



HAL
open science

Optimal transportation and stationary measures for Randomly Iterated Function Systems

Benoît Kloeckner

► **To cite this version:**

Benoît Kloeckner. Optimal transportation and stationary measures for Randomly Iterated Function Systems. 2019. hal-02276750v1

HAL Id: hal-02276750

<https://hal.science/hal-02276750v1>

Preprint submitted on 3 Sep 2019 (v1), last revised 27 May 2021 (v3)

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

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

Optimal transportation and stationary measures for Randomly Iterated Function Systems

Benoît R. Kloeckner *

September 3, 2019

In this article, we show how ideas, methods and results from optimal transportation can be used to study various aspects of the stationary measures of Iterated Function Systems equipped with a probability distribution. We recover a classical existence and uniqueness result under a contraction-on-average assumption, prove moment bounds and generalized moment bounds, consider the convergence of the empirical measure of an associated Markov chain, prove in many cases the Lipschitz continuity of the stationary measure when the system is perturbed, with as a consequence a “linear response formula” at almost every parameter of the perturbation, and prove singularity of the stationary measure in some cases where the classical dimension bound coincides with the dimension of the ambient space.

1 Introduction

Let (X, d) be a complete separable metric space (endowed with its Borel σ -algebra for all measurability purposes) and $\Phi = \{\phi_i : i \in I\}$ be an *Iterated Functions System* (IFS), i.e. a family of continuous maps $\phi_i : X \rightarrow X$ indexed by a set I , either countable or endowed with a standard σ -algebra. The set of probability measures on X is denoted by $\mathcal{P}(X)$.

Hutchinson [Hut81] introduced such IFS to produce fractals: under a contraction hypothesis, there is a unique compact subset K_Φ of X such that

$$K_\Phi = \bigcup_{i \in I} \phi_i(K_\Phi).$$

*Université Paris-Est, Laboratoire d'Analyse et de Matématiques Appliquées (UMR 8050), UPEM, UPEC, CNRS, F-94010, Créteil, France

The proof is very simple: one shows that the map $K \mapsto \cup_i \phi_i(K)$ is a contraction in the Hausdorff metric, and applies the Banach fixed point theorem. Given additionally $\eta \in \mathcal{P}(I)$, one is interested in existence, uniqueness and properties of a measure $\mu \in \mathcal{P}(X)$ such that

$$\mu = \int (\phi_i)_* \mu \, d\eta(i) \quad (1)$$

i.e. $\int f(x) \, d\mu(x) = \int \int f \circ \phi_i(x) \, d\mu(x) \, d\eta(i)$ for all $f \in C_b(X)$, the set of bounded continuous functions $X \rightarrow \mathbb{R}$. Such a measure shall be called a *stationary* measure for the pair (Φ, η) , which we shall call a *Randomly Iterated Functions System* (RIFS). When the ϕ_i are contractions, again existence and uniqueness mostly follow from the Banach fixed point theorem; Hutchinson used the now-called Wasserstein distance of exponent 1 (in its dual formulation, restricted to compactly supported measures). Afterward, the contraction hypothesis has been quite relaxed to a form of contraction on average. The now most common technique is to use backward iteration of the system, which to an infinite word $\omega = i_1 \dots i_k \dots$ with letters in I and a point $x \in X$ associate the points $x_k = \phi_{i_1} \circ \phi_{i_2} \circ \dots \circ \phi_{i_k}(x)$. One shows (under appropriate hypotheses) that if the i_k are drawn randomly independently with law η , then x_k converges almost surely to a random point not depending on x , whose law is the stationary measure. As we shall see, using general Wasserstein distances one can use the fixed point theorem approach and get moment estimates at the same time.

More generally, the goal of the present article is to apply tools and ideas from optimal transportation in this context, to show the variety of information they provide with simple (while not always elementary) proofs. General results are stated and proved all along the article, but to give a taste of the main results we start with their applications to specific examples.

1.1 Tail estimates for a parabolic-hyperbolic example

For $p, a \in (0, 1)$, consider the RIFS given by

$$\begin{aligned} I &= \{0, 1\}, & \eta(\{0\}) &= p & \eta(\{1\}) &= 1 - p \\ \phi_0(x) &= ax & \phi_1(x) &= x + 1 & \forall x \in \mathbb{R} \end{aligned} \quad (2)$$

When ϕ_0 and ϕ_1 are seen as Möbius transformations (i.e. extended as homography of the real projective line, with ∞ as a fixed point) or as hyperbolic isometries (i.e. extended to the Poincaré upper half plane of \mathbb{C} with its hyperbolic metric), ϕ_0 is *hyperbolic* (one attractive and one repulsive fixed points on the projective line) while ϕ_1 is *parabolic* (a single fixed point on the projective line, which is repulsive on one side and attractive on the other side). This example is interesting in particular because it is not uniformly contracting, and cannot be made so in any set of coordinates because of the parabolic fixed point at infinity. Let $\mu_{a,p}$ denote the unique stationary measure of this RIFS; it is concentrated on $[0, +\infty)$. Its dimension and absolute continuity have for example been studied in [NSB02]. As an application of the generalized moment estimate obtained from the use of Wasserstein distances, we obtain an exponential tail estimate for $\mu_{a,p}$, sharp up to a polynomial factor.

Corollary A. For all $p, a \in (0, 1)$, there exist $c, C > 0$ such that for all $t \geq 1$:

$$c(1-p)^t \leq \mu_{a,p}([t, +\infty)) \leq Ct^{\frac{1}{1-a}}(1-p)^t.$$

(The main result is the upper bound, the lower bound being trivial.)

1.2 Linear response for Bernoulli convolutions

The *Bernoulli convolution* μ_λ (where $\lambda \in (0, 1)$) is defined as the stationary measure of the following most classical RIFS (Φ^λ, η) :

$$\begin{aligned} I &= \{0, 1\}, & \eta(\{0\}) &= \eta(\{1\}) = \frac{1}{2} \\ \phi_0^\lambda(x) &= \lambda x, & \phi_1^\lambda(x) &= \lambda x + (1 - \lambda) \quad \forall x \in \mathbb{R}. \end{aligned}$$

(The precise value $1 - \lambda$ of the translation part in ϕ_1^λ has no particular relevance –as soon as it is not zero– we chose this value to have a fixed attractor $[0, 1]$ but it bears no consequences on the result below.)

We shall prove that the map $\lambda \mapsto \mu_\lambda$ is Lipschitz in the Wasserstein distances of all exponents; for exponents larger than 1, thanks to the differentiation theorem of Ambrosio, Gigli and Savaré [AGS08] this implies an almost-everywhere *linear response formula*¹: the map $\lambda \mapsto \mu_\lambda$ can be differentiated in some precise sense at almost-all λ , and while we do not get an explicit expression for the differential we show that it takes a specific form.

Corollary B. The family of Bernoulli convolutions $(\mu_\lambda)_{\lambda \in (\frac{1}{2}, 1)}$ is differentiable almost everywhere. More specifically, there exist a family $(w_\lambda)_{\lambda \in (\frac{1}{2}, 1)}$ of measurable functions from $[0, 1]$ to \mathbb{R} such that for Lebesgue-almost all $\lambda \in (\frac{1}{2}, 1)$ and for all \mathcal{C}^1 compactly supported $f : \mathbb{R} \rightarrow \mathbb{R}$,

$$\frac{d}{dt} \int f d\mu_t \Big|_{t=\lambda} = \int_0^1 f'(x) w_\lambda(x) dx.$$

Moreover, for Lebesgue-almost all $\lambda > 1/\sqrt{2}$ we have $w_\lambda \in L^q([0, 1])$ for all $q > 1$.

In fact, we also have differentiability for almost-all $\lambda \in (0, \frac{1}{2})$, but to get the expression above we use that μ_λ is absolutely continuous for almost all $\lambda > \frac{1}{2}$, see below.

1.3 Singularity of stationary measures via approximate overlaps

The most important question about Bernoulli convolutions is to determine for which λ is μ_λ absolutely continuous (and also for which λ it has Hausdorff dimension 1); we recall a few facts and refer the reader to [PSS00] for more details on these. Important recent works on this topic include [Hoc14, Shm19, Var19b, Var19a].

It is easy to see that for $\lambda < \frac{1}{2}$, μ_λ is supported on a compact Cantor set of dimension < 1 ; that $\mu_{\frac{1}{2}}$ is the uniform measure on $[0, 1]$, and that the support of μ_λ is $[0, 1]$ for

¹This terminology has been coined in dynamical system, see e.g. [Rue98, Rue09, BS12].

all $\lambda \geq \frac{1}{2}$. Erdős proved [Erd39] that when $\lambda \in (\frac{1}{2}, 1)$ is the inverse of a *Pisot number*, then μ_λ is singular with respect to Lebesgue; while Solomyak [Sol95] proved that for Lebesgue-almost all $\lambda \in (\frac{1}{2}, 1)$, μ_λ is absolutely continuous. Moreover, for almost all $\lambda \in (1/\sqrt{2}, 1)$, the density of μ_λ is continuous (see (8.1) in [PSS00]).

While Erős used Fourier analysis to prove its singularity result, another common method is to use a dimension upper bound given by the ratio of the entropy of η and the opposite to the Lyapunov exponent of (Φ, η) . Clearly, if the dimension of a measure on \mathbb{R}^d is less than d , then the measure cannot be absolutely continuous. Except in the simplest cases, to use the dimension bound effectively one relies on *exact overlaps*, i.e. equality between two compositions of maps of Φ with different indices (for example in the above example $\phi_1^\lambda \phi_0^\lambda \phi_0^\lambda = \phi_0^\lambda \phi_1^\lambda \phi_1^\lambda$ when $\lambda = \frac{\sqrt{5}-1}{2}$, the inverse of the golden ratio). We show that one can use *approximate overlaps*, if not to improve the dimension bound, at least to show in some critical cases that the stationary measure is not absolutely continuous.

Corollary C. *Let $\ell \geq 3$ be an integer. There exist a set $T = T(\ell) \subset \mathbb{R}$ with 2^{\aleph_0} elements such that for all $t_2 \in T$, for all $(t_3, \dots, t_{\ell-1}) \in \mathbb{R}^{\ell-3}$, the stationary measure μ of the RIFS*

$$I = \{0, 1, \dots, \ell - 1\}, \quad \eta(\{i\}) = \frac{1}{\ell} \quad \forall i \in I$$

$$\phi_0(x) = \frac{x}{\ell}, \quad \phi_1(x) = \frac{x}{\ell} + 1, \quad \phi_i(x) = \frac{x}{\ell} + t_i \quad \forall i \geq 2$$

is not absolutely continuous with respect to Lebesgue measure.

Note that the dimension bound alluded to above only yields here the useless inequality $\dim \mu \leq 1$ (the entropy is $\log \ell$ and the Lyapunov exponent $\log 1/\ell$): we are in the *critical* case. What we do is to choose T to ensure many approximate overlaps and use them to show that μ can be approximated by discrete measures slightly too well for it to be absolutely continuous. The same method can be used in higher dimension as well, see Theorem 5.2.

In the case $\ell = 3$, this result does not bring new information since Kenyon proved that for all irrational t_2 , the support of μ has zero Lebesgue measure [Ken97].

