



**HAL**  
open science

# Optimal transportation and dynamics of maps acting on measures, with an emphasis on expanding circle maps

Benoit Kloeckner

► **To cite this version:**

Benoit Kloeckner. Optimal transportation and dynamics of maps acting on measures, with an emphasis on expanding circle maps. 2010. hal-00494397v1

**HAL Id: hal-00494397**

**<https://hal.science/hal-00494397v1>**

Preprint submitted on 23 Jun 2010 (v1), last revised 20 May 2015 (v4)

**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.

---

# OPTIMAL TRANSPORTATION AND DYNAMICS OF MAPS ACTING ON MEASURES, WITH AN EMPHASIS ON EXPANDING CIRCLE MAPS

*by*

Benoît Kloeckner

---

**Abstract.** — Given a map that acts on a compact polish space, one constructs by push-forward a map that acts on the set of probability measures of the space, endowed with the weak topology. We use optimal transportation (mainly with quadratic cost) to study such actions. In a first part we consider the case of an expanding circle map, and show that its action on measures is topologically transitive and has infinite topological entropy. Moreover, we compute its derivative (with respect to the structure of infinite dimensional manifold on the set of measures formalized by Gigli) at its unique absolutely continuous invariant measure. In a second part we generalize one of these results by proving that if a map has positive topological entropy, then its action on measures has infinite topological entropy. It is a consequence of a geometric embedding theorem that is interesting by itself: given any metric space  $X$  and any positive integer  $n$ , there is a bi-Lipschitz embedding of  $X^n$  into the Wasserstein space of  $X$ .

## 1. Introduction

The theory of optimal transport has drawn much attention in recent years. Its applications to geometry and PDEs have in particular been largely disseminated. In this paper, we would like to show its effectiveness in a dynamical context.

The paper is divided in two parts. In the first one, whose goal is to show on a concrete example what one can expect from optimal transport in this context, we are interested in arguably the simplest dynamical system where the action on measures is significantly different from the

action on points, namely expanding circle maps. The second part, on the contrary, gives a very general result on topological entropy.

### 1.1. Results of the first part. —

*1.1.1. An important model example.* — Let us first consider the usual degree 2 self-covering map of the circle  $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$  defined by

$$\Phi(x) = 2x \pmod{1}.$$

It acts on the set  $\mathcal{P}(\mathbb{S}^1)$  of Borel probability measures, endowed with the topology of weak convergence, by the push-forward map  $\Phi_{\#}$ .

A map like  $\Phi$  can act by composition on the right on a Hilbertian function space of the circle (e.g. Sobolev spaces). The adjoint of this map is usually called a Perron-Frobenius operator (or transfer operator), and a great deal of effort has been made to understand these operators, especially their spectral properties (see for example [Bal00]). One can consider  $\Phi_{\#}$  as an analogue for (possibly singular) measures of the Perron-Frobenius operator of  $\Phi$ .

First we study some topological aspects of the dynamics of  $\Phi_{\#}$ .

**Theorem 1.1.** — *The map  $\Phi_{\#} : \mathcal{P}(\mathbb{S}^1) \rightarrow \mathcal{P}(\mathbb{S}^1)$  is topologically transitive and has infinite topological entropy.*

This is restated below as Theorems 2.4 and 3.1; here, following [KH95] we say that a map is topologically transitive if it has a dense orbit.

Then we study the first-order dynamics of  $\Phi_{\#}$  near the uniform measure  $\lambda$ .

**Theorem 1.2.** — *The Gâteaux derivative of  $\Phi_{\#}$  at  $\lambda$  exists (in a sense to be precised) and is twice the Perron-Frobenius operator of  $\Phi$  acting on a suitable  $L^2$ -like function space. Its spectrum is the disc of radius 2.*

In particular,  $\Phi_{\#}$  is very far from being hyperbolic. This result is detailed as Theorem 4.1 and Proposition 4.4 below. We shall also see that  $\Phi_{\#}$  is not Fréchet differentiable.

Let us explain briefly the role of optimal transport in these results. We use the quadratic cost, which gives birth to the so-called  $L^2$  Wasserstein metric. Since the circle is compact, it metrizes the weak topology. We first use it to quantify and control approximations in the proof of topological transitivity; at this point, optimal transportation is convenient but far from necessary though. Second, the very definition of topological entropy needs such a metric. Another, much more specific feature of this

metric is that it can be used to give some sort of infinite-dimensional manifold structure to  $\mathcal{P}(\mathbb{S}^1)$ . We use the point of view developed by Gigli in [Gig09a]. Lemma 4.2 illustrated by figure 3 is the key tool in computing the derivative and seems interesting by itself.

Note that one can also rely on absolutely continuous curves to define the almost everywhere differentiability of maps, see in particular [Gig09b] where this method is applied to the exponential map. Other previous uses of variants of this manifold structure include the definition of gradient flows, as in the pioneering [Ott01] and in [AGS08], and of curvature, as in [Lot08].

*1.1.2. General expanding circle maps.* — The first generalization of this problem is to consider the action on measures of expanding circle maps. It is easily noticed that the proofs of topological conjugacy and topological entropy extends to the higher degree model maps  $x \mapsto dx \pmod{1}$ . In fact, we restrict to  $d = 2$  only to simplify notations. Since expanding circle maps are topologically conjugate to one of these models, Theorem 1.1 extends to all of them without any extra work. Theorem 1.2, however, implies the differential as well as the metric structure of the circle. In the last section of the paper, given a general  $C^2$  expanding map  $\varphi$ , we compute the derivative of  $\varphi_{\#}$  at its unique absolutely continuous invariant measure (Theorem 5.1). Instead of writing down the expression here, let us simply state the following.

**Theorem 1.3.** — *If  $\varphi$  is a  $C^2$  expanding circle map, it has a Gâteaux derivative at its unique invariant absolutely continuous measure, whose adjoint operator (in a suitable  $L^2$ -like function space) is  $u \mapsto \varphi' u \circ \varphi$ .*

In particular this derivative is a multiple of the Perron-Frobenius operator only when  $\varphi'$  is constant, that is when  $\varphi$  is a model map.

*1.1.3. Prospects.* — The next natural question, not addressed at all here, concerns the dynamical properties of the action on measures of higher dimensional hyperbolic dynamical systems like Anosov maps or flows, or of discontinuous systems like interval exchange maps.

Note that it has been proposed to study the dynamics of a map acting on measures using a different topology than the weak one in [Sli08]. The proposed topology is that induced by the  $L^\infty$  Wasserstein metric as introduced in [CDPJ08]. Since this metric dominates the  $L^2$  Wasserstein metric, the topological entropy of  $\Phi_{\#}$  is also infinite in this setting. Since the  $L^\infty$  Wasserstein metric is not separable,  $\Phi_{\#}$  has no dense orbit with

respect to this topology, but the proof of Theorem 2.4 shows that it satisfies the following form of topological transitivity: given any two points, there is an orbit that passes arbitrarily near both these points. Note that this is equivalent to topological transitivity for separable metrizable spaces that are moreover locally compact (see [KH95] Lemma 1.4.2) or complete (see [Sil92]).

**1.2. Results of the second part.** — The dynamical goal of the second part is to prove the following.

**Theorem 1.4.** — *Assume  $X$  is compact and  $\varphi : X \rightarrow X$  is a continuous map with positive topological entropy. Then the action of  $\varphi_{\#}$  on the set of probability measures on  $X$ , endowed with the weak topology, has infinite topological entropy.*

It shall be very easily deduced from a detailed version (Theorem 5.3) of the following geometric result, which involves the  $L^p$  Wasserstein space  $\mathscr{W}_p(X)$  with arbitrary exponent.

**Theorem 1.5.** — *Let  $X$  be any metric space,  $p \in [1, \infty)$  and  $n$  be any positive integer. Then there is a bi-Lipschitz embedding of  $X^n$  into  $\mathscr{W}_p(X)$ .*

This echoes other embedding or non-embedding theorems into Wasserstein spaces. For example, it is a simple matter to prove that as soon as  $X$  contains a complete geodesic, then one can isometrically embed Euclidean balls of arbitrary radius and dimension into  $\mathscr{W}_2(X)$ . Under the same assumption, it is also possible to construct bi-Lipschitz embeddings of Euclidean spaces of arbitrary dimension, see [Klo10]. However, if  $X$  is a globally CAT(−1) space, then it is not possible to embed  $\mathbb{R}^2$  isometrically in  $\mathscr{W}_2(X)$  [BK10].

**1.3. Recalls and notations.** — The most convenient point of view here is to construct the circle as the quotient  $\mathbb{R}/\mathbb{Z}$ . We shall often and without notice write a real number  $x \in [0, 1)$  to mean its image by the canonical projection. We proceed similarly for intervals of length less than 1. Sometimes, it shall be more convenient to use the fundamental domain  $[-1/2, 1/2)$ ; we shall make this choice explicitly when needed.

Recall that the push-forward of a measure is defined by  $\Phi_{\#}\mu(A) = \mu(\Phi^{-1}A)$  for all Borelian set  $A$ .

