



A geometric study of Wasserstein spaces: Euclidean spaces

Benoit Kloeckner

► To cite this version:

Benoit Kloeckner. A geometric study of Wasserstein spaces: Euclidean spaces. *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze*, 2010, IX (2), pp.297-323. [⟨10.2422/2036-2145.2010.2.03⟩](#). [⟨hal-00275067v3⟩](#)

HAL Id: hal-00275067

<https://hal.science/hal-00275067v3>

Submitted on 15 Apr 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A GEOMETRIC STUDY OF WASSERSTEIN SPACES: EUCLIDEAN SPACES

by

Benoît Kloeckner

Abstract. — In this article we consider Wasserstein spaces (with quadratic transportation cost) as intrinsic metric spaces. We are interested in usual geometric properties: curvature, rank and isometry group, mostly in the case of Euclidean spaces. Our most striking result is that the Wasserstein space of the line admits “exotic” isometries, which do not preserve the shape of measures.

Contents

1. Introduction	1
2. The Wasserstein space	4
3. Geodesics	8
4. Curvature	11
5. Isometries: the case of the line	14
6. Isometries: the higher-dimensional case	18
7. Ranks	22
8. Open problems	24
References	25

1. Introduction

The concept of optimal transportation recently raised a growing interest in links with the geometry of metric spaces. In particular the L^2 Wasserstein space $\mathcal{W}_2(X)$ have been used by Von Renesse and Sturm [15], Sturm [17] and Lott and Villani [12] to define certain curvature conditions on a metric space X . Many useful properties are inherited from X by $\mathcal{W}_2(X)$ (separability, completeness, geodesicness, some non-negative curvature conditions) while some other are not, like local compacity.

2000 Mathematics Subject Classification. — 54E70, 28A33.

Key words and phrases. — Wasserstein distance, optimal transportation, isometries, rank.

In this paper, we study the geometry of Wasserstein spaces as intrinsic spaces. We are interested, for example, in the isometry group of $\mathcal{W}_2(X)$, in its curvature and in its rank (the greatest possible dimension of a Euclidean space that embeds in it). In the case of the Wasserstein space of a Riemannian manifold, itself seen as an infinite-dimensional Riemannian manifold, the Riemannian connection and curvature have been computed by Lott [13]. See also [18] where Takatsu studies the subspace of Gaussian measures in $\mathcal{W}_2(\mathbb{R}^n)$, and [1] where Ambrosio and Gigli are interested in the second order analysis on $\mathcal{W}_2(\mathbb{R}^n)$, in particular its parallel transport.

The Wasserstein space $\mathcal{W}_2(X)$ contains a copy of X , the image of the isometric embedding

$$\begin{aligned} E : X &\rightarrow \mathcal{W}_2(X) \\ x &\mapsto \delta_x \end{aligned}$$

where δ_x is the Dirac mass at x . Moreover, given an isometry φ of X one defines an isometry $\varphi_\#$ of $\mathcal{W}_2(X)$ by $\varphi_\#(\mu)(A) = \mu(\varphi^{-1}(A))$. We thus get an embedding

$$\# : \text{Isom } X \rightarrow \text{Isom } \mathcal{W}_2(X)$$

These two elementary facts connect the geometry of $\mathcal{W}_2(X)$ to that of X .

One could expect that $\#$ is onto, *i.e.* that all isometries of $\mathcal{W}_2(X)$ are induced by those of X itself. Elements of $\#(\text{Isom } X)$ are called *trivial* isometries. Let us introduce a weaker property: a self-map Φ of $\mathcal{W}_2(X)$ is said to *preserve shapes* if for all $\mu \in \mathcal{W}_2(X)$, there is an isometry φ of X (that depends upon μ) such that $\Phi(\mu) = \varphi_\#(\mu)$. An isometry that does not preserve shapes is said to be *exotic*.

Our main result is the surprising fact that $\mathcal{W}_2(\mathbb{R})$ admits exotic isometries. More precisely we prove the following.

Theorem 1.1. — *The isometry group of $\mathcal{W}_2(\mathbb{R})$ is a semidirect product*

$$(1) \quad \text{Isom } \mathbb{R} \ltimes \text{Isom } \mathbb{R}$$

Both factors decompose: $\text{Isom } \mathbb{R} = \mathbb{Z}/2\mathbb{Z} \ltimes \mathbb{R}$ and the action defining the semidirect product (1) is simply given by the usual action of the left $\mathbb{Z}/2\mathbb{Z}$ factor on the right \mathbb{R} factor, that is $(\varepsilon, v) \cdot (\eta, t) = (\eta, \varepsilon t)$ where $\mathbb{Z}/2\mathbb{Z}$ is identified with $\{\pm 1\}$.

In (1), the left factor is the image of $\#$ and the right factor consist in all isometries that fix pointwise the set of Dirac masses. In the decomposition of the latter, the $\mathbb{Z}/2\mathbb{Z}$ factor is generated by a non-trivial involution that preserves shapes, while the \mathbb{R} factor is a flow of exotic isometries.

The main tool we use is the explicit description of the geodesic between two points μ_0, μ_1 of $\mathcal{W}_2(\mathbb{R})$ 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 $\mathcal{W}_2(\mathbb{R})$ 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 $\mathcal{W}_2(X)$. We also use the fact that $\mathcal{W}_2(\mathbb{R})$ has vanishing curvature in the sense of Alexandrov.

Let us describe roughly the non-trivial isometries that fix pointwise the set of Dirac masses. On the one hand, the non-trivial isometry generating the $\mathbb{Z}/2\mathbb{Z}$ factor

is defined as follows: a measure μ is mapped to its symmetric with respect to its center of mass. On the other hand, the 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 (but of course not in $\mathcal{W}_2(\mathbb{R})$) to δ_x (where x is the center of mass of μ), see Proposition 5.4.

The case of the line seems very special. For example, $\mathcal{W}_2(\mathbb{R}^n)$ admits non-trivial isometries but all of them preserve shapes.

Theorem 1.2. — *If $n \geq 2$, the isometry group of $\mathcal{W}_2(\mathbb{R}^n)$ is a semidirect product*

$$(2) \quad \text{Isom}(\mathbb{R}^n) \ltimes \text{O}(n)$$

where the action of an element $\psi \in \text{Isom}(\mathbb{R}^n)$ on $\text{O}(n)$ is the conjugacy by its linear part $\vec{\psi}$.

The left factor is the image of $\#$ and each element in the right factor fixes all Dirac masses and preserves shapes.

The proof relies on Theorem 1.1, some elementary properties of L^2 optimal transportation in \mathbb{R}^n and Radon's Theorem [14].

We see that the quotient $\text{Isom} \mathcal{W}_2(\mathbb{R}^n) / \text{Isom} \mathbb{R}^n$ is compact if and only if $n > 1$. The higher-dimensional Euclidean spaces are more rigid than the line for this problem, and we expect most of the other metric spaces to be even more rigid in the sense that $\#$ is onto.

Another consequence of the study of complete geodesics concerns the rank of $\mathcal{W}_2(\mathbb{R}^n)$.

Theorem 1.3. — *There is no isometric embedding of \mathbb{R}^{n+1} into $W_2(\mathbb{R}^n)$.*

It is simple to prove that despite Theorem 1.3, large pieces of \mathbb{R}^n can be embedded into $\mathcal{W}_2(\mathbb{R})$, which has consequently infinite weak rank in a sense to be precised. As a consequence, we get for example:

Proposition 1.4. — *If X is any Polish geodesic metric space that contains a complete geodesic, then $\mathcal{W}_2(X)$ 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. An explicit example is computed in [2] (Example 7.3.3); more generally, if X contains a rhombus (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 $\mathcal{W}_2(X)$ is not uniquely geodesic, and in particular not CAT(0), even if X itself is strongly negatively curved.

