CONTRACTION IN THE WASSERSTEIN METRIC FOR SOME MARKOV CHAINS, AND APPLICATIONS TO THE DYNAMICS OF EXPANDING MAPS.

by

Benoît R. Kloeckner, Artur O. Lopes & Manuel Stadlbauer

Abstract. — We employ techniques from optimal transport in order to prove decay of transfer operators associated to iterated functions systems and expanding maps, giving rise to a new proof without requiring a Doeblin-Fortet (or Lasota-Yorke) inequality.

Our main result is the following. Suppose T is an expanding transformation acting on a compact metric space M and $A : M \to \mathbb{R}$ a given fixed Hölder function, and denote by $\mathscr{L}$ the Ruelle operator associated to A. We show that if $\mathscr{L}$ is normalized (i.e. if $\mathscr{L}(1) = 1$), then the dual transfer operator $\mathscr{L}^*$ is an exponential contraction on the set of probability measures on M with the 1-Wasserstein metric.

Our approach is flexible and extends to a relatively general setting, which we name Iterated Contraction Systems. We also derive from our main result several dynamical consequences; for example we show that Gibbs measures depends in a Lipschitz-continuous way on variations of the potential.

1. Introduction and statement of the main results

It has already been noticed that the 1-Wasserstein distance issued from optimal transportation theory is very convenient to prove exponential contraction properties for Markov chains (see e.g. [**HM08, Sta13, Oll09**]). In this article, we observe that this idea applies very effectively to the dynamics of expanding maps: indeed the dual transfer operator of an expanding map with respect to a normalized potential can be seen as a Markov chain, for which we prove exponential contraction. We shall notably deduce from this result several Lipschitz stability results for expanding maps: stability of Gibbs measures in terms of a variation of the potential, stability of the maximal entropy measure in terms of a variation of the map, etc.

By these results and the simplicity of the proofs, we hope that the present article will make a clear case about the usefulness of the application of coupling techniques and objects from optimal transport to dynamical systems and thermodynamical formalism (general references for this last topic are [**PP90**] and [**Bal00**]).

Note that a similar coupling has been used in e.g. [**BFG99**] in order to show decay of correlations for Gibbs measures of low-regularity potential in the case of the shift, but their arguments are not using methods from optimal transport. While we stick here to the more standard case of Hölder potentials, we take a more geometric point of view that allows us firstly to handle a much broader family of dynamical systems and secondly to derive a number of corollaries. Namely, the contraction in the Wasserstein metric easily implies a spectral gap and decay of correlations, but also the stability results alluded to above.

Our main result and method of proof are also similar to a recent result of the third named author for some random Markov shifts ([**Sta13**]); again the present result is less general in some aspects and more general in others since in here we only consider non-random dynamical systems but are able to cover a wide range of expanding maps and iterated function systems.

We consider the following setting: let (Ω, d) be a compact metric space, $k \in \mathbb{N}$ and F a map which assigns to $x \in \Omega$ a k-multiset $F(x) \subset \Omega$. That is, allowing multiple occurrences of elements, $F(x)$ contains k elements (a typical example is given by $F(x) = T^{-1}(x)$ where T is a k-to-1 map). We then refer to F as a *k-iterated contraction system (ICS)* if there exists $\theta < 1$ such that for all $x, y \in \Omega$ there exists a bijection $x_i \mapsto y_i$ between $F(x)$ and $F(y)$ with $d(x_i, y_i) \leq \theta d(x, y)$ for all $i = 1, \dots, k$. We will say that a transformation T of Ω is a *regular expanding* map if $T^{-1}(\{x\})$ defines an ICS once its elements are given suitable multiplicities. For more details we refer to section 2.

Observe that this class of dynamical systems contains, among others, expanding local diffeomorphisms of compact Riemannian manifolds and iterated function systems (IFS) given by k contractions on Ω. A general reference for IFS is [**MU03**].

The transfer operator with respect to a given continuous function $A : \Omega \to \mathbb{R}$ is defined as usual by, for $f : \Omega \to \mathbb{R}$ continuous,

$$\mathscr{L}(f)(x) = \sum_{y \in F(x)} e^{A(y)} f(y).$$

Furthermore, let ρ refer to the spectral radius of $\mathscr{L}$ acting on continuous functions and suppose that $h : \Omega \to \mathbb{R}$ is strictly positive and Lipschitz continuous with $\mathscr{L}(h) = \rho h$; we will show that such an h exists and is unique up to multiplication by constants in proposition 3.1 and corollary 5.2 below. Then the *normalized operator* defined by

$$\mathbb{P}(x) = \mathscr{L}(h \cdot f)(x)/\rho h(x)$$

satisfies $\mathbb{P}(1) = 1$ and is conjugate to $\mathscr{L}$ up to the constant ρ; the iterates are related through $\rho^n h \cdot \mathbb{P}^n(f) = \mathscr{L}^n(h \cdot f)(x)$. By uniqueness of h, $\mathbb{P}(x)$ is uniquely determined by F and A. Also note that in case of an ICS which is defined through a map T, the above operator can be obtained by substituting A by the *normalized potential* $A + \log h - \log h \circ T - \log \rho$.

General references on Transport Theory and the Wasserstein distance are [**Vil03**], [**Vil09**], [**AGS08**] and [**Gig11**].

***Theorem 1.1* (Contraction property)**. — *Let F be an iterated contraction system with contraction ratio $\theta \in (0, 1)$ and let A be a Lipschitz-continuous potential on Ω. Then the dual $\mathbb{P}^*$ of the normalized transfer operator $\mathbb{P}$ is exponentially contracting on probability measures in the Wasserstein metric. That is, for all $n \in \mathbb{N}$ and all $\mu, \nu \in \mathscr{P}(\Omega)$ we have*

$$W_1((\mathbb{P}^*)^n\mu, (\mathbb{P}^*)^n\nu) \leq C\lambda^n W_1(\mu, \nu).$$

where C and $\lambda < 1$ are constants depending only on θ, the Lipschitz constant $\mathrm{Lip}(A)$ *and* $\operatorname{diam}\Omega$.

There are several features of this result that we wish to stress before giving applications. First, there is no dimension restriction: our purely metric arguments are very flexible and do not depend on a Doeblin-Fortet inequality (also known as Ionescu Tulcea-Marinescu or Lasota-Yorke inequality, [**DF37**]), so that the proof also applies to, say, expanding circle maps and expanding maps on higher-dimensional manifolds.

This metric setting also enables us to extend the result from Lipschitz to Hölder regularity without difficulty: the result applies equally well to Ω endowed with the metric d^α when $\alpha \in (0, 1]$, and any potential which is α-Hölder in the metric d. The conclusion then involves the 1-Wasserstein metric W_α of d^α (also known as the α-Wasserstein metric of d), but if needed one can use the obvious inequalities

$$W_1 \leq (\operatorname{diam}\Omega)^{1-\alpha}\, W_\alpha \leq (\operatorname{diam}\Omega)^{1-\alpha}\, W_1^\alpha.$$

We only state our results with respect to Lipschitz regularity to avoid making the notation heavier.

Note that the constants C and λ are explicit, though convoluted (and λ may be much closer to 1 than θ).

The Wasserstein metric is in our opinion a natural metric (for example it metrizes the weak-$*$ topology on probability measures when Ω is compact), but its relevance is much deeper: we will state below several corollaries whose proofs rely on the metric being W_1, but whose statement are free from any reference to optimal transport.

Let us now give some consequences of Theorem 1.1. Unless stated otherwise, we always consider an iterated contraction system F with contraction ratio $\theta \in (0,1)$ on a phase space Ω and a Lipschitz potential A, we denote by $\mathscr{L}$ the transfer operator and by $\mathbb{P}$ its normalization. The dependency of constants on $\operatorname{Lip}(A)$, θ, $\operatorname{diam}\Omega$ will be kept implicit and C, λ will always denote the constants given in Theorem 1.1.

The first obvious consequence of the contraction is that $\mathbb{P}^*$ fixes a unique probability measure μ_A; note that in case F is given by an expanding map T, this μ_A is the well-known invariant Gibbs measure associated with the potential A.

We proceed with a property of classical flavor.

