A GEOMETRIC STUDY OF WASSERSTEIN SPACES: EMBEDDING POWERS

by

Benoît Kloeckner

Abstract. — The Wasserstein spaces $\mathscr{W}_p(X)$ of a metric space X are sets of sufficiently concentrated measures, endowed with a metric defined using optimal transportation. When X is compact, $\mathscr{W}_p(X)$ is a metrization of the set of probability measures on X endowed with the weak topology.

In this article we prove that X^k always admits a bi-Lipschitz embedding into $\mathscr{W}_p(X)$, with explicit and nearly optimal constants depending on k and p only. This result has an application in dynamics: we show that if X is compact and φ is a self-map of positive entropy, then its action $\varphi_\#$ on the probability measures of X has positive metric mean dimension (with respect to the Wasserstein metrics). This refines the easy result that $\varphi_\#$ has infinite entropy.

1. Introduction

This article inscribes itself in a series, partly joint with Jérôme Bertrand, where given a metric space (X,d) we study some geometric properties of its *Wasserstein spaces* $\mathscr{W}_p(X)$. These spaces of measures are in some sense geometric measure theory versions of L^p spaces (see Section 2 for precise definitions). Here we investigate an embedding question without any further assumption on X.

Several embedding and non-embedding results are proved in previous articles in the series for special classes of spaces X, in the most important case $p=2$. On the first hand, it is easy to see that if X contains a complete geodesic (that is, an isometric embedding of $\mathbb{R}$), then $\mathscr{W}_2(X)$

contains isometric embeddings of open Euclidean cone of arbitrary dimension [**Klo10a**]. In particular it contains isometric embeddings of Euclidean balls of arbitrary dimension and radius, and bi-Lipschitz embeddings of $\mathbb{R}^k$ for all k. On the other hand, if X is negatively curved and simply connected, $\mathscr{W}_2(X)$ does not contain any isometric embedding of $\mathbb{R}^2$ [**BK10**].

Here we prove that the powers of X always embed in a bi-Lipschitz way into $\mathscr{W}_p(X)$.

1.1. The embedding result. — The space X^k can be endowed with several equivalent metrics, for example

$$d_p\big(\bar{x} = (x_1, \dots, x_k)\,,\, \bar{y} = (y_1, \dots, y_k)\big) = \left(\sum_{i=1}^{k} d(x_i, y_i)^p\right)^{1/p}$$

and

$$d_\infty\big(\bar{x}, \bar{y}\big) = \max_{1 \leqslant i \leqslant k} d(x_i, y_i)$$

which come out naturally in the proof; moreover d_∞ is well-suited to the dynamical application below.

***Theorem 1.1* (embedding theorem).** — *Let X be any metric space, $p \in [1, \infty)$ and k be any positive integer. There exists a map $f : X^k \to \mathscr{W}_p(X)$ such that for all $\bar{x}, \bar{y} \in X^k$:*

$$\frac{1}{k(2^k - 1)^{\frac{1}{p}}}\, d_p(\bar{x}, \bar{y}) \leqslant \mathrm{W}_p(f(\bar{x}), f(\bar{y})) \leqslant \left(\frac{2^{k-1}}{2^k - 1}\right)^{\frac{1}{p}} d_p(\bar{x}, \bar{y})$$

and that intertwins dynamical systems in the following sense: given any measurable self-map φ of X, denoting by φ_k the induced map on X^k and by $\varphi_\#$ the induced map on measures, it holds

$$f \circ \varphi_k = \varphi_\# \circ f.$$

Note that since $d_\infty \leqslant d_p \leqslant k^{\frac{1}{p}} d_\infty$ similar bounds hold with d_∞; in fact the lower bound that comes from the proof is in term of d_∞ and is slightly better:

$$\frac{1}{k^{1-\frac{1}{p}}(2^k - 1)^{\frac{1}{p}}}\, d_\infty(\bar{x}, \bar{y}) \leqslant \mathrm{W}_p(f(\bar{x}), f(\bar{y})).$$

This result is proved in Section 3. Let us recall that the push-forward of a measure is defined by $\varphi_\#\mu(A) = \mu(\varphi^{-1}A)$ for all Borelian set A.