Organization of the paper. — Sections 2 to 4 collect some properties needed in the sequel. Theorem 1.1 is proved in Section 5, Theorem 1.2 in Section 6. Section 7 is devoted to the ranks of $\mathcal{W}_2(\mathbb{R})$ and $\mathcal{W}_2(\mathbb{R}^n)$, and we end in Section 8 with some open questions.

Acknowledgements. — I wish to thank 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. I am also indebted to Yann Ollivier for advises and pointing out some inaccuracies and mistakes in preliminary versions of this paper.

2. The Wasserstein space

In this preliminary section we recall well-known general facts on $\mathcal{W}_2(X)$. One can refer to [19, 20] for further details and much more. Note that the denomination “Wasserstein space” is debated and historically inaccurate. However, it is now the most common denomination and thus an occurrence of the self-applying theorem of Arnol’d according to which a mathematical result or object is usually attributed to someone that had little to do with it.

2.1. Geodesic spaces. — Let X be a Polish (*i.e.* complete and separable metric) 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 $\mathcal{W}_2(X)$ 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 $d_{\mathcal{W}}$ defined by:

$$d_{\mathcal{W}}^2(\mu_0, \mu_1) = \inf \int_{X \times X} d^2(x, y) \Pi(dxdy)$$

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 $d_{\mathcal{W}}^2(\mu_0, \mu_1)$ and a transportation plan achieving this minimum is an optimal coupling.

An optimal coupling is said to be *deterministic* if it can be written under the form $\Pi(dxdy) = \mu(dx) \mathbb{1}[y = Tx]$ where $T : X \rightarrow X$ is a measurable map and $\mathbb{1}[A]$ is 1 if A is satisfied and 0 otherwise. This means that the coupling does not split mass: all the mass at point x is moved to the point Tx . One usually write $\Pi = (\text{Id} \times T)_{\#} \mu$. Of course, for Π to be a coupling between μ and ν , the relation $\nu = T_{\#} \mu$ must hold.

Under the assumptions we put on X , the metric space $\mathcal{W}_2(X)$ 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 $\mathcal{W}_2(X)$ in the following way. Let $C([0, 1], X)$ be the set of continuous curves $[0, 1] \rightarrow X$, let $g : X \times X \rightarrow C([0, 1], X)$ be the application that maps (x, y) to the constant speed geodesic between these

points, and for each $t \in [0, 1]$ let $e(t) : C([0, 1], X) \rightarrow X$ be the map $\gamma \mapsto \gamma(t)$. Then $t \mapsto e(t)_{\#} g_{\#} \Pi$ 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 $\mathcal{W}_2(X)$ associated to the optimal coupling Π . Moreover, all geodesics are obtained that way.

Note that for most spaces X , the optimal coupling is not unique for all pairs of probability measures, and $\mathcal{W}_2(X)$ is therefore not uniquely geodesic even if X is.

One of our goal is to determine whether the Dirac measures can be detected inside $\mathcal{W}_2(X)$ by purely geometric properties, so that we can link the isometries of $\mathcal{W}_2(X)$ 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:

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

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 .

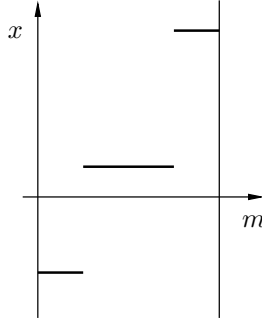


FIGURE 1. Inverse distribution function of a combination of three Dirac masses

Let μ_0 and μ_1 be two points of $\mathcal{W}_2(\mathbb{R})$, and let F_0, F_1 be their distribution functions. Then the distance between μ_0 and μ_1 is given by

$$(3) \quad d^2(\mu_0, \mu_1) = \int_0^1 (F_0^{-1}(m) - F_1^{-1}(m))^2 dm$$

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

$$(4) \quad F_t^{-1} = (1 - t)F_0^{-1} + tF_1^{-1}$$

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 are 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 were linear, it would be equivalent to leave the mass on $[\varepsilon, 1]$ where it is and move the remainder from $[0, \varepsilon]$ to $[1, 1 + \varepsilon]$. If the cost function were concave, then the latter solution would be better than the former.

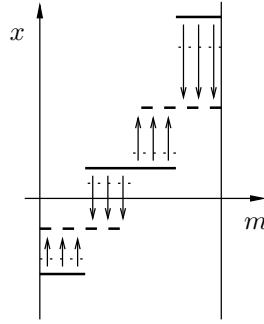


FIGURE 2. A geodesic between two atomic measures: the mass moves with speed proportional to the length of the arrows.

2.3. Higher dimensional Euclidean spaces. — The Monge-Kantorovich problem is far more intricate in \mathbb{R}^n ($n \geq 2$) than in \mathbb{R} . The major contributions of Knott and Smith [11, 16] and Brenier [3, 4] give a quite satisfactory characterization of optimal couplings and their unicity when the two considered measures μ and ν are absolutely continuous (with respect to the Lebesgue measure). We shall not give details of these works, for which we refer to [19, 20] again. Let us however consider some toy cases, which will prove useful later on. Missing proofs can be found in [10], Section 2.1.2.

We consider \mathbb{R}^n endowed with its canonical inner product and norm, denoted by $|\cdot|$.

Translations. — Let T_v be the translation of vector v and assume that $\nu = (T_v)_\# \mu$. Then the unique optimal coupling between μ and ν is deterministic, equal to $(\text{Id} \times T_v)_\# \mu$, and therefore $d_W(\mu, \nu) = |v|$. This means that the only most economic way to move the mass from μ to ν is to translate each bit of mass by the vector v . This is a quite intuitive consequence of the convexity of the cost. In particular, the geodesic between μ and ν can be extended for all times $t \in \mathbb{R}$. This happens only in this case as we shall see later on.

Dilations. — Let D_x^λ be the dilation of center x and ratio λ and assume that $\nu = (D_x^\lambda)_\# \mu$. Then the unique optimal coupling between μ and ν is deterministic, equal

to $(\text{Id} \times D_x^\lambda)_\# \mu$. In particular,

$$d_{\mathcal{W}}(\mu, \nu) = |1 - \lambda^2|^{\frac{1}{2}} d_{\mathcal{W}}(\mu, \delta_x)$$

As a consequence, the geodesic between μ and ν is unique and made of homothetic of μ , and can be extended only to a semi-infinite interval: it cannot be extended beyond δ_x (unless μ is a Dirac mass itself).

Orthogonal measures. — Assume that μ and ν are supported on orthogonal affine subspaces V and W of \mathbb{R}^n . Then if Π is any coupling, assuming $0 \in V \cap W$, we have

$$\begin{aligned} \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \Pi(dxdy) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} (|x|^2 + |y|^2) \Pi(dxdy) \\ &= \int_V |x|^2 \mu(dx) + \int_W |y|^2 \nu(dy) \end{aligned}$$

therefore the cost is the same whatever the coupling.

