STUDIA MATHEMATICA 205 (3) (2011)

The rate of convergence for iterated function systems

by

Maciej Ślęczka (Katowice)

Abstract. Iterated function systems with place-dependent probabilities are considered. It is shown that the rate of convergence of transition probabilities to a unique invariant measure is geometric.

1. Introduction. We consider the evolution of distributions due to the action of randomly chosen transformations, so called iterated function systems (briefly IFS) with place-dependent probabilities. Iterated function systems are a fast developing topic with applications to different areas such as fractals [1, 2, 8, 13] and learning models [4, 11]. The reader is referred to [3, 10, 19] for historical remarks.

In [3] Barnsley et al. proved the existence and uniqueness of an attractive invariant measure for IFS on locally compact spaces. Szarek [20, 21] generalized this result to Polish spaces. In this paper, we are interested in the rate of convergence to the invariant measure. If the probabilities are constant then this rate is known to be geometric, since, as shown by Lasota [12], the corresponding Markov operator is contractive in the Wasserstein norm. A similar result can be found in [16]. In the general case of place-dependent probabilities this idea breaks down. Our approach is based on the coupling technique (see [22] and [6, 7, 15, 25]).

The organization of the paper is as follows. In Section 2 we introduce notation and give some basic definitions. Section 3 contains our results and is divided into several parts. In Section 3.1 we formulate the main theorem (Theorem 3.1) which states that the iterated function systems under consideration are mixing at a geometric rate. In Sections 3.2 and 3.3 we construct a coupling for IFS and formulate some technical lemmas. Finally, in Section 3.4 we prove the main theorem.

2010 *Mathematics Subject Classification*: Primary 60J05; Secondary 37A25.
Key words and phrases: iterated function system, invariant measure, rate of convergence, coupling.

DOI: 10.4064/sm205-3-1

2. Preliminaries

2.1. Notation and basic definitions. Let (X,d) be a *Polish space*, i.e. a separable and complete metric space. We denote by $B(X)$ the space of all bounded Borel measurable functions $f : X \to \mathbb{R}$, equipped with the supremum norm, and by $\mathcal{B}_X$ the family of all Borel subsets of X. Let $\mathcal{M}_{\mathrm{fin}}(X)$ and $\mathcal{M}_1(X)$ be the sets of nonnegative measures on X such that $\mu(X) < \infty$ and $\mu(X) = 1$, respectively. The elements of $\mathcal{M}_1(X)$ are called *probability measures*. The elements of $\mathcal{M}_{\mathrm{fin}}(X)$ such that $\mu(X) \le 1$ are called *subprobability measures*. We denote by $\mathcal{M}_{\mathrm{sig}}(X)$ the family of all signed measures,

$$\mathcal{M}_{\mathrm{sig}}(X) = \{\mu_1 - \mu_2 : \mu_1, \mu_2 \in \mathcal{M}_{\mathrm{fin}}(X)\}.$$

Fix $\bar{x} \in X$. Let $\mathcal{M}_1^1(X)$ denote the set of all probability measures on X with the first moment finite:

$$\mathcal{M}_1^1(X) = \Big\{\mu \in \mathcal{M}_1(X) : \int_X d(x,\bar{x})\,\mu(dx) < \infty\Big\}.$$

By the triangle inequality this family is independent of the choice of $\bar{x}$.

To simplify the notation we will write

$$\langle f, \mu\rangle = \int_X f(x)\,\mu(dx) \quad \text{for } f \in B(X),\ \mu \in \mathcal{M}_{\mathrm{sig}}(X).$$

Let δ_x be the *Dirac measure* concentrated at $x \in X$.

For $\mu \in \mathcal{M}_{\mathrm{fin}}(X)$ the *support* of μ is

$$\operatorname{supp}\mu = \{x \in X : \mu(B(x,r)) > 0 \text{ for every } r > 0\},$$

where $B(x,r)$ is the open ball in X with center at x and radius r.

If μ is a measure on X, Y is a measurable space and $f : X \to Y$ is a measurable transformation, then $f^*\mu$ denotes the measure on Y defined by

$$f^*\mu(B) = \mu(f^{-1}(B)) \quad \text{for } B \in \mathcal{B}_Y.$$

Spaces of measures on X are equipped with several norms. For $\mu \in \mathcal{M}_{\mathrm{sig}}(X)$ we denote by $\|\mu\|$ the total variation of μ. In particular, if μ is nonnegative then $\|\mu\|$ is the total mass of μ.

In $\mathcal{M}_{\mathrm{sig}}(X)$ we also introduce the *Fortet–Mourier norm*

$$\|\mu\|_{\mathcal{L}} = \sup\{|\langle f, \mu\rangle| : f \in \mathcal{L}\},$$

where $\mathcal{L} = \{f \in B(X) : |f(x) - f(y)| \le 1 \text{ and } |f(x)| \le 1 \text{ for } x, y \in X\}$. The space $\mathcal{M}_1(X)$ with the metric $\|\mu_1 - \mu_2\|_{\mathcal{L}}$ is complete (see [5]).

In the space $\mathcal{M}_{\mathrm{sig}}^1(X)$ of all signed measures with the first moment finite we introduce the *Wasserstein norm*

$$\|\mu\|_{\mathcal{H}} = \sup\{|\langle f, \mu\rangle| : f \in \mathcal{H}\},$$

where $\mathcal{H} = \{f \in B(X) : |f(x) - f(y)| \le d(x,y) \text{ for } x, y \in X\}$. The space $\mathcal{M}_1^1(X)$ with the metric $\|\mu_1 - \mu_2\|_{\mathcal{H}}$ is complete (see [23]). We have

$$\|\mu\|_{\mathcal{L}} \le \|\mu\|_{\mathcal{H}} \quad \text{for } \mu \in \mathcal{M}_1^1(X).$$

2.2. Measures on the pathspace. Coupling. We consider a family $\{\mathbf{Q}_x : x \in X\}$ of measures on X. We assume that for every Borel subset $A \subset X$ the function $x \mapsto \mathbf{Q}_x(A)$ is measurable. If $n, m \in \mathbb{N}$ are fixed, and $\{\mathbf{Q}_x : x \in X\}$ is a family of measures on X^n and $\{\mathbf{R}_x : x \in X\}$ a family of measures on X^m, we can define a family $\{\mathbf{RQ}_x : x \in X\}$ of measures on $X^n \times X^m$ by

$$(\mathbf{RQ}_x)(A \times B) = \int_A \mathbf{R}_{z_n}(B)\, \mathbf{Q}_x(dz), \tag{1}$$

where $A \in \mathcal{B}_{X^n}$, $B \in \mathcal{B}_{X^m}$ and $z = (z_1, \ldots, z_n) \in X^n$.