Structure of the article. Section 2 is a very brief introduction to optimal transportation and the Wasserstein distances. In Section 3, we start by using the completeness of Wasserstein spaces to prove existence and uniqueness of stationary measures under a hypothesis of contraction on average; this is a classical result, but we obtain moment estimates for free. Then we deal with generalized moments (Section 3.3, containing the proof of Corollary A), study the convergence toward the stationary measure of the empirical measure of the random walks associated to the RIFS (Section 3.5), and consider stationary measures in the more general setting of skew-products (Section 3.6).

In section 4 we study how much the stationary measure depends on the underlying RIFS, leading to the general linear response formula and the proof of Corollary C.

In Section 5, we explain how the problem of *quantization* of measures is related to the classical dimension bound, and give a criterion for singularity of a stationary measure leading to Corollary C.

2 Notation and definition of Wasserstein distances

We sometimes use $f \lesssim g$ as an alternative to Landau's notation $f = O(g)$, i.e. to express that for some C and all x , $f(x) \leq Cg(x)$. Sometimes we use C to denote such constants, and its value can change from line to line.

Let us now introduce briefly the Wasserstein distances, issued from optimal transportation theory. We only mention here statements that will be used several times or are relevant to several parts of the text. Below, it will happen several times that we use a result of the literature in an crucial way without giving its full statement; we shall only do so when we can give a precise reference, use the result as it is stated without modification, and when restating it would be somewhat redundant with the corollary we get from it. This makes the present article not as self-contained as it could be, but is consistent with the purpose of showing what optimal transportation can bring to the subject and encourage the reader to learn more about it. For details and proofs of the claims made in this section, see for example [Vil09].

Let us fix a reference point $x_0 \in X$. This choice can be arbitrary and has no conceptual bearing, but can be subject to optimization in some cases. For each $q \in (0, +\infty)$, the q -th moment of $\mu \in \mathcal{P}(X)$ is

$$m_{x_0}^q(\mu) := \int d(x, x_0)^q d\mu(x) \in [0, +\infty].$$

The set of probability measures μ of finite q -th moment is denoted by $\mathcal{P}_q(X)$ and does not depend on x_0 .

Given measures $\mu_0, \mu_1 \in \mathcal{P}(X)$, the set of *transport plans* or *couplings* is the set $\Gamma(\mu_0, \mu_1)$ of measures $\gamma \in \mathcal{P}(X \times X)$ such that $\gamma(A \times X) = \mu_0(A)$ and $\gamma(X \times B) = \mu_1(B)$ for all measurable $A, B \subset X$. The number $\gamma(A \times B)$ can be interpreted as the amount of mass moved from A to B under the plan γ .

One defines the *total cost* and *Wasserstein distance* of exponent q between two probability measures by:

$$\begin{aligned} C_q(\nu_0, \nu_1) &= \inf_{\gamma \in \Gamma(\nu_0, \nu_1)} \int d(x, y)^q d\gamma(x, y) \\ W_q(\nu_0, \nu_1) &= C_q(\nu_0, \nu_1)^{\min(1, \frac{1}{q})} \end{aligned}$$

Observe that the Wasserstein distance of exponent $q < 1$ is actually the Wasserstein distance of exponent 1 of (X, d^q) . The cost and the Wasserstein distance are finite as soon as $\mu_0, \mu_1 \in \mathcal{P}_q(X)$, and $(\mathcal{P}_q(X), W_q)$ is a complete metric space. Convergence in the Wasserstein distance is stronger than weak-* convergence when X is not compact; if X is compact, then W_q metrizes the weak-* topology.

Note that when $\mu_0 = \delta_x$ is a Dirac mass, there is only one possible coupling: $\Gamma(\delta_x, \mu_1) = \{\delta_x \otimes \mu_1\}$; therefore the q -th moment of μ can be expressed as $m_{x_0}^q(\mu) = C_q(\delta_{x_0}, \mu)$.

We denote $\text{Lip}(\phi)$ the Lipschitz constant of a map $\phi : X \rightarrow \mathbb{R}$, i.e.

$$\text{Lip}(\phi) = \inf \left\{ C \geq 0 : |\phi(x) - \phi(y)| \leq C d(x, y) \forall x, y \in X \right\} \in [0, +\infty].$$

The *Kantorovich duality* expresses that W_q coincides with the value of a “dual” optimization problem:

$$W_1(\mu_0, \mu_1) = \sup_{\text{Lip}(f) \leq 1} \left| \int f \, d\mu_0 - \int f \, d\mu_1 \right|.$$

Similarly, for all $q < 1$ the Wasserstein distance W_q can be expressed as the maximal differences between the integrals of q -Hölder functions with Hölder constant 1, since q -Hölder functions are precisely the Lipschitz function of the metric d^q .

3 Wasserstein contraction and its consequences

Under a suitable assumptions, one can prove that the *dual transfer operator* associated to an RIFS is contracting in some Wasserstein distance. The completeness of the Wasserstein spaces thus makes it easy to prove existence and uniqueness of a stationary measure in $\mathcal{P}_q(X)$ (some technicalities are needed to prove uniqueness on the whole of $\mathcal{P}(X)$; Hutchinson restricts to compactly supported stationary measures). This is not an original result, but this contraction has several other nice consequences: we have a control on moments of the stationary measure, a good convergence of the empirical measure, and later we shall also obtain a Lipschitz dependency of the stationary measure under perturbation of the RIFS.

3.1 Contracting dual operator

We consider the “dual transfer operator” L^* defined on $\mathcal{P}(X)$ by

$$L^* \mu = \int (\phi_i)_* \mu \, d\eta(i),$$

i.e. $L^* \mu$ is the law of x_{n+1} if $(x_n)_n$ is a Markov chain jumping from x to $\phi_i(x)$ with probability $d\eta(i)$ and $x_n \sim \mu$. A stationary measure is precisely a $\mu \in \mathcal{P}(X)$ such that $L^* \mu = \mu$.

The “transfer operator” (also known as Markov operator), acting on suitable spaces of functions $X \rightarrow \mathbb{R}$ is defined by

$$L f(x) = \int f \circ \phi_i(x) \, d\eta(x);$$

the transfer operator acts for example on the space of bounded measurable functions. It is a positive operator fixing each constant function, so that when $a \leq f \leq b$ with $a, b \in \mathbb{R}$, then also $a \leq L f \leq b$. The duality relation between L and L^* is

$$\int f \, d(L^* \mu) = \int L f \, d\mu$$

and is a direct consequence of Fubini's theorem.

The dual transfer operator has a natural extension to couplings, which we denote in the same way: given $\gamma \in \Gamma(\mu_0, \mu_1)$, we define

$$L^* \gamma = \int (\phi_i \times \phi_i)_* \gamma \, d\eta(i) \in \Gamma(L^* \mu_0, L^* \mu_1)$$

Our first main result is the following, where the control of the moments is our focus. The first hypothesis below asks for contraction *on L^q average*, while the second hypothesis asks the maps not to translate a point too far away.

Theorem 3.1. *Let (Φ, η) be a RIFS on a complete metric space (X, d) and fix any $x_0 \in X$. Assume that for some $q > 0$, $A > 0$, $\lambda \in (0, 1)$ we have*

$$\int d(\phi_i(x), \phi_i(y))^q \, d\eta(i) \leq \lambda d(x, y)^q \quad \forall x, y \in X \quad (3)$$

$$\int d(x_0, \phi_i(x_0))^q \, d\eta(i) \leq A \quad (4)$$

Then (Φ, η) has a unique stationary measure $\mu \in \mathcal{P}(X)$; moreover μ has finite q -th moment:

$$m_{x_0}^q(\mu) \leq \begin{cases} \frac{A}{1-\lambda} & \text{when } q \leq 1 \\ \frac{A}{(1-\lambda^{\frac{1}{q}})^q} & \text{when } q \geq 1 \end{cases}.$$

The proof is split into a few lemmas.

Lemma 3.2. *The dual transfer operator preserves $\mathcal{P}_q(X)$ and is a contraction of ratio $\bar{\lambda} := \lambda^{\min(1, \frac{1}{q})}$.*

Proof. Let $\mu_0, \mu_1 \in \mathcal{P}_q(X)$ and choose an optimal coupling $\gamma \in \Gamma(\mu_0, \mu_1)$ for W_q . Then

$$\begin{aligned} \int d(x, y)^q \, d(L^* \gamma)(x, y) &= \iint d(\phi_i(x), \phi_i(y))^q \, d\eta(i) \, d\gamma(x, y) \\ &\leq \lambda \int d(x, y)^q \, d\gamma(x, y) \end{aligned}$$

so that $W_q(L^* \mu_0, L^* \mu_1) \leq \bar{\lambda} W_q(\mu_0, \mu_1)$.

In particular, all elements of $\mathcal{P}_q(X)$ are sent at finite W_q -distance from $L^* \delta_{x_0}$ and we only have left to prove that $L^* \delta_{x_0} \in \mathcal{P}_q(X)$, which follows from (4):

$$C_q(\delta_{x_0}, L^* \delta_{x_0}) = \int d(x, y)^q \, d(\delta_{x_0} \otimes L^* \delta_{x_0})(x, y) = \int d(x_0, \phi_i(x_0))^q \, d\eta(i) \leq A.$$

□

Lemma 3.3. *There is a unique stationary measure in $\mathcal{P}_q(X)$. Moreover for all $\nu \in \mathcal{P}_q(X)$, we have $L^{*k} \nu \rightarrow \mu$ in the distance W_q , exponentially fast.*

Proof. Follows from Lemma 3.2 and the Banach fixed point theorem. \square

Lemma 3.4. *For all continuous bounded function $f : X \rightarrow \mathbb{R}$, we have $L^k f(x) \rightarrow \int f d\mu$ for all $x \in X$.*

Proof. Since $\delta_x \in \mathcal{P}_q(X)$ we have

$$L^k f(x) = \int L^k f d\delta_x = \int f d(L^{*k} \delta_x) \rightarrow \int f d\mu$$

\square

Lemma 3.5. *Every stationary measure has finite q -th moment, therefore there is a unique stationary measure in $\mathcal{P}(X)$.*

Proof. For each $n \in \mathbb{N}$, define a continuous bounded function by

$$f_n(x) = \min(d(x_0, x)^q, n).$$

For all stationary measure $\mu' \in \mathcal{P}(X)$ and all $n, k \in \mathbb{N}$ we have

$$\int f_n d\mu' = \int f_n d(L^{*k} \mu') = \int L^k f_n d\mu'$$

Since $L^k f_n$ is bounded between 0 and n for all k , we can apply the dominated convergence theorem as $k \rightarrow \infty$, so that by Lemma 3.4

$$\int f_n d\mu' = \int \lim_{k \rightarrow \infty} L^k f_n d\mu' = \int \left(\int f_n d\mu \right) d\mu' = \int f_n d\mu \leq \int d(x_0, x)^q d\mu < \infty.$$

The monotone convergence theorem applied to f_n as $n \rightarrow \infty$ then shows that

$$\int d(x_0, x)^q d\mu' \leq \int d(x_0, x)^q d\mu < \infty,$$

so that $\mu' \in \mathcal{P}_q(X)$. Lemma 3.3 shows that $\mu' = \mu$. \square

Lemma 3.6. *The unique stationary measure μ satisfies $m_{x_0}^q(\mu) \leq A/(1 - \bar{\lambda})^{\max(1, q)}$.*

Proof. Setting $\bar{A} = A^{\min(1, \frac{1}{q})}$ and using $L^* \mu = \mu$ we get:

$$\begin{aligned} W_q(\delta_{x_0}, \mu) &\leq W_q(\delta_{x_0}, L^* \delta_{x_0}) + W_q(L^* \delta_{x_0}, L^* \mu) \\ &\leq \bar{A} + \bar{\lambda} W_q(\delta_{x_0}, \mu) \\ (1 - \bar{\lambda}) W_q(\delta_{x_0}, \mu) &\leq \bar{A}. \end{aligned}$$

If $q \leq 1$, $\bar{A} = A$ and $\bar{\lambda} = \lambda$; we get $\int d(x_0, x)^q d\mu \leq A/(1 - \lambda)$. If $q \geq 1$; $\bar{A} = A^{\frac{1}{q}}$ and $\bar{\lambda} = \lambda^{\frac{1}{q}}$ and we get $\left(\int d(x_0, x)^q d\mu \right)^{\frac{1}{q}} \leq A^{\frac{1}{q}}/(1 - \lambda^{\frac{1}{q}})$. \square

Theorem 3.1 follows at once from Lemmas 3.3, 3.5 and 3.6.

