of \( B_s(\mu, \nu) \) for all \( s \in (0, 1) \). In fact, we will only use that \( B_{1/2}(\mu, \nu) \) is a singleton. Let \( \pi \) be an optimal transport plan for which

\[
\alpha = d_p^p(\mu, \nu) = \sum_{(x, y) \in \text{supp}(\pi)} \rho(x, y)\pi(x, y).
\]
First assume indirectly that there exists an \((x', y') \in \text{supp}(\pi)\) for which \(k := \varrho(x', y') > 1\). Let us choose a \(k\)-long path between \(x'\) and \(y'\) and denote by \(x'_+\) the first point after \(x'\), and by \(y'_-\) the last point before \(y'\). Set \(c = \frac{\pi(x', y')}{2}\) and modify \(\xi_2\) along the path as follows

\[
\xi := \xi_2 - c\delta_{x'} + c\delta_{x'_+} + c\delta_{y'_-} - c\delta_{y'}.
\]

Obviously, \(\xi \neq \xi_2\) and it follows from the construction that

\[
d_p^p(\mu, \xi) \leq d_p^p(\mu, \xi_2) - k^p c + (k - 1)^p c + 1^p c \leq d_p^p(\mu, \xi_2) \leq \frac{\alpha}{2}.
\]

A similar calculation shows that \(d_p^p(\xi, \nu) \leq \frac{\alpha}{2}\), and thus \(\xi \in B_{\frac{\alpha}{2}}(\mu, \nu)\), a contradiction.

Now we know that \(\varrho(x, y) = 1\) for all \((x, y) \in \text{supp}(\pi)\) in (20). Assume indirectly that \(\text{supp}(\pi)\) has at least two different elements \((x_1, y_1)\) and \((x_2, y_2)\). Set \(c := \min\{\pi(x_1, y_1), \pi(x_2, y_2)\}\) and modify \(\xi_2\) as follows:

\[
\xi := \xi_2 + c\delta_{x_1} + c\delta_{y_1} - c\delta_{x_2} - c\delta_{y_2}.
\]

Again, it is obvious that \(\xi \neq \xi_2\). In order to give an upper estimation for \(d_p^p(\mu, \xi)\) let us define the coupling \(\tilde{\pi} \in \Pi(\mu, \xi)\):

\[
\tilde{\pi}(x, y) = \begin{cases} 
\pi(x_1, y_1) + c & \text{if } (x, y) = (x_1, y_1), \\
\pi(x_2, y_2) - c & \text{if } (x, y) = (x_2, y_2), \\
\pi(x, y) & \text{otherwise}.
\end{cases}
\]

It follows from the definition of \(\tilde{\pi}\) that \(\text{supp}(\tilde{\pi}) = \text{supp}(\pi)\) and that

\[
d_p^p(\mu, \xi) \leq C_{\varrho, p}(\pi) = \sum_{(x, y) \in \text{supp}(\pi)} \frac{\pi(x, y)}{2} = \frac{\alpha}{2}.
\]

Similarly, \(d_p^p(\xi, \nu) \leq \frac{\alpha}{2}\), and thus \(\xi\) belongs to \(B_{\frac{\alpha}{2}}(\mu, \nu)\), a contradiction. The only remaining possibility is that there exists \(u, v \in X\) such that \(\varrho(u, v) = 1\), \(\text{supp}(\pi) = \{(u, v)\}\), and \(\pi(u, v) = \alpha\) in (20), which means exactly that (A) holds. This proves (A) \iff (B).
If (A) holds then we will say that \( \mu \) and \( \nu \) are \( \alpha \)-neighbouring, or we will use the notation \( \mu \equiv \nu \ [\alpha] \). (Note that \( \equiv \) is not an equivalence relation, as it is not reflexive and not transitive.) Observe immediately that \( \mu \equiv \nu \ [\alpha] \iff \Phi(\mu) \equiv \Phi(\nu) \ [\alpha] \). Indeed, this follows from (A)\( \iff \) (B) and the fact that \( B_s(\mu, \nu) \) is a singleton if and only if \( B_s(\Phi(\mu), \Phi(\nu)) \) is a singleton (here we used that \( \Phi \) is a bijection).