***Corollary 1.2* (Spectral gap).** — *The action on Lipschitz functions of $\mathbb{P}$ is exponentially contracting on a complement of the set of constant functions (which by normalization is the 1-eigenspace of $\mathbb{P}$).*

More precisely, for each Lipschitz function $\zeta : \Omega \to \mathbb{R}$ with $\int \zeta \,\mathrm{d}\mu_A = 0$, we have

$$\|\mathbb{P}^n \zeta\|_{\mathrm{Lip}} \leq C_2(\zeta)\lambda^n$$

where $C_2(\zeta) = C(1 + \operatorname{diam}\Omega)\operatorname{Lip}(\zeta)$ and $\|\cdot\|_{\mathrm{Lip}} = \|\cdot\|_\infty + \operatorname{Lip}(\cdot)$ denotes the Lipschitz norm.

We now turn to stability results (see Section 6).

***Corollary 1.3* (Lipschitz-continuity of the Gibbs map)**

Assume that A, B are normalized Lipschitz potentials for the same ICS F and let μ_A and μ_B refer to the corresponding Gibbs measures. Then

$$W_1(\mu_A, \mu_B) \leq C_3 \|A - B\|_\infty$$

where $C_3 = \frac{C}{1-\lambda} \operatorname{diam}\Omega$.[(1)] *In particular, for any Lipschitz test function φ, we have*

$$\Big| \int \varphi \,\mathrm{d}\mu_A - \int \varphi \,\mathrm{d}\mu_B \Big| \leq C_3 \operatorname{Lip}(\varphi)\, \|A - B\|_\infty.$$

Note that if we translate this in α-Hölder potentials, the Gibbs map is still locally *Lipschitz* on the space of α-Hölder potentials, with the space of measures endowed with W_α. The estimate with test functions then stands for α-Hölder test functions.

[(1)] Only $\operatorname{Lip}(A)$ appears in C and λ, by no accident: we only need to control one Lipschitz constant, not both.

We turn to result which are specific to the case of regular expanding maps; i.e. we now assume that F is obtained from a map T. First, Corollary 1.3 implies the following.

Corollary 1.4 **(Continuity of the metric entropy)**
If A and B are normalized Lipschitz potentials, then

$$|h(\mu_A) - h(\mu_B)| \leq C_4 \|A - B\|_\infty$$

where $C_4 = \frac{C\,\mathrm{Lip}(A)}{1-\lambda} \operatorname{diam}\Omega + 1$.

We are also able to deal with variations of the map T; as an illustration of our method, we concentrate on a simple case where potential variation will not interfere. We will use the following notation for the uniform distance between maps acting on the same space:

$$d_\infty(T_1, T_2) := \sup_{x\in\Omega} d(T_1(x), T_2(x)).$$

Corollary 1.5 **(Continuity of the maximal entropy measure)**
Let T_1 and T_2 be two C^1 expanding maps on the same manifold Ω with the same number k of sheets, assume that one of them is $1/\theta$-expanding, and let μ_i be the maximal entropy measure of T_i for $i = 1, 2$.
If $\|T_1 - T_2\|_\infty \leq \frac{1}{4}\operatorname{sys}(\Omega)$ then

$$W_1(\mu_1, \mu_2) \leq C_5\, d_\infty(T_1, T_2)$$

where $C_5 = \frac{2C}{1-\lambda}$ and C is computed with $\mathrm{Lip}(A) = 0$.

In the above result, $\operatorname{sys}(\Omega)$ denotes the systole of the manifold Ω, i.e. the length of the shortest non-homotopically trivial curve (see [**Gro81**] for general results and references on the topic). The restriction on $\|T_1 - T_2\|_\infty$ can possibly be waived; e.g. it would be sufficient to prove that the space of expanding maps on a manifold is connected by small jumps.

It is also very likely that Corollary 1.5 extends in some form to many other classes of expanding maps (e.g. piecewise uniformly expanding interval maps), but we do not have a general argument that would avoid a cumbersome list of specific results; its main part is a general result, Corollary 6.2 below.

Note that Corollary 1.5 deals with the regularity of a natural invariant measure in terms of a varying expanding map, in the same spirit of many previous works (see [**Rue98**], [**BS08**], [**Bal08**], [**HM10**] and [**BCV12**]) in which the absolutely invariant measure was considered. These papers are all in the so called Linear Response Theory. Here, the maximal entropy measures we deal with are most of the time singular with respect to Lebesgue measure and singular one with respect to the other, a setting where many previous approaches are difficult to apply.

Our method depends on an argument which only applies to operators $\mathscr{L}^*$ when they map probability measures to probability measures. Therefore, it is essential to normalize these operators, thus to have a Ruelle-Perron-Frobenius theorem in the setting of ICS. This is the role of Proposition 3.1, and it is worth noting that the method of proof, even though obviously inspired by the construction of conformal measures in [**DU91**], seems to be new. Unfortunately, as we would otherwise need to control the variations of the map the sends a potential to its normalized counterpart, we have to require that the potentials in corollaries 1.4 and 1.5 are already normalized.

Note that below, we introduce a pretty general framework which enables us to treat IFS in the same setting as expanding maps; our main motivation for this is simply to treat expanding maps on manifolds and piecewise uniformly expanding (onto) maps together; but an IFS comes naturally with a transfer operator, to which most of the above results apply. In particular, it is possible to deduce from our results that two self similar IFS which are close one to another have their "natural measures" close one to the other.

2. Definitions and examples

In this section we introduce the precise setting in which we will work. We tried to set unified notation applicable in as broad a generality as possible, which explains why our definitions are not totally standard.

2.1. Iterated contraction systems. — Iterated contraction systems, to be defined below, are a natural generalization of iterated function systems. The only departure from the usual setting is that instead of considering a finite set of contracting maps, we consider one multiset-valued map with contraction properties. The reason for this choice is that it makes this notation immediately applicable to expanding maps, see Section 2.3

Definition 2.1. — We shall define a *multiset* with k elements (or k-multiset) as the orbit of a k-tuple under the action of the permutation group S_k; we will denote a multiset using the usual set braces, repeating elements if needed: for example $\{1, 2, 2, 5\}$ is a multiset with 4 elements.

The set of elements of a multiset is called its *underlying set*.

Then the *multiplicity function* $\mathbf{1}_A$ of a multiset A whose elements are in some "universal" set Ω is the functions which maps every element of Ω to its multiplicity as an element of A; the multiplicity function contains all the information on A. The *sum* of multisets A and B is the multiset $A \uplus B$ whose multiplicity function is $\mathbf{1}_A + \mathbf{1}_B$.

A *bijection* f between k-multisets A and B is the data of k pairs (a_i, b_i) such that $A = \{a_1, \dots, a_k\}$ and $B = \{b_1, \dots, b_k\}$; beware that the functional notation $b_i = f(a_i)$ would be misleading as we could have $a_i = a_j$ while $f(a_i) \neq f(a_j)$; we therefore sometimes write $f(i) = (a_i, b_i)$, with the understanding that for any permutation π, the map $f_\pi := i \mapsto (a_{\pi(i)}, b_{\pi(i)})$ is identified with f.

The set of all k-multisets whose elements are taken in some set Ω is denoted by $\mathscr{M}_k(\Omega)$.

When summing and multiplying over multisets, each element appears in the sum as many times as it appears in the multiset:

$$\sum_{x\in\{1,2,2,5\}} x = 1+2+2+5.$$

Definition 2.2. — Let Ω be a complete metric space, k be a positive integer, and F be a map $\Omega \to \mathscr{M}_k(\Omega)$.

We say that F is an *iterated contraction system* (*ICS* for short, k-ICS or ICS with k terms if we want to make k explicit) if there is a number $\theta \in (0,1)$ (called *contraction ratio*) such that for all $x, y \in \Omega$ there is a bijection $f = (x_i, y_i)_i$ between $F(x)$ and $F(y)$ such that for all i,

$$d(x_i, y_i) \leq \theta d(x, y).$$

The iterates of F are the ICS $F^t : \Omega \to \mathscr{M}_{k^t}(\Omega)$ (where $t \in \mathbb{N}$) defined by

$$F^1 = F \quad \text{and} \quad F^{n+1}(x) = \biguplus_{y\in F^n(x)} F(y);$$

note that θ^n is a contraction ratio for F^n.

If A is a subset of Ω, we denote by $F(A)$ the union of all the underlying sets of the $F(a)$, when a runs over A.

