Optimal transportation and stationary measures for iterated function systems

BY BENOÎT R. KLOECKNER

Univ Paris Est Creteil, CNRS, LAMA, F-94010 Creteil, France
Univ Gustave Eiffel, LAMA, F-77447 Marne-la-Vallée, France.
e-mail: benoit.kloeckner@u-pec.fr

Abstract

In this paper 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 generalised moment bounds from which tail estimates can be deduced, consider the convergence of the empirical measure of an associated Markov chain, and 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.

---

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 $\phi_i : X \to X$ 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

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

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

$$\mu = \int (\phi_i)_* \mu \, \mathrm{d}\eta(i) \tag{1·1}$$

i.e. $\int f(x) \, \mathrm{d}\mu(x) = \iint f \circ \phi_i(x) \, \mathrm{d}\mu(x) \, \mathrm{d}\eta(i)$ for all $f \in C_b(X)$, the set of bounded continuous functions $X \to \mathbb{R}$. Such a measure is called a *stationary* measure for the pair (Φ, η), which is sometimes called an Iterated Functions System *with probabilities* but that we will

call an IFS for simplicity, since we will only consider this case. 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). As we shall see, using general Wasserstein distances one can use the fixed point theorem approach and get moment estimates at the same time.

We shall also be concerned with the attractivity of the stationary measure μ. Assume $(I_k)_{k \in \mathbb{N}}$ are independent, identically distributed random variables of law η and X_0 is a random variable independent from them. Then one constructs a Markov chain, named the "chaos game" by Barnsley [BDM$^+$88], by setting $X_{k+1} = \phi_{I_k}(X_k)$. Two sequences of measures are then to be studied: the laws of the X_k, and the "empirical measures" $1/k \sum_{j=1}^{k} \delta_{X_j}$.

The goal of this paper 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; we shall also get inspiration from the "thermodynamical formalism"; more precisely we use the transfer operator (also known as the Markov operator in the present context) in a crucial way. Results are stated and proved throughout, but provide to motivation, at the end of this introduction we give the most striking application, in the well-known example of Bernoulli convolutions.

1·1. *Brief comparison of the literature with the present results*

The stationary measures of IFS have been investigated in many times, and this overview is necessarily partial. In [BE88], one of the early works popularising the field, the existence and uniqueness of the stationary measure was proven under a hypothesis of contraction on average, and moment estimates where given. As a warm-up, we will reprove their main results using the point of view of optimal transport in Section 3. In Section 4, using the transfer operator we obtain a new generalised moment estimate, see e.g. Corollary 4·2.

Ergodicity, i.e. weak convergence of the empirical measures toward the stationary measure, is an important property as it enables a Monte Carlo Markov Chain approach to approximate the stationary measure. Ergodicity was proved in [Elt87, Elt90], see also [FM98, Sza03]. In Section 5, we give almost optimal bounds for convergence of the empirical measure in the L^1 Wasserstein distance and other related metrics, thus providing a "quantitative ergodicity" statement (Theorem 5·1). This result is new, but follows immediately from a general result on Markov chains obtained previously in [Klo18a]; it is mentioned as an illustration of the wealth of tools available.

Tail estimates for the stationary measure have been largely investigated, notably for affine IFS in [Kes73, Gol91] and more recently [Kev16]. These works use assumptions that induce a polynomial tail; in Section 4 we will be interested in quite general and easy to obtain, but less precise, tail estimates. In particular we will consider a case with exponential tail, which (obviously) does not fit Kesten's assumptions (Corollary 4·3).

In the case of two affine transformations of the line, one contracting and one expanding, [BMS06] studied more general invariant measures, where the random choice of indices need not be independent but is defined by a shift-invariant measure on $I^{\mathbb{N}}$. In Section 7, we use a modified Wasserstein distance to generalise this setting further to skew products; under a uniform contraction assumption, we get explicit convergence rates towards the stationary measure. For such generalisations, ergodicity has been studied in [Elt90, SS98].

In addition to the aforementioned [MS10], let us cite two other works studying IFS and Wasserstein distances. First [GMN16] focusses on the computation of the stationary measure, with certified bounds on the Wasserstein distance from the approximation. Second, [Fra15] computes the exact value of the Wasserstein distance between the stationary measures of different affine IFS on the line. In Section 6 we study how much the stationary measure depends on the underlying IFS; this is less precise but more general than the previous cited article. Our interest in such bounds is that they enable us to use a sophisticated Rademacher theorem for measures to obtain a linear response formula (see Corollary 1·1). While linear response has become a classical subject in dynamical systems, to my knowledge this result is the first of this kind in the context of IFS; it has one weakness and one strength: it is only proven for *almost* all parameters, but we get a derivative in quite a strong sense.

Let us conclude with some directions we do not pursue in this work. A prominent topic in the study of IFS on $\mathbb{R}^d$ is to determine the dimension of their stationary measure and whether it is absolutely continuous with respect to Lebesgue measure —this includes open questions even for some of the simplest IFS. Citing all relevant works would be daunting and, while we will use Solomyak's theorem [Sol95], we will not be primarily concerned with such questions here. We thus refer the reader to the recent survey [Var18] and references therein.

More general conditions ensuring existence, uniqueness and attractivity of the stationary measure have been sought, a notable example being the local contraction condition introduced by [Ste99], where the focus is on the behaviour of the backward iteration. For IFS satisfying this local contraction assumption, exponential convergence of the law in the Wasserstein distance to the stationary measure was proven in [MS10]. In a very general setting, without the contraction assumption, [BV11] manages to prove that the attractor of an IFS is obtained as the limit of a random orbit. The case of place-dependent probabilities is for example considered in [BDEG88, Sza03].

Inspired by the thermodynamical formalism for expanding dynamical systems, statistical properties (e.g. Central Limit Theorem, Invariance Principle) were studied in [Pei93, Pol01, Wal07, SW13]. Our strong reliance on transfer operators is quite similar in spirit, with the same inspiration from expanding dynamical systems, but this paper focuses on different results.

1·2. *Linear response for Bernoulli convolutions*

The *Bernoulli convolution* μ_β (where $\beta \in (0, 1)$) is defined as the stationary measure of the following classical IFS (Φ^λ, η):

$$I = \{0, 1\}, \qquad \eta(\{0\}) = \eta(\{1\}) = \frac{1}{2}$$

$$\phi_0^\lambda(x) = \lambda x, \qquad \phi_1^\lambda(x) = \lambda x + (1 - \lambda) \qquad \forall x \in \mathbb{R}.$$

(The precise value $1 - \lambda$ of the translation part in ϕ_1^λ has no particular relevance —as soon as it is not zero— with this value the attractor is $[0, 1]$ for all $\lambda > 1/2$, but it bears no consequence on the result below.)

We shall prove in Section 6 that the map $\lambda \mapsto \mu_\lambda$ is Lipschitz in the Wasserstein distances W_q of all exponents q; for $q > 1$, thanks to the differentiation theorem of Ambrosio, Gigli and Savaré [AGS08] this implies an almost-everywhere *linear response formula* (a terminology coined in dynamical systems, see e.g. [Rue98, Rue09, BS12]): 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. We state the result here with $q = 2$, see Corollary 6·6 for a more general result. Let Id denote the identity map of $\mathbb{R}$, so that for a function w and number x, ε we have $(\mathrm{Id} + \varepsilon w)(x) = x + \varepsilon w(x)$. Let $\varphi_* \mu$ denote the push-forward of a measure μ by a map φ.

