

---

# OPTIMAL TRANSPORT AND DYNAMICS OF EXPANDING CIRCLE MAPS ACTING ON MEASURES

*by*

Benoît Kloeckner

---

**Abstract.** — In this article we compute the derivative of the action on probability measures of an expanding circle map at its absolutely continuous invariant measure. The derivative is defined using optimal transport: we use the rigorous framework set up by N. Gigli to endow the space of measures with a kind of differential structure.

It turns out that 1 is an eigenvalue of infinite multiplicity of this derivative, and we deduce that the absolutely continuous invariant measure can be deformed in many ways into atomless, nearly invariant measures. As a consequence, we obtain counter-examples to an infinitesimal version of Furstenberg's conjecture.

We also show that the action of standard self-covering maps on measures has positive metric mean dimension.

Some time after the publication of a first version of this article [Klo13], I found an application of the results to an infinitesimal version of Furstenberg's conjecture, as well as an error in a Lemma (which could be corrected without affecting the main results). Both are to be published together in *Ergodic Theory and Dynamical system*, and the present article is a consolidated version which combines the original article with the error corrected, and the additional material.

---

This work was supported by the Agence Nationale de la Recherche, grant ANR-11-JS01-0011.

## 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. 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.

Another goal of the paper is to exemplify the rigorous differential structure defined by N. Gigli [Gig09a], for the simplest possible compact manifold. Note that one can use 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]. But to our knowledge, no example of explicit derivative of a measure-defined map at a given point had been computed before.

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

$$\Phi_d(x) = dx \pmod{1}.$$

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  $\Phi_{d\#}$ .

A map like  $\Phi_d$  can act by composition on the right on a function space (e.g. Sobolev spaces). The adjoint of this map is usually called a Perron-Frobenius operator or a 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  $\Phi_{d\#}$  as an analogue for possibly singular measures of the Perron-Frobenius operator of  $\Phi_d$ .

As pointed out by the referee of a previous version of this paper, using the finite-to-one maps

$$(x_1, \dots, x_n) \mapsto \frac{1}{n}\delta_{x_1} + \dots + \frac{1}{n}\delta_{x_n}$$

it is easy to prove that  $\Phi_{d\#}$  is topologically transitive and has infinite topological entropy. To refine this last remark, we shall prove that  $\Phi_{d\#}$

has positive metric mean dimension (a metric dynamical invariant of infinite-entropy maps).

**Theorem 1.1.** — *For all integers  $d \geq 2$  and all exponents  $p \in [1, +\infty)$  we have*

$$\text{mdim}_M(\Phi_{d\#}, W_p) \geq p(d-1)$$

where  $W_p$  is the Wasserstein metric with cost  $|\cdot|^p$ .

The definition of Wasserstein metrics is given below; for the definition of metric mean dimension and the proof of the above result, see Section 2. Except in this result, we shall only use the quadratic Wasserstein metric ( $p = 2$ ), which we will denote by  $W$ .

Our main goal is to study the first-order dynamics of  $\Phi_{d\#}$  near the uniform measure  $\lambda$ . The precise setting will be exposed latter; let us just give a few elements. The tangent space  $T_\mu$  to  $\mathcal{P}(\mathbb{S}^1)$  at a measure  $\mu$  that is absolutely continuous with a bounded and bounded away from zero density identifies with the Hilbert space  $L_0^2(\mu)$  of all vector fields  $v : \mathbb{S}^1 \rightarrow \mathbb{R}$  that are  $L^2$  with respect to  $\mu$ , and such that  $\int v \lambda = 0$ . More generally, if  $\mu$  is atomless  $T_\mu$  identifies with a Hilbert subspace  $L_0^2(\mu)$  of  $L^2(\mu)$ .

We have a kind of exponential map:  $\exp_\mu(v) = \mu + v := (\text{Id} + v)_\# \mu$ . Then we say that a map  $f$  acting on  $\mathcal{P}(\mathbb{S}^1)$  has Gâteaux derivative  $L$  at  $\mu$  if  $f(\mu)$  has no atom and  $L : L_0^2(\mu) \rightarrow L_0^2(f(\mu))$  is a continuous linear operator such that for all  $v$  we have

$$W(f(\mu + tv), f(\mu) + tLv) = o(t)$$

We then write  $D_\mu f = L$ .

Our first differentiability result is the following.

**Theorem 1.2.** — *The map  $\Phi_{d\#}$  has a Gâteaux derivative at  $\lambda$ , equal to  $d$  times the Perron-Frobenius operator of  $\Phi_d$  acting on  $L_0^2(\lambda)$ . In particular its spectrum is the disc of radius  $d$  and all numbers of modulus  $< d$  are eigenvalues with infinite multiplicity.*

This result is detailed as Theorem 4.1 and Proposition 4.4 below. We shall also see that  $\Phi_{d\#}$  is not Fréchet differentiable.

**1.2. General expanding maps.** — The next step is to consider the action on measures of expanding circle maps. In Section 5, given a general  $C^2$  expanding map  $\Phi$ , we compute the derivative of  $\Phi_\#$  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  $\Phi$  is a  $C^2$  expanding circle map,  $\Phi_{\#}$  has a Gâteaux derivative at its unique invariant absolutely continuous measure  $\rho\lambda$ , whose adjoint operator in  $L_0^2(\rho\lambda)$  is  $u \mapsto \Phi' u \circ \Phi$ .*

In particular this derivative is a multiple of the Perron-Fronenius operator (on  $L_0^2(\rho\lambda)$ ) only when  $\Phi'$  is constant, that is when  $\Phi$  is a model map. Using general results in the spectral theory of transfer operator, it is however possible to prove that 1 is always an eigenvalue of infinite multiplicity, with continuous eigenfunctions.

**1.3. Nearly invariant measures.** — The spectral study of  $D_\lambda(\Phi_{\#})$  gives us large families of nearly invariant measures, with Lipschitz parametrization.

**Theorem 1.4.** — *For all integers  $n$ , there is a bi-Lipschitz embedding  $F : B^n \rightarrow \mathcal{P}(\mathbb{S}^1)$  mapping 0 to the absolutely continuous invariant measure  $\rho\lambda$  of  $\Phi$  such that*

$$W(\Phi_{\#}(F(a)), F(a)) = o(|a|).$$

*As a consequence, for all  $\varepsilon > 0$  and all integer  $K$  there is a radius  $r > 0$  such that for all  $k \leq K$  and all  $a \in B^n(0, r)$  the following holds:*

$$W(\Phi_{\#}^k(F(a)), F(a)) \leq \varepsilon|a|.$$

Here  $B^n$  denotes the unit Euclidean ball centered at 0 and  $W$  is the quadratic Wasserstein distance (whose definition is recalled below).

It is easy to construct invariant measures near the absolutely continuous one, for example supported on a union of periodic orbits. One can also consider convex sums  $(1 - a)\rho\lambda + a\mu$  where  $\mu$  is any invariant measure and  $a \ll 1$ . But note that the curves  $a \mapsto (1 - a)\rho\lambda + a\mu$  need not be rectifiable, let alone Lipschitz. Bernoulli measures are also examples; they are singular, atomless, fully supported invariant measures of  $\Phi_d$  that can be arbitrary close to  $\lambda$ .

The nearly invariant measures above seem of a different nature, and a natural question is how regular they are. They are given by push-forwards of the uniform measure by continuous functions; for example in

the model case a one parameter family is given by

$$\left(\text{Id} + t \sum_{\ell=0}^{\infty} d^{-\ell} \cos(2\pi d^{\ell} \cdot)\right)_{\#} \lambda$$

where  $t \in [0, \varepsilon)$ . This makes it easy to prove that almost all of them are atomless.

**Proposition 1.5.** — *If  $\mu$  is an atomless measure and  $v \in L^2(\mu)$ , for all but a countable number of values of  $t \in [0, 1]$ , the measure  $\mu + tv = (\text{Id} + tv)_{\#} \mu$  has no atom.*

*In particular, with the notation of Theorem 1.4, the map  $F$  can be chosen such that for almost all  $a$  the measure  $F(a)$  has no atom.*

That the first part of this result implies the second part shall become clear during the proof of Theorem 1.4 in Section 6, where we construct  $F$ .

This leaves open the following, antagonist questions.

**Question 1.** — *Is the measure  $F(a)$  absolutely continuous for most, or at least some  $a \neq 0$ ?*

**Question 2.** — *Is the measure  $F(a)$  invariant for most, or at least some  $a \neq 0$ ?*

The next natural questions, 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.

#### 1.4. An infinitesimal version of Furstenberg's conjecture. —

While circle expanding maps have in many respects become toy-models in the category of hyperbolic dynamical systems, a prominent question concerning them is still open for over half a century.

**Conjecture 1.6 (Furstenberg).** — *If an atomless probability measure  $\mu$  on the circle  $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$  is invariant under both*

$$\Phi_2 : x \mapsto 2x \pmod{1} \quad \text{and} \quad \Phi_3 : x \mapsto 3x \pmod{1}$$

*then  $\mu$  is equal to the Lebesgue measure  $\lambda$ .*

In the above conjecture, one can replace 2 and 3 by two multiplicatively independent integers. Even the above case is wide open in general, though a theorem of Rudolph asserts that Furstenberg’s conjecture holds for measures  $\mu$  having positive entropy for one of the maps  $\Phi_2$  or  $\Phi_3$  [Rud90] (see also [Joh92]). Many other results related to this question can be found in the literature, among which [HS12, BLMV09]; the interested reader can for example use the answers to the MathOverflow question [uea] as pointers.