Let us give a few motivating examples.

Example 2.3. — Consider an IFS, that is a family of k contracting maps $F_1, \dots, F_k$ of Ω. The multiset valued map defined by $F(x) = \{F_1(x), \dots, F_k(x)\}$ is an ICS: the bijection between $F(x)$ and $F(y)$ is simply given by the pairs $(F_i(x), F_i(y))$. The contraction ratio of F is the largest contraction ratio of the F_i.

This is a very particular kind of ICS, since we have globally defined sections of F (i.e., maps that selects continuously for each x an element of $F(x)$); but $F(x)$ is not a set whenever two F_i's take the same value at x.

Example 2.4. — Consider the map

$$T : x \mapsto 2x \mod 1$$

acting on $S^1 = \mathbb{R}/\mathbb{Z}$, and for each $x \in S^1$ let $F(x) = T^{-1}(\{x\})$. Then F is an ICS with contraction ratio $1/2$.

This is a very particular kind of ICS, since $F(x)$ is always a set; but as is well-known we do not have globally defined sections, so that it is not possible to obtain F from an IFS. However, this ICS has the nice property that each x admits a neighborhood on which sections can be defined (we say that F admits local sections).

Example 2.5. — The following map acting on the closed unit disc of $\mathbb{C}$ is an ICS with contraction ratio 1/2:

$$F : re^{2i\pi\alpha} \mapsto \left\{ \frac{r}{2}e^{i\pi\alpha}, \frac{r}{2}e^{i\pi(\alpha+1)} \right\}$$

Note that $F(x)$ is a set except when $x = 0$, as $F(0) = \{0, 0\}$. This ICS does not even admit *local* sections around the origin.

Just like an IFS, an ICS admits a unique attractor, i.e. a non-empty compact set A such that $A = F(A)$ (proof: the map $A \mapsto F(A)$ is a contraction in the Hausdorff metric, thus has a unique fixed point). Moreover this attractor can be approximated by iterating F on any given non-empty compact set.

2.2. Markov chains associated to an ICS and potentials. — Let F be an ICS on a complete metric space Ω; up to restricting F to its attractor, we assume that Ω is compact and that $\Omega = F(\Omega)$.

Definition 2.6. — A Markov chain on Ω is said to be *compatible* with F if at each $x \in \Omega$, its kernel $P(x, \cdot)$ is supported on the underlying set of $F(x)$. In other words, if the position at time t of the Markov chain is x, we ask that with probability one the position at time $t + 1$ is an element of $F(x)$.

Note that compatibility only depends on the underlying set-valued map of F. We will be interested by very specific compatible Markov chains, where the transition probabilities are given by a normalization of a potential function only depending on the target points: these Markov chains indeed occur in the thermodynamical formalism, which is our main motivation.

Definition 2.7. — A *potential* is simply a continuous function $A : \Omega \to \mathbb{R}$; it is said to be *normalized* with respect to F if for all $x \in \Omega$ we have

$$\sum_{y \in F(x)} e^{A(y)} = 1,$$

where we sum over the *multi*set $F(x)$.

The *Markov chain associated to* a normalized potential A is defined by letting $m \cdot e^{A(y)}$ be the transition probability from x to y whenever y is an element of $F(x)$ of multiplicity m.

We denote by $\mathscr{L}^*_{F,A}$ (leaving asside any subscripts that are clear from the context) the operator on finite, signed measures, defined by

$$\int \varphi(x)\,\mathrm{d}(\mathscr{L}^*\mu)(x) = \int \sum_{y\in F(x)} e^{A(y)}\varphi(y)\,\mathrm{d}\mu(x)$$

whenever φ is a continuous test function. In other words, $\mathscr{L}^*$ is the dual of the transfer operator defined by

$$\mathscr{L}\varphi(x) = \sum_{y\in F(x)} e^{A(y)}\varphi(y).$$

Note that, if A is normalized, then $\mathscr{L}(1) = 1$ and $\mathscr{L}^*$ maps probability measures to probability measures.

In case of a non-normalized potential, the associated Markov chain is obtained through a normalization of $\mathscr{L}$ through the constrution of an invariant function in proposition 3.1 as shown below (see definition 3.2).

The simplest example of a normalized potential is the constant one: $A(y) = -\log k$ where k is the number of terms of F. For example if F is an IFS with uniform contraction ratio, the stationary probability of the Markov chain associated to A is the usual canonical measure on the fractal attractor defined by F.

Other examples are easy to construct when F is an IFS with the "strong separation property": F has global sections $F_1,\dots,F_k$ with disjoint images, and any sufficiently negative continuous function on $F_1(\Omega)\cup\dots\cup F_{k-1}(\Omega)$ can be extended to a normalized potential by suitably choosing its values on $F_k(\Omega)$.

2.3. The case of expanding maps. — The definition of expanding maps may vary in the literature; the one we adopt fits what we will need in the proof of the contraction property, and includes in the same framework shifts, some IFS, classical smooth expanding maps, piecewise expanding unimodal maps and other examples.

Definition 2.8. — If Ω is a compact metric space, a continuous map $T : \Omega\to\Omega$ is said to be *regular expanding* if $T^{-1} : x\mapsto T^{-1}(\{x\})$ is the underlying set-valued map of a k-ICS F, where $k = \max\{\#T^{-1}(\{x\}) \mid x\in\Omega\}$.

We say that T has k sheets, and if θ is a contraction ratio of F then we say that T is $\frac{1}{\theta}$-expanding.

It is not clear from this definition that F is uniquely defined by T; but in the cases we will consider, the set of points x having a maximal number of inverse images is dense in Ω, so that F is in fact uniquely defined by T.

Example 2.9. — Let Ω be a compact Riemannian manifold, and $T : \Omega \to \Omega$ be a C^1 map such that $\|D_x T(v)\| \geq \frac{1}{\theta}\|v\|$ for some $\theta \in (0,1)$ and all $(x,v) \in TM$. Then T is regular expanding; indeed T is a local diffeomorphism, thus a covering map and $F(x) = T^{-1}(x)$ defines an IFS: the uniformly expanding property of $D_x T$ easily ensures the contracting property for F, using the lifting property on a minimizing geodesic from x to y to pair their inverse images.

Note that few manifolds admit expanding maps, an obvious example being the torus of any dimension. The keyword here is "infra-nil-manifold", but we will not elaborate on this topic.

Example 2.10. — Let $\Omega = [a,b]$ be a closed interval, and $T : \Omega \to \Omega$ be a piecewise C^1 expanding unimodal map; that is, for some $c \in (a,b)$ the map T is C^1 with $T' > 1$ on $[a,c]$ and C^1 with $T' < -1$ on $[c,b]$, and we have $T(a) = T(b) = a$ and $T(c) = b$.

Then T is regular expanding; it has 2 sheets and is $(\min|T'|)^{-1}$-expanding, and its associated ICS F is in fact an IFS (the linear order on $[a,b]$ enables one to define global sections). For all $x \neq b$, $F(x)$ has two distinct elements while $F(b) = \{c,c\}$.

More examples of this kind are provided by letting $T(x)$ zig-zag between a and b more than once, or by considering higher-dimensional analogues, such as the following triangle foldings.

Example 2.11. — Let Ω be a simplex in $\mathbb{R}^d$ which is subdivided into a tiling of smaller simplices. Consider a map φ defined on the vertices of this simplicial decomposition, with values in the set of vertices of Ω, and not mapping two adjacent vertices to the same vertex. Define a map $T : \Omega \to \Omega$ by extending affinely the map φ over each subsimplex. If all of these affine maps are dilating (e.g. if the subsimplices are all small enough), then T is a regular expanding map which has as many sheets as there are simplices in the decomposition.

An explicit example is given by a right-angled isocele triangle, which is folded along the altitude issued from the right-angled vertex and then rotated and dilated into the original triangle.

Just like the piecewise expanding unimodal maps above, all these examples can be considered both as IFS and expanding maps.

Example 2.12. — Let $F_1, \dots F_k : \Omega \to \Omega$ be an IFS on some compact space Ω, assume the strong separation property (i.e. the $F_i(\Omega)$ are pairwise disjoints) and up to restriction, assume Ω is the attractor (i.e. $\Omega = F_1(\Omega) \cup \cdots \cup F_k(\Omega)$). Define on Ω the map T that sends $x \in F_i(\Omega)$ to $F_i^{-1}(x)$. Then T is obviously a regular expanding map.