Let $\{\mathbf{P}_x : x \in X\}$ be a family of probability or, more generally, subprobability measures on X. We assume that for every Borel subset $A \subset X$ the transformation $x \mapsto \mathbf{P}_x(A)$ is measurable. If, in addition, all $\mathbf{P}_x$ are probability measures then $\{\mathbf{P}_x : x \in X\}$ is a *transition probability function*. For given $x \in X$ and $A \in \mathcal{B}_X$, $\mathbf{P}_x(A)$ is the probability of transition from x to A.

We define a family of measures on the *pathspace* X^∞ in the following way. For fixed $x \in X$ one-dimensional distributions $\{\mathbf{P}_x^n : n \in \mathbb{N}_0\}$ are defined inductively:

$$\mathbf{P}_x^0 = \delta_x, \qquad \mathbf{P}_x^{n+1}(\cdot) = \int_X \mathbf{P}_z(\cdot)\, \mathbf{P}_x^n(dz). \tag{2}$$

Two- and higher-dimensional distributions are constructed by iterating (1). We denote the resulting family of subprobability measures on X^∞ by $\{\mathbf{P}_x^\infty : x \in X\}$.

Suppose that $\{\mathbf{P}_x : x \in X\}$ is the *transition probability function for a Markov chain* $(\xi_n)_{n\in\mathbb{N}_0}$ taking values in X, i.e. if ν is the distribution of ξ_n then $\int_X \mathbf{P}_x(\cdot)\, \nu(dx)$ is the distribution of ξ_{n+1}. Then the measure $\mathbf{P}_x^\infty$ is the distribution in X^∞ of the chain $(\xi_n)_{n\in\mathbb{N}_0}$ assuming that $\xi_0 = x$.

For any subprobability measure $\mu \in \mathcal{M}_{\mathrm{fin}}(X)$ we define on X^∞ the measure

$$\mathbf{P}_\mu^\infty(\cdot) = \int_X \mathbf{P}_x^\infty(\cdot)\, \mu(dx).$$

The one-dimensional distributions of $\mathbf{P}_\mu^\infty$ are denoted by $\mathbf{P}_\mu^n$ for $n \in \mathbb{N}_0$.

If μ is the starting distribution of a Markov chain $(\xi_n)_{n\in\mathbb{N}_0}$ with transition function $\{\mathbf{P}_x : x \in X\}$ then $\mathbf{P}_\mu^\infty$ is the distribution of $(\xi_n)_{n\in\mathbb{N}_0}$ in X^∞.

Definition 2.1. A *coupling* for a transition probability function $\{\mathbf{P}_x : x \in X\}$ is a family $\{\mathbf{C}_{x,y} : x, y \in X\}$ of probability measures on $X \times X$ such that

$$\mathbf{C}_{x,y}(A \times X) = \mathbf{P}_x(A) \quad \text{and} \quad \mathbf{C}_{x,y}(X \times A) = \mathbf{P}_y(A)$$

for all $x, y \in X$ and $A \in \mathcal{B}_X$.

If $\{\mathbf{P}_x : x \in X\}$ is the transition function for a Markov chain $(\xi_n)_{n\in\mathbb{N}_0}$ then we say that $\{\mathbf{C}_{x,y} : x, y \in X\}$ is a *coupling for* $(\xi_n)_{n\in\mathbb{N}_0}$.

2.3. Iterated function systems. An *iterated function system* is given by a sequence of continuous transformations

$$S_i : X \to X \quad \text{for } i = 1, \ldots, N$$

and a probability vector $(p_1(x), \ldots, p_N(x))$, $x \in X$, i.e.,

$$p_i(x) \geq 0 \quad \text{and} \quad \sum_{i=1}^{N} p_i(x) = 1 \quad \text{for } x \in X.$$

We assume that $p_i : X \to [0, 1]$, $i = 1, \ldots, N$, are continuous functions. Such a system is briefly denoted by $(S_i, p_i)_{i=1}^N$.

The action of an IFS can be described as follows. We choose an initial point $x_0 \in X$ and we randomly select from $\{1, \ldots, N\}$ an integer according to the distribution $(p_1(x_0), \ldots, p_N(x_0))$. When a number k_0 is drawn we define $x_1 = S_{k_0}(x_0)$. Having x_1 we select k_1 with probabilities $(p_1(x_1), \ldots, p_N(x_1))$ and we define $x_2 = S_{k_1}(x_1)$ and so on.

We assume that the iterated function system $(S_i, p_i)_{i=1}^N$ has the following properties:

(i) there exists $\alpha \in (0, 1)$ such that

$$\sum_{i=1}^{N} p_i(x) d(S_i(x), S_i(y)) \leq \alpha d(x, y) \quad \text{for } x, y \in X,$$

(ii) there exists $\delta > 0$ such that for every $x, y \in X$,

$$\sum_{i\,:\,d(S_i(x),S_i(y))\leq \alpha d(x,y)} p_i(x) p_i(y) \geq \delta,$$

(iii) $p_i : X \to [0, 1]$, $i = 1, \ldots, N$, are Hölder-continuous, i.e. there exist constants $L_1 > 0$ and $\nu_i \in (0, 1]$ such that

$$|p_i(x) - p_i(y)| \leq L_i d(x, y)^{\nu_i} \quad \text{for } x, y \in X,\ i = 1, \ldots, N.$$

Condition (i) was introduced by Isaac [9]. It is sometimes expressed in a slightly more general form (see [3, 19]): there exists $0 < r_1 < 1$ such that

$$\prod_{i=1}^{N} d(S_i(x), S_i(y))^{p_i(x)} \leq r_1 d(x, y) \quad \text{for } x, y \in X.$$

Werner [24] shows that if $\min_{i=1,\ldots,N} \inf_{x\in X} p_i(x) > 0$ then the above inequality and (i) are equivalent. It seems that if one requires the geometric rate of convergence, then some kind of contraction on average cannot be avoided. Indeed, Öberg [14] shows that without contraction on average, even if there exists an attractive invariant measure, convergence to this measure is not necessarily geometric.

