ITERATED FUNCTION SYSTEMS. A CRITICAL SURVEY

MARIUS IOSIFESCU

In the last 30 years or so, the phrase 'iterated function system' has become more and more frequent in mathematical papers and in very many publications of applied people. As in many other instances, the notion of an iterated function system (IFS) is not a new one. Actually, we are faced with the renaming of an old concept, as shown in the first section of the present paper. However, it should be accepted that the study of this notion has been very much deepened under its new clothes. This survey is thus intended to present the state of the art of the IFS notion, its connections with other concepts, as well as to point out to some open problems. The paper is divided into six sections and two appendices. The first section sketches a historical perspective starting from the simplest case of a finite number of self-mappings. Section 2 introduces the general case of an arbitrary family of self-mappings obeying an i.i.d. mechanism. In Section 3 the existence and uniqueness of a stationary distribution are studied while in Section 4 almost sure convergence properties of the backward process are proved. Section 5 is devoted to a study of the support of the stationary distribution. Section 6 takes up the more general case of an arbitrary family of self-mappings obeying a strictly stationary mechanism instead of an i.i.d. one as in the five previous sections. The appendices collect some classical concepts and results on metrics and distances in metric spaces that we are using in the paper.

AMS 2000 Subject Classification: 60G10, 60J05.

Key words: iterated function system, metric space, Markov process, stationary probability, weak convergence, almost sure convergence, strictly stationary process.

1. A SIMPLE BASIC CASE

The simplest iterated function system (IFS)

$$(p,\, (u_i)_{1\le i\le m})$$

is defined by a finite collection of measurable self-mappings $u_i : W \to W$, $1 \le i \le m$, $m \in \mathbf{N}_+ := \{1, 2, \dots\}$, $m \ge 2$, of a metric space W with metric d and Borel σ-algebra $\mathcal{B}_W$, and a constant probability vector $p = (p_i)_{1\le i\le m}$. This allows to define a sequence $(\zeta_n)_{n\in\mathbf{N}}$ of W-valued random variables by the

MATH. REPORTS **11(61)**, *3* (2009), 181–229

recursive equation

$$\zeta_n = u_{\xi_n}(\zeta_{n-1}), \quad n \in \mathbf{N}_+,$$

where $\zeta_0 = w_0$ (arbitrarily given in W) and $(\xi_n)_{n\in\mathbf{N}_+}$ is a sequence of $\{1,\ldots,m\}$-valued random variables with common probability distribution p, on the infinite product probability space

$$(\{1,\ldots,m\}, \mathcal{P}\{1,\ldots,m\}, (p_i)_{1\le i\le m})^{\mathbf{N}_+}.$$

[This clearly implies that $(\xi_n)_{n\in\mathbf{N}_+}$ is an i.i.d. sequence.] Then

$$\zeta_n = u_{\xi_n} \circ \cdots \circ u_{\xi_1}(w_0), \quad n \in \mathbf{N}_+,$$

and it easy to see that $(\zeta_n)_{n\in\mathbf{N}}$ is a W-valued Markov process starting at $w_0 \in W$ with transition function

$$P(w, A) = \sum_{i\in A_w} p_i, \quad w \in W,\ A \in \mathcal{B}_W,$$

where $A_w = \{1 \le i \le m \mid u_i(w) \in A\} = \{1 \le i \le m \mid w \in u_i^{-1}(A)\}$. So, the transition operator U of our process is defined by

$$Uf(w) = \int_W P(w, \mathrm{d}w')\, f(w') = \sum_{i=1}^m p_i f(u_i(w)), \quad w \in W,\ f \in B(W),$$

the last equation being easily verified starting with indicator functions $f = I_A$, $A \in \mathcal{B}_W$. Here, $B(W)$ is the linear space of complex-valued bounded $\mathcal{B}_W$-measurable functions defined on W.

A lot of work has been devoted to such iterated function systems $(p, (u_i)_{1\le i\le m})$ in the last three decades. As already mentioned, IFS is not at all a new concept. It only became fashionable in the framework of fractals and chaos but, before that, it appeared as the simplest case of a random system with complete connections and, in particular, as the Bush-Mosteller model for learning with experimenter-controlled-events [see, e.g., Herkenrath, Iosifescu, and Rudolph [23] as well as the review MR 932532 (90b:60078) of Barnsley and Elton [6]; above all see Iosifescu and Grigorescu [29, Chapter 1].

Even if objects now defined as fractals have been known to artists and mathematicians for centuries, the word ‘fractal’ was coined by Benoit Mandelbrot in the late 1970s to designate a set whose Hausdorff dimension is not an integer. In less formal terms, a fractal object is one that is self-similar and sub-divisible: subsections of it are similar in some sense to the whole object while no matter how small is a subdivision of it, this contains no less details than the whole.