COROLLARY 1·1. *The family of Bernoulli convolutions* $(\mu_\lambda)_{\lambda \in (0,1)}$ *is differentiable almost everywhere in the quadratic Wasserstein space, meaning that there exist a family* $(v_\lambda)_{\lambda \in (0,1)}$ *of* $L^2(\mu_\lambda)$ *functions such that for Lebesgue-almost all* $\lambda \in (1/2, 1)$:

$$W_2 \left(\mu_{\lambda+\varepsilon}, (\mathrm{Id} + \varepsilon v_\lambda)_* \mu_\lambda \right) = o(\varepsilon) \qquad as \ \varepsilon \longrightarrow 0. \tag{1·2}$$

As a consequence, there exist a family $(w_\lambda)_{\lambda \in (\frac{1}{2}, 1)}$ *of measurable functions and a Lebesgue-negligible set* E *such that for all* $\lambda \in (1/2, 1) \setminus E$ *and for all smooth compactly supported* $f : \mathbb{R} \to \mathbb{R}$,

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

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

2. *Notation and definition of Wasserstein distances*

Let us now introduce briefly the Wasserstein distances originating in optimal transportation theory. We only mention here statements that will be used several times or are relevant to several parts of the text. On several occasions below we will use results from the literature in a crucial way without giving their full statements; 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 this paper 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 optimisation in some cases. For each $q \in (0, +\infty)$, the *qth moment* of $\mu \in \mathcal{P}(X)$ is

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

The set of probability measures μ of finite qth 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:

$$C_q(\nu_0, \nu_1) = \inf_{\gamma \in \Gamma(\nu_0, \nu_1)} \int d(x, y)^q \, d\gamma(x, y)$$

$$W_q(\nu_0, \nu_1) = C_q(\nu_0, \nu_1)^{\min(1, \frac{1}{q})}.$$

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 metrises 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 qth moment of μ can be expressed as $m_{x_0}^q(\mu) = C_q(\delta_{x_0}, \mu)$.

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

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

The *Kantorovich duality* expresses that W_1 coincides with the value of a "dual" optimisation problem:

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

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 suitable assumptions, one can prove that the *dual transfer operator* associated to an IFS 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 $\mathcal{P}(X)$; Huntchinson restricts to compactly supported stationary measures). Later we shall also use this contraction property to get Lipschitz continuity of the stationary measure under perturbation of the IFS.

3·1. *Contracting dual operator*

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

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

i.e. $\mathrm{L}^* \mu$ 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 $\mathrm{L}^* \mu = \mu$.

The "transfer operator" (also known as the Markov operator), acting for example on the space of bounded measurable functions $X \to \mathbb{R}$, is defined by

$$\mathrm{L} f(x) = \int f \circ \phi_i(x) \, d\eta(i).$$

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 $a \leq \mathrm{L} f \leq b$. The duality relation between L and L^* is

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

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

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

We first recover the following slight extension of the main results of [BE88], the difference being that only finite index sets were considered there. The first hypothesis requires the IFS to be contracting on L^q average, while the second ensures that the maps do not shift some point too much (this was automatically granted in the finite index case, see Section 3·3 for examples showing the importance of this second hypothesis).

THEOREM 3·1 (variant of Barnsley, Elton 1988). *Let* (Φ, η) *be a IFS on a complete metric space* (X, d) *and fix any* $x_0 \in X$. *Assume that for some* $q > 0$, $A > 0$, $\rho \in (0, 1)$ *the following holds*:

$$\int d(\phi_i(x), \phi_i(y))^q \, \mathrm{d}\eta(i) \leq \rho \, d(x, y)^q \qquad \forall x, y \in X \tag{3·1}$$

$$\int d(x_0, \phi_i(x_0))^q \, \mathrm{d}\eta(i) \leq A. \tag{3·2}$$

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

$$m_{x_0}^q(\mu) \leq \begin{cases} \dfrac{A}{1 - \rho} & \text{when } q \leq 1 \\[3mm] \dfrac{A}{\left(1 - \rho^{\frac{1}{q}}\right)^q} & \text{when } q \geq 1. \end{cases}$$

The proof is split into several lemmas. It has some similarity with the original proof of Barnsley and Elton, our point here being to show that the Wasserstein distances are quite convenient, and enable us to use the dual transfer operator throughout the proof without introducing backward iterations; see also [Ios09] where a slightly different metric is used.

LEMMA 3·2. *The dual transfer operator preserves* $\mathcal{P}_q(X)$ *and is a contraction of ratio no more than* $\bar{\rho} := \rho^{\min\left(1, \frac{1}{q}\right)}$.

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

$$\int d(x, y)^q \, \mathrm{d}(\mathrm{L}^* \gamma)(x, y) = \iint d(\phi_i(x), \phi_i(y))^q \, \mathrm{d}\eta(i) \, \mathrm{d}\gamma(x, y)$$

$$\leq \rho \int d(x, y)^q \, \mathrm{d}\gamma(x, y)$$

so that $\mathrm{W}_q(\mathrm{L}^* \mu_0, \mathrm{L}^* \mu_1) \leq \bar{\rho} \, \mathrm{W}_q(\mu_0, \mu_1)$.

In particular, all elements of $\mathcal{P}_q(X)$ are sent a finite W_q distance from $\mathrm{L}^* \delta_{x_0}$ and we only have left to prove that $\mathrm{L}^* \delta_{x_0} \in \mathcal{P}_q(X)$, which follows from (3·2):

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

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 $\mathrm{L}^{*k} \nu \to \mu$ in the distance W_q, exponentially fast.*

Proof. Follows from Lemma 3·2 and the Banach fixed point theorem.

LEMMA 3·4. $\mathrm{L}^k f(x) \to \int f \, \mathrm{d}\mu$ *for all continuous bounded functions $f : X \to \mathbb{R}$ and for all $x \in X$.*

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

$$\mathrm{L}^k f(x) = \int \mathrm{L}^k f \, \mathrm{d}\delta_x = \int f \, \mathrm{d}(\mathrm{L}^{*k} \delta_x) \longrightarrow \int f \, \mathrm{d}\mu.$$

LEMMA 3·5. *Every stationary measure has finite qth 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

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

Let $\mu' \in \mathcal{P}(X)$ be a stationary measure, the qth moment of which we do not assume to be finite. For all $n, k \in \mathbb{N}$,

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

Since $\mathrm{L}^k f_n$ is bounded between 0 and n for all k, we can apply the dominated convergence theorem as $k \to \infty$, so that by Lemma 3·4

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

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

$$\int d(x_0, x)^q \, \mathrm{d}\mu' \le \int d(x_0, x)^q \, \mathrm{d}\mu < \infty,$$

so that $\mu' \in \mathcal{P}_q(X)$. Since both μ and μ' are stationary measures of finite qth moment, Lemma 3·3 shows that $\mu' = \mu$.

LEMMA 3·6. *The unique stationary measure μ satisfies $m_{x_0}^q(\mu) \le A/(1 - \bar{\rho})^{\max(1, q)}$.*

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

$$\mathrm{W}_q(\delta_{x_0}, \mu) \leq \mathrm{W}_q(\delta_{x_0}, L^* \delta_{x_0}) + \mathrm{W}_q(L^* \delta_{x_0}, L^* \mu)$$
$$\leq \bar{A} + \bar{\rho}\, \mathrm{W}_q(\delta_{x_0}, \mu)$$
$$(1 - \bar{\rho})\, \mathrm{W}_q(\delta_{x_0}, \mu) \leq \bar{A}.$$

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

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, AH16]. For each $\omega = (i_1, \ldots, i_n) \in \{0, 1\}^n$ we set $\phi_\omega = \phi_{i_1} \circ \phi_{i_2} \circ \cdots \phi_{i_n}$ (here the order of composition, forward or backward, has no particular relevance).