Assumption (ii) is standard (see [3, 20, 21]). Note, however, that Stenflo [18] gives an example of an IFS which preserves a unique attractive invariant measure and convergence to this measure is geometric, yet condition (ii) is not necessarily satisfied.

In order to obtain uniqueness of an invariant measure for an iterated function system one needs some regularity assumptions on the probabilities p_i. Indeed, Stenflo [17] gives an example of an IFS with continuous p_i's for which there exist two invariant measures. Our assumption (iii) is slightly stronger than Dini continuity which implies uniqueness of the invariant measure (see [3]).

Remark. Conditions (i)–(iii) guarantee existence of an attractive invariant measure for $(S_i, p_i)_{i=1}^N$ (see [21]). We will show that convergence to this measure is at a geometric rate.

We introduce a new metric in the space X:

$$d_1(x,y) = \sum_{i=1}^N L_i d(x,y)^{\nu_i} \quad \text{for } x,y \in X.$$

The metrics d and d_1 are equivalent, and so the classes of sets bounded in d and d_1 are equal. The space (X, d_1) is a Polish space.

Define $a = \alpha^{\min\{\nu_i : i=1,\ldots,N\}}$. Since the functions $[0,\infty) \ni t \mapsto t^{\nu_i}$, $i = 1,\ldots,N$, are concave, we obtain:

(i) for every $x, y \in X$,

$$\sum_{i\,:\,d_1(S_i(x),S_i(y)) \le a d_1(x,y)} p_i(x)p_i(y) \ge \delta, \tag{3}$$

(ii) for every $x, y \in X$,

$$\sum_{i=1}^N p_i(x) d_1(S_i(x), S_i(y)) \le a d_1(x,y), \tag{4}$$

(iii) for every $x, y \in X$,

$$\sum_{i=1}^N |p_i(x) - p_i(y)| \le d_1(x,y). \tag{5}$$

For an iterated function system $(S_i, p_i)_{i=1}^N$ we define a family $\{\mathbf{P}_x : x \in X\}$ of probability measures on X by

$$\mathbf{P}_x = \sum_{i=1}^{N} p_i(x)\delta_{S_i(x)} \quad \text{for } x \in X.$$

Since for every $A \in \mathcal{B}_X$ the function $x \mapsto \mathbf{P}_x(A)$ is measurable, the family $\{\mathbf{P}_x : x \in X\}$ may be viewed as a transition probability function on X. For any given $\mu \in \mathcal{M}_1(X)$ we define a measure $\mathbf{P}_\mu \in \mathcal{M}_1(X)$ by

$$\mathbf{P}_\mu(A) = \int_X \mathbf{P}_x(A)\,\mu(dx).$$

We have

$$\mathbf{P}_\mu^{n+1} = \mathbf{P}_{\mathbf{P}_\mu^n} \quad \text{for } n \in \mathbb{N}_0,$$

where $\mathbf{P}_\mu^n$ is the nth marginal of $\mathbf{P}_\mu^\infty$.

A measure $\mu \in \mathcal{M}_1(X)$ is called *invariant* if $\mathbf{P}_\mu = \mu$.

A Lyapunov function V for the iterated function system is given by

$$V(x) = d_1(x, \bar{x}) \quad \text{for } x \in X,$$

where $\bar{x} \in X$ is fixed. From (4) it follows that

$$\langle V, \mathbf{P}_\mu\rangle \le a\langle V, \mu\rangle + c \quad \text{for } \mu \in \mathcal{M}_1^1(X), \tag{6}$$

where $c = \max_{i=1,\dots,N} d_1(\bar{x}, S_i(\bar{x}))$.

We define $\bar{V} : X^2 \to [0,\infty)$ by

$$\bar{V}(x,y) = V(x) + V(y) \quad \text{for } x, y \in X.$$

Fix $\mu, \nu \in \mathcal{M}_1^1(X)$. Let $b \in \mathcal{M}_1(X^2)$ be such that

$$b(A \times X) = \mu(A) \quad \text{and} \quad b(X \times A) = \nu(A) \quad \text{for } A \in \mathcal{B}_X.$$

Let $\bar{b} \in \mathcal{M}_1(X^2)$ be such that

$$\bar{b}(A \times X) = \mathbf{P}_\mu(A) \quad \text{and} \quad \bar{b}(X \times A) = \mathbf{P}_\nu(A) \quad \text{for } A \in \mathcal{B}_X.$$

From (6) it follows that

$$\langle \bar{V}, \bar{b}\rangle \le \langle \bar{V}, b\rangle + 2c. \tag{7}$$

On the space $\mathcal{M}_{\mathrm{fin}}^1(X^2)$ of finite measures on X^2 with the first moment finite, we define the linear functional

$$\phi(b) = \int_{X^2} d_1(x,y)\, b(dx, dy) \quad \text{for } b \in \mathcal{M}_{\mathrm{fin}}^1(X^2). \tag{8}$$

For every $b \in \mathcal{M}_{\mathrm{fin}}^1(X^2)$, we have

$$\phi(b) \le \langle \bar{V}, b\rangle. \tag{9}$$

3. Results

3.1. Main theorem. Now we formulate the main result of the paper. Its proof is given in Section 3.4.

THEOREM 3.1. *There exist constants* $q \in (0,1)$ *and* $C > 0$ *such that*

$$\|\mathbf{P}_x^n - \mathbf{P}_y^n\|_{\mathcal{L}} \le q^n C(1 + \bar{V}(x,y))$$

for all $x, y \in X$ *and* $n \in \mathbb{N}$.

