Regular Articles

Isometric rigidity of Wasserstein spaces over Euclidean spheres ☆

György Pál Gehér [a], Aranka Hrušková [b,c], Tamás Titkos [b,d], Dániel Virosztek [b,*]

[a] *Department of Mathematics and Statistics, University of Reading, Whiteknights, P.O. Box 220, Reading RG6 6AX, United Kingdom*
[b] *HUN-REN Alfréd Rényi Institute of Mathematics Reáltanoda u. 13-15, Budapest 1053, Hungary*
[c] *Central European University, Nádor u. 9, Budapest 1051, Hungary*
[d] *Corvinus University of Budapest, Department of Mathematics, Fővám tér 13-15, Budapest 1093, Hungary*

A B S T R A C T

We study the structure of isometries of the quadratic Wasserstein space $\mathcal{W}_2(\mathbb{S}^n, \| \cdot \|)$ over the sphere endowed with the distance inherited from the norm of $\mathbb{R}^{n+1}$. We prove that $\mathcal{W}_2(\mathbb{S}^n, \| \cdot \|)$ is isometrically rigid, meaning that its isometry group is isomorphic to that of $(\mathbb{S}^n, \| \cdot \|)$. This is in striking contrast to the non-rigidity of its ambient space $\mathcal{W}_2(\mathbb{R}^{n+1}, \| \cdot \|)$ but in line with the rigidity of the geodesic space $\mathcal{W}_2(\mathbb{S}^n, \triangleleft)$. One of the key steps of the proof is the use of mean squared error functions to mimic displacement interpolation in $\mathcal{W}_2(\mathbb{S}^n, \| \cdot \|)$. A major difficulty in proving rigidity for quadratic Wasserstein spaces is that one cannot use the Wasserstein potential technique. To illustrate its general power, we use it to prove isometric rigidity of $\mathcal{W}_p(\mathbb{S}^1, \| \cdot \|)$ for $1 \leq p < 2$.

Contents

1. Motivation and main result

In recent years, there has been considerable activity in characterising isometries of various metric spaces of probability measures. See e.g. [2–4,6,7,9–21,24,27] for results about the total variation, Lévy, Kuiper, Lévy-Prokhorov, Kolmogorov-Smirnov, and Wasserstein metrics. Among these, an interesting result is due to Kloeckner. In [18, Theorem 1.1 and Theorem 1.2], he shows that the quadratic Wasserstein space $\mathcal{W}_2\left(\mathbb{R}^{n+1}, \|\cdot\|\right)$, where $\|\cdot\|$ stands for the metric induced by the norm, exhibits the rare phenomenon of not being isometrically rigid, meaning that not all isometries of $\mathcal{W}_2\left(\mathbb{R}^{n+1}, \|\cdot\|\right)$ are induced by an isometry of $\left(\mathbb{R}^{n+1}, \|\cdot\|\right)$. In this paper, we consider the metric subspace $(\mathbb{S}^n, \|\cdot\|)$ of the base space $\left(\mathbb{R}^{n+1}, \|\cdot\|\right)$ and prove that the non-rigidity does not carry over: the exotic isometries of $\mathcal{W}_2\left(\mathbb{R}^{n+1}, \|\cdot\|\right)$ send measures supported on $\mathbb{S}^n$ to measures supported also outside of $\mathbb{S}^n$, while we gain no new exotic isometries by restricting to this smaller metric space. In general, when H is an arbitrary Borel subset of $\mathbb{R}^{n+1}$, then $\mathcal{W}_p(H, \|\cdot\|)$ embeds isometrically into $\mathcal{W}_p\left(\mathbb{R}^{n+1}, \|\cdot\|\right)$, but this does not necessarily imply that there exists such a natural embedding for their isometry groups. To see an example, we mention the case of the real line $(\mathbb{R}, |\cdot|)$ with the subset $H = [0,1]$ (see [12, Theorem 2.5 and Theorem 3.7] for details): the isometry group of $\mathcal{W}_1([0,1], |\cdot|)$ is the Klein group, which cannot be embedded by a group homomorphism into the isometry group of $\mathcal{W}_1(\mathbb{R}, |\cdot|)$, which is isomorphic to the isometry group of the real line.

Finally, we draw attention to Santos-Rodríguez's paper [24] and our recent work [15]. In [24], the author considers (among others) Wasserstein spaces with $p > 1$ whose underlying metric space is a rank-one symmetric space, which class contains the sphere $\mathbb{S}^n$ with the spherical distance $\triangleleft$, while in [15], we considered finite-dimensional tori and spheres with their geodesic distances for all parameters $p \geq 1$. Together, these two papers show that $\mathcal{W}_p(\mathbb{S}^n, \triangleleft)$ is isometrically rigid for all $p \geq 1$. As explained above, in this paper, we replace the angular distance $\triangleleft$ with another natural metric: the distance inherited from the norm of $\mathbb{R}^{n+1}$. We focus on the case of $p = 2$ because this is the only parameter value for which the ambient space $\mathcal{W}_p\left(\mathbb{R}^{n+1}, \|\cdot\|\right)$ is *not* rigid. We expect that for $p \neq 2$, techniques similar to the ones used in [14] would lead to a proof of isometric rigidity. The situation is analogous to the case of the real line and the unit interval: the quadratic Wasserstein space is not rigid over $\mathbb{R}$ but it is rigid over the compact subset $[0,1]$, see [18, Theorem 1.1] and [12, Theorem 2.6]. Our main result reads as follows.

Theorem 1.1. *For all $n \in \mathbb{N}$, the quadratic Wasserstein space $\mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$ is isometrically rigid. That is, for any isometry $\Psi \colon \mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right) \to \mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$, there exists an isometry $\psi \colon \mathbb{S}^n \to \mathbb{S}^n$ such that $\Psi = \psi_\#$.*

In our recent works [14,15], recovering measures from their Wasserstein potentials — see (2.3) for precise definition — turned out to be a powerful method to prove isometric rigidity. However, this method cannot be used in the case of $\mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$, as shown by the following simple example. Let δ_x denote the Dirac measure concentrated at $x \in \mathbb{S}^n$, let $\mu_z := \frac{1}{2}(\delta_z + \delta_{-z})$ for $z \in \mathbb{S}^n$, and note that for any $x \in \mathbb{S}^n$ we have $d_{\mathcal{W}_2}^2(\mu_z, \delta_x) = \frac{1}{2}(\|x - z\|^2 + \|x + z\|^2) = 2$ independently of x and z — see (2.1) for the precise definition of the p-Wasserstein distance $d_{\mathcal{W}_p}$. This means that every element of the set $\{\mu_z \mid z \in \mathbb{S}^n\}$ has the same Wasserstein potential function, and hence potentials do not determine measures uniquely in general.

Our complimentary result Theorem 3.1 demonstrates sensitivity of the Wasserstein potential method to the parameter value p. Namely, we show that, at least in the special case of $\mathbb{S}^1$, measures are uniquely determined by their potentials if $1 \leq p < 2$, and hence $\mathcal{W}_p\left(\mathbb{S}^1, \|\cdot\|\right)$ is isometrically rigid.

2. The Wasserstein space $\mathcal{W}_p\left(\mathbb{S}^n, \|\cdot\|\right)$ and the Wasserstein potential

In this section, we recall all the necessary notions and notations. Let (Y, ρ) be a complete and separable metric space, $p \geq 1$ a fixed real number, and $\mathcal{P}(Y)$ the set of all Borel probability measures on Y. The p-Wasserstein space $\mathcal{W}_p(Y, \rho)$, where $p \in [1, \infty)$, is then defined as the set

$$\left\{ \mu \in \mathcal{P}(Y) \,\middle|\, \exists \hat{y} \in Y : \int_Y \rho(y, \hat{y})^p \, \mathrm{d}\mu(y) < \infty \right\}$$

of probability measures endowed with the *p-Wasserstein metric*

$$d_{\mathcal{W}_p}(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \iint_{Y \times Y} \rho(x, y)^p \, \mathrm{d}\pi(x, y) \right)^{1/p}, \tag{2.1}$$

where the infimum is taken over the set $\Pi(\mu, \nu)$ of all couplings of μ and ν. A Borel probability measure π on $Y \times Y$ is called a *coupling* of μ and ν if $\pi(A \times Y) = \mu(A)$ and $\pi(Y \times B) = \nu(B)$ for all Borel sets $A, B \subseteq Y$. For more details about Wasserstein spaces, we refer the reader to the comprehensive textbooks [1,8,23,26]. Now we only mention that optimal couplings always exist, and the infimum in (2.1) becomes minimum [1, Theorem 1.5]. Furthermore, finitely supported measures are dense in Wasserstein spaces, see, e.g., [26, Theorem 6.18].