For a detailed introduction on optimal transport, the interested reader can for example consult [Vil03]. Let us give an overview of the properties we shall need. Given an exponent  $p \in [1, \infty)$ , if  $(X, d)$  is a general metric space, assumed to be polish (complete separable) to avoid measurability issues and endowed with its Borel  $\sigma$ -algebra, its  $L^p$  Wasserstein space is the set  $\mathscr{W}_p(X)$  of probability measures  $\mu$  on  $X$  whose  $p$ -th moment is finite:

$$\int d(x_0, x)^p \mu(dx) \quad \text{for some, hence all } x_0 \in X$$

endowed with the following metric: given  $\mu, \nu \in \mathscr{W}_p(X)$  one sets

$$d(\mu, \nu) = d_p(\mu, \nu) = \left( \inf_{\Pi} \int_{X \times X} d(x, y)^p \Pi(dxdy) \right)^{1/p}$$

where the infimum is over all probability measures  $\Pi$  on  $X \times X$  that projects to  $\mu$  on the first factor and to  $\nu$  on the second one. Such a measure is called a transport plan between  $\mu$  and  $\nu$ , and is said to be optimal when it achieves the infimum. In this setting, an optimal transport plan always exist. Note that when  $X$  is compact, the set  $\mathscr{W}_p(X)$  is equal to the set  $\mathscr{P}(X)$  of all probability measures on  $X$ .

The name “transport plan” is suggestive: it is a way to describe what amount of mass is transported from one region to another. We shall often use this point of view, for example by writing that “ $\Pi$  moves a mass at least  $m$  from  $A$  to  $B$ ” instead of writing that  $\Pi(A \times B) \geq m$ .

The function  $d$  is a metric, called the  $(L^p)$  Wasserstein metric, and when  $X$  is compact it induces the weak topology. Being defined by an infimum, it is usually easy to bound from above, which makes it particularly suitable to prove the topological transitivity. Kantorovich duality can be used to give lower bounds, but here we shall rely on more direct methods since we do not need a very high precision.

Note that the above makes sense also for measure having the same finite mass different than 1. This shall be useful to construct transport plan by parts.

We shall give some details on the differentiable structure on  $\mathscr{P}(\mathbb{S}^1)$  in Section 4.

**PART I**  
**EXPANDING CIRCLE MAPS**

In this first part, we prove Theorems 1.1, 4.1 and 1.3. The emphasis is on the model map  $\Phi$ , but the end of the part is devoted to more general expanding circle maps.

**2. Topological transitivity**

In the proof of the topological transitivity of  $\Phi_{\#}$ , the fixed point  $\delta_0$  (the Dirac mass at 0) is used as a reference point. The proof goes roughly as follows (see figure 1). First, we prove that near every measure lies a measure mapped to  $\delta_0$  by some iterate of  $\Phi_{\#}$ . This measure is obtained by a dyadic discretization. Then we prove that every measure has an iterated antecedent close to  $\delta_0$ . The last intermediate step is to prove that if  $\nu$  is close to  $\delta_0 = \Phi_{\#}^k \mu$ , then  $\nu = \Phi_{\#}^k \mu'$  for some measure  $\mu'$  very close to  $\mu$ . Finally, the separability and completeness of  $\mathcal{P}(S^1)$  lead to the conclusion.

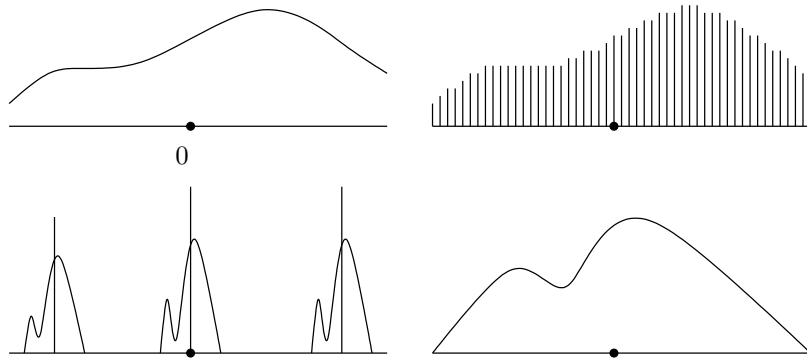


FIGURE 1. To construct a measure close to a given measure (upper left) and having an iterate close to another (lower right), one dyadically discretizes the first one (upper right, Lemma 2.1), then replace the Dirac masses with copies of a concentrated version of the second one (lower left, zoomed in, Lemmas 2.2 and 2.3).

**Lemma 2.1.** — Let  $\mu \in \mathcal{P}(\mathbb{S}^1)$  and  $\varepsilon > 0$ . There is a measure  $\mu'$  and an integer  $k$  such that  $d(\mu, \mu') < \varepsilon$  and  $\Phi_{\#}^k \mu' = \delta_0$ .

*Proof.* — Let  $k$  be such that  $2^{-k-1} < \varepsilon$ , and let  $\mu'$  be the measure supported on the points  $i2^{-k}$ ,  $0 \leq i < 2^k$  and whose mass at point  $i2^{-k}$  is

$$\mu([(i-1/2)2^{-k}, (i+1/2)2^{-k})).$$

There is an obvious coupling between  $\mu$  and  $\mu'$ , where every bit of mass is moved by at most  $2^{-k-1}$ . Therefore,  $d(\mu, \mu') < \varepsilon$ . Moreover, since  $\Phi(i2^{-k}) = 0$ , we have  $\Phi_{\#}^k \mu' = \delta_0$ .  $\square$

**Lemma 2.2.** — Let  $\mu \in \mathcal{P}(\mathbb{S}^1)$  and  $\varepsilon > 0$ . There is a measure  $\nu$  and an integer  $k$  such that  $d(\nu, \delta_0) < \varepsilon$  and  $\Phi_{\#}^k \nu = \mu$ .

*Proof.* — Let  $k$  be such that  $2^{-k-1} < \varepsilon$ . Denote by  $\Phi^{-1}$  the discontinuous right inverse to  $\Phi$  defined by  $\Phi^{-1}(x) = x/2$  for  $x \in [-1/2, 1/2)$ . Then  $\nu = \Phi_{\#}^{-k} \mu$  is a  $\Phi^k$ -antecedent of  $\mu$  and is supported on  $[2^{-k-1}, 2^{-k-1})$ , so that  $d(\nu, \delta_0) < \varepsilon$ .  $\square$

**Lemma 2.3.** — Let  $\mu \in \mathcal{P}(\mathbb{S}^1)$  such that for some integer  $k$ , one has  $\Phi_{\#}^k(\mu) = \delta_0$ . Then for all  $\nu \in \mathcal{P}(\mathbb{S}^1)$ , there is a  $\mu' \in \mathcal{P}(\mathbb{S}^1)$  such that  $\Phi_{\#}^k \mu' = \nu$  and  $d(\mu, \mu') \leq 2^{-k} d(\delta_0, \nu)$ .

*Proof.* — Let  $\nu' = \Phi_{\#}^{-k} \nu$  as defined above, and let  $\mu' = \mu * \nu'$  where  $*$  denotes the convolution of measures:  $\mu'(A) = \int_{\mathbb{S}^1} \nu'(A-x) \mu(dx)$ , where  $A-x$  is the translation by  $-x$  of the Borelian set  $A$ .

From a transport plan  $\Pi$  between measures  $\alpha$  and  $\beta$ , it is easy to construct a transport plan  $\Pi'$  between  $\mu * \alpha$  and  $\mu * \beta$  that has the same cost: simply take  $\Pi'(A \times B) = \int_{\mathbb{S}^1} \Pi(A-x, B-x) \mu(dx)$ .

It follows that  $d(\mu, \mu') \leq d(\delta_0, \nu') = 2^{-k} d(\delta_0, \nu)$ . Moreover one has

$$\begin{aligned} (\Phi_{\#}^k \mu) * (\Phi_{\#}^k \alpha)(A) &= \int_{\mathbb{S}^1} \alpha(\Phi^{-k}(A-x)) \Phi_{\#}^k \mu(dx) \\ &= \int_{\mathbb{S}^1} \alpha(\Phi^{-k}(A - \Phi^k(x))) \mu(dx) \\ &= \int_{\mathbb{S}^1} \alpha(\Phi^{-k}(A) - x) \mu(dx) \\ &= \Phi_{\#}^k(\mu * \alpha)(A) \end{aligned}$$

so that, in particular,  $\Phi_{\#}^k(\mu') = \delta_0 * \nu = \nu$ .  $\square$



**Theorem 2.4.** — *The map  $\Phi_{\#}$  acts topologically transitively on  $\mathcal{P}(\mathbb{S}^1)$ .*

*Proof.* — The space  $\mathcal{P}(\mathbb{S}^1)$  is complete and separable since  $\mathbb{S}^1$  is so (see e.g. [Vil03]). Therefore, it is sufficient to prove that given any two open sets  $U, V \in \mathcal{P}(\mathbb{S}^1)$ , there is a positive integer  $n$  such that  $\Phi_{\#}^n(U) \cap V \neq \emptyset$  (see [Sil92]).

Choose  $\mu \in U$  and  $\nu \in V$  and let  $\varepsilon > 0$  be a number such that  $B(\mu, \varepsilon) \subset U$ .

