A geometric study of the Wasserstein space of the line

Benoit Kloeckner

A GEOMETRIC STUDY OF THE WASSERSTEIN SPACE OF THE LINE

BENOÎT KLOECKNER

1. Introduction

The concept of optimal transportation raised recently a growing interest in link with the geometry of metric spaces. In particular the L^2 Wasserstein space $W^2(X)$ have been used in [6] and [8, 9] to define certain curvature conditions on a metric space X. Many useful properties are inherited from X by $W^2(X)$ (separability, completeness, geodesicness, some non-negative curvature conditions) while some other are not, like the local compacity.

In this paper, we aim at starting a geometric study of Wasserstein spaces as intrinsic spaces. We are interested, for example, in the isometry group of $W^2(X)$, in its curvature and in its rank (the greatest possible dimension of a Euclidean space that embeds in it).

We shall concentrate on the case where X is the real line for several reason. First, it is arguably the simplest geodesic space, thus a natural choice to start our study. Second, it has some very specific features which play a rôle in our main result (for example, we shall see that $W^2(\mathbb{R})$ is one of the very few CAT(0) Wasserstein spaces). Last, since any complete simple geodesic in a metric space is isometric to the line, one can hope to deduce information on many Wasserstein spaces from the study of $W^2(\mathbb{R})$.

The Wasserstein space $W^2(X)$ contains an isometric embedding of X: $x \mapsto \delta_x$ where δ_x is the Dirac mass at x. Moreover any isometry of X acts isometrically on $W^2(X)$ in a natural way: $\phi(\mu)(A) = \mu(\phi^{-1}(A))$, giving an embedding $\operatorname{Isom} X \to \operatorname{Isom} W^2(X)$. These two elementary facts enable to connect the geometry of $W^2(X)$ to that of X.

Our main result concerns the isometries of $W^2(\mathbb{R})$. One could expect that the embedding $\operatorname{Isom} X \to \operatorname{Isom} W^2(X)$ is onto, *i.e.* that all isometries of $W^2(\mathbb{R})$ are induced by those of $\mathbb{R}$ itself. Surprisingly this happens to be false at least for the line.

Theorem 1.1 — *The isometry group of* $W^2(\mathbb{R})$ *is a semidirect product*

$$\operatorname{Isom}(\mathbb{R}) \rtimes \mathbb{R}$$

where the right factor is an isometric flow that fixes each Dirac mass.

The main tool we use is the explicit description of the geodesic between two points μ_0, μ_1 of $W^2(\mathbb{R})$ that follows from the fact that the unique optimal transportation plan between μ_0 and μ_1 is the non-decreasing rearrangement. It implies that most of the geodesics in $W^2(\mathbb{R})$ are not complete, and we rely on this fact to give a metric characterization of Dirac masses and of linear combinations of two Dirac masses, among all points of $W^2(X)$. We also use the fact that $W^2(\mathbb{R})$ has vanishing curvature in the sense of Alexandrov.

This "exotic" isometric flow tends to put all the mass on one side of the center of gravity (that must be preserved), close to it, and to send a small bit of mass far away on the other side (so that the Wasserstein distance to the center of mass is preserved). In particular, under this flow any measure μ converges weakly to δ_x (where x is the center of mass of μ), see Proposition 5.3.

Another consequence of the study of maximal geodesics concerns the rank of $W^2(\mathbb{R})$.

Theorem 1.2 — *There is no isometric, totally geodesic embedding of $\mathbb{R}^2$ into $W_2(\mathbb{R})$.*

It is simple to prove that despite Theorem 1.2, large pieces of $\mathbb{R}^n$ can be embedded into $W^2(\mathbb{R})$, which has consequently infinite weak ranks in a sense to be precised. As a consequence, we get for example:

Proposition 1.3 — *If X is any Polish geodesic metric space that contains a complete geodesic, then $W^2(X)$ is not δ-hyperbolic.*

This is not surprising, since it is well-known that the negative curvature assumptions tend not to be inherited from X by its Wasserstein space. For example, if X contains a lozenge (four distinct points x_1, x_2, x_3, x_4 so that $d(x_i, x_i + 1)$ is independent of the cyclic index i) then $W^2(X)$ is not uniquely geodesic, and in particular not CAT(0), even if X itself is strongly negatively curved.

Organization of the paper. We start in Section 2 by some recalls and notations. Section 3 is devoted to the study of maximal extensions of geodesic segments and convex hulls. Section 4 is devoted to the curvature of $W^2(\mathbb{R})$ and to an extension theorem that follows. In section 5 we prove Theorem 1.1 and describe the unique non trivial isometric flow on $W^2(\mathbb{R})$. We turn to the study of ranks in Section 6, and end with several open problems in Section 7.

Acknowledgements. I wish to thanks all speakers of the workshop on optimal transportation held in the Institut Fourier in Grenoble, especially Nicolas Juillet with whom I had numerous discussion on Wasserstein spaces, and its organizer Hervé Pajot.

2. The wasserstein space

In this preliminary section we recall general facts on $W^2(X)$. One can refer to [10] for further details and much more.

2.1. **Geodesic spaces.** Let X be a Polish (*i.e.* complete separable) space, and assume that X is geodesic, that is: between two points there is a rectifiable curve whose length is the distance between the considered points. Note that we only consider *globally* minimizing geodesics, and that a geodesic is always assumed to be parametrized proportionally to arc length.