An *isometric embedding* between metric spaces (X, d) and (Y, ρ) is a map $\phi \colon (X, d) \to (Y, \rho)$ which preserves distances, i.e., a map such that $d(x, x') = \rho(\phi(x), \phi(x'))$ for all $x, x' \in X$. We shall use the term *isometry* for a surjective isometric embedding from a metric space onto itself. It is important to note that if (X, d) is a compact metric space, then every isometric embedding $\phi \colon (X, d) \to (X, d)$ is surjective and hence an isometry [5, Theorem 1.6.14].

For a Borel-measurable map $\psi \colon Y \to Y$, its push-forward $\psi_\# \colon \mathcal{W}_p(Y, \rho) \to \mathcal{W}_p(Y, \rho)$ is defined by $\left(\psi_\#(\mu)\right)(A) := \mu(\psi^{-1}[A])$, where $A \subseteq Y$ is a Borel set and $\psi^{-1}[A] = \{x \in X \,|\, \psi(x) \in A\}$. In particular, when $\psi \colon Y \to Y$ is an isometry, then so is $\psi_\#$ by the very definition of the Wasserstein distance, giving rise to a canonical embedding of the isometries of (Y, ρ) to the isometries of $\mathcal{W}_p(Y, \rho)$.

In this paper, we consider the compact metric space $(\mathbb{S}^n, \|\cdot\|)$, where

$$\mathbb{S}^n := \{x \in \mathbb{R}^{n+1} \colon \|x\| = 1\}$$

is the unit sphere of $\mathbb{R}^{n+1}$ equipped with the distance inherited from the Euclidean norm of $\mathbb{R}^{n+1}$, that is, $\mathrm{dist}(x, y) = \|x - y\|$. The point $-x$ is called the antipodal of x. Since $(\mathbb{S}^n, \|\cdot\|)$ is bounded, the Wasserstein space $\mathcal{W}_p(\mathbb{S}^n, \|\cdot\|)$ is the entire set $\mathcal{P}(\mathbb{S}^n)$ endowed with the distance

$$d_{\mathcal{W}_p}(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \iint_{\mathbb{S}^n \times \mathbb{S}^n} \|x - y\|^p \, \mathrm{d}\pi(x, y) \right)^{1/p}. \tag{2.2}$$

We write $\mathcal{W}_p(\mathbb{S}^n, \|\cdot\|)$ instead of the usual $\mathcal{W}_p(\mathbb{S}^n)$ notation to avoid any confusion with the results in [15,24]. As the Wasserstein distance metrizes the weak convergence of probability measures over bounded metric spaces (see, e.g., [25, Theorem 7.12]), by Prokhorov's theorem, $(\mathbb{S}^n, \|\cdot\|)$ being compact tells us that $\mathcal{W}_p(\mathbb{S}^n, \|\cdot\|)$ is compact too — see also Remark 6.19 in [26]. This implies that every isometric embedding of $\mathcal{W}_p(\mathbb{S}^n, \|\cdot\|)$ into itself is an isometry.

For a measure $\mu \in \mathcal{P}(\mathbb{S}^n)$, its support $\mathrm{supp}(\mu)$ is the set of all points $x \in \mathbb{S}^n$ for which every open neighbourhood of x has positive measure. As usual, δ_x denotes the Dirac measure supported on the single point $x \in \mathbb{S}^n$.

The question arises whether it is possible to identify a measure if we know its distance from all Dirac measures. (Recall that $d_{\mathcal{W}_p}(\delta_x, \delta_y) = \|x - y\|$ for all $x, y \in \mathbb{S}^n$ and thus the set of all Dirac measures is an isometric copy of the underlying metric space.) To answer this question, we first introduce the notion of *Wasserstein potential* $\mathcal{T}_\mu^{(p)}$. For a given $\mu \in \mathcal{W}_p(\mathbb{S}^n, \|\cdot\|)$, the Wasserstein potential is the function

$$\mathcal{T}_{\mu}^{(p)} : \mathbb{S}^n \to \mathbb{R}; \qquad \mathcal{T}_{\mu}^{(p)}(x) := d_{\mathcal{W}_p}^p (\delta_x, \mu) = \int_{\mathbb{S}^n} \|x - y\|^p \, \mathrm{d}\mu(y). \tag{2.3}$$

Now, the question above can be rephrased as follows: *does the Wasserstein potential determine the measure uniquely?*

3. Does the Wasserstein potential determine the measure uniquely?

The answer to this question is no, in general. A prominent example is $\mathcal{W}_2 (\mathbb{S}^n, \|\cdot\|)$ where measures supported on antipodal points with both weights equal to $\frac{1}{2}$ have the same (constant) potential function — see the example in Section 1, after Theorem 1.1. Beyond this, exotic isometries of $\mathcal{W}_2(\mathbb{R}, |\cdot|)$ (see [18, Section 5.1 and Section 5.2]) are also counterexamples.

However, we will now prove that it does in the case of $\mathbb{S}^1 \simeq \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ equipped with the distance function $r(z, \omega) = \left|\frac{1}{2}(z - \omega)\right|$ for $1 \le p < 2$. This normalization of the distance is consistent with the one used in [27]. In this section, we assume that $1 \le p < 2$, and we recall that the p-Wasserstein distance of $\mu, \nu \in \mathcal{W}_p(\mathbb{T})$ in this case is

$$d_{\mathcal{W}_p} (\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu, \nu)} \iint_{\mathbb{T} \times \mathbb{T}} \left| \frac{1}{2}(z - \omega) \right|^p \, \mathrm{d}\pi (z, \omega) \right)^{1/p},$$

and therefore for any $z \in \mathbb{T}$, the Wasserstein potential is of the form

$$\mathcal{T}_{\mu}^{(p)}(z) = d_{\mathcal{W}_p}^p (\delta_z, \mu) = \int_{\mathbb{T}} \left| \frac{1}{2}(z - \omega) \right|^p \, \mathrm{d}\mu (\omega). \tag{3.1}$$

We showed in [15] that Fourier analytic methods can sometimes solve the problem of rigidity in a very elegant way. For example, we showed that isometric rigidity of $\mathcal{W}_2(\mathbb{T}, \triangleleft)$ can be proved by using the Fourier transform of the Wasserstein potential, however, the same method fails in the case $\mathcal{W}_1(\mathbb{T}, \triangleleft)$. As we will see, if we endow $\mathbb{T}$ with the distance $r(z, \omega) = \left|\frac{1}{2}(z - \omega)\right|$, then the situation changes: the same method works to prove isometric rigidity of $\mathcal{W}_1(\mathbb{T}, r)$, but fails in the case $\mathcal{W}_2(\mathbb{T}, r)$.

Now we recall the very basics of Fourier analysis on the abelian group $\mathbb{T}$. The main reason for doing so is to fix the notation. The continuous *characters* of $\mathbb{T}$ are exactly the power functions with an integer exponent. That is, if $\varphi_k(z) = z^k$ for all $k \in \mathbb{Z}$ and Γ is the *dual group* (i.e., the group of all continuous characters), then $\Gamma = \{\varphi_k : \mathbb{T} \to \mathbb{C} \mid k \in \mathbb{Z}\}$, and $\Gamma \cong \mathbb{Z}$. The group $\mathbb{T}$ is compact, hence it admits a unique *Haar probability measure* λ, which can be expressed explicitly as

$$\mathrm{d}\lambda (z) = \frac{\mathrm{d}z}{2\pi i z}.$$

The *Fourier transform* of a (complex-valued) function $f \in L^1 (\mathbb{T}, \lambda)$ is defined by

$$\hat{f}(k) = \int_{\mathbb{T}} f \overline{\varphi_k} \mathrm{d}\lambda = \frac{1}{2\pi i} \int_{\mathbb{T}} f(z) z^{-(k+1)} \mathrm{d}z \qquad (k \in \mathbb{Z}).$$

Let us denote the set of all (complex-valued) measures of finite total variation by $\mathbb{M}(\mathbb{T})$. The Fourier transform of $\mu \in \mathbb{M}(\mathbb{T})$ is defined by

$$\hat{\mu}(k) = \int_{\mathbb{T}} \overline{\varphi_n} \mathrm{d}\mu = \int_{\mathbb{T}} z^{-k} \, \mathrm{d}\mu (z) \qquad (k \in \mathbb{Z}).$$

We note that L^1 functions can be naturally identified with absolutely continuous measures (with respect to the Haar measure), see [22, Subsection 1.3.4.]. The convolution of L^1 functions f and g is defined by

$$(f * g)(z) = \int_{\mathbb{T}} f(z\omega^{-1}) g(\omega) \, d\lambda(\omega) \tag{3.2}$$

and the convolution of $f \in L^1(\mathbb{T}, \lambda)$ and $\mu \in \mathbb{M}(\mathbb{T})$ is defined by

$$(f * \mu)(z) = \int_{\mathbb{T}} f(z\omega^{-1}) \, d\mu(\omega). \tag{3.3}$$

It is a key identity that the Fourier transform factorizes the convolution, that is,

$$\widehat{f * \nu} = \hat{f} \cdot \hat{\nu}. \tag{3.4}$$

Now we are ready to state and prove the main result of this section. It says that if $1 \le p < 2$, then the Wasserstein space $\mathcal{W}_p(\mathbb{T}, r)$ is isometrically rigid.