As an illustration, let us consider a simple case studied for example in [BMS06]. For each $\omega = (i_1, \dots, i_n) \in \{0, 1\}^n$ we set $\phi_\omega = \phi_{i_1} \circ \phi_{i_2} \circ \dots \circ \phi_{i_n}$ (here the order of composition, forward or backward, has no particular relevance).

Corollary 3.7. *Let $a \in (0, 1)$ and $b \in (1, \frac{1}{a})$. The RIFS on the line given by*

$$\begin{aligned} I = \{0, 1\}, \quad \eta_p(\{0\}) = \frac{1}{2} \quad \eta_p(\{1\}) = \frac{1}{2} \\ \phi_0(x) = ax \quad \phi_1(x) = bx + 1 \quad \forall x \in \mathbb{R} \end{aligned} \quad (5)$$

has a unique stationary measure μ , which has finite moments of all orders $q \in (0, q_0)$ where q_0 is the unique solution in $(0, +\infty)$ of $a^{q_0} + b^{q_0} = 2$. More precisely

$$m_0^q(\mu) \leq \frac{2}{2 - (a^q + b^q)}.$$

Moreover, for all $q \in (0, \min(1, q_0))$, all q -Hölder-continuous function $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ and all x_0 ,

$$\left| \frac{1}{2^n} \sum_{\omega \in \{0, 1\}^n} f(\phi_\omega(x_0)) - \int f \, d\mu \right| \leq C \lambda^n$$

where $\lambda = \frac{1}{2}(a^q + b^q) < 1$ and $C = \text{Hol}_q(f)/(1 - \lambda)$.

The notation $\text{Hol}_q(f)$ denotes the best Hölder constant of f ; note that here we ask f to be *globally* q -Hölder, implying it has a growth at infinity of the order of x^q at most. Of course, unbalanced versions of this example (i.e. with $\eta(\{0\}) \neq \eta(\{1\})$) can be studied in the same way.

Proof. First note that the function $q \mapsto a^q + b^q$ is convex, and the assumptions ensure it is decreasing in some interval $(0, q_1)$ and goes to $+\infty$ when $q \rightarrow +\infty$, so that this function takes the value 2 at exactly two points, 0 and q_0 .

For all $q > 0$, we have (4) with $A = 1$ and

$$\int |\phi_i(x) - \phi_i(y)|^q \, d\eta(i) = \frac{1}{2}(a^q + b^q)|x - y|^q$$

so that when $q < q_0$, (3) is satisfied with $\lambda = \frac{1}{2}(a^q + b^q)$. The claim on existence, uniqueness and moments of μ thus follows from Theorem 3.1.

The convergence of empirical averages of f toward its integral with respect to μ follow from Lemma 3.2, observing

$$\frac{1}{2^n} \sum_{\omega \in \{0, 1\}^n} f(\phi_\omega(x_0)) = L^n f(x_0).$$

Indeed as in the proof of Lemma 3.4 we have

$$\begin{aligned} \left| L^n f(x_0) - \int f \, d\mu \right| &= \left| \int f \, d(L^{*n} \delta_{x_0}) - \int f \, d\mu \right| \\ &\leq \text{Hol}_q(f) W_q(L^{*n} \delta_{x_0}, \mu) \\ &\leq \text{Hol}_q(f) \lambda^n W_q(\delta_{x_0}, \mu). \end{aligned}$$

and $W_q(\delta_{x_0}, \mu) = m_{x_0}^q(\mu) \leq \frac{1}{1 - \lambda}$. □

3.2 Two simple tools

To apply Theorem 3.1, it can be convenient to use the following statements.

Proposition 3.8. *If (Φ, η) satisfies (3) and (4), then for all $q' \in (0, q)$ it also satisfies them with constants q' , $\lambda' := \lambda^{q'/q}$ and $A' := A^{q'/q}$.*

Proof. Follows from the Jensen inequality applied to the concave function $r \mapsto r^{q'/q}$. \square

Proposition 3.9. *If (Φ, η) satisfies*

$$\int \log \text{Lip}(\phi_i) \, d\eta(i) < 0 \quad \text{and} \quad \exists p > 0, \int \text{Lip}(\phi_i)^p \, d\eta(i) < +\infty,$$

then there exists $q > 0, \lambda \in (0, 1)$ such that (3) holds.

This enables us to recover a more usual ‘contracting on average’ hypothesis, used for example in [DF99].

Proof. The idea is simply to differentiate $\int \text{Lip}(\phi_i)^t \, d\eta(i)$ with respect to t at $t = 0$; we shall use truncation to differentiate under the integral sign.

For all $n \in \mathbb{N}$, consider the functions $f_n : I \rightarrow \mathbb{R}$ and $F_n : (0, p] \rightarrow \mathbb{R}$ defined by

$$f_n(i) = \max(\text{Lip}(\phi_i), 1/n) \quad \text{and} \quad F_n(t) = \int f_n(i)^t \, d\eta(i).$$

Since $-\log n \leq \log f_n(i) \leq \frac{1}{p} \text{Lip}(\phi_i)^p$ for all i, n , the functions $\log f_n$ are η -integrable. The monotone convergence theorem implies that

$$\int \log f_n(i) \, d\eta(i) \rightarrow \int \log \text{Lip}(\phi_i) \, d\eta(i) \in [-\infty, 0)$$

so that for some $n \in \mathbb{N}$ we have $\int \log f_n(i) \, d\eta(i) \in (-\infty, 0)$. Now $F_n(0) = 1$ and for all $t \in [0, \frac{p}{2}]$:

$$-\log(n) \max\left(1, f_n(i)^{\frac{p}{2}}\right) \leq \frac{d}{dt} f_n(i)^t = \log f_n(i) \cdot f_n(i)^t \leq \frac{2}{p} f_n(i)^{t+\frac{p}{2}}$$

is η -integrable, uniformly so over t (n being fixed above). The function F_n is thus differentiable on $[0, \frac{p}{2}]$, with $F_n'(0) = \int \log f_n(i) \, d\eta_i < 0$. We conclude that there is some $q \in (0, \frac{p}{2})$ such that $F_n(q) \in (0, 1)$. Now

$$\int \text{Lip}(\phi_i)^q \, d\eta_i \leq \int f_n(i)^q \, d\eta_i = F_n(q) < 1$$

which readily implies (3). \square

3.3 Generalized moment estimates

In some cases, the stationary measure of an IFS will not be compactly supported, but will have finite moment of all order; it then makes sense to develop tools to estimate exponential, sub-exponential or super-exponential moments. In practice, the following simple result will be quite efficient.

Proposition 3.10. *Let (Φ, η) be an RIFS on X and $\varphi, \psi : X \rightarrow [0, +\infty)$ two functions that are bounded on every bounded subset of X (most usually, they will factor through $d(x_0, \cdot)$). Denote by L the transfer operator, i.e. $L f(x) = \int f(\phi_i(x)) d\eta(i)$. Assume μ is a stationary measure for (Φ, η) and $\int \psi d\mu < +\infty$.*

If there exist $\theta \in (0, 1)$ and $B \geq 0$ such that for all $x \in X$, $L\varphi(x) \leq \theta\varphi(x) + B\psi$, then $\int \varphi d\mu < +\infty$ and more precisely

$$\int \varphi d\mu \leq \frac{B}{1-\theta} \int \psi d\mu.$$

Proof. For all $m \in [0, +\infty)$, set $\varphi_m(x) = \min(\varphi(x), m)$. We have $L^* \mu = \mu$ since μ is stationary, and by positivity

$$L\varphi_m \leq \min(L\varphi, m) \leq \min(\theta\varphi + B\psi, m) \leq \min(\theta\varphi, m) + B\psi = \theta\varphi_{\frac{m}{\theta}} + B\psi.$$

Applying duality we get:

$$\int \varphi_m d\mu = \int \varphi_m dL^* \mu = \int L\varphi_m d\mu \leq \int (\theta\varphi_{\frac{m}{\theta}} + B\psi) d\mu$$

from which we deduce

$$\int (\varphi_m - \theta\varphi_{\frac{m}{\theta}}) d\mu \leq B \int \psi d\mu.$$

For each $x \in X$, the function $m \mapsto \varphi_m(x) - \theta\varphi_{\frac{m}{\theta}}(x)$ is non-decreasing and converges to $(1-\theta)\varphi(x)$. The monotone convergence theorem ensures we can pass to the limit in the above inequality, leading precisely to the claimed inequality. \square

Let us now consider the stationary measures $\mu_{a,p}$ defined in Section 1.1. Theorem 3.1 shows that $\mu_{a,p}$ has finite moment of all orders. Indeed (3) and (4) are satisfied for all $q \geq 1$ with $x_0 = 0$, $\lambda = 1 - p + pa^q$ (which is less than 1 since $a < 1$), $A = 1 - p$. Let us use Proposition 3.10 to show that $\mu_{a,p}$ has some finite exponential moments.

Corollary 3.11. *For all $p, a \in (0, 1)$ and all $b < \log \frac{1}{1-p}$ we have*

$$\int e^{bx} d\mu_{a,p}(x) \leq \left(\frac{p}{1 - (1-p)e^b} \right)^{\frac{1}{1-a}}$$

Proof. Let L^* be the dual transfer operator and $\varphi(x) = e^{bx}$. We have

$$L^* \varphi(x) = p(\varphi(x))^a + (1-p)e^b \varphi(x).$$

Pick any θ strictly between $(1-p)e^b$ and 1: for some $C > 0$ (which could be computed explicitly), $L^* \varphi(x) \leq \theta \varphi(x)$ whenever $x \geq C$. Since φ is increasing, we have $L^* \varphi \leq \theta \varphi + e^{bC}$ and we can thus apply Proposition 3.10 with $\psi \equiv 1$, obtaining

$$\int e^{bx} d\mu_{a,p} \leq \frac{e^{bC}}{1-\theta} m_0^1(\mu_{a,p}) < \infty.$$

In particular $\int \varphi^a d\mu_{a,p} < +\infty$ and we can apply Proposition 3.10 again, to the more precise inequality $L^* \varphi \leq (1-p)e^b \varphi + p\varphi^a$. We get

$$\begin{aligned} \int \varphi d\mu_{a,p} &\leq \frac{p}{1-(1-p)e^b} \int \varphi^a d\mu_{a,p} \\ &\leq \frac{p}{1-(1-p)e^b} \left(\int \varphi d\mu_{a,p} \right)^a \\ \int \varphi d\mu_{a,p} &\leq \left(\frac{p}{1-(1-p)e^b} \right)^{\frac{1}{1-a}} \end{aligned}$$

□

The bound $\log \frac{1}{1-p}$ is optimal, as shown by Corollary A which we now prove.