By Lemma 2.1, there is a measure  $\mu_0$  and a number  $k$  such that  $d(\mu_0, \mu) < \varepsilon/2$  and  $\Phi_{\#}^k \mu_0 = \delta_0$ . By Lemma 2.2, there is a measure  $\nu_0$  and a number  $l$  such that  $d(\nu_0, \delta_0) < 2^{k-1}\varepsilon$  and  $\Phi_{\#}^l \nu_0 = \nu$ . By Lemma 2.3, there is a measure  $\mu_1 = \mu_0 * \nu_0$  such that  $d(\mu_1, \mu_0) < \varepsilon/2$  and  $\Phi_{\#}^k \mu_1 = \nu_0$ . In particular, we get that  $d(\mu_1, \mu) < \varepsilon$  and  $\Phi_{\#}^{k+l} \mu_1 = \nu$ , so that  $\nu \in \Phi_{\#}^{k+l}(U) \cap V$ .  $\square$

### 3. Topological entropy

In this section we consider the topological entropy of  $\Phi_{\#}$ . Let us briefly recall a definition (for more details, [KH95] is an extensive reference). Given a map  $f : X \rightarrow X$  acting on a metric space, for any  $n \in \mathbb{N}$  one defines a new metric on  $X$  by

$$d_n(x, y) := \max\{d(f^k(x), f^k(y)); 0 \leq k \leq n\}.$$

Given  $\varepsilon > 0$ , one says that a subset  $S$  of  $X$  is  $(n, \varepsilon)$ -separated if  $d_n(x, y) \geq \varepsilon$  whenever  $x \neq y \in S$ . Denoting by  $N(f, \varepsilon, n)$  the maximal size of a  $(n, \varepsilon)$ -separated set, the topological entropy of  $f$  is defined as

$$h(f) := \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n} \log N(f, \varepsilon, n).$$

Note that this limit exists since  $\limsup_{n \rightarrow +\infty} \frac{1}{n} \log N(f, \varepsilon, n)$  is nonincreasing in  $\varepsilon$ . The adjective “topological” is relevant since  $h(f)$  does not depend upon the distance on  $X$ , but only on the topology it defines.

The topological entropy is in some sense a global measure of the dependence on initial condition of the considered dynamical system.

The map  $\Phi$  is a classical example, whose topological entropy is  $\log 2$ .

**Theorem 3.1.** — *The topological entropy of  $\Phi_{\#}$  is infinite.*

This result is not surprising since the Wasserstein space can in some ways be seen as an infinite product of the original space, and  $\Phi_{\#}$  as the

diagonal action of  $\Phi$  on this product. This point of view is developed in the second part, whose content is more geometric, and where a large generalization of Theorem 3.1 is proved. We nevertheless provide here a simpler and more direct proof in the particular case of  $\Phi_{\#}$ .

*Proof.* — To construct a large  $(n, \varepsilon)$ -separated set, we proceed as follows: we start with the point  $\delta_0$ , and choose a  $\varepsilon$ -separated set of its inverse images. Then we inductively choose  $\varepsilon$ -separated sets of inverse images of each elements of the set previously defined. Doing this, we need not control the distance between inverse images of two different elements.

Let  $k > 1$  be an integer ( $\varepsilon$  will be exponential in  $-k$ ). Let  $A_k$  be the set all  $\mu \in \mathcal{P}(\mathbb{S}^1)$  such that  $\mu((1 - 2^{-k}, 1)) = 0$  and  $\mu([0, 1/2]) \geq 1/2$ . These conditions are designed to bound from below the distances between the antecedents to be constructed: a given amount of mass (second condition) will have to travel a given distance (first condition).

An element  $\mu \in A_k$  decomposes as  $\mu = \mu_h + \mu_t$  where  $\mu_h$  is supported on  $[0, 1 - 2^{-k+1}]$  and  $\mu_t$  is supported on  $(1 - 2^{-k+1}, 1 - 2^{-k})$ . Let  $e_0$  and  $e_1$  be the right inverses to  $\Phi$  defined onto  $[0, 1/2)$  and  $[1/2, 1)$  respectively. For all integer  $\ell$  between  $2^{k-1}$  and  $2^k$ , define

$$\mu_{\ell} = (e_0)_{\#}(\ell 2^{-k} \mu_h + \mu_t) + (e_1)_{\#}((1 - \ell 2^{-k}) \mu_h)$$

(see figure 2). It is a probability measure on  $\mathbb{S}^1$ , lies in  $A_k$  and  $\Phi_{\#} \mu_{\ell} = \mu$ . Moreover, if  $\ell' \neq \ell$  then any transport plan from  $\mu_{\ell}$  to  $\mu_{\ell'}$  has to move a mass at least  $2^{-k-1}$  by a distance at least  $2^{-k}$ . Therefore,  $d(\mu_{\ell}, \mu_{\ell'}) \geq 2^{-3k/2-1/2}$ .

Let  $\varepsilon = 2^{-3k/2-1/2}$  and define  $S_n$  inductively as follows. First,  $S_0 = \{\delta_0\}$ . Given  $S_n \subset A_k$ ,  $S_{n+1}$  is the set of all  $\mu_{\ell}$  constructed above, where  $\mu$  runs through  $S_n$ .

By construction,  $S_n$  has  $n^{2^{k-1}} = n^{2^{-4/3} \varepsilon^{-2/3}}$  elements. Let  $\mu, \nu$  be two distinct elements of  $S_n$  and  $m$  be the greatest index such that  $\Phi_{\#}^m \mu \neq \Phi_{\#}^m \nu$ . Since  $\Phi_{\#}^n \mu = \delta_0 = \Phi_{\#}^n \nu$ ,  $m$  exists and is at most  $n - 1$ . The measures  $\mu' = \Phi_{\#}^m \mu$  and  $\nu' = \Phi_{\#}^m \nu$  both lie in  $S_{n-m}$  and have the same image. Therefore, they are  $\varepsilon$ -separated. This shows that  $S_n$  is  $(n, \varepsilon)$ -separated.

It follows that  $\frac{1}{n} \log N(\Phi_{\#}, \varepsilon, n) \geq 2^{-4/3} \varepsilon^{-2/3}$ , which goes to  $+\infty$  when  $\varepsilon \rightarrow 0$ .  $\square$

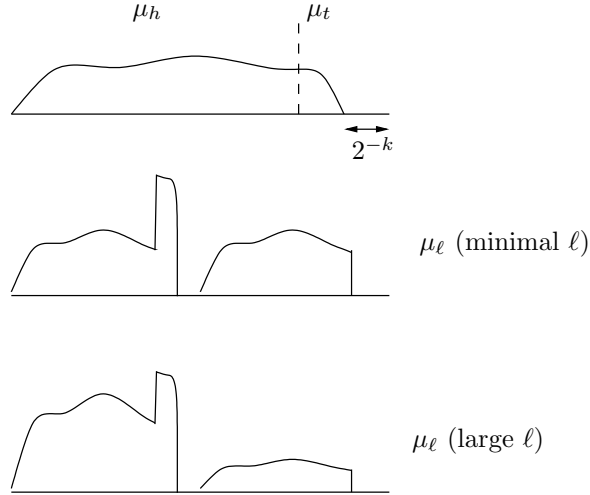


FIGURE 2. Construction of separated antecedents of a given measure.

#### 4. First-order dynamics near the uniform measure

In this section we recall without proofs the “differential structure” induced on  $\mathcal{P}(\mathbb{S}^1)$  by the Wasserstein metric, following the point of view of [Gig09a], and we show that  $\Phi_\#$  is (weakly) differentiable at the point  $\lambda$ . Its differential map  $D_\lambda \Phi_\#$  is an explicit, simple endomorphism of a Hilbert space, and we shall give a brief study of its spectrum.

Note that considering the Wasserstein space of a Riemannian manifold as an infinite-dimensional Riemannian manifold dates back to Otto [Ott01]. However, in many ways it stayed a formal view until the work of Gigli.

**4.1. The first-order differential structure on measures.** — This subsection contains no novelty, but only recalls the aforementioned work of Gigli in the particular case of the circle. Note that as is customary in these topics, by a geodesic we mean a non-constant globally minimizing geodesic segment or line, parametrized proportionally to arc length.

Given  $\mu \in \mathcal{P}(\mathbb{S}^1)$ , there are several equivalent ways to define its tangent space  $T_\mu$ . In fact,  $T_\mu$  has a vectorial structure only when  $\mu$  is atomless; otherwise it is only a tangent cone. Note that the atomless condition has to be replaced by a more intricate one in higher dimension.

The most Riemannian way to construct  $T_\mu$  is to use the exponential map. Let  $L^2(\mu)$  be the set of probability measures on the tangent bundle  $T\mathbb{S}^1$  that are mapped to  $\mu$  by the canonical projection. Given  $\xi, \zeta \in L^2(\mu)$ , one defines

$$d(\xi, \zeta) = \left( \inf_{\Pi} \int_{T\mathbb{S}^1 \times T\mathbb{S}^1} d(x, y)^2 \Pi(dx dy) \right)^{1/2}$$

where  $d$  is any metric whose restriction to the fibers is the Riemannian distance (here the fibers are isometric to  $\mathbb{R}$ ), and the infimum is over transport plans  $\Pi$  that are mapped to the identity  $(\text{Id} \times \text{Id})_{\#}\mu$  by the canonical projection on  $\mathbb{S}^1 \times \mathbb{S}^1$ . This means that we allow only to move the mass *along* the fibers. Equivalently, one can disintegrate  $\xi$  and  $\zeta$  along  $\mu$ , writing  $\xi = \int \xi_x \mu(dx)$  and  $\zeta = \int \zeta_x \mu(dx)$ , with  $(\xi_x)_{x \in \mathbb{S}^1}$  and  $(\zeta_x)_{x \in \mathbb{S}^1}$  two families of probability measures on  $T_x\mathbb{S}^1 \simeq \mathbb{R}$  uniquely defined up to a set of measure zero. Then one gets

$$d(\xi, \zeta)^2 = \int_{\mathbb{S}^1} d(\xi_x, \zeta_x)^2 \mu(dx)$$

where one integrates the squared Wasserstein metric (with respect to the Riemannian metric, that is  $|\cdot|$ ).