When an IFS does not have the strong separation property, we do not usually get a well-defined expanding map. This is not a big issue since our real focus here is on the random *backward* orbits, which are well-defined for all IFS even when they have big overlaps.

Example 2.13. — Let $\Omega = \{1, \dots, k\}^{\mathbb{N}}$ endowed with the metric

$$d_\theta(x, y) = \theta^{i(x,y)}$$

where $x = (x_j)_j, y = (y_j)$ and $i(x, y) = \min\{j \in \mathbb{N} \mid x_j \neq y_j\}$ for any fixed $\theta < 1$. The shift map $\sigma : \Omega \to \Omega$ is then obviously a regular expanding map, with k sheets and expanding ratio $\frac{1}{\theta}$.

The present framework does not cover subshifts of finite type, first because we assume a bijection between $F(x)$ and $F(y)$ for all x, y (but it might be possible to use the multiset approach to solve this issue), second because we ask a bijection $(x_i, y_i)_i$ between $F(x)$ and $F(y)$ that pairs only close elements together. It might be possible to extend the proof of the contraction property below to the case when the *average* distance between x_i and y_i is small, but at best at the cost of some technical complication.

2.4. Iterates of the transfer operator. — We will need to consider iterates of the transfer operator, so let us fix some notation and prove a useful estimate, to be used several times below.

Assume that F is an iterated contraction system and $A : \Omega \to \mathbb{R}$ is Lipschitz. For each $x \in \Omega$ consider the following multiset $\bar{F}^t(x)$ of *admissible sequences* with respect to F, of length t and starting at x: $\bar{F}^t(x)$ contains each sequence $s = (x_0 = x, x_1, x_2, \dots, x_t)$ with $x_{n+1} \in F(x_n)$ for all $0 < n < t$. Furthermore, the sequence $(x_0 = x, x_1, x_2, \dots, x_t)$ occurs with multiplicity given by the product of the multiplicities of x_{n+1} in $F(x_n)$, for $0 < n < t$. This multiset is in a natural bijection with $F^t(x)$, but refines it by identifying the orbits followed from x to each of the elements of $F^t(x)$.

Then for each admissible sequence $s = (x, x_1, \dots, x_t)$ of length t, we define

$$A^t(s) := \sum_{n=1}^{t} A(x_n)$$

so that, for $\varphi : \Omega \to \mathbb{R}$ continuous,

$$\mathscr{L}_A^t \varphi(x) = \sum_{s=(x,x_1 x_2,\dots,x_t) \in \bar{F}^t(x)} e^{A^t(s)} \varphi(x_t).$$

By definition of an ICS, for all x and y there is a bijection between $\bar{F}^t(x)$ and $\bar{F}^t(y)$ such that for all admissible $s = (x, x_1, x_2, \dots, x_t)$, the corresponding $r = (y, y_1, y_2, \dots, y_t)$ satisfies $d(x_n, y_n) \leq \theta^n d(x, y)$ for all n. As A is Lipschitz, we hence have that

$$|A^t(s) - A^t(r)| = \Big|\sum_{n=1}^{t} A(x_n) - \sum_{n=1}^{t} A(y_n)\Big| \le \sum_{n=1}^{t} \mathrm{Lip}(A) d(x_n, y_n)$$
$$\le \mathrm{Lip}(A) \sum_{n=1}^{t} \theta^n d(x,y) \le \frac{\mathrm{Lip}(A)}{1-\theta} d(x,y)$$

For all t, all x, y, and all appropriately paired $s = (x, x_1, \dots, x_t) \in \bar{F}^t(x)$ and $r = (y, y_1, \dots, y_t) \in \bar{F}^t(y)$ we therefore have

$$e^{A^t(s) - A^t(r)} \le e^{Md(x,y)}, \tag{1}$$

where $M = \mathrm{Lip}(A)(1-\theta)^{-1}$.

3. Normalized potentials and operators

For a given Lipschitz continuous potential A and an ICS F, we now construct an $\mathscr{L}_{F;A}$-invariant function. Recall that the spectral radius of $\mathscr{L}_{F;A}$ acting on the space of continuous functions $C(\Omega)$ with respect to the norm $\|f\|_\infty := \sup_{x\in\Omega} |f(x)|$, is

$$\rho = \lim_{n\to\infty} \left(\sup_{f \in C(\Omega), f\neq 0} \frac{\|\mathscr{L}^n(f)\|_\infty}{\|f\|_\infty} \right)^{\frac{1}{n}}$$

Proposition 3.1. — *Assume that F is an iterated contraction system and $A : \Omega \to \mathbb{R}$ is Lipschitz. Then there exists a strictly positive, Lipschitz continuous function h such that $\mathscr{L}(h) = \rho h$.*

Proof. — We begin with the construction of ρ. Note that by compactness of Ω, A is bounded from above and below. In particular, for $t \in \mathbb{N}$,

$$k^t e^{t \min_{x\in\Omega} A(x)} \le \mathscr{L}^t(1)(x) \le k^t e^{t \max_{x\in\Omega} A(x)}$$

for all $x \in \Omega$. Hence, for a fixed $x_0 \in \Omega$,

$$\tilde{\rho} := \limsup_{t\to\infty} (\mathscr{L}^t(1)(x_0))^{1/t}$$

is bounded away from 0 and ∞. Note that we immediately have $\tilde{\rho} \le \rho$, but we will get equality later.

Now, fix a bijection $(s^i, r^i)_{1\le i\le k^t}$ as above between $\bar{F}^t(x)$ and $\bar{F}^t(y)$. Then

$$\left|\mathscr{L}^t(1)(x) - \mathscr{L}^t(1)(y)\right| \le \sum_i \left|e^{A^t(s_i)} - e^{A^t(r_i)}\right| \le \sum_i e^{A^t(s_i)} \left|1 - e^{A^t(r_i) - A^n(s_i)}\right|$$
$$\le \left|e^{Md(x,y)} - 1\right| \mathscr{L}^n(1)(x) \le \tilde{M} \mathscr{L}^n(1)(x) d(x,y), \tag{2}$$

with $\tilde{M} = (\exp(M\operatorname{diam}(\Omega)) - 1)/\operatorname{diam}(\Omega)$. This estimate has several important consequences. First of all, as the diameter of Ω is bounded, it follows that

$$\sup\{\mathscr{L}^n(1)(x)/\mathscr{L}^n(1)(y) : x, y \in \Omega, n \in \mathbb{N}\} < \infty, \tag{3}$$

which implies that $\tilde{\rho}$ does not depend on the choice of x_0; in particular, $\tilde{\rho} = \rho$.

Hence, the radius of convergence of the power series

$$\sum_{n=1}^{\infty} s^n \mathscr{L}^n(1)(x)$$

is equal to $1/\rho$ for all $x \in \Omega$. Moreover, following Denker and Urbanski ([**DU91**]), there exists a sequence (a_n) with $a_1 = 1$, $a_{n+1} \geq a_n$ and $\frac{a_{n+1}}{a_n} \to 1$ such that

$$\sum_{n=1}^{\infty} a_n s^n \mathscr{L}^n(1)(x) \begin{cases} = \infty & s \geq 1/\rho \\ < \infty & s < 1/\rho. \end{cases}$$

Note that (a_n) might be chosen independently from $x \in \Omega$ by (3). For $0 < s < 1/\rho$, define

$$h_s(x) := \frac{\sum_{n=1}^{\infty} a_n s^n \mathscr{L}^n(1)(x)}{\sum_{n=1}^{\infty} a_n s^n \mathscr{L}^n(1)(x_0)}.$$

It follows from (3) that $\|h_s\|_\infty$ is uniformly bounded, and from (2) that $|h_s(x) - h_s(y)| \leq \tilde{M} h_s(x) d(x,y)$. Hence, by Arzéla-Ascoli, there exists a sequence (s_m) with $s_m \nearrow 1/\rho$ and a Lipschitz function h such that $\lim_m \|h_{s_m} - h\|_\infty = 0$ and $|h(x) - h(y)| \leq \tilde{M} h(x) d(x,y)$.