COROLLARY 3.2. *Let* q *and* C *be as in the statement of Theorem* 3.1. *For all* $\mu, \nu \in \mathcal{M}_1^1(X)$ *and* $n \in \mathbb{N}$ *we have*

$$\|\mathbf{P}_\mu^n - \mathbf{P}_\nu^n\|_{\mathcal{L}} \le Cq^n(1 + \langle V, \mu\rangle + \langle V, \nu\rangle).$$

Moreover, there exists a unique invariant measure $\mu_* \in \mathcal{M}_1^1(X)$ *such that for all* $\mu \in \mathcal{M}_1^1(X)$ *and* $n \in \mathbb{N}$,

$$\|\mathbf{P}_\mu^n - \mu_*\|_{\mathcal{L}} \le q^n \widetilde{C}(1 + \langle V, \mu\rangle),$$

where $\widetilde{C} > 0$ *is constant.*

Proof of Corollary. First observe that for $f \in \mathcal{L}$ and $\mu, \nu \in \mathcal{M}_1^1(X)$,

$$\int_{X^2} \langle f, \mathbf{P}_x^n - \mathbf{P}_y^n\rangle\, \mu \times \nu(dx, dy) = \langle f, \mathbf{P}_\mu^n\rangle - \langle f, \mathbf{P}_\nu^n\rangle.$$

Theorem 3.1 gives

$$|\langle f, \mathbf{P}_\mu^n - \mathbf{P}_\nu^n\rangle| \le \int_{X^2} \|\mathbf{P}_x^n - \mathbf{P}_y^n\|_{\mathcal{L}}\, \mu \times \nu(dx, dy) \le q^n C(1 + \langle V, \mu\rangle + \langle V, \nu\rangle)$$

for all $f \in \mathcal{L}$. This inequality, combined with the fact that $\mathcal{M}_1(X)$ equipped with the metric induced by the Fortet–Mourier norm is complete, gives the existence of an invariant measure μ_*. The proof is finished. ■

3.2. Construction of a coupling for IFS. We can now construct a coupling for the iterated function system $(S_i, p_i)_{i=1}^N$. On X^2 we define a subprobability transition function by

$$\mathbf{Q}_{x,y} = \sum_{i=1}^N \min\{p_i(x), p_i(y)\}\delta_{(S_i(x), S_i(y))} \quad \text{for } x, y \in X.$$

For $B \in \mathcal{B}_X$ we have

$$\mathbf{Q}_{x,y}(B \times X) = \sum_{i=1}^N \min\{p_i(x), p_i(y)\}\delta_{S_i(x)}(B) \le \mathbf{P}_x(B),$$

$$\mathbf{Q}_{x,y}(X \times B) = \sum_{i=1}^N \min\{p_i(x), p_i(y)\}\delta_{S_i(y)}(B) \le \mathbf{P}_y(B).$$

For given $b \in \mathcal{M}_{\mathrm{fin}}(X^2)$ we define $\mathbf{Q}_b \in \mathcal{M}_{\mathrm{fin}}(X^2)$ by

$$\mathbf{Q}_b(A) = \int_{X^2} \mathbf{Q}_{x,y}(A)\, b(dx, dy),$$

where $A \in \mathcal{B}_{X^2}$. As above, we have

$$\mathbf{Q}_b^{n+1} = \mathbf{Q}_{\mathbf{Q}_b^n},$$

where $\{\mathbf{Q}_b^n : n \in \mathbb{N}_0\}$ is the family of one-dimensional distributions of $\mathbf{Q}_b^\infty$.

From (4) it follows that

$$\phi(\mathbf{Q}_b) \le a\phi(b) \quad \text{for } b \in \mathcal{M}_{\mathrm{fin}}(X^2). \tag{10}$$

LEMMA 3.3. *There exists a family $\{\mathbf{R}_{x,y} : x, y \in X\}$ of measures on X^2 such that if we define*

$$\mathbf{C}_{x,y} = \mathbf{Q}_{x,y} + \mathbf{R}_{x,y} \quad \text{for } x, y \in X$$

then:

(i) *for every $A \in \mathcal{B}_{X^2}$ the function $(x, y) \mapsto \mathbf{R}_{x,y}(A)$ is measurable,*
(ii) *the measures $\mathbf{R}_{x,y}$ are nonnegative for $x, y \in X$,*
(iii) *each $\mathbf{C}_{x,y}$ is a probability measure and consequently $\{\mathbf{C}_{x,y} : x, y \in X\}$ defines a transition probability function on X^2,*
(iv) *for all $B \in \mathcal{B}_X$ and $x, y \in X$ we have*

$$\mathbf{C}_{x,y}(B \times X) = \mathbf{P}_x(B) \quad \text{and} \quad \mathbf{C}_{x,y}(X \times B) = \mathbf{P}_y(B).$$

Proof. A short computation shows that the measure $\mathbf{R}_{x,y}$ defined on rectangles by

$$\mathbf{R}_{x,y}(A \times B) = \frac{1}{1 - \mathbf{Q}_{x,y}(X^2)}(\mathbf{P}_x(A) - \mathbf{Q}_{x,y}(A \times X))(\mathbf{P}_y(B) - \mathbf{Q}_{x,y}(X \times B))$$

when $\mathbf{Q}_{x,y}(X^2) < 1$, and $\mathbf{R}_{x,y}(A \times B) = 0$ when $\mathbf{Q}_{x,y}(X^2) = 1$, has all the required properties. ∎

The above lemma states that one can construct a coupling $\{\mathbf{C}_{x,y} : x, y \in X\}$ for $\{\mathbf{P}_x : x \in X\}$ such that $\mathbf{Q}_{x,y} \le \mathbf{C}_{x,y}$. The idea of splitting the transition function of a coupled chain into two parts, $\mathbf{C}_{x,y} = \mathbf{Q}_{x,y} + \mathbf{R}_{x,y}$ where $\mathbf{Q}_{x,y}$ is contractive (see [10]), is due to Hairer [6].

By the procedure described in (1) and (2) we obtain a family of probability measures $\mathbf{C}_{x,y}^\infty$ on $X^\infty \times X^\infty$ with marginals $\mathbf{P}_x^\infty$ and $\mathbf{P}_y^\infty$.