There is a natural cone structure on  $L^2(\mu)$ , extending the scalar multiplication on the tangent bundle: letting  $D_r$  be the dilation of ratio  $r$  on  $T\mathbb{S}^1$ , one defines  $r \cdot \xi := (D_r)_{\#}\xi$ .

The exponential map  $\exp : T\mathbb{S}^1 \rightarrow \mathbb{S}^1$  now gives a map  $\exp_{\#} : L^2(\mu) \rightarrow \mathcal{P}(\mathbb{S}^1)$ . The point is that not for all  $\xi \in L^2(\mu)$ , is there a  $\varepsilon > 0$  such that  $t \mapsto \exp_{\#}(t \cdot \xi)$  defines a geodesic of  $\mathcal{P}(\mathbb{S}^1)$  on  $[0, \varepsilon)$ . Consider for example  $\mu = \lambda$ , and  $\xi$  be defined by  $\xi_x \equiv 1$ . Then  $\exp_{\#}(t \cdot \xi) = \lambda$  for all  $t$ : one rotates all the mass while letting it in place would be more efficient.

The first definition is that  $T_\mu$  is the closure in  $L^2(\mu)$  of the subset of all  $\xi$  such that  $\exp_{\#}(t \cdot \xi)$  defines a geodesic for small enough  $t$ .

Let us now give another definition, assuming  $\mu$  has no atom. Given a smooth function  $f : \mathbb{S}^1 \rightarrow \mathbb{R}$ , its gradient  $\nabla f : \mathbb{S}^1 \rightarrow T\mathbb{S}^1$  can be used to push  $\mu$  to an element  $\xi_f = (\nabla f)_{\#}\mu$  of  $L^2(\mu)$ . This element has the property that  $\exp_{\#}(t \cdot \xi) = (\text{Id} + t\xi_f)_{\#}\mu$  defines a geodesic for small enough  $t$  (the time bound depending on the maximum of  $|\nabla f|$ ). Now, the closure in  $L^2(\mu)$  of the image of the map  $f \mapsto \xi_f$  is equal to  $T_\mu$ .

In particular, this means that as soon as  $\mu$  is atomless, the disintegration  $(\xi_x)_x$  of an element of  $T_\mu$  writes  $\xi_x = \delta_{v(x)}$  for some function  $v$  and

$\mu$ -almost all  $x$ . Moreover,  $v$  must have mean zero (with respect to the uniform measure  $\lambda$ , since this condition is inherited from the fact that  $v = \nabla f$  for  $\xi$  in a dense set). We get that  $T_\mu$  can be identified with the set  $L_0^2(\mu)$  of functions  $v : \mathbb{S}^1 \rightarrow \mathbb{R}$  that are square-integrable and of mean zero with respect to  $\lambda$ . When  $\mu$  is the uniform measure, we write  $L_0^2$  instead of  $L_0^2(\lambda)$ .

For simplicity, given  $v \simeq \xi \in L_0^2(\mu) \simeq T_\mu$  we shall denote  $\exp_{\#}(t \cdot \xi)$  by  $\mu + tv$ .

Let us end this subsection with a remark. One could consider simpler to just take the smooth functions of  $\mathbb{S}^1$  as coordinates to define a smooth structure on  $\mathcal{P}(\mathbb{S}^1)$ . First, it is important to understand that optimal transportation is about pushing mass, not (directly) about recording the variation of density at each point (and to describe the variation of density under a transport plan shall need some work in the sequel). Second, if one considered smooth function as coordinates, then a path of the form  $\gamma_t = t\delta_x + (1-t)\delta_y$  should be considered smooth since  $\int f(u)\gamma_t(du)$  would depend smoothly on  $t$  for all smooth  $f$ . However, the Wasserstein distance between  $\gamma_t$  and  $\gamma_s$  is the order of  $\sqrt{|t-s|}$ , so that  $\gamma_t$  is not rectifiable (it has infinite length)! The least to expect from smooth paths is to be rectifiable, so that the point of view “functions as coordinates” has to be dismissed.

**4.2. Differentiability of  $\Phi_{\#}$ .** — We now turn to the main result of this section.

**Theorem 4.1.** — *Let  $\mathcal{L} : L_0^2 \rightarrow L_0^2$  be the linear operator defined by  $\mathcal{L}v(x) = v(x/2) + v((x+1)/2)$ . Then  $\mathcal{L}$  is the derivative of  $\Phi_{\#}$  at  $\lambda$  in the following sense: for all  $v \in L_0^2 \simeq T_\lambda$ , one has*

$$d(\Phi_{\#}(\lambda + tv), \lambda + t\mathcal{L}(v)) = o(t).$$

First, we recognize in  $\mathcal{L}$  twice the Perron-Frobenius operator of  $\Phi$ , that is the adjoint of the map  $u \mapsto u \circ \Phi$ , acting on the space  $L_0^2$ . Second, we only get a Gâteaux derivative, when one would prefer a Fréchet one, that is a formula of the kind

$$d(\Phi_{\#}(\lambda + v), \lambda + \mathcal{L}(v)) = o(|v|).$$

However, we shall see that such a uniform bound does not hold, at least not in  $L^2$  norm (which is the natural norm here). This shall be discussed later on.

The main point to prove in the above theorem is the following estimate.

**Lemma 4.2.** — Given vector fields  $v_1, \dots, v_n \in L^2$  and positive numbers  $\alpha_1, \dots, \alpha_n$  adding up to 1, one has

$$d\left(\lambda + t \sum_i \alpha_i v_i, \sum_i \alpha_i (\lambda + t v_i)\right) = o(t).$$

*Proof.* — We prove the case  $n = 2$  since the general case can then be deduced by a straightforward induction. Let  $\varepsilon$  be any positive number. Let  $\bar{v}_1$  and  $\bar{v}_2$  be piecewise constant functions that approximate  $v_1$  and  $v_2$  in  $L^2$  norm:  $|v_i - \bar{v}_i|_2 \leq \varepsilon$ . Then  $((\text{Id} + v_i) \times (\text{Id} + \bar{v}_i))_{\#} \lambda$  is a transport plan from  $\lambda + v_i$  to  $\lambda + \bar{v}_i$ , whose cost is precisely  $|v_i - \bar{v}_i|_2^2$ . This shows that  $d(\lambda + v_i, \lambda + \bar{v}_i) \leq \varepsilon$ . We can moreover assume that  $\bar{v}_i$  are constant on each interval of the form  $[i/k, (i+1)/k)$  for some fixed  $k$  (depending only upon  $v_1$  and  $v_2$ ).

To see what happens on such an interval  $I$ , temporarily denoting by  $v_1$  and  $v_2$  the values taken by the functions  $\bar{v}_i$  on  $I$ , let us construct for  $t$  small enough an economic transport plan from  $(\text{Id} + t(\alpha_1 v_1 + \alpha_2 v_2))_{\#} \lambda_I$  to  $\alpha_1 (\text{Id} + t v_1)_{\#} \lambda_I + \alpha_2 (\text{Id} + t v_2)_{\#} \lambda_I$ . Assuming that  $t$  is small enough to ensure that the intervals  $(\text{Id} + t v_1)(I)$  and  $(\text{Id} + t v_2)(I)$  meet, one can simply let the common mass in place and move at each side a mass  $\alpha_1 \alpha_2 |v_1 - v_2| t$  by a distance at most  $|v_1 - v_2| t$  (see figure 3; this is not optimal but sufficient for our purpose). This transport plan has a cost  $t^3 \alpha_1 \alpha_2 |v_1 - v_2|^3$ .

By adding one such plan for each interval  $[i/k, (i+1)/k)$ , we get a transport plan from  $(\text{Id} + t(\alpha_1 \bar{v}_1 + \alpha_2 \bar{v}_2))_{\#} \lambda$  to  $\alpha_1 (\text{Id} + t \bar{v}_1)_{\#} \lambda + \alpha_2 (\text{Id} + t \bar{v}_2)_{\#} \lambda$  whose cost is  $ct^3$  where  $c = \alpha_1 \alpha_2 \int |\bar{v}_1 - \bar{v}_2|^3$ . Note that even if the  $v_i$  are only  $L^2$ ,  $\bar{v}_i$  are bounded. Now we have

$$d(\lambda + t(\alpha_1 \bar{v}_1 + \alpha_2 \bar{v}_2), \alpha_1 (\lambda + t \bar{v}_1) + \alpha_2 (\lambda + t \bar{v}_2)) \leq \sqrt{ct^3}$$

so that, for  $t$  small enough,

$$d(\lambda + t(\alpha_1 \bar{v}_1 + \alpha_2 \bar{v}_2), \alpha_1 (\lambda + t \bar{v}_1) + \alpha_2 (\lambda + t \bar{v}_2)) \leq \varepsilon t$$

By triangular inequality, it follows that

$$d(\lambda + t(\alpha_1 v_1 + \alpha_2 v_2), \alpha_1 (\lambda + t v_1) + \alpha_2 (\lambda + t v_2)) \leq 3\varepsilon t.$$