We now exploit the divergence in order to show that $\mathscr{L}(h) = \rho h$. Let $\varepsilon > 0$ and choose N_ε such that $|a_{n-1}/a_n - 1| < \varepsilon$ for all $n > N_\varepsilon$. Set $Q(s) := \sum_{n=1}^{\infty} a_n s^n \mathscr{L}^n(1)(x_0)$. We then have by divergence of $Q(s)$ that

$$\begin{aligned} |\mathscr{L}(h)(x) - \rho h(x)| &\leq \lim_{m\to\infty} \frac{1}{Q(s_m)} \left| \sum_{n=2}^{\infty} (a_{n-1} s_m^{n-1} - a_n s_m^n \rho) \mathscr{L}^n(1)(x) \right| \\ &= \lim_{m\to\infty} \frac{\rho}{Q(s_m)} \left| \sum_{n=N_\varepsilon}^{\infty} \left(\frac{a_{n-1}}{\rho a_n s_m} - 1 \right) a_n s_m^n \mathscr{L}^n(1)(x) \right| \\ &\leq \rho h(x) \sup_{n \geq N_\varepsilon} \lim_{m\to\infty} \left| \frac{a_{n-1}}{\rho a_n s_m} - 1 \right| \leq \varepsilon \rho h(x). \end{aligned}$$

Hence, $\mathscr{L}(h) = \rho h$. □

We now employ the above proposition in order to associate a Markov chain and a corresponding Markov operator to a given ICS F and a potential A.

Definition 3.2. — The *Markov chain associated to* the Lipschitz potential A is defined by letting $m \cdot e^A(y)h(y)/\rho h(x)$ be the transition probability from x to y whenever y is an element of $F(x)$ of multiplicity m, where ρ and h are as in proposition 3.1.

We denote by $\mathbb{P}^*_{F,A,h}$ (leaving again aside any subscripts that are clear from the context) the operator on finite, signed measures, defined by

$$\int \varphi(x)\,\mathrm{d}(\mathbb{P}^*\mu)(x) = \int \sum_{y\in F(x)} e^{A(y)}\frac{h(y)}{\rho h(x)}\varphi(y)\,\mathrm{d}\mu(x)$$

whenever φ is a continuous test function. In other words, $\mathbb{P}^*$ is the dual of the operator defined by

$$\mathbb{P}\varphi(x) = \sum_{y\in F(x)} e^{A(y)}\frac{h(y)}{\rho h(x)}\varphi(y) = \frac{\mathscr{L}(h\varphi)(x)}{\rho h(x)}.$$

We refer to $\mathbb{P}$ and $\mathbb{P}^*$ as the *normalized operators* with respect to A and h. As above, since $\mathbb{P}(1) = 1$, the dual $\mathbb{P}^*$ leaves invariant the subspace of probability measures.

As a preparation for the the proofs below, we now analyze the regularity of the iterates of $\mathbb{P}$. For $s(x, x_1, \dots, x_t) \in \bar{F}_t$ as defined above, set

$$A^t_h(s) = A^t(s) + \log h(x_t) - \log h(x) - n\log\rho.$$

As it easily can be seen, we then have that

$$\mathbb{P}^t\varphi(x) = \sum_{s=(x,\dots x_t),s\in\bar{F}^t(x)} e^{A^t_h(s)}\varphi(x_t).$$

Furthermore, for $r, s \in \bar{F}_t$ appropriately paired with $r = (y, y_1, \dots, y_t)$, it follows that

$$\begin{aligned} e^{A^t_h(s)-A^t_h(r)} &= e^{A^t(s)-A^t(r)}\frac{h(x_t)}{h(y_t)}\frac{h(y)}{h(x)} \\ &\leq e^{Md(x,y)}\left(1+\tilde{M}\theta^t d(x,y)\right)\left(1+\tilde{M}d(x,y)\right) \\ &\leq e^{(M+2\tilde{M})d(x,y)} = e^{M'd(x,y)}, \end{aligned} \tag{4}$$

where $M' = M + 2\tilde{M}$.

4. Optimal transport and Wasserstein metric

Let us briefly introduce the definition of the 1-Wasserstein metric (the only one that we will use here) and recall some of its basic properties.

Let Ω be a compact metric space. The 1-Wasserstein distance is defined on the set $\mathscr{P}(\Omega)$ of (Borel) probability measures on Ω by

$$W_1(\mu,\nu) = \inf_{\pi\in\Gamma(\mu,\nu)} \int_{\Omega\times\Omega} d(x,y)\,\mathrm{d}\pi(x,y)$$

where $\Gamma(\mu,\nu)$ is the set of measures on $\Omega\times\Omega$ whose marginals are μ and ν. Elements of $\Gamma(\mu,\nu)$ are called *transport plans* from μ to ν or *couplings*.

Let us quote a few basic properties: W_1 is indeed a metric; the infimum in its definition is always attained by some transport plan, then called optimal and generally not unique; the topology induced by W_1 is the weak-* topology (this is only true because Ω is compact).

Whenever it is needed, we will write W_1^d to stress the underlying metric d; when no confusion is expected, we will simply use the same decoration on the distance and the Wasserstein distance (e.g. W_1' will denote the Wasserstein distances with respect to a metric d'). Note that the definition of W_1 extends to all pair of positive measures having the same total mass.

There are many ways to design good couplings, let us give a simple but useful technical result.

Proposition 4.1. — *Assume that there are sets $A_1,\dots A_n$ such that the probability measures μ and ν are concentrated on the union of the A_i. Let $c = \max_i \operatorname{diam}(A_i)$, $C = \operatorname{diam}(\cup A_i)$ and $m = \sum_i \min(\mu(A_i),\nu(A_i))$. Then*

$$W_1(\mu,\nu) \le mc + (1-m)C.$$

Proof. — We let π be a coupling of μ and ν that moves a mass at most m between different A_i's, i.e. such that

$$\pi(\{(x,y)|\exists i \text{ such that both } x,y\in A_i\}) \ge m.$$

Once this transport plan is constructed, we compute

$$\begin{aligned}\int_{\Omega\times\Omega} d(x,y)\,\mathrm{d}\pi(x,y) &= \int_{\cup_i A_i\times A_i} d(x,y)\,\mathrm{d}\pi(x,y) + \int_{\Omega\setminus\cup_i A_i\times A_i} d(x,y)\,\mathrm{d}\pi(x,y)\\ &\le mc + (1-m)C.\end{aligned}$$

To construct π, we first note that it is possible to decompose μ into

$$\mu = \sum_i (\mu_i^{\mathrm{in}} + \mu_i^{\mathrm{out}})$$

where the $\mu_i^{\mathrm{in/out}}$ are concentrated on A_i and $\mu_i^{\mathrm{in}}(A_i) = \min(\mu(A_i),\nu(A_i))$ (and similarly for ν). Then we set

$$\pi = \sum_i \mu_i^{\mathrm{in}}\otimes\nu_i^{\mathrm{in}} + \big(\sum_i \mu_i^{\mathrm{out}}\big)\otimes\big(\sum_i \nu_i^{\mathrm{out}}\big).$$

□

The following proposition is also more or less folklore and very useful; it appears for example in a proof in [**HM08**].

Proposition 4.2. — *Let P be a linear operator on the set of measures on Ω (assumed to be compact for simplification), such that P is continuous in the weak-* topology and maps probability measures to probability measures.*

If for some $C > 0$ and all x, y in some dense subset of Ω we have

$$W_1(P(\delta_x), P(\delta_y)) \leq Cd(x, y)$$

then for all $\mu, \nu \in \mathscr{P}(\Omega)$ we also have

$$W_1(P(\mu), P(\nu)) \leq CW_1(\mu, \nu).$$

Proof. — Let us give a slight variation of the Hairer-Mattingly proof, using density of finitely supported measures: we only have to prove $W_1(P(\mu), P(\nu)) \leq CW_1(\mu, \nu)$ when $\mu = \sum_{i\in I} a_i \delta_{x_i}$ and $\nu = \sum_{j\in J} b_j \delta_{y_j}$ and x_i, y_j are in the dense subset of Ω we are given. Let

$$\tilde{\pi} = \sum_{i,j} c_{i,j} \delta_{(x_i, y_j)}$$

be an optimal transport plan from μ to ν, and for each (i, j), let $\pi_{i,j}$ be an optimal transport plan from $P(\delta_{x_i})$ to $P(\delta_{y_j})$.