For the measures $\{\mathbf{Q}_{x,y} : x, y \in X\}$ we define, as above, a family $\{\mathbf{Q}_{x,y}^\infty : x, y \in X\}$ of subprobability measures on $X^\infty \times X^\infty$.

Suppose that a Markov chain Φ on X^2, with transition function $\{\mathbf{C}_{x,y} : x, y \in X\}$, stays at (x_0, y_0) at time n. The next step of Φ can be drawn according to the measure $\mathbf{Q}_{x_0,y_0}$ or $\mathbf{R}_{x_0,y_0}$ with probability $\|\mathbf{Q}_{x_0,y_0}\|$ or $\|\mathbf{R}_{x_0,y_0}\|$, respectively. In order to distinguish these cases we introduce the augmented space $X \times \{0,1\}$ and the transition functions

$$\widehat{\mathbf{C}}_{x,y,\theta} = \mathbf{Q}_{x,y} \times \delta_1 + \mathbf{R}_{x,y} \times \delta_0,$$
$$\widehat{\mathbf{Q}}_{x,y,\theta} = \mathbf{Q}_{x,y} \times \delta_1, \qquad \widehat{\mathbf{R}}_{x,y,\theta} = \mathbf{R}_{x,y} \times \delta_0,$$

where $(x,y) \in X^2$ and $\theta \in \{0,1\}$. The parameter $\theta \in \{0,1\}$ is responsible for choosing either the measure $\mathbf{Q}_{x,y}$ or $\mathbf{R}_{x,y}$. For example, if the chain stays in $X^2 \times \{1\}$ at time n, it means that the last step was drawn according to $\mathbf{Q}_{u,v}$ for some $(u,v) \in X^2$.

Obviously

$$\pi^*_{X^2} \widehat{\mathbf{C}}_{x,y,\theta} = \mathbf{C}_{x,y} \quad \text{for every } (x,y,\theta) \in X^2 \times \{0,1\},$$

where $\pi_{X^2} : X^2 \times \{0,1\} \to X^2$ is the projection.

Fix $(x_0,y_0) \in X^2$. The transition probabilities $\{\mathbf{C}_{x,y} : x,y \in X\}$ define on X^2 the Markov chain Φ starting from (x_0,y_0). The transition probabilities $\{\widehat{\mathbf{C}}_{x,y,\theta} : x,y \in X,\ \theta \in \{0,1\}\}$ define a Markov chain $\widehat{\Phi}$ on the augmented space $X^2 \times \{0,1\}$. We adopt the convention that the starting distribution of $\widehat{\Phi}$ is $\widehat{\mathbf{C}}^0_{x_0,y_0} = \delta_{(x_0,y_0,1)}$.

As above, on $(X^2 \times \{0,1\})^\infty$ we construct the measure $\widehat{\mathbf{C}}^\infty_{x_0,y_0}$, associated with $\widehat{\Phi}$. From now on, we assume that the chains Φ and $\widehat{\Phi}$, taking values in X^2 and $X^2 \times \{0,1\}$, are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Expectation with respect to $\mathbf{C}^\infty_{x_0,y_0}$ or $\widehat{\mathbf{C}}^\infty_{x_0,y_0}$ is denoted by $\mathbb{E}_{x_0,y_0}$.

3.3. Lemmas. For any fixed $\varepsilon \in (0, 1-a)$ we define

$$K_\varepsilon = \{(x,y) \in X^2 : \bar{V}(x,y) < 2c/\varepsilon\}.$$

Let $\rho : (X \times X)^\infty \to \mathbb{N}$ denote the time of the first visit in K_ε, i.e.

$$\rho((x_n,y_n)_{n\in\mathbb{N}_0}) = \inf\{n \geq 1 : (x_n,y_n) \in K_\varepsilon\}.$$

Lemma 3.4. *For every $\gamma \in (0,1)$ there exist constants $C_1, C_2 > 0$ such that for every $(x_0,y_0) \in X^2$,*

$$\mathbb{E}_{x_0,y_0}((a+\varepsilon)^{-\gamma\rho}) \leq C_1 \bar{V}(x_0,y_0) + C_2. \tag{11}$$

Proof. Fix $(x_0,y_0) \in X^2$. Let $\Phi = (X_n,Y_n)_{n\in\mathbb{N}_0}$ be the Markov chain starting from (x_0,y_0) and with transition function $\{\mathbf{C}_{x,y} : x,y \in X\}$. Let $\mathcal{F}_n \subset \mathcal{F}$, $n \in \mathbb{N}_0$, be the filtration in Ω associated with Φ. Define

$$A_n = \{\omega \in \Omega : (X_i(\omega), Y_i(\omega)) \notin K_\varepsilon \text{ for } i = 1,\ldots,n\} \quad \text{for } n \in \mathbb{N}_0.$$

We have $A_{n+1} \subset A_n$ and $A_n \in \mathcal{F}$, $n \in \mathbb{N}$. From (7) it follows that

$$\mathbf{1}_{A_n} \cdot \mathbb{E}(\bar{V}(X_{n+1},Y_{n+1}) \mid \mathcal{F}_n) \leq \mathbf{1}_{A_n} \cdot (a\bar{V}(X_n,Y_n) + 2c) \leq \mathbf{1}_{A_n} \cdot (a+\varepsilon)\bar{V}(X_n,Y_n)$$

$\mathbb{P}$-a.e. in Ω. Further,

$$\int_{A_n} \bar V(X_n,Y_n)\,d\mathbb{P} \le \int_{A_{n-1}} \bar V(X_n,Y_n)\,d\mathbb{P} = \int_{A_{n-1}} \mathbb{E}(\bar V(X_n,Y_n)\,|\,\mathcal{F}_{n-1})\,d\mathbb{P}$$
$$\le \int_{A_{n-1}} [a\bar V(X_{n-1},Y_{n-1})+2c]\,d\mathbb{P}$$
$$\le (a+\varepsilon)\int_{A_{n-1}} \bar V(X_{n-1},Y_{n-1})\,d\mathbb{P}$$

and so

$$\int_{A_n} \bar V(X_n,Y_n)\,d\mathbb{P} \le (a+\varepsilon)^{n-1}\int_{A_1} \bar V(X_1,Y_1)\,d\mathbb{P} \le (a+\varepsilon)^n[a\bar V(x_0,y_o)+2c].$$