□

*Proof of Theorem 4.1.* — Remark that

$$\Phi_{\#}(\lambda + t v) = \frac{1}{2} (\lambda + 2t v(\cdot/2)) + \frac{1}{2} (\lambda + 2t v((\cdot + 1)/2))$$

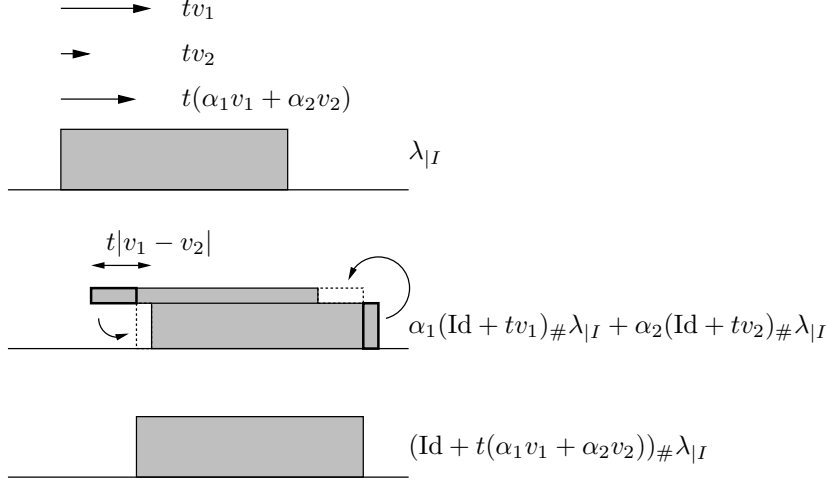


FIGURE 3. The cost of this transport plan has the order of magnitude  $t^3$

and apply the preceding lemma.  $\square$

Let us prove that we cannot hope for the Fréchet differentiability of  $\Phi_\#$ .

**Proposition 4.3.** — *For all positive  $\varepsilon$ , there is a map  $v \in L_0^2$  that satisfies the following:*

1.  $|v|_2 \leq \varepsilon$ ,
2.  $\mathcal{L}v = 0$  so that  $\lambda + \mathcal{L}v = \lambda$ , and
3.  $d(\Phi_\#(\lambda + v), \lambda) \geq c\varepsilon$

for some constant  $c$  independent of  $\varepsilon$  and  $v$ .

*Proof.* — Let  $k$  be a positive integer, to be precised later on. Let  $v$  be the piecewise affine map defined as follows (see figure 4):  $v(x) = 1/(4k) - y$  when  $x = i/(2k) + y$  with  $y \in [0, 1/(2k))$  and  $0 \leq i < k$  an integer, and  $v(x) = -1/(4k) + y$  when  $x = i/(2k) + y$  with  $y \in [0, 1/(2k))$  and  $k \leq i < 2k$ . We have  $|v|_2^2 = (4k)^{-2}/3$  so that taking  $k \geq \frac{\sqrt{3}}{4}\varepsilon^{-1}$  ensures point 1. Moreover, 2 is straightforward, and have left to prove that  $k$  chosen with the order of  $\varepsilon^{-1}$  gives 3.

On any small enough interval  $I$ , if  $w$  is an affine function of slope  $-1$  with a zero at the center of  $I$ , then  $\lambda|_I + w$  is a Dirac mass at the center of  $I$  (each element of mass is moved to the center). If  $w$  has slope 1, then

the mass moves in the other direction, and  $\lambda|_I + w$  is uniform of density  $1/2$  on the interval  $I'$  having the same center than  $I$  and twice as long. By combining these two observations, one deduces that

$$\mu := \Phi_{\#}(\lambda + v) = 1/2\lambda + \sum_{i=1}^k \frac{1}{2k} \delta_{\frac{i-1/2}{k}}.$$

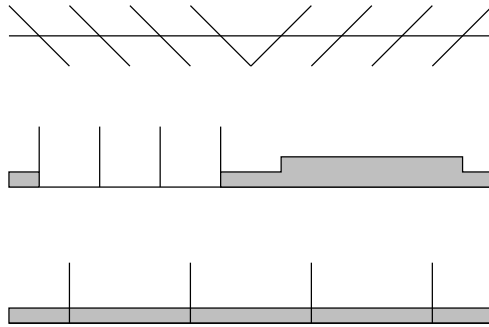


FIGURE 4. The case  $k = 4$ . Up: the graph of  $v$ ; middle:  $\lambda + v$ ; down:  $\Phi_{\#}(\lambda + v)$ .

Each interval of the form  $I_i = [(i - 5/8)/k, (i - 3/8)/k]$  is given by  $\lambda$  a mass  $1/(4k)$ . The discrete part of  $\mu$  consists in a Dirac mass of weight  $1/(2k)$  at the center of each  $I_i$ . Any transport plan from  $\mu$  to  $\lambda$  must therefore move a mass at least  $1/(4k)$  from each of these Dirac masses to the outside of  $I_k$ , so that a total mass at least  $1/4$  has to move a distance at least  $1/(8k)$ . From this it follows that  $d(\lambda, \mu) \geq 1/(16k)$ . When  $k$  is chosen with the order of  $\varepsilon^{-1}$ , this distance has at least the order of  $\varepsilon$ , as required. More precisely, the constant  $c$  is (up to rounding up)  $(4\sqrt{3})^{-1}$ .  $\square$

**4.3. Spectral study of  $\mathcal{L}$ .** — Let us compute the spectrum of  $\mathcal{L} = D_{\lambda}(\Phi_{\#})$ . The following proposition is very elementary and probably not new, but we produce a proof for the sake of completeness.

**Proposition 4.4.** — *A number  $\alpha$  is an eigenvalue of  $\mathcal{L}$  if and only if  $|\alpha| < 2$ . Moreover, each eigenvalue has an infinite-dimensional eigenspace. Last, the spectrum of  $\mathcal{L}$  is the closed disc of radius 2.*

In particular, it is to be noted that  $\lambda$  is not an hyperbolic fixed point of  $\Phi_{\#}$ .



The proof of Proposition 4.4 consist simply in using Fourier series to show that (up to a multiplicative constant)  $\mathcal{L}$  is conjugated to a countable product of the shift on  $\ell^2(\mathbb{N})$ .

*Proof.* — Let  $c_k$  denote the function  $x \mapsto \cos(2\pi kx)$  defined on the circle, and  $s_k : x \mapsto \sin(2\pi kx)$ . Then it is readily checked that  $\mathcal{L}c_k = \mathcal{L}s_k = 0$  when  $k$  is odd, and  $\mathcal{L}c_k = 2c_{k/2}$ ,  $\mathcal{L}s_k = 2s_{k/2}$  when  $k$  is even.

Let  $\sigma$  be the shift of the Hilbert space  $\ell^2 = \ell^2(\mathbb{N})$  of  $\mathbb{N}$ -indexed square integrable sequences: if  $\underline{x} = (x_0, x_1, x_2, \dots)$  then  $\sigma\underline{x} = (x_1, x_2, x_3, \dots)$ . Let  $\sigma^{\mathbb{N}}$  be the direct product of  $\sigma$ , acting on  $(\ell^2)^{\mathbb{N}}$ . Then the map

$$\begin{aligned} \Psi : (\ell^2)^{\mathbb{N}} &\rightarrow L_0^2 \\ X = (\underline{x}^0, \underline{x}^1, \underline{x}^2, \dots) &\mapsto \sum_{i,j \in \mathbb{N}} x_j^{2i} c_{(2i+1)2^j} + x_j^{2i+1} s_{(2i+1)2^j} \end{aligned}$$

is an isomorphism (and even an isometry) that intertwins  $\sigma^{\mathbb{N}}$  and  $\frac{1}{2}\mathcal{L}$ . The spectral study of  $\mathcal{L}$  therefore reduces to that of  $\sigma$ .

A non-zero eigenvector of  $\sigma$ , associated to an eigenvalue  $\alpha$ , must have the form  $(x, \alpha x, \alpha^2 x, \dots)$  with  $x \neq 0$ . Such a sequence is square integrable if and only if  $|\alpha| < 1$ . Moreover the operator norm of  $\sigma$  is 1, so that its complex spectrum is a subset of the closed unit disc. Since the spectrum is closed, and contains the set of eigenvalues, it is equal to the closed unit disc.  $\square$

**4.4. Discussion on the non-Fréchet differentiability and the non-hyperbolicity.** — The counter-example to the Fréchet differentiability of  $\Phi_{\#}$  at  $\lambda$  has high total variation, and it is likely that using a norm that controls variations (e.g. a Sobolev norm) on (a subspace of)  $T_{\lambda}$  shall provide a uniform error bound.

Moreover, up to multiplication by 2 the derivative  $\mathcal{L}$  is the Perron-Frobenius operator of  $\Phi$ , and such operators have far more subtle spectral properties when defined over Sobolev spaces.

For these two reasons, it seems that one could search for a modification of optimal transport that would give a manifold structure to  $\mathcal{P}(\mathbb{S}^1)$ , in such a way that  $T_{\lambda}$  identifies with a Sobolev space. A way to achieve this could be to penalize not only the distance by which a transport plan moves mass, but also the distorsion, that is the variation of the pairwise distances of the elements of mass. This should impose more regularity to optimal transport plans.