COROLLARY 3·7. *Let $a \in (0, 1)$ and $b \in (1, 1/a)$. The IFS on the line given by*

$$I = \{0, 1\}, \qquad \eta_p(\{0\}) = \frac{1}{2} \qquad \eta_p(\{1\}) = \frac{1}{2} \tag{3·3}$$
$$\phi_0(x) = ax \qquad \phi_1(x) = bx + 1 \qquad \forall x \in \mathbb{R}$$

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 $f : \mathbb{R}_+ \to \mathbb{R}$ and all x_0,

$$\left| \frac{1}{2^n} \sum_{\omega \in \{0,1\}^n} f(\phi_\omega(x_0)) - \int f \, d\mu \right| \leq C \rho^n, \tag{3·4}$$

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

We write $\mathrm{Hol}_q(f)$ for the least possible Hölder constant of f; note that 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 that it is decreasing on some interval $(0, q_1)$ and goes to $+\infty$ when $q \to +\infty$, so that this function takes the value 2 at exactly two points, 0 and q_0.

For all $q > 0$, we have (3·2) 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 $q < q_0$, (3·1) is satisfied with $\rho = (a^q + b^q)/2$. 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 follows from Lemma 3·2, observing

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

Indeed as in the proof of Lemma 3·4 we have

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

and $\mathrm{W}_q(\delta_{x_0}, \mu) = m_{x_0}^q(\mu) \leq 1/[1-\rho]$.

3·2. *Some simple tools*

We present here a few statements that prove convenient for applications of Theorem 3·1.

LEMMA 3·8. *If (Φ, η) satisfies (3·1) and (3·2), then for all $q' \in (0, q)$ it also satisfies them with constants q', $\rho' := \rho^{q'/q}$ and $A' := A^{q'/q}$.*

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

The next two complementary results enable us to reduce the conditions of Theorem 3·1 to other "contracting on average" hypotheses, including the one used for example in [DF99]. Similar lemmas can be found in [BE88] in the case of finite index set.

LEMMA 3·9. *If (Φ, η) satisfies*

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

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

Proof. The idea is simply to differentiate $\int \mathrm{Lip}(\phi_i)^t \, \mathrm{d}\eta(i)$ 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 $f_n : I \to \mathbb{R}$ and $F_n : (0, p] \to \mathbb{R}$ defined by

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

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

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

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

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

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

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

which readily implies (3·1).

LEMMA 3·10. *Assume that there exist $L \geq 1$ and $r \in (0, 1)$ such that $\mathrm{Lip}(\phi_i) \leq L$ for all $i \in I$ and*

$$\exp\left(\int \log d(\phi_i(x), \phi_i(y))\, d\eta(i)\right) \leq r d(x, y)$$

for all $x, y \in X$. Then there exists $q > 0$, $\rho \in (0, 1)$ such that (3·1) holds.

Proof. Applying the second order Taylor formula with Lagrange remainder to $q \mapsto a^q$ ensures that

$$a^q \leq 1 + q \log(a) + Cq^2$$

for all $a \in (0, L)$ and all $q \in (0, 1]$, where $C = \sup\{a(\log a)^2/2 \colon x \in (0, L]\}$. Fix $x \neq y \in X$ and for all $i \in I$ apply this to $a = d(\phi_i(x), \phi_i(y))/d(x, y)$ and integrate with respect to η to get:

$$\int \left(\frac{d(\phi_i(x), \phi_i(y))}{d(x, y)}\right)^q\, d\eta(i) \leq 1 + q \int \log\left(\frac{d(\phi_i(x), \phi_i(y))}{d(x, y)}\right)\, d\eta(i) + Cq^2$$
$$\leq 1 + q \log(r) + Cq^2,$$

where C is independent of x, y. Since $\log(r) < 0$, we can find q such that $1 + q \log(r) + Cq^2 < 1$ and we are done.

3·3. *A heavy tail of translations*

To illustrate the role of assumption (3·2) in Theorem 3·1 let us consider the following example on $[0, +\infty)$:

$$I = \mathbb{N}, \qquad \eta(\{n\}) = p_n,$$
$$\forall n > 0, \forall x \in \mathbb{R}\colon \quad \phi_n(x) = x + n, \qquad \phi_0(x) = ax; \tag{3·5}$$

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 (3·1) is satisfied for any $q > 0$ with $\rho = 1 - (1 - a^q)p_0$. As in the previous example, ρ decreases from 1 when $q \to 0$ to $1 - p_0$ when $q \to \infty$; however the translation part influences the moments of the stationary measure. By a direct application of Theorem 3·1, we get:

PROPOSITION 3·11. *Let $q > 0$; if $\sum n^q p_n < +\infty$ then (3·5) has a unique stationary measure μ, and μ has finite qth moment; if $\sum n^q p_n = +\infty$, then any stationary measure of (3·5) has infinite qth moment.*

Giving $(p_n)_n$ a heavy tail and taking $a \ll 1$ we get examples with very quick convergence in low-exponent Wasserstein distances 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 $p_n \sim c/n(\log n)^2$? Does there exist a stationary measure?

4. *Generalised moment estimates*

In some cases, the stationary measure of an IFS will not be compactly supported, but will have finite moments 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 4·1. *Let (Φ, η) be an IFS on X and $\varphi, \psi : X \to [0, +\infty)$ two functions that are bounded on every bounded subset of X (usually, they will be of the form $h(d(x_0, \cdot))$ for some function h). 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 $\varphi_m(x) = \min(\varphi(x), m)$. Since μ is stationary, $L^* \mu = \mu$ 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:

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

from which we deduce

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

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 that we can pass to the limit in the above inequality, leading precisely to the claimed inequality.

Let us now treat a specific example for which to my knowledge no precise tail estimate has been derived yet. For $p, a \in (0, 1)$, consider the IFS given by

$$
\begin{aligned}
I = \{0, 1\}, \qquad & \eta(\{0\}) = p \qquad \eta(\{1\}) = 1 - p \\
& \phi_0(x) = ax \qquad \phi_1(x) = x + 1 \qquad \forall x \in \mathbb{R}.
\end{aligned}
\tag{4·1}
$$

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* (two fixed points on the projective line, one attractive and one repulsive) 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. Moreover it does not fit into Kesten and Goldie's framework for a polynomial tail, the parabolic map sits at the frontier between contraction (then the support would be bounded) and dilation (then only some moments would be finite, and a polynomial tail would be a possibility).

Let $\mu_{a,p}$ denote the unique stationary measure of this IFS; it is concentrated on $[0, +\infty)$. Its local structure has for example been studied in [NSB02, AH16]. Theorem 3·1 shows that $\mu_{a,p}$ has finite moments of all orders. Indeed (3·1) and (3·2) are satisfied for all $q \geq 1$ with $x_0 = 0$, $\rho = 1 - p + pa^q$ (which is less than 1 since $a < 1$), $A = 1 - p$. Since the maps of the IFS are affine, the transfer operator will behave very nicely with exponentials, and we get an explicit formula for the exponential moments.

COROLLARY 4·2. *For all $p, a \in (0, 1)$ and all $b < \log(1/(1 - p))$,*

$$\int e^{bx} \, \mathrm{d}\mu_{a,p}(x) = \prod_{k=0}^{\infty} \frac{p}{1 - (1 - p)e^{a^k b}} < \infty.$$

In particular there is a constant $C(a, p)$ independent of b such that

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

Proof. Let L be the dual transfer operator and $\varphi_b(x) = e^{bx}$; then

$$\mathrm{L}\,\varphi_b(x) = p(\varphi_b(x))^a + (1 - p)e^b \varphi_b(x). \tag{4·2}$$

Pick any θ strictly between $(1 - p)e^b$ and 1: for some $C > 0$ (which could be computed explicitly), $\mathrm{L}\,\varphi_b(x) \leq \theta \varphi_b(x)$ whenever $x \geq C$. Since φ_b is increasing, $\mathrm{L}\,\varphi_b \leq \theta \varphi_b + e^{bC}$ and we can thus apply Proposition 4·1 with $\psi \equiv 1$, obtaining