To see how relevant our results can be in the context of Furstenberg conjecture, let us consider how one can approach this kind of problem in a differential geometric setting.

Furstenberg’s conjecture is a strong rigidity statement; in differential geometry, a common strategy to attack such questions is to aim to weaker rigidity statements. A first weakening would be to ask whether the point known to have a given property of interest (here:  $\lambda$ ) is, rather than unique, at least isolated among points with this property? If this stays out of reach, then can we prove that it is not possible to deform this point, i.e. to find a non-constant continuous path starting at this point inside the set defined by the given property? A further weakening is to ask for first-order rigidity, i.e. to ask whether we can use the tangent space and derivatives to prove that no  $C^1$  deformation can exist in the considered set. In the case of Furstenberg conjecture, we have a space  $\mathcal{P}(\mathbb{S}^1)$  and two rather rich subspaces, the sets of atomless invariant measures for  $\Phi_2$  and  $\Phi_3$ . Let us denote these sets of fixed measures by  $I_2$  and  $I_3$ ; then the conjecture is that  $I_2 \cap I_3 = \lambda$ . Imagine for a moment that  $I_2$  and  $I_3$  are some sort of differentiable submanifolds of  $\mathcal{P}(\mathbb{S}^1)$ ; then the various above weakenings of Furstenberg’s conjecture would take the form of the following questions:

1. Is  $\lambda$  isolated in  $I_2 \cap I_3$ ?
2. Is  $\lambda$  the sole point in its path-connected component inside  $I_2 \cap I_3$ ?
3. Must a  $C^1$  curve starting at  $\lambda$  and lying inside  $I_2 \cap I_3$  be constant?

Finally, to prove a positive answer to this third weakening, the most common approach would be to prove that the intersection  $I_2 \cap I_3$  is “first-order rigid” at  $\lambda$ , in the sense that the tangent spaces  $T_\lambda I_2$  and  $T_\lambda I_3$  intersect trivially.

Since  $I_2$  and  $I_3$  are defined (if we forget momentarily the atomless condition) as sets of fixed points for  $\Phi_{2\#}$  and  $\Phi_{3\#}$ , the first-order rigidity question would reduce to ask whether the spaces  $E_2, E_3 \subset T_\lambda \mathcal{P}(\mathbb{S}^1)$  of

invariant vectors for the derivatives  $D_\lambda\Phi_{2\#}$  and  $D_\lambda\Phi_{3\#}$  intersect trivially. Even if all the above speculation turns out to be wrong (e.g.  $I_2$  and  $I_3$  could not be anything close to submanifolds), this last question is perfectly defined in the differential setting alluded to above, and can be considered an infinitesimal version of Furstenberg’s conjecture. As a consequence of the previous results, we will prove that this question has a negative answer.

**Theorem 1.7.** — *The vector space  $E_2 \cap E_3 \subset T_\lambda\mathcal{P}(\mathbb{S}^1)$  of tangent vectors at  $\lambda$  that are simultaneously invariant under both  $D_\lambda\Phi_{2\#}$  and  $D_\lambda\Phi_{3\#}$  is infinite-dimensional.*

*The vector space  $\bigcap_{d=2}^\infty E_d$  of tangent vectors at  $\lambda$  that are simultaneously invariant under all the  $D_\lambda\Phi_{d\#}$  is 2-dimensional.*

Formulated as it is in terms of the Wasserstein metric, this result could feel very abstract and potentially artificial, so let us give a direct corollary that contains no reference to optimal transport or abstract differential geometric setting. The idea behind this corollary goes back to an insight of Otto [Ott01] related to the point of view of Benamou and Brenier [BB00] and developed in [AGS08]: by integration, smooth test functions can serve as a kind of (weak) coordinates on  $\mathcal{P}(\mathbb{S}^1)$ ; for simplicity this corollary is phrased in a restricted way, only using that  $\bigcap_{d=2}^\infty E_d$  is not reduced to 0.

**Corollary 1.8.** — *There exists a path of probability measures  $(\mu_t)_{t \in (-\varepsilon, \varepsilon)}$  with  $\mu_0 = \lambda$ , continuous in the weak topology, with  $\mu_t$  atomless for almost all  $t$ , such that:*

$$\frac{d}{dt} \int_{\mathbb{S}^1} \psi_0 d\mu_t \Big|_{t=0} \neq 0$$

for some smooth function  $\psi_0 : \mathbb{S}^1 \rightarrow \mathbb{R}$ , and

$$\frac{d}{dt} \int_{\mathbb{S}^1} \psi d\mu_t \Big|_{t=0} = \frac{d}{dt} \int_{\mathbb{S}^1} \psi d(\Phi_{d\#}\mu_t) \Big|_{t=0}$$

for all smooth functions  $\psi : \mathbb{S}^1 \rightarrow \mathbb{R}$  and all integer  $d \geq 2$ .

**Remark 1.9.** — 1. The first condition ensures that  $\mu_t$  depends significantly on  $t$  (in particular, it avoids the degenerate and obvious choice  $\mu_t \equiv \lambda$ ), while the second condition expresses that for small  $t$ ,  $\mu_t$  is “almost invariant” under all the push-forward maps  $\Phi_{d\#}$ . Of

course, this condition can be rewritten

$$\frac{d}{dt} \int_{\mathbb{S}^1} \psi d\mu_t \Big|_{t=0} = \frac{d}{dt} \int_{\mathbb{S}^1} \psi \circ \Phi_d d\mu_t \Big|_{t=0}.$$

2. This corollary is intrinsically much weaker than the theorem, as differentiability in the sense of Wasserstein distance implies differentiability of the integrals of test functions, but the converse implication does not hold. For example, a curve of the form  $(t\mu + (1-t)\nu)_t$  is usually not differentiable (or even rectifiable) in the differential structure induced by  $W_2$ , while the integral of any test function depends affinely on  $t$ . Nevertheless, I do not know a simpler way to get Corollary 1.8 even when restricting  $d$  to  $\{2, 3\}$ . Even if the Furstenberg conjecture were false and there were an atomless probability measure  $\mu \neq \lambda$  invariant by  $\Phi_2$  and  $\Phi_3$ , the curve  $(t\lambda + (1-t)\mu)_t$  would not work as these measures are not positive for negative  $t$ .
3. One could try to extend this infinitesimal argument to the construction of families of counter-examples to the Furstenberg conjecture: if one of the invariant vectors we found could be extended to a vector field preserved by both  $\Phi_2$  and  $\Phi_3$ , then the integral curve issued from  $\lambda$  would be entirely made of invariant measures for both  $\Phi_2$  and  $\Phi_3$ . However, it would be incredibly bold to conjecture this extension to be possible: we do not even know whether  $\Phi_{2\#}$  is differentiable at any non-absolutely continuous measure. Note also that this extension cannot be expected at all for the full semi-group  $\mathbb{N}$ , as it is known that the Lebesgue measure is the only atomless measure invariant under all  $\Phi_d$  (this holds more generally for large enough sub-semigroups of  $\mathbb{N}$ , see [EF08]). One can still dream of making this approach work for finitely generated multiplicative sub-semigroups, as this case is very different from larger sub-semigroups: in the former case, the remainder in the first-order Taylor formula for the  $\Phi_{d\#}$  at  $\lambda$  can be made uniform over the generators (for a fixed simultaneously invariant tangent vector).

**1.5. 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.

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

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:

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

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

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

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 exists. 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.

The function  $W_p$  is a metric, called the  $(L^p)$  Wasserstein metric, and when  $X$  is compact it induces the weak topology. We sometimes denote  $W_2$  simply by  $W$ .

## 2. Metric mean dimension

Metric mean dimension is a metric invariant of dynamical systems introduced by Lindenstrauss and Weiss [LW00], that refines topological entropy for infinite-entropy systems.

Let us briefly recall the definitions. Given a map  $f : X \rightarrow X$  acting on a compact 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

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

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 dependance on initial condition of the considered dynamical system. The map  $\Phi_d$  is a classical example, whose topological entropy is  $\log d$ .

Now, the metric mean dimension is

$$\text{mdim}_M(f, d) := \liminf_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{\log N(f, \varepsilon, n)}{n |\log \varepsilon|}.$$

It is zero as soon as topological entropy is finite. Note that this quantity does depend upon the metric; here we shall use  $W_p$ . Lindenstrauss and Weiss define the metric mean dimension using covering sets rather than separated sets, but this does not matter since their sizes are comparable.

Let us prove Theorem 1.1: the metric mean dimension of  $\Phi_{d\#}$  is at least  $p(d-1)$  when  $\mathcal{P}(\mathbb{S}^1)$  is endowed with the  $W_p$  metric. In another paper [Klo10a], we prove the same kind of result, replacing  $\Phi_d$  by any map having positive entropy. However Theorem 1.1 has a better constant and its proof is simpler.

*Proof of Theorem 1.1.* — To construct a large  $(n, \varepsilon)$ -separated set, we proceed as follows: we start with the point  $\delta_0$ , and choose an  $\varepsilon$ -separated set of its antecedents. Then we inductively choose  $\varepsilon$ -separated sets of antecedents of each elements of the set previously defined. Doing this, we need not control the distance between antecedents of two different elements.