Define $\pi = \sum_{i,j} c_{i,j} \pi_{i,j}$; it transports $P(\mu)$ to $P(\nu)$ and we have

$$\begin{aligned}
\int_{\Omega\times\Omega} d(x, y) \,\mathrm{d}\pi(x, y) &= \sum_{i,j} c_{i,j} \int d(x, y) \,\mathrm{d}\pi_{i,j}(x, y) \\
&= \sum_{i,j} c_{i,j} W_1\big(P(\delta_{x_i}), P(\delta_{y_j})\big) \\
&\leq C \sum_{i,j} c_{i,j} d(x_i, y_j) \\
&= CW_1(\mu, \nu)
\end{aligned}$$

proving the claim. □

5. Proof of the main result and first applications

We are now in position to prove the main theorem. Throughout this section, assume that F is a k-ICS with contraction ratio θ, A is a Lipschitz potential on the attractor Ω of F, and $\mathbb{P}$ and $\mathbb{P}^*$ are defined as in definition 3.2. For the reader's convenience, the statement of Theorem 1.1 is repeated.

***Theorem 5.1* (Contraction property).** — *The normalized operator $\mathbb{P}^*$ is exponentially contracting on probability measures: There exist constants $C =$*

$C(\mathrm{Lip}(A),\theta,\operatorname{diam}\Omega)$ *and* $\lambda=\lambda(\mathrm{Lip}(A),\theta,\operatorname{diam}\Omega)<1$ *such that for all* $n\in\mathbb{N}$ *and all* $\mu,\nu\in\mathscr{P}(\Omega)$ *we have*

$$W_1((\mathbb{P}^*)^n\mu,(\mathbb{P}^*)^n\nu)\leq C\lambda^n W_1(\mu,\nu).$$

Proof. — We use three reductions of the problem. First, it is sufficient to prove Theorem 5.1 for some iterate $(\mathbb{P}^*)^t$ of the dual of the normalized operator (using the continuity of the operator and the flexibility given by the constant C). Second, it is sufficient to prove it when Ω is endowed with any metric d' which is Lipschitz-equivalent to d_θ (again using the constant C to absorb the ratio between the two metrics); an important point is that we can choose the metric d' depending on A. Last, thanks to Proposition 4.2, we only need to prove it when μ and ν are Dirac measures.

So, it is sufficient to find $t\in\mathbb{N}$, a metric d' equivalent to d and a number $\lambda'\in(0,1)$ such that for all $x,y\in\Omega$ we have

$$W_1'((\mathbb{P}^*)^t\delta_x,(\mathbb{P}^*)^t\delta_y)\leq\lambda' d'(x,y)$$

where W_1' is the Wasserstein metric associated to the distance d'.

The principal idea is to apply Proposition 4.1; let us define

$$d'(x,y)=\begin{cases}\theta^{-N}d_\theta(x,y) & \text{if } d_\theta(x,y)\leq\theta^N\cdot\operatorname{diam}\Omega\\ \operatorname{diam}\Omega & \text{otherwise}\end{cases}$$

for some N to be specified later. This metric will make Proposition 4.1 more effective because it localizes the Wasserstein metric to some small scale (all displacements are now equivalent as soon as they are somewhat big).

Now fix a positive integer t. Moreover, for $x,y\in\Omega$, fix a bijection $(s^i,r^i)_{1\leq i\leq k^t}$ between $\bar{F}^t(x)$ and $\bar{F}^t(y)$ as in Section 2.4 and apply a slight variant of Proposition 4.1: let π refer to a transport plan from

$$(\mathbb{P}^*)^t\delta_x=\sum_i e^{A_h^t(s^i)}\delta_{x_t}$$

to

$$(\mathbb{P}^*)^t\delta_y=\sum_i e^{A_h^t(r^i)}\delta_{y_t}$$

that moves a mass at least (cf. estimate (4))

$$\begin{aligned}m(x,y):=\sum_i\min(e^{A_h^t(s^i)},e^{A_h^t(r^i)})&\geq\sum_i e^{A_h^t(s^i)}e^{-M'd(x,y)}\\ =e^{-M'd(x,y)}\geq\ e^{-M'd(x,y)}\end{aligned}$$

by a distance at most $d'(x_t^i, y_t^i) \le \theta^{t-N} d'(x,y)$ and moves the rest of the mass by a distance at most $\operatorname{diam}\Omega$. We get

$$\begin{aligned} W_1'\Big((\mathbb{P}^*)^t\delta_x, (\mathbb{P}^*)^t\delta_y\Big) &\le e^{-M'd(x,y)}\theta^{t-N}d'(x,y) + (1-e^{-M'd(x,y)})\operatorname{diam}\Omega \\ &\le \theta^{t-N}d'(x,y) + (1-e^{-M'd(x,y)})\operatorname{diam}\Omega, \end{aligned}$$

which is at most

$$\begin{cases} (\theta^{t-N} + M'\cdot\operatorname{diam}\Omega\cdot\theta^N)d'(x,y) & \text{when } d'(x,y) < \operatorname{diam}\Omega \\ \big(\theta^{t-N} + 1 - e^{-M'\operatorname{diam}\Omega}\big)\cdot\operatorname{diam}\Omega & \text{when } d'(x,y) = \operatorname{diam}\Omega \end{cases}$$

First note that the expressions above only depend on the parameters θ, $\operatorname{diam}\Omega$, $\operatorname{Lip}(A)$. Now, taking N large enough and then t large enough ensures that the right-hand-side is at most $\lambda' d'(x,y)$ for some uniform $\lambda' < 1$. □

If A already is a normalized potential, the constants in the above theorem can be determined rather explicitly. Namely, it is not difficult to see that one can take for example

$$C = \theta^{-N}\Big(\theta + \frac{M}{1-\theta}\operatorname{diam}\Omega\Big)^{2t}$$

(recall that $M = \operatorname{Lip}(A)(1-\theta)^{-1}$) and

$$\lambda = \Big(1 - \frac{1}{2}e^{-\frac{M}{1-\theta}\operatorname{diam}\Omega}\Big)^{\frac{1}{t}}$$

where N is the solution to

$$\theta^N\frac{M}{1-\theta}\operatorname{diam}\Omega = 1 - e^{-\frac{M}{1-\theta}\operatorname{diam}\Omega}$$

and t is such that

$$\theta^t \le \theta^{2N}\frac{M}{2(1-\theta)}\operatorname{diam}\Omega.$$

We will always assume $C \ge 1$ if needed, as it will prove convenient and that if the above expression where less than 1, we can always increase C and this would allow us to improve λ.

5.1. Proof of the existence of a spectral gap. — Through duality, it is now easy to prove Corollary 1.2 and deduce uniqueness of h.

Corollary 5.2 **(Spectral gap).** — *Let μ be the fixed point of $\mathbb{P}^*$ in $\mathscr{P}(\Omega)$ (i.e. the invariant Gibbs measure associated to F and A); for each Lipschitz function $\zeta : \Omega \to \mathbb{R}$ such that $\int \zeta \,\mathrm{d}\mu = 0$, we have*

$$\|\mathbb{P}^n\zeta\|_{\mathrm{Lip}} \le C_2(\zeta)\lambda^n$$

where $C_2(\zeta) = (1 + \operatorname{diam}\Omega) C \operatorname{Lip}(\zeta)$ and C, λ are the constants given by Theorem 5.1. In particular, the function h in proposition 3.1 is unique up to multiplication by constants.