One defines the Wasserstein space of X as the set $W^2(X)$ of Borel probability measures μ on X that satisfy

$$\int_X d^2(x_0, x)\mu(dx) < +\infty$$

for some (hence all) point $x_0 \in X$, equipped by the distance d_W defined by:

$$d_W^2(\mu_0, \mu_1) = \inf \int_{X\times X} d^2(x, y)\Pi(dx, dy)$$

where the infimum is taken over all coupling Π of μ_0, μ_1. A coupling realizing this infimum is said to be optimal, and there always exists an optimal coupling.

The idea behind this distance is linked to the Monge-Kantorovitch problem: given a unit quantity of goods distributed in X according to μ_0, what is the most economical way to displace them so that they end up distributed according to μ_1, when the cost to move a unit of good from x to y is given by $d^2(x, y)$? The minimal cost is $d_W^2(\mu_0, \mu_1)$ and a transportation plan achieving this minimum is an optimal coupling.

Under the assumptions we put on X, the metric space $W^2(X)$ is itself Polish and geodesic. If moreover X is uniquely geodesic, then to each optimal coupling Π between μ_0 and μ_1 is associated a unique geodesic in $W^2(X)$ in the following way. Let $C([0, 1], X)$ be the set of continuous curves $[0, 1] \to X$, let $g : X \times X \to C([0, 1], X)$ be the application that maps (x, y) to the unit speed geodesic between these points, and for each $t \in [0, 1]$ let $e^t : C([0, 1], X) \to X$ be the map $\gamma \mapsto \gamma(t)$. Then $t \mapsto e^t_\sharp g_\sharp \Pi$ is a geodesic between μ_0 and μ_1. Informally, this means that we choose randomly a couple (x, y) according to the joint law Π, then take the time t of the geodesic $g(x, y)$. This gives a random point in X, whose law is μ_t, the time t of the geodesic in $W^2(X)$ associated to the optimal coupling Π.

Note that for most spaces X, the optimal coupling between two probability measures is not unique, and $W^2(X)$ is therefore not uniquely geodesic even if X is.

Let us name the isometric and totally geodesic embedding of X into $W^2(X)$:

$$I : x \mapsto \delta_x$$

where δ_x is the Dirac mass at point x. One of our goal is to determine wether the Dirac measures can be detected inside $W^2(X)$ by purely geometric properties, so that we can link the isometries of $W^2(X)$ to those of X.

2.2. **The line.** Given the distribution function

$$F : x \mapsto \mu(]-\infty, x])$$

of a probability measure μ, one defines its left-continuous inverse:

$$\begin{array}{rcl} F^{-1} :]0,1[& \to & \mathbb{R} \\ m & \mapsto & \sup\{x \in \mathbb{R}\,;\, F(x) \leq m\} \end{array}$$

that is a non-decreasing, left-continuous function; $\lim_0 F^{-1}$ is the infimum of the support of μ and $\lim_1 F^{-1}$ its supremum. A discontinuity of F^{-1} happens for each interval that does not intersect the support of μ, and F^{-1} is constant on an interval for each atom of μ.

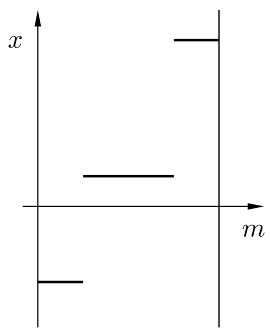

FIGURE 1. Inverse distribution function of a combination of three Dirac masses

Let μ_0 and μ_1 be two points of $W_2(\mathbb{R})$, and let F_0, F_1 be their repartition functions. Then the distance between μ_0 and μ_1 is given by

$$d^2(\mu_0, \mu_1) = \int_0^1 \left(F_0^{-1}(m) - F_1^{-1}(m)\right)^2 dm \tag{1}$$

and there is a unique constant speed geodesic $(\mu_t)_{t\in[0,1]}$, where μ_t has a distribution function F_t defined by

$$F_t^{-1} = (1-t)F_0^{-1} + tF_1^{-1} \tag{2}$$

This means that the best way to go from μ_0 to μ_1 is simply to rearrange increasingly the mass, a consequence of the convexity of the cost function. For example, if μ_0 and μ_1 where uniform measures on $[0,1]$ and $[\varepsilon, 1+\varepsilon]$, then the optimal coupling is deterministic given by the translation $x \mapsto x+\varepsilon$. That is: the best way to go from μ_0 to μ_1 is to shift every bit of mass by ε. If the cost function where linear, it would be equivalent to leave the mass on $[\varepsilon, 1]$ in place and move the remainder from $[0,\varepsilon]$ to $[1, 1+\varepsilon]$. If the cost function where concave, then the latter solution would be better than the former.

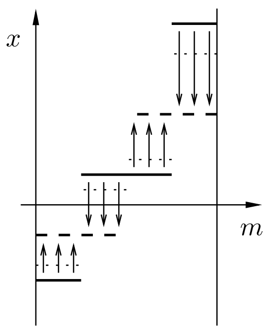

FIGURE 2. A geodesic between two atomic measures: the mass moves with speed proportional to the length of the arrows.