Chaos is a subject brought forward by the study of nonlinear dynamics and has connections with fractal geometry. Chaotic systems are characterized by major changes in their behaviour caused by minor changes in the parameters that control them. Often used to illustrate the concept is the “butterfly effect": the breeze produced by the beating of a butterfly's wings may eventually generate a hurricane.

In Crilly, Earnshow, and Jones (Eds.) [12] the reader will find a lot of interesting material on fractals, chaos, and their interrelationship, as well as many references. See also the site `www.superfractals.com`. For historical material see [1].

Sketchily (for more details see further on Section 5), using an IFS, a fractal can be constructed as follows. Assume (W, d) is a bounded subset of $\mathbf{R}^2$ in view of computer graphics applications, and the u_i, $1 \leq i \leq m$, are contraction self-mappings of W, i.e.,

$$d\left(u_i(w'), u_i(w'')\right) \leq r\, d(w', w'')$$

for any $w', w'' \in W$ and $1 \leq i \leq m$, where r is a positive number strictly less than 1. For any compact subset A of W define

$$\mathcal{S}(A) = \bigcup_{1 \leq i \leq m} u_i\left(A\right).$$

Then there is a unique compact subset K of W such that $\mathcal{S}(K) = K$. This is called the *attractor* of the self-mappings u_i, $1 \leq i \leq m$, or of the *deterministic* IFS $(u_i)_{1 \leq i \leq m}$ (note that no use was still made of the probability vector $p = (p_i)_{1 \leq i \leq m}$), and in many cases has a fractal structure. Equally, for any compact subset A of W, the sequence $(A_n)_{n \in \mathbf{N}}$, where $A_0 = A$ and $A_n = \mathcal{S}\left(A_{n-1}\right)$, $n \in \mathbf{N}_+$, converges to K in the Hausdorff metric (see A2.2) regardless of the choice of A. In applications, the u_i are usually taken to be affine, that is of the form $u_i\left(w\right) = M_i w + b_i$, $1 \leq i \leq m$, for $w \in W \subset \mathbf{R}^2$, where the M_i are 2×2 matrices and the b_i two-dimensional real vectors. In such a case, K is encoded by $6m$ real numbers. Traditional fractals as the middle thirds Cantor set and the Sierpinski triangle (or arrowhead or gasket) can be generated in this way.

A sequence $(w_n)_{n \in \mathbf{N}_+}$ of points in W is called an *orbit* of the deterministic IFS $(u_i)_{1 \leq i \leq m}$ if $w_{n+1} = u_{i_n}(w_n)$, where $i_n \in I$, $n \in \mathbf{N}_+$. Then K above is an attractor in the sense of dynamical systems, since every orbit does approach K as $n \to \infty$. Moreover, for any $w \in K$ there are $i_n = i_n(w) \in \{1, \ldots, m\}$, $n \in \mathbf{N}_+$, such that the sequence $(u_{i_n} \circ \cdots \circ u_{i_1}(w_0))_{n \in \mathbf{N}_+}$ converges to w as $n \to \infty$ for any $w_0 \in W$. This clearly connects IFS with chaos as described above.

There are several procedures to plot the attractor K on a computer screen. See, e.g., Bressloff and Stark [9]. In this reference a neural network formulation of a deterministic IFS can also be found.

Note that a deterministic IFS can only generate a black and white image. Instead, a (random) IFS $((u_i)_{1 \leq i \leq m}, p)$ as defined before is able to generate both colour and grey images. See again Bressloff and Stark (op. cit.). In such a framework, the attractor K of the deterministic IFS $(u_i)_{1 \leq i \leq m}$ is the support of the unique invariant measure of the Markov chain $(\zeta_n)_{n\in\mathbf{N}}$ introduced at the beginning of this section. Let us conclude it by mentioning that if W is a separable complete metric space, then any transition function

$$Q : W \times \mathcal{B}_W \to [0,1], \quad (w, A) \to Q\,(w, A)$$

can be represented as a measure $\mathcal{P}$ supported by some subset $(f_i)_{i\in Y}$ of the set of all measurable self-mappings of W, to mean that

$$Q\,(w, A) = \mathcal{P}\left(\{f \in (f_i)_{i\in Y} : f(w) \in A\}\right)$$

for any $w \in W$ and $A \in \mathcal{B}_W$. So, an IFS $((u_i)_{1\le i\le m}, p)$ is a very special case of this general context. See Section 2 for more details.

2. THE GENERAL I.I.D. CASE

In this section we will take up a more general case. At the expense of some notational complication, nothing prevents us to consider an arbitrary measurable space instead of the finite set $\{1, \ldots, m\}$.