Theorem 3.1. *Let $p \in [1, 2)$ be a real number and let $\Psi \colon \mathcal{W}_p(\mathbb{T}, r) \to \mathcal{W}_p(\mathbb{T}, r)$ be an isometry. Then there exists an isometry $\psi \colon (\mathbb{T}, r) \to (\mathbb{T}, r)$ such that $\Psi = \psi_\#$.*

Proof. First observe that the diameter of $\mathcal{W}_p(\mathbb{T}, r)$ is 1, and $d_{\mathcal{W}_p}(\mu, \nu) = 1$ if and only if $\mu = \delta_x$ and $\nu = \delta_{-x}$ for some $x \in \mathbb{T}$. Since Ψ is an isometry, we have

$$1 = d_{\mathcal{W}_p}(\delta_x, \delta_{-x}) = d_{\mathcal{W}_p}(\Psi(\delta_x), \Psi(\delta_{-x})) \tag{3.5}$$

for all $x \in \mathbb{T}$, which implies that $\Psi(\delta_x)$ is a Dirac measure as well.

Let us define the map $\psi \colon \mathbb{T} \to \mathbb{T}$ via the identity $\Psi(\delta_x) = \delta_{\psi(x)}$ – this means that Ψ coincides with $\psi_\#$ on the set of Dirac measures. The map $\psi \colon (\mathbb{T}, r) \to (\mathbb{T}, r)$ is in fact an isometry:

$$r(\psi(x), \psi(y)) = d_{\mathcal{W}_p}(\delta_{\psi(x)}, \delta_{\psi(y)}) = d_{\mathcal{W}_p}(\Psi(\delta_x), \Psi(\delta_y)) = d_{\mathcal{W}_p}(\delta_x, \delta_y) = r(x, y)$$

for all $x, y \in \mathbb{T}$, and $(\mathbb{T}, r)$ is compact. These together combine into that $(\psi^{-1})_\# \circ \Psi$ is an isometry which fixes all Dirac measures. If we now prove that any isometry of $\mathcal{W}_p(\mathbb{T}, r)$ which fixes all Dirac measures must be the identity, we are done: in that case, $(\psi^{-1})_\# \circ \Psi = \mathrm{id}_{\mathcal{W}_p(\mathbb{T}, r)}$, i.e., $\Psi = \psi_\#$ as claimed.

From now on, let us assume that $\Phi \colon \mathcal{W}_p(\mathbb{T}, r) \to \mathcal{W}_p(\mathbb{T}, r)$ is an isometry such that $\Phi(\delta_z) = \delta_z$ for all $z \in \mathbb{T}$. Then we have

$$\mathcal{T}_\mu^{(p)}(z) = d_{\mathcal{W}_p}^p(\delta_z, \mu) = d_{\mathcal{W}_p}^p(\Phi(\delta_z), \Phi(\mu)) = d_{\mathcal{W}_p}^p(\delta_z, \Phi(\mu)) = \mathcal{T}_{\Phi(\mu)}^p(z)$$

for all $z \in \mathbb{T}$ and $\mu \in \mathcal{W}_p(\mathbb{T}, r)$. The question is whether this implies $\mu = \Phi(\mu)$. The proof will be done once we prove that a measure $\mu \in \mathcal{W}_p(\mathbb{T}, r)$ is uniquely determined by its Wasserstein potential. To this end, assume that μ and ν are two measures such that

$$\mathcal{T}_\mu^{(p)}(z) = \mathcal{T}_\nu^p(z) \quad \text{for all } z \in \mathbb{T}. \tag{3.6}$$

We need to show that (3.6) implies $\mu = \nu$. Let us introduce the map

$$f_p(z) := \left| \frac{1}{2}(z - 1) \right|^p. \tag{3.7}$$

Then by (3.1) and (3.3) one can observe that $\mathcal{T}_\mu^{(p)}(z) = (f_p * \mu)(z)$ holds for all $z \in \mathbb{T}$ and $\mu \in \mathcal{W}_p(\mathbb{T})$. Indeed, we have

$$
\mathcal{T}_\mu^{(p)}(z) = d_{\mathcal{W}_p}^p (\delta_z, \mu) = \int_{\mathbb{T}} \left| \frac{1}{2}(z - \omega) \right|^p \, d\mu(\omega)
$$
$$
= \int_{\mathbb{T}} \left| \frac{1}{2} \left(z\omega^{-1} - 1 \right) \right|^p \, d\mu(\omega) = \int_{\mathbb{T}} f_p\left(z\omega^{-1} \right) \, d\mu(\omega) = (f_p * \mu)(z). \tag{3.8}
$$

The key observation is that the Fourier transform of f_p does not vanish anywhere, that is, $\hat{f}_p(n) \neq 0$ for all $n \in \mathbb{Z}$. For $n = 0$, we have

$$
\hat{f}_p(0) = \int_{\mathbb{T}} \left| \frac{1}{2}(z - 1) \right|^p \, d\lambda(z) > 0,
$$

while for $n \neq 0$, we use that

$$
f_p(z) = \left(\left| \frac{1}{2}(z - 1) \right|^2 \right)^{\frac{p}{2}} = \left(\frac{1}{4} \left(2 - z - z^{-1} \right) \right)^{\frac{p}{2}} = \left(1 - \frac{1}{4} \left(2 + z + z^{-1} \right) \right)^{\frac{p}{2}},
$$

and by the binomial series expansion we get that

$$
f_p(z) = \sum_{k=0}^\infty \binom{\frac{p}{2}}{k} \left(\frac{-1}{4} \left(2 + z + z^{-1} \right) \right)^k, \tag{3.9}
$$

where $\binom{\frac{p}{2}}{0} = 1$ and $\binom{\frac{p}{2}}{k} = \frac{\prod_{j=0}^{k-1} \left(\frac{p}{2} - j \right)}{k!}$. Using that the sign of $\binom{\frac{p}{2}}{k} (-1)^k$ is negative for all $k \geq 1$, equality (3.9) can be written as

$$
f_p(z) = 1 - \left\{ \frac{p}{2} \cdot \frac{2 + z + z^{-1}}{4} + \sum_{k=2}^\infty \left(\frac{\frac{p}{2} \prod_{j=1}^{k-1} (j - \frac{p}{2})}{k!} \left(\frac{2 + z + z^{-1}}{4} \right)^k \right) \right\}. \tag{3.10}
$$

It is a useful feature of the group $\mathbb{T}$ that the Fourier series of a function coincides with its power series. Therefore, the above binomial expansion gives us useful information about $\hat{f}_p$, namely, $\hat{f}_p(k)$ coincides with the coefficient of z^k in the expansion (3.9).

Let us note that for $n \neq 0$, the coefficient of z^n must be strictly negative because the expressions $\frac{p}{2}, 1 - \frac{p}{2}, 2 - \frac{p}{2}, \dots$ are all positive – here we use the assumption that $p < 2$. So we obtained that $\hat{f}_p(0) > 0$ and $\hat{f}_p(n) < 0$ for $n \neq 0$ which means that $\hat{f}_p(n) \neq 0$ for all $n \in \mathbb{Z}$.

By (3.8), the assumption that $\mathcal{T}_\mu^{(p)}(z) = \mathcal{T}_\nu^p(z)$ for all $z \in \mathbb{T}$ implies that $f_p * \mu = f_p * \nu$. By (3.4), this means that $\hat{f}_p \cdot \hat{\mu} = \hat{f}_p \cdot \hat{\nu}$. Since $\hat{f}_p(n) \neq 0$ for every n, we can deduce that $\hat{\mu} = \hat{\nu}$, but the Fourier transform completely determines the measure [22, Chapter 1], hence $\mu = \nu$, and the proof is done. $\square$

4. Isometric rigidity of $\mathcal{W}_2 (\mathbb{S}^n, \| \cdot \|)$ — proof of Theorem 1.1

The assumption $p < 2$ was crucial in the previous section, and therefore the quadratic case cannot be handled with the same Fourier-analytic technique. In this section, we use a method that allows us to prove isometric rigidity in the quadratic case not only over the circle but over higher-dimensional spheres too. We start this section with three propositions which will be utilized later in the proof of Theorem 1.1.