We shall see in Section 4 that the constants cannot be improved much for general spaces, but that for some specific spaces, a bi-Lipschitz map with a lower bound polynomial in k can be constructed. This map however does not enjoy the intertwining property.

1.2. A dynamical consequence. — The intertwining property is desirable, since it makes it possible to deduce the following corollary from the embedding theorem.

Corollary 1.2. — *If X is compact and $\varphi : X \to X$ is a continuous map with positive topological entropy, then $\varphi_\#$ has positive metric mean dimension. More precisely*

$$\operatorname{mdim}_M(\varphi_\#, \mathrm{W}_p) \geqslant p\frac{\mathrm{h}_{\mathrm{top}}(\varphi)}{\log 2}.$$

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

Note that the constant in Proposition 1.2 is not optimal in the case of multiplicative maps $\times d$ acting on the circle: in [**Klo10b**] we prove the lower bound $p(d-1)$ (instead of $p\log_2 d$ here).

The following question is natural: is the (topological) mean dimension of $\varphi_\#$ positive as soon as φ has positive entropy? Can this be determined at least for some map φ?

Acknowledgements. — I warmly thank Antoine Gournay for a very interesting discussion, and in particular for introducing me to metric mean dimension.

2. Wasserstein spaces

For a detailled 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, usually assumed to be polish (complete separable) to avoid mesurability issues although this plays no role here, and endowed with its Borel σ-algebra, its L^p *Wasserstein space* is the set $\mathscr{W}_p(X)$ of probability measures μ on X whose p-th moment is finite:

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

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

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

where the infimum is over all probability measures Π on $X\times X$ that projects to μ on the first factor and to ν on the second one. Such a measure is called a transport plan between μ and ν, and is said to be optimal when it achieves the infimum. The function d^p is called the cost function, and the value of $\int_{X\times X} d^p(x,y)\,\Pi(dxdy)$ is the total cost of Π.

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 and W_p metrizes the weak topology.

The name “transport plan” is suggestive: it is a way to describe what amount of mass is transported from one region to another.

3. Proof of the embedding theorem

The first power of X embeds isometrically by $x\to\delta_x$ where δ_x is the Dirac mass at a point. The idea behind the choice of f is to encode a tuple by a measure supported on its elements, without adding any extra symmetry: one should be able to distinct $f(a,b,\dots)$ from $f(b,a,\dots)$. Define the map

$$\begin{array}{rrcl} f: & X^k & \to & \mathscr{W}_p(X) \\ & \bar x=(x_1,\dots,x_k) & \mapsto & \alpha\sum_{i=1}^k \frac{1}{2^i}\delta_{x_i}\end{array}$$

where $\alpha=1/(1-2^{-k})$ is a normalizing constant. This choice of masses moreover ensures that different subsets of the tuple have different masses. Moreover, the intertwining property is obvious since $\varphi_\#(\delta_x)=\delta_{\varphi(x)}$.

Lemma 3.1. — *The map f is $(\alpha/2)^{\frac1p}$-Lipschitz when X^k is endowed with the metric d_p.*

Proof. — 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 L^p cost is

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

so that $\mathrm{W}_p(f(\bar x),f(\bar y))\leqslant(\alpha/2)^{\frac1p}d_p(\bar x,\bar y)$. □

Our goal is now to bound $\mathrm{W}_p(f(\bar{x}), f(\bar{y}))$ from below. This bound is not surprising, and the proof is not really hard. But the point is that since the support of $f(\bar{x})$ and $f(\bar{y})$ can meet, it is not obvious that an optimal transport plan must move a given amount of mass by a given distance. We shall use a combinatorial description of transport plans for this purpose. The cost of all transport plans below are computed with respect to the cost d^p, where p is fixed.