Let  $k \gg 1$  and  $\alpha > 0$  be integers;  $\varepsilon$  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  $\mu([0, 1/d]) \geq 1/2$ . 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  $[0, 1 - d2^{-k}]$  and  $\mu_t$  is supported on  $(1 - d2^{-k}, 1 - 2^{-k})$ . Let  $e_1, \dots, e_d$  be the right inverses to  $\Phi$  defined onto  $[0, 1/d), [1/d, 2/d), \dots, [(d-1)/d, 1)$  respectively. For all integer tuples  $\ell = (\ell_1, \dots, \ell_d)$  such that  $\ell_1 \geq 2^{\alpha k - 1}$

and  $\sum \ell_i = 2^{\alpha k}$ , define

$$\mu_\ell = e_{1\#}(\ell_1 2^{-\alpha k} \mu_h + \mu_t) + \sum_{i>1} e_{i\#}(\ell_i 2^{-\alpha k} \mu_h)$$

(see figure 1 that illustrates the case  $d = 2$ ). It is a probability measure on  $\mathbb{S}^1$ , lies in  $A_k$  and  $\Phi_{d\#}(\mu_\ell) = \mu$ . Moreover, if  $\ell' \neq \ell$  then the masses given by  $\mu_{\ell'}$  and  $\mu_\ell$  to some interval  $[(i-1)/d, i/d)$  differ by at least  $2^{-\alpha k - 1}$ , and the  $2^{-k}d^{-1}$ -long rightmost part of each such interval is given no mass by  $\mu_\ell$  and  $\mu_{\ell'}$ . It follows that any transport plan  $\Pi$  from  $\mu_\ell$  to  $\mu_{\ell'}$  has to give a mass at least  $2^{-\alpha k - 1}$  to the set of pairs  $(x, y) \in \mathbb{S}^1$  such that  $|x - y| \geq 2^{-k}d^{-1}$  (one sometimes says that  $\Pi$  moves a mass at least  $2^{-\alpha k - 1}$  by a distance at least  $2^{-k}d^{-1}$ ). Therefore,

$$W_p(\mu_\ell, \mu_{\ell'}) \geq d^{-1} 2^{-k(\alpha/p+1)-1/p}.$$

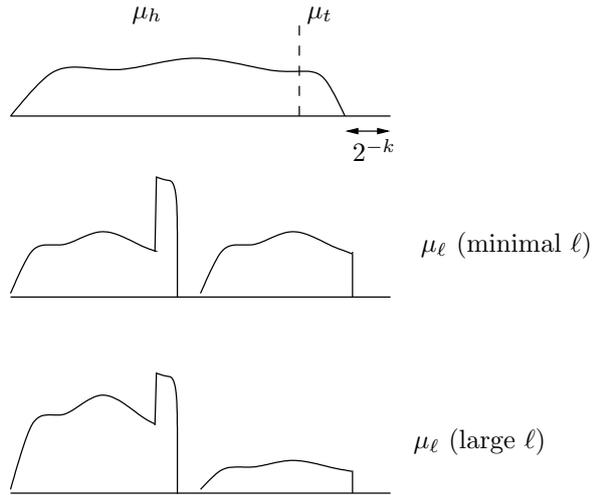


FIGURE 1. Construction of separated antecedents of a given measure.

Let  $\varepsilon = d^{-1} 2^{-k(\alpha/p+1)-1/p}$  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,  $S_{n+1}$  has at least  $C 2^{\alpha k(d-1)}$  times as many elements as  $S_n$ , for some constant  $C$  depending only on  $d$ . Then  $S_n$  has at least  $C^n 2^{n\alpha k(d-1)}$  elements. Let  $\mu, \nu$  be two distinct elements of  $S_n$  and  $m$  be the greatest index such that  $\Phi_{d\#}^m \mu \neq \Phi_{d\#}^m \nu$ . Since  $\Phi_{d\#}^n \mu = \delta_0 = \Phi_{d\#}^n \nu$ ,  $m$

exists and is at most  $n - 1$ . The measures  $\mu' = \Phi_{d\#}^m \mu$  and  $\nu' = \Phi_{d\#}^m \nu$  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

$$\begin{aligned} \frac{\log N(\Phi_{d\#}, \varepsilon, n)}{n |\log \varepsilon|} &\geq \frac{C}{|\log \varepsilon|} + \frac{\alpha(d-1)}{\frac{\alpha}{p} + 1} \left( \frac{-\frac{1}{p} - \frac{\log d}{\log 2}}{|\log \varepsilon|} + 1 \right) \\ &\geq \frac{\alpha(d-1)}{\frac{\alpha}{p} + 1} (1 + o(1)) + o(1). \end{aligned}$$

In the case of a general  $\varepsilon$ , we get the same bound on  $\log N$  up to an additive term  $n\alpha(d-1)\log 2$ , so that

$$\text{mdim}_M(\Phi_{d\#}, W_p) \geq \frac{\alpha(d-1)}{\frac{\alpha}{p} + 1}.$$

By taking  $\alpha \rightarrow \infty$  we get  $\text{mdim}_M(\Phi_{d\#}, W_p) \geq p(d-1)$ .  $\square$

### 3. The first-order differential structure on measures

In this section we give a short account on the work of Gigli [**Gig09a**] in the particular case of the circle. Note that considering the Wasserstein space of a Riemannian manifold as an infinite-dimensionnal Riemannian manifold dates back to the work of Otto [**Ott01**]. However, in many ways it stayed at the formal level until the work of Gigli.

**3.1. Why bother with this setting?**— Before getting started, let us explain why we do not simply use the natural affine structure on  $\mathcal{P}(\mathbb{S}^1)$ , the tangent space at a point simply consisting on signed measures having zero total mass. Similarly, one could consider simpler to just take the smooth functions of  $\mathbb{S}^1$  as coordinates to define a smooth structure on  $\mathcal{P}(\mathbb{S}^1)$ .

The first argument against these points of view is that optimal transportation is about pushing mass, not (directly) about recording the variation of density at each point.

More important, these simple ideas would lead to a path of the form  $\gamma_t = t\delta_x + (1-t)\delta_y$  being smooth. However, the Wasserstein distance between  $\gamma_t$  and  $\gamma_s$  has the order of  $\sqrt{|t-s|}$ , so that  $\gamma_t$  is not rectifiable (it has infinite length)! This also holds, for example, for convex sums of measures with different supports.

One could argue that the previous paths can be made Lipschitz by using  $W_1$  instead of  $W_2$ , so let us give another argument: in the affine structure, the Lebesgue measure does not have a tangent space but only a tangent cone since  $\lambda + t\mu$  is not a positive measure for all small  $t$  unless  $\mu \ll \lambda$ . If one wants to consider singular measures in the same setting than regular ones, the  $W_2$  setting seems to be the right tool.

Note that it will appear that the differential structure on  $\mathcal{P}(\mathbb{S}^1)$  depends not only on the differential structure of the circle, but also on its metric. This should not be considered surprising: in finite dimension, the fact that the differential structures are defined independently of any reference to a metric comes from the equivalence of norms in Euclidean space: here, in infinite dimension, even the simple formula  $W(f(\mu + tv), f(\mu) + tD_x f(v)) = o(t)$  involves a metric in a crucial way.

One could also be surprised that this differential structure involving the metric of the circle could be preserved by expanding maps of non-constant derivative. This point shall be cleared in Section 5, see Proposition 5.2 and the discussion before it.

**3.2. The exponential map.** — 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  $\mathcal{P}(T\mathbb{S}^1)_\mu$  be the set of probability measures on the tangent bundle  $T\mathbb{S}^1$  that are mapped to  $\mu$  by the canonical projection.

Given  $\xi, \zeta \in \mathcal{P}(T\mathbb{S}^1)_\mu$ , one defines

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

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  $(\text{Id}, \text{Id})_{\#}\mu$  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 sets of measure zero. Then one gets

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

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

There is a natural cone structure on  $\mathcal{P}(T\mathbb{S}^1)_\mu$ , extending the scalar multiplication on the tangent bundle: letting  $D_r$  be the dilation of ratio  $r$  along fibers, acting 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

$$\exp_\# : \mathcal{P}(T\mathbb{S}^1)_\mu \rightarrow \mathcal{P}(\mathbb{S}^1).$$

The point is that not for all  $\xi \in \mathcal{P}(T\mathbb{S}^1)_\mu$ , 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  $\mathcal{P}(T\mathbb{S}^1)_\mu$  of the subset of all  $\xi$  such that  $\exp_\#(t \cdot \xi)$  defines a geodesic for small enough  $t$ .

**3.3. Another definition of the tangent space.** — Let us now give another definition, assuming  $\mu$  is atomless. We denote by  $|\cdot|_{L^2(\mu)}$  the norm defined by the measure  $\mu$ , and by  $|\cdot|_2$  the usual  $L^2$  norm defined by the Lebesgue measure  $\lambda$ .

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  $\mathcal{P}(T\mathbb{S}^1)_\mu$ . This element has the property that  $\exp_\#(t \cdot \xi) = (\text{Id} + t \xi_f)_\# \mu$  defines a geodesic for small enough  $t$ , with a time bound depending on  $\nabla f$  and not on  $\mu$ . More precisely, the geodesicness holds as soon as no mass is moved at a distance exceeding  $1/2$ , and no element of mass crosses another one, and these conditions translate to  $t(\nabla f)'(x) \geq -1$  for all  $x$ . This is a particular case of Kantorovich duality, see for example [Vil09], especially figure 5.2.