Balanced combinations of two Dirac masses. — Assume that $\mu = 1/2\delta_{x_0} + 1/2\delta_{y_0}$ and $\nu = 1/2\delta_{x_1} + 1/2\delta_{y_1}$. A coupling between μ and ν is entirely determined by the amount $m \in [0, 1/2]$ of mass sent from x_0 to x_1 . The cost of the coupling is

$$\frac{1}{2}|x_1 - y_0|^2 + \frac{1}{2}|x_0 - y_1|^2 - 2m(y_0 - x_0) \cdot (y_1 - x_1)$$

thus the optimal coupling is unique and deterministic if $(y_0 - x_0) \cdot (y_1 - x_1) \neq 0$, given by the map $(x_0, y_0) \mapsto (x_1, y_1)$ if $(y_0 - x_0) \cdot (y_1 - x_1) > 0$ and by the map $(x_0, y_0) \mapsto (y_1, x_1)$ if $(y_0 - x_0) \cdot (y_1 - x_1) < 0$ (figure 3). Of course if $(y_0 - x_0) \cdot (y_1 - x_1) = 0$, then all coupling have the same cost and are therefore optimal.

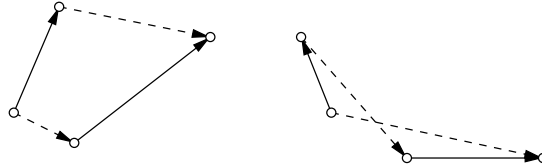


FIGURE 3. Optimal coupling between balanced combinations of two Dirac masses. Continuous arrows represent the vectors $y_0 - x_0$ and $y_1 - x_1$ while dashed arrows represent the optimal coupling.

If the combinations are not balanced (the mass is not equally split between the two point of the support), then the optimal coupling is easy to deduce from the preceding computation. For example if $(y_0 - x_0) \cdot (y_1 - x_1) > 0$ then as much mass as possible must be sent from x_0 to x_1 , and this determines the optimal coupling.

This example has a much more general impact than it might seem: it can be generalized to the following (very) special case of the *cyclical monotonicity* (see for example [20], Chapter 5) which will prove useful in the sequel.

Lemma 2.1. — *If Π is an optimal coupling between any two probability measures on \mathbb{R}^n , then*

$$(y_0 - x_0) \cdot (y_1 - x_1) \geq 0$$

holds whenever (x_0, x_1) and (y_0, y_1) are in the support of Π .

2.4. Spaces of nonpositive curvature. — In this paper we shall consider two curvature conditions. The first one is a negative curvature condition, the δ -hyperbolicity introduced by Gromov (see for example [5]). 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 sectional 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 locally CAT(0) spaces.

A geodesic CAT(0) Polish space X is also called a *Hadamard space*. 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 $\mu \in \mathcal{M}_2(X)$ 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 is \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 the use of such a formula.

We thus get a map $P : \mathcal{M}_2(X) \rightarrow X$ that maps any L^2 probability measure to its center of mass. Obviously, P is a left inverse to E and one can hope to use this map to link closer the geometry of $\mathcal{M}_2(X)$ to that of X . That's why our questions, unlike most of the classical ones in optimal transportation, might behave more nicely when the curvature is non-positive than when it is non-negative.

3. Geodesics

The content of this section, although difficult to locate in the bibliography, is part of the folklore and does not pretend to originality. We give proofs for the sake of completeness.

3.1. Case of the line. — We now consider the geodesics of $\mathcal{M}_2(\mathbb{R})$. Our first goal is to determine on which maximal interval they can be extended.

Maximal extension. — Let μ_0, μ_1 be two points of $\mathcal{M}_2(\mathbb{R})$ 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 $\mathcal{M}_2(\mathbb{R})$ is uniquely geodesic, there is a unique maximal interval on which γ can be extended into a geodesic, denoted by $I(\mu_0, \mu_1)$.

Lemma 3.1. — *One has*

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

where F_t^{-1} is defined by the formula (4). It is a closed interval. If one of its bound t_0 is finite, then μ_{t_0} does not have bounded 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 $\mathcal{W}_2(\mathbb{R})$, then its probability measure belongs to $\mathcal{W}_2(\mathbb{R})$ too.

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

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

The fact that $I(\mu_0, \mu_1)$ is closed follows from the stability of non-decreasing functions 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, and cannot have a bounded density. \square

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 $[0, +\infty[$ (in the latter 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.

Proposition 3.2. — *In $\mathcal{W}_2(\mathbb{R})$:*

- (1) *any geodesic ray issued from a Dirac mass can be extended 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 $\mathcal{W}_2(\mathbb{R})$:*

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

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

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

are all zero, for all positive times t the functions F_t^{-1} defined by formula (4) 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 $\mathcal{W}_2(\mathbb{R})$ 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 becoming negative 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). \square

Convex hulls of totally atomic measures. — Define in $\mathcal{W}_2(\mathbb{R})$ the following sets:

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

Recall that 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, and is equal to $\bigcup_{n \in \mathbb{N}} Y_n$ where $Y_0 = Y$ and Y_{n+1} is obtained by adding to Y_n all points lying on a geodesic whose endpoints are in Y_n .

Since Δ_1 is the image of the isometric embedding $E : \mathbb{R} \rightarrow \mathcal{W}_2(\mathbb{R})$, it is a convex set. This is not the case of Δ_n is $n > 1$. In fact, we have the following.

Proposition 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 $x_1 < x_2 < \dots < x_{n+1}$. 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

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

where $\alpha = \frac{a_n}{a_{n+1}}$. \square

In particular, we get the following noteworthy fact that will prove useful latter on.

Proposition 3.4. — *The convex hull of Δ_n is dense in $\mathcal{W}_2(\mathbb{R})$ if $n > 1$.*

Proof. — Follows from Proposition 3.3 since the set of totally atomic measures $\bigcup_n \Delta_n$ is dense in $\mathcal{W}_2(\mathbb{R})$. \square

3.2. Complete geodesics in higher dimension. — In \mathbb{R}^n , the optimal coupling and thus the geodesics are not as explicit as in the case of the line. It is however possible to determine which geodesic can be extended to all times in \mathbb{R} .

Lemma 3.5. — *Let $\mu = (\mu_t)_{t \in I}$ be a geodesic in $\mathcal{W}_2(\mathbb{R}^n)$ associated to an optimal coupling Π between μ_0 and μ_1 . Then for all times r and s in I and all pair of points $(x_0, x_1), (y_0, y_1)$ in the support of Π , the following hold:*

$$|u|^2 + (r+s)u \cdot v + rs|v|^2 \geq 0$$

where $u = y_0 - x_0$ and $v = y_1 - x_1 - (y_0 - x_0)$.

Proof. — Let us introduce the following notations: for all pair of points $a_0, a_1 \in \mathbb{R}^n$, $a_t = (1-t)a_0 + ta_1$ and $\Pi_{r,s}$ is the law of the random variable (X_r, X_s) where (X_0, X_1) is any random variable of law Π . As we already said, $\Pi_{r,s}$ is an optimal coupling of μ_r, μ_s whose corresponding geodesic is the restriction of (μ_t) to $[r, s]$.

Since $\Pi_{r,s}$ is optimal, according to the cyclical monotonicity (see Lemma 2.1) one has $(y_r - x_r) \cdot (y_s - x_s) \geq 0$.

But with the above notations, one has $y_r - x_r = u + rv$ and $y_s - x_s = u + sv$, and we get the desired inequality. \square

Let us show why this Lemma implies that the only complete geodesics are those obtained by translation. There are immediate consequences on the rank of $\mathcal{W}_2(\mathbb{R}^n)$, see Theorem 1.3 and Section 7.

Proposition 3.6. — *Let $\mu = (\mu_t)_{t \in \mathbb{R}}$ be a geodesic in $\mathcal{W}_2(\mathbb{R}^n)$ defined for all times. Then there is a vector u such that $\mu_t = (T_{tu})_{\#}\mu_0$.*

This result holds even if $n = 1$, as stated in Proposition 3.2.

Proof. — It is sufficient to find a u such that $\mu_1 = (T_u)_{\#}\mu_0$, since then there is only one geodesic from μ_0 to μ_1 .