Proof. — We first control the uniform norm of $\mathbb{P}^n\zeta$ (this is the part where we need ζ to have vanishing μ-average): for all $x \in \Omega$ we have

$$\begin{aligned}|\mathbb{P}^n\zeta(x)| &= \left|\int \mathbb{P}^n\zeta(y)\,\mathrm{d}\delta_x(y) - \int \zeta\,\mathrm{d}\mu\right| = \left|\int \zeta(y)\,\mathrm{d}\big(\mathbb{P}^{*n}\delta_x\big)(y) - \int \zeta\,\mathrm{d}\mu\right| \\ &\le \operatorname{Lip}(\zeta) W_1(\mathbb{P}^{*n}\delta_x, \mu) \le \operatorname{Lip}(\zeta)\cdot C\lambda^n W_1(\delta_x, \mu) \\ &\le C \operatorname{diam}\Omega \cdot \operatorname{Lip}(\zeta)\cdot\lambda^n.\end{aligned}$$

Next we control with the same kind of trick the Lipschitz constant of $\mathbb{P}^n\zeta$ (this part holds whatever the integral of ζ): for all x, y we have

$$\begin{aligned}|\mathbb{P}^n\zeta(x) - \mathbb{P}^n\zeta(y)| &= \left|\int \mathbb{P}^n\zeta\,\mathrm{d}\delta_x - \int \mathbb{P}^n\zeta\,\mathrm{d}\delta_y\right| \\ &= \left|\int \zeta\,\mathrm{d}\big(\mathbb{P}^{*n}\delta_x\big) - \int \zeta\,\mathrm{d}\big(\mathbb{P}^{*n}\delta_y\big)\right| \\ &\le \operatorname{Lip}(\zeta) W_1(\mathbb{P}^{*n}\delta_x, \mathbb{P}^{*n}\delta_y) \le \operatorname{Lip}(\zeta)\cdot C\lambda^n d(x,y).\end{aligned}$$

This also implies that $\mathbb{P}f = f$ if and only if f is a constant function. Hence, $\mathscr{L}(f) = f$ if and only if f is a multiple of h given by proposition 3.1. □

Observe that this result for example implies that an expression like

$$\sum_{n=0}^{\infty} \mathbb{P}^n\zeta$$

is a well-defined Lipschitz function whenever $\int \zeta\, h\mathrm{d}\mu = 0$. This expression moreover defines a bounded inverse to the operator $I - \mathbb{P}_{F,A}$ restricted to 0-average functions.

When F is induced by a map T, it is also classical to deduce an exponential decay of correlations from the spectral gap; however, in our general setting and given the way $\mathbb{P}$ is defined, we would need to extend to regular expanding maps the classical relation

$$\int f\circ T\cdot g\,\mathrm{d}\mu = \int f\cdot \mathbb{P}(g)\,\mathrm{d}\mu$$

(for all $f \in L^1(\mu)$ and g continuous). This is certainly doable, but needs to carefully handle measurable selections; to keep the present article relatively short, we prefer to postpone these details to a further study of ICS and regular expanding maps.

6. Stability of the Gibbs map

Unless otherwise specified, we assume throughout this section that the potentials are already normalized (i.e. $\mathscr{L} = \mathbb{P}$) in order to be able to give accessible proofs which reveal the interplay between coupling techniques and thermodynamic formalism. Moreover, this also allows to give relatively explicit controls on the associated constants.

6.1. General results. — In order prove that the map which sends an ICS F and a normalized potential A to the Gibbs measure $\mu_{F,A}$ is locally Lipschitz, we first need to prove the stability of the dual transfer operator.

The uniform norm $\|\cdot\|_\infty$ is defined as usual for potentials, and a similar distance is defined for ICS with the same number of terms defined on a common metric space X by:

$$d_\infty(F_1, F_2) = \sup_{x\in X} \inf_{(y_1^j, y_2^j)_j} \sup_j d(y_1^j, y_2^j)$$

where the infimum is taken over all bijections between the multisets $F_1(x)$ and $F_2(x)$. In other words, $d_\infty(F_1, F_2) \leq D$ exactly when for all x, it is possible to pair the elements of $F_1(x)$ and $F_2(x)$ such that no two paired elements are more than D apart.

Proposition 6.1. — *Let F_1, F_2 be two ICS with k terms defined on the same compact metric space X. Let A_1, A_2 be potentials defined on X which are assumed to be normalized with respect to F_1 and F_2 respectively. Let $\mathscr{L}_i = \mathscr{L}_{F_i,A_i}$ be the transfer operator defined by (F_i, A_i) on the set of continuous functions from X to $\mathbb{R}$. Then for any probability measure μ on X, we have*

$$W_1(\mathscr{L}_1^*\mu, \mathscr{L}_2^*\mu) \leq \operatorname{diam} X \cdot \|A_1 - A_2\|_\infty + (\operatorname{Lip}(A_2) \operatorname{diam} X + 1) d_\infty(F_1, F_2).$$

This inequality is not optimal from the proof below, but is good enough for small variations and easy to state. Note that by symmetry, $\operatorname{Lip}(A_2)$ can be replaced by $\operatorname{Lip}(A_1)$, the point being that we only need to control one of the Lipschitz constants.

Proof. — Reasoning as in the proof of Proposition 4.2, we see that it is sufficient to prove this inequality when $\mu = \delta_x$ is a Dirac mass. In this case, we have

$$\mathscr{L}_i^* \delta_x = \sum_{j=1}^k e^{A_i(y_i^j)} \delta_{y_i^j};$$

where $y_i^1, \dots y_i^k$ are the elements of $F_i(x)$, numbered such that $d(y_1^j, y_2^j) \leq d_\infty(F_1, F_2)$ for all j. There is a transport plan between these two measures that moves as much mass as possible from each of the y_1^j to y_2^j. This plan moves an amount of mass

$$m(x) := \sum_j \min(e^{A_1(y_1^j)}, e^{A_2(y_2^j)})$$

by a distance at most $d_\infty(F_1, F_2)$, and the rest of the mass is moved by at most $\operatorname{diam} X$.

We have for all j:

$$\begin{aligned} A_2(y_2^j) &\geq A_2(y_1^j) - \operatorname{Lip}(A_2) d(y_1^j, y_2^j) \\ &\geq A_1(y_1^j) - \|A_1 - A_2\|_\infty - \operatorname{Lip}(A_2) d_\infty(F_1, F_2) \end{aligned}$$

so that

$$e^{A_2(y_2^j)} \geq e^{A_1(y_1^j)} e^{-\|A_1 - A_2\|_\infty - \operatorname{Lip}(A_2) d_\infty(F_1,F_2)},$$

from which it comes (using the normalization $\sum e^{A_1(y_1^j)} = 1$)that

$$1 - m(x) \leq \|A_1 - A_2\|_\infty + \operatorname{Lip}(A_2) d_\infty(F_1, F_2).$$

We get that the plan under consideration has cost less than

$$m(x) d_\infty(F_1, F_2) + \operatorname{diam} X \left(\|A_1 - A_2\|_\infty + \operatorname{Lip}(A_2) d_\infty(F_1, F_2)\right)$$

and bounding $m(x)$ by 1 yields the claimed inequality. □

Combining this estimate with the contraction property, we obtain that the Gibbs measure depends on the ICS and the potential in a locally Lipschitz way.

Corollary 6.2. — *Let F_1, F_2 be two ICS with k terms defined on the same compact metric space X.*[(2)] *Let A_1, A_2 be potentials defined on X which are assumed to be normalized with respect to F_1 and F_2 respectively. Let μ_i be the Gibbs measure associated with (F_i, A_i), i.e. the unique probability measure invariant under $\mathscr{L}_i^* = \mathbb{P}_i^*$.*

If F_2 has contraction ratio θ then we have

$$W_1(\mu_1, \mu_2) \leq \frac{C}{1-\lambda} \big(\operatorname{diam} X \cdot \|A_1 - A_2\|_\infty + (\operatorname{Lip}(A_2) \operatorname{diam} X + 1) d_\infty(F_1, F_2) \big)$$

where C, λ are the constants given by Theorem 5.1 in terms of $\operatorname{diam} X$, *θ and* $\operatorname{Lip}(A_2)$.

Note that if we vary both pairs (F_i, A_i), we only get a *locally* Lipschitz control, as C and λ both get poor when $\operatorname{Lip}(A_2)$ goes to infinity, or θ goes to 1. But if we fix one of them, (F_2, A_2) say, then we get a globally uniform control of the distance between the Gibbs measures.

---

(2) with possibly different attractors Ω_1, Ω_2.