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

This shows that φ_b and φ_{ab} are integrable with respect to $\mu_{a,p}$; using

$$\int \varphi_b \, \mathrm{d}\mu_{a,p} = \int \varphi_b \, \mathrm{d}\,\mathrm{L}^* \mu_{a,p} = \int \mathrm{L}\,\varphi_b \, \mathrm{d}\mu_{a,p}$$

and (4·2) we get

$$\int \varphi_b \, \mathrm{d}\mu_{a,p} = p \int \varphi_{ab} \, \mathrm{d}\mu_{a,p} + (1 - p)e^b \int \varphi_b \, \mathrm{d}\mu_{a,p}$$

$$\int \varphi_b \, \mathrm{d}\mu_{a,p} = \frac{p}{1 - (1 - p)e^b} \int \varphi_{ab} \, \mathrm{d}\mu_{a,p}$$

Applying this equality to $\varphi_{a^k b}$, an induction yields

$$\int \varphi_b \, \mathrm{d}\mu_{a,p} = \left(\prod_{k=0}^{K-1} \frac{p}{1 - (1 - p)e^{a^k b}} \right) \int \varphi_{a^K b} \, \mathrm{d}\mu_{a,p}.$$

for all $K \in \mathbb{N}$. We only have left to let $K \to \infty$: indeed $\varphi_{a^K b} \to 1$ and the monotone convergence theorem gives the desired formula.

The upper bound is then obtained by using $a^k b \leq a^k \log(1/(1-p))$ and setting

$$C(a, p) = p \prod_{k=1}^{\infty} \frac{p}{1 - (1-p)^{1-a^k}};$$

observe for the convergence that the logarithm of the kth factor is asymptotic to $((1-p)/p)(\log(1/(1-p)))a^k$.

We obtain from this an exponential tail estimate for $\mu_{a,p}$, sharp up to a linear factor.

COROLLARY 4·3. *For all $p, a \in (0, 1)$, there exist $c, C > 0$ such that for all $t \geq 1$:*

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

Proof. 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 4·2 and Chebyshev's inequality: for all $t > 0$ and all $b < \log(1/(1-p))$,

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

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

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

yielding the bound $\mu_{a,p}([t, +\infty)) \leq C(a, p)p(1 + 1/t)^t(1+t)(1-p)^t = O\big(t(1-p)^t\big)$.

PROBLEM 4·1. *Find an asymptotic for the tail of $\mu_{a,p}$, in the spirit of [Kes73] and [Gol91]. Note these works give (in a 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 4·3 already gives 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 sensitivity on t is thus strong.*

5. *Convergence rate for the empirical measure*

We now turn to the "chaos game" [BDM$^+$88]: $(X_k)_{k \in \mathbb{N}}$ is a Markov chain obtained by choosing random indices $(I_k)_{k \geq 1}$ independently 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). Quantitative results on the convergence of the laws of the X_k are abundant; formula (3·4) and Theorem 7·2 are examples, see also [MS10] for IFS satisfying a weak, local contraction assumption. We are interested here in the behaviour of the *empirical averages* (also known as *Birkhoff sums* in the field of dynamical systems)

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

for a given suitable 'observable" f, and of the empirical measure $\hat{\mu}_n$. The ergodic theorem of Elton [Elt90] states that, under hypotheses similar to those of Theorem 3·1, $\hat{\mu}_n$ converges almost surely to the stationary measure μ in the weak-$*$ topology. We shall be interested in *quantitative* ergodicity, i.e. in giving explicit estimates on the rate of convergence. For this, we will have to restrict to observables f with some regularity, and slower rates are expected for the convergence of $\hat{\mu}_n$ in the Wasserstein metric than for the empirical average of a single function. Indeed, the former gives a simultaneous control for *all* function in a certain class, and it is quite likely that the empirical averages of *some* of these deviate from the standard long-term behaviour. However, by asking for enough regularity we will be able to get a uniform rate close to $1/\sqrt{n}$, which in general cannot be surpassed even for a single f because of the Central Limit Theorem.

Before turning to this, let us mention that while $\hat{\mu}_n(f)$ cannot in general converge to $\mu(f)$ at a faster rate than $1/\sqrt{n}$, one can get strong *concentration* results, i.e. prove that $\hat{\mu}_n(f)$ is very likely to be very close to its expectation. Indeed, in a very general setting Ollivier introduced in [Oll09] Markov chains of *positive Ricci curvature*. The inspiration comes from Riemannian geometry, where positive Ricci curvature can roughly be translated to the following property: given two points x, y, the uniform measures on the balls $B(x, r)$, $B(y, r)$ are closer one to another (in Wasserstein distance) than x to y, by a factor $\rho < 1$. On such a manifold, the Markov chain that jumps from x to a uniform random point in $B(x, r)$ will thus have a unique stationary measure with exponential convergence. Ollivier's definition simply generalises this to arbitrary Markov chains on metric spaces. That the dual transfer operator is a contraction in the Wasserstein distance W_q for some $q \in (0, 1]$ is a property equivalent to $(\mathrm{X}_k)_{k \in \mathbb{N}}$ having positive Ricci curvature in the sense of Ollivier on the space (X, d^q), and implies strong concentration properties of the empirical averages $\hat{\mu}_n(f)$ whenever $f : X \to \mathbb{R}$ is a q-Hölder function; see [JO10] for effective and completely explicit results (that depend on many specific quantities that may vary between examples).

Let us consider the following metrics between measures defined on $\mathbb{R}^d$:

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

where s is any positive integer and $\mathcal{C}_1^s$ is the set of $\mathcal{C}^{s-1}$ functions $\mathbb{R}^d \to \mathbb{R}$ with all derivatives not greater than 1, and with their derivatives of order $(s - 1)$ 1-Lipschitz. In particular, $\|\cdot\|_{\mathcal{C}_1^1} = \mathrm{W}_1$. The following result shows that we can control the empirical averages of all observables of $\mathcal{C}_1^s$ simultaneously, with very good bounds when s is large enough compared to the dimension.

THEOREM 5·1. *Let (Φ, η) be an IFS on $\mathbb{R}^d$ satisfying (3·1) and (3·2) with $q = 1$ and preserving a compact domain D of $\mathbb{R}^d$.*

Let $(\mathrm{X}_k)_{k \in \mathbb{N}}$ be a Markov chain driven by (Φ, η), with $\mathrm{X}_0 \in D$, and consider the empirical measures

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

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

$$\mathbb{E}\left[\|\hat{\mu}_n - \mu\|_{\mathcal{C}_1^s}\right] \leq C \begin{cases} \dfrac{(\log n)^{\frac{d}{2s+1}}}{\sqrt{n}} & when\ 2s > d \\[2ex] \dfrac{\log n}{\sqrt{n}} & when\ 2s = d \\[2ex] \dfrac{(\log n)^{d-2s+\frac{s}{d}}}{n^{\frac{s}{d}}} & when\ 2s < d. \end{cases} \tag{5·1}$$

Proof. Formula (5·1) is the conclusion of theorem A in [Klo18a], whose hypotheses are compactness of the domain and exponential contraction of the Markov chain in the metric W_1, which follows from Lemma 3·2 with $q = 1$.

Remark 5·2. These rates cannot be improved, except possibly for the logarithm factors (see [Klo18a] for this and other considerations, including concentration bounds $\mathbb{P}[\|\hat{\mu}_n - \mu\|_{\mathcal{C}_1^s} > M_n] \leq \varepsilon_n$ for appropriate rates M_n, ε_n).

Remark 5·3. If for some $\rho \in (0, 1)$ every map ϕ_i of Φ is ρ-Lipschitz and $\{\|\phi_i(0)\| : i \in I\}$ is bounded, then there is a compact domain preserved by Φ, so that the hypotheses of Theorem 5·1 are satisfied. Indeed let $\Delta = \sup\{\|\phi_i(0)\| : i \in I\}$ and let B be the ball B of center 0 and radius $R = \Delta/(1 - \rho)$. Whenever $\|x\| \leq R$,

$$\|\phi_i(x)\| \leq \|\phi_i(x) - \phi_i(0)\| + \|\phi_i(0)\| \leq \rho\|x\| + \Delta \leq R$$

so that $\phi_i(B) \subset B$ for all i.