Consider any pair of points $(x_0, x_1), (y_0, y_1)$ in the support of the coupling Π between μ_0 and μ_1 that defines the restriction of μ to $[0, 1]$. Define $u = y_0 - x_0$ and $v = y_1 - x_1 - (y_0 - x_0)$. If $v \neq 0$, then there are real numbers $r < s$ such that $|u|^2 + (r+s)u \cdot v + rs|v|^2 < 0$. Then the coupling $\Pi_{r,s}$ between μ_r and μ_s that defines the restriction of μ to $[r, s]$, defined as above, cannot be optimal. This is a contradiction with the assumption that μ is a geodesic.

Therefore, for all $(x_0, x_1), (y_0, y_1)$ in the support of Π one has $y_0 - x_0 = y_1 - x_1$. This amounts to say that Π is deterministic, given by a translation of vector $u = y_0 - x_0$. \square

4. Curvature

Once again, this section mainly collects some facts that are already well-known but shall be used on the sequel.

More details on the (sectional) curvature of metric spaces are available for example in [6] or [9]. We shall consider the curvature of $\mathcal{W}_2(\mathbb{R})$, in the sense of Alexandrov. Given any three points x, y, z in a geodesic metric space X , there is up to congruence

a unique comparison triangle x', y', z' in \mathbb{R}^2 , 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 to the distance between the corresponding points in the comparison triangle, see figure 4.

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

$$(5) \quad d^2(y, \gamma(t)) \leq (1-t)d^2(y, \gamma(0)) + td^2(y, \gamma(1)) - t(1-t)t\ell(\gamma)^2$$

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

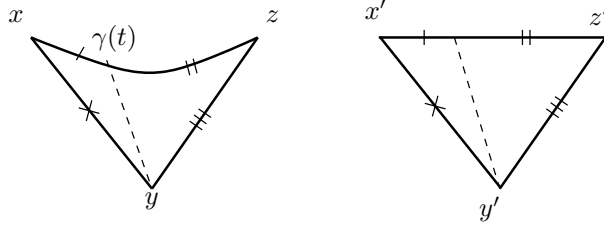


FIGURE 4. The CAT(0) inequality: the dashed segment is shorter in the triangle xyz than in the comparison triangle on the right.

One says that X has *vanishing curvature* if equality holds for all x, y, z, γ, t :

$$(6) \quad d^2(y, \gamma(t)) = (1-t)d^2(y, \gamma(0)) + td^2(y, \gamma(1)) - t(1-t)t\ell(\gamma)^2$$

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 $\mathcal{W}_2(\mathbb{R})$ has vanishing Alexandrov curvature.*

Proof. — It follows from the expression (3) of the distance in $\mathcal{W}_2(\mathbb{R})$: if we denote by A, B, C the inverse distribution functions of the three considered points $x, y, z \in \mathcal{W}_2(\mathbb{R})$, we get:

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

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

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

□

We shall use the vanishing curvature of $\mathcal{W}_2(\mathbb{R})$ 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 $C(Y)$.*

Proof. — Let $\varphi : Y \rightarrow Y$ the isometry to be extended. Let x, y be any points lying each on one geodesic segment $\gamma, \tau : [0, 1] \rightarrow X$ whose endpoints are in Y . Consider the unique geodesics γ', τ' that satisfy $\gamma'(0) = \varphi(\gamma(0))$, $\gamma'(1) = \varphi(\gamma(1))$, $\tau'(0) = \varphi(\tau(0))$, $\tau'(1) = \varphi(\tau(1))$ 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 $d(x', y') = d(x, y)$.

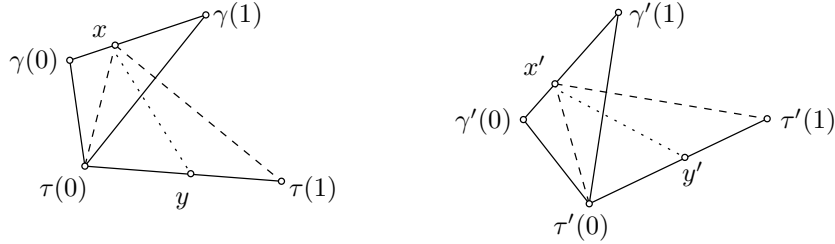


FIGURE 5. All triangles being flat, the distance is the same between x' and y' and between x and y .

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

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

Higher dimensional case. — Proposition 4.1 does not hold in $\mathcal{W}_2(\mathbb{R}^n)$. In effect, there are pairs of geodesics that meet at both endpoints (take measures whose support lie on orthogonal subspaces of \mathbb{R}^n). Taking a third point in one of the two geodesics, one gets a triangle in $\mathcal{W}_2(\mathbb{R}^n)$ whose comparison triangle has its three vertices on a line. This implies that $\mathcal{W}_2(\mathbb{R}^n)$ is not CAT(0). The situation is in fact worse than that: in any neighborhood U of any point of $\mathcal{W}_2(\mathbb{R}^n)$ one can find two different geodesics that meet at their endpoints. One can say that this space has positive sectional curvature at arbitrarily small scales.

5. Isometries: the case of the line

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 Proposition 3.2.

Lemma 5.1. — *An isometry of $\mathcal{W}_2(\mathbb{R})$ 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 Proposition 3.2, the points $\mu \in \Delta_1$ are the only ones to satisfy: every maximal 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 $\mathcal{W}_2(\mathbb{R})$.

Second, let us prove that the point $\mu \in \Delta_2$ are the only ones that satisfy: any geodesic μ_t such that $\mu = \mu_0$ and 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, denote by F_t the distribution function of μ_t and write $\mu = a\delta_x + (1-a)\delta_y$ where $x < y$. Then if μ_1 does not write $\mu_1 = a\delta_{x_1} + (1-a)\delta_{y_1}$ with $x_1 < y_1$, either

- there are two reals b, c such that $a < b < c < 1$ and $F_1^{-1}(b) < F_1^{-1}(c)$,
- there are two reals b, c such that $0 < b < c < a$ and $F_1^{-1}(b) < F_1^{-1}(c)$, or
- μ_1 is a Dirac mass.

In the first two cases, μ_t is not defined for $t < 0$, and in the third one, it is not defined for $t > 1$. If μ_1 does write $\mu_1 = a\delta_{x_1} + (1-a)\delta_{y_1}$, 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 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

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

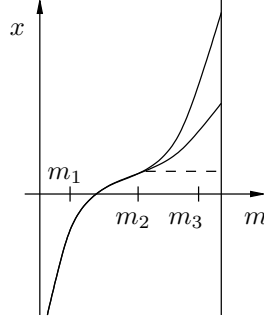


FIGURE 6. The geodesic between these inverse distributions is defined for negative times, more precisely until it reaches the dashed line.

on $[m_2, 1)$ (see figure 6). 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

Now we consider isometries of Δ_2 , to which all isometries of $\mathcal{W}_2(\mathbb{R})$ shall be reduced. Any point $\mu \in \Delta'_2$ writes under the form

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

where x is its center of mass, σ is the distance between μ and its center of mass, and p is any real number. In probabilistic terms, if μ is the law of a random variable then x is its expected value and σ^2 its variance.

Lemma 5.2. — *An isometry of $\mathcal{W}_2(\mathbb{R})$ that fixes each point of Δ_1 must restrict to Δ'_2 to a map of the form:*

$$\Phi(\varphi) = \mu(x, \sigma, p) \mapsto \mu(x, \sigma, \varphi(p))$$

for some $\varphi \in \text{Isom}(\mathbb{R})$. Any such map is an isometry of Δ_2 .