Proof of Corollary A. The lower bound follows from the simple observation that for all $n \in \mathbb{N}$ we have $\mu_{a,p}([n+1, n+2]) \geq (1-p)\mu_{a,p}([n, n+1])$, thus $\mu_{a,p}([n, n+1]) \geq (1-p)^n \mu_{a,p}([0, 1])$. Since $\mu_{a,p}$ is a probability measure, there is some $n \in \mathbb{N}$ such that $\mu_{a,p}([0, a^{-n}]) > 0$, and $\mu_{a,p}([0, 1]) \geq p^n \mu_{a,p}([0, a^{-n}]) > 0$.

The upper bound follows from Corollary 3.11 and Chebyshev's inequality: for all $t > 0$ and all $b < \log \frac{1}{1-p}$, we have

$$\mu_{a,p}([t, +\infty)) \leq e^{-bt} \left(\frac{p}{1-(1-p)e^b} \right)^{\frac{1}{1-a}}.$$

Given t , we can choose b in order to optimize the above inequality. An elementary computation leads to take

$$e^b = \frac{1}{1-p} \cdot \frac{(1-a)t}{1+(1-a)t},$$

leading to a bound asymptotically equivalent to

$$\frac{(p(1-a)t)^{\frac{1}{1-a}}}{(1-p)^t}.$$

□

Problem 3.12. Find an asymptotic for the tail of $\mu_{a,p}$, in the spirit of [Kes73] and [Gol91]. Note these works give (in a slightly different context) a precise asymptotic $\mu([t, +\infty)) = f(t) + o(f(t))$ with f a polynomial function; this can thus be written $\mu([t, +\infty)) = f(t + o(t))$. While Corollary A already give an estimation of the form $\mu_{a,p}([t, +\infty)) = g(t + o(t))$, it might be difficult to get $\mu_{a,p}([t, +\infty)) = g(t) + o(g(t))$ since g is exponential and the sensibility on t is thus strong.

3.4 A heavy tail of translations

To illustrate the role of assumption (4) in Theorem 3.1 let us consider the following example on $[0, +\infty)$ (simply denoted by (Φ, η) throughout this subsection):

$$\begin{aligned} I &= \mathbb{N}, & \eta(\{n\}) &= p_n \\ \forall n > 0, \forall x \in \mathbb{R} : & \phi_0(x) = ax & \phi_n(x) &= x + n \end{aligned} \tag{6}$$

where $p_0 > 0$ and, of course, $p_n \geq 0$ and $\sum_{n \geq 0} p_n = 1$.

We have good contraction properties: assumption (3) is satisfied for any $q > 0$ with $\lambda = 1 - (1 - a^q)p_0$. As in the previous example, λ decreases from 1 when $q \rightarrow 0$ to $1 - p_0$ when $q \rightarrow \infty$, and thus improves with q . However the translation part can weight on the moments of the stationary measure. By a direct application of Theorem 3.1, we get:

Proposition 3.13. *Let $q > 0$; if $\sum n^q p_n < +\infty$ then (Φ, η) has a unique stationary measure μ , and μ has finite q -th moment; if $\sum n^q p_n = +\infty$, then any stationary measure of (Φ, η) has infinite q -th moment.*

Giving $(p_n)_n$ a heavy tail and taking $a \ll 1$ we get examples with very quick convergence in low-exponent Wasserstein metric but only few finite moments. This begs the question: what happens when $\sum n^q p_n = \infty$ for all $q > 0$, e.g. when $p_n \sim 1/n(\log n)^2$ as $n \rightarrow \infty$? Does there exist a stationary measure?

3.5 Statistical properties

Assume $(X_k)_{k \in \mathbb{N}}$ is a Markov chain obtained by choosing randomly independently indices $(I_k)_{k \geq 1}$ with law η , and setting $X_k = \phi_{I_k}(X_{k-1})$; we shall say that $(X_k)_{k \in \mathbb{N}}$ is *driven* by (Φ, η) . The fact that the dual transfer operator is a contraction in the Wasserstein distance W_q for some $q \in (0, 1]$ is equivalent to $(X_k)_{k \in \mathbb{N}}$ to have *positive Ricci curvature* in the sense of Ollivier [Oll09] in the metric space (X, d^q) ; this for example implies strong concentration properties of the empirical averages

$$\frac{1}{n} \sum_{k=1}^n f(X_k) =: \hat{\mu}_n(f)$$

whenever $f : X \rightarrow \mathbb{R}$ is a q -Hölder function, see [JO10] for effective and completely explicit results (that depends on many specific quantities that may vary between examples).

When it is possible to restrict to a compact domain of Euclidean space, one can also obtain explicit convergence speed for the empirical measure $\hat{\mu}_n$ itself toward the stationary measure μ , in terms of W_1 or in terms of other distance defined by duality [Klo18a]. Specifically, we shall consider the following metrics between measures:

$$\|\nu_0 - \nu_1\|_{\mathcal{C}_1^s} = \sup_{f \in \mathcal{C}_1^s} \left| \int f d\nu_0 - \int f d\nu_1 \right|$$

where s is any positive integer and \mathcal{C}_1^s is the set of \mathcal{C}^{s-1} functions $\mathbb{R}^n \rightarrow \mathbb{R}$ with all derivatives not greater than 1, and with their derivatives of order $(s-1)$ 1-Lipschitz. In particular, $\|\cdot\|_{\mathcal{C}_1^1}$ matches exactly W_1 .

Theorem 3.14. *Let (Φ, η) be an RIFS on \mathbb{R}^d such that for some $\lambda \in (0, 1)$ every map ϕ_i of Φ is λ -Lipschitz and $\{\|\phi_i(0)\| : i \in I\}$ is bounded. Let $(x_k)_{k \in \mathbb{N}}$ be a Markov chain driven by (Φ, η) , and consider the empirical measures*

$$\hat{\mu}_n := \frac{1}{n} \sum_{k=1}^n \delta_{x_k}.$$

Then there exists a constant $C > 0$ such that for all $n \in \mathbb{N}$,

$$\mathbb{E} \left[\|\hat{\mu}_n - \mu\|_{c_1^s} \right] \leq C \begin{cases} \frac{(\log n)^{\frac{d}{2s+1}}}{\sqrt{n}} & \text{when } 2s > d \\ \frac{\log n}{\sqrt{n}} & \text{when } 2s = d \\ \frac{(\log n)^{d-2s+\frac{s}{d}}}{n^{\frac{s}{d}}} & \text{when } 2s < d. \end{cases} \quad (7)$$

These rates cannot be improved, except possibly for the logarithm factors (see [Klo18a] for this and other considerations, including a comparison with concentrations obtained for a fixed observable f).

Proof. Since Lemma 3.2 with $q = 1$ ensures the Markov chain $(x_k)_{k \in \mathbb{N}}$ is exponentially contracting, this is a direct application of Theorem A in [Klo18a]; the only point to check is that we can restrict to a compact subset of \mathbb{R}^d . This follows easily from the hypotheses: let $D = \sup\{\|\phi_i(0)\| : i \in I\}$; then the ball B of center 0 and radius $R = D/(1 - \lambda)$ satisfies $\phi_i(B) \subset B$ for all i . Indeed whenever $\|x\| \leq R$, we have

$$\|\phi_i(x)\| \leq \|\phi_i(x) - \phi_i(0)\| + \|\phi_i(0)\| \leq \lambda\|x\| + D \leq R.$$

□

Using relaxed assumptions such as (3) and (4) would be nice, but the rates would necessarily be altered given the stationary measure would not usually have finite moments of all orders (without some moment condition, one cannot expect even an optimal approximation by a discrete measure supported on n point to achieve the rate $1/n^{\frac{1}{d}}$ in W_1 when $d > 2$; see [FG15] and [DM19] for rates of convergence of empirical measures under various moment assumptions).

3.6 Stationary measures beyond products

While the case of RIFS as defined above, where the randomness is materialized by a sequence of independent random variables of law $\eta \in \mathcal{P}(I)$, is most commonly studied, there has been some interest to generalize this setting. A first generalization is to replace the i.i.d. sequence by a stationary Markov chain; a further generalization is to draw the infinite word $\omega = \omega_0 \dots \omega_k \dots$ randomly with law an arbitrary shift-invariant measure $\nu \in \mathcal{P}(I^{\mathbb{N}})$ – the case of RIFS corresponding to the independent Bernoulli product $\nu = \eta^{\otimes \mathbb{N}}$; then one can consider a yet further generalization where the shift is replaced by an arbitrary measure-preserving dynamical system.

3.6.1 Skew-products

We still consider (X, d) a complete metric space, and we additionally fix a standard measured space (Y, \mathcal{A}) (i.e. it is isomorphic to $[0, 1]$ with its Borel σ -algebra) equipped with a probability measure ν , and a ν -preserving map $S : Y \rightarrow Y$. A *skew-product map* over S with fiber X is a map

$$\begin{aligned} \Psi : X \times Y &\rightarrow X \times Y \\ (x, y) &\mapsto (\psi_y(x), S(y)) \end{aligned}$$

where $(x, y) \mapsto \psi_y(x)$ is a measurable map. While, as we have seen above, an RIFS can be studied dynamically by looking at a random orbit $x_0, x_{n+1} = \phi_{I_n}(x_n)$ where $(I_n)_{n \geq 1}$ are i.i.d. random variables of law η , in the present setting the corresponding random sequence of points is given by $x_{n+1} = \psi_{S^n(Y)}(x_n)$ where Y is a random element of Y with law ν , taking the place of the whole sequence (I_1, I_2, \dots) . In other words, RIFS correspond to the particular case when $Y = I^{\mathbb{N}}$, $\nu = \eta^{\otimes \mathbb{N}}$, S is the shift $y = (y_0, y_1, \dots) \mapsto S(y) = (y_1, y_2, \dots)$ and $\psi_y(x) = \phi_{y_0}(x)$. Note that Ψ carries the information of what are X, Y and S ; when we refer to this setting, we shall therefore call (Ψ, ν) a *skew product*.

We denote by π^X, π^Y the projection maps from $X \times Y$ to each factor; a measure $\mu \in \mathcal{P}(X)$ is said to be a *stationary measure* of the skew product (Ψ, ν) when there exist a measure $\hat{\nu} \in \mathcal{P}(X \times Y)$ such that:

$$\hat{\nu} \text{ is } \Psi\text{-invariant,} \quad \pi_*^Y \hat{\nu} = \nu, \quad \text{and} \quad \pi_*^X \hat{\nu} = \mu.$$

In the case of a RIFS, this coincides with the previous definition of stationary measure. The measure $\hat{\nu}$ as above shall be called a *lift* of ν . The basic question we want to address under specific assumptions is whether there exist a unique stationary measure; a positive answer will follow from the uniqueness of the lift of ν .