The first proposition, which can be found also in the Appendix of [14] (see the proof of Lemma 3.13 there), helps us understand how a translation affects the Wasserstein distance. For $\mu \in \mathcal{P}\left(\mathbb{R}^{n+1}\right)$ and $v \in \mathbb{R}^{n+1}$, the translation of μ by v is the measure $(t_v)_{\#}\mu \in \mathcal{P}\left(\mathbb{R}^{n+1}\right)$ where $t_v \colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$, $x \mapsto x + v$ is the translation by v. Recall that elements of $\mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$ can be considered as elements of $\mathcal{W}_2(\mathbb{R}^{n+1}, \|\cdot\|)$ because $\mathcal{W}_2(\mathbb{S}^n, \|\cdot\|)$ naturally embeds into $\mathcal{W}_2(\mathbb{R}^{n+1}, \|\cdot\|)$. The barycenter of $\mu \in \mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$ is defined to be the point $m(\mu) = \int_{\mathbb{S}^n} x \, d\mu(x) \in \mathbb{R}^{n+1}$. Note that the barycenter $m(\mu)$ defined above is also the unique point in $\mathbb{R}^{n+1}$ satisfying $\langle m(\mu), z \rangle = \int_{\mathbb{R}^{n+1}} \langle x, z \rangle \, d\mu(x)$ for all $z \in \mathbb{R}^{n+1}$, which equation is used to define the barycenter in an infinite-dimensional setting. However, in finite dimensions, the above direct definition, which does not refer to the dual space, is available.

Proposition 4.1. *Let* $\mu, \nu \in \mathcal{W}_2\left(\mathbb{R}^{n+1}\right)$ *and* $v \in \mathbb{R}^{n+1}$. *Then we have*

$$d_{\mathcal{W}_2}^2\left((t_v)_{\#}\mu, \nu\right) = d_{\mathcal{W}_2}^2(\mu, \nu) + \langle v, v + 2m(\mu) - 2m(\nu) \rangle. \tag{4.1}$$

In particular, substituting $v = m(\nu) - m(\mu)$ *gives*

$$d_{\mathcal{W}_2}^2(\mu, \nu) = d_{\mathcal{W}_2}^2\left((t_{-m(\mu)})_{\#}\mu, (t_{-m(\nu)})_{\#}\nu\right) + \|m(\nu) - m(\mu)\|^2. \tag{4.2}$$

Subsequently, ν *is a translated version of* μ *if and only if* $d_{\mathcal{W}_2}(\mu, \nu) = \|m(\nu) - m(\mu)\|$.

Proof. For any $\pi \in \Pi(\mu, \nu)$ and $v \in \mathbb{R}^{n+1}$, we have $(t_{(v,0)})_{\#}\pi \in \Pi\left((t_v)_{\#}\mu, \nu\right)$, and vice versa. (Here, 0 stands for $0 \in \mathbb{R}^{n+1}$.) Hence

$$
\begin{aligned}
d_{\mathcal{W}_2}^2\left((t_v)_{\#}\mu, \nu\right) &= \inf_{\pi \in \Pi(\mu, \nu)} \iint_{\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}} \|x - y\|^2 \, d\left((t_{(v,0)})_{\#}\pi\right)(x, y) \\
&= \inf_{\pi \in \Pi(\mu, \nu)} \iint_{\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}} \|x + v - y\|^2 \, d\pi(x, y) \\
&= \inf_{\pi \in \Pi(\mu, \nu)} \iint_{\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}} \left(\|x - y\|^2 + \|v\|^2 + 2\langle x, v \rangle - 2\langle y, v \rangle\right) \, d\pi(x, y) \\
&= d_{\mathcal{W}_2}^2(\mu, \nu) + \|v\|^2 + 2\int_{\mathbb{R}^{n+1}} \langle x, v \rangle \, d\mu(x) - 2\int_{\mathbb{R}^{n+1}} \langle y, v \rangle \, d\nu(y),
\end{aligned}
$$

which gives (4.1). The identity (4.2) follows if we translate both arguments in the left-hand side by the vector $m(\nu)$. $\square$

In quadratic Wasserstein spaces over uniquely geodesic spaces, the α-weighted mean squared-error function

$$\rho \mapsto (1 - \alpha)d_{\mathcal{W}_2}^2(\mu, \rho) + \alpha d_{\mathcal{W}_2}^2(\nu, \rho)$$

defined by μ and ν has a unique minimizer — provided that the optimal coupling of μ and ν is unique — which is the *displacement convex combination* or *displacement interpolation* of μ and ν with weights $(1 - \alpha)$ and α [25,26]. Intuitively, this is the measure that we obtain if we start moving μ to ν according to the optimal transport plan, but stop at proportion α of the journey. A great challenge concerning $(\mathbb{S}^n, \|\cdot\|)$ is that it has no geodesics at all, and hence the quadratic Wasserstein space $\mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$ has no geodesics either. Still, mean squared-error functions make perfect sense on $\mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$, they are invariant under isometries in an appropriate sense, and hence if the measures μ and ν defining them are fixed by an isometry Φ, then so are

the unique minimizers — if they exist. We will prove in Proposition 4.2 that on $(\mathbb{S}^n, \|\cdot\|)$, the minimizer of the α-weighted squared-error function is the projection of the displacement interpolation onto the sphere. This is similar to how for a measure $\mu \in \mathcal{P}(\mathbb{S}^n)$, its closest Dirac measure supported on a point in $\mathbb{R}^{n+1}$ is $\delta_{m(\mu)}$, while among those supported on $\mathbb{S}^n$, it is the projection of $\delta_{m(\mu)}$. We are going to exploit this characterisation in Step 6 of the proof of Theorem 1.1.

We will use the following projection of the α-weighted mean of two points x, y onto $\mathbb{S}^n$ frequently:

$$p_\alpha(x, y) := \frac{(1-\alpha)x + \alpha y}{\|(1-\alpha)x + \alpha y\|} \qquad (\alpha \in [0, 1],\ x, y \in \mathbb{S}^n).$$

Note that $p_\alpha(x, y)$ is not defined when $\alpha = \frac{1}{2}$ and $x = -y$.

Let us define the cost $c_\alpha \colon \mathbb{S}^n \times \mathbb{S}^n \to [0, 2]$ by

$$c_\alpha(x, y) := \min_{z \in \mathbb{S}^n} \left\{ (1-\alpha)\|x - z\|^2 + \alpha\|z - y\|^2 \right\} = 2(1 - \|(1-\alpha)x + \alpha y\|). \tag{4.3}$$

If $\pi \in \mathcal{P}(\mathbb{S}^n \times \mathbb{S}^n)$ is a coupling of μ and ν, then let $\rho_\pi^{(\alpha)} \in \mathcal{W}_2(\mathbb{S}^n, \|\cdot\|)$ be defined to be the displacement interpolation in $\mathbb{R}^{n+1}$ at time α between μ and ν according to the plan π, projected to $\mathbb{S}^n$. Formally,

$$\rho_\pi^{(\alpha)} := (\{(x, y) \mapsto p_\alpha(x, y)\})_\# \pi. \tag{4.4}$$

Now let $\mu, \nu \in \mathcal{W}_2(\mathbb{S}^n, \|\cdot\|)$ and consider the α-weighted mean squared error

$$Q_\alpha^{\mu,\nu} \colon \mathcal{W}_2(\mathbb{S}^n, \|\cdot\|) \to [0, \infty);\ \rho \mapsto Q_\alpha^{\mu,\nu}(\rho) := (1-\alpha)d_{\mathcal{W}_2}^2(\mu, \rho) + \alpha d_{\mathcal{W}_2}^2(\nu, \rho). \tag{4.5}$$

Proposition 4.2. *Let $\mu, \nu \in \mathcal{W}_2(\mathbb{S}^n, \|\cdot\|)$ be such that there is a unique optimal transport plan π_* for them with respect to the cost c_α defined in (4.3). Suppose that $\alpha \neq \frac{1}{2}$ or $\alpha = \frac{1}{2}$ and $\pi_*(\{(z, -z) : z \in \mathbb{S}^n\}) = 0$. Then the mean squared error $Q_\alpha^{\mu,\nu}$ defined in (4.5) has a unique minimizer which is equal to $\rho_{\pi_*}^{(\alpha)}$, the push-forward of π_* by p_α — see (4.4) for the precise definition. If $\pi_*(\{(z, -z) : z \in \mathbb{S}^n\}) > 0$ and $\alpha = \frac{1}{2}$, then $\rho_{\pi_*}^{(\alpha)} = \rho_{\pi_*}^{(\frac{1}{2})}$ is not well-defined and $Q_\alpha^{\mu,\nu} = Q_{\frac{1}{2}}^{\mu,\nu}$ has infinitely many minimizers.*