### 5. First-order dynamics for general expanding maps

In this last section, we consider a general map  $\varphi : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ , assumed to be  $C^2$  and expanding, *i.e.*  $|\varphi'| > 1$ . Such a map is a self-covering, and has a unique absolutely continuous invariant measure (see e.g. [KH95]) which has a positive and  $C^1$  density [Krz77], denoted by  $\rho$ . The measure itself is denoted by  $\rho\lambda$ . Note that as sets,  $L^2(\rho\lambda) = L^2$ , although they differ as Hilbert spaces. All integrals where the variable is implicit are with respect to the Lebesgue measure  $\lambda$ .

The result is as follows.

**Theorem 5.1.** — *The map  $\varphi_{\#}$  has a Gâteaux derivative at  $\rho\lambda$ , which –remembering  $T_{\rho\lambda} \simeq L_0^2(\rho\lambda)$ – is given by*

$$\begin{aligned} \mathcal{K} : L_0^2(\rho\lambda) &\rightarrow L_0^2(\rho\lambda) \\ v &\mapsto \sum_{y \in \varphi^{-1}(x)} \frac{\rho(y)}{\rho(x)} v(y) - \frac{\int v \varphi' \frac{\rho}{\rho \circ \varphi}}{\rho \int 1/\rho} \end{aligned}$$

The adjoint operator of  $\mathcal{K}$  in  $L_0^2(\rho\lambda)$  is given by

$$\mathcal{K}^* u = \varphi' u \circ \varphi.$$

The end of the section is devoted to the proof of this result.

First, as in the case of  $\Phi_{\#}$ , Lemma 4.2 shows that for  $v \in L_0^2(\rho\lambda)$ ,

$$(1) \quad d\left(\varphi_{\#}(\rho\lambda + tv), \rho\lambda + t\tilde{\mathcal{K}}v\right) = o(t)$$

where

$$\tilde{\mathcal{K}}v(x) = \sum_{y \in \varphi^{-1}(x)} \frac{\rho(y)}{\rho(x)} v(y)$$

is the first term in the expression of  $\mathcal{K}$ . In words, each of the inverse image of  $x$  gives a contribution to the local displacement of mass that is proportional to  $v(y)$  and to  $\rho(y)$ .

This seems very similar to the case of  $\Phi_{\#}$ , except that  $\tilde{\mathcal{K}}$  need not map  $L_0^2(\rho\lambda)$  to itself! Let us stress, once again, that the condition that  $v \in L_0^2(\rho\lambda)$  has mean zero is to be understood *with respect to the uniform measure*  $\lambda$ , since it translates the *metric* property of being (close to) the gradient of a smooth function. This does not prevent Equation (1) to make sense, but shows that  $\tilde{\mathcal{K}}v$  cannot be considered as the directional derivative of  $\varphi_{\#}$  since it does not belong to  $T_{\rho\lambda} = L_0^2(\rho\lambda)$ . In fact, we

shall see that there is another vector field, that lies in  $L_0^2(\rho\lambda)$  and gives the same pushed measure (at least at order 1).

**Proposition 5.2.** — *Given  $\tilde{w} \in L^2(\rho\lambda)$  and assuming that  $\tilde{w}$  is  $C^1$ , there is a  $C^1$  vector field  $w \in L_0^2(\rho\lambda)$  such that  $d(\rho\lambda + t\tilde{w}, \rho\lambda + tw) = o(t)$ . Moreover,  $w$  is given by*

$$w = \tilde{w} + \frac{\int \tilde{w}}{\rho \int 1/\rho}.$$

*Proof.* — To find the suitable  $w$ , let us compute the density  $\tilde{\rho}_t$  of  $\rho\lambda + t\tilde{w}$ . Note that only for small  $t$  is this an absolutely continuous measure, as shown by the example in figure 4. Let  $t$  be small enough so that  $\text{Id} + t\tilde{w}$  is a diffeomorphism (a sufficient condition is that  $|t\tilde{w}'| < 1$ ). Then for all continuous function  $f$ , one has

$$\begin{aligned} \int f(x)(\rho\lambda + t\tilde{w})(dx) &= \int f(x)(\text{Id} + t\tilde{w})_{\#}(\rho\lambda)(dx) \\ &= \int f(x + t\tilde{w}(x))\rho(x)dx \\ &= \int f(y)\frac{\rho}{1 + t\tilde{w}'} \circ (\text{Id} + t\tilde{w})^{-1}(y)dy \end{aligned}$$

by a change of variable. It follows that

$$\tilde{\rho}_t = \frac{\rho}{1 + t\tilde{w}'} \circ (\text{Id} + t\tilde{w})^{-1}.$$

Up to a  $o(t)$  term (which, as all other remainder terms that follows, depends upon  $\rho$ ,  $w$ , possibly  $1/\rho$  and their derivative, all of which are bounded data of our problem), this is equal to

$$\eta_t = \frac{\rho}{1 + t\tilde{w}'} \circ (\text{Id} - t\tilde{w})$$

so that

$$\left. \frac{d}{dt} \right|_{t=0} \tilde{\rho}_t(x) = -\tilde{w}(x)\rho'(x) - \tilde{w}'(x)\rho(x).$$

There exists exactly one vector field  $w$  that is  $C^1$ , has mean zero, and such that  $(\rho w)' = (\rho\tilde{w})'$ : it is given by the claimed formula.

Now we have to prove that if two vector fields induce the same first order perturbation on  $\rho$ , then the distance between the pushed measure is small. To this end, we shall construct a deterministic transport plan with small cost.

Given a probability measure with continuous positive density  $\eta$ , the map  $x \mapsto \int_0^x \eta(u) du$  defines a diffeomorphism of the circle, that maps  $\eta\lambda$  to  $\lambda$ . Call  $\tilde{T}_t$  the map obtained for  $\eta = \tilde{\rho}_t$  and  $T_t$  that given by  $\eta = \rho_t$ . Then the map  $U_t = T_t^{-1} \circ \tilde{T}_t$  is a diffeomorphism that sends  $\tilde{\rho}_t\lambda$  to  $\rho_t\lambda$ . Moreover,  $U_t(x)$  is the first (and only)  $y \in [0, 1)$  such that  $\int_0^y \rho_t(u) du = \int_0^x \tilde{\rho}_t(u) du$ . But since  $\rho_t$  and  $\tilde{\rho}_t$  have the same  $t$ -derivative at 0, we get that the number  $\varepsilon_t := |\rho_t - \tilde{\rho}_t|_\infty$  is an  $o(t)$ . This implies that  $|U_t(x) - x| \leq \varepsilon_t / \min(\tilde{\rho}_t)$  for all  $x$ . Integrating, it comes that  $d(\tilde{\rho}_t\lambda, \rho_t\lambda) = o(t)$ . In other words,  $d(\rho\lambda + t\tilde{w}, \rho\lambda + tw) = o(t)$ .  $\square$

Note that we did not bother to prove the unicity of  $w$ : Gigli's construction shows that the first order perturbation of the measure (with respect to the  $L^2$  Wasserstein metric) characterizes a tangent vector in  $T_\mu$ , see Theorem 5.5 in [Gig09a]. This is one of the convincing features of the framework of Gigli: it gives a way to select one vector field among a set of vector fields that give curves of measures tangent one to each other.

Now if one considers the “centering” operator  $\mathcal{C} : L^2(\rho\lambda) \rightarrow L_0^2(\rho\lambda)$  defined by

$$\mathcal{C}v = v - \frac{\int v}{\rho \int 1/\rho},$$

the derivative of  $\varphi_\#$  at  $\rho\lambda$  is given by the composition  $\mathcal{C}\tilde{\mathcal{H}}$ . Indeed, the previous proposition shows this for a  $C^1$  argument, but  $C^1$  vector fields are dense in  $L_0^2(\rho\lambda)$  and the involved operators are continuous.

To get the expression of  $\mathcal{H}$  given in Theorem 5.1, one only need a change of variable: denoting by  $\varphi_i^{-1}$  ( $i = 1, 2, \dots, d$ ) the right inverses to  $\varphi$  that are onto intervals  $[a_1 = 0, a_2), [a_2, a_3), \dots, [a_d, a_{d+1} = 1)$  one has

$$\begin{aligned} \int \tilde{\mathcal{H}}v &= \sum_i \int \frac{\rho \circ \varphi_i^{-1}}{\rho} v \circ \varphi_i^{-1} \\ &= \sum_i \int_{a_i}^{a_{i+1}} \frac{\rho}{\rho \circ \varphi} v \varphi' \\ &= \int v \varphi' \frac{\rho}{\rho \circ \varphi}. \end{aligned}$$

The computation of the adjoint is a similar change of variable that we omit. Note that the adjoint of the extension to  $L^2(\rho\lambda)$  of  $\mathcal{H}$  (with the

same expression) is

$$u \mapsto \varphi' u \circ \varphi - \frac{\varphi' \int u}{\rho \circ \varphi \int 1/\rho}$$

and the second term vanishes in  $L_0^2(\rho\lambda)$ . The first term is also the adjoint in  $L^2(\rho\lambda)$  of  $\tilde{\mathcal{K}}$ , and this adjoint preserves  $L_0^2(\rho\lambda)$ . In other words,  $\mathcal{K}$  is the adjoint in  $L_0^2(\rho\lambda)$  of the adjoint in  $L^2(\rho\lambda)$  of  $\tilde{\mathcal{K}}$ .