Remark 5·4. Relaxing the assumption to any number $q \in (0, 1]$ is possible, but needs some adaptation from [Klo18a]. Theorem A from there asks contraction of L^* in the W_1 metric, but the proof actually first reduces to contraction in the W_α for some $\alpha \in (0, 1)$. Later α is optimised, but the optimal value goes to zero when $n \to \infty$, so only the value of the constant C would affected.

Dispensing from the compactness assumption is also certainly possible, but the rates would necessarily be altered given the stationary measure might not 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).

6. *Dependence of the stationary measure on the IFS and linear response*

In this section we seek to quantify how close the stationary measures of two slightly different IFS must be. To this end, we need to introduce a way to quantify the distance between IFS; 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 might consider IFS with different index sets. 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:

$$d_{x_0}(\phi, \psi) := \min \left\{ \varepsilon \geq 0 \,\middle|\, \forall x \in X \colon d(\phi(x), \psi(x)) \leq \varepsilon + \varepsilon d(x, x_0) \right\}$$
$$= \sup_{x \in X} \frac{d(\phi(x), \psi(x))}{1 + d(x, x_0)}.$$

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

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

6·1. *Lipschitz regularity of the stationary measure*

THEOREM 6·1. *Consider two IFS* (Φ_0, η_0) *and* (Φ_1, η_1) *such that the first one satisfies (3·1) and (3·2) for some* q, ρ_0, A_0 *(and thus has a unique stationary measure* μ_0*), and the second one has at least one stationary measure* μ_1 *with finite* q*th moment* $m_{x_0}^q(\mu_1)$*. Then*

$$W_q(\mu_0, \mu_1) \leq C \, W_{x_0, q}\left((\Phi_0, \eta_0), (\Phi_1, \eta_1)\right),$$
$$\textit{where } C = \begin{cases} \dfrac{1 + m_{x_0}^q(\mu_1)}{1 - \rho_0} & \text{when } q \leq 1 \\[3mm] 2^{1 - \frac{1}{q}} \dfrac{(1 + m_{x_0}^q(\mu_1))^{\frac{1}{q}}}{1 - \rho_0^{\frac{1}{q}}} & \text{when } q > 1. \end{cases}$$

In particular, if we fix $\Phi = \Phi_0 = \Phi_1$ *and restrict to measures* η *satisfying (3·1) and (3·2) the map* $\eta \mapsto \mu$ *(which is well-defined by Theorem 3·1) is locally Lipschitz.*

Remark 6·2. When the second IFS satisfies (3·1) and (3·2) with constants q, ρ_1, A_1, we can choose to apply the result after exchanging them to optimise; however using only the moment estimate of Theorem 3·1 this is expected to only provide small improvements, since both spectral gaps (i.e. $1 - \rho_0$ and $1 - \rho_1$) are then involved in denominators.

Remark 6·3. When the second ISF has several stationary measures with finite qth moment, Theorem 6·1 shows that they *all* lie within small distance of μ_0.

Remark 6·4. For a fixed family of contractions Φ and varying probabilities η_t, Theorem 6·1 gives a Lipschitz regularity that is stronger than what can be obtained in a similar case, for the thermodynamical formalism of expanding dynamical systems (see Section 6 of [GKLM18]). When considering probabilities $\eta_t(i, x)$ that depend on the point x and a time parameter t, the family (μ_t) of stationary measures should thus not be expected to be more than Hölder-continuous with respect to t.

Let us now turn to the proof of Theorem 6·1. We denote by L_k the transfer operator of the IFS (Φ_k, η_k) $(k \in \{0, 1\})$.

LEMMA 6·5. *For all IFS* (Φ_0, η_0), (Φ_1, η_1) *and all* $\nu \in \mathcal{P}_q(X)$

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

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

Proof. If $C_{x_0, q}\big((\Phi_0, \eta_0), (\Phi_1, \eta_1)\big) = \infty$, the statement is emptily true. Assume otherwise, and let $\gamma \in \Gamma(\eta_0, \eta_1)$ be an optimal coupling. Let $\bar{\nu} = (\mathrm{Id}, \mathrm{Id})_* \nu \in \mathcal{P}(X \times X)$ 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

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

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

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

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

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

Proof of Theorem 6·1. Apply Lemma 6·5 to $\nu = \mu_1$ and use that L_0^* is a contraction (Lemma 3·2, recall $\bar{\rho}_0 = \rho_0^{\min(1, \frac{1}{q})}$):

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

6·2. *Linear response*

The Rademacher theorem ensures that Lipschitz functions $[a, b] \to \mathbb{R}$ are differentiable Lebesgue almost-everywhere; a similar result has been proven by Ambrosio, Gigli and Savaré [AGS08] for maps $[a, b] \to \mathcal{P}_q(\mathbb{R}^n)$ for $q > 1$, where one has of course to make

precise what "differentiable" means. Together with Theorem 6·1, this provides a "linear response formula" in many cases. We use $u \cdot w$ to denote the scalar product of two vectors $u, w \in \mathbb{R}^n$.

COROLLARY 6·6 (Linear Response). *Let $(\Phi_t, \eta_t)_{t \in [a,b]}$ be a curve of IFS on $\mathbb{R}^n$ (endowed with the Euclidean metric, the origin O serving as reference point), assume that for some $q > 1$:*

i. *$t \mapsto (\Phi_t, \eta_t)_t$ is Lipschitz in $W_{O,q}$;*
ii. *there exist $\rho_+ \in (0, 1)$ and $A_+ > 0$ such that for all $t \in [a, b]$ the IFS (Φ_t, η_t) satisfies hypotheses (3·1) and (3·2) with parameters q, $\rho_t \le \rho_+$ and $A_t \le A_+$ 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;*
iii. *for Lebesgue almost all t, $\|v_t\| \in L^q(\mu_t)$, and $|v_t|^{q-2}v_t$ can be approximated by gradients of smooth functions $\mathbb{R}^n \to \mathbb{R}$ in the $L^{q'}(\mu_t)$ norm where $q' = q/(q-1)$;*
iv. *$\mathrm{d}\mu_t/\mathrm{d}t + \nabla \cdot (v_t \mu_t) = 0$ weakly on $\mathbb{R} \times \mathbb{R}^n$, i.e. for all smooth compactly supported $F : \mathbb{R} \times \mathbb{R}^n$;*

$$\int_{\mathbb{R}} \int_{\mathbb{R}^n} \left(\frac{\mathrm{d}}{\mathrm{d}t} F(t, x) + \nabla_x F(t, x) \cdot v_t(x) \right) \mathrm{d}\mu_t \, \mathrm{d}t = 0;$$

v. *for Lebesgue almost all t_0,*

$$W_q \left(\mu_{t+\varepsilon}, (\mathrm{Id} + \varepsilon v_t)_* \mu_t \right) = o(\varepsilon).$$

As a consequence of v, *at almost every $t_0 \in [a, b]$, for all compactly supported smooth functions $f : \mathbb{R}^n \to \mathbb{R}$:*

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

Note that here $q > 1$ is needed to ensure strict convexity in the optimal transport problem. Property iii may seem rather exotic; it is more easily explained when $q = 2$: the approximation of v_t by gradient of smooth functions is a way to formulate that the "curl with respect to μ_t" of this vector field vanishes, which relates to optimality in an "infinitesimal" transport problem. On $\mathbb{R}$ it is vacuous but in higher dimension it is important as it ensures uniqueness of v_t.

The proof is only an application of some results in [AGS08], but we detail some classical arguments to better show how optimal transportation is related to linear response formulas.