Proof. We proceed by establishing a lower bound for (4.5) and taking care of the case of equality. Let $\rho \in \mathcal{W}_2(\mathbb{S}^n, \|\cdot\|)$ be arbitrary, and let $\pi_{\mu,\rho}$ and $\pi_{\rho,\nu}$ be optimal transport plans (w.r.t. the quadratic distance) between μ and ρ, and ρ and ν, respectively. Let $\pi_{\mu,\rho,\nu} \in \mathcal{P}(\mathbb{S}^n \times \mathbb{S}^n \times \mathbb{S}^n)$ be the *gluing* of $\pi_{\mu,\rho}$ and $\pi_{\rho,\nu}$ — see [25, Lemma 7.6] for the precise definition. Then $\pi_{\mu,\nu} := (\pi_{\mu,\rho,\nu})_{1,3} \in \mathcal{P}(\mathbb{S}^n \times \mathbb{S}^n)$ is a coupling of μ and ν. Now

$$\begin{aligned}
Q_\alpha^{\mu,\nu}(\rho) &= (1-\alpha)d_{\mathcal{W}_2}^2(\mu, \rho) + \alpha d_{\mathcal{W}_2}^2(\nu, \rho) \\
&= (1-\alpha) \iint_{\mathbb{S}^n \times \mathbb{S}^n} \|x - z\|^2\, d\pi_{\mu,\rho}(x, z) + \alpha \iint_{\mathbb{S}^n \times \mathbb{S}^n} \|z - y\|^2\, d\pi_{\rho,\nu}(z, y) \\
&= \iiint_{\mathbb{S}^n \times \mathbb{S}^n \times \mathbb{S}^n} (1-\alpha)\|x - z\|^2 + \alpha\|z - y\|^2\, d\pi_{\mu,\rho,\nu}(x, z, y) \\
&\geq \iiint_{\mathbb{S}^n \times \mathbb{S}^n \times \mathbb{S}^n} c_\alpha(x, y)\, d\pi_{\mu,\rho,\nu}(x, z, y) = \iint_{\mathbb{S}^n \times \mathbb{S}^n} c_\alpha(x, y)\, d\pi_{\mu,\nu}(x, y).
\end{aligned} \tag{4.6}$$

The inequality (4.6) is saturated if and only if $z = p_\alpha(x, y)$ for $\pi_{\mu,\rho,\nu}$-a.e. $(x, z, y) \in (\mathbb{S}^n)^3$, that is, if $\rho = \rho_{\pi_{\mu,\nu}}^{(\alpha)}$. Moreover, the right-hand side of (4.6) is minimal if and only if $\pi_{\mu,\nu} = \pi_*$. Consequently,

$$Q_\alpha^{\mu,\nu}(\rho) \geq \min_{\pi \in \Pi(\mu,\nu)} \left\{ \iint_{\mathbb{S}^n \times \mathbb{S}^n} c_\alpha(x,y) \, \mathrm{d}\pi(x,y) \right\} = \iint_{\mathbb{S}^n \times \mathbb{S}^n} c_\alpha(x,y) \, \mathrm{d}\pi_*(x,y)$$

and the only ρ realizing this minimum is $\rho_{\pi_*}^{(\alpha)}$. On the other hand, if π_* puts weight on antipodal points, that is, $\pi_* (\{(z,-z) : z \in \mathbb{S}^n\}) > 0$, and $\alpha = \frac{1}{2}$, then we have an infinite collection of minimizing measures by the theorems of Thales and Pythagoras — or by a simple direct computation. $\quad\square$

In the next proposition, we consider the case when the first argument of p_α is fixed and clarify injectivity and surjectivity properties of p_α as α varies from 0 to 1.

Proposition 4.3. *Let $\alpha \in \left(0,\frac{1}{2}\right) \cup \left(\frac{1}{2},1\right]$ and let $N \in \mathbb{S}^n$ be arbitrary but fixed — it may be considered as the "north pole". Let $p_\alpha(N,\cdot) \colon \mathbb{S}^n \to \mathbb{S}^n$ be the map sending u to*

$$p_\alpha(N,u) = \frac{(1-\alpha)N + \alpha u}{\|(1-\alpha)N + \alpha u\|}.$$

Then for $\alpha \in \left(\frac{1}{2},1\right]$, $p_\alpha(N,\cdot)$ is bijective. For $\alpha \in \left(0,\frac{1}{2}\right)$, $p_\alpha(N,\cdot)$ is neither surjective nor injective: it is 2-to-1 for almost all points of $\mathbb{S}^n$. Finally, $p_{\frac{1}{2}}(N,\cdot) \colon \mathbb{S}^n \setminus \{-N\} \to \mathbb{S}^n \setminus \{-N\}$ is injective, and its range is the open "upper" hemisphere $\{z \in \mathbb{S}^n \mid \langle z,N \rangle > 0\}$.

Proof. When considering $p_\alpha(N,u)$, we can assume without loss of generality that $N = (0,0,\dots,0,1)$ and $u = (\cos\theta, 0, \dots, 0, \sin\theta)$ for some $\theta \in (-\pi, \pi]$. Let $c_u^{(\alpha)}$ be the normalising constant $\|(1-\alpha)N + \alpha u\|$. Note that $c_u^{(\alpha)} > 0$ if and only if $(\alpha, u) \neq \left(\frac{1}{2}, -N\right)$. Whenever $(\alpha, u) \neq \left(\frac{1}{2}, -N\right)$, we have that

$$c_u^{(\alpha)} p_u(N,u) = (\alpha \cos\theta, 0, \dots, 0, (1-\alpha) + \alpha \sin\theta),$$

and so for a fixed $\alpha \neq \frac{1}{2}$, setting

$$x_\theta := \alpha \cos\theta$$

$$y_\theta := (1-\alpha) + \alpha \sin\theta,$$

we see that (x_θ, y_θ) satisfy $x_\theta^2 + (y_\theta - (1-\alpha))^2 = \alpha^2$, i.e., they lie on the circle of radius α with the centre at $(0, 1-\alpha)$. For any point $u = (\cos\theta, 0, \dots, 0, \sin\theta)$, its image $p_\alpha(N,u)$ is the projection of $(x_\theta, 0, \dots, 0, y_\theta)$ onto $\mathbb{S}^n$, i.e., the point obtained as the intersection of $\mathbb{S}^n$ and the half-line from $(0, \dots, 0)$ through $(x_\theta, 0, \dots, 0, y_\theta)$. Now the statements of this proposition are easy to see from Fig. 1. $\quad\square$

Now we turn to the proof of Theorem 1.1 which, for the sake of clarity, we divide into six steps.

Step 1. Similarly as in the proof of Theorem 3.1, we first understand the action of Ψ on the set of Dirac measures. The maximal distance in $\mathcal{W}_2 (\mathbb{S}^n, \|\cdot\|)$ is 2, and is attained only on pairs of Dirac measures that are concentrated on antipodal points. Since

$$2 = d_{\mathcal{W}_2}(\mu,\nu) = d_{\mathcal{W}_2}(\Psi(\mu), \Psi(\nu)),$$

we get that $\Psi(\delta_x)$ is a Dirac measure for all $x \in \mathbb{S}^n$. Since Ψ and Ψ^{-1} are both isometries, the map $\psi \colon \mathbb{S}^n \to \mathbb{S}^n$ defined by $\Psi(\delta_x) = \delta_{\psi(x)}$ is a bijection, and furthermore, since