An interesting feature of the expression of  $\mathcal{K}^*$  is that it does not involve the invariant measure. It could therefore be useful to study the spectrum of  $\mathcal{K}$ .

## PART II

### EMBEDDING POWERS OF A SPACE IN ITS WASSERSTEIN SPACE

In this part we prove and discuss the following.

**Theorem 5.3.** — *Let  $X$  be any metric space,  $p \in [1, \infty)$  and  $n$  be any positive integer. Then there is a bi-Lipschitz embedding of  $X^n$  into  $\mathcal{W}_p(X)$ .*

*More precisely, if we endow  $X^n$  with the metric given by*

$$d(\bar{x} = (x_1, \dots, x_n), \bar{y} = (y_1, \dots, y_n)) = \left( \sum_{i=1}^n d(x_i, y_i)^p \right)^{1/p}.$$

*then there exist a map  $f : X^n \rightarrow \mathcal{W}_p(X)$  such that*

$$\frac{1}{n(2^n - 1)^{1/p}} d(\bar{x}, \bar{y}) \leq d(f(\bar{x}), f(\bar{y})) \leq \left( \frac{2^{n-1}}{2^n - 1} \right)^{1/p} d(\bar{x}, \bar{y}).$$

*for all  $\bar{x}, \bar{y} \in X^n$  and that intertwines induced dynamical systems in the following sense: for all continuous map  $\varphi : X \rightarrow X$ , its diagonal action  $\varphi^n : X^n \rightarrow X^n$  and the push-forward  $\varphi_\# : \mathcal{W}_p(X) \rightarrow \mathcal{W}_p(X)$  satisfy  $f \circ \varphi^n = \varphi_\# \circ f$ .*

## 6. A dynamical consequence

Let us first deduce from Theorem 5.3 the dynamical result claimed in the introduction: if  $X$  is compact and  $\varphi : X \rightarrow X$  is a continuous map with positive topological entropy, then the action of  $\varphi_{\#}$  on  $\mathcal{P}(X)$  has infinite topological entropy.

*Proof of Theorem 1.4.* — By compactity,  $\mathcal{W}_p(X)$  is the set of all probability measures on  $X$  and the Wassertein distance metrizes the weak topology. Fix some positive integer  $n$ . The map  $f : X^n \rightarrow \mathcal{W}_p(X)$  given by the theorem is a homeomorphism on its image that conjugates  $\varphi^n$  and  $(\varphi_{\#})|_{f(X^n)}$ . It follows, denoting topological entropy by  $h$  that (see [KH95] Proposition 3.1.7):

$$h(\varphi_{\#}) \geq h((\varphi_{\#})|_{f(X^n)}) = h(\varphi^n) = nh(\varphi)$$

for all  $n$ , so that  $h(\varphi_{\#}) = \infty$ .  $\square$

## 7. Proof of the main result

Let us now prove Theorem 5.3. The first power of  $X$  embeds isometrically by  $x \rightarrow \delta_x$  where  $\delta_x$  is the Dirac mass at a point. The key of the proof lies in the way of encoding a tuple by a measure, without adding any extra symmetry: one should be able to distinct  $f(a, b, \dots)$  from  $f(b, a, \dots)$ . The idea is to attribute masses to the different elements of the tuple, that are not only different but also cannot be combined to the same amount by adding some of them.

Define

$$\begin{aligned} f : \quad X^n &\rightarrow \mathcal{W}_p(X) \\ \bar{x} = (x_1, \dots, x_n) &\mapsto \alpha \sum_{i=1}^n 2^{-i} \delta_{x_i} \end{aligned}$$

where  $\alpha = 1/(1 - 2^{-n})$  is a normalizing constant.

*Step 0:*  $d(f(\bar{x}), f(\bar{y}))$  is bounded from above and intertwines dynamical systems. — There is an obvious transport plan from an image  $f(\bar{x})$  to another  $f(\bar{y})$ , given by  $\alpha \sum_i 2^{-i} \delta_{x_i} \otimes \delta_{y_i}$ . Its cost is

$$\alpha \sum_i 2^{-i} d(x_i, y_i)^p \leq \alpha/2 \sum_i d(x_i, y_i)^p$$

so that  $d(f(\bar{x}), f(\bar{y})) \leq (\alpha/2)^{1/p} d(\bar{x}, \bar{y})$ . The intertwining property is obvious from the construction of  $f$ , since  $\varphi_{\#}(\delta_x) = \delta_{\varphi(x)}$ .

The point is now to bound  $d(f(\bar{x}), f(\bar{y}))$  from below. Let  $i_0$  be an index that maximizes  $d(x_i, y_i)$ . Then  $d(\bar{x}, \bar{y}) \leq n^{1/p} d(x_{i_0}, y_{i_0})$  and we only need to bound  $d(f(\bar{x}), f(\bar{y}))$  in terms of  $d(x_{i_0}, y_{i_0})$ . For this we shall show that an optimal transport plan must send, possibly “in several steps”, some mass from  $x_{i_0}$  to  $y_{i_0}$ . Assume from now on that  $d(x_{i_0}, y_{i_0}) > 0$ , since otherwise  $\bar{x} = \bar{y}$ .

*Step 1: a transport plan defines an oriented graph.* — Let  $\Pi$  be an optimal transport plan from  $f(\bar{x})$  to  $f(\bar{y})$ . Since the source and target are discrete measures,  $\Pi$  must be discrete too:

$$\Pi = \sum_{x,y \in V} \alpha m_{xy} \delta_x \otimes \delta_y$$

where  $V$  is the set of all  $x_i$ 's and  $y_i$ 's, and  $m_{xy}$  are non-negative coefficients. One gets an oriented graph without loops by letting an edge  $x \rightarrow y$  as soon as  $m_{xy} > 0$ .

*Step 2:  $\Pi$  can be chosen so as to define a forest.* — First, the optimality of  $\Pi$  implies that the graph defined above contains no oriented cycle. If there were points  $v_1, v_2, \dots, v_k$  in  $V$  such that  $v_k = v_1$  and  $m_i := m_{v_i v_{i+1}} > 0$  for all  $i < k$ , then by soustracting the minimal value of  $m_i$  to each of them one would get a new transport plan from  $f(\bar{x})$  to  $f(\bar{y})$ , cheaper than  $\Pi$ .

Now, there can be non-oriented cycles. Let us show that given such a cycle, it is possible to modify  $\Pi$  into a new transport plan, having the same cost, so that the corresponding graph is that of  $\Pi$  with one edge of the cycle removed. Inductively, it will then possible to chose an optimal transport plan whose graph is a forest.

A non-oriented cycle is determined by two sets of vertices  $v_1, \dots, v_k$  and  $w_1, \dots, w_k$  and two sets of oriented paths  $P_i : v_i \rightarrow w_i$ ,  $Q_i : v_i \rightarrow w_{i+1}$  where  $w_{k+1} := w_1$ , see Figure 5.

A path  $P$  is described by a tuple of edges  $P = (e_1, \dots, e_l)$ , the endpoint  $e_i^+$  of  $e_i$  being equal to the starting point  $e_{i+1}^-$  of  $e_{i+1}$ . The *cost* of  $P$  is the cost of a unit mass travelling along  $P$ , that is  $c(P) = \sum_{i=1}^l d(e_i^-, e_i^+)^p$ . The *flow* of  $P$  according to  $\Pi$  is the amount of mass travelling along  $P$ , that is  $\phi(P) = \min_i m_{e_i^- e_i^+}$ . Taking a minimal non-oriented cycle, we can assume that no two paths among all  $P_i$ 's and  $Q_i$ 's share an edge.

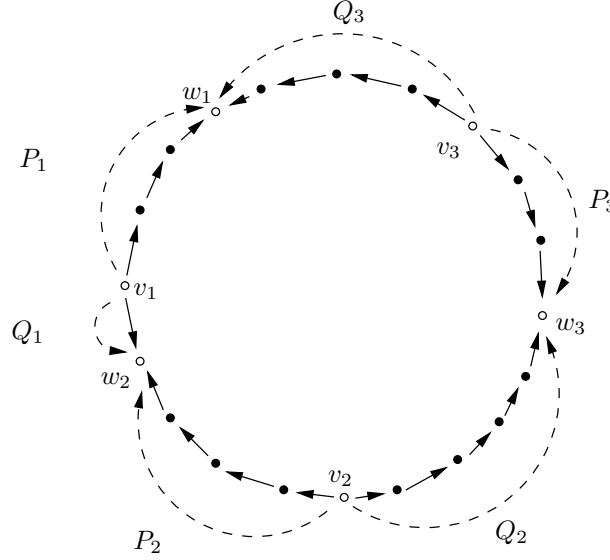


FIGURE 5. A non-oriented cycle:  $v_i$ 's and  $w_i$ 's are the vertices where the edges change orientation.

One can construct a new transport plan from  $f(\bar{x})$  to  $f(\bar{y})$  by adding a small  $\varepsilon$  to all  $m_{xy}$ 's labelling an edge of a  $P_i$ , and soustracting the same  $\varepsilon$  to all  $m_{xy}$ 's labelling an edge of a  $Q_i$ . This operation adds  $\varepsilon$  to  $\phi(P_i)$  and  $-\varepsilon$  to  $\phi(Q_i)$ , thus it adds  $\varepsilon \sum_i c(P_i) - c(Q_i)$  to the cost of  $\Pi$ .