Now assume that \( \mu = \delta_u \) is a Dirac measure and choose a \( v \in X \) such that \( g(u, v) = 1 \). Since \( \delta_u \) and \( \delta_v \) are 1-neighbouring, so are \( \Phi(\delta_u) \) and \( \Phi(\delta_v) \), that is, \( \Phi(\mu) = \delta_u \) and \( \Phi(\nu) = \delta_v \) for some \( \hat{u}, \hat{v} \in X \) with \( g(\hat{u}, \hat{v}) = 1 \). On the one hand, this implies that \( \Phi(\mu) \) is a Dirac measure. On the other hand, since \( \Phi^{-1} \) is an isometry as well, we can obtain that \( \Phi \) maps the set of Dirac masses bijectively onto itself.

Furthermore, the function \( \psi : X \to X \) defined by
\[
\Phi(\delta_x) := \delta_{\psi(x)} \quad (x \in X)
\]
is an isometry. As the map \( \widetilde{\Phi} := \psi \#^{-1} \circ \Phi \) fixes all Dirac measures and
\[
\widetilde{\Phi}(\mu) = \mu \quad (\mu \in \mathcal{W}_p(X))
\]
would imply \( \Phi = \psi \# \), we can assume without loss of generality that \( \psi(x) = x \) and hence \( \Phi(\delta_x) = \delta_x \) for all \( x \in X \). Our aim now is to show that \( \Phi(\mu) = \mu \) holds for all \( \mu \in \mathcal{W}_p(X) \).

Step 2. Our next task is to prove by induction that \( \Phi \) leaves a dense set of finitely supported measures fixed. We label the points of \( X \) as follows: let \( x_1 \) be an arbitrary element of \( X \) and let us denote the subspace \( \{x_1\} \) by \( X_1 \). If a collection of elements \( X_n = \{x_1, \ldots, x_n\} \) is already chosen for some \( n \geq 1 \), choose an \( x_{n+1} \) such that \( \min\{g(x_{n+1}, y) \mid y \in X_n\} = 1 \). (Using the language of the underlying graph \( G(X, E) \): first choose an arbitrary vertex \( x_1 \in X \) and then in each step choose an \( x_{n+1} \in X \) which is connected by an edge with at least one element of \( \{x_1, \ldots, x_n\} \).) If \( \mu \in \mathcal{F}(X) \) then \( \text{supp}(\mu) \subset X_N \) for some large enough \( N \in \mathbb{N} \), and therefore it is enough to show that measures supported in \( X_n \) are fixed by \( \Phi \) for all \( n \in \mathbb{N} \). For \( n = 1 \) this follows from our assumption, since \( \delta_{x_1} \) is the only measure that is supported in \( X_1 \).

If \( n = 2 \), then any element supported in \( X_2 \) can be written as \( \mu = (1-s)\delta_{x_1} + s\delta_{x_2} \) for some \( s \in [0, 1] \). If \( s = 0 \) or \( s = 1 \) then \( \mu \) is a Dirac measure and thus \( \Phi_{\mu} = \mu \) holds. Assume now that \( 0 < s < 1 \). Since \( \delta_{x_1} \equiv \delta_{x_2} \ [1] \), we have \( B_s(\delta_{x_1}, \delta_{x_2}) = \{(1-s)\delta_{x_1} + s\delta_{x_2}\} = \{\mu\} \) according to (A)\( \iff \) (B), and \( \Phi(\delta_x) = x_i \) \((i = 1, 2)\) imply \( \Phi(\mu) = \mu \) as it was pointed out in (7)-(8).