Now, let  $L_0^2(\mu)$  be the set of all vector fields  $v \in L^2(\mu)$  that are  $L^2(\mu)$ -approximable by gradient of smooth functions. Then the image of the map  $v \mapsto (\text{Id}, v)_\# \mu$  defined on  $L_0^2(\mu)$  with value in  $\mathcal{P}(T\mathbb{S}^1)_\mu$  is precisely  $T_\mu$ . 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,  $v$  is  $L^2(\mu)$ -approximable by gradient of smooth functions; note that among smooth vector fields, gradients are

characterized by  $\int \nabla f \lambda = 0$ . We shall freely identify the tangent space with  $L_0^2(\mu)$  whenever  $\mu$  has no atom.

In the important case when  $\mu = \rho \lambda$  for some positive continuous density  $\rho$ , a vector field  $v \in L^2(\mu)$  is approximable by gradient of smooth functions if and only if  $\int v \lambda = 0$ . We get that in this case,  $T_\mu$  can be identified with the set of functions  $v : \mathbb{S}^1 \rightarrow \mathbb{R}$  that are square-integrable with respect to  $\mu$  and of mean zero with respect to  $\lambda$ . When  $\mu$  is the uniform measure, we write  $L_0^2$  instead of  $L_0^2(\lambda)$ . Note that if  $v \in L^2(\mu)$  has neither its negative part nor its positive part  $\lambda$ -integrable, then it can be approximated in  $L^2(\mu)$  norm by gradient of smooth functions, and that if  $\mu$  has not full support, then  $L_0^2(\mu) = L^2(\mu)$ .

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

This point of view is convenient, in particular because the distance between exponential curves issued from  $\mu$  can be estimated easily:

$$W(\mu + tv, \mu + tw) \underset{t \rightarrow 0}{\sim} t|v - w|_{L^2(\mu)}.$$

Note that when  $v$  is differentiable, then by geodesicness for  $t$  small enough we have

$$W(\mu, \mu + tv) = t|v|_{L^2(\mu)}$$

and not only an equivalence. This will prove useful in the next subsection where several measures and vector fields will be involved.

**3.4. Two properties.** — We shall prove that the exponential map can be used to construct bi-Lipschitz embeddings of small, finite-dimensional balls into  $\mathcal{P}(\mathbb{S}^1)$ , then we shall study how the density of an absolutely continuous measure evolves when pushed by a small vector field.

The following natural result shall be used in the proof of Theorem 1.4.

**Proposition 3.1.** — *Given  $\mu \in \mathcal{P}(\mathbb{S}^1)$  and  $(v_1, \dots, v_n)$  continuous, linearly independent vector fields in  $L_0^2(\mu)$ , there is an  $\eta > 0$  such that the map  $B^n(0, \eta) \rightarrow \mathcal{P}(\mathbb{S}^1)$  defined by  $E(a) = \mu + \sum a_i v_i$  is bi-Lipschitz.*

The difficulty is only technical: we already know that  $E$  is bi-Lipschitz along rays and we need some uniformity in the distance estimates to prove the global bi-Lipschitzness. The continuity hypothesis is not satisfactory but is all we need in the sequel.

Note that we did not assume that  $\mu$  has no atom; when it has,  $L_0^2(\mu)$  (still defined as the closure in  $L^2(\mu)$  of gradients of smooth functions) is not the tangent cone  $T_\mu \mathcal{P}(\mathbb{S}^1)$  but only a part of it. Note that if  $v$  is a

$C^1$  vector field of vanishing  $\lambda$ -mean,  $(\mu + tv)_t$  still defines a geodesic as long as  $tv' \geq -1$ .

*Proof.* — Let  $a, b \in B^n$ . The plan  $(\text{Id} + \sum a_i v_i, \text{Id} + \sum b_i v_i)_\# \lambda$  transports  $E(a)$  to  $E(b)$  at a cost

$$\left| \sum (a_i - b_i) v_i \right|_2^2 \leq \left( \sum |v_i|_2^2 \right) |a - b|^2$$

so that  $E$  is Lipschitz.

Up to a linear change of coordinates, we assume that the  $v_i$  form an orthonormal family of  $L_0^2(\mu)$ . To bound the distance between  $E(a)$  and  $E(b)$  from below, we shall design a vector field  $\tilde{v}$  such that pushing  $E(a)$  by  $\tilde{v}$  gives a measure close to  $E(b)$ .

Choose  $\varepsilon > 0$  such that for all  $i$  we have

$$|x - y| \leq \varepsilon \Rightarrow |v_i(x) - v_i(y)| \leq \frac{1}{4\sqrt{n}}.$$

Assume moreover  $\varepsilon < 1/8$ .

Let  $w_i$  be gradient of smooth functions such that  $|v_i - w_i|_\infty \leq \varepsilon$ . Let  $\eta > 0$  be small enough to ensure  $2\sqrt{n}\eta \leq 1$  and  $w'_i \geq -(4n\eta)^{-1}$  for all  $i$ .

Fix  $a, b \in B^n(0, \eta)$  and introduce two maps defined by  $\psi(y) = y + \sum a_i v_i(y)$  and  $\tilde{\psi}(y) = y + \sum a_i w_i(y)$ . Note that  $\tilde{\psi}' \geq 1/2$  so that  $\tilde{\psi}$  is a diffeomorphism and  $\tilde{\psi}^{-1}$  is 2-Lipschitz. Let  $\tilde{v} = \sum (b_i - a_i) v_i \circ \tilde{\psi}^{-1}$ .

On the first hand, given any  $y \in \mathbb{S}^1$ , we have

$$|\tilde{\psi}(y) - \psi(y)| \leq |a| \left( \sum (w_i(y) - v_i(y))^2 \right)^{1/2} \leq |a| \sqrt{n} \varepsilon$$

so that

$$|y - \tilde{\psi}^{-1} \psi(y)| \leq 2\sqrt{n} |a| \varepsilon \leq \varepsilon$$

and

$$\left| v_i(\tilde{\psi}^{-1} \psi(y)) - v_i(y) \right| \leq \frac{1}{4\sqrt{n}}.$$

It follows that

$$\left| \sum (b_i - a_i) (v_i(\tilde{\psi}^{-1} \psi(y)) - v_i(y)) \right| \leq \frac{1}{4} |b - a|,$$

and therefore

$$(1) \quad \left| \tilde{v} \circ \psi - \sum (b_i - a_i) v_i \right|_{L^2(\nu)} \leq \frac{1}{4} |b - a|$$

where  $\nu$  could be any probability measure. We shall take  $\nu = \mu + \sum a_i v_i$ .

Similarly,

$$\begin{aligned}
|\tilde{v}|_{L^2(\nu)} &= \left( \int \tilde{v}^2(x) (\psi_{\#}\mu)(dx) \right)^{1/2} \\
&= \left( \int \tilde{v}^2(\psi x) \mu(dx) \right)^{1/2} \\
&= \left| \sum (b_i - a_i) v_i \tilde{\psi}^{-1} \psi \right|_{L^2(\mu)} \\
&\geq \frac{3}{4} \left| \sum (b_i - a_i) v_i \right|_{L^2(\mu)} \\
(2) \quad |\tilde{v}|_{L^2(\nu)} &\geq \frac{3}{4} |b - a|.
\end{aligned}$$

On the other hand, we have

$$W\left(\mu + \sum a_i v_i, \mu + \sum b_i v_i\right) \geq W(\nu, \nu + \tilde{v}) - W\left(\nu + \tilde{v}, \mu + \sum b_i v_i\right).$$

Let  $\tilde{w} = \sum (b_i - a_i) v_i \circ \tilde{\psi}^{-1}$ . We have  $|\tilde{v} - \tilde{w}|_{\infty} \leq \varepsilon |b - a|$ . In particular,  $|\tilde{w}|_{L^2(\nu)} \geq \frac{5}{8} |b - a|$ . The choice of  $\eta$  ensures that  $\tilde{w}' \geq -1$ , so that

$$W(\nu, \nu + \tilde{w}) = |\tilde{w}|_{L^2(\nu)} \geq \frac{5}{8} |b - a|.$$

Since  $W(\nu + \tilde{v}, \nu + \tilde{w}) \leq |\tilde{v} - \tilde{w}|_{\infty}$  we get

$$(3) \quad W(\nu, \nu + \tilde{v}) \geq \frac{1}{2} |b - a|.$$

Finally, since  $\nu + \tilde{v} = (\psi + \tilde{v}\psi)_{\#}\mu$ , (1) shows that

$$W\left(\nu + \tilde{v}, \mu + \sum b_i v_i\right) \leq \frac{1}{4} |b - a|$$

so that

$$W\left(\mu + \sum a_i v_i, \mu + \sum b_i v_i\right) \geq \frac{1}{4} |b - a|.$$

□

**Proposition 3.2.** — *Let  $\rho$  be a  $C^1$  density and  $v : \mathbb{S}^1 \rightarrow \mathbb{R}$  be a  $C^1$  vector field. Then for  $t \in \mathbb{R}$  small enough  $\rho\lambda + tv$  is absolutely continuous and its density  $\rho_t$  is continuous and satisfy*

$$\rho_t(x) = \rho(x) - t(\rho v)'(x) + o(t)$$

where the remainder term is independent of  $x$ .

*Proof.* — Let  $t$  be small enough so that  $\text{Id} + tv$  is a diffeomorphism. By a change of variable, we see that

$$\begin{aligned}\rho_t &= \frac{\rho}{1 + tv'} \circ (\text{Id} + tv)^{-1} \\ &= (\rho(1 - tv')) \circ (\text{Id} - tv) + o(t) \\ &= \rho - t(\rho'v + v'\rho) + o(t)\end{aligned}$$

where the  $o(t)$  term depends upon  $\rho$  and  $v$  but is uniform in  $x$ .  $\square$

Note that the  $o(t)$  depends in particular on the moduli of continuity of  $v'$  and  $\rho'$  and need not be an  $O(t^2)$  unless  $v$  and  $\rho$  are  $C^2$ .