Finally, by the Chebyshev inequality, for every $n\in\mathbb{N}$ we obtain

$$\mathbb{P}((X_1,Y_1)\notin K_\varepsilon,\dots,(X_n,Y_n)\notin K_\varepsilon) = \int_{A_{n-1}} \mathbb{P}((X_n,Y_n)\notin K_\varepsilon\,|\,\mathcal{F}_{n-1})\,d\mathbb{P} \tag{12}$$
$$\le \frac{\varepsilon}{2c}\int_{A_{n-1}} \mathbb{E}(\bar V(X_n,Y_n)\,|\,\mathcal{F}_{n-1})\,d\mathbb{P} \le (a+\varepsilon)^n\frac{\varepsilon}{2c}[a\bar V(x_0,y_0)+2c].$$

Fix $\gamma\in(0,1)$. Inequality (12) implies (11) with properly chosen constants C_1 and C_2, which finishes the proof. ∎

For every $r>0$ we define

$$C_r = \{(x,y)\in X^2 : d_1(x,y)<r\}.$$

LEMMA 3.5. *If a measure $b\in\mathcal{M}_{\mathrm{fin}}(X^2)$ is such that* $\operatorname{supp} b\subset C_r$ *then*

$$\mathbf{Q}_b(C_{ar}) \ge \delta\|b\|.$$

Proof. Fix $r>0$ and $(x,y)\in C_r$. From (3) we obtain

$$\sum_{i\,:\,d_1(S_i(x),S_i(y))\le ad_1(x,y)} \min\{p_i(x),p_i(y)\} \ge \sum_{i\,:\,d_1(S_i(x),S_i(y))\le ad_1(x,y)} p_i(x)p_i(y) \ge \delta$$

and so

$$\mathbf{Q}_{x,y}(C_{ar}) \ge \delta.$$

Now, let $b\in\mathcal{M}_{\mathrm{fin}}(X^2)$ satisfy $\operatorname{supp} b\subset C_r$. Then

$$\mathbf{Q}_b(C_{ar}) = \int_{\operatorname{supp} b} \mathbf{Q}_{u,v}(C_{ar})\,b(du,dv) \ge \delta\|b\|. \ ∎$$

LEMMA 3.6. *For every $\varepsilon\in(0,1-a)$ there exists $n_0\in\mathbb{N}$ such that*

$$\|\mathbf{Q}^\infty_{x,y}\| \ge \frac{1}{2}\delta^{n_0} \quad \textit{for } (x,y)\in K_\varepsilon.$$

Proof. For every $(x,y)\in X^2$ we have

$$\|\mathbf{Q}_{x,y}\| + \sum_{i\,:\,p_i(x)\ge p_i(y)} |p_i(x)-p_i(y)| = 1,$$

which implies by (5) that

$$\|\mathbf{Q}_{x,y}\| \geq 1 - d_1(x,y).$$

For any $b \in \mathcal{M}_{\mathrm{fin}}(X^2)$ we thus obtain

$$\|\mathbf{Q}_b\| \geq \|b\| - \phi(b),$$

where ϕ is defined in (8). Since (10) holds, iterating the above inequality leads to

$$\|\mathbf{Q}_b^{n+1}\| \geq \|b\| - \Big(\sum_{k=0}^{n} a^k\Big)\phi(b) \quad \text{for } n \in \mathbb{N}.$$

This implies that

$$\|\mathbf{Q}_b^{n}\| \geq \|b\| - \frac{\phi(b)}{1-a} \quad \text{for } n \in \mathbb{N}. \tag{13}$$

Let $r = \frac{1}{2}(1-a)$ and $\operatorname{supp} b \subset C_r$. From (13) it follows that

$$\|\mathbf{Q}_b^{\infty}\| \geq \|b\|/2. \tag{14}$$

Now fix $\varepsilon \in (0, 1-a)$. We have $K_\varepsilon \subset C_{2c/\varepsilon}$ by the definition of these sets. If we define $n_0 = \min\{n \geq 1 : a^n \cdot 2c/\varepsilon < r\}$ then $C_{a^{n_0}\cdot 2c/\varepsilon} \subset C_r$. By Lemma 3.5 we obtain

$$\mathbf{Q}_{x,y}^{n_0}(C_r) \geq \mathbf{Q}_{x,y}^{n_0}(C_{a^{n_0}\cdot 2c/\varepsilon}) \geq \delta^{n_0}$$

for every $(x,y) \in K_\varepsilon$. Finally (14) shows that $\|\mathbf{Q}_{x,y}^{\infty}\| \geq \frac{1}{2}\delta^{n_0}$. ■

Definition 3.7. The *coupling time* $\tau : (X^2 \times \{0,1\})^\infty \to \mathbb{N}_0$ is defined by

$$\tau((x_n, y_n, \theta_n)_{n\in\mathbb{N}_0}) = \inf\{n \geq 1 : \forall_{k\geq n} \theta_k = 1\}.$$

The coupling time of the Markov chain $\widehat{\Phi}$ is the random time from which the movement of $\widehat{\Phi}$ is controlled exclusively by the measures $\{\mathbf{Q}_{x,y} : x, y \in X\}$. The proof of the following lemma can be found in [15], but in a different setup. For the convenience of the reader we give it here.