Proof. — Let Φ be an isometry of $\mathcal{W}_2(\mathbb{R})$ that fixes each point of Δ_1 .

A computation gives the following expression for the distance between two measures in Δ'_2 :

$$d_{\mathcal{W}}^2(\mu(x, \sigma, p), \mu(y, \rho, q)) = (x - y)^2 + \sigma^2 + \rho^2 - 2\sigma\rho e^{|p-q|}$$

Since Φ is an isometry, it preserves the center of mass and variance. The preceding expression shows that it must preserve the euclidean distance between p and q 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 . \square

Lemma 5.3. — *Let $\psi : x \rightarrow \varepsilon x + v$ and $\varphi : p \rightarrow \eta p + t$ be isometries of \mathbb{R} . Then*

$$\#(\psi)\Phi(\varphi)\#(\psi)^{-1}(\mu(x, \sigma, p)) = \mu(x, \sigma, \eta p + \varepsilon t)$$

Proof. — It follows from a direct computation:

$$\begin{aligned}
\#(\psi)\Phi(\varphi)\#(\psi)^{-1}(\mu(x, \sigma, p)) &= \#(\psi)\Phi(\varphi)(\mu(\varepsilon x - \varepsilon v, \sigma, \varepsilon p)) \\
&= \#(\psi)(\mu(\varepsilon x - \varepsilon v, \sigma, \eta \varepsilon p + t)) \\
&= \mu(x, \sigma, \eta p + \varepsilon t)
\end{aligned}$$

□

We are now able to prove our main result.

Proof of Theorem 1.1. — First Lemma 5.1 says that any isometry of $\mathcal{W}_2(\mathbb{R})$ acts on Δ_1 and Δ_2 .

Let \mathcal{L} be $\#(\text{Isom } \mathbb{R})$ and \mathcal{R} be the subset of $\text{Isom } \mathcal{W}_2(\mathbb{R})$ consisting of isometries that fix Δ_1 pointwise. Then \mathcal{R} is a normal subgroup of $\text{Isom } \mathcal{W}_2(\mathbb{R})$.

Let Ψ be an isometry of $\mathcal{W}_2(\mathbb{R})$. It acts isometrically on Δ_1 , thus there is an isometry ψ of \mathbb{R} such that $\#(\psi)\Psi \in \mathcal{R}$. In particular, $\text{Isom } \mathcal{W}_2(\mathbb{R}) = \mathcal{L}\mathcal{R}$. Since $\mathcal{L} \cap \mathcal{R}$ is reduced to the identity, we do have a semidirect product $\text{Isom } \mathcal{W}_2(\mathbb{R}) = \mathcal{L} \ltimes \mathcal{R}$.

According to Proposition 4.2, each map $\Phi(\varphi) : \mu(x, \sigma, p) \mapsto \mu(x, \sigma, \varphi(p))$ extends into an isometry of $\overline{C(\Delta_2)}$, which is $\mathcal{W}_2(\mathbb{R})$ by Proposition 3.4. We still denote by $\Phi(\varphi)$ this extension. Proposition 3.4 also shows that an isometry of $\mathcal{W}_2(\mathbb{R})$ is entirely determined by its action on Δ_2 . The description of \mathcal{R} now follows from Lemma 5.2.

If σ denotes the symmetry around $0 \in \mathbb{R}$, then $\Phi(\sigma)$ maps a measure $\mu \in \mathcal{W}_2(\mathbb{R})$ to its symmetric with respect to its center of mass, thus preserves shapes. Any other $\varphi \in \text{Isom}(\mathbb{R})$ is a translation or the composition of σ and a translation.

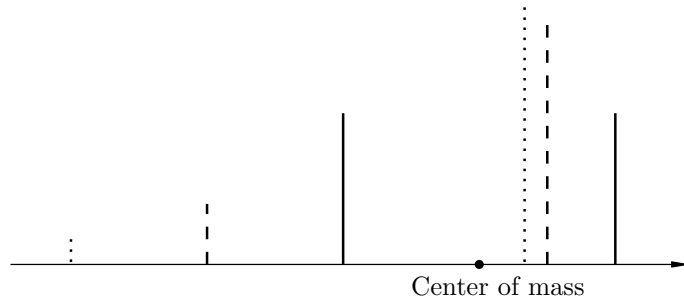
By the *exotic isometry flow* of $\text{Isom } \mathcal{W}_2(\mathbb{R})$ we mean the flow of isometries $\Phi^t = \Phi(\varphi^t)$ obtained when $\varphi^t : p \rightarrow p + t$ is a translation. This flow does not preserve shapes as is seen from its expression in Δ_2 .

At last, Lemma 5.3 gives the asserted description of the semidirect product. □

5.2. Behaviour of the exotic isometry 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.

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 .

Direct 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 $x_1 \leq x_2 \leq x_3$,

FIGURE 7. Image of a point of Δ_2 by Φ^2 (dashed) and Φ^3 (dotted).

then

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

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

Proposition 5.4. — *Let μ be any point of $\mathcal{W}_2(\mathbb{R})$ and x its center of mass. If t goes to $\pm\infty$, then $\Phi^t(\mu)$ converges weakly to δ_x .*

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

Lemma 5.5. — *If γ^t and ν^t are in $\mathcal{W}_2(\mathbb{R})$ 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 $1-2\varepsilon$. \square

Now we are able to prove the proposition on larger and larger subsets of $\mathcal{W}_2(\mathbb{R})$. First, it is obvious on Δ_2 . If it holds on Δ_n , the preceding Lemma together with Proposition 3.3 implies that it holds on Δ_{n+1} . To prove it on the whole of $\mathcal{W}_2(\mathbb{R})$, the density of the subset of $\bigcup_n \Delta_n$ consisting of measures having center of mass x , and a diagonal process are sufficient. \square

6. Isometries: the higher-dimensional case

To show that the exotic isometry flow of $\mathcal{W}_2(\mathbb{R})$ is exceptional, let us consider the higher-dimensional case : there are isometries of $\mathcal{W}_2(\mathbb{R}^n)$ that fix pointwise the set of Dirac masses, but there are not so many and they preserve shapes.

6.1. Existence of non-trivial isometries. — We start with the existence of non-trivial isometries, that however preserves shapes.

Proposition 6.1. — *If φ is a linear isometry of \mathbb{R}^n , then the map*

$$\Phi(\varphi) : \mu \mapsto \varphi_{\#}(\mu - g) + g$$

where g denotes both the center of mass of μ and the corresponding translation, is an isometry of $\mathcal{W}_2(\mathbb{R}^n)$ (see figure 8).

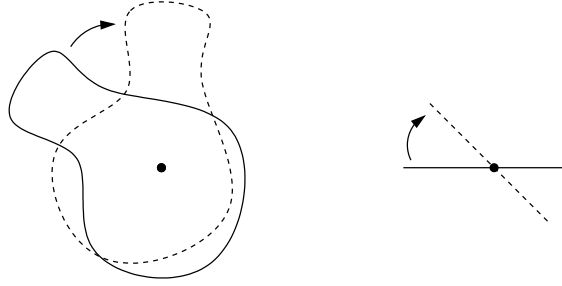


FIGURE 8. Example of a non-trivial isometry that preserves shapes.

Note that we need φ to be linear, thus we do not get as many non-trivial isometries as in $\mathcal{W}_2(\mathbb{R})$. Moreover, all isometries constructed this way preserve shapes.

Proof. — We only need to check the case of absolutely continuous measures μ, ν since they form a dense subset of $\mathcal{W}_2(\mathbb{R}^n)$. In that case, there is a unique optimal coupling that is deterministic, given by a map $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that $T_{\#}(\mu) = \nu$. Denote by g and h the centers of mass of μ and ν . Let us show that there is a good coupling between $\mu' = \Phi(\varphi)(\mu)$ and $\nu' = \Phi(\varphi)(\nu)$.