2.3. **Spaces of nonpositive curvature.** We shall consider two curvature conditions. The first one is a negative curvature condition, the δ-hyperbolicity introduced by Gromov. A geodesic space is said to be δ-hyperbolic (where δ is a non-negative number) if in any triangle, any point of any of the sides is at distance at most δ from one of the other two sides. For example, the real hyperbolic space is δ-hyperbolic (the value of δ depending on the value of the curvature), a tree is 0-hyperbolic and the euclidean spaces of dimension at least 2 are not δ-hyperbolic for any δ.

The second condition is the classical non-positive curvature condition CAT(0), detailed in Section 4, that roughly means that triangles are thinner in X than in the euclidean plane. Euclidean spaces, any Riemannian manifold having non-positive sectional curvature are examples of CAT(0) spaces.

A geodesic Polish space X is said to be a *Hadamard space* if it is simply connected and CAT(0). A Hadamard space is uniquely geodesic, and admits a natural boundary at infinity. The feature that interests us most is the following classical result: if X is a Hadamard space, given $\mu \in W^2(X)$ there is a unique point $x_0 \in X$, called the center of mass of μ, that minimizes the quantity $\int_X d^2(x_0, x)\mu(dx)$. If $X = \mathbb{R}^n$ endowed with the canonical scalar product, then the center of mass is of course $\int_{\mathbb{R}^n} x\mu(dx)$ but in the general case, the lack of an affine structure on X prevents to use this last formula.

We thus get a map $P : W^2(X) \to X$ that maps any L^2 probability measure to its center of mass. Obviously, P is a left inverse to I and one can hope to use this map to link closer the geometry of $W^2(X)$ to that of X. That's why our questions, unlike most of the classical ones in optimal transportation, might behave nicer when the curvature is non-positive than when it is non-negative.

3. Geodesics

3.1. **Maximal extension of geodesics.**

We now consider the geodesics of $W^2(\mathbb{R})$ to determine on which maximal interval they can be extended.

Let μ_0, μ_1 be two points of $W^2(\mathbb{R})$ and F_0, F_1 their distribution functions. Let $(\mu_t)_{t\in[0,1]}$ be the geodesic between μ_0 and μ_1. Since $W^2(\mathbb{R})$ is uniquely geodesic, there is a unique maximal interval on which γ can be extended into a geodesic, denoted by $\iota(\mu_0, \mu_1)$.

Lemma 3.1 — *One has*

$$\iota(\mu_0, \mu_1) = \{t \in \mathbb{R}\,;\, F_t^{-1} \text{ is non-decreasing}\}$$

where F_t^{-1} is defined by the formula (2). It is a closed interval. If one of its bound t_0 is finite, then μ_{t_0} has a point of infinite density with respect to the Lebesgue measure.

Proof. Any non-decreasing left continuous function is the inverse distribution function of some probability measure. If such a function is obtained by an affine combination of probabilities belonging to $W^2(\mathbb{R})$, then its probability measure belongs to $W^2(\mathbb{R})$ too.

Moreover, an affine combination of two left continuous function is left continuous, so that

$$\iota(\mu_0, \mu_1) = \{t \in \mathbb{R}\,;\, F_t^{-1} \text{ is non-decreasing}\}.$$

The fact that $\iota(\mu_0, \mu_1)$ is closed follows from the stability of non-decreasingness under pointwise convergence.

If the minimal slope

$$\inf \left\{ \frac{F_t^{-1}(m) - F_t^{-1}(m')}{m - m'} \,;\, 0 < m < m' < 1 \right\}$$

is positive for some t, then it stays positive in a neighborhood of t. Thus, a finite bound of $I(\mu_0, \mu_1)$ must have zero minimal slope, which corresponds to a point of infinite density. □

A geodesic is said to be complete if it is defined for all times. We also consider geodesic rays, defined on an interval $[0, T]$ or $[0, T[$ where T can be $+\infty$ (in which case we say that the ray is complete), and geodesic segments, defined on a closed interval.

It is easy to deduce a number of consequences from Lemma 3.1.

Lemma 3.2 — *In $W^2(\mathbb{R})$:*

1. *any geodesic ray issued from a Dirac mass can be extend to a complete ray,*
2. *no geodesic ray issued from a Dirac mass can be extended for negative times, except if all of its points are Dirac masses,*
3. *up to normalizing the speed, the only complete geodesics are those obtained by translating a point of $W^2(\mathbb{R})$:*

$$\mu_t(A) = \mu_0(A - t),$$

Proof. The inverse distribution function of a Dirac mass δ_x is the constant function F_0^{-1} with value x. Since it slopes

$$\frac{F_0^{-1}(m) - F_0^{-1}(m')}{m - m'} \qquad 0 < m < m' < 1$$

are all zero, for all positive times t the functions F_t^{-1} defined by formula (2) for any non-decreasing F_1^{-1} are non-decreasing. However, for $t < 0$ the F_t^{-1} are not non-decreasing if F_1^{-1} is not constant, we thus get (1) and (2).

Consider a point μ_0 of $W^2(\mathbb{R})$ defined by an inverse distribution function F_0^{-1}, and consider a complete geodesic (μ_t) issued from μ_0. Let F_t^{-1} be the inverse distribution function of μ_t. Then, since μ_t is defined for all times $t > 0$, the slopes of F_1^{-1} must be greater than those of F_0^{-1}:

$$F_0^{-1}(m) - F_0^{-1}(m') \leq F_1^{-1}(m) - F_1^{-1}(m') \quad \forall m < m'$$

otherwise, when t increases, some slope of F_t^{-1} will decrease linearly in t, thus vanishing in finite time.