Now assume that \( \Phi(\xi) = \xi \) holds whenever \( \text{supp}(\xi) \subseteq X_n = \{x_1, \ldots, x_n\} \), and choose a finitely supported measure \( \mu = \sum_{i=1}^{n+1} \mu(x_i)\delta_{x_i} \) with the following properties: for all \( x, y \in \text{supp}(\mu) \): \( x \neq y \) implies \( \mu(x) \neq \mu(y) \), and for all pairwise different elements \( x, y, z \in \text{supp}(\mu) \): \( \mu(x) + \mu(y) \neq \mu(z) \). If \( \mu(x_{n+1}) = 0 \) then \( \text{supp}(\mu) \subseteq X_n \), if \( \mu(x_{n+1}) = 1 \) then \( \mu = \delta_{x_{n+1}} \). In both cases \( \Phi(\mu) = \mu \) by assumption, so it remains
to deal with the case $0 < \mu(x_{n+1}) < 1$. According to the construction, there exists an $u \in X_n$ such that $\rho(u, x_{n+1}) = 1$. Now set $c := \mu(u) + \mu(x_{n+1})$ and define two measures $\mu_*$ and $\mu^*$ as follows

$$
\mu_* := \mu - \mu(u)\delta_u + \mu(u)\delta_{x_{n+1}} \\
\mu^* := \mu - \mu(x_{n+1})\delta_{x_{n+1}} + \mu(x_{n+1})\delta_u
$$

For $\mu_*$ and $\mu^* \geq 0$ we have $\mu_* \equiv \mu^* [c]$ and

$$
\mu_*(u) = \mu^*(x_n) = 0, \quad \mu_*(x_{n+1}) = \mu^*(u) = c, \quad \mu_* \equiv \mu [\mu(u)], \quad \mu^* \equiv \mu [\mu(x_{n+1})].
$$

Teleporting $\mu_*$ to $\mu^*$ along the curve $\gamma$

Furthermore, the curve $\gamma : [0, c] \to \mathcal{W}_p(X)$ defined by $\gamma(t) := \mu^* + t\delta_{x_{n+1}} - t\delta_u$ connects $\mu^*$ with $\mu_*$ in a way that $\gamma(\mu(x_{n+1})) = \mu$, and for all $t \in [0, c]$ \{\{ $\mu(x_{n+1})$\}: $\gamma(t) \equiv \mu [t - \mu(x_{n+1})]$]. Now consider $\Phi(\mu_*)$ and $\Phi(\mu^*)$. Since $\mu_* \equiv \mu^* [c]$, we have $\Phi(\mu_*) \equiv \Phi(\mu^*) [c]$, and thus there exists a $v \in \text{supp}(\mu^*)$ and a $w \in X$ such that

$$
\Phi(\mu_*) = \mu_* \equiv c\delta_v + c\delta_w.
$$

At this point we can conclude that $\mu^*(v) \geq c$, and that for all $t \in [0, c]$ (and thus for $\gamma(\mu(x_{n+1})) = \mu$ as well)

$$
\Phi(\gamma(t)) = \mu^* - t\delta_v + t\delta_w.
$$

Indeed, for all $t \in (0, 1)$ we have $\gamma(t) \equiv \mu^* [t] \iff \Phi(\gamma(t)) \equiv \Phi(\mu^*) [t]$, and similarly, $\gamma(t) \equiv \mu_* [c - t] \iff \Phi(\gamma(t)) \equiv \Phi(\mu_*) [c - t]$ which forces (30). Now issue an other curve $\Gamma : [0, \mu^*(v))] \to \mathcal{W}_p(X)$ from $\mu^*$

$$
\Gamma(t) := \mu^* - t\delta_v + t\delta_w.
$$

Obviously, $\Gamma|_{[0, c]} = \Phi[\gamma]$ (where $\Phi[\gamma](t) := \Phi(\gamma(t))$ for all $t \in [0, c]$), and for all $t \in [0, \mu^*(v)) \setminus \{\mu(x_{n+1})\}$ we have $\Gamma(t) \equiv \Phi(\mu) [t - \Phi(x_{n+1})]$. Assume indirectly that $c < \mu^*(v)$ and consider the curve $\Phi^{-1}[\Gamma] : [0, \mu^*(v))] \to \mathcal{W}_p(X)$