#### 4. First-order dynamics in the model case

In this section we show that  $\Phi_{d\#}$  is (weakly) differentiable at the point  $\lambda$ . Its derivative is an explicit, simple endomorphism of a Hilbert space, and we shall give a brief study of its spectrum.

**Theorem 4.1.** — *Let  $\mathcal{L}_d : L_0^2 \rightarrow L_0^2$  be the linear operator defined by*

$$\mathcal{L}_d v(x) = v(x/d) + v((x+1)/d) + \cdots + v((x+d-1)/d).$$

*Then  $\mathcal{L}_d$  is the derivative of  $\Phi_{d\#}$  at  $\lambda$  in the following sense: for all  $v \in L_0^2 \simeq T_\lambda$ , one has*

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

First, we recognize in  $\mathcal{L}_d$  a multiple of the Perron-Frobenius operator of  $\Phi_d$ , 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

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

However, we shall see that such a uniform bound does not hold. However, one easily gets uniform remainder terms in restriction to any finite-dimensional subspace of  $L_0^2$ .

**4.1. Differentiability of  $\Phi_{d\#}$ .** — The main point to prove in the above theorem is the following estimate; this is where the original article contained a mistake.<sup>(1)</sup>

**Lemma 4.2.** — *Given a Hölder continuous and positive density  $\rho$ , vector fields  $v_1, \dots, v_n \in L^2(\rho\lambda)$  and positive numbers  $\alpha_1, \dots, \alpha_n$  summing up to 1, one has*

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

The positivity assumption may not be necessary, but at the very least simplifies the proof.

*Proof.* — We prove the case  $n = 2$  since the general case can then be deduced by induction.

Let  $\varepsilon$  be any positive number, and consider vector fields  $\bar{v}_i$  ( $i = 1, 2$ ) that are constant on the intervals  $[j/k_1, (j+1)/k_1)$  for some  $k_1$  and all  $j < k_1$  and such that  $\|\bar{v}_i - v_i\|_{L^2(\rho\lambda)} \leq \varepsilon$ . Note that  $k_1$  and the  $\bar{v}_i$  are chosen to depend only on  $\varepsilon$ , not on  $t$ ; in particular  $\|\bar{v}_1 - \bar{v}_2\|_\infty$  is finite and independent of  $t$ .

Now consider any value of  $t$ , to be taken small enough a few times below. Let  $k = k(t)$  be a multiple of  $k_1$  having the magnitude of  $(1/t)^{1/(1+\beta/2)}$  where  $\beta$  is the Hölder exponent of  $\rho$ , say  $kt^{1/(1+\beta/2)} \in [1, 2]$ .

We define  $\bar{\rho}$  as the density that is constant on each  $I_j = [\frac{j}{k}, \frac{j+1}{k})$ , of value  $\bar{\rho}_j := k \int_{I_j} \rho \, d\lambda$ . Denoting by  $C$  the Hölder constant of  $\rho$ , we get

$$\|\rho - \bar{\rho}\|_\infty \leq Ck^{-\beta}$$

We denote by  $\bar{v}_i(j)$  the value of  $\bar{v}_i$  on  $I_j$ ; observe that when  $t$  is small, these values are the same on many successive intervals since  $k$  is much larger than  $k_1$ .

Let us first bound above  $W(\rho\lambda, \bar{\rho}\lambda)$ . We consider the monotone rearrangement fixing 0 as transport plan; by definition of  $\bar{\rho}$ , it preserves each

<sup>(1)</sup>More precisely, in [Klo13] the right-hand side of the first inline equation in Lemma 4.2 should be  $\varepsilon t + 2^{-3/2}\varepsilon$  rather than  $(1 + 2^{-3/2})\varepsilon t$ . This mistake can be corrected by estimating how well a piecewise constant density with  $k$  pieces of equal length can approximate the given density. Then the issue is moved to the main argument: in order to ultimately get a  $o(t)$  remainder, we need to take advantage of the presence of many overlaps (as in Figure 2), which only exist if  $k$  increases not too fast with respect to  $t$ . This can be ensured by adding the regularity hypothesis. We shall only use Lemma 4.2 for positive  $C^1$  densities, so this hypothesis is harmless.

$I_i$ . To simplify notation, let us bound the cost due to the mass located in  $I_0$ , the other intervals behaving in exactly the same way. The cumulative distribution functions of  $\rho\lambda$  and  $\bar{\rho}\lambda$  are given by

$$F(x) = \int_0^x \rho \, d\lambda \quad \text{and} \quad G(x) = x\bar{\rho}_0.$$

The monotone rearrangement is given on  $I_0$  by the map  $T = G^{-1} \circ F$ , so that the contribution of  $I_0$  to its cost is

$$\begin{aligned} \int_0^{\frac{1}{k}} |T(x) - x|^2 \rho(x) \, dx &= \int_0^{\frac{1}{k}} \left| \frac{1}{\bar{\rho}_0} \int_0^x \rho \, d\lambda - \frac{1}{\bar{\rho}_0} \int_0^x \bar{\rho}_0 \, d\lambda \right|^2 \rho(x) \, dx \\ &\leq \int_0^{\frac{1}{k}} \left| \frac{1}{\bar{\rho}_0} \int_0^x |\rho - \bar{\rho}_0| \, d\lambda \right|^2 \rho(x) \, dx \\ &\leq \int_0^{\frac{1}{k}} \frac{C^2 x^2}{\bar{\rho}_0^2 k^{2\beta}} \rho(x) \, dx \\ &\leq \frac{C^2}{\bar{\rho}_0 k^{3+2\beta}} \end{aligned}$$

Since the mass lying in  $I_0$  is  $\bar{\rho}_0/k$  (for both densities), the ratio cost per mass is bounded above by

$$\frac{C^2}{\bar{\rho}_0^2 k^{2+2\beta}} \leq \frac{C^2}{(\min \rho)^2 k^{2+2\beta}}.$$

Since this holds in all intervals  $I_i$ , the overall cost is bounded by the same value, so that

$$W(\rho\lambda, \bar{\rho}\lambda) \leq \frac{C}{\min \rho} \frac{1}{k^{1+\beta}}.$$

The same argument also yields

$$W(\rho\lambda + v, \bar{\rho}\lambda + v) \leq \frac{C}{\min \rho} \frac{1}{k^{1+\beta}}$$

for any vector field  $v$  which is constant on each  $I_j$ : indeed, if  $\Pi$  is a transport plan from  $\rho\lambda$  to  $\bar{\rho}\lambda$ , then  $(\text{Id} + v, \text{Id} + v)_\# \Pi$  is a transport plan from  $\rho\lambda + v$  to  $\bar{\rho}\lambda + v$  whose cost is not greater than the cost of  $\Pi$  (for each bit of mass moved from  $x$  to  $T(x)$  by  $\Pi$ , this new plan moves the same amount of mass from  $x + v(x)$  to  $T(x) + v(T(x))$ ); the hypothesis that  $v$  is constant on each  $I_j$  then ensures that  $v(T(x)) = v(x)$ ). Applying this

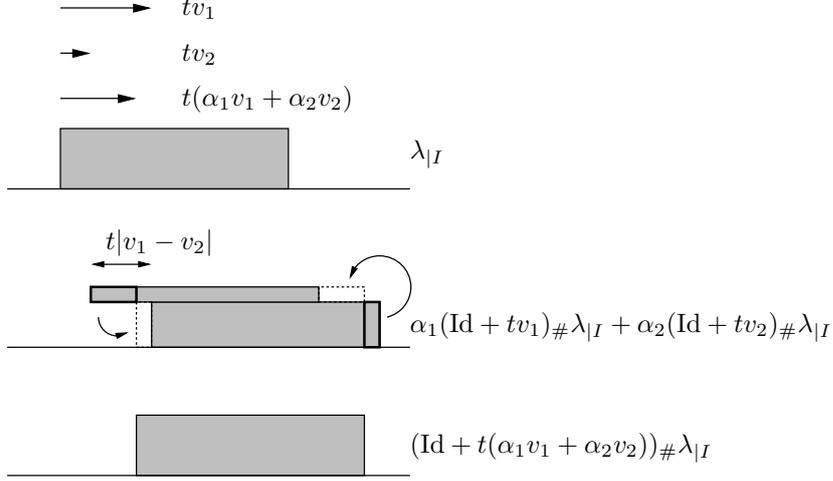


FIGURE 2. The cost of this transport plan has the order of magnitude  $t^3$

to  $v = t\sum\alpha_i\bar{v}_i$  we get

$$W(\rho\lambda + t\sum\alpha_i\bar{v}_i, \bar{\rho}\lambda + t\sum\alpha_i\bar{v}_i) \leq \frac{C}{\min\rho} \frac{1}{k^{1+\beta}}.$$

Applying the same reasoning to each  $v = t\bar{v}_i$  separately and concatenating the corresponding transport plan also yields

$$W(\sum\alpha_i(\rho\lambda + t\bar{v}_i), \sum\alpha_i(\bar{\rho}\lambda + t\bar{v}_i)) \leq \frac{C}{\min\rho} \frac{1}{k^{1+\beta}}.$$

We will now prove the bound

$$W(\bar{\rho}\lambda + t\sum\alpha_i\bar{v}_i, \sum\alpha_i(\bar{\rho}\lambda + t\bar{v}_i)) \leq t^{3/2}k^{1/2}\|\bar{v}_1 - \bar{v}_2\|_\infty^{3/2}.$$

For this, on each  $I_j$  the construction pictured in Figure 2 gives a transport plan from  $(\text{Id} + t(\alpha_1 v_1(j) + \alpha_2 v_2(j)))_\# \bar{\rho}_j \lambda|_{I_j}$  to  $\alpha_1(\text{Id} + tv_1(j))_\# \bar{\rho}_j \lambda|_{I_j} + \alpha_2(\text{Id} + tv_2(j))_\# \bar{\rho}_j \lambda|_{I_j}$  whose cost is at most  $t^3 \bar{\rho}_j |\bar{v}_1(j) - \bar{v}_2(j)|^3$  for a contribution to the mass of  $\bar{\rho}_j/k$ . More precisely, temporarily denoting by  $\rho$ ,  $v_1$  and  $v_2$  the values taken by the functions  $\bar{\rho}$  and  $\bar{v}_i$  on  $I_j$ , one simply let the common mass in place and moves at each side a mass  $\alpha_1 \alpha_2 \rho |v_1 - v_2| t$  by a distance at most  $|v_1 - v_2| t$ ; this is not optimal but sufficient for our purpose.