Proof. By Theorem 6·1, the family of stationary measures $(\mu_t)_{t \in [a,b]}$ is Lipschitz, in particular absolutely continuous. Thus theorem 8.3.1 of [AGS08] applies, giving iii and iv (note the formulation (8.1.4) for the interpretation of the continuity equation, and see above definition 5.1.11 that $\mathrm{Cyl}(\mathbb{R}^n)$ is the space of smooth compactly supported functions). Proposition 8.4.6 of [AGS08] gives v and we are left with proving the differentiation formula for $\int f \, \mathrm{d}\mu_t$.

If we fix f, the weak derivative given in iv is sufficient to obtain the derivative of $t \mapsto \mu_t(f)$ almost everywhere; but it could be that the negligible set for which the formula fails turns out to depend on f. We therefore use v: fix any t at which it holds and f a smooth compactly supported function. By Jensen's inequality, since $q > 1$ we

have $W_1(\mu_{t+\varepsilon}, (\mathrm{Id} + \varepsilon v_t)_* \mu_t) = o(\varepsilon)$, and since f is Lipschitz the dual formulation of the Wasserstein distance yields

$$\int f \, d\mu_{t+\varepsilon} = \int f(x + \varepsilon v_t(x)) \, d\mu_t + o(\varepsilon).$$

The second order Taylor formula ensures that $f(x + \varepsilon v_t(x)) = f(x) + \varepsilon v_t(x) \cdot \nabla f(x) + O(\varepsilon^2 v_t(x)^2)$ where the implied constant is uniform in x. When $q \geq 2$, we thus get

$$\int f \, d\mu_{t+\varepsilon} = \int f \, d\mu_t + \varepsilon \int \nabla f \cdot v_t \, d\mu_t + O\left(\varepsilon^2 \int v_t^2 \, d\mu_t\right) + o(\varepsilon)$$

with $v_t \in L^2(\mu_t)$, giving the desired derivative at t. When $q \in (1, 2)$, we argue as follows.

Let $\alpha \in (1, 3 - 2/q)$ and set $B_\varepsilon = \{x \in \mathbb{R}^n \mid \varepsilon^2 v_t(x)^2 > \varepsilon^\alpha v_t(x)^q\}$ and $G_\varepsilon = \mathbb{R}^n \setminus B_\varepsilon$. By Chebyshev's inequality,

$$\mu_t(B_\varepsilon) = \mu_t(\{v^q > \varepsilon^{-(2-\alpha)\frac{q}{2-q}}\}) \leq \varepsilon^{(2-\alpha)\frac{q}{2-q}} \int v_t^q \, d\mu_t = O(\varepsilon^\beta) = o(\varepsilon),$$

where $\beta = (2 - \alpha)a/(2 - q)$. It follows:

$$
\begin{aligned}
\int f \, d\mu_{t+\varepsilon} &= \int_{B_\varepsilon} f(x + \varepsilon v_t(x)) \, d\mu_t + \int_{G_\varepsilon} f(x + \varepsilon v_t(x)) \, d\mu_t + o(\varepsilon) \\
&= O(\mu_t(B_\varepsilon)) + \int_{G_\varepsilon} f \, d\mu_t + \varepsilon \int_{G_\varepsilon} \nabla f \cdot v_t \, d\mu_t + \varepsilon^\alpha \int_{G_\varepsilon} v_y^q \, d\mu_t + o(\varepsilon) \\
&= \int f \, d\mu_t + \varepsilon \int \nabla f \cdot v_t \, d\mu_t + O(\mu_t(B_\varepsilon)) + o(\varepsilon) \\
&= \int f \, d\mu_t + \varepsilon \int \nabla f \cdot v_t \, d\mu_t + o(\varepsilon)
\end{aligned}
$$

as desired.

6·3. *The case of Bernoulli convolutions*

Corollary 1·1 will follow from Corollary 6·6. We take as reference point $x_0 = O = 0 \in \mathbb{R}$; recall that the family (Φ^λ, η) of IFS defining Bernoulli convolutions is given in Section 1·2.

We start by the following explicit Lipschitz estimate.

PROPOSITION 6·7. *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. Given any $q \geq 1$ and $\lambda, \lambda' \in (0, 1)$,

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

Considering the identity coupling of η with itself, defined by $\gamma(\{(0, 0)\}) = \gamma(\{(1, 1)\}) = 1/2$ and $\gamma(\{(0, 1)\}) = \gamma(\{(1, 0)\}) = 0$, we get $W_{0,q}\left((\Phi^\lambda, \eta), (\Phi^{\lambda'}, \eta)\right) \leq |\lambda - \lambda'|$; moreover (Φ^λ, η) satisfies (3·1) and (3·2) with constants $\rho = \lambda^q$ and $A = 1$; and $m_{x_0}^q(\mu_{\lambda'}) \leq 1$.

Theorem 6·1 then ensures that

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

Proof of Corollary 1·1. Fix any $q > 1$; as seen above, $W_{0,q}\left((\Phi^\lambda, \eta), (\Phi^{\lambda'}, \eta)\right) \le |\lambda - \lambda'|$ and in restriction to any interval $[1/2, 1 - \varepsilon]$ with $\varepsilon > 0$ we have (3·1) and (3·2) with uniform bounds $\rho_+ = (1 - \varepsilon)^q$ and $A_+ = 1$.

Corollary 6·6 provides us with a family $(v_\lambda)_{\lambda \in (0,1)}$ of vector fields on $\mathbb{R}$, which can be identified with $\|v_\lambda\| \in L^q(\mu_\lambda)$ functions, such that for Lebesgue-almost all $\lambda \in (0, 1)$, $W_q(\mu_{\lambda+\varepsilon}, (\mathrm{Id} + \varepsilon v_\lambda)_*\mu_\lambda) = o(\varepsilon)$. When $\lambda > 1/2$, up to further restricting to a subset of full Lebesgue measure for the parameter λ, Solomyak's theorem [Sol95] ensures that μ_λ is absolutely continuous with respect to the Lebesgue measure; let us denote its density by g_λ. For almost all $\lambda \in (1/2, 1)$ and all smooth compactly supported test function $f : \mathbb{R} \to \mathbb{R}$ we get

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

which is the desired formula with $w_\lambda = v_\lambda g_\lambda$. Moreover, for almost-all $\lambda > 1/\sqrt{2}$, g_λ is bounded ([Sol95], Corollary 1), implying $w_\lambda \in L^q(\mu_\lambda)$ and then $w_\lambda \in L^q([0, 1])$.

Now, w_λ seems to depend on the choice of q. But if $\tilde{w}_\lambda$ 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 $\tilde{w}_\lambda$ must differ by a constant, and we thus must have $w_\lambda = \tilde{w}_\lambda$. It follows that there is a single w_λ, belonging to all $L^q([0, 1])$ simultaneously.