$$
\Phi^{-1}[\Gamma](t) := \Phi^{-1}(\Gamma(t)).
$$

Then $\Phi^{-1}[\Gamma]|_{[0, c]} = \gamma$ holds, and $\mu_*$ is an inner point of $\Phi^{-1}[\Gamma]$. Moreover, $\Gamma(t) \equiv \Phi(\mu) [t - \mu(x_{n+1})]$ implies $\Phi^{-1}[\Gamma](t) \equiv \Phi^{-1}(\Phi(\mu)) [t - \mu(x_{n+1})]$, that is, $\gamma$ can be extended through $\mu_*$ by measures which are in neighbouring relation with $\mu$, a contradiction. Indeed, to continue $\gamma$ we need to add more weight to $x_{n+1}$. But $\mu_*(u) = 0$, so we should relocate mass from a point $x \in X \setminus \{x_{n+1}, u\}$ which would ruin
the neighbouring relation with $\mu$. This contradiction guarantees that $\mu^*(u) = \mu^*(v)$ holds. From this we can conclude by our initial assumptions on $\mu$ that $u = v$. If $\mu(u) = 0$ and $u \neq v$ then

$$\mu(v) = \mu^*(v) = \mu^*(u) = \mu(u) + \mu(x_{n+1}) = \mu(x_{n+1}) > 0$$

and thus $x_{n+1}, v \in \text{supp}(\mu)$, $x_{n+1} \neq v$, but $\mu(x_{n+1}) = \mu(v)$. If $\mu(u) \neq 0$ and $u \neq v$ then $\mu(v) = \mu^*(v) > 0$ implies $u, v, x_{n+1} \in \text{supp}(\mu)$ are pairwise different points and

$$\mu(v) = \mu^*(v) = \mu^*(u) = \mu(u) + \mu(x_{n+1})$$

holds, a contradiction. Having that $u = v$ we can easily finish this step. On the one hand, we have

$$d_p(\delta_{x_{n+1}}, \mu) = d_p(\delta_{x_{n+1}}, \mu^*) - \mu(x_{n+1})$$

and

$$d_p(\Phi(\delta_{x_{n+1}}, \mu)) = d_p(\delta_{x_{n+1}}, \mu^*) - \mu(x_{n+1}) + \rho^p(x_{n+1}, w)\mu(x_{n+1}).$$

On the other hand, we have

$$d_p(\delta_{x_{n+1}}, \mu) = d_p(\Phi(\delta_{x_{n+1}}, \mu)), \Phi(\mu)) = d_p(\delta_{x_{n+1}}, \Phi(\mu)).$$

Obviously, (36) and (37) imply $x_{n+1} = w$ which means that $\Phi(\mu) = \mu$.

Step 3. In the previous step we proved that $\Phi(\mu) = \mu$ holds if (i) no two points of the support have the same mass and (ii) $\mu(x) + \mu(y) \neq \mu(z)$ holds for all $x, y, z \in \text{supp}(\mu)$ such that $x \neq y, x \neq z, y \neq z$. Since $\mathcal{F}(X)$ is dense in $\mathcal{W}_p(X)$, it is enough to show that every $\nu \in \mathcal{F}(X)$ can be approximated by such measures. Fix an $\varepsilon > 0$ and a measure $\nu = \sum_{i=1}^N a_i\delta_{x_i}$. Set

$$K := \max\{|x_i - x_j| | 1 \leq i, j \leq N, i \neq j\},$$

and consider the cube

$$C = \prod_{i=1}^n \left[ a_i - \frac{\varepsilon}{K_pN}, a_i + \frac{\varepsilon}{K_pN} \right] \subseteq \mathbb{R}^n.$$ 