Let T' be the map defined by

$$T'(\varphi(x - g) + g) = \varphi(Tx - h) + h$$

By construction $T'_\#(\mu') = \nu'$. Moreover the cost of the coupling $(\text{Id} \times T')_\# \mu'$ is

$$\begin{aligned}
A &= \int_{\mathbb{R}^n} |\varphi(x - g) + g - \varphi(Tx - h) - h|^2 \mu(dx) \\
&= \int_{\mathbb{R}^n} \left(|Tx - x|^2 + 2|g - h|^2 + 2(x - Tx) \cdot (h - g) + 2\varphi(x - Tx) \cdot (g - h) \right. \\
&\quad \left. + 2\varphi(h - g) \cdot (g - h) \right) \mu(dx) \\
&= d_{\mathcal{W}}^2(\mu, \nu) + 2|g - h|^2 + 2(g - h) \cdot (h - g) + 2\varphi(g - h) \cdot (g - h) \\
&\quad + 2\varphi(h - g) \cdot (g - h) \\
&= d_{\mathcal{W}}^2(\mu, \nu)
\end{aligned}$$

This shows that μ' and ν' are at distance at most $d_{\mathcal{W}}(\mu, \nu)$. Applying the same reasoning to $\Phi(\varphi)^{-1}$, we get that $d_{\mathcal{W}}(\mu', \nu') = d_{\mathcal{W}}(\mu, \nu)$ and $\Phi(\varphi)$ is an isometry. \square

6.2. Semidirect product decomposition. — The Dirac masses are the only measures such that any geodesic issued from them can be extended for all times (given any other measure, the geodesic pointing to any Dirac mass cannot be extended past it, see in Section 2.3 the paragraph on dilations). As a consequence, an isometry of $\mathcal{W}_2(\mathbb{R}^n)$ must globally preserve the set of Dirac masses

As in the case of the line, if we let $\mathcal{L} = \# \text{Isom}(\mathbb{R}^n)$ and \mathcal{R} be the set of isometries of $\mathcal{W}_2(\mathbb{R}^n)$ that fix each Dirac mass, then $\text{Isom } \mathcal{W}_2(\mathbb{R}^n) = \mathcal{L} \ltimes \mathcal{R}$. We proved above that \mathcal{R} contains a copy of $O(n)$, and is in particular non-trivial.

Moreover, if one conjugates a $\Phi(\varphi)$ by some $\#(\psi)$, where $\varphi \in O(n)$ and $\psi \in \text{Isom } \mathbb{R}^n$, one gets the map $\Phi(\vec{\psi}\varphi\vec{\psi}^{-1})$ (it is sufficient to check this on some measure μ in the easy cases when ψ is a translation or fixes the center of mass of μ). Therefore, the action of \mathcal{L} on $O(n) \subset \mathcal{R}$ in the semidirect product is as asserted in Theorem 1.2.

To deduce Theorem 1.2, we are thus left with proving that an isometry that fixes pointwise all Dirac masses must be of the form $\Phi(\varphi)$ for some $\varphi \in O(n)$.

6.3. Measures supported on subspaces. — The following lemma will be used to prove that isometries of $\mathcal{W}_2(\mathbb{R}^n)$ must preserve the property of being supported on a proper subspace.

Lemma 6.2. — *Let $\mu, \nu \in \mathcal{W}_2(\mathbb{R}^n)$, denote by g, h their centers of mass and let $\sigma = d_{\mathcal{W}}(\mu, \delta_g)$ and $\rho = d_{\mathcal{W}}(\nu, \delta_h)$. The equality*

$$(7) \quad d_{\mathcal{W}}^2(\mu, \nu) = d^2(g, h) + \sigma^2 + \rho^2$$

holds if and only if there are two orthogonal affine subspaces L and M such that $\mu \in \mathcal{W}_2(L)$ and $\nu \in \mathcal{W}_2(M)$.

Proof. — Let us first prove that $d^2(g, h) + \sigma^2 + \rho^2$ is the cost B of the independent coupling $\Pi = \mu \otimes \nu$:

$$\begin{aligned}
B &:= \int_{\mathbb{R}^n \times \mathbb{R}^n} d^2(x, y) \Pi(dxdy) \\
&= \int_{\mathbb{R}^n \times \mathbb{R}^n} |(x - g) - (y - h) + (g - h)|^2 \Pi(dxdy) \\
&= \sigma^2 + \rho^2 + d^2(g, h) - 2 \left(\int_{\mathbb{R}^n} (x - g) \mu(dx) \right) \cdot \left(\int_{\mathbb{R}^n} (y - h) \nu(dy) \right) \\
&\quad + 2(g - h) \cdot \left(\int_{\mathbb{R}^n} (x - g) \mu(dx) \right) - 2(g - h) \cdot \left(\int_{\mathbb{R}^n} (y - h) \nu(dy) \right) \\
&= \sigma^2 + \rho^2 + d^2(g, h)
\end{aligned}$$

since by definition $g = \int x \mu(dx)$ and $h = \int y \nu(dy)$.

As a consequence, (7) holds if and only if the independent coupling is optimal.

If μ has two point x, y in its support and ν has two points z, t in its support such that (xy) is not orthogonal to (zt) , then either $(x - y) \cdot (z - t) < 0$ or $(x - y) \cdot (t - z) < 0$. Then by cyclical monotonicity (see Lemma 2.1) the support of an optimal coupling cannot contain (x, z) and (y, t) (in the first case) or (z, x) and (y, t) (in the second case) and thus cannot be the independent coupling.

As a consequence, (7) holds if and only if μ and ν are supported on two orthogonal affine subspaces. \square

Lemma 6.3. — *Isometries of $\mathcal{W}_2(\mathbb{R}^n)$ send hyperplane supported-measures on hyperplane supported measures. Moreover, if two measures are supported on parallel hyperplanes, then their images by any isometry are supported by parallel hyperplanes*

Proof. — Let $\mu \in \mathcal{W}_2(\mathbb{R}^n)$ be supported by some hyperplane H and Φ be an isometry of $\mathcal{W}_2(\mathbb{R}^n)$. Let ν be any measure that is supported on a line orthogonal to H , and that is not a Dirac mass. Then (7) holds (with the same notation as above).

Let μ' and ν' denote the images of μ and ν by Φ . We know that Φ must map δ_g and δ_h to Dirac masses $\delta_{g'}$ and $\delta_{h'}$. Since the center of mass of an element of $\mathcal{W}_2(\mathbb{R}^n)$ is uniquely defined as its projection on the set of Dirac masses, g' is the center of mass of μ' and h' is that of ν' . We get, denoting by σ' and ρ' the distances of μ' and ν' to their centers of mass,

$$\begin{aligned}
d_{\mathcal{W}}^2(\mu', \nu') &= d_{\mathcal{W}}^2(\mu, \nu) \\
&= d^2(g, h) + \sigma^2 + \rho^2 \\
&= d^2(g', h') + \sigma'^2 + \rho'^2
\end{aligned}$$

which implies that μ' and ν' are supported on orthogonal subspaces L and M of \mathbb{R}^n . Since ν' is not a Dirac mass, M is not a point and L is contained in some hyperplane H' .

Moreover, if μ_1 is another measure supported on a hyperplane parallel to H , then its image μ'_1 is supported on some subspace orthogonal to M . It follows that we can find a hyperplane parallel to H' that contains the support of μ'_1 . \square

We are now equipped to end the proof of Theorem 1.2 by induction on the dimension.