Lemma 3.8. *There exist constants $\widetilde{q} \in (0,1)$ and $C_3 > 0$ such that for every $(x,y) \in X^2$ we have*

$$\mathbb{E}_{x,y}(\widetilde{q}^{\,-\tau}) \leq C_3(1 + \bar{V}(x,y)).$$

Proof. Fix $\varepsilon \in (0, 1-a)$ and $x, y \in X$. For simplicity we denote $\beta = (a+\varepsilon)^{-1/2}$. Let ρ be the random time of the first visit in K_ε. Define the random time of the nth visit in K_ε:

$$\rho_1 = \rho, \qquad \rho_{n+1} = \rho_n + \rho \circ T_{\rho_n} \quad \text{for } n > 1,$$

where T_n are the shift operators on $(X^2 \times \{0,1\})^\infty$, i.e.

$$T_n((x_k, y_k, \theta_k)_{k\in\mathbb{N}_0}) = (x_{k+n}, y_{k+n}, \theta_{k+n})_{k\in\mathbb{N}_0}.$$

Since, by Lemma 3.4, every ρ_n is $\mathbf{C}^\infty_{x,y}$-a.e. finite, the strong Markov property gives

$$\mathbb{E}_{x,y}(\beta^\rho \circ T_{\rho_n} \mid \mathcal{F}_{\rho_n}) = \mathbb{E}_{(X_{\rho_n},Y_{\rho_n})}(\beta^\rho) \quad \text{for } n \in \mathbb{N},$$

where $\mathcal{F}_{\rho_n}$ is the σ-algebra in $(X^2 \times \{0,1\})^\infty$ generated by ρ_n, and $\Phi = (X_n, Y_n)_{n\in\mathbb{N}_0}$ is the Markov chain with distribution $\mathbf{C}^\infty_{x,y}$.

It follows from Lemma 3.4 and the definition of K_ε that

$$\mathbb{E}_{x,y}(\beta^{\rho_{n+1}}) = \mathbb{E}_{x,y}(\beta^{\rho_n}\mathbb{E}_{(X_{\rho_n},Y_{\rho_n})}(\beta^\rho)) \le \mathbb{E}_{x,y}(\beta^{\rho_n})\left[C_1\frac{2c}{\varepsilon} + C_2\right].$$

Setting $\eta = C_1\frac{2c}{\varepsilon} + C_2$ we obtain

$$\mathbb{E}_{x,y}(\beta^{\rho_{n+1}}) \le \eta^n \mathbb{E}_{x,y}(\beta^\rho) \le \eta^n[C_1\bar{V}(x,y) + C_2]. \tag{15}$$

Define

$$\widehat{\tau}((x_n,y_n,\theta_n)_{n\in\mathbb{N}_0}) = \inf\{n \ge 1 : (x_n,y_n) \in K_\varepsilon,\ \forall_{k\ge n}\,\theta_k = 1\},$$
$$\sigma = \inf\{n \ge 1 : \widehat{\tau} = \rho_n\}.$$

Lemma 3.6 implies that there exists $n_0 \in \mathbb{N}$ such that

$$\widehat{\mathbf{C}}^\infty_{x,y}(\sigma > n) \le (1 - \delta^{n_0}/2)^n \quad \text{for } n \in \mathbb{N}. \tag{16}$$

Let $p > 1$. From the Hölder inequality, (15) and (16) it follows that

$$\begin{aligned}
\mathbb{E}_{x,y}(\beta^{\widehat{\tau}/p}) &\le \sum_{k=1}^\infty \mathbb{E}_{x,y}(\beta^{\rho_k/p}\mathbf{1}_{\sigma=k}) \le \sum_{k=1}^\infty [\mathbb{E}_{x,y}(\beta^{\rho_k})]^{1/p}\widehat{\mathbf{C}}^\infty_{x,y}(\sigma = k)^{1-1/p} \\
&\le [C_1\bar{V}(x,y) + C_2]^{1/p}\eta^{-1/p}\sum_{k=1}^\infty \eta^{k/p}(1-\delta^{n_0}/2)^{(k-1)(1-1/p)} \\
&= [C_1\bar{V}(x,y) + C_2]^{1/p}\eta^{-1/p}(1-\delta^{n_0}/2)^{-(1-1/p)} \\
&\quad \times \sum_{k=1}^\infty \left[\left(\frac{\eta}{1-\delta^{n_0}/2}\right)^{1/p}(1-\delta^{n_0}/2)\right]^k.
\end{aligned}$$

Choosing sufficiently large p and setting $\widetilde{q} = \beta^{-1/p}$ we observe that

$$\mathbb{E}_{x,y}(\widetilde{q}^{\,-\widehat{\tau}}) = \mathbb{E}_{x,y}(\beta^{\widehat{\tau}/p}) \le (1 + \bar{V}(x,y))C_3$$

for some $C_3 > 1$. Since $\tau \le \widehat{\tau}$, the proof is complete. ∎

3.4. Proof of Theorem 3.1. For every $n \in \mathbb{N}$ we define

$$A_{n/2} = \{t \in (X^2 \times \{0,1\})^\infty : \tau(t) \le n/2\},$$
$$B_{n/2} = \{t \in (X^2 \times \{0,1\})^\infty : \tau(t) > n/2\}.$$

Since $A_{n/2}$ and $B_{n/2}$ are disjoint and their union is the whole pathspace $(X^2 \times \{0,1\})^\infty$ for every $n \in \mathbb{N}$, we can write

$$\widehat{\mathbf{C}}^\infty_{x,y} = \widehat{\mathbf{C}}^\infty_{x,y}|_{A_{n/2}} + \widehat{\mathbf{C}}^\infty_{x,y}|_{B_{n/2}} \quad \text{for } n \in \mathbb{N}.$$

From the fact that $\|\cdot\|_{\mathcal{L}} \le \|\cdot\|_{\mathcal{H}}$ it follows that

$$(17)\qquad \|\mathbf{P}_x^n - \mathbf{P}_y^n\|_{\mathcal{L}} = \sup_{f\in\mathcal{L}} \Big| \int_{X^2} (f(z_1) - f(z_2))\,(\pi_{X^2}^*\pi_n^*\widehat{\mathbf{C}}_{x,y}^\infty)(dz_1, dz_2)\Big|$$
$$\le \sup_{f\in\mathcal{H}} \Big| \int_{X^2} (f(z_1) - f(z_2))\,(\pi_{X^2}^*\pi_n^*(\widehat{\mathbf{C}}_{x,y}^\infty|_{A_{n/2}}))(dz_1, dz_2)\Big| + 2\widehat{\mathbf{C}}_{x,y}^\infty(B_{n/2}),$$

where $\pi_n : (X^2\times\{0,1\})^\infty \to X^2\times\{0,1\}$ are projections on the nth component.