But since μ_t is also defined for all $t < 0$, the slopes of F_1^{-1} must be lesser than those of F_0^{-1}. They are therefore equal, and the two inverse distribution function are equal up to an additive constant. The geodesic μ_t is the translation of μ_0 and we proved (3). □

3.2. **Convex hulls of totally atomic measures.** Define in $W^2(\mathbb{R})$ the following sets:

$$\begin{aligned}\Delta_1 &= \{\delta_x\,;\, x \in \mathbb{R}\}\\ \Delta_n &= \{\sum_{i=1}^{n} a_i\delta_{x_i}\,;\, x_i \in \mathbb{R}, \sum a_i = 1\}\\ \Delta'_{n+1} &= \Delta_{n+1} \setminus \Delta_n\end{aligned}$$

Then Δ_1 is simply the image of the natural embedding $\mathbb{R} \to W^2(\mathbb{R})$. In particular, it is a convex set. This is not the case of Δ_n is $n > 1$. In fact, we have the following.

Lemma 3.3 — *If $n > 1$, any point μ of Δ_{n+1} lies on a geodesic segment with endpoints in Δ_n. Moreover, the endpoints can be chosen with the same center of mass than that of μ.*

Proof. If $\mu \in \Delta_n$ the result is obvious. Assume $\mu = \sum a_i\delta_{x_i}$ is in Δ'_{n+1}. We can assume further that $x_1 < x_2 < \cdots < x_{n+1}$. Consider the measures

$$\begin{aligned}\mu_{-1} &= \sum_{i<n-1} a_i\delta_{x_i} + (a_{n-1} + a_n)\delta_{x_{n-1}} + a_{n+1}\delta_{x_{n+1}}\\ \mu_1 &= \sum_{i<n-1} a_i\delta_{x_i} + a_{n-1}\delta_{x_{n-1}} + (a_n + a_{n+1})\delta_{x_{n+1}}.\end{aligned}$$

Then μ lies on the geodesic segment from μ_{-1} to μ_1. To get a constant center of mass, one considers the geodesic

$$\mu_t = \sum_{i<n-1} a_i\delta_{x_i} + a_{n-1}\delta_{x_{n-1}+t} + a_n\delta_{x_n-\alpha t}$$

where $\alpha = \frac{a_{n-1}}{a_n}$. □

If X is a Polish geodesic space and C is a subset of X, one says that C is convex if every geodesic segment whose endpoints are in C lies entirely in C. The convex hull of a subset Y is the least convex set $C(Y)$ that contains Y. It is well defined since the intersection of two convex sets is a convex set.

Lemma 3.3 provides the following noteworthy fact that will prove useful latter on.

Lemma 3.4 — *The convex hull of Δ_n is dense in $W^2(\mathbb{R})$ if $n > 1$.*

Proof. Follows from Lemma 3.3 since the set of totally atomic measures $\bigcup_n \Delta_n$ is dense in $W^2(\mathbb{R})$. □

4. Curvature

More details on the (sectional) curvature of metric spaces are available for example in [3] or [5]. We shall consider the curvature of $W^2(\mathbb{R})$, in the sense of Alexandrov. Given any three points x, y, z in a geodesic metric space X, there is in $\mathbb{R}^2$ up to congruence a unique comparison triangle x', y', z', that is a triangle that satisfies $d(x, y) = d(x', y')$, $d(y, z) = d(y', z')$, and $d(z, x) = d(z', x')$.

One says that X has *non-positive curvature* (in the sense of Alexandrov), or is CAT(0), if for all x, y, z the distances between two points on sides of this triangle is lesser than or equal the distance of corresponding points in the comparison triangle.

Equivalently, X is CAT(0) if for any triangle x, y, z, any geodesic γ such that $\gamma(0) = x$ and $\gamma(1) = y$, and any $t \in [0, 1]$, the following inequality holds:

$$d^2(y, \gamma(t)) \leq (1-t)d^2(y, \gamma(0)) + td^2(y, \gamma(1)) - t(1-t)t\ell(\gamma)^2 \tag{3}$$

where $\ell(\gamma)$ denotes the length of γ, that is $d(x, z)$.

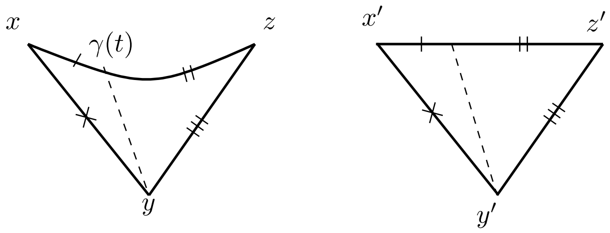

Figure 3. The CAT(0) inequality: the dashed segment is shorter in the triangle xyz than in the comparison triangle on the right.

One says that X has *vanishing curvature* (in the sense of Alexandrov) if equality holds for all x, y, z, γ, t:

$$d^2(y, \gamma(t)) = (1-t)d^2(y, \gamma(0)) + td^2(y, \gamma(1)) - t(1-t)t\ell(\gamma)^2 \tag{4}$$

This is equivalent to the condition that for any triangle x, y, z in X and any point $\gamma(t)$ on any geodesic segment between x and z, the distance between y and $\gamma(t)$ is *equal* to the corresponding distance in the comparison triangle.