6.4. Non-existence of exotic isometries: case of the plane. — Given a line $L \subset \mathbb{R}^2$, denote by $\mathcal{W}_2(L)$ the subset of $\mathcal{W}_2(\mathbb{R}^2)$ consisting of all measures whose support is a subset of L . An optimal coupling between two points of $\mathcal{W}_2(L)$ must have its support in $L \times L$, thus $\mathcal{W}_2(L)$ endowed with the restriction of the distance of $\mathcal{W}_2(\mathbb{R}^2)$ is isometric to $\mathcal{W}_2(\mathbb{R})$. More precisely, given any isometry $\psi : \mathbb{R} \rightarrow L$, we get an isometry $\psi_\# : \mathcal{W}_2(\mathbb{R}) \rightarrow \mathcal{W}_2(L)$.

Lemma 6.3 ensures that any isometry Φ maps a line-supported measure to a line-supported measure, and that the various measures in $\mathcal{W}_2(L)$ are mapped to measures supported on parallel lines.

We assume from now on that Φ fixes each Dirac mass and, up to composing it with some $\Phi(\varphi)$, that it preserves globally $\mathcal{W}_2(L)$ for some L (the axis $\mathbb{R} \times \{0\}$ say). We can moreover assume that its restriction to $\mathcal{W}_2(L)$ is a Φ^t for some t .

Lemma 6.4. — *Let Φ be an isometry of $\mathcal{W}_2(\mathbb{R}^2)$ that fixes Dirac masses, preserves globally $\mathcal{W}_2(L)$ and such that its restriction to this subspace is the time t of the exotic isometric flow.*

Then t must be 0 and up to composing with some $\Phi(\varphi)$, we can assume that Φ preserves $\mathcal{W}_2(M)$ for all line M .

Proof. — We identify a measure $\mu \in \mathcal{W}_2(\mathbb{R})$ with its image by the usual embedding that identifies \mathbb{R} with the axis L . We denote by θL the rotate of L by an angle θ around the origin.

Denote by $\mu(x, \sigma, p, \theta)$ the combination of two Dirac masses that is the image of $\mu(x, \sigma, p)$ if $\theta = 0$, and its rotate around x by an angle θ otherwise. If $\theta \leq \pi/2$, one gets

$$(8) \quad d_{\mathcal{W}}^2(\mu(0, 1, p, 0), \mu(0, 1, q, \theta)) = 2 - 2e^{|p-q|} \cos \theta$$

This shows in particular that the measures supported on θL and with center of mass 0 must be mapped to measures supported on $\pm \theta L$. Up to composing with $\Phi(\varphi)$ where φ is the orthogonal symmetry with respect to L , we can assume that the measures supported on $\frac{\pi}{3}L$ and with center of mass 0 are mapped to measures supported on $\frac{\pi}{3}L$.

Then the measures supported on θL and with center of mass 0 must be mapped to measures supported on a line that crosses L and $\frac{\pi}{3}L$ with angles $\pm \theta$ and $\pm(\theta - \frac{\pi}{3})$. They are therefore mapped to measures supported on θL .

Using the same argument and Lemma 6.3, we get that Φ must preserve $\mathcal{W}_2(M)$ for all lines M .

Moreover, from equation (8) we deduce that if the restriction of Φ to $\mathcal{W}_2(L)$ is the time t of the exotic isometric flow, then for all $\theta < \pi/2$ its restriction to θL also is. But applying the same reasoning to $\frac{\pi}{3}L$ and $\frac{2\pi}{3}L$, then to $\frac{2\pi}{3}L$ and $\pi L = L$ we see that the restriction of Φ to $\mathcal{W}_2(L)$ with the reversed orientation must be the time t of the exotic isometric flow. This implies $t = -t$ thus $t = 0$. \square

The case $n = 2$ of Theorem 1.2 is now reduced to the following.

Lemma 6.5. — *If an isometry Φ of $\mathcal{M}_2(\mathbb{R}^2)$ fixes pointwise the set of line-supported measures, then it must be the identity.*

Proof. — This is a consequence of Radon's theorem, which asserts that a function (compactly supported and smooth, say) in \mathbb{R}^n is characterized by its integrals along all hyperplanes [14] (also see [8]).

Given $\mu \in \mathcal{M}_2(\mathbb{R}^2)$, one can determine by purely metric means its orthogonal projection on any fixed line L : it is its metric projection, that is the unique ν supported on L that minimizes the distance $d_{\mathcal{M}}(\mu, \nu)$.

Now if μ has smooth density and is compactly supported, then the integral of its density along any line L is exactly the density at point $M \cap L$ of its projection onto any line M orthogonal to L .

Therefore, Φ must fix every measure $\mu \in \mathcal{M}_2(\mathbb{R}^2)$ that has a smooth density and is compactly supported. They form a dense set of $\mathcal{M}_2(\mathbb{R}^2)$ thus Φ must be the identity. \square

6.5. Non-existence of exotic isometries: general case. — We end the proof of Theorem 1.2 by an induction on the dimension. There is nothing new compared to the case of the plane, so we stay sketchy.

Let Φ be an isometry of $\mathcal{M}_2(\mathbb{R}^n)$ that fixes pointwise the Dirac masses. It must map every hyperplane-supported measure to a hyperplane-supported measure. Using a non-trivial isometry, we can assume that for some hyperplane L , Φ globally preserves the set $\mathcal{M}_2(L)$ of measures supported on L . Thanks to the induction hypothesis, we can compose Φ with another non-trivial isometry to ensure that Φ fixes $\mathcal{M}_2(L)$ pointwise.

Let μ be a measure supported on some hyperplane $M \neq L$. Let M' be a hyperplane supporting $\Phi(\mu)$. Then as in the case of the plane, it is easy to show that the dihedral angle of (L, M) equals that of (L, M') . Moreover, all measures supported on $L \cap M$ are fixed by Φ , and we conclude that $M' = M$ (up to composition with $\Phi(\varphi)$ where φ is the orthogonal symmetry with respect to L).

The same argument shows that Φ preserves $\mathcal{M}_2(M)$ for all hyperplanes M .

A measure supported on M is determined, if its dihedral angle with L different from $\pi/2$, by its orthogonal projection onto L . since Φ fixes $\mathcal{M}_2(L)$ pointwise, it must fix $\mathcal{M}_2(M)$ pointwise as well. When M is orthogonal to L , the use of a third hyperplane not orthogonal to M nor L yields the same conclusion.

Now that we know that Φ fixes every hyperplane-supported measure, we can use the Radon Theorem to conclude that it is the identity.

7. 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 embedding of \mathbb{R}^k into X .

As a consequence of Proposition 3.6, we get the following result announced in the introduction.

Theorem 7.1. — *The space $\mathcal{M}_2(\mathbb{R}^n)$ has rank n .*

Proof. — An isometric embedding $e : \mathbb{R}^{n+1} \rightarrow \mathcal{W}_2(\mathbb{R}^n)$ must map a geodesic to a geodesic, since they are precisely those curves γ satisfying

$$d(\gamma(t), \gamma(s)) = v|t - s|$$

for some constant v . The union of complete geodesics through any point μ in the image of e would contain a copy of \mathbb{R}^{n+1} , but Proposition 3.6 shows that this union is isometric to \mathbb{R}^n . \square

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

Definition 7.2. — 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 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 .