3.1. Labelled graphs. — To describe transport plans, we shall use *labelled graphs*, defined as tuples $G = (V, E, m, m_0, m_1)$ where V is a finite subset of X, E is a set of couples $(x, y) \in V^2$ where $x \neq y$ (so that G is an oriented graph without loops), m is a function $E \to [0, 1]$ and m_0, m_1 are functions $V \to [0, 1]$. An element of V will usually be denoted by x if its thought of as a starting point, y if its thought of as a final point, and v if no such assumption is made.

To any transport plan between finitely supported measures, one can associate a labelled graph as follows.

Definition 3.2. — Let μ, ν be probability measures supported on finite sets $A, B \subset X$ and let Π be any transport plan from μ to ν. We define a labelled graph G^Π by: $V^\Pi = A \cup B$,

$$E^\Pi = \operatorname{supp} \Pi \setminus \Delta = \big\{(x, y) \in X^2 \,\big|\, x \neq y \text{ and } \Pi(\{x, y\}) > 0\big\},$$

$m^\Pi(x, y) = \Pi(\{x, y\})$, $m_0^\Pi(x) = \mu(\{x\})$ and $m_1^\Pi(y) = \nu(\{y\})$.

In other words, the graph encodes the initial and final measures and the amount of mass moved from any given point in $\operatorname{supp} \mu$ to any given point in $\operatorname{supp} \nu$. The transport plan itself can be retrieved from its graph; for example its cost is

$$c_p(\Pi) = \sum_{e \in E} \left(m^\Pi(e)\right)^p.$$

Not every labelled graph encodes a transport plan between two measures. We say that G is *admissible* if:

- $\sum_V m_0(v) = \sum_V m_1(v) = 1$,
- for all $e \in E$, $m(e) > 0$,
- for all $v \in V$, $m_0(v) + \sum_{e=(x,v)\in E} m(e) - \sum_{e=(v,y)\in E} m(e) = m_1(v)$, $\sum_{e=(x,v)\in E} m(e) \leqslant m_1(v)$ and $\sum_{e=(v,y)\in E} m(e) \leqslant m_0(v)$.

A labelled graph is admissible if and only if it is the graph of some transport plan. The next steps of the proof shall give some information on the graphs of optimal plans.

3.2. The graph of some optimal plan is a forest. — Let us introduce some notation related to a given labelled graph G. Given an edge $e \in E$, one denotes its starting point by e^+ and its ending point by e^-. A *path* is a tuple of edges $P = (e_1, \ldots, e_l)$ such that e_i has an endpoint in common with e_{i+1} for all i. If moreover $e_i^+ = e_{i+1}^-$ holds for all i, we say that P is an *oriented path*. We define the *unitary cost* of P as the cost of a unit mass travelling along P, that is $c(P) = \sum_{i=1}^{l} d(e_i^-, e_i^+)^p$, and the *flow* of P as the amount of mass travelling along P, that is $\phi(P) = \min_i m(e_i)$. Cycles and oriented cycles are defined in an obvious, similar way; a graph is a *forest* if it contains no cycle.

Lemma 3.3. — *If* Π *is an optimal plan between any two finitely supported measures* μ, ν*, then* G^Π *contains no oriented cycle.*

Proof. — This is a direct consequence of the so-called cyclic monotony of optimal plans: if there where points $v_1, v_2, \ldots, v_n$ in V^Π such that $v_n = v_1$ and $m(i) := m^\Pi(v_i, v_{i+1}) > 0$ for all $i < n$, then by soustracting the minimal value of m_i to each of them one would get an new admissible labelled graph with $m_0 = m_0^\Pi$ and $m_1 = m_1^\Pi$ and cost less than the cost of G^Π. This new graph would give a new transport plan from μ to ν, cheaper than Π. □

An optimal plan can a priori have non-oriented cycles, but up to changing the plan (without changing its cost), we can assume it does not.