Proposition 4.1 — *The space $W^2(\mathbb{R})$ has vanishing curvature in the sense of Alexandrov.*

Proof. It follows from the expression (1) of the distance in $W^2(\mathbb{R})$: if we denote by A, B, C the inverse distribution functions of the three considered points $x, y, z \in W^2(\mathbb{R})$, we get:

$$\begin{aligned} d^2(y,\gamma(t)) &= \int_0^1 (B-(1-t)A-tC)^2 \\ &= \int_0^1 \big[(1-t)^2(B-A)^2+t^2(B-C)^2 \\ &\quad +2t(1-t)(B-A)(B-C)\big] \end{aligned}$$

and using that $(1-t)^2 = (1-t) - t(1-t)$ and $t^2 = t - t(1-t)$,

$$\begin{aligned} d^2(y,\gamma(t)) &= (1-t)d^2(y,x)+td^2(y,z)-t(1-t)\int_0^1 \big[(B-A)^2 \\ &\quad +(B-C)^2-2(B-A)(B-C)\big] \\ &= (1-t)d^2(y,x)+td^2(y,z)-t(1-t)d^2(x,z). \end{aligned}$$

□

We shall use the vanishing curvature of $W^2(\mathbb{R})$ by means of the following result, where all subsets of X are assumed to be endowed with the induced metric (that need therefore not be inner).

PROPOSITION 4.2 — *Let X be a Polish uniquely geodesic space with vanishing curvature. If Y is a subset of X and $C(Y)$ is the convex hull of Y, then any isometry of Y can be extended into an isometry of $\overline{C(Y)}$.*

Proof. Let $\phi : Y \to Y$ the isometry to be extended. Let x, y be any points lying each on one geodesic segment $\gamma, \tau : [0,1] \to X$ whose endpoints are in Y. Consider the unique geodesics γ', τ' that satisfy $\gamma'(0) = \phi(\gamma(0))$, $\gamma'(1) = \phi(\gamma(1))$ $\tau'(0) = \phi(\tau(0))$, $\tau'(1) = \phi(\tau(1))$ and the points x', y' lying on them so that $d(x', \gamma'(0)) = d(x, \gamma(0))$, $d(x', \gamma'(1)) = d(x, \gamma(1))$, and the same for y'. This makes sense since, ϕ being an isometry on Y, γ' has the length of γ and τ' that of τ. We shall prove that $d(x', y') = d(x, y)$.

The vanishing of curvature implies that $d(x', \tau'(0)) = d(x, \tau(0))$: the triangles $\gamma(0), \gamma(1), \tau(0)$ and $\gamma'(0), \gamma'(1), \tau'(0)$ have the same comparison triangle. Similarly $d(x', \tau'(1)) = d(x, \tau(1))$. Now $x, \tau(0), \tau(1)$ and $x', \tau'(0), \tau'(1)$ have the same comparison triangle, and the vanishing curvature assumption implies $d(x', y') = d(x, y)$.

In particular, if $x = y$ then $x' = y'$. We can thus extend ϕ to the union of geodesic segments whose endpoints are in Y by mapping any such x to the corresponding x'. This is well-defined, and an isometry.

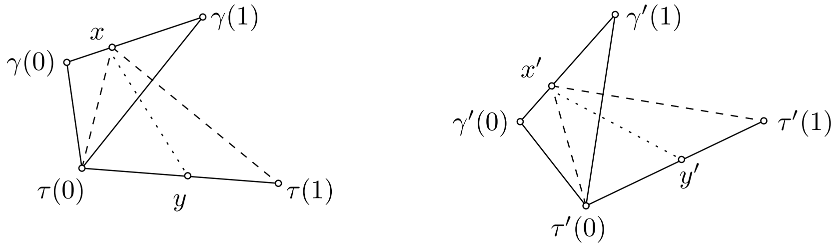

FIGURE 4. All triangles being flat, the distance is the same between x' and y' and between x and y.

Repeating this operation we can extend ϕ into an isometry of $C(Y)$. But X being complete, the continuous extension of ϕ to $\overline{C(Y)}$ is well-defined and an isometry. □

Note that the same result holds with the same proof when the curvature is constant but non-zero.

5. ISOMETRIES

5.1. **Existence and unicity of the non-trivial isometric flow.** In this section, we prove Theorem 1.1.

Let us start with the following consequence of Lemma 3.2.

LEMMA 5.1 — *An isometry of $W^2(\mathbb{R})$ must globally preserve the sets Δ_1 and Δ_2.*

Proof. We shall exhibit some geometric properties that characterize the points of Δ_1 and Δ_2 and must be preserved by isometries.

First, according to Lemma 3.2, the points $\mu \in \Delta_1$ are the only ones to satisfy : every geodesic ray starting at μ is complete. Since an isometry must map a geodesic (ray, segment) to another, this property is preserved by isometries of $W^2(\mathbb{R})$.

Second, let us prove that the point $\mu \in \Delta_2$ are the only ones that satisfy: any geodesic μ_t such that $\mu = \mu_0$, that can be extended to a maximal interval $[-T, +\infty)$ with $-\infty < T < 0$, has its endpoint μ_T in Δ_1.