Definition 3.15. We say that Ψ *contracts the fibers* whenever there exist $\lambda \in (0, 1)$ such that for all $y \in Y$, the map ψ_y is λ -Lipschitz.

We say that Ψ has *bounded displacement* if for some $x_0 \in X$, there exist an $A > 0$ such that the set $d(x_0, \psi_y(x_0)) \leq A$ for all $y \in Y$.

Observe that when $(\psi_y)_{y \in Y}$ is an equicontinuous family, e.g. when Ψ contracts the fibers, in the definition of bounded displacement “for some x_0 ” could be equivalently replaced by “for all x_0 ” (up to changing the value of A).

The main result of this section is the following.

Theorem 3.16. *Let Ψ be a skew-product map on $X \times Y$ that contracts the fibers and has bounded displacement. Each S -invariant $\nu \in \mathcal{P}(Y)$ has a unique lift, and in particular the skew product (Ψ, ν) has a unique stationary measure μ , which moreover has bounded support.*

Let $(x_k)_{k \in \mathbb{N}}$ be a stochastic process associated to (Ψ, ν) as above, with x_0 independent from Y and of arbitrary law $\tilde{\mu}_0 \in \mathcal{P}_q(X)$ for some $q > 0$, and let $\tilde{\mu}_k \in \mathcal{P}(X)$ be the law of x_k . Then for all $k \in \mathbb{N}$,

$$W_q(\tilde{\mu}_k, \mu) \leq D\tilde{\lambda}^k$$

where

$$\tilde{\lambda} = \lambda^{\min(q,1)} \in (0,1), \quad D = m_{x_0}^q(\tilde{\mu}_0)^{\min(1,\frac{1}{q})} + \left(\frac{A}{1-\lambda}\right)^{\min(1,q)},$$

and A, λ are the constants in the bounded displacement and fiber contraction hypotheses.

3.6.2 Fiber-wise Wasserstein distance

The main tool to prove Theorem 3.16 is a variation of Wasserstein distance that is adapted to a projection map and the inverse images of a given measure on its target space. This notion was at the heart of [Klo18b], from which we adapt the relevant definitions and properties. Theorem A from [Klo18b] is not immediately applicable here since X need not be compact, $\text{diam}(\Psi^n(X \times \{y\}))$ might be infinite for all n , and Y is not even a topological space; but the adaptation is relatively straightforward.

Fix any $\nu \in \mathcal{P}(Y)$ and let $\mathcal{P}^\nu := (\pi_*^Y)^{-1}(\nu) \subset \mathcal{P}(X \times Y)$ be the *fiber* of ν , i.e. the set of measures on $X \times Y$ with second marginal equal to ν . Reminding that we fixed a point $x_0 \in X$, given any $\rho \in \mathcal{P}^\nu$ and $q > 0$ we define its q -th moment by

$$m_{x_0}^q(\rho) = \int d(x, x_0)^q d\rho(x, y)$$

where the integral is over the whole product $X \times Y$ but distances are recorded only “along the fibers”, i.e. over the X factor. We let \mathcal{P}_q^ν be the subset of \mathcal{P}^ν consisting of measure of finite q -th moment (this set does not depend on x_0).

The product $X^2 \times Y$ identifies with what was noted Δ_π in [Klo18b] (pairs of point in the total space that project to the same point on the base Y); we consider the maps

$$\pi_{02} : (x, x', y) \mapsto (x, y) \quad \pi_{12} : (x, x', y) \mapsto (x', y) \quad \pi_2 : (x, x', y) \mapsto y.$$

For all $\rho_0, \rho_1 \in \mathcal{P}^\nu$ let $\Gamma^\nu(\rho_0, \rho_1) := \{\gamma \in \mathcal{P}(X^2 \times Y) \mid (\pi_{02*})\gamma = \rho_0 \text{ and } (\pi_{12*})\gamma = \rho_1\}$ (playing the role of Γ_π in [Klo18b], where we had chosen to emphasize the projection map rather than the image measure) and define

$$C_q^\nu(\rho_0, \rho_1) = \inf_{\gamma \in \Gamma^\nu(\rho_0, \rho_1)} \int d(x, x')^q d\gamma(x, x', y)$$

$$W_q^\nu(\rho_0, \rho_1) = \left(C_q^\nu(\rho_0, \rho_1)\right)^{\min(1, \frac{1}{q})}.$$

The following basic result is proven in the same way as in [Klo18b].

Proposition 3.17. *For all $\rho_0, \rho_1 \in \mathcal{P}^\nu$, the set $\Gamma^\nu(\rho_0, \rho_1)$ is non-empty. If the moments $m_{x_0}^q(\rho_i)$ are finite for $i \in \{0, 1\}$, then $W^\nu(\rho_0, \rho_1) < \infty$. Moreover, if $(\xi_y)_{y \in Y}$ and $(\zeta_y)_{y \in Y}$ are the disintegrations of ρ_0 and ρ_1 with respect to π^Y , then*

$$W_q^\nu(\rho_0, \rho_1) = \begin{cases} \left(\int W_q(\xi_y, \zeta_y)^q d\nu(y) \right)^{\frac{1}{q}} & \text{when } q \geq 1 \\ \int W_q(\xi_y, \zeta_y) d\nu(y) & \text{when } q \leq 1. \end{cases} \quad (8)$$

Last, W_q^ν is a complete metric on the set \mathcal{P}_q^ν .

Proof. Let $(\xi_y)_{y \in Y}$ and $(\zeta_y)_{y \in Y}$ be the disintegrations of ρ_0 and ρ_1 with respect to π^Y (identifying X with the fibers of π^Y , $(\xi_y)_{y \in Y}$ is thus a family of measures on X characterized by $\int f(x, y) d\xi_y(x) d\nu(y) = \int f(x, y) d\rho_0(x, y)$ for all continuous bounded function $f : X \times Y \rightarrow \mathbb{R}$.)

From any measurable choice of $y \mapsto \gamma_y \in \Gamma(\xi_y, \zeta_y)$ (e.g. $\gamma_y = \xi_y \otimes \zeta_y$) we can build an element γ of $\Gamma^\nu(\rho_0, \rho_1)$ by setting $\int f(x, x', y) d\gamma(x, x', y) = \iint f(x, x', y) d\gamma_y(x, x') d\nu(y)$. In particular $\Gamma^\nu(\rho_0, \rho_1)$ is non-empty.

Conversely, given any $\gamma \in \Gamma^\nu(\rho_0, \rho_1)$ its disintegration with respect to π_2 is a family $(\gamma_y)_{y \in Y}$ of measures on $X \times X$, and by testing γ against integrands of the form $f(x)g(y)$ and $f(x')g(y)$ one sees that $\gamma_y \in \Gamma(\xi_y, \zeta_y)$ for ν -almost all y .

Since $\int d(x, x')^q d\gamma(x, x', y) = \iint d(x, x')^q d\gamma_y(x, x') d\nu(y) \geq \int C_q(\xi_y, \zeta_y) d\check{\mu}(y)$, taking an infimum we get $C_q^\nu(\mu_0, \mu_1) \geq \int C_q(\xi_y, \zeta_y) d\check{\mu}$.

For each y , the set of optimal transport plans from ξ_y to ζ_y is compact (see e.g. the proof of Theorem 4.1 in [Vil09]), thus by the measurable selection theorem there is a measurable family $(\gamma_y)_{y \in Y}$ such that for ν -almost all $y \in Y$, $\int d(x, x')^q d\gamma_y(x, x') = C_q(\xi_y, \zeta_y)$. It follows $C_q^\nu(\rho_0, \rho_1) \leq \int C_q(\xi_y, \zeta_y) d\nu$ and (8) is proven.

To complete the proof, it remains to be seen that $C_q^\nu(\rho_0, \rho_1) < \infty$ and that W_q^ν is a metric making \mathcal{P}_q^ν a complete space. The triangular inequality follows from (8), and then finiteness is obtained by observing

$$W_q^\nu(\rho_0, \rho_1) \leq W_q^\nu(\rho_0, \delta_{x_0} \otimes \nu) + W_q^\nu(\delta_{x_0} \otimes \nu, \rho_1) = m_{x_0}^q(\rho_0) + m_{x_0}^q(\rho_1).$$

Finally, The Riesz-Fischer Theorem for metric-space valued functions ensures that W_q^ν is a complete metric on \mathcal{P}_q^ν , seen *via* disintegration as a closed subset of the space of maps $Y \rightarrow \mathcal{P}_q(X)$. \square

3.6.3 Proof of Theorem 3.16

Let ν be any S -invariant probability measure on Y . We first observe that the fiber contraction and bounded displacement properties ensure that Ψ_* preserves \mathcal{P}_q^ν for all q . These uniform assumptions also ensure that for some bounded set $B \subset X$, the set $B \times Y$ is an *absorbing invariant set*, i.e. $\Psi(B \times Y) \subset B \times Y$ and for all $(x, y) \in X \times Y$ there is some $k \in \mathbb{N}$ such that $\Psi^k(x, y) \in B \times Y$. Let indeed $A > 0$ be such that for all y , $d(x_0, \psi_y(x_0)) \leq A$, fix any $\varepsilon > 0$, set $R = (1 + \varepsilon)A/(1 - \lambda)$ and let $B = B(x_0, R)$ be the ball of center x_0 and radius R in X ; then for all $x, y \in X \times Y$ we have

$$\begin{aligned} d(x_0, \psi_y(x)) &\leq d(x_0, \psi_y(x_0)) + d(\psi_y(x_0), \psi_y(x)) \\ &\leq A + \lambda d(x_0, x). \end{aligned}$$

When $x \in B$, the right-hand side is at most $A + \lambda R = \frac{1 + \varepsilon \lambda}{1 - \lambda} A < R$, proving the $B \times Y$ is Ψ -invariant. When $x \notin B$, we have $A < \frac{1 - \lambda}{1 + \varepsilon} d(x_0, x)$ and the right-hand side is at most

$$\left(\frac{1 - \lambda}{1 + \varepsilon} + \lambda \right) d(x_0, x) = \frac{1 + \varepsilon \lambda}{1 + \varepsilon} d(x_0, x)$$

where $\frac{1 + \varepsilon \lambda}{1 + \varepsilon} < 1$, proving the absorbing property with $k \simeq \log d(x_0, x)$.