Lemma 3.4. — *Between any two finitely supported measures* μ, ν*, there is an optimal plan* Π *such that* G^Π *is a forest.*

Proof. — Let Π be any optimal plan from μ to ν, and let $G_0 = G^\Pi$ be its graph.

A non-oriented cycle is determined by two sets of vertices $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ and two sets of oriented paths $P_i : x_i \to y_i$, $Q_i : x_i \to y_{i+1}$ where $y_{n+1} := y_1$, see Figure 1.

Consider a minimal non-oriented cycle of G_0, so that no two paths among all P_i's and Q_i's share an edge.

One can construct a new admissible labelled graph G_1, with the same vertex labels m_0 and m_1 than G, by adding a small ε to all $m(e)$ where

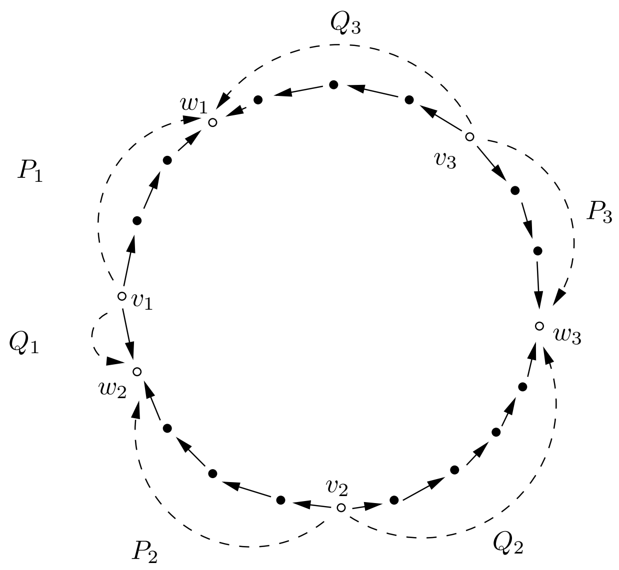

FIGURE 1. A non-oriented cycle: x_i's and y_i's are the vertices where the edges change orientation.

e appears in some P_i, and soustracting the same ε to all $m(e)$ where e appears in some Q_i. This operation adds ε 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 Π.

Since Π 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 ε equal to plus or minus the minimal value of all $m(e)$ where e appears in a P_i or in a Q_i, one designs the wanted new admissible graph G_1.

Now, G_1 has its edge set included in the edge set of G, with at least one less oriented cycle. By repeating this operation, one constructs an admissible labelled graph G without cycle, that has the same total cost and the same vertex labels than G_0. The transport plan defined by G is therefore optimal, from μ to ν. □

The non-existence of cycle has an important consequence.

Lemma 3.5. — *Let Π be a transport plan between two finitely supported measures μ and ν, whose graph is a forest. If there is some real number r such that all $m_0^\Pi(v)$ and all $m_1^\Pi(v)$ are integer multiples of r, then all $m^\Pi(e)$ are integer multiples of r.*

Proof. — Let $G_0 = G^\Pi = (V, E, m, m_0, m_1)$. If G_0 has no edge, then we are done. Otherwise, G_0 has a leaf, that is a vertex x_0 connected to exactly one vertex y_0, by an edge e_0. Assume for example that $e_0 = (x_0, y_0)$ (the other case is treated similarly). Then $m(e_0) = m_0(x_0) - m_1(x_0)$ is an integer multiple of r.

Define $G_1 = (V, E \setminus \{e_0\}, m', m'_0, m'_1)$ where:
- $m'(e) = m(e)$ for all $e \in E \setminus \{e_0\}$,
- $m'_0(x_0) = m_0(x_0) + m(e_0)$,
- $m'_0(x) = m_0(x)$ for all $x \in V \setminus \{x_0\}$,
- $m'_1(y_0) = m_1(y_0) - m(e_0)$,
- $m'_1(y) = m_1(y)$ for all $y \in V \setminus \{y_0\}$.