7. *Stationary measures beyond products*

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

7·1. *Skew-products*

We still consider (X, d) a complete metric space, and we additionally fix a standard measure 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 $S : Y \to Y$. A *skew-product map* over S with fiber X is a map

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

where $(x, y) \mapsto \psi_y(x)$ is a measurable map. While, as we have seen above, an IFS can be studied dynamically by looking at a random orbit X_0, $X_{n+1} = \phi_{I_n}(X_n)$ where $(I_n)_{n \ge 1}$ 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, IFS 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 exists a measure $\hat{\nu} \in \mathcal{P}(X \times Y)$ such that:

$$\hat{\nu} \text{ is } \Psi\text{-invariant}, \qquad \pi^Y_* \hat{\nu} = \nu, \quad \text{and} \quad \pi^X_* \hat{\nu} = \mu.$$

In the case of an IFS, this coincides with the previous definition of stationary measure. The measure $\hat{\nu}$ as above will 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 7·1. We say that Ψ *contracts the fibers* whenever there exist $\rho \in (0, 1)$ such that for all $y \in Y$, the map ψ_y is ρ-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 7·2. *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 $k \in \mathbb{N}$,

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

where

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

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

7·2. *Fiber-wise Wasserstein distance*

The main tool to prove Theorem 7·2 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,

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. Recalling that we fixed a point $x_0 \in X$, given any $\sigma \in \mathcal{P}^\nu$ and $q > 0$ we define its qth moment by

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

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 $\mathcal{P}_q^\nu$ be the subset of $\mathcal{P}^\nu$ consisting of measures of finite qth 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

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

For all $\sigma_0, \sigma_1 \in \mathcal{P}^\nu$ let $\Gamma^\nu(\sigma_0, \sigma_1) := \{\gamma \in \mathcal{P}(X^2 \times Y) \mid (\pi_{02*})\gamma = \sigma_0$ and $(\pi_{12*})\gamma = \sigma_1\}$ (playing the role of Γ_π in [Klo18b], where we had chosen to emphasize the projection map rather than the image measure) and define

$$C_q^\nu(\sigma_0, \sigma_1) = \inf_{\gamma \in \Gamma^\nu(\sigma_0, \sigma_1)} \int d(x, x')^q \, d\gamma(x, x', y)$$

$$W_q^\nu(\sigma_0, \sigma_1) = \left(C_q^\nu(\sigma_0, \sigma_1) \right)^{\min(1, \frac{1}{q})}.$$

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

PROPOSITION 7·3. *For all $\sigma_0, \sigma_1 \in \mathcal{P}^\nu$, the set $\Gamma^\nu(\sigma_0, \sigma_1)$ is non-empty. If the moments $m_{x_0}^q(\sigma_i)$ are finite for $i \in \{0, 1\}$, then $W^\nu(\sigma_0, \sigma_1) < \infty$. Moreover, if $(\xi_y)_{y \in Y}$ and $(\zeta_y)_{y \in Y}$ are the disintegrations of σ_0 and σ_1 with respect to π^Y, then*

$$W_q^\nu(\sigma_0, \sigma_1) = \begin{cases} \left(\int W_q(\xi_y, \zeta_y)^q \, d\nu(y) \right)^{\frac{1}{q}} & \text{when } q \geq 1 \\[2mm] \int W_q(\xi_y, \zeta_y) \, d\nu(y) & \text{when } q \leq 1. \end{cases} \qquad (7·1)$$

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

Proof. Let $(\xi_y)_{y \in Y}$ and $(\zeta_y)_{y \in Y}$ be the disintegrations of σ_0 and σ_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 characterised by $\int f(x, y) \, d\xi_y(x) \, d\nu(y) = \int f(x, y) \, d\sigma_0(x, y)$ for all continuous bounded functions $f : X \times Y \to \mathbb{R}$.)

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 $\Gamma^\nu(\sigma_0, \sigma_1)$ by setting $\int f(x, x', y) \, d\gamma(x, x', y) = \int\int f(x, x', y) \, d\gamma_y(x, x') \, d\nu(y)$. In particular $\Gamma^\nu(\sigma_0, \sigma_1)$ is non-empty.

Conversely, given any $\gamma \in \Gamma^\nu(\sigma_0, \sigma_1)$ 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 $\int d(x, x')^q \, \mathrm{d}\gamma(x, x', y) = \iint d(x, x')^q \, \mathrm{d}\gamma_y(x, x') \, \mathrm{d}\nu(y) \geq \int \mathrm{C}_q(\xi_y, \zeta_y) \, \mathrm{d}\check{\mu}(y)$, taking an infimum we get $\mathrm{C}_q^\nu(\mu_0, \mu_1) \geq \int \mathrm{C}_q(\xi_y, \zeta_y) \, \mathrm{d}\check{\mu}$.

For each y, the set of optimal transport plans from ξ_y to ζ_y is compact (see e.g. [Vil09, proof of Theorem 4.1]), thus by the measurable selection theorem there is a measurable family $(\gamma_y)_{y \in Y}$ such that for ν-almost all $y \in Y$, $\int d(x, x')^q \, \mathrm{d}\gamma_y(x, x') = \mathrm{C}_q(\xi_y, \zeta_y)$. It follows $\mathrm{C}_q^\nu(\sigma_0, \sigma_1) \leq \int \mathrm{C}_q(\xi_y, \zeta_y) \, \mathrm{d}\nu$ and (7·1) is proven.

To complete the proof, it remains to be seen that $\mathrm{C}_q^\nu(\sigma_0, \sigma_1) < \infty$ and that W_q^ν is a metric making $\mathcal{P}_q^\nu$ a complete space. The triangular inequality follows from (7·1), and then finiteness is obtained by observing

$$\mathrm{W}_q^\nu(\sigma_0, \sigma_1) \leq \mathrm{W}_q^\nu(\sigma_0, \delta_{x_0} \otimes \nu) + \mathrm{W}_q^\nu(\delta_{x_0} \otimes \nu, \sigma_1) = m_{x_0}^q(\sigma_0) + m_{x_0}^q(\sigma_1).$$

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

7·3. *Proof of Theorem 7·2*

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 $\mathcal{P}_q^\nu$ 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 $R = (1 + \varepsilon)A/(1 - \rho)$ and let $B = B(x_0, R)$ be the ball of center x_0 and radius R in X; then for all $x, y \in X \times Y$

$$d(x_0, \psi_y(x)) \leq d(x_0, \psi_y(x_0)) + d(\psi_y(x_0), \psi_y(x))$$
$$\leq A + \rho d(x_0, x).$$

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

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

where $(1 + \varepsilon\rho)/(1 + \varepsilon) < 1$, proving the absorbing property with $k \simeq \log d(x_0, x)$.

Let $\sigma_0, \sigma_1 \in \mathcal{P}_q^\nu$. We consider the map $X^2 \times Y \to X^2 \times Y$ 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 $\gamma \in \Gamma^\nu(\sigma_0, \sigma_1)$ we have $\Psi_{2*}\gamma \in \Gamma^\nu(\Psi_*\sigma_0, \Psi_*\sigma_1)$. Observing

$$\int d(x, x')^q \, \mathrm{d}\Psi_{2*}\gamma(x, x', y) = \int d(\psi_y(x), \psi_y(x'))^q \, \mathrm{d}\gamma(x, x', y)$$
$$\leq \rho^q \int d(x, x')^q \, \mathrm{d}\gamma(x, x', y)$$

and taking an infimum, we see that

$$\mathrm{W}_q^\nu(\Psi_*\sigma_0, \Psi_*\sigma_1) \leq \rho^{\min(1,q)} \, \mathrm{W}_q^\nu(\sigma_0, \sigma_1),$$

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 qth 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 $\mathcal{P}^\nu$, and that its first marginal μ is supported on a bounded set. Explicitly, by letting ε above go to 0, we obtain that μ is concentrated on $B(x_0, A/(1-\rho))$.

Consider now the stochastic process $(X_k)_{k \in \mathbb{N}}$. Let $\sigma_0 := \tilde\mu_0 \otimes \nu$ be the law of (X_0, Y); then the law of $(X_k, S^k(Y)) = \Psi^k(X_0, Y)$ is $\sigma_k = \Psi_*^k(\sigma_0)$, by definition has first marginal $\tilde\mu_k$, and by invariance has second marginal ν. Since Ψ_* is a contraction in $\mathcal{P}_q^\nu \ni \sigma_0$, we obtain that

$$W_q^\nu(\sigma_k, \hat\nu) \le \rho^{k \min(1,q)} W_q^\nu(\sigma_0, \hat\nu). \tag{7.2}$$

On the first hand, using the transport plan obtained by projecting an optimal $\gamma \in \Gamma^\nu(\sigma_k, \hat\nu)$ on the first two variables, we get $W_q(\tilde\mu_k, \mu) \le W_q^\nu(\sigma_k, \hat\nu)$. On the other hand,

$$\begin{aligned} W_q^\nu(\sigma_0, \hat\nu) &\le W_q^\nu(\tilde\mu_0 \otimes \nu, \delta_{x_0} \otimes \nu) + W_q^\nu(\delta_{x_0} \otimes \nu, \hat\nu) \\ &\le m_{x_0}^q(\tilde\mu_0)^{\min(1, \frac{1}{q})} + \left(A/(1-\rho)\right)^{\min(1,q)} \end{aligned}$$

since $\hat\nu$ is concentrated on $B(x_0, A/(1-\rho)) \times Y$. Together with (7.2), this concludes the proof of Theorem 7.2.