This property is obviously satisfied by points of Δ_1. It is also satisfied by every points of Δ_2'. Indeed, write $\mu = a\delta_x + b\delta_y$ where $x < y$. Then if μ_1 does not write $\mu_1 = a\delta_{x_1} + b\delta_{y_1}$ with $x_1 < y_1$, either μ_t is not defined for $t > 1$ or it is not defined for $t < 0$. If μ_1 does write $\mu_1 = a\delta_{x_1} + b\delta_{y_1}$, then either $|y_1 - x_1| = |y - x|$ and μ_t is defined for all t, or $|y_1 - x_1| < |y - x|$ and μ_t is only defined until a finite positive

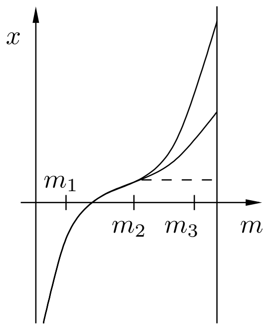

FIGURE 5. The geodesic between these inverse distributions is defined for negative times, more precisely until it reaches the dashed line.

time, or $|y_1 - x_1| > |y - x|$ and μ_t is defined from a finite negative time T where $\mu_T \in \Delta_1$.

Now if $\mu \notin \Delta_2$, its inverse distribution function F^{-1} takes three different values at some points $m_1 < m_2 < m_3$. Consider the geodesic between μ and the measure μ' whose inverse distribution function F'^{-1} coincide with F^{-1} on $[m_1, m_2]$ but is defined by

$$F'^{-1}(m) - F^{-1}(m_2) = 2(F^{-1}(m) - F^{-1}(m_2))$$

on $[m_2, 1)$. Then this geodesic is defined for all positive times, but stops at some nonpositive time T. Since F^{-1} takes different values at m_2 and m_3, one can extend the geodesic for small negative times and $T < 0$. But the inverse distribution function of the endpoint μ_T must take the same values than that of μ in m_1 and m_2, thus $\mu_T \notin \Delta_1$. $\square$

Up to composing with an isometry of $\mathbb{R}$, one is reduced to consider the isometries of $W^2(\mathbb{R})$ that fix each Dirac mass.

Any point $\mu \in \Delta_2'$ writes under the form

$$\mu = \mu(x, \sigma, e) = \frac{1}{1+e^2}\delta_{x-\sigma e} + \frac{e^2}{1+e^2}\delta_{x+\sigma/e}$$

where x is its center of mass, σ is the distance between μ and its center of mass, and $e > 0$.

LEMMA 5.2 — *An isometry of $W^2(\mathbb{R})$ that fixes each point of Δ_1 must restrict to Δ_2' to a map of the form:*

$$\Phi^t = \mu(x, \sigma, e) \mapsto \mu(x, \sigma, te)$$

for some $t \in (0, +\infty)$. Any such map is an isometry of Δ_2.

Proof. Let Φ be an isometry of $W^2(\mathbb{R})$ that fixes each point of Δ_1.

A computation gives the following expression for the distance between two measures in Δ_2':

$$d^2(\mu(x,\sigma,e),\mu(y,\rho,f)) = (x-y)^2+\sigma^2+\rho^2-2\sigma\rho\frac{f}{e}$$

whenever $e \geq f$.

Since Φ is an isometry, it preserves the center of mass and variance. The preceding expression shows that it must preserve the ratio e/f for any two measures $\mu(x,\sigma,e),\mu(y,\rho,f)$, and that this condition is sufficient to make Φ an isometry of Δ_2. □

Now Theorem 1.1 follows from Lemma 5.2, Proposition 4.2 and Lemma 3.4: any map $\mu(x,\sigma,e)\mapsto\mu(x,\sigma,te)$ with $t>0$ extends into an isometry of $\overline{C(\Delta_2)} = W^2(\mathbb{R})$. Any isometry of $W^2(\mathbb{R})$ can be assumed, by composition with an element of $\mathrm{Isom}(\mathbb{R})$, to fix Δ_1 pointwise and then, by composition with an element of this isometric flow, to fix Δ_2 pointwise. Then it must be the identity on the convex hull of Δ_2, thus on $W^2(\mathbb{R})$.

5.2. **Behaviour of the non-trivial isometric flow.** The definition of Φ^t is constructive, but not very explicit outside Δ_2. On Δ_2, the flow tends to put most of the mass on the right of the center of mass, very close to it, and send a smaller and smaller bit of mass far away on the left.

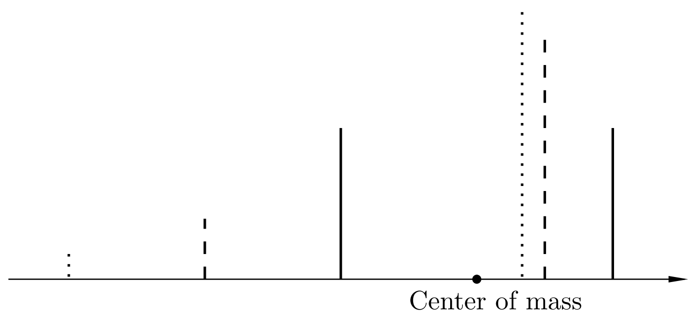

FIGURE 6. Image of a point of Δ_2 by Φ^2 (dashed) and Φ^3 (dotted).