Then G_1 is still admissible (with different starting and ending measures μ' and ν', though), and all $m'_0(v), m'_1(v)$ are integer multiples of r. By induction, we are reduced to the case of an edgeless graph. □

3.3. End of the proof. — Now we are ready to bound $\mathrm{W}_p(f(\bar{x}), f(\bar{y}))$ from below in terms of $d_\infty(\bar{x}, \bar{y})$. Let i_0 be an index that maximizes $d(x_i, y_i)$ and let Π be an optimal transport plan from $f(\bar{x})$ to $f(\bar{y})$ whose graph $G = (V, E, m, m_0, m_1)$ is a forest.

Lemma 3.6. — *With the notation above, there is a path in G connecting x_{i_0} to y_{i_0}*

Proof. — The choice of f shows that all $m_0(v), m_1(v)$ are integer multiples of $\alpha 2^{-k}$, so that all $m(e)$ are integer multiples of $\alpha 2^{-k}$. Let $n(e), n_0(v), n_1(v) \in \mathbb{N}$ be such that $m(e) = n(e)\alpha 2^{-k}$, $m_0(v) = n_0(v)\alpha 2^{-k}$ and $m_1(v) = n_1(v)$. Then the only $v \in V = \operatorname{supp} f(\bar{x}) \cup \operatorname{supp} f(\bar{y})$ such that $n_0(v)$ contains 2^{k-i_0} in its base-2 expansion is x_{i_0}. Similarly, the only $w \in V$ such that $n_1(w)$ contains 2^{k-i_0} in its base-2 expansion is y_{i_0}. Let $E' \subset E$ be the set of edges e such that $n(e)$ contains 2^{k-i_0} in its base-2 expansion.

Any vertex v such that $n_0(v) - n_1(v)$ does not contain 2^{k-i_0} in its base-2 expansion must be adjacent to an even number of edges of E'. Therefore the non-oriented graph induced by E' is Eulerian, with exactly two points of odd degree: x_{i_0} and y_{i_0}. In particular, x_{i_0} and y_{i_0} are connected by a path in E'. □

Let P_0 be a minimal path between x_{i_0} and y_{i_0}. Each endpoint of each edge in this path has to be some y_i, all distinct by minimality, so that P_0 has length at most k. It follows by a convexity argument that $c(P_0)$

is at least $k(d(x_{i_0}, y_{i_0})/k)^p$. Moreover $\phi(P) \geqslant \alpha 2^{-k}$ so that the cost of Π is at least $\alpha 2^{-k} d(x_{i_0}, y_{i_0})^p / k^{p-1}$. We get

$$\mathrm{W}_p(f(\bar{x}), f(\bar{y})) \geqslant \frac{\alpha^{\frac{1}{p}} 2^{-\frac{k}{p}}}{k^{1-\frac{1}{p}}} d_\infty(\bar{x}, \bar{y}) \geqslant \frac{1}{k(2^k-1)^{\frac{1}{p}}} d_p(\bar{x}, \bar{y})$$

which ends the proof of Theorem 1.1.

4. Discussion of the embedding constants

One can wonder if the constants in Theorem 1.1 are optimal. We shall see in the simplest possible example that they are off by at most a polynomial factor, then see how they can be improved in a specific case.

Proposition 4.1. — *Let $X = \{0, 1\}$ where the two elements are at distance 1 and consider a map $g : X^k \to \mathscr{W}_p(X)$ such that*

$$m\, d_p(\bar{x}, \bar{y}) \leqslant \mathrm{W}_p(g(\bar{x}), g(\bar{y})) \leqslant M\, d_p(\bar{x}, \bar{y})$$

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

$$m \leqslant \frac{1}{(2^k-1)^{\frac{1}{p}}} \quad \textit{and} \quad \frac{M}{m} \geqslant \left(\frac{2^k-1}{k}\right)^{\frac{1}{p}}.$$

Moreover there is a map whose constants satisfy $m = (2^k-1)^{-\frac{1}{p}}$ and $M/m \leqslant (2^k-1)^{\frac{1}{p}}$.