Let us recall that a map $f : Y \rightarrow X$ is said to be a quasi-isometric embedding if there are constants $C > 1, D > 0$ such that for all $y, z \in Y$ the following holds :

$$C^{-1}d(y, z) - D \leq d(f(y), f(z)) \leq Cd(y, z) + D.$$

The notion of loose rank is relevant in a large class of metric spaces, including discrete spaces (the Gordian space [7], 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 7.3. — *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

Proposition 7.4. — *The semi-global rank and the loose rank of $\mathcal{W}_2(\mathbb{R})$ are infinite.*

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

Moreover the map

$$\begin{aligned} \mathbb{R}_{\leq}^k &\rightarrow \mathcal{W}_2(\mathbb{R}) \\ (x_1, \dots, x_k) &\mapsto \sum \frac{1}{k} \delta_{x_i} \end{aligned}$$

is an isometric embedding.

Since \mathbb{R}_{\leq}^k contains arbitrarily large balls, $\mathcal{W}_2(\mathbb{R})$ must have infinite semi-global rank.

Moreover, since \mathbb{R}_{\leq}^k 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

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

Now the vertical projection from $\{x_1 = 0\}$ to \mathcal{C} is bi-Lipschitz. There is therefore a bi-Lipschitz embedding of \mathbb{R}^{k-1} in $\mathcal{W}_2(\mathbb{R})$ and, *a fortiori*, a quasi-isometric embedding of \mathbb{Z}^{k-1} . Therefore $\mathcal{W}_2(\mathbb{R})$ has infinite loose rank. \square

7.1. Ranks of other spaces. — The ranks of $\mathcal{W}_2(\mathbb{R})$ have an influence on those of many spaces due to the following lemma.

Lemma 7.5. — *If X and Y are Polish geodesic spaces, any isometric embedding $\varphi : X \rightarrow Y$ induces an isometric embedding $\varphi_{\#} : \mathcal{W}_2(X) \rightarrow \mathcal{W}_2(Y)$.*

As usual, $\varphi_{\#}$ is defined by: $\varphi_{\#}\mu(A) = \mu(\varphi^{-1}(A))$ for all measurable $A \subset Y$.

Proof. — Since φ is isometric, for any $\mu \in \mathcal{W}_2(X)$, $\varphi_{\#}\mu$ is in $\mathcal{W}_2(Y)$. Moreover any optimal transportation plan in X is mapped to an optimal transportation plan in Y (note that a coupling between two measures with support in $\varphi(X)$ must have its support contained in $\varphi(X) \times \varphi(X)$). Integrating the equality $d(\varphi(x), \varphi(y)) = d(x, y)$ yields the desired result. \square

Corollary 7.6. — *If X is a Polish geodesic space that contains a complete geodesic, then $\mathcal{W}_2(X)$ has infinite semi-global rank and infinite loose rank. As a consequence, $\mathcal{W}_2(X)$ is not δ -hyperbolic.*

Proof. — Follows from the preceding Lemma, Proposition 7.4 and Proposition 7.3. \square

This obviously applies to $\mathcal{W}_2(\mathbb{R}^n)$.

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

$$P : \mathcal{W}_2(X) \rightarrow X$$

could give a higher bound on the rank of $\mathcal{W}_2(X)$ 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 $\mathbb{R}H^2$ 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 $\mathcal{W}_2(\mathbb{R}H^2)$ that is mapped by P to a curve with the same endpoints than γ , but is different from it. Therefore, it cannot be a geodesic.

8. Open problems

Since the higher-dimensional Euclidean spaces are more rigid (has few non-trivial isometries) than the line, we expect other spaces to be even more rigid.

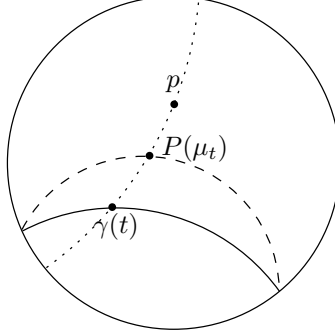


FIGURE 9. The projection P maps a geodesic of $\mathcal{W}_2(\mathbb{R}H^2)$ to a non-geodesic curve (dashed) in $\mathbb{R}H^2$.

Question 1. — Does it exist a Polish (or Hadamard) space $X \neq \mathbb{R}$ such that $\mathcal{W}_2(X)$ admits exotic isometries?

Does it exist a Polish (or Hadamard) space $X \neq \mathbb{R}^n$ such that $\mathcal{W}_2(X)$ admits non-trivial isometries?

In any Hadamard space X , isometries of $\mathcal{W}_2(X)$ 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 in this case.

For general spaces, even the following seems not obvious.

Question 2. — Does it exist a Polish space X whose Wasserstein space $\mathcal{W}_2(X)$ 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 $\mathcal{W}_2(X)$ 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 $\mathcal{W}_2(X)$?

References

- [1] L. AMBROSIO & N. GIGLI — “Construction of the parallel transport in the wasserstein space”, *Methods Appl. Anal.* **15** (2008), no. 1, p. 1–30.
- [2] L. AMBROSIO, N. GIGLI & G. SAVARÉ — *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005.
- [3] Y. BRENIER — “Décomposition polaire et réarrangement monotone des champs de vecteurs”, *C. R. Acad. Sci. Paris Sér. I Math.* **305** (1987), no. 19, p. 805–808.
- [4] ———, “Polar factorization and monotone rearrangement of vector-valued functions”, *Comm. Pure Appl. Math.* **44** (1991), no. 4, p. 375–417.
- [5] M. R. BRIDSON & A. HAEFLIGER — *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319, Springer-Verlag, Berlin, 1999.

- [6] D. BURAGO, Y. BURAGO & S. IVANOV – *A course in metric geometry*, Graduate Studies in Mathematics, vol. 33, American Mathematical Society, Providence, RI, 2001.
- [7] J.-M. GAMBAUDO & É. GHYS – “Braids and signatures”, *Bull. Soc. Math. France* **133** (2005), no. 4, p. 541–579.
- [8] S. HELGASON – *The Radon transform*, Progress in Mathematics, vol. 5, Birkhäuser Boston, Mass., 1980.
- [9] J. JOST – *Nonpositive curvature: geometric and analytic aspects*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 1997.
- [10] N. JUILLET – “Optimal transport and geometric analysis in heisenberg groups”, Ph.D. Thesis, Université Grenoble 1 et Rheinische Friedrich-Wilhelms-Universität Bonn, December 2008.
- [11] M. KNOTT & C. S. SMITH – “On the optimal mapping of distributions”, *J. Optim. Theory Appl.* **43** (1984), no. 1, p. 39–49.
- [12] J. LOTT & C. VILLANI – “Ricci curvature for metric-measure spaces via optimal transport”, *Ann. of Math. (2)* **169** (2009), no. 3, p. 903–991.
- [13] J. LOTT – “Some geometric calculations on Wasserstein space”, *Comm. Math. Phys.* **277** (2008), no. 2, p. 423–437.
- [14] J. RADON – “Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten”, *Ber. Verh. Sächs. Akad. Wiss. Leipzig, Math-Nat. Kl.* **69** (1917), p. 262–277.
- [15] M.-K. VON RENESSE & K.-T. STURM – “Transport inequalities, gradient estimates, entropy, and Ricci curvature”, *Comm. Pure Appl. Math.* **58** (2005), no. 7, p. 923–940.
- [16] C. S. SMITH & M. KNOTT – “Note on the optimal transportation of distributions”, *J. Optim. Theory Appl.* **52** (1987), no. 2, p. 323–329.
- [17] K.-T. STURM – “On the geometry of metric measure spaces. I, II”, *Acta Math.* **196** (2006), no. 1, p. 65–131, 133–177.
- [18] A. TAKATSU – “On wasserstein geometry of the space of gaussian measures”, arXiv:0801.2250, 2008.
- [19] C. VILLANI – *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
- [20] ———, *Optimal transport old and new*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009.