The flow Φ^t preserves Δ_3 as its elements are the only ones to lie on a geodesic segment having both endpoints in Δ_2. Similarly, elements of Δ_n are the only ones to lie on a geodesic segment having an endpoint in Δ_2 and another in Δ_{n-1}, therefore Φ^t preserves Δ_n for all n.

Computations enable one to find formulas for Φ^t on Δ_n, but the expressions one gets are not so nice. For example, if $\mu = \frac{1}{3}\delta_{x_1}+\frac{1}{3}\delta_{x_2}+$

$\frac{1}{3}\delta_{x_3}$ where $x_1 \leq x_2 \leq x_3$, then

$$\begin{aligned}\Phi^t(\mu) &= \frac{1}{1+2t^2}\,\delta_{x_1+\frac{1}{3}(1-t)(x_3-x_1)+\frac{1}{3}(1-t)(x_2-x_1)} \\ &\quad+\frac{\frac{3}{2}t^2}{\left(1+\frac{1}{2}t^2\right)(1+2t^2)}\,\delta_{x_1+\frac{1}{3}(1-t)(x_3-x_1)+\frac{1}{3}(1+t^{-1}+t)(x_2-x_1)} \\ &\quad+\frac{\frac{1}{2}t^2}{1+\frac{1}{2}t^2}\,\delta_{x_1+\frac{1}{3}(1+2t^{-1})(x_3-x_1)+\frac{1}{3}(1-t^{-1})(x_2-x_1)}\end{aligned}$$

In order to get some intuition about Φ^t, let us prove the following.

PROPOSITION 5.3 — *Let μ be any point of $W^2(\mathbb{R})$ and x be its center of mass. If t goes to 0 or $+\infty$, then $\Phi^t(\mu)$ converges weakly to δ_x.*

Proof. We shall only consider the case when $t \to +\infty$ since the other one is symmetric. Let us start with a lemma.

LEMMA 5.4 — *If γ^t and ν^t are in $W^2(\mathbb{R})$ and both converge weakly to δ_x when t goes to $+\infty$, and if μ^t is in the geodesic segment between γ^t and ν^t for all t, then μ^t converges weakly to δ_x when t goes to $+\infty$.*

Proof. It is a direct consequence of the form of geodesics: if γ^t and ν^t both charge an interval $[x-\eta, x+\eta]$ with a mass at least $1-\varepsilon$, then μ^t must charge this interval with a mass at least $1-2\varepsilon$. □

Now we are able to prove the proposition on larger and larger subsets of $W^2(\mathbb{R})$. First, it is obvious on Δ_2. If it holds on Δ_n, the preceding Lemma together with Lemma 3.3 implies that it holds on Δ_{n+1}. To prove it on the whole of $W^2(\mathbb{R})$, the density of the subset of $\bigcup_n \Delta_n$ consisting of measures having center of mass x, and a diagonal process are sufficient. □

6. RANKS

One usually defines the *rank* of a metric space X as the supremum of the set of positive integers k such that there is an isometric, totally geodesic embedding of $\mathbb{R}^k$ into X. We say that an embedding is *totally geodesic* if any geodesic of the domain is mapped on a geodesic of the range.

6.1. Ranks of $W^2(\mathbb{R})$.

As a consequence of Lemma 3.2, we get the following result announced in the introduction.

THEOREM 6.1 — *The space $W^2(\mathbb{R})$ has rank 1.*

Proof. The image of an isometric, totally geodesic embedding of $\mathbb{R}^2$ would contain two different, crossing, complete geodesics, which is prevented by Lemma 3.2. $\square$

However, one can define less restrictive notions of rank as follows.

Definition 6.2 — *Let X be a Polish space.*

The semi-global rank *of X is defined as the supremum of the set of positive integers k such that for all $r \in \mathbb{R}^+$, there is an isometric, totally geodesic embedding of the ball of radius r of $\mathbb{R}^k$ into X.*

The loose rank *of X is defined as the supremum of the set of positive integers k such that there is a quasi-isometric embedding of $\mathbb{Z}^k$ into X.*

The notion of loose rank is relevant in a large class of metric spaces, including discrete spaces (the Gordian space [4], or the Cayley graph of a finitely presented group for example). We chose not to call it "coarse rank" due to the previous use of this term by Kapovich, Kleiner and Leeb.

The semi-global rank is motivated by the following simple result.

Proposition 6.3 — *A geodesic space X that has semi-global rank at least 2 is not δ-hyperbolic.*

Proof. Since X contains euclidean disks of arbitrary radius, it also contains euclidean equilateral triangles of arbitrary diameter. In such a triangle, the maximal distance between a point of an edge and the other edges is proportional to the diameter, thus is unbounded in X. $\square$

Consider the subset $\mathbb{R}^k_\leq = \{(x_1,\dots,x_k)\,;\,x_1 \leq x_2 \leq, \dots \leq x_k\}$ of $\mathbb{R}^k$. It is a closed, convex cone.

Lemma 6.4 — *The map*

$$\begin{array}{rcl} \mathbb{R}^k_\leq & \to & W^2(\mathbb{R}) \\ (x_1,\dots,x_k) & \mapsto & \sum \frac{1}{k}\delta_{x_i} \end{array}$$

is an isometric, totally geodesic embedding.