The intersection of $C$ with the hyperplane $P = \{(c_1, \ldots, c_N) | \sum_{i=1}^N c_i = 1\}$ is a large set in the sense that it cannot be covered by the union of those (finitely many) subspaces which are defined by the restrictions (i) and (ii). For all uncovered $(c_1, \ldots, c_N)$ we can obtain a measure $\mu = \sum_{i=1}^N c_i\delta_{x_i}$ such that $d_p(\mu, \nu) \leq \varepsilon$. Indeed, since we have to transport at most $\frac{\varepsilon}{K_pN}$ weight from at most $K$ distance for each $x_i$, we get

$$d_p(\mu, \nu) \leq \sqrt{NK_p\frac{\varepsilon}{K_pN}} = \varepsilon.\text{ This means that the set of these measures is dense in } \mathcal{W}_p(X), \text{ which completes the proof.}$$

\end{document}
3. Closing remarks

A very natural question was raised by König in [18]: which groups are isomorphic to the automorphism group of a graph? Of course, one can replace graphs with other mathematical structures, for example with Wasserstein spaces, and ask the same question. Since in the metric context, automorphisms are in particular isometries, the corresponding question reads as follows: which groups are isomorphic to the isometry group of a complete separable metric space? Using some famous results in graph theory, the answer for countable groups follows easily from our main theorem. (For an analogous result for autohomeomorphism groups see [4, Theorem 7].)

**Corollary 2.** Let $H$ be any countable group and $p \geq 1$ any real number. Then there exists a metric space $X$ such that $\text{Isom}(\mathcal{W}_p(X)) \cong H$.

**Proof.** As an extension of Frucht’s theorem [8], de Groot proved that every countable group $H$ is isomorphic to the automorphism group of a countable graph $G$ (see [4, comments on p.96]). Let $(X, \varrho)$ be the graph metric space associated to $G$ and consider the $p$-Wasserstein space $\mathcal{W}_p(X)$. As it was proved in the main theorem, $\text{Isom}(\mathcal{W}_p(X)) \cong \text{Isom}(X)$, and thus it is enough to show that $\text{Isom}(X) \cong \text{Aut}(G)$. But this is trivial, because on the one hand, any automorphism of $G$ is an isometry of $X$ by definition, and therefore $\text{Isom}(X) \supseteq \text{Aut}(G)$. On the other hand if $f \in \text{Isom}(X)$ then $x$ and $y$ ($x \neq y$) are adjacent if and only if $f(x)$ and $f(y)$ are adjacent if and only if $\varrho(x, y) = \varrho(f(x), f(y)) = 1$, and $x$ and $y$ are non-adjacent if and only if $f(x)$ and $f(y)$ are non-adjacent if and only if $\varrho(x, y) = \varrho(f(x), f(y)) \geq 2$, thus $f$ is an automorphism of $G$ and $\text{Isom}(X) \subseteq \text{Aut}(G)$.

We remark that there is no uniqueness in Frucht’s and de Groot’s theorems. In fact, Izbicki proved in [15] that there are uncountably many infinite graphs realizing any finite symmetry group.) We also remark that de Groot’s theorem is valid for non-countable groups as well, however, we do not know the smallest possible order of the representing graph. If the cardinality of the vertex set is bigger than $\aleph_0$, then the graph corresponding graph metric space is not separable, and thus our proof of rigidity fails as the set of finitely supported measures is not dense anymore. Therefore the following question remains open:

**Problem 3.** Given an uncountable group $G$, does there exists a $p$-Wasserstein space whose isometry group is $G$?
References

[1] L. Ambrosio, N. Gigli, A user’s guide to optimal transport. Modelling and optimisation of flows on networks, Lecture Notes in Math., 2062, Fond. CIME/CIME Found. Subser., Springer, Heidelberg, 2013.

[2] J. Bertrand, B. Kloeckner, A geometric study of Wasserstein spaces: isometric rigidity in negative curvature, Int. Math. Res. Notices 5 (2016), 1368–1386.