The last estimates we need are

$$W(\rho\lambda + t\sum\alpha_i v_i, \rho\lambda + t\sum\alpha_i \bar{v}_i) \leq \varepsilon t$$

and

$$W(\sum \alpha_i(\rho\lambda + t\bar{v}_i), \sum \alpha_i(\rho\lambda + tv_i)) \leq \varepsilon t.$$

They are both obtained, as in the original proof, by observing that for any measure  $\mu$  and any vector fields  $v, \bar{v}$  in  $L^2(\mu)$ , the transport plan  $(\text{Id} + v, \text{Id} + \bar{v})_{\#}\mu$  has cost exactly  $\|v - \bar{v}\|_{L^2(\mu)}$ .

Using the triangle inequality to combine all these estimates, we get

$$\begin{aligned} W\left(\rho\lambda + t\sum \alpha_i v_i, \sum \alpha_i(\rho\lambda + tv_i)\right) \\ \leq 2\varepsilon t + t^{\frac{3}{2}} k^{\frac{1}{2}} \|\bar{v}_1 - \bar{v}_2\|_{\infty}^{\frac{3}{2}} + \frac{2C}{\min \rho} \frac{1}{k^{1+\beta}} \\ = 2\varepsilon t + O\left(t^{\frac{4+3\beta}{4+\beta}}\right) + O\left(t^{\frac{2+2\beta}{2+\beta}}\right) = o(t). \end{aligned}$$

□

*Proof of Theorem 4.1.* — Recall that

$$\begin{aligned} \Phi_{d\#}(\lambda + tv) &= \frac{1}{d}(\lambda + dt v(\cdot/d)) + \frac{1}{d}(\lambda + dt v((\cdot + 1)/d)) \\ &\quad + \cdots + \frac{1}{d}(\lambda + dt v((\cdot + d - 1)/d)) \end{aligned}$$

and apply the preceding lemma. □

Let us prove that we cannot hope for the Fréchet differentiability of  $\Phi_{d\#}$ . We only treat the case  $d = 2$  for simplicity.

**Proposition 4.3.** — *For all positive  $\varepsilon$ , there is a vector field  $v \in L^2_0$  that satisfies the following:*

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

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 3):  $v(x) = 1/(4k) - y$  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 we are left to prove that if  $k$  is of order  $\varepsilon^{-1}$ , then property 3 holds.

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

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

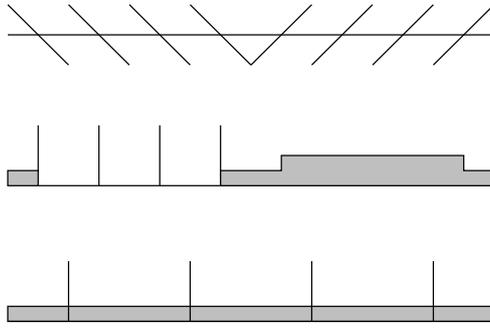


FIGURE 3. The case  $k = 4$ . Up: the graph of  $v$ ; middle:  $\lambda + v$ ; down:  $\Phi_{\#}(\lambda + v)$ .

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  $I_i$ , so that a total mass at least  $1/4$  has to move a distance at least  $1/(8k)$ . From this it follows that  $W(\lambda, \mu) \geq 1/(16k)$ . When  $k$  is chosen with the order of  $\varepsilon^{-1}$ , this distance has at least the order of  $\varepsilon$ , as required.  $\square$

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

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

The proof of Proposition 4.4 consist simply in using Fourier series to show that (up to a multiplicative constant)  $\mathcal{L}_d$  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  $\mathcal{L}_d c_k = \mathcal{L}_d s_k = 0$  when  $d$  does not divide  $k$ , and  $\mathcal{L}_d c_k = dc_{k/d}$ ,  $\mathcal{L}_d s_k = ds_{k/d}$  when  $d|k$ .

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 diagonally on the space  $(\ell^2)^{(\mathbb{N})}$  of sequences  $X = (\underline{x}^0, \underline{x}^1, \underline{x}^2, \dots)$  such that  $\underline{x}^i \in \ell^2$  and  $\sum |\underline{x}^i|_2^2 < \infty$ . Then the map  $\Psi : (\ell^2)^{(\mathbb{N})} \rightarrow L_0^2$  defined by

$$\begin{aligned} \Psi(X) = & \sum_{i,j \in \mathbb{N}} x_j^{2(d-1)i} c_{(di+1)dj} + x_j^{2(d-1)i+1} c_{(di+2)dj} \\ & + \dots + x_j^{2(d-1)i+d-2} c_{(di+d-1)dj} \\ & + x_j^{2(d-1)i+d-1} s_{(di+1)dj} + x_j^{2(d-1)i+d} s_{(di+2)dj} \\ & + \dots + x_j^{2(d-1)i+2d-3} s_{(di+d-1)dj} \end{aligned}$$

is an isomorphism (and even an isometry) that intertwins  $\sigma^{\mathbb{N}}$  and  $\frac{1}{d}\mathcal{L}_d$ . The spectral study of  $\mathcal{L}_d$  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.  $\square$

**4.3. Discussion of the non-Fréchet differentiability.** — 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  $d$  the derivative  $\mathcal{L}_d$  is the Perron-Frobenius operator of  $\Phi_d$ , 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 distortion, 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 section, we consider a general map  $\Phi : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ , assumed to be  $C^2$  and expanding, *i.e.*  $|\Phi'| > 1$ . 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  $\Phi_{\#}$  has a Gâteaux derivative  $\mathcal{L} : L_0^2(\rho\lambda) \rightarrow L_0^2(\rho\lambda)$  at  $\rho\lambda$ , given by*

$$\mathcal{L}v(x) = \sum_{y \in \Phi^{-1}(x)} \frac{\rho(y)}{\rho(x)} v(y) - \frac{\int v \Phi' \frac{\rho}{\rho \circ \Phi}}{\rho(x) \int 1/\rho}$$

Moreover the adjoint operator of  $\mathcal{L}$  in  $L_0^2(\rho\lambda)$  is given by

$$\mathcal{L}^*u = \Phi' u \circ \Phi.$$

**5.1. Proof of Theorem 5.1.** — First, as in the case of  $\Phi_{d\#}$ , Lemma 4.2 shows that for  $v \in L_0^2(\rho\lambda)$ ,

$$(4) \quad d\left(\Phi_{\#}(\rho\lambda + tv), \rho\lambda + t\tilde{\mathcal{L}}v\right) = o(t)$$

where

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

is the first term in the expression of  $\mathcal{L}$ . In words, each of the antecedents 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  $\Phi_{d\#}$ , except that  $\tilde{\mathcal{L}}$  need not to 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 (4) to

make sense, but shows that  $\tilde{\mathcal{L}}v$  cannot be considered as the directional derivative of  $\Phi_{\#}$  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  $W(\rho\lambda + t\tilde{w}, \rho\lambda + tw) = o(t)$ . Moreover,  $w$  is given by*

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

*Proof.* — This is a direct application of Proposition 3.2: we search for a  $w$  such that  $(\rho w)' = (\rho \tilde{w})'$ , so that the densities  $\rho_t$  and  $\tilde{\rho}_t$  of  $\rho\lambda + tw$  and  $\rho\lambda + t\tilde{w}$  are  $L^\infty$  and therefore  $L^1$  close one to the other. This ensures that  $W(\rho\lambda + t\tilde{w}, \rho\lambda + tw) \leq |\rho_t - \tilde{\rho}_t| = o(t)$ .

But 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.  $\square$

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].

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  $\Phi_{\#}$  at  $\rho\lambda$  is given by the composition  $\mathcal{C}\tilde{\mathcal{L}}$ . 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 in the  $L^2(\rho\lambda)$  topology.