$$\|\psi(x) - \psi(y)\| = d_{\mathcal{W}_2} \left(\delta_{\psi(x)}, \delta_{\psi(y)} \right) = d_{\mathcal{W}_2} \left(\Psi(\delta_x), \Psi(\delta_y) \right) = d_{\mathcal{W}_2} \left(\delta_x, \delta_y \right) = \|x - y\|,$$
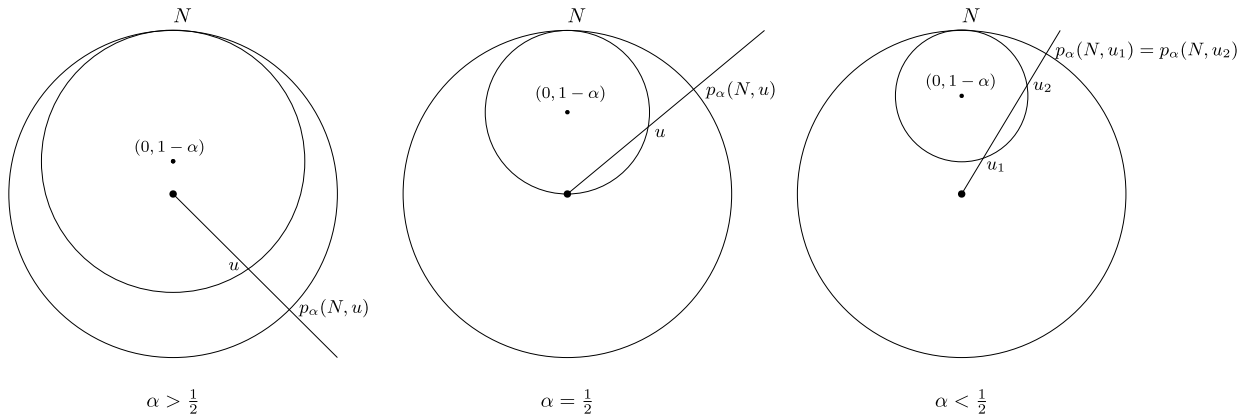
Fig. 1. $p_\alpha(N, u)$ lies on the $\mathbb{S}^1$ spanned by N and u, at the spherical projection of $c_u^{(\alpha)} p_\alpha(N, u)$. The bigger circle displayed in each of the cases is $\mathbb{S}^1$, the smaller one is $c_\cdot^{(\alpha)} p_\alpha(N, \cdot)$.

it is in fact an isometry. Just as before, we will be done once we prove that an isometry of $\mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$ which fixes all Dirac measures is necessarily the identity, because then in particular, $\left(\psi^{-1}\right)_\# \circ \Psi = \mathrm{id}_{\mathcal{W}_2(\mathbb{S}^n, \|\cdot\|)}$, and so $\Psi = \psi_\#$ as claimed. From now on, we assume that Φ is an isometry of $\mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$ such that $\Phi(\delta_x) = \delta_x$, and our aim is to show that $\Phi(\mu) = \mu$ for all $\mu \in \mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$.

Step 2. Next we claim that Φ preserves the barycenter of measures.

For any $\mu \in \mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$ and $x \in \mathbb{S}^n$, we have

$$d_{\mathcal{W}_2}^2(\mu, \delta_x) = \int_{\mathbb{S}^n} \|y - x\|^2 \mathrm{d}\mu(y) = 2 - 2\left\langle x, \int_{\mathbb{S}^n} y\, \mathrm{d}\mu(y)\right\rangle = 2(1 - \langle x, m(\mu)\rangle). \qquad (4.7)$$

This implies that

$$2(1 - \langle x, m(\mu)\rangle) = d_{\mathcal{W}_2}^2(\mu, \delta_x) = d_{\mathcal{W}_2}^2(\Phi(\mu), \Phi(\delta_x)) = d_{\mathcal{W}_2}^2(\Phi(\mu), \delta_x) = 2\left(1 - \langle x, m(\Phi(\mu))\rangle\right).$$

But then

$$\langle x, m(\Phi(\mu))\rangle = \langle x, m(\mu)\rangle$$

and hence

$$\langle x, m(\Phi(\mu)) - m(\mu)\rangle = 0. \qquad (4.8)$$

Since for a fixed μ, equation (4.8) holds for every $x \in \mathbb{S}^n$, we conclude that $m\left(\Phi\left(\mu\right)\right) = m\left(\mu\right)$.

Step 3. Now we prove that measures supported on two points are mapped to measures supported on two points. We first show that for all $\mu, \nu \in \mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$,

$$\mathrm{affspan}\left(\mathrm{supp}\left(\Phi(\mu)\right)\right) \perp \mathrm{affspan}\left(\mathrm{supp}\left(\Phi(\nu)\right)\right)$$

holds if and only if

$$\mathrm{affspan}\left(\mathrm{supp}(\mu)\right) \perp \mathrm{affspan}\left(\mathrm{supp}(\nu)\right).$$

Kloeckner proved in [18, Lemma 6.2] that orthogonality of supports can be characterized by the metric in the ambient space $\mathcal{W}_2\left(\mathbb{R}^{n+1}, \|\cdot\|\right)$. Namely,

$$d^2_{\mathcal{W}_2(\mathbb{R}^{n+1})}(\mu, \nu) = \|m(\mu) - m(\nu)\|^2 + d^2_{\mathcal{W}_2(\mathbb{R}^{n+1})}\left(\mu, \delta_{m(\mu)}\right) + d^2_{\mathcal{W}_2(\mathbb{R}^{n+1})}\left(\nu, \delta_{m(\nu)}\right)$$

holds if and only if there exist two orthogonal affine subspaces $L, M \subset \mathbb{R}^{n+1}$ such that $\operatorname{supp}(\mu) \subseteq L$ and $\operatorname{supp}(\nu) \subseteq M$. We proceed by showing that the isometries of $\mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$ leave the $\mathcal{W}_2(\mathbb{R}^{n+1})$-distance of a measure from the Dirac mass concentrated on its barycenter invariant, that is,

$$d_{\mathcal{W}_2(\mathbb{R}^{n+1})}\left(\mu, \delta_{m(\mu)}\right) = d_{\mathcal{W}_2(\mathbb{R}^{n+1})}\left(\Phi(\mu), \delta_{m(\Phi(\mu))}\right)$$

for any $\mu \in \mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$. Indeed, a direct computation very similar to (4.7) shows that

$$d^2_{\mathcal{W}_2(\mathbb{R}^{n+1})}\left(\mu, \delta_{m(\mu)}\right) = 1 - \|m(\mu)\|^2 \text{ and } d^2_{\mathcal{W}_2(\mathbb{R}^{n+1})}\left(\Phi(\mu), \delta_{m(\Phi(\mu))}\right) = 1 - \|m(\Phi(\mu))\|^2,$$

which implies our statement as we have shown $m(\Phi(\mu)) = m(\mu)$ in Step 2. Hence for any $\mu, \nu \in \mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$,

$$\|m(\mu) - m(\nu)\|^2 + d^2_{\mathcal{W}_2}\left(\mu, \delta_{m(\mu)}\right) + d^2_{\mathcal{W}_2}\left(\nu, \delta_{m(\nu)}\right)$$
$$= \|m(\Phi(\mu)) - m(\Phi(\nu))\|^2 + d^2_{\mathcal{W}_2}\left(\Phi(\mu), \delta_{m(\Phi(\mu))}\right) + d^2_{\mathcal{W}_2}\left(\Phi(\nu), \delta_{m(\Phi(\nu))}\right),$$

meaning that orthogonally supported measures must be mapped to orthogonally supported measures by Φ.

A maximal set of measures whose supports are one-dimensional and pairwise orthogonal must therefore be mapped to a set of measures whose supports are zero- or one-dimensional. But zero-dimensionally supported measures are exactly the Dirac masses, to which only Dirac masses can be mapped by Φ, and so one-dimensionally supported measures must be mapped to one-dimensionally supported measures. Continuing similarly, we would see more generally that the affine dimension of the support is preserved by Φ, but since on the sphere, one-dimensionally supported measures are exactly the two-point supported measures, the one-dimensional case is enough to prove our statement.

Step 4. We proceed with showing that measures supported on two points are fixed by Φ. Let us introduce the notation $\Delta_2'(\mathbb{S}^n)$ for the set of all elements in $\mathcal{W}_2\left(\mathbb{S}^n, \|\cdot\|\right)$ with a two-point support, set $\widetilde{\mu} := \left(t_{-m(\mu)}\right)_{\#}\mu$ for all $\mu \in \Delta_2'(\mathbb{S}^n)$, and $\Delta_{2,0}'(\mathbb{S}^n) := \left\{\widetilde{\mu} \in \mathcal{P}(\mathbb{R}^{n+1}) : \mu \in \Delta_2'(\mathbb{S}^n)\right\}$. By Step 3, $\Phi|_{\Delta_2'(\mathbb{S}^n)} \colon \Delta_2'(\mathbb{S}^n) \to \Delta_2'(\mathbb{S}^n)$ is an isometric embedding. By Proposition 4.1 we know that for all $\mu, \nu \in \Delta_2'(\mathbb{S}^n)$,

$$d^2_{\mathcal{W}_2}\left(\widetilde{\mu}, \widetilde{\nu}\right) = d^2_{\mathcal{W}_2}\left(\mu, \nu\right) - \|m(\mu) - m(\nu)\|^2$$
$$= d^2_{\mathcal{W}_2}\left(\Phi(\mu), \Phi(\nu)\right) - \|m(\Phi(\mu)) - m(\Phi(\nu))\|^2 = d^2_{\mathcal{W}_2}\left(\widetilde{\Phi(\mu)}, \widetilde{\Phi(\nu)}\right).$$

Consequently, $\widetilde{\mu} = \widetilde{\nu}$ holds if and only if $\widetilde{\Phi(\mu)} = \widetilde{\Phi(\nu)}$, in other words, $\Phi(\nu)$ is a translate of $\Phi(\mu)$ if and only if ν is a translate of μ.