Since  $\Pi$  is optimal, one cannot reduce its cost by this operation. This implies that  $\sum_i c(P_i) - c(Q_i) = 0$ . By operating as above with  $\varepsilon$  equal to plus or minus the minimal value of all  $m_{xy}$  labelling a  $P_i$  or a  $Q_i$ , one designs a new optimal transport plan whose graph has the chosen cycle broken.

We therefore assume from now on that the graph of  $\Pi$  has no cycle, i.e. is a forest.

*Step 3:  $\Pi$  moves mass by an amount bounded from below.*— Let us prove, by induction on the number of edges, that the  $m_{xy}$ 's are integer multiples of  $2^{-n}$ . We start with  $(V, E)$  being the graph defined by  $\Pi$ , with each vertex labelled by the couple  $(m_0(v), m_1(v))$  where

$$m_0(v) = \sum_{i|x_i=v} 2^{-i} \quad m_1(v) = \sum_{i|y_i=v} 2^{-i}$$



are, up to the constant  $\alpha$ , the initial and final mass at  $v$ , and edges labelled by the  $m_{xy}$ . During the induction, we consider modifications of this graph that are not directly related to a transport plan from  $f(\bar{x})$  to  $f(\bar{y})$  anymore, but that still satisfy the mass preservation condition  $\sum_x m_{xv} - \sum_y m_{vy} = m_1(v) - m_0(v)$ . Also, during the whole process the  $m_0(v)$  and  $m_1(v)$  shall be integer multiples of  $2^{-n}$ .

The graph  $(V, E)$  is a forest. The base case is when it has no edge, in which case all edge labels satisfy our condition by emptyness. Otherwise, the graph has a leaf, that is a vertex  $v_0$  that is incident to precisely one edge  $e_0$ . Assume for example that  $v_0 = e_0^-$  (the other case is treated similarly). Then the label of  $e_0$  is an integer multiple of  $2^{-n}$ , since it is equal to  $m_0(v_0) - m_1(v_0)$ . Now consider the graph obtained from  $(V, E)$  by deleting  $v_0$  and  $e_0$ , and soustracting the label of  $e_0$  from  $m_1(e_0^+)$ . This graphs still satisfies the mass preservation condition and the labels of vertices still are integer multiples of  $2^{-n}$ . Moreover the labels of edges are unchanged.

By inductively pruning the leafs of  $(V, E)$ , we can therefore conclude that the  $m_{xy}$ 's all are integer multiples of  $2^{-n}$ .

*Step 4: some mass has to travel from  $x_{i_0}$  to  $y_{i_0}$ .*— Now is the time to use the way we designed  $f$ . Write all  $2^n m_{xy}$  in binary, and consider the set  $E'$  of all edges such that this expansion contains the term  $2^{n-i_0}$ . The only vertex such that  $2^n m_0(v)$  contains  $2^{n-i_0}$  in its binary expansion is  $x_{i_0}$ , and the only one such that  $2^n m_1(v)$  does is  $y_{i_0}$ . It follows by the preservation of mass condition that the graph  $(V, E')$  has even degree at each vertex, except  $x_{i_0}$  and  $y_{i_0}$ . But a classical, simple result of graph theory asserts that the number of odd-degree vertices of a graph must be even. Applying this to the connected components of  $(V, E')$ , we get that  $x_{i_0}$  and  $y_{i_0}$  must be in the same component.

Thus,  $E'$  must contain a (possibly non-oriented) path  $P_0$  between  $x_{i_0}$  and  $y_{i_0}$ . Each endpoint of each edge in this path has to be some  $y_i$ , so that the path has length at most  $n$ . It follows by a convexity argument that  $c(P_0)$  (defined as for an oriented path) is at least  $n(d(x_{i_0}, y_{i_0})/n)^p$ . In particular, the cost of  $\Pi$  is at least  $\alpha 2^{-n} d(x_{i_0}, y_{i_0})^p / n^{p-1}$  and

$$d(f(\bar{x}), f(\bar{y})) \geq \frac{\alpha^{1/p} 2^{-n/p}}{n^{1-1/p}} d(x_{i_0}, y_{i_0}) \geq \frac{1}{n(2^n - 1)^{1/p}} d(\bar{x}, \bar{y})$$

which ends the proof.

### 8. Discussion of the embedding constants

One can wonder if the constants in Theorem 5.3 are optimal. The simplest possible example shows that they are off by at most a polynomial factor.

**Proposition 8.1.** — *Let  $X = \{0, 1\}$  where the two elements are at distance 1 and consider a map  $f : X^n \rightarrow \mathscr{W}_p(X)$  such that*

$$m d(\bar{x}, \bar{y}) \leq d(f(\bar{x}), f(\bar{y})) \leq M d(\bar{x}, \bar{y})$$

for all  $\bar{x}, \bar{y} \in X^n$  and some positive constants  $m, M$ . Then

$$\frac{M}{m} \geq \left( \frac{2^n - 1}{n} \right)^{1/p}.$$

Moreover there is a map whose constants satisfy  $M/m \leq (2^n - 1)^{1/p}$ .

*Proof.* — By homogeneity, it is sufficient to consider  $p = 1$ , in which case  $X^n$  is the  $n$ -dimensional discrete hypercube endowed with the combinatorial metric: two elements are at a distance equal to the number of bits by which they differ. Moreover  $\mathscr{W}_1(X)$  identifies with the segment  $[0, 1]$  endowed with the usual metric  $|\cdot|$ : a number  $t$  corresponds to the measure  $t\delta_0 + (1 - t)\delta_1$ .

The diameter of  $X^n$  is  $n$ , so that the diameter of  $f(X^n)$  is at most  $Mn$ . Since  $f(X^n)$  has  $2^n$  elements, by the pigeon-hole principle at least two of them are at distance at most  $(2^n - 1)^{-1}Mn$ . Since the distance between their inverse images is at least 1, we get  $m \leq (2^n - 1)^{-1}Mn$  so that  $M/m \geq (2^n - 1)/n$ .

To get a map  $f$  with  $M/m = (2^n - 1)$ , it suffices to use a gray code: it is an enumeration  $x_1, x_2, \dots, x_{2^n}$  of the elements of  $X^n$ , such that to consequent elements are adjacent (see [Ham80]). Letting  $f(x_i) := (i - 1)/(2^n - 1)$  we get a map with  $M \leq 1$  and  $m = (2^n - 1)^{-1}$ .  $\square$

Note that one could improve the lower bound on  $M/m$  by a factor asymptotically of the order of  $2^{1/p}$  by using the fact that every element in  $X^n$  has a opposite, that is an element at distance  $n$  from it.

Let us end with two open questions: are there some spaces for which the bounds can be significantly improved? What are the optimal bounds in the case of  $\mathbb{R}$ ?

**Acknowledgements.** — I would like to thank Frédéric Faure, Étienne Ghys, Nicola Gigli and Nicolas Juillet for interesting discussions and their comments on earlier versions of this paper.

### References

- [AGS08] L. AMBROSIO, N. GIGLI & G. SAVARÉ – *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.
- [Bal00] V. BALADI – *Positive transfer operators and decay of correlations*, Advanced Series in Nonlinear Dynamics, vol. 16, World Scientific Publishing Co. Inc., River Edge, NJ, 2000.
- [BK10] J. BERTRAND & B. KLOECKNER – “A geometric study of wasserstein spaces: Hadamard spaces”, in preparation, 2010.
- [CDPJ08] T. CHAMPION, L. DE PASCALE & P. JUUTINEN – “The  $\infty$ -Wasserstein distance: local solutions and existence of optimal transport maps”, *SIAM J. Math. Anal.* **40** (2008), no. 1, p. 1–20.
- [Gig09a] N. GIGLI – “On the inverse implication of Brenier-McCann theorems and the structure of  $(P_2(M), W_2)$ ”, available at <http://cvgmt.sns.it/people/gigli/>, 2009.
- [Gig09b] ———, “Second order analysis on  $(P_2(M), W_2)$ ”, to appear in *Memoirs of the AMS*, available at <http://cvgmt.sns.it/people/gigli/>, 2009.
- [Ham80] R. W. HAMMING – *Coding and information theory*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1980.
- [KH95] A. KATOK & B. HASSELBLATT – *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, vol. 54, Cambridge University Press, Cambridge, 1995, With a supplementary chapter by Katok and Leonardo Mendoza.
- [Klo10] B. KLOECKNER – “A geometric study of wasserstein spaces: Euclidean spaces”, *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **IX** (2010), no. 2, p. 297–323.
- [Krz77] K. KRZYŻEWSKI – “Some results on expanding mappings”, *Dynamical systems, Vol. II*—Warsaw, Soc. Math. France, Paris, 1977, p. 205–218. Astérisque, No. 50.
- [Lot08] J. LOTT – “Some geometric calculations on Wasserstein space”, *Comm. Math. Phys.* **277** (2008), no. 2, p. 423–437.

- [Ott01] F. OTTO – “The geometry of dissipative evolution equations: the porous medium equation”, *Comm. Partial Differential Equations* **26** (2001), no. 1-2, p. 101–174.
- [Sil92] S. SILVERMAN – “On maps with dense orbits and the definition of chaos”, *Rocky Mountain J. Math.* **22** (1992), no. 1, p. 353–375.
- [Sli08] S. SLIJEPEVIĆ – “Stability of invariant measures”, arXiv:0811.0109v1, 2008.
- [Vil03] C. VILLANI – *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.