Proof. — By homogeneity, it is sufficient to consider $p = 1$, in which case X^k is the k-dimensional discrete hypercube endowed with the Hamming 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^k is k, so that the diameter of $g(X^k)$ is at most Mk. Since $g(X^k)$ has 2^k elements, by the pigeon-hole principle at least two of them are at distance at most $(2^k-1)^{-1}Mk$. Since the distance between their inverse images is at least 1, we get $m \leqslant (2^k-1)^{-1}Mk$ so that $M/m \geqslant (2^k-1)/k$. The pigeon-hole principle also gives $m \leqslant (2^k-1)^{-1}$ simply by using that $\mathscr{W}_1(X)$ has diameter 1.

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

Note that in Proposition 4.1, one could improve the lower bound on M/m by a factor asymptotically of the order of $2^{\frac{1}{p}}$ by using the fact that every element in X^n has an opposite, that is an element at distance n from it.

Let us give an example where the constants are much better.

Example 4.2. — Let $X = \{0,1\}^{\mathbb{N}}$ with the following metric: given $x = (x^1, x^2, \dots) \neq y = (y^1, y^2, \dots)$ in X, $d(x,y) = 2^{-i}$ where i is the least index such that $x^i \neq y^i$. Then given k, let ℓ be the least integer such that $2^\ell \geqslant k$ and let $w_1, \dots, w_k \in \{0,1\}^\ell$ be distinct words on ℓ letters. For $x = (x^1, x^2, \dots) \in X$ and $w = (w^1, \dots, w^\ell) \in \{0,1\}^\ell$, define wx as the element $(w^1, w^2, \dots, w^\ell, x^1, x^2, \dots)$ of X.

Now let $g : X^k \to \mathscr{W}_p(X)$ be defined by

$$g\big(x = (x_1, \dots, x_k)\big) = \sum_{i=1}^{k} \frac{1}{k} \delta_{w_i x_i}.$$

For all $x, y \in X$ and all $i \neq j$, we have $d(w_i x, w_j y) \geqslant 2^{-\ell} \geqslant d(w_i x, w_i y)$. It follows that

$$\begin{aligned} \mathrm{W}_p(g(\bar{x}), g(\bar{y})) &= \left(\frac{1}{k} \sum_i 2^{-p\ell} d^p(x_i, y_i)\right)^{\frac{1}{p}} \\ &= \frac{1}{k^{\frac{1}{p}} 2^\ell} d_p(\bar{x}, \bar{y}). \end{aligned}$$

For this example, we have $M = m$ and moreover m has only the order of $k^{-1-\frac{1}{p}}$ instead of being exponentially small.

This example could be generalised to more general spaces, for example the middle-third Cantor set. What is important is: that the various components of a given depth are separated by a distance at least the diameter of the components; that the metric does not decrease too much between $d(x,y)$ and $d(wx, wy)$ (any bound that is exponential in the length of w would do).

5. Dynamical consequences

In this section, X is asumed to be compact. Given a continuous map $\varphi : X \to X$, for any $n \in \mathbb{N}$ one defines a new metric on X by

$$d_{[n]}(x,y) := \max\{d(\varphi^i(x),\varphi^i(y)); 0 \leqslant i \leqslant n\}.$$

Given $\varepsilon > 0$, one says that a subset S of X is (n,ε)-separated if $d_{[n]}(x,y) \geqslant \varepsilon$ whenever $x \neq y \in S$. Denoting by $N(\varphi,\varepsilon,n)$ the maximal size of a (n,ε)-separated set, the topological entropy of φ is defined as

$$\mathrm{h_{top}}(\varphi) := \lim_{\varepsilon\to 0} \limsup_{n\to+\infty} \frac{\log N(\varphi,\varepsilon,n)}{n}.$$

Note that this limit exists since $\limsup_{n\to+\infty} \frac{1}{n} \log N(\varphi,\varepsilon,n)$ is nonincreasing in ε. The adjective "topological" is relevant since $\mathrm{h_{top}}(\varphi)$ 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 $\times d : x \mapsto dx \mod 1$ acting on the circle is a classical example, whose topological entropy is $\log d$.