Let a measure $\mu \in \Delta_2'(\mathbb{S}^n)$ be fixed. We can assume without loss of generality that

$$\operatorname{supp}(\mu) = \{(\cos\theta, 0, \ldots, 0, \sin\theta), (\cos\theta, 0, \ldots, 0, -\sin\theta)\}$$

for some $\theta \in (0, \pi/2]$. In this case,

$$\frac{1}{2}\sum_{x \in \operatorname{supp}(\mu)} x = (\cos\theta, 0, \ldots, 0)$$

and

$$\left(\text{affspan}\left(\text{supp}(\mu)\right) - \frac{1}{2} \sum_{x \in \text{supp}(\mu)} x \right)^{\perp} = \{(v_1, \ldots, v_n, 0) : v_1, \ldots, v_n \in \mathbb{R}\}.$$

Define $\vec{\mu} := \{v \in \mathbb{R}^{n+1} : (t_v)_{\#}\mu \in \mathcal{P}(\mathbb{S}^n)\}$, and observe that $\vec{\mu}$ is the set of those vectors $(v_1, \ldots, v_{n+1})$ such that

$$\|(v_1 + \cos\theta, v_2, \ldots, v_n, v_{n+1} + \sin\theta)\| = \|(v_1 + \cos\theta, v_2, \ldots, v_n, v_{n+1} - \sin\theta)\| = 1.$$

Then we get that $v \in \vec{\mu}$ if and only if

$$(v_1 + \cos\theta)^2 + v_2^2 + \cdots + v_n^2 + v_{n+1}^2 \pm 2v_{n+1}\sin\theta = 1 - \sin^2\theta.$$

Since $\sin\theta \neq 0$, this holds exactly when $v_{n+1} = 0$ and $(v_1 + \cos\theta)^2 + v_2^2 + \cdots + v_n^2 = \cos^2\theta$, i.e., the first n coordinates span an $n-1$-dimensional sphere with radius $|\cos\theta|$ centred at $(-\cos\theta, 0, \ldots, 0)$, or they are just the singleton containing $0 \in \mathbb{R}^n$ in the case $\cos\theta = 0$. In other words,

$$\vec{\mu} = -(\cos\theta, 0, \ldots, 0) + |\cos\theta| \cdot (\mathbb{S}^n \cap \{(v_1, \ldots, v_n, 0) : v_1, \ldots, v_n \in \mathbb{R}\})$$

$$= -\left(\frac{1}{2} \sum_{x \in \text{supp}(\mu)} x \right) + \left\| \frac{1}{2} \sum_{x \in \text{supp}(\mu)} x \right\| \cdot \left(\mathbb{S}^n \cap \left(\text{affspan}\left(\text{supp}(\mu)\right) - \frac{1}{2} \sum_{x \in \text{supp}(\mu)} x \right)^{\perp} \right).$$

As Φ maps the translates of μ to the translates of $\Phi(\mu)$, there is an $\eta \in \Delta'_{2,0}(\mathbb{S}^n) \subset \mathcal{P}(\mathbb{R}^{n+1})$ such that

$$\Phi\left((t_v)_{\#}\mu\right) = (t_{v+m(\mu)})_{\#}\eta \qquad (v \in \vec{\mu}). \tag{4.9}$$

We emphasise that η does not depend on v. It follows that $\vec{\mu} + m(\mu) + \text{supp}(\eta) \subset \mathbb{S}^n$. But by plugging $v = 0 \in \mathbb{R}^{n+1}$ to (4.9), we get that $\text{supp}\left(\Phi(\mu)\right) = m(\mu) + \text{supp}(\eta)$, and so the previous line becomes $\vec{\mu} + \text{supp}\left(\Phi(\mu)\right) \subset \mathbb{S}^n$. By the definition of $\overrightarrow{\Phi(\mu)}$, this means that $\vec{\mu} \subseteq \overrightarrow{\Phi(\mu)}$.

For any μ with $\text{diam}(\text{supp}(\mu)) < 2$, we have that $\cos\theta \neq 0$, and so $\vec{\mu}$ is an $n-1$-dimensional sphere, implying that $\vec{\mu} = \overrightarrow{\Phi(\mu)}$ and $\text{supp}(\mu) = \text{supp}\left(\Phi(\mu)\right)$. Now μ and $\Phi(\mu)$ are probability measures with the same 2-point support and the same barycenter, and so $\mu = \Phi(\mu)$. Finally, $\mu = \Phi(\mu)$ for all $\mu \in \Delta'_2(\mathbb{S}^n)$ by continuity of Φ.

Step 5. Now assume that $\mu = \sum_{i=1}^{m} \lambda_i \delta_{x_i}$ where $x_i \neq -x_j$ for all $1 \leq i < j \leq m$. Such measures form a dense subset of $\mathcal{W}_2(\mathbb{S}^n, \|\cdot\|)$. We claim that

$$\text{supp}\left(\Phi(\mu)\right) \subseteq \{x_1, \ldots, x_m\} \cup \{-x_1, \ldots, -x_m\}$$

and $\left(\Phi(\mu)\right)(\{x_i, -x_i\}) = \mu(\{x_i\})$ for all $1 \leq i \leq m$.

The proof of this claim relies on preserving the mass of *bisectors* which are defined as follows: for $u, v \in \mathbb{S}^n$, the corresponding bisector is

$$B(u, v) := \{y \in \mathbb{S}^n : \|u - y\| = \|v - y\|\} \cong \mathbb{S}^{n-1}.$$

To start, we apply Lemma 3.17 from [14] with $E = \mathbb{R}^{n+1}$, $p = 2$, $x \in \mathbb{S}^n$, $a = 1$ and $b = -1$ to obtain that

$$\mu\left(B(x, -x)\right) = \max\left\{\alpha : d_{\mathcal{W}_2}\left(\mu, \alpha\delta_x + (1-\alpha)\delta_{-x}\right) = m_\mu\right\} - \min\left\{\alpha : d_{\mathcal{W}_2}\left(\mu, \alpha\delta_x + (1-\alpha)\delta_{-x}\right) = m_\mu\right\},$$

where $m_\mu := \min\{d_{\mathcal{W}_2}(\mu, \alpha\delta_x + (1-\alpha)\delta_{-x}) : 0 \le \alpha \le 1\}$ and $B(x, -x) \cong \mathbb{S}^{n-1}$ is the bisector between x and $-x$, i.e., the set of all points equidistant from x and $-x$. But since $\Phi(\alpha\delta_x + (1-\alpha)\delta_{-x}) = \alpha\delta_x + (1-\alpha)\delta_{-x}$ for all $\alpha \in [0,1]$ by Steps 1 and 4, we get that $m_\mu = m_{\Phi(\mu)}$, and subsequently $\mu(B(x, -x)) = (\Phi(\mu))(B(x, -x))$ for all $x \in \mathbb{S}^n$. Since for every $x \in \mathbb{S}^n$, $B(x, -x)$ is an $n-1$-dimensional subsphere of $\mathbb{S}^n$, and every $n-1$-dimensional subsphere of $\mathbb{S}^n$ is of the form $B(x, -x)$ for some $x \in \mathbb{S}^n$, the previous sentence says that $\mu(S) = (\Phi(\mu))(S)$ for every subsphere S of codimension 1.