Let $\rho_0, \rho_1 \in \mathcal{P}_q^\nu$. We consider the map $X^2 \times Y \rightarrow X^2 \times Y$ defined by

$$\Psi_2(x, x', y) = (\psi_y(x), \psi_y(x'), S(y)).$$

For $i \in \{0, 1\}$ we have $\pi_{i2} \circ \Psi_2 = \Psi \circ \pi_{i2}$; as a consequence, for any $\gamma \in \Gamma^\nu(\rho_0, \rho_1)$ we have $\Psi_{2*}\gamma \in \Gamma^\nu(\Psi_*\rho_0, \Psi_*\rho_1)$. Observing

$$\begin{aligned} \int d(x, x')^q d\Psi_{2*}\gamma(x, x', y) &= \int d(\psi_y(x), \psi_y(x'))^q d\gamma(x, x', y) \\ &\leq \lambda^q \int d(x, x')^q d\gamma(x, x', y) \end{aligned}$$

and taking an infimum, we see that

$$W_q^\nu(\Psi_*\rho_0, \Psi_*\rho_1) \leq \lambda^{\min(1, q)} W_q^\nu(\rho_0, \rho_1),$$

in particular Ψ_* induces a contraction on the complete metric space $(\mathcal{P}_q^\nu, W_q^\nu)$. Therefore, there exists a unique Ψ -invariant lift $\hat{\nu}$ of ν having finite q -th moment. By considering different q , we already see that the measure $\hat{\nu}$ has finite moments of all orders but, since $B \times Y$ is absorbing, any Ψ -invariant measure is concentrated on $B \times Y$. This proves that $\hat{\nu}$ is the unique Ψ -invariant lift of ν on the whole of \mathcal{P}^ν , and that its first marginal μ is supported on a bounded set. Explicitly, by letting ε above go to 0, we have that μ is concentrated on $B(x_0, A/(1 - \lambda))$.

Consider now the stochastic process $(X_k)_{k \in \mathbb{N}}$. Let $\rho_0 := \tilde{\mu}_0 \otimes \nu$ be the law of (X_0, Y) ; then the law of $(X_k, S^k(Y)) = \Psi^k(X_0, Y)$ is $\rho_k = \Psi_*^k(\rho_0)$, by definition has first marginal $\tilde{\mu}_k$, and by invariance has second marginal ν . Since Ψ_* is a contraction in $\mathcal{P}_q^\nu \ni \rho_0$, we obtain that

$$W_q^\nu(\rho_k, \hat{\nu}) \leq \lambda^{k \min(1, q)} W_q^\nu(\rho_0, \hat{\nu}). \quad (9)$$

On the first hand, using the transport plan obtained by projecting an optimal $\gamma \in \Gamma^\nu(\rho_k, \hat{\nu})$ on the first two variables, we get $W_q(\tilde{\mu}_k, \mu) \leq W_q^\nu(\rho_k, \hat{\nu})$. On the other hand, we have

$$\begin{aligned} W_q^\nu(\rho_0, \hat{\nu}) &\leq W_q^\nu(\tilde{\mu}_0 \otimes \nu, \delta_{x_0} \otimes \nu) + W_q^\nu(\delta_{x_0} \otimes \nu, \hat{\nu}) \\ &\leq m_{x_0}^q (\tilde{\mu}_0)^{\min(1, \frac{1}{q})} + (A/(1 - \lambda))^{\min(1, q)} \end{aligned}$$

since $\hat{\nu}$ is concentrated on $B(x_0, A/(1 - \lambda)) \times Y$. Together with (9), this concludes the proof of Theorem 3.16.

4 Dependence of the stationary measure on the RIFS and linear response

In this section we seek to quantify how close the stationary measures of two slightly different RIFS must be. To this end, we need to introduce a way to quantify the distance between RIFS; it is both natural and effective to use an adaptation of Wasserstein

distances. There are two points to consider in this adaptation: first, which metric to use for maps; second, how to take into account that we need not consider IFS with the same index set. The second point is easily dealt with, by considering couplings $\gamma \in \mathcal{P}(I_0 \times I_1)$ of measures $\eta_0 \in \mathcal{P}(I_0)$ and $\eta_1 \in \mathcal{P}(I_1)$. There is much flexibility to address the first point; taking the uniform distance $d_\infty(\phi, \psi) = \sup_{x \in X} d(\phi(x), \psi(x))$ is ill suited to the non-compact case, as for example the map $x \mapsto ax + b$ acting on \mathbb{R} would not depend continuously on the parameters a, b : changing a the slightest bit would yield a map infinitely far from the original one. We therefore consider a pointed Lipschitz distance, notably suitable for Lipschitz IFS:

$$\begin{aligned} d_{x_0}(\phi, \psi) &:= \min \left\{ \varepsilon \geq 0 \mid \forall x \in X : d(\phi(x), \psi(x)) \leq \varepsilon + \varepsilon d(x, x_0) \right\} \\ &= \sup_{x \in X} \frac{d(\phi(x), \psi(x))}{1 + d(x, x_0)}. \end{aligned}$$

This defines a metric on the space of Lipschitz maps $X \rightarrow X$, and we construct from it the Wasserstein-like distance $W_{x_0, q}$ (possibly taking the value ∞) between RIFS:

$$\begin{aligned} C_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right) &:= \inf_{\gamma \in \Gamma(\eta_0, \eta_1)} \int d_{x_0}(\phi_i, \psi_j)^q d\gamma(i, j) \\ W_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right) &:= C_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right)^{\min(1, \frac{1}{q})}. \end{aligned}$$

4.1 Lipschitz regularity of the stationary measure

Theorem 4.1. *Consider two RIFS (Φ_0, η_0) and (Φ_1, η_1) such that the first one satisfies (3) and (4) for some q, λ_0, A_0 (and thus has a unique stationary measure μ_0), and such that the second one has at least one stationary measure μ_1 with finite q -th moment $m_{x_0}^q(\mu_1)$. Then we have*

$$\begin{aligned} W_q(\mu_0, \mu_1) &\leq C W_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right) \\ \text{where } C &= \begin{cases} \frac{1 + m_{x_0}^q(\mu_1)}{1 - \lambda_0} & \text{when } q \leq 1 \\ 2^{1 - \frac{1}{q}} \frac{(1 + m_{x_0}^q(\mu_1))^{\frac{1}{q}}}{1 - \lambda_0^{\frac{1}{q}}} & \text{when } q > 1 \end{cases} \end{aligned}$$

In particular, if we fix $\Phi = \Phi_0 = \Phi_1$ and restrict to measures η satisfying (3) and (4) the map $\eta \mapsto \mu$ (which is well-defined by Theorem 3.1) is locally Lipschitz.

Remark 4.2. When the second RIFS satisfy (3) and (4) with constants q, λ_1, A_1 , we can choose to apply the result after exchanging them to optimize; however using only the moment estimate of Theorem 3.1 this is expected to provide small improvements, since both spectral gaps (i.e. $1 - \lambda_0$ and $1 - \lambda_1$) are involved in denominators.

Remark 4.3. When the second RISF has several stationary measures with finite q -th moment, Theorem 4.1 shows that they *all* lie within small distance of μ_0 .

Let us now turn to the proof of Theorem 4.1. We denote by L_k the transfer operator of the RIFS (Φ_k, η_k) ($k \in \{0, 1\}$).

Lemma 4.4. *For all RIFS (Φ_0, η_0) , (Φ_1, η_1) and all $\nu \in \mathcal{P}_q(X)$ we have*

$$W_q(L_0^* \nu, L_1^* \nu) \leq D W_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right)$$

$$\text{where } D = \begin{cases} 1 + m_{x_0}^q(\nu) & \text{when } q \leq 1 \\ 2^{1-\frac{1}{q}} (1 + m_{x_0}^q(\nu))^{\frac{1}{q}} & \text{when } q > 1 \end{cases}$$

Proof. If $C_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right) = \infty$, the statement is empty true. Assume otherwise, and let $\gamma \in \Gamma(\eta_0, \eta_1)$ be an optimal coupling. Let $\bar{\nu} = (\text{Id}, \text{Id})_* \nu \in \mathcal{P}(X \times X)$ be the trivial coupling of ν with itself. Then $\bar{\gamma} := \int (\phi_i, \psi_j)_* \bar{\nu} d\gamma(i, j)$ is a coupling of $L_0^* \nu$ and $L_1^* \nu$, so that

$$\begin{aligned} C_q(L_0^* \nu, L_1^* \nu) &\leq \int d(x, y)^q d\bar{\gamma}(x, y) \\ &= \iint d(\phi_i(x), \psi_j(y))^q d\bar{\nu}(x, y) d\gamma(i, j) \\ &= \iint d(\phi_i(x), \psi_j(x))^q d\nu(x) d\gamma(i, j) \\ &\leq \iint \left(d_{x_0}(\phi_i, \psi_j)(1 + d(x, x_0)) \right)^q d\nu(x) d\gamma(i, j) \\ &\leq \int d_{x_0}(\phi_i, \psi_j)^q d\gamma(i, j) \int (1 + d(x, x_0))^q d\nu(x). \end{aligned}$$

When $q \leq 1$, using $(1 + r)^q \leq 1 + r^q$ we obtain

$$\begin{aligned} W_q(L_0^* \nu, L_1^* \nu) &\leq W_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right) \int (1 + d(x, x_0))^q d\nu(x) \\ &\leq W_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right) (1 + m_{x_0}^q(\nu)) \end{aligned}$$

while when $q \geq 1$, using $(1 + r)^q \leq 2^{q-1}(1 + r^q)$ we get

$$\begin{aligned} W_q(L_0^* \nu, L_1^* \nu) &\leq W_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right) \left(2^{q-1} \int (1 + d(x, x_0))^q d\nu(x) \right)^{\frac{1}{q}} \\ &\leq W_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right) \cdot 2^{1-\frac{1}{q}} (1 + m_{x_0}^q(\nu))^{\frac{1}{q}}. \end{aligned}$$

□

Proof of Theorem 4.1. Apply Lemma 4.4 to $\nu = \mu_1$ and use that L_0^* is a contraction (Lemma 3.2, recall $\bar{\lambda}_0 = \lambda_0^{\min(1, \frac{1}{q})}$):

$$\begin{aligned} W_q(\mu_0, \mu_1) &\leq W_q(\mu_0, L_0^* \mu_1) + W_q(L_0^* \mu_1, \mu_1) \\ &= W_q(L_0^* \mu_0, L_0^* \mu_1) + W_q(L_0^* \mu_1, L_1^* \mu_1) \\ &\leq \bar{\lambda}_0 W_q(\mu_0, \mu_1) + D W_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right) \\ W_q(\mu_0, \mu_1) &\leq \frac{D}{1 - \bar{\lambda}_0} W_{x_0, q} \left((\Phi_0, \eta_0), (\Phi_1, \eta_1) \right) \end{aligned}$$

□

4.2 Linear response

The Rademacher theorem ensures that Lipschitz functions $[a, b] \rightarrow \mathbb{R}$ are differentiable Lebesgue almost-everywhere; a similar result has been proven by Ambrosio, Gigli and Savaré [AGS08] for maps $[a, b] \rightarrow \mathcal{P}_q(\mathbb{R}^n)$ for $q > 1$, where one has of course to make precise what “differentiable” means. Together with Theorem 4.1, this provides a “linear response formula” in many cases.