REFERENCES

[AGS08] L. AMBROSIO, N. GIGLI and G. SAVARÉ. Gradient flows: in metric spaces and in the space of probability measures. (Springer, 2008).

[AH16] A. ANCKAR and G. HÖGNÄS. The fine structure of the stationary distribution for a simple Markov process. *Probability on algebraic and geometric structures, Contemp. Math.*, vol. 668 (Amer. Math. Soc., Providence, RI, 2016), pp. 1–12.

[BDEG88] M. F. BARNSLEY, S. G. DEMKO, J. H. ELTON and J. S. GERONIMO. Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities. *Ann. Inst. H. Poincaré Probab. Statist.* **24** (1988), no. 3, 367–394.

[BDM+88] M. F. BARNSLEY, R. L. DEVANEY, B. B. MANDELBROT, H.-O. PEITGEN, D. SAUPE and R. F. VOSS. The science of fractal images, (Springer-Verlag, New York, 1988), With contributions by Yuval Fisher and Michael McGuire.

[BE88] M. F. BARNSLEY and J. H. ELTON. A new class of Markov processes for image encoding. *Adv. in Appl. Probab.* **20** (1988), no. 1, 14–32.

[BMS06] V. BERGELSON, M. MISIUREWICZ and S. SENTI. Affine actions of a free semigroup on the real line. *Ergodic Theory Dynam. Systems* **26** (2006), no. 5, 1285–1305.

[BS12] V. BALADI and D. SMANIA. Linear response for smooth deformations of generic nonuniformly hyperbolic unimodal maps. *Ann. Sci. Éc. Norm. Sup.* (4) **45** (2012), no. 6, 861–926 (2013).

[BV11] M. F. BARNSLEY and A. VINCE. The chaos game on a general iterated function system. *Ergodic Theory Dynam. Systems* **31** (2011), no. 4, 1073–1079.

[DF99] P. DIACONIS and D. FREEDMAN. Iterated random functions. *SIAM review* **41** (1999), no. 1, 45–76.

[DM19] J. DEDECKER and F. MERLEVÈDE. Behavior of the empirical Wasserstein distance in $\mathbb{R}^d$ under moment conditions. *Electron. J. Probab.* **24** (2019).

[Elt87] J. H. ELTON. An ergodic theorem for iterated maps. *Ergodic Theory Dynam. Systems* **7** (1987), no. 4, 481–488.

[Elt90] J. H. ELTON. A multiplicative ergodic theorem for Lipschitz maps. *Stochastic Process. Appl.* **34** (1990), no. 1, 39–47.

[FG15] N. FOURNIER and A. GUILLIN. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** (2015), no. 3-4, 707–738.

[FM98] B. FORTE and F. MENDIVIL. A classical ergodic property for IFS: a simple proof. *Ergodic Theory Dynam. Systems* **18** (1998), no. 3, 609–611.

[Fra15]   J. M. FRASER. First and second moments for self-similar couplings and Wasserstein distances. *Math. Nachr.* **288** (2015), no. 17-18, 2028–2041.

[GKLM18] P. GIULIETTI, B. KLOECKNER, A. O. LOPES and D. MARCON. The calculus of thermodynamical formalism. *J. Eur. Math. Soc.* (JEMS) **20** (2018), no. 10, 2357–2412.

[GMN16]   S. GALATOLO, M. MONGE and I. NISOLI. Rigorous approximation of stationary measures and convergence to equilibrium for iterated function systems. *J. Phys. A* **49** (2016), no. 27, 274001, 22.

[Gol91]   C. M. GOLDIE. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1** (1991), no. 1, 126–166.

[Hut81]   J. E. HUTCHINSON. Fractals and self-similarity. *Indiana Univ. Math. J.* **30** (1981), no. 5, 713–747.

[Ios09]   M. IOSIFESCU. Iterated function systems. *A critical survey, Math. Rep. (Bucur.)* **11(61)** (2009), no. 3, 181–229.

[JO10]   A. JOULIN and Y. OLLIVIER. Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.* **38** (2010), no. 6, 2418–2442.

[Kes73]   H. KESTEN. Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131** (1973), no. 1, 207–248.

[Kev16]   P. KEVEI. A note on the Kesten-Grincevičius-Goldie theorem. *Electron. Commun. Probab.* **21** (2016), Paper No. 51, 12.

[Klo18a]   B. R. KLOECKNER. Empirical measures: regularity is a counter-curse to dimensionality. arXiv:1802.04038, (2018).

[Klo18b]   B. R. KLOECKNER. Extensions with shrinking fibers. arXiv:1812.08437, (2018).

[MS10]   N. MADRAS and D. SEZER. Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances. *Bernoulli* **16** (2010), no. 3, 882–908.

[NSB02]   M. NICOL, N. SIDOROV and D. BROOMHEAD. On the fine structure of stationary measures in systems which contract-on-average. *J. Theoret. Probab.* **15** (2002), no. 3, 715–730.

[Oll09]   Y. OLLIVIER. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256** (2009), no. 3, 810–864.

[Pei93]   M. PEIGNÉ. Iterated function systems and spectral decomposition of the associated Markov operator. *Fascicule de probabilités, Publ. Inst. Rech. Math. Rennes, vol. 1993, Univ. Rennes I, Rennes*, 1993, p. 28.

[Pol01]   M. POLLICOTT. Contraction in mean and transfer operators. *Dyn. Syst.* **16** (2001), no. 1, 97–106.

[Rue98]   D. 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.

[Rue09]   D. RUELLE. A review of linear response theory for general differentiable dynamical systems. *Nonlinearity* **22** (2009), no. 4, 855–870.

[Sol95]   B. SOLOMYAK. On the random series $\sum \pm \lambda^n$ (an Erdős problem). *Ann. of Math.* (2) **142** (1995), no. 3, 611–625.

[SS98]   D. S. SILVESTROV and Ö. STENFLO. Ergodic theorems for iterated function systems controlled by regenerative sequences. *J. Theoret. Probab.* **11** (1998), no. 3, 589–608.

[Ste99]   D. STEINSALTZ. Locally contractive iterated function systems. *Ann. Probab.* **27** (1999), no. 4, 1952–1979.

[SW13]   S. I. SANTOS and C. WALKDEN. Distributional and local limit laws for a class of iterated maps that contract on average. *Stoch. Dyn.* **13** (2013), no. 2, 1250019, 28.

[Sza03]   T. SZAREK. Invariant measures for nonexpensive Markov operators on Polish spaces. Dissertationes Math. (Rozprawy Mat.) **415** (2003), 62, Dissertation, Polish Academy of Science, Warsaw, (2003).

[Var18]   P. P. VARJÚ. Recent progress on Bernoulli convolutions, European Congress of Mathematics. *Eur. Math. Soc., Zürich*, (2018), pp. 847–867.

[Vil09]   C. VILLANI. Optimal transport, Grund. Math. Wiss. [Fundamental Principles of Mathematical Sciences], vol. 338 (Springer-Verlag, Berlin, 2009), Old and new.

[Wal07]   C. P. WALKDEN. Invariance principles for iterated maps that contract on average. *Trans. Amer. Math. Soc.* **359** (2007), no. 3, 1081–1097.