For every $\widetilde{x} \in \{x_1, \ldots, x_m\} = \mathrm{supp}(\mu)$, there exists a sequence $(S_j)_{j\in\mathbb{N}}$ of $n-1$-dimensional subspheres of $\mathbb{S}^n$ such that $S_j \cap \mathrm{supp}(\mu) = \{\widetilde{x}\}$ for every j, and the intersection of any n subspheres is trivial, that is, $\bigcap_{k=1}^n S_{j_k} = \{\widetilde{x}, -\widetilde{x}\}$ for any choice of $j_1 < j_2 < \cdots < j_n$. Therefore,

$$\mu(\{\widetilde{x}\}) = \mu(S_j) = (\Phi(\mu))(S_j) \qquad (j \in \mathbb{N}),$$

and we are in the right position to prove that

$$\Phi(\mu)(\{\widetilde{x}, -\widetilde{x}\}) = \mu(\{\widetilde{x}\}).$$

The inequality $\Phi(\mu)(\{\widetilde{x}, -\widetilde{x}\}) \le (\Phi(\mu))(S_j) = \mu(\{\widetilde{x}\})$ holds because $\{\widetilde{x}, -\widetilde{x}\} \subseteq S_j$. If $n = 1$, then in fact $\{\widetilde{x}, -\widetilde{x}\} = S_j$, and we are done. If $n \ge 2$, assume indirectly that $\Phi(\mu)(\{\widetilde{x}, -\widetilde{x}\}) < \mu(\{\widetilde{x}\})$, and let $\varepsilon > 0$ denote the gap between the two sides of this strict inequality. Now we have

$$(\Phi(\mu))(S_j \setminus \{\widetilde{x}, -\widetilde{x}\}) = (\Phi(\mu))(S_j) - \Phi(\mu)(\{\widetilde{x}, -\widetilde{x}\}) = \varepsilon$$

for every $j \in \mathbb{N}$. The fact that $\bigcap_{k=1}^n S_{j_k} = \{\widetilde{x}, -\widetilde{x}\}$ for every $j_1 < j_2 < \cdots < j_n$ implies that the family of sets $(S_j \setminus \{\widetilde{x}, -\widetilde{x}\})_{j=1}^\infty$ covers any point of $\mathbb{S}^n$ at most n times. This means that $\sum_{j=1}^\infty (\Phi(\mu))(S_j \setminus \{\widetilde{x}, -\widetilde{x}\})$ is bounded from above by $n \cdot (\Phi(\mu))(\mathbb{S}^n) = n$, which is a contradiction as $(\Phi(\mu))(S_j \setminus \{\widetilde{x}, -\widetilde{x}\}) = \varepsilon$ for every $j \in \mathbb{N}$ and $\sum_{j=1}^\infty \varepsilon = \infty$.

Step 6. A crucial consequence of the claim made in Step 5 is that the isometry Φ fixes all measures that are supported within an open hemisphere of $\mathbb{S}^n$. Indeed, we learned from Step 2 that Φ preserves the barycenter of measures, and from Step 5 that the only possible action Φ can do is to send some mass from a point to its antipodal point. But if a measure is supported on an open hemisphere, then the transport of any mass to its antipodal point would change the barycenter.

Suppose that $\mu(\{-N\}) = 0$. Then the spherical projection $\rho_{\delta_N \otimes \mu}^{(\frac{1}{2})}$ of the displacement convex combination of δ_N and μ is well-defined, and since it is supported on the upper hemisphere, $\Phi\left(\rho_{\delta_N \otimes \mu}^{(\frac{1}{2})}\right) = \rho_{\delta_N \otimes \mu}^{(\frac{1}{2})}$. Let us now consider the sets

$$A := \left\{ Q_{\frac{1}{2}}^{\delta_N, \mu}(\rho) : \rho \in \mathcal{P}(\mathbb{S}^n) \right\}$$

$$= \left\{ Q_{\frac{1}{2}}^{\Phi(\delta_N), \Phi(\mu)}(\Phi(\rho)) : \rho \in \mathcal{P}(\mathbb{S}^n) \right\}$$

$$= \left\{ Q_{\frac{1}{2}}^{\delta_N, \Phi(\mu)}(\kappa) : \kappa \in \mathcal{P}(\mathbb{S}^n) \right\} =: B$$

where the last equality follows from the surjectivity of the isometry Φ. Since $A = B$, necessarily

$$\min B = \min A = Q_{\frac{1}{2}}^{\delta_N, \mu}\left(\rho_{\delta_N \otimes \mu}^{(\frac{1}{2})}\right) = Q_{\frac{1}{2}}^{\Phi(\delta_N), \Phi(\mu)}\left(\Phi\left(\rho_{\delta_N \otimes \mu}^{(\frac{1}{2})}\right)\right) = Q_{\frac{1}{2}}^{\delta_N, \Phi(\mu)}\left(\Phi\left(\rho_{\delta_N \otimes \mu}^{(\frac{1}{2})}\right)\right).$$

The fact that B has a unique minimizer implies by the second statement of Proposition 4.2 that $\Phi(\mu)(\{-N\}) = 0$. Consequently — let us now use the first statement of Proposition 4.2 —, the unique minimizer of $Q_\alpha^{\delta_N, \Phi(\mu)}$ is $\rho_{\delta_N \otimes \Phi(\mu)}^{(\frac{1}{2})}$, and hence

$$\rho^{(\frac{1}{2})}_{\delta_N \otimes \Phi(\mu)} = \Phi\left(\rho^{(\frac{1}{2})}_{\delta_N \otimes \mu}\right) = \rho^{(\frac{1}{2})}_{\delta_N \otimes \mu}. \tag{4.10}$$

By injectivity of $p_{\frac{1}{2}}(N, \cdot)$ on $\mathbb{S}^n \setminus \{-N\}$, see Proposition 4.3, for every measure $\nu \in \mathcal{P}(\mathbb{S}^n)$ supported within the upper hemisphere, there is a unique measure $\kappa \in \mathcal{P}(\mathbb{S}^n)$ such that ν is the spherical projection $\rho^{(\frac{1}{2})}_{\delta_N \otimes \kappa}$ of the displacement convex combination of δ_N and κ. Therefore, (4.10) implies that $\Phi(\mu) = \mu$, which completes the proof.

Acknowledgments

References

[1] L. Ambrosio, N. Gigli, A User's Guide to Optimal Transport. Modelling and Optimisation of Flows on Networks, Lecture Notes in Math., vol. 2062, Springer, Heidelberg, 2013, Fond. CIME/CIME Found. Subser.

[2] Z.M. Balogh, T. Titkos, D. Virosztek, Isometries and isometric embeddings of Wasserstein spaces over the Heisenberg group, 2023, arXiv:2303.15095, manuscript.

[3] Z.M. Balogh, T. Titkos, D. Virosztek, Isometric rigidity of the Wasserstein space $\mathcal{W}_1(\mathbf{G})$ over Carnot groups, 2023, arXiv:2305.05492, manuscript.

[4] J. Bertrand, B. Kloeckner, A geometric study of Wasserstein spaces: isometric rigidity in negative curvature, Int. Math. Res. Not. 5 (2016) 1368–1386.

[5] D. Burago, Y. Burago, S. Ivanov, A Course in Metric Geometry, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, Providence, RI, 2001.

[6] G. Dolinar, B. Kuzma, D. Mitrovic, Isometries of probability measures with respect to the total variation distance, J. Math. Anal. Appl. 507 (2021) 125829.

[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] A. Figalli, F. Glaudo, An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows, EMS Textbooks in Mathematics, EMS Press, Berlin, 2021.

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

[10] Gy.P. Gehér, T. Titkos, A characterisation of isometries with respect to the Lévy-Prokhorov metric, Ann. Sc. Norm. Super. Pisa, Cl. Sci. XIX (2019) 655–677.

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

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

[13] Gy.P. Gehér, T. Titkos, D. Virosztek, Dirac masses and isometric rigidity, RIMS Kokyuroku 2125 (2020) 34–41.

[14] Gy.P. Gehér, T. Titkos, D. Virosztek, The isometry group of Wasserstein spaces: the Hilbertian case, J. Lond. Math. Soc. 106 (2022) 3865–3894.

[15] Gy.P. Gehér, T. Titkos, D. Virosztek, Isometric rigidity of Wasserstein tori and spheres, Mathematika 69 (2023) 20–32.

[16] Gy.P. Gehér, T. Titkos, D. Virosztek, On the exotic isometry flow of the quadratic Wasserstein space over the real line, Linear Algebra Appl. 693 (2024) 39–48.

[17] G. Kiss, T. Titkos, Isometric rigidity of Wasserstein spaces: the graph metric case, Proc. Am. Math. Soc. 150 (2022) 4083–4097.

[18] B. Kloeckner, A geometric study of Wasserstein spaces: Euclidean spaces, Ann. Sc. Norm. Super. Pisa, Cl. Sci. (2010) 297–323.

[19] L. Molnár, Kolmogorov-Smirnov isometries and affine automorphisms of spaces of distribution functions, Cent. Eur. J. Math. 9 (2011) 789–796.

[20] L. Molnár, P. Szokol, Kolmogorov-Smirnov isometries of the space of generalized distribution functions, Math. Slovaca 64 (2014) 433–444.

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

[22] W. Rudin, Fourier Analysis on Groups, Interscience Tracts in Pure and Applied Mathematics, vol. 12, Interscience Publishers, New York-London, 1962.

[23] F. Santambrogio, Optimal Transport for Applied Mathematicians, Progress in Nonlinear Differential Equations and Their Applications, vol. 87, Birkhauser, Basel, 2015.

[24] J. Santos-Rodríguez, On isometries of compact L^p–Wasserstein spaces, Adv. Math. 409 (2022) 108632.

[25] C. Villani, Topics in Optimal Transportation, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.

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

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