Corollary 4.5 (Linear Response). *Let $(\Phi_t, \eta_t)_{t \in [a, b]}$ be a curve of RIFS on \mathbb{R}^n (endowed with the Euclidean metric, the origin O serving as reference point), assume that for some $q > 1$,*

- i. $(\Phi_t, \eta_t)_t$ is Lipschitz in $W_{O, q}$, i.e. there exist some $C > 0$ such that for all $t, t' \in [a, b]$,*

$$W_{O, q} \left((\Phi_t, \eta_t), (\Phi_{t'}, \eta_{t'}) \right) \leq C |t - t'|,$$

- ii. there exist $\lambda_+ \in (0, 1)$ and $A_+ > 0$ such that for all $t \in [a, b]$ the RIFS (Φ_t, η_t) satisfies hypotheses (3) and (4) with parameters q , $\lambda_t \leq \lambda_+$ and $A_t \leq A_+$*

and let μ_t denote the unique stationary measure of (Φ_t, η_t) . Then there exist a family $(v_t)_{t \in [a, b]}$ of measurable vector fields on \mathbb{R}^n such that for Lebesgue almost all t :

- i. $\|v_t\| \in L^q(\mu_t)$ and $|v_t|^{q-2} v_t$ can be approximated by gradients of smooth functions $\mathbb{R}^n \rightarrow \mathbb{R}$ in the $L^{q'}(\mu_t)$ norm where $q' = q/(q-1)$,*
- ii. $\frac{d}{dt} \mu_t + \nabla \cdot (v_t \mu_t) = 0$ weakly, i.e. for almost all $t_0 \in [a, b]$ and all C^1 compactly supported functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$:*

$$\frac{d}{dt} \int f \, d\mu_t \Big|_{t=t_0} = \int \nabla f \cdot v_{t_0} \, d\mu_{t_0}.$$

Note that here $q > 1$ is needed to ensure strict convexity in the optimal transport problem.

Proof. By Theorem 4.1, the family of stationary measures $(\mu_t)_{t \in [a, b]}$ is Lipschitz, in particular absolutely continuous (with bounded derivative). Thus Theorem 8.3.1 of [AGS08] applies, giving precisely the claim (note the formulation (8.1.4) for the interpretation of the continuity equation). \square

4.3 The case of Bernoulli convolutions

Corollary B will follow from Corollary 4.5. We take as reference point $x_0 = O = 0 \in \mathbb{R}$ and recall that the family (Φ^λ, η) of RIFS defining Bernoulli convolutions is given in Section 1.2.

Proof of Corollary B. Fix any $q > 1$ and let us check the hypotheses of Corollary 4.5

First, given $\lambda, \lambda' \in (0, 1)$,

$$d_{x_0}(\phi_0^\lambda, \phi_0^{\lambda'}) = d_{x_0}(\phi_1^\lambda, \phi_1^{\lambda'}) = |\lambda - \lambda'|,$$

so that the identity coupling of η with itself, defined by $\gamma(\{(0, 0)\}) = \gamma(\{(1, 1)\}) = \frac{1}{2}$ and $\gamma(\{(0, 1)\}) = \gamma(\{(1, 0)\}) = 0$, implies $C_{x_0, q}((\Phi^\lambda, \eta), (\Phi^{\lambda'}, \eta)) \leq |\lambda - \lambda'|^q$. Second, in restriction to any interval $[\frac{1}{2}, 1 - \varepsilon]$ with $\varepsilon > 0$ we have (3) and (4) with uniform bounds $\lambda_+ = 1 - \varepsilon$ and $A_+ = 1$.

Corollary 4.5 provides us with a family $(v_\lambda)_{\lambda \in (0, 1)}$ of vector fields on \mathbb{R} , which can be identified with functions $\mathbb{R} \rightarrow \mathbb{R}$, such that $\|v_\lambda\| \in L^q(\mu_\lambda)$ and $\frac{d}{d\lambda}\mu_\lambda + (v_\lambda\mu_\lambda)' = 0$ weakly for almost all λ . When $\lambda > \frac{1}{2}$, up to further restricting to a subset of full Lebesgue measure for the parameter λ , by Solomyak's Theorem [Sol95] we have μ_λ absolutely continuous with density denoted by g_λ . For a \mathcal{C}^1 compactly supported test function $f : \mathbb{R} \rightarrow \mathbb{R}$ we get

$$\frac{d}{dt} \int f d\mu_t \Big|_{t=\lambda} = \int_0^1 f'(x) v_\lambda(x) g_\lambda(x) dx,$$

which is the desired formula with $w_\lambda = v_\lambda g_\lambda$. Moreover, for almost-all $\lambda > 1/\sqrt{2}$, as mentioned above g_λ is continuous, hence bounded. Then $w_\lambda \in L^q(\mu_\lambda)$ so that $w_\lambda \in L^q([0, 1])$.

Now, w_λ *a priori* depends on the choice of q . But if \tilde{w}_λ is another suitable choice for the same λ (and possibly different q), extending both of them by 0 outside $[0, 1]$, we have $\int_{\mathbb{R}} f' w_\lambda = \int_{\mathbb{R}} f' \tilde{w}_\lambda$ for all test functions f . As a consequence the extensions of w_λ and \tilde{w}_λ must differ by a constant, and we thus must have $w_\lambda = \tilde{w}_\lambda$. It follows that we have a single w_λ belonging to all $L^q([0, 1])$ simultaneously. \square

We also got the following Lipschitz estimate along the way.

Proposition 4.6. *For all $q \geq 1$ and all $\lambda, \lambda' \in (0, 1)$:*

$$W_q(\mu_\lambda, \mu_{\lambda'}) \leq \frac{2}{1 - \lambda} |\lambda - \lambda'|.$$

In particular, $\lambda \mapsto \mu_\lambda$ is Lipschitz in the metrics W_q on each interval of the form $[0, 1 - \varepsilon]$ where $\varepsilon > 0$.

Proof. On the one hand, we proved above $W_{0, q}((\Phi^\lambda, \eta), (\Phi^{\lambda'}, \eta)) \leq |\lambda - \lambda'|$; on the other hand, the RIFS (Φ^λ, η) obviously satisfies (3) and (4) for any $q > 0$ and constants λ^q and $A = 1$; moreover $m_{x_0}^q(\mu_{\lambda'}) \leq 1$. Theorem 4.1 then ensures that

$$W_q(\mu_\lambda, \mu_{\lambda'}) \leq \frac{2^{1 - \frac{1}{q}} 2^{\frac{1}{q}}}{1 - \lambda} |\lambda - \lambda'|.$$

\square

5 Approximate overlaps and absolute continuity

5.1 Dimension, overlaps and approximation

Consider a RIFS (Φ, η) with finite index set I ; for simplicity, we shall write η_i for $\eta(\{i\})$. Given a positive integer n , for each word $\omega = (i_1, \dots, i_n) \in I^n$ one sets $\phi_\omega = \phi_{i_1} \circ \dots \circ \phi_{i_n}$ and $\eta_\omega = \eta_{i_1} \cdot \eta_{i_2} \cdot \dots \cdot \eta_{i_n}$; this gives rise to a new IFS $\Phi^n := (\phi_\omega)_{\omega \in I^n}$ and a new probability distribution $\eta^n = (\eta_\omega)_{\omega \in I^n}$. One defines the *upper Lyapunov exponent* of (Φ, η) as

$$\chi = \chi(\Phi, \eta) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\omega \in I^n} \eta_\omega \log \text{Lip}(\phi_\omega)$$

(the limit exists by sub-additivity) and the entropy of η as

$$h = H(\eta) := \sum_{i \in I} \eta_i \log \frac{1}{\eta_i}.$$

Then assuming $h < +\infty$ and $-\infty < \chi < 0$, one can prove

$$\dim \mu \leq \min \left(d, \frac{h}{-\chi} \right) \tag{10}$$

A folklore conjecture (see e.g. [Hoc14]) asserts that when $X = \mathbb{R}$ and the ϕ_i are similarities (in which case $\chi = \sum \eta_i \log r_i$, with r_i the contraction ratio of ϕ_i) then there is equality in (10) unless there are *exact overlaps*, i.e. there exist $n \in \mathbb{N}$ and words $\omega \neq \omega' \in I^n$ such that $\phi_\omega = \phi_{\omega'}$. When exact overlaps occur, μ can be expressed as the stationary measure for the IFS made of all *distinct* compositions of n maps among the ϕ_i with the corresponding probability vector; then the Lyapunov exponent becomes $n\chi$ but the entropy is less than nh , and the bound (10) may improve.

An informal way to understand (10) is to first consider the case when η is uniform and all r_i are equal. A best approximation of a measure of dimension α by a discrete measure supported on N points should be away from the approximated measure by an error of the order of $1/N^{\frac{1}{\alpha}}$; the discrete measure $\sum_{\omega \in I^n} \eta_\omega \phi_\omega(x_0)$ has at most $|I|^n$ points, and is at W_1 distance at most $\simeq e^{n\chi}$ from μ , we get $\alpha \leq h/(-\chi)$ in the limit $n \rightarrow \infty$. Then (10) is a natural generalization in the general case. This point of view immediately suggests that *approximate* overlaps could in principle result in a lower dimension bounds: instead of considering $\sum_{\omega \in I^n} \eta_\omega \phi_\omega(x_0)$ as the approximation measure, one could take advantage of approximate overlaps to merge the points $\phi_\omega(x_0)$ that cluster closely enough, reducing the number of support points (and the entropy). It seems difficult to find a RIFS made of similarities where there are enough approximate overlaps to imply a dimension drop (finding one without exact overlap would disprove the folklore conjecture mentioned above). It is the goal of this section to show that this observation can still be used in some particular circumstances to prove that some stationary measures are not absolutely continuous.

5.2 Approximate overlaps can prevent absolute continuity

We shall use the following simple definition.

Definition 5.1. The *approximate cardinal* of the n -th iteration of Φ at x_0 up to uncertainty $\varepsilon > 0$ is defined as the integer

$$N(\Phi, x_0, n, \varepsilon) = \min \left\{ |E| : E \subset \mathbb{R}^d, \forall \omega \in I^n, \exists y \in E : \|\phi_\omega(x_0) - y\| \leq \varepsilon \right\}$$

We could instead ask that *most* (instead of *all*) points $\phi_\omega(x_0)$ can be approximated by some $y \in E$ and it would also be relevant to consider instead the least entropy of an approximation of $\sum_{\omega \in I^n} \eta_\omega \delta_{\phi_\omega(x_0)}$; however the above definition will suffice to treat the case we are interested in.

Theorem 5.2. Let $\ell \geq 3$ be an integer and set $I = \{0, 1, \dots, \ell - 1\}$. Let $\Phi = (\phi_i)_{i \in I}$ be an IFS on \mathbb{R}^d made of ℓ similarities of ratio $(\frac{1}{\ell})^{\frac{1}{d}}$, let η be the uniform measure on I , and let μ be the stationary measure. The Lyapunov exponent is then $\chi = \frac{1}{d} \log \frac{1}{\ell}$, the entropy is $h = \log \ell$ and $h/(-\chi) = d$.

If there exist an increasing sequence of positive integers $(n_k)_{k \in \mathbb{N}}$ and a sequence of positive reals $(\varepsilon_k)_{k \in \mathbb{N}}$ with $\varepsilon_k = O(e^{\chi n_k})$, such that

$$N(\Phi, x_0, n_k, \varepsilon_k) = o(e^{-\chi d n_k})$$

then μ is singular with respect to the Lebesgue measure.