To get the expression of  $\mathcal{L}$  given in Theorem 5.1, one only needs a change of variable: denoting by  $\Phi_i^{-1}$  ( $i = 1, 2, \dots, d$ ) the right inverses to

$\Phi$  that are onto intervals  $[a_1 = 0, a_2), [a_2, a_3), \dots, [a_d, a_{d+1} = 1)$  one has

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

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  $\mathcal{L}$  (with the same expression) is

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

and the second term vanishes when  $u$  is in  $L_0^2(\rho\lambda)$ . The first term is also the adjoint in  $L^2(\rho\lambda)$  of  $\tilde{\mathcal{L}}$ , and this adjoint preserves  $L_0^2(\rho\lambda)$ . In other words,  $\mathcal{L}$  is the adjoint in  $L_0^2(\rho\lambda)$  of the adjoint in  $L^2(\rho\lambda)$  of  $\tilde{\mathcal{L}}$ . An interesting feature of the expression of  $\mathcal{L}^*$  is that it does not involve the invariant measure.

**5.2. Spectral study.** — Even if  $\mathcal{L}$  is not a multiple of the Perron-Frobenius operator of  $\Phi$ , its first term  $\tilde{\mathcal{L}}$  is a weighted transfert operator, with weight  $g = \frac{\rho}{\rho \circ \Phi}$ . According to Theorem 2.5 in [Bal00], every number of modulus less than  $R_g = \lim_n (\sup \tilde{\mathcal{L}}^n 1)^{1/n}$  is an eigenvalue of infinite multiplicity with continuous eigenfunctions.

**Proposition 5.3.** — *We have  $R_g \geq \min \Phi' > 1$ , and therefore there is an infinite linearly independent family  $(v_i)_i$  of continuous functions in  $L_0^2(\rho\lambda)$  such that  $\mathcal{L}v_i = v_i$ .*

*Proof.* — Let  $m = \min \Phi'$ : we have  $m > 1$  and, since  $\rho\lambda$  is invariant,

$$\rho(x) = \sum_{y \in \Phi^{-1}(x)} \frac{\rho(y)}{\Phi'(y)} \leq \frac{1}{m} \sum_{y \in \Phi^{-1}(x)} \rho(y)$$

It follows that for all positive continuous function  $f$ ,

$$\tilde{\mathcal{L}}f = \sum_{y \in \Phi^{-1}(x)} \frac{\rho(y)}{\rho(x)} f(y) \geq m |\inf f|;$$

in particular,  $R_g \geq m > 1$  and there is a linearly independent infinite family  $u_0, u_1, \dots, u_i \dots$  of continuous 1-eigenfunctions of  $\tilde{\mathcal{L}}$ . If not all  $u_i$  have mean 0 (with respect to Lebesgue's measure  $\lambda$ ), assume the mean of  $u_0$  is not zero and let  $v_i = u_i - \alpha_i u_0$  where  $\alpha_i$  is chosen such that  $\int v_i \lambda = 0$ . Otherwise, simply put  $v_i = u_i$ .

Now, since  $\tilde{\mathcal{L}}v_i = v_i$  and  $v_i$  has mean zero, we get  $\mathcal{L}v_i = \tilde{\mathcal{L}}v_i = v_i$ .  $\square$

In the same way, we see that all numbers less than  $m > 1$  are eigenvalues of  $\mathcal{L}$  (with infinite multiplicity and continuous eigenfunctions).

## 6. Nearly invariant measures

In this section we prove Theorem 1.4 and Proposition 1.5.

**6.1. Construction.** — Fix some positive integer  $n$  and let  $v_1, \dots, v_n$  be continuous, linearly independent eigenfunctions for  $\mathcal{L} = D_{\rho\lambda}(\Phi_{\#})$ .

For all  $a = (a_1, \dots, a_n) \in B^n(0, \eta)$ , define  $E(a) = \rho\lambda + \sum_i a_i v_i \in \mathcal{P}(\mathbb{S}^1)$  and using Proposition 3.1, choose  $\eta$  small enough to ensure that  $E$  is bi-Lipschitz. Then define  $F(a) = E(\eta a)$  on the unit ball  $B^n$ .

**Proposition 6.1.** — *We have*

$$W(\Phi_{\#}(F(a)), F(a)) = o(|a|)$$

and, as a consequence, for all  $\varepsilon > 0$  and all integer  $K$ , there is a radius  $r$  such that for all  $k \leq K$  and all  $a \in B^n(0, c)$  the following holds:

$$W(\Phi_{d\#}^k(F(a)), F(a)) \leq \varepsilon|a|.$$

*Proof.* — Since we have restricted ourselves to a finite-dimensional space, we have  $W(\Phi_{\#}(\rho\lambda + \eta \sum a_i v_i), \rho\lambda + \eta \sum a_i \mathcal{L}(v_i)) = o(|a|)$  and, since  $\mathcal{L}(v_i) = v_i$ , we get  $W(\Phi_{\#}(F(a)), F(a)) = o(|a|)$ .

The second inequality follows easily. The map  $\Phi_{\#}$  is  $L$ -Lipschitz for some  $L > 1$  ( $L = d$  in the model case,  $L > d$  otherwise). For all  $\varepsilon > 0$  and for all integer  $K$ , let  $r > 0$  be small enough to ensure that

$$|a| < \delta \Rightarrow W(\Phi_{\#}(F(a)), F(a)) \leq \frac{L-1}{L^{k-1}-1} \varepsilon|a|.$$

Then

$$\begin{aligned}
W(\Phi_{\#}^k(F(a)), F(a)) &\leq \sum_{\ell=1}^{k-1} W(\Phi_{\#}^{\ell}(F(a)), \Phi_{\#}^{\ell-1}(F(a))) \\
&\leq \sum_{\ell=1}^{k-1} L^{\ell-1} W(\Phi_{d\#}(F(a)), F(a)) \\
&\leq \varepsilon|a|.
\end{aligned}$$

□

This ends the proof of Theorem 1.4. It would be interesting to have explicit control on  $r$  in terms of  $\varepsilon$ ,  $n$  and  $K$ , and in particular to replace the  $o(|a|)$  by a  $O(|a|^{\alpha})$  for some  $\alpha > 1$ . This seems uneasy because, even in the model case where  $v_i$  are explicit, we can approximate them by  $C^{\infty}$  vector fields  $w_i$  with a good control on  $(-w'_i)^{-1}$  and  $w'$ , but only bad bounds on  $w''$  (and therefore on the modulus of continuity of  $w'$ ).

**6.2. Regularity.** — Let us prove that given  $\mu$  an atomless measure and  $v \in L_0^2(\mu)$  (or, indifferently,  $v \in L^2(\mu)$ ), for all but countably many values of the parameter  $t$ , the measure  $\mu + tv$  has no atom.

*Proof of Proposition 1.5.* — By a line in  $T\mathbb{S}^1 \simeq \mathbb{S}^1 \times \mathbb{R}$ , we mean the image of a non-horizontal line of  $\mathbb{R}^2$  by the quotient map  $(x, y) \mapsto (x \bmod 1, y)$ . We sometimes refer to a line by an equation of one of its lifts in  $\mathbb{R}^2$ .

The measure  $\mu + tv$  has an atom at  $s$  if and only if the measure  $\Gamma = (\text{Id}, v)_{\#}\mu$  defined on  $T\mathbb{S}^1$  gives a positive mass to the line  $(x + ty = s)$ . Since  $\mu$  has no atom, neither does  $\Gamma$ , and since two lines intersect in a countable set, the intersection of two lines is  $\Gamma$ -negligible. It follows that there can be at most  $n$  different lines that are given a mass at least  $1/n$  by  $\Gamma$ . In particular, at most countably many lines are given a positive mass by  $\Gamma$ , and the result follows. □

For a general  $L^2$  vector field, we cannot hope for more. The following folklore example shows a  $L_0^2$  function such that  $\lambda + tv$  is stranger to  $\lambda$  for almost all  $t$ .

**Example 6.2.** — Let  $K$  be a four-corner Cantor set of  $\mathbb{R}^2$ . More precisely,  $A, B, C, D$  are the vertices of a square,  $S_A, S_B, S_C, S_D$  are the homotheties of coefficient  $1/4$  centered at these points, and  $K$  is the unique

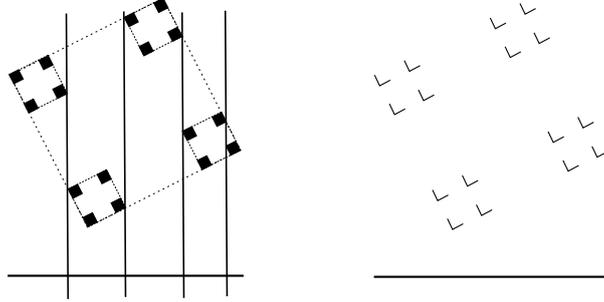


FIGURE 4. A square Cantor set that projects vertically to a segment, but projects in almost all directions to negligible sets. On the right, an approximation of the graph of the function  $v$ .

fixed point of the map defined on compact sets  $M \subset \mathbb{R}^2$  by

$$\mathcal{S}(M) = S_A(M) \cup S_B(M) \cup S_C(M) \cup S_D(M).$$

The Cantor set  $K$  projects on a well-chosen line to an interval, see figure 4, while in almost all directions it projects to  $\lambda$ -negligible sets, see e.g. [PSS03] for a proof. Choose the square so that  $K$  projects vertically to  $[0, 1]$  (identified to  $\mathbb{S}^1$ ), and for  $x \in [0, 1]$  define  $v(x)$  as the least  $y$  such that  $(x, y) \in K$ . Then  $v$  is  $L^2$  and, up to a vertical translation, we can even assume that  $v \in L_0^2$ . But for almost all  $t$ , the measure  $\lambda + tv$  is concentrated into a negligible set.