Proof. — Consider

$$u_n := \sup_{\mu \in \mathscr{P}(X)} W_1(\mathscr{L}_1^{*n}\mu, \mathscr{L}_2^{*n}\mu);$$

from the previous proposition we know that

$$u_1 \leq \operatorname{diam} X \cdot \|A_1 - A_2\|_\infty + (\operatorname{Lip}(A_2) \operatorname{diam} X + 1) d_\infty(F_1, F_2).$$

Given any probability measure μ on X, we have

$$\begin{aligned}
&W_1(\mathscr{L}_1^{*(n+1)}\mu, \mathscr{L}_2^{*(n+1)}\mu) \\
&\leq W_1\big(\mathscr{L}_1^{*n}(\mathscr{L}_1^*\mu), \mathscr{L}_2^{*n}(\mathscr{L}_1^*\mu)\big) + W_1\big(\mathscr{L}_2^{*n}(\mathscr{L}_1^*\mu), \mathscr{L}_2^{*n}(\mathscr{L}_2^*\mu)\big) \\
&\leq u_n + C\lambda^n W_1(\mathscr{L}_1^*\mu, \mathscr{L}_2^*\mu) \\
&\leq u_n + C\lambda^n u_1.
\end{aligned}$$

Then by induction on n we get

$$u_n \leq (C\lambda^{n-1} + \cdots + C\lambda^2 + C\lambda + 1)u_1 \leq \frac{C}{1-\lambda} u_1.$$

For any fixed probability μ, when n goes to ∞, we have $\mathscr{L}_i^{*n}\mu \to \mu_i$ so that we get

$$W_1(\mu_1, \mu_2) \leq \liminf u_n \leq \frac{C}{1-\lambda} u_1$$

as desired. □

We can now easily deduce the results announced in the introduction starting with the following.

Proof of Corollary 1.3. — We simply apply Corollary 6.2 to $F_1 = F_2 = F$ and A, B, getting:

$$W_1(\mu_A, \mu_B) \leq \frac{C}{1-\lambda} \operatorname{diam} \Omega \cdot \|A - B\|_\infty.$$

The consequence in term of test functions follows by duality. □

6.2. Application to expanding maps. — Let us now see how the above can be used to prove Corollaries 1.4 and 1.5 above for expanding maps with respect to normalized potentials.

Proof of Corollary 1.4. — Since A and B are normalized, the spectral radii of the $\mathscr{L}_A$ and $\mathscr{L}_B$ are equal to 1. Furthermore, μ_A and μ_B are equilibrium states (see, e.g., [**Wal78**]). Hence, $h(\mu_A) = -\int A \, \mathrm{d}\mu_A$ and $h(\mu_B) = -\int B \, \mathrm{d}\mu_B$.

Using the previous inequality we get:

$$\begin{aligned}|h(\mu_A) - h(\mu_B)| &\leq \Big| \int A \,\mathrm{d}\mu_A - \int A \,\mathrm{d}\mu_B \Big| + \int |A - B| \,\mathrm{d}\mu_B \\ &\leq \operatorname{Lip}(A) W_1(\mu_A, \mu_B) + \|A - B\|_\infty \\ &\leq \Big(\frac{C \operatorname{Lip}(A)}{1-\lambda} \operatorname{diam} \Omega + 1 \Big) \|A - B\|_\infty.\end{aligned}$$

□

To prove Corollary 1.5, we mainly have to show how the ICS F depends on the given expanding map T. This is the part where we restrict to C^1 expanding maps on manifolds.

Lemma 6.3. — *Let T_1, T_2 be C^1 expanding map on the same manifold Ω and assume that $\|T_1 - T_2\|_\infty \leq \frac{1}{4} \operatorname{sys}(\Omega)$. Then the ICS $F_i : x \mapsto T_i^{-1}(x)$ satisfy*

$$d_\infty(F_1, F_2) \leq 2 d_\infty(T_1, T_2).$$

Proof. — First, recall that both T_1 and T_2 are self-covering maps of Ω.

Let $x \in \Omega$ be any point, and let

$$\{x_1, \dots, x_k\} := T_1^{-1}(x) = F_1(x).$$

For all $j \in \{1, \dots, k\}$, let γ_j be a shortest geodesic from x to $T_2(x_j)$ and denote by γ_j^{-1} the same curve parametrized in the other direction; note that these curves have length at most $d_\infty(T_1, T_2)$. We construct a curve $\tilde{\gamma}_j$ in Ω as follows.

First, $\tilde{\gamma}_j^1$ is the lift of γ_j with respect to the covering map T_1 that starts at x_j. Its endpoint is mapped by T_1 to $T_2(x_1)$. Second, $\tilde{\gamma}_j^1$ is the lift of γ_j^{-1} with respect to the covering map T_2 that starts at the endpoint of $\tilde{\gamma}_j^1$; its endpoint is denoted by y_j and we have $T_2(y_j) = x$. Then $\tilde{\gamma}_j$ is the concatenation of $\tilde{\gamma}_j^1$ and $\tilde{\gamma}_j^2$.

By construction, $\tilde{\gamma}_j$ links $x_j \in F_1(x)$ to $y_j \in F_2(x)$ and, since the T_i are expanding, has length at most $2d_\infty(T_1, T_2)$. Our assumption on the distance between the T_i ensures that the y_j are pairwise distinct, so that $F_2(x) = \{y_1, \dots, y_k\}$; the conclusion then follows from the definition of the uniform distance between ICS. □

Proof of Corollary 1.5. — It is well-known (see, e.g., [**Wal78**]) that the maximal entropy measure of T_i is the Gibbs measure associated to the constant potential $A = -\log k$ where k is the number of sheets of T_i. We only have to apply Corollary 6.2 with $A_1 = A_2 = A$ (so that in particular $\operatorname{Lip}(A) = 0$), using the previous Lemma to control $d_\infty(F_1, F_2)$, to get the desired conclusion. □

References

[AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[Bal00] Viviane Baladi. *Positive transfer operators and decay of correlations*, volume 16 of *Advanced Series in Nonlinear Dynamics.* World Scientific Publishing Co., Inc., River Edge, NJ, 2000.

[Bal08] Viviane Baladi. Linear response despite critical points. *Nonlinearity*, 21(6):T81–T90, 2008.

[BCV12] Thiago Bomfim, Armando Castro, and Paulo Varandas. Differentiability of thermodynamical quantities in non-uniformly expanding dynamics, 2012. arXiv:1205.5361v3.

[BFG99] Xavier Bressaud, Roberto Fernández, and Antonio Galves. Decay of correlations for non-Hölderian dynamics. A coupling approach. *Electron. J. Probab.*, 4:no. 3, 19 pp. (electronic), 1999.

[BS08] Viviane Baladi and Daniel Smania. Linear response formula for piecewise expanding unimodal maps. *Nonlinearity*, 21(4):677–711, 2008.

[DF37] Wolfgang Doeblin and Robert Fortet. Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France*, 65:132–148, 1937.

[DU91] Manfred Denker and Mariusz Urbański. On the existence of conformal measures. *Trans. Amer. Math. Soc.*, 328(2):563–587, 1991.

[Gig11] Nicola Gigli. *Introduction to optimal transport: theory and applications.* Publicações Matemáticas do IMPA. [IMPA Mathematical Publications]. Instituto Nacional de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 2011. 28o Colóquio Brasileiro de Matemática. [28th Brazilian Mathematics Colloquium].

[Gro81] Mikhael Gromov. *Structures métriques pour les variétés riemanniennes*, volume 1 of *Textes Mathématiques.* CEDIC, Paris, 1981. Edited by J. Lafontaine and P. Pansu.

[HM08] Martin Hairer and Jonathan C. Mattingly. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.*, 36(6):2050–2091, 2008.

[HM10] Martin Hairer and Andrew J. Majda. A simple framework to justify linear response theory. *Nonlinearity*, 23(4):909–922, 2010.

[MU03] R. Daniel Mauldin and Mariusz Urbański. *Graph directed Markov systems*, volume 148 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 2003. Geometry and dynamics of limit sets.

[Oll09] Yann Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.

[PP90] William Parry and Mark Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque*, (187-188):268, 1990.

[Rue98] David Ruelle. General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium. *Phys. Lett. A*, 245(3-4):220–224, 1998.

[Sta13] Manuel Stadlbauer. Coupling methods for random topological markov chains, 2013. arxiv:1312.6033.

[Vil03] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

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

[Wal78] Peter Walters. Invariant measures and equilibrium states for some mappings which expand distances. *Trans. Amer. Math. Soc.*, 236:121–153, 1978.

Benoît R. Kloeckner
Artur O. Lopes
Manuel Stadlbauer