Proof. By assumption, for each k there exists a set $E_k \subset \mathbb{R}^d$ with $N_k = N(\Phi, x_0, n_k, \varepsilon_k)$ elements and a map $f : I^{n_k} \rightarrow E$ such that for each $\omega \in I^{n_k}$, $\|\phi_\omega(x_0) - f(\omega)\| \leq \varepsilon_k \lesssim e^{\chi n_k}$. Let $\mu_k = \sum_{\omega \in I^{n_k}} \ell^{-n_k} \delta_{\phi_\omega(x_0)}$ and $\mu'_k = \sum_{\omega \in I^{n_k}} \ell^{-n_k} \delta_{f(\omega)}$. On the one hand by using the transport plan naturally defined by f , we have $W_1(\mu_k, \mu'_k) \leq \varepsilon_k \lesssim e^{\chi n_k}$. On the other hand, since $\mu_k = L^{n_k}(\delta_{x_0})$ where L is the dual transfer operator, $W_1(\mu_k, \mu) \leq e^{\chi n_k} m_{x_0}^1(\mu)$ (μ is compactly supported given the assumptions, thus has finite first moment). The triangular inequality then implies $W_1(\mu'_k, \mu) \lesssim e^{\chi d n_k}$.

We know that μ'_k is supported on N_k points; by a classical quantization theorem (see [GL00], Theorem 6.2), if μ were absolutely continuous we would have $W_1(\mu, \mu'_k) \gtrsim 1/N_k^{\frac{1}{d}}$, i.e. $N_k \gtrsim e^{-\chi n_k}$. This is precisely contradicted by the assumption. \square

Remark 5.3. There are several important restrictions in this result: that the ϕ_i have equal ratio, that η is uniform, and that the dimension bound is exactly d . They can probably be lifted if we ask $N(\Phi, x_0, n_k, \varepsilon_k) = O(e^{c d n_k})$ for some $c < -\chi$, or more precisely an equivalent bound on the minimal discrete entropy of an ε_k -approximation of μ_k , in which case we expect to get a suitably improved dimension bound for μ . However, we do not know an example where enough approximate overlaps occur. This explains our need for the equality $h/(-\chi) = d$. If the ratios r_i of the ϕ_i were not equal, then we would have $W_1(\mu_k, \mu) \lesssim \bar{r}^{n_k}$ with \bar{r} the arithmetic mean of the r_i , while e^{χ} is their geometric mean. Therefore the speed of convergence would not be quick enough to conclude. Using W_q with small q would give better results, but would still

be insufficient without a better bound on N_k . Last, to tackle the case of non-uniform η we would have to use “variable rate” quantization, i.e. approximation results where the size of a discrete measure is measured by its entropy and not the cardinal of its support. In this case, bounds of the kind $N_k = o(e^{-\chi d n_k})$ do not prove μ to be singular, but only that its density (if it exists) is concentrated on small sets (it has infinite entropy relative to Lebesgue).

5.3 A case of application

Let us construct an example where Theorem 5.2 applies, proving Corollary C. We set $d = 1$, $\Phi = (\phi_i)_{i \in I}$ with $I = \{0, 1, \dots, \ell - 1\}$, $\phi_i(x) = \frac{x}{\ell} + t_i$, $t_0 = 0$, $t_1 = 1$, t_2 to be determined and t_i arbitrary for $i \geq 3$, and $\eta(\{i\}) = \frac{1}{\ell}$ for all i . The Lyapunov exponent is thus $\chi = \log \frac{1}{\ell}$.

Let $(d_k)_{k \geq 1}$ be a sequence of positive integers, assumed to increase *very* quickly, more precisely $d_{k+1} \geq k(d_k + 1)\ell^{d_k+1}$ for all k large enough; assume further that for all k , d_{k+1} is a multiple of $d_k + 1$. Finally, set

$$t_2 = \left(\sum_{j=1}^{\infty} \frac{1}{\ell^{d_j}} \right)^{-1} \in [\ell - 1, +\infty).$$

Let $\tilde{\omega}_k = 100 \cdots 0 \in I^{d_{k+1}}$ and $\omega_k = i_1^k i_2^k \cdots i_{d_{k+1}}^k \in I^{d_{k+1}}$ be defined by $i_{d_j+1}^k = 1$ for all $j \leq k$ and $i_m^k = 0$ for all other indices m . Then

$$\phi_{\omega_k}(0) = t_2 \sum_{j=1}^k \frac{1}{\ell^{d_j}} \quad \text{and} \quad \phi_{\tilde{\omega}_k}(0) = 1,$$

so that for all x and for k large enough,

$$\begin{aligned} |\phi_{\omega_k}(x) - \phi_{\tilde{\omega}_k}(x)| &= \frac{\sum_{j=k+1}^{\infty} \frac{1}{\ell^{d_j}}}{\sum_{j=1}^{\infty} \frac{1}{\ell^{d_j}}} \\ &= \frac{t_2}{\ell^{d_{k+1}}} \left(\sum_{j=k+1}^{\infty} \frac{1}{\ell^{d_j - d_{k+1}}} \right) \\ &\lesssim \frac{1}{\ell^{d_{k+1}}} = e^{\chi d_{k+1}}. \end{aligned}$$

We choose $n_k = d_{k+1}$, and since all ϕ_i are contracting we deduce that for all pairs of words $\tau_1, \tau_2 \in I^{n_k}$ which have the form $\tau_1 = \alpha \omega_k \beta$ and $\tau_2 = \alpha \tilde{\omega}_k \beta$, we also have

$$|\phi_{\tau_1}(0) - \phi_{\tau_2}(0)| = O(e^{\chi n_k}).$$

Given $\omega \in I^{n_k}$, decompose ω into n_k/d_k blocks of length d_k and replace each block equal

to $\tilde{\omega}_k$ with a copy of ω_k ; call the result ω' and set $f(\omega) = \phi_{\omega'}(0)$. This yields

$$\begin{aligned} N(\Phi, n_k, Ce^{\chi n_k}) &\leq (\ell^{d_k+1} - 1)^{\frac{n_k}{d_k+1}} \\ &\leq \ell^{n_k} e^{\frac{n_k}{d_k+1} \log(1 - \ell^{-d_k-1})} \\ &\leq \ell^{n_k} e^{-C \frac{d_k+1}{(d_k+1)\ell^{d_k+1}}} \\ &\leq \ell^{n_k} e^{-Ck} = o(\ell^{n_k}) = o(e^{-\chi n_k}). \end{aligned}$$

Theorem 5.2 applies, and we deduce that the stationary measure of (Φ, η) is not absolutely continuous.

The liberty in the choice of the tail of the sequence, $(d_k)_{k \geq k_0}$ enables to produce 2^{\aleph_0} possible values for t_2 , so that we proved Corollary C. The additional freedom given for the beginning $(d_k)_{1 \leq k \leq k_0}$ of the sequence is a coquetry that can be used to cover a wider range of values.

References

- [AGS08] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows: in metric spaces and in the space of probability measures*, Springer, 2008. [1.2](#), [4.2](#), [4.2](#)
- [BMS06] Vitaly Bergelson, Michał Misiurewicz, and Samuel Senti, *Affine actions of a free semigroup on the real line*, Ergodic Theory Dynam. Systems **26** (2006), no. 5, 1285–1305. [3.1](#)
- [BS12] Viviane Baladi and Daniel Smania, *Linear response for smooth deformations of generic nonuniformly hyperbolic unimodal maps*, Ann. Sci. Éc. Norm. Supér. (4) **45** (2012), no. 6, 861–926 (2013). MR 3075107 [1](#)
- [DF99] Persi Diaconis and David Freedman, *Iterated random functions*, SIAM review **41** (1999), no. 1, 45–76. [3.2](#)
- [DM19] Jérôme Dedecker and Florence Merlevède, *Behavior of the empirical Wasserstein distance in \mathbb{R}^d under moment conditions*, Electron. J. Probab. **24** (2019). [3.5](#)
- [Erd39] Paul Erdős, *On a family of symmetric bernoulli convolutions*, Amer. J. Math. **61** (1939), no. 4, 974–976. [1.3](#)
- [FG15] Nicolas Fournier and Arnaud Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probab. Theory Related Fields **162** (2015), no. 3-4, 707–738. [3.5](#)
- [GL00] Siegfried Graf and Harald Luschgy, *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics, vol. 1730, Springer-Verlag, Berlin, 2000. MR 1764176 (2001m:60043) [5.2](#)

- [Gol91] Charles M Goldie, *Implicit renewal theory and tails of solutions of random equations*, Ann. Appl. Probab. **1** (1991), no. 1, 126–166. [3.12](#)
- [Hoc14] Michael Hochman, *On self-similar sets with overlaps and inverse theorems for entropy*, Ann. Math. **180** (2014), 773–822. [1.3](#), [5.1](#)
- [Hut81] John E. Hutchinson, *Fractals and self-similarity*, Indiana Univ. Math. J. **30** (1981), no. 5, 713–747. MR 625600 [1](#)
- [JO10] Aldéric Joulin and Yann Ollivier, *Curvature, concentration and error estimates for Markov chain Monte Carlo*, Ann. Probab. **38** (2010), no. 6, 2418–2442. MR 2683634 [3.5](#)
- [Ken97] Richard Kenyon, *Projecting the one-dimensional Sierpinski gasket*, Israel J. Math. **97** (1997), 221–238. MR 1441250 [1.3](#)
- [Kes73] Harry Kesten, *Random difference equations and renewal theory for products of random matrices*, Acta Math. **131** (1973), no. 1, 207–248. [3.12](#)
- [Klo18a] Benoît R. Kloeckner, *Empirical measures: regularity is a counter-curse to dimensionality*, arXiv:1802.04038, 2018. [3.5](#), [3.5](#)
- [Klo18b] ———, *Extensions with shrinking fibers*, arXiv:1812.08437, 2018. [3.6.2](#)
- [NSB02] Matthew Nicol, Nikita Sidorov, and David Broomhead, *On the fine structure of stationary measures in systems which contract-on-average*, J. Theoret. Probab. **15** (2002), no. 3, 715–730. [1.1](#)
- [Oll09] Yann Ollivier, *Ricci curvature of Markov chains on metric spaces*, J. Funct. Anal. **256** (2009), no. 3, 810–864. MR 2484937 [3.5](#)
- [PSS00] Yuval Peres, Wilhelm Schlag, and Boris Solomyak, *Sixty years of bernoulli convolutions*, Fractal geometry and stochastics II, Springer, 2000, pp. 39–65. [1.3](#)
- [Rue98] David Ruelle, *General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium*, Phys. Lett. A **245** (1998), no. 3-4, 220–224. MR 1642617 [1](#)
- [Rue09] ———, *A review of linear response theory for general differentiable dynamical systems*, Nonlinearity **22** (2009), no. 4, 855–870. MR 2486360 [1](#)
- [Shm19] Pablo Shmerkin, *On Furstenberg’s intersection conjecture, self-similar measures, and the L^q norms of convolutions*, Ann. of Math. (2) **189** (2019), no. 2, 319–391. MR 3919361 [1.3](#)
- [Sol95] Boris Solomyak, *On the random series $\sum \pm \lambda^n$ (an Erdős problem)*, Ann. of Math. (2) **142** (1995), no. 3, 611–625. MR 1356783 [1.3](#), [4.3](#)

- [Var19a] Péter P. Varjú, *Absolute continuity of Bernoulli convolutions for algebraic parameters*, J. Amer. Math. Soc. **32** (2019), no. 2, 351–397. MR 3904156 [1.3](#)
- [Var19b] ———, *On the dimension of Bernoulli convolutions for all transcendental parameters*, Ann. of Math. (2) **189** (2019), no. 3, 1001–1011. MR 3961088 [1.3](#)
- [Vil09] Cédric Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new. MR MR2459454 [2](#), [3.6.2](#)