## 7. The infinitesimal version of Furstenberg's theorem

In this section we prove Theorem 1.7 and Corollary 1.8.

*Proof of Theorem 1.7.* — We only need to look closely at the expressions of  $\mathcal{L}_d := D_\lambda \Phi_{d\#}$ . As indicated in the proof of Proposition 4.4, setting  $c_k(x) = \cos(2\pi kx)$  and  $s_k(x) = \sin(2\pi kx)$ , we have

$$\mathcal{L}_d(c_k) = \mathcal{L}_d(s_k) = 0$$

when  $k$  is not a multiple of  $d$ , and

$$\mathcal{L}_d(c_{md}) = dc_m, \quad \mathcal{L}_d(s_{md}) = ds_m$$

when  $m$  is a positive integer. Note that the union of  $(c_k)_{k \geq 1}$  and  $(s_k)_{k \geq 1}$  is a Hilbert basis of  $T_\lambda \mathcal{P}(\mathbb{S}^1)$ .

It follows that the 1-eigenspace  $E_d$  of  $\mathcal{L}_d$  is generated by the functions

$$\sum_{m \geq 0} d^{-m} c_{d^m j} \quad \text{and} \quad \sum_{m \geq 0} d^{-m} s_{d^m j}$$

where  $j$  runs over positive integers not multiple of  $d$ .

Taking intersections, we first see that  $E_2 \cap E_3$  is generated by the functions

$$\sum_{n, m \geq 1} 2^{-n} 3^{-m} c_{2^n 3^m j} \quad \text{and} \quad \sum_{n, m \geq 1} 2^{-n} 3^{-m} s_{2^n 3^m j}$$

where  $j$  is any positive integer prime with 6. In particular,  $E_2 \cap E_3$  is infinite-dimensional.

We also get that  $\cap_{d \geq 2} E_d$  is generated by

$$\sum_{p \geq 1} p^{-1} c_p \quad \text{and} \quad \sum_{p \geq 1} p^{-1} s_p$$

and is thus 2-dimensional. Note that these functions are indeed in  $L^2(\lambda)$ , and therefore represent tangent vectors at  $\lambda$ .  $\square$

The proof of the corollary is also easy; it relies mostly on the following pointwise version of the continuity equation. We do not claim any novelty in the Lemma below, but still provide a simple proof of the simple case we need.

**Lemma 7.1.** — *Assume that  $(\mu_t)_t$  is a curve of probability measures on  $\mathbb{S}^1$  which is differentiable at 0 with tangent vector  $v \in T_{\mu_0} \mathcal{P}(\mathbb{S}^1)$ , in the sense that*

$$W_2(\mu_t, \mu_0 + tv) = o(t)$$

(be reminded that  $\mu_0 + tv := (\text{Id} + tv)_\# \mu_0$  is the image of  $tv$  by the exponential map at  $\mu_0$ ).

Then for all smooth function  $\psi$ , we have

$$\frac{d}{dt} \int \psi d\mu_t \Big|_{t=0} = \int \psi' v d\mu_0.$$

Note that this lemma is mostly relevant in the case when  $\mu_0$  is regular in the sense of Gigli (i.e., in dimension 1, atomless) since then all “tangent vectors” at  $\mu_0$  are indeed represented by a vector field  $v \in L^2(\mu_0)$ . In a more general manifold, the same result would hold with  $\psi$  a compactly supported function, and  $\nabla \psi$  instead of  $\psi'$ .

*Proof.* — First, we observe that denoting by  $\Pi_t$  an optimal transport plan from  $\mu_t$  to  $\mu_0 + tv$  we have

$$\begin{aligned} \left| \int \psi \, d\mu_t - \int \psi \, d(\mu_0 + tv) \right| &= \left| \int \psi(x) \, d\Pi_t(x, y) - \int \psi(y) \, d\Pi_t(x, y) \right| \\ &\leq \int |\psi(x) - \psi(y)| \, d\Pi(x, y) \\ &\leq \sup(\psi') \int d(x, y) \, d\Pi(x, y) \\ &\leq \sup(\psi') \left( \int d(x, y)^2 \, d\Pi(x, y) \right)^{\frac{1}{2}} \left( \int 1 \, d\Pi(x, y) \right)^{\frac{1}{2}} \\ &= \sup(\psi') \, W_2(\mu_t, \mu_0 + tv) \\ &= o(t) \end{aligned}$$

so that we can use  $\mu_0 + tv$  to estimate the derivative of the integral of  $\psi$ .

Next, we have

$$\begin{aligned} \int \psi \, d(\mu_0 + tv) - \int \psi \, d\mu_0 &= \int (\psi \circ (\text{Id} + tv) - \psi) \, d\mu_0 \\ &= \int (\psi(x + tv(x)) - \psi(x)) \, d\mu_0 \\ &= \int \left( t\psi'(x)v(x) + \frac{1}{2}\psi''(c_x)t^2v(x)^2 \right) \, d\mu_0 \\ &= t \int \psi'v \, d\mu_0 + O(t^2). \end{aligned}$$

Note that the finiteness of  $\|v\|_{L^2(\mu_0)}$  is part of the definition of the tangent space at  $\mu_0$ .

Now the claimed equality follows readily from these two estimates.  $\square$

*Proof of Corollary 1.8.* — Let  $v \in \cap_{d \geq 2} E_d$  be a non-zero tangent vector at  $\lambda$  invariant under all the  $D_\lambda \Phi_{d\#}$ , and define  $\mu_t = \lambda + tv$ .

By definition of the tangent space at  $\lambda$ ,  $\int v \, d\lambda = 0$  so that  $v$  has a well-defined antiderivative. Let  $\psi_0$  be a smooth approximation of one of its anti-derivatives, so that  $\int \psi_0'v \, d\lambda \simeq \int v^2 \, d\lambda$  is non-zero.

Then the pointwise continuity equation implies that

$$\frac{d}{dt} \int \psi_0 \, d\mu_t \Big|_{t=0} = \int \psi_0'v \, d\lambda \neq 0.$$

Moreover, the invariance of  $v$  means that the curve  $(\Phi_{d\#}\mu_t)$  is also differentiable at  $t = 0$ , with tangent vector  $v$ . In consequence, we get for

all smooth function  $\psi$ :

$$\frac{d}{dt} \int \psi d(\Phi_{d\#}\mu_t) \Big|_{t=0} = \int \psi' v d\lambda = \frac{d}{dt} \int \psi d\mu_t \Big|_{t=0}.$$

The weak continuity of  $(\mu_t)$  is obvious, and the fact that  $\mu_t$  is atomless for almost all  $t$  is Proposition 1.5.  $\square$

**Acknowledgements.** — I am indebted to Artur Oscar Lopes for his numerous questions and comments on the various versions of this paper, and it is a pleasure to thank him.

I also wish to thank Frédéric Faure, Étienne Ghys, Nicola Gigli, Antoine Gournay, Nicolas Juillet and Hervé Pajot for interesting discussions and their comments on earlier versions of this paper, and the anonymous referees for their corrections and the constructive criticism of one of them.

### 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.
- [BB00] J.-D. BENAMOU & Y. BRENIER – “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem”, *Numer. Math.* **84** (2000), no. 3, p. 375–393.
- [BLMV09] J. BOURGAIN, E. LINDENSTRAUSS, P. MICHEL & A. VENKATESH – “Some effective results for  $\times a \times b$ ”, *Ergodic Theory Dynam. Systems* **29** (2009), no. 6, p. 1705–1722.
- [EF08] M. EINSIEDLER & A. FISH – “Rigidity of measures invariant under the action of a multiplicative semigroup of polynomial growth on  $\mathbb{T}$ ”, 2008, arXiv:0804.3586 [math.DS].
- [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.

- [HS12] M. HOCHMAN & P. SHMERKIN – “Local entropy averages and projections of fractal measures”, *Ann. of Math. (2)* **175** (2012), no. 3, p. 1001–1059.
- [Joh92] A. S. A. JOHNSON – “Measures on the circle invariant under multiplication by a nonlacunary subsemigroup of the integers”, *Israel J. Math.* **77** (1992), no. 1-2, p. 211–240.
- [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.
- [Klo10a] B. KLOECKNER – “A geometric study of wasserstein spaces: embedding products”, preprint, 2010.
- [Klo10b] ———, “A geometric study of wasserstein spaces: Euclidean spaces”, *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **IX** (2010), no. 2, p. 297–323.
- [Klo13] ———, “Optimal transport and dynamics of expanding circle maps acting on measures”, *Ergodic Theory Dynam. Systems* **33** (2013), no. 2, p. 529–548.
- [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.
- [LW00] E. LINDENSTRAUSS & B. WEISS – “Mean topological dimension”, *Israel J. Math.* **115** (2000), p. 1–24.
- [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.
- [PSS03] Y. PERES, K. SIMON & B. SOLOMYAK – “Fractals with positive length and zero Buffon needle probability”, *Amer. Math. Monthly* **110** (2003), no. 4, p. 314–325.
- [Rud90] D. J. RUDOLPH – “ $\times 2$  and  $\times 3$  invariant measures and entropy”, *Ergodic Theory Dynam. Systems* **10** (1990), no. 2, p. 395–406.
- [uea] USER39115 ET AL. – “Furstenberg  $\times 2 \times 3$  conjecture, bibliography”, <http://mathoverflow.net/questions/161517/furstenberg-times-2-times-3-conjecture-bibliography>.
- [Vil03] C. VILLANI – *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.

- [Vil09] ———, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new.

---

BENOÎT KLOECKNER