By iterative application of (10) we have

$$(18)\qquad \sup_{f\in\mathcal{H}} \Big| \int_{X^2} (f(z_1) - f(z_2))\,(\pi_{X^2}^*\pi_n^*(\widehat{\mathbf{C}}_{x,y}^\infty|_{A_{n/2}}))(dz_1, dz_2)\Big|$$
$$\le \phi(\pi_{X^2}^*\pi_n^*(\widehat{\mathbf{C}}_{x,y}^\infty|_{A_{n/2}})) \le a^{n/2}\phi(\pi_{X^2}^*\pi_{n/2}^*(\widehat{\mathbf{C}}_{x,y}^\infty|_{A_{n/2}})).$$

It follows from (7) and (9) that

$$(19)\qquad \phi(\pi_{X^2}^*\pi_{n/2}^*(\widehat{\mathbf{C}}_{x,y}^\infty|_{A_{n/2}})) \le a^{n/2}\bar{V}(x,y) + \frac{2c}{1-a}.$$

Then combining (17), (18) and (19) gives

$$(20)\qquad \|\mathbf{P}_x^n - \mathbf{P}_y^n\|_{\mathcal{L}} \le a^{n/2}\Big[a^{n/2}\bar{V}(x,y) + \frac{2c}{1-a}\Big] + 2\widehat{\mathbf{C}}_{x,y}^\infty(B_{n/2}).$$

From Lemma 3.8 and the Chebyshev inequality we get

$$(21)\qquad \widehat{\mathbf{C}}_{x,y}^\infty(B_{n/2}) = \widehat{\mathbf{C}}_{x,y}^\infty(\{\tau > n/2\}) \le \widetilde{q}^{n/2}C_4(1+\bar{V}(x,y))$$

for some $\widetilde{q}\in(0,1)$ and $C_4 > 0$. Finally, it follows from (20) and (21) that

$$\|\mathbf{P}_x^n - \mathbf{P}_y^n\|_{\mathcal{L}} \le a^{n/2}C_5(1+\bar{V}(x,y)) + \widetilde{q}^{n/2}C_4(1+\bar{V}(x,y)),$$

where $C_5 = \max\{1, 2c/(1-a)\}$. Setting $q = \max\{a^{1/2}, \widetilde{q}^{1/2}\}$ completes the proof. ∎

Remark. The general idea of the above proof is due to M. Hairer. A central role is played by inequality (20), called the *coupling inequality*, which can be found in [6].

References

[1] M. F. Barnsley, *Fractals Everywhere*, Academic Press, New York, 1988.

[2] —, *Superfractals*, Cambridge Univ. Press, New York, 2006.

[3] 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 24 (1988), 367–394.

[4] J. H. Elton and M. Piccioni, *Iterated function systems arising from recursive estimation problems*, Probab. Theory Related Fields 91 (1992), 103–114.

[5] S. Ethier and T. Kurtz, *Markov Processes*, Wiley, New York, 1986.

[6] M. Hairer, *Exponential mixing properties of stochastic PDEs through asymptotic coupling*, Probab. Theory Related Fields 124 (2002), 345–380.

[7] —, *Coupling stochastic PDEs*, in: XIVth International Congress on Mathematical Physics, World Sci., 2005, 281–289.

[8] J. Hutchinson, *Fractals and self-similarity*, Indiana Univ. Math. J. 30 (1981), 713–747.

[9] R. Isaac, *Markov processes and unique stationary probability measures*, Pacific J. Math. 12 (1962), 273–286.

[10] T. Kaijser, *On a new contraction condition for random systems with complete connections*, Rev. Roumaine Math. Pures Appl. 26 (1981), 1075–1117.

[11] S. Karlin, *Some random walks arising in learning models*, Pacific J. Math. 3 (1953), 725–756.

[12] A. Lasota, *From fractals to stochastic differential equations*, in: Chaos—The Interplay between Stochastic and Deterministic Behaviour (Karpacz, 1995), P. Garbaczewski et al. (eds.), Lecture Notes in Phys. 457, Springer, 1995, 235–255.

[13] A. Lasota and M. C. Mackey, *Chaos, Fractals and Noise. Stochastic Aspects of Dynamics*, Springer, 1994.

[14] A. Öberg, *Approximation of invariant measures for random iterations*, Rocky Mountain J. Math. 36 (2006), 273–301.

[15] C. Odasso, *Exponential mixing for stochastic PDEs: the non-additive case*, Probab. Theory Related Fields 140 (2008), 41–82.

[16] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems, Vol. II: Applications*, Springer, 1998.

[17] Ö. Stenflo, *A note on a theorem of Karlin*, Statist. Probab. Lett. 54 (2001), 183–187.

[18] —, *Ergodic theorems for Markov chains represented by iterated function systems*, Bull. Polish Acad. Sci. Math. 49 (2001), 27–43.

[19] —, *Uniqueness of invariant measures for place-dependent random iterations of functions*, in: Fractals in Multimedia (Minneapolis, MN, 2002), IMA Vol. Math. Appl. 132, Springer, New York, 2002, 13–32.

[20] T. Szarek, *Invariant measures for Markov operators with application to function systems*, Studia Math. 154 (2003), 207–222.

[21] —, *Invariant measures for nonexpansive Markov operators on Polish spaces*, Dissertationes Math. 415 (2003), 62 pp.

[22] H. Thorisson, *Coupling, Stationarity and Regeneration*, Springer, 2000.

[23] C. Villani, *Optimal Transport: Old and New*, Grundlehren Math. Wiss. 338, Springer, 2009.

[24] I. Werner, *Contractive Markov systems*, J. London Math. Soc. 71 (2005), 236–258.

[25] L.-S. Young, *Recurrence times and rates of mixing*, Israel J. Math. 110 (1999), 153–188.

Maciej Ślęczka
Institute of Mathematics
Silesian University
Bankowa 14
40-007 Katowice, Poland
E-mail: sleczka@ux2.math.us.edu.pl

Received April 4, 2009
Revised version June 9, 2011 (6599)