Now, the metric mean dimension is

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

It is zero as soon as topological entropy is finite. Note that 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 now prove that when $\mathrm{h_{top}}(\varphi) > 0$, then $\varphi_\# : \mathscr{W}_p(X) \to \mathscr{W}_p(X)$ has positive metric mean dimension.

Proof of Corollary 1.2. — Let $\varepsilon, \eta > 0$ and k be such that $\eta \geqslant k(2^k - 1)^{\frac{1}{p}}\varepsilon$. If A is a (n,η)-separated set for (X,φ,d) then $A^k \subset X^k$ is a (n,η) separated set for (X^k,φ_k,d_∞). Then Theorem 1.1 shows that $f(A^k)$ is a (n,ε)-separated set for $(\mathscr{W}_p(X),\varphi_\#,\mathrm{W}_p)$, so that

$$N(\varphi_\#,\varepsilon,n) \geqslant \left(N(\varphi,k(2^k-1)^{1/p}\varepsilon,n)\right)^k.$$

Let $H < \mathrm{h_{top}}(\varphi)$ and $\beta < 1$. For all $\varepsilon > 0$ small enough, and for arbitrarily large integer n we have $N(\varphi,\varepsilon,n) \geqslant \exp(nH)$. Define

$$k = \left\lfloor \frac{\beta p(-\log\varepsilon)}{\log 2} \right\rfloor;$$

then $k(2^k-1)^{1/p}\varepsilon = O\left((-\log\varepsilon)\varepsilon^{1-\beta}\right) \to 0$ when $\varepsilon\to 0$. Therefore, for all small enough ε, there are arbitrarily large n such that

$$\begin{aligned}
N(\varphi_\#,\varepsilon,n) &\geqslant \exp(nHk) \\
&\geqslant \exp\left(nH\left(\frac{\beta p}{\log 2}(-\log\varepsilon)-1\right)\right) \\
\frac{\log N(\varphi_\#,\varepsilon,n)}{n(-\log\varepsilon)} &\geqslant \frac{H\beta p}{\log 2} - \frac{H}{-\log\varepsilon} \\
\operatorname{mdim}_M(\varphi_\#,\mathrm{W}_p) &\geqslant \frac{H\beta p}{\log 2}
\end{aligned}$$

Letting $H\to \mathrm{h_{top}}(\varphi)$ and $\beta\to 1$ gives

$$\operatorname{mdim}_M(\varphi_\#,\mathrm{W}_p) \geqslant p\frac{\mathrm{h_{top}}(\varphi)}{\log 2}$$

as claimed. □

In the case of the shift on $\{0,1\}^{\mathbb{N}}$, one could want to use the better bound obtained in Example 4.2. But the map g defined there does not intertwin φ_k and $\varphi_\#$, and the method above does not apply.

References

[BK10] J. Bertrand & B. Kloeckner – "A geometric study of wasserstein spaces: Hadamard spaces", preprint, 2010.

[Ham80] R. W. Hamming – *Coding and information theory*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1980.

[Klo10a] 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.

[Klo10b] ———, "Optimal transport and dynamics of circle expanding maps acting on measures", preprint, 2010.

[LW00] E. Lindenstrauss & B. Weiss – "Mean topological dimension", *Israel J. Math.* **115** (2000), p. 1–24.

[Vil03] C. Villani – *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.

Benoît Kloeckner