[3] G. Chartrand, D. Kay, A Characterization of Certain Ptolemaic Graphs, Canadian Journal of Mathematics 17 (1965), 342–346.

[4] J. de Groot, Groups represented by homeomorphism groups I., Math. Ann. 138 (1959), 80–102.

[5] G. De Palma, M. Marvian, D. Trevisan and S. Lloyd, The quantum Wasserstein distance of order 1, arXiv preprint, arXiv:2009.04469, 2020.

[6] G. Dolinar, B. Kuzma, D. Mitrivic, Isometries of probability measures with respect to the total variation distance Manuscript, 2021.

[7] G. Dolinar, L. Molnár, Isometries of the space of distribution functions with respect to the Kolmogorov–Smirnov metric, J. Math. Anal. Appl. 348 (2008), 494–498.

[8] R. Frucht, Herstellung von Graphen mit vorgegebenen abstrakter Gruppe, Compositio Math. 6 (1938), 239–250.

[9] P. Jiradilok, S. Kamtue, Transportation distance between probability measures on the infinite regular tree, manuscript, arXiv:2107.09876, 2021.

[10] Gy.P. Gehér, Surjective Kuiper isometries, Houston J. Math. 44 (2018), 263–281.

[11] Gy.P. Gehér, 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.

[12] Gy.P. Gehér, T. Titkos, D. Virosztek, On isometric embeddings of Wasserstein spaces – the discrete case, J. Math. Anal. Appl. 480 (2019), 123435.

[13] Gy.P. Gehér, T. Titkos, D. Virosztek, Isometric study of Wasserstein spaces – the real line, Trans. Amer. Math. Soc. 373 (2020), 5855–5883.

[14] Gy. P. Gehér, T. Titkos, D. Virosztek, The isometry group of Wasserstein spaces: the Hilbertian case, manuscript, arXiv:2102.02037, 2021.

[15] H. Izbicki, Unendliche Graphen endlichen Grades mit vorgegebenen Eigenschaften, Monatshefte für Mathematik 63 (1959), 298–301.

[16] B. Kloeckner, A geometric study of Wasserstein spaces: Euclidean spaces, Annali della Scuola Normale Superiore di Pisa - Classe di Scienze IX (2010), 297–323.

[17] B. Kloeckner, A generalization of Hausdorff dimension applied to Hilbert cubes and Wasserstein spaces, J. Topol. Anal. 4 (2012), 203–235.

[18] D. König, Theorie der endlichen und unendlichen Graphen, Akad. Verlag, Leipzig 1936.

[19] L. Molnár, Lévy isometries of the space of probability distribution functions, J. Math. Anal. Appl. 380 (2011), 847–852.

[20] D.S. Ornstein, An Application of Ergodic Theory to Probability Theory, The Annals of Probability 1 (1973), 43–58.

[21] F. Rhodes, Some characterizations of the chessboard metric and the city block metric, Pattern Recognition Letters 11 (1990), 669–675.

[22] F. Santambrogio, Optimal Transport for Applied Mathematicians, Progress in Nonlinear Differential Equations and Their Applications 87, Birkhauser Basel (2015).
[23] J. Santos-Rodríguez, Isometric rigidity of compact Wasserstein spaces, manuscript, arXiv:2102.08725, 2021.

[24] C. Villani, Optimal Transport: Old and New, Grundlehren der mathematischen Wissenschaften, Springer, 2009.

[25] D. Virosztek, Maps on probability measures preserving certain distances — a survey and some new results., Acta Sci. Math. (Szeged) 84 (2018), 65–80.

Gergely Kiss, Alfréd Rényi Institute of Mathematics – Eötvös Loránd Research Network, Réáltanoda u. 13-15., Budapest H-1053, Hungary.

Email address: kiss.gergely@renyi.hu

Tamás Titkos, Alfréd Rényi Institute of Mathematics – Eötvös Loránd Research Network, Réáltanoda 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.hu