Proof. Straightforward. $\square$

Corollary 6.5 — *The space $W^2(\mathbb{R})$ has infinite semi-global rank and infinite loose rank.*

Proof. Since $\mathbb{R}^k_\leq$ contains arbitrarily large balls, the preceding Lemma implies that $W^2(\mathbb{R})$ has infinite semi-global rank.

Moreover, since $\mathbb{R}^k_\leq$ is a convex cone of non-empty interior, it contains a circular cone. Such a circular cone is conjugate by a linear (and thus bi-Lipschitz) map to the cone

$$\mathscr{C} = \{x_1^2 = \sum_{i\geq 2} x_i^2\}.$$

Now the vertical projection from $\{x_1 = 0\}$ to $\mathscr{C}$ is bi-Lipschitz. There is therefore a bi-Lipschitz embedding of $\mathbb{R}^{k-1}$ in $W^2(\mathbb{R})$ and, a fortiori, a quasi-isometric embedding of $\mathbb{Z}^{k-1}$. Therefore $W^2(\mathbb{R})$ has infinite loose rank. □

6.2. **Ranks of other spaces.** The ranks of $W^2(\mathbb{R})$ have an influence on those of many spaces due to the following lemma.

Lemma 6.6 — *If X and Y are Polish geodesic spaces, any isometric and totally geodesic embedding $\phi : X \to Y$ induces an isometric and totally geodesic embedding $\phi_\sharp : W^2(X) \to W^2(Y) : \mu \mapsto \mu(\phi\cdot)$.*

Proof. Since ϕ is isometric, for any $\mu \in W^2(X)$, $\phi_\sharp\mu$ is in $W^2(Y)$. Moreover any optimal transport in X is mapped to an optimal transport in Y (note that a coupling between two measures with support in $\phi(X)$ must have its support contained in $\phi(X) \times \phi(X)$). Since ϕ is totally geodesic, and due to the way geodesics in $W^2(X)$ are obtained from geodesics and an optimal transport, $\phi_\sharp$ is a totally geodesic embedding. □

Corollary 6.7 — *If X is a Polish geodesic space that contains a complete geodesic, then $W^2(X)$ has infinite semi-global rank and infinite loose rank. As a consequence, $W^2(X)$ is not δ-hyperbolic.*

Proof. Follows from the preceding Lemma, Corollary 6.5 and Proposition 6.3. □

One could hope that in Hadamard spaces, the projection to the center of mass

$$P : W^2(X) \to X$$

could give a higher bound on the rank of $W^2(X)$ by means of that of X. However, P need not map a geodesic on a geodesic. For example, if one consider on the real hyperbolic plane $\mathbb{R}\mathrm{H}^2$ the measures $\mu_t = 1/2\delta_p + 1/2\delta_{\gamma(t)}$ where p is a fixed point and $\gamma(t)$ is a geodesic, then μ_t is a geodesic of $W^2(\mathbb{R}\mathrm{H}^2)$ that is mapped to a curve with the same endpoints than γ, that can thus not be a geodesic.

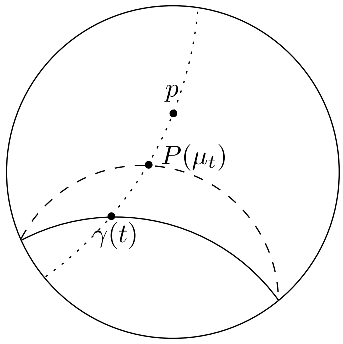

FIGURE 7. The projection P maps a geodesic of $W^2(\mathbb{R}\mathrm{H}^2)$ to a non-geodesic curve (dashed) in $\mathbb{R}\mathrm{H}^2$.

7. Open problems

Theorem 1.1 might not hold in $\mathbb{R}^n$, since the optimal transportation plan is far more intricate in higher dimension (see [1, 2, 7]). However in any Hadamard space X, isometries of $W^2(X)$ must preserve the set of Dirac masses (the proof is the same than in $\mathbb{R}$), and this fact could help get a grip on the problem.

QUESTION 1 — *Given a Hadamard space X (for example $\mathbb{R}^n$), does there exist isometries of $W^2(X)$ that are not induced by an isometry of X ?*

This question can of course be asked for any Polish space, but this might be more difficult. Even the following seems not obvious.

QUESTION 2 — *Does there exist a Polish space X whose Wasserstein space $W^2(X)$ possess an isometry that does not preserve the set of Dirac masses ?*

Last, when X is Hadamard, one could hope to use the projection P to link the rank of $W^2(X)$ to the loose rank of X.

QUESTION 3 — *If X is a Hadamard space, is the loose rank of X an upper bound for the rank of $W^2(X)$?*

References

[1] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(19):805–808, 1987.

[2] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[3] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[4] Jean-Marc Gambaudo and Étienne Ghys. Braids and signatures. *Bull. Soc. Math. France*, 133(4):541–579, 2005.

[5] Jürgen Jost. *Nonpositive curvature: geometric and analytic aspects*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1997.

[6] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. To appear in *Ann. Math.*.

[7] L. Rüschendorf and S. T. Rachev. A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.*, 32(1):48–54, 1990.

[8] Karl-Theodor Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.

[9] Karl-Theodor Sturm. On the geometry of metric measure spaces. II. *Acta Math.*, 196(1):133–177, 2006.

[10] Cédric Villani. *Optimal transport, old and new*. Springer.