Let W always be a metric space with metric d and Borel σ-algebra $\mathcal{B}_W$, $(X, \mathcal{X})$ an arbitrary measurable space, $u : W \times X \to W$ a $(\mathcal{B}_W \otimes \mathcal{X}, \mathcal{B}_W)$-measurable mapping, and p a probability measure on $\mathcal{X}$. Write $u_x(w) := u(w, x)$, $x \in X$, and note that for any $x \in X$ we have a $\mathcal{B}_W$-measurable self-mapping $u_x : W \to W$. The pair

$$(p, (u_x)_{x\in X}) \tag{2.1}$$

is called an *iterated function system* (*IFS*), as in the case where X is a finite set. Similarly to the latter case, on a probability space $(\Omega, \mathcal{K}, \mathbb{P}_p)$ consider the W-valued sequence $(\zeta_n)_{n\in\mathbf{N}}$ defined by $\zeta_0 = w_0$ (arbitrarily given in W) and

$$\zeta_n = u_{\xi_n} \circ \cdots \circ u_{\xi_1}(w_0), \quad n \in \mathbf{N}_+, \tag{2.2}$$

where $(\xi_n)_{n\in\mathbf{N}_+}$ is an i.i.d. X-valued sequence with common $\mathbb{P}_p$-distribution p. To mark dependence on w_0, we shall occasionally write $\zeta_n^{w_0}$ to denote the random variables defined by (2.2). Again, $(\zeta_n)_{n\in\mathbf{N}}$ is a Markov process starting at $w_0 \in W$ with transition function P defined by $P(w, A) = p(A_w)$, $w \in W$, $A \in \mathcal{B}_W$, where $A_w := \{x \in X \mid u_x(w) \in A\} = \{x \in X \mid w \in u_x^{-1}(A)\}$. The transition operator U of our process is now defined by

$$Uf(w) = \int_W P(w, \mathrm{d}w')f(w') = \int_X f(u_x(w))p(\mathrm{d}x), \ w \in W, \ f \in B(W), \tag{2.3}$$

the last equation being again easily verified starting with indicator functions $f = I_A,\ A \in \mathcal{B}_W$. More generally,

$$U^n f(w) = \int_W f(w')P^n(w, \mathrm{d}w')$$

$$= \int_X \cdots \int_X p(\mathrm{d}x_1)\cdots p(\mathrm{d}x_n) f(u_{x_n} \circ \cdots \circ u_{x_1}(w)), \quad w \in W,$$

for any $n \in \mathbf{N}_+$ and $f \in B(W)$, where P^n is the n-step transition function associated with P. The probabilistic meaning of $U^n f(w)$ is that it is the mean value of $f(\zeta_n^w)$ under $\mathbb{P}_p$ for any $n \in \mathbf{N}_+$, $f \in B(W)$, and $w \in W$.

We shall also consider the more general case where $w_0 \in W$ is chosen at random according to a given probability distribution. More precisely, on a probability space $(\Omega, \mathcal{K}, \mathbb{P}_{\lambda,p})$ let w_0 be a W-valued random variable with probability distribution $\lambda \in \mathrm{pr}(\mathcal{B}_W)$ [= the collection of all probability measures on $\mathcal{B}_W$], that is independent of the $\xi_i,\ i \in \mathbf{N}_+$, which always are i.i.d. with common $\mathbb{P}_{\lambda,p}$-distribution p. In this case, $(\zeta_n)_{n\in\mathbf{N}}$ defined by (2.2) is still a W-valued Markov process with initial distribution λ and transition function P. Consequently, its transition operator is always U. Clearly, the probability $\mathbb{P}_{\lambda,p}$ reduces to $\mathbb{P}_p$ when $\lambda = \delta_w$ = probability measure concentrated at some $w \in W$.

Note that U is a bounded linear operator of norm 1 on $B(W)$, which is a Banach space when endowed with the supremum norm

$$\|f\| = \sup |f(w)|, \quad f \in B(W).$$

Under a natural continuity assumption, namely that for p-almost all $x \in X$ the self-mapping $u_x : W \to W$ is continuous, the same assertion holds for U acting on $C(W)$ [= the linear space of complex-valued bounded continuous functions defined on W], also endowed with the supremum norm. The only thing needing proof is that U is now a Feller operator, to mean that $Uf \in C(W)$ for any $f \in C(W)$. To proceed, fix arbitrarily $w \in W$ and consider any sequence $(w_n)_{n\in\mathbf{N}_+}$ in W such that $w_n \to w$ as $n \to \infty$. Clearly, according to the assumption made, for any $f \in C(W)$ and any $x \in X$ not contained in the p-null exceptional set we have

$$\lim_{n\to\infty} f\left(u_x(w_n)\right) = f\left(u_x(w)\right).$$

Then, by bounded convergence,

$$\lim_{n\to\infty} \int_X p\left(\mathrm{d}x\right) f\left(u_x(w_n)\right) = \int_X p\left(\mathrm{d}x\right) f\left(u_x(w)\right),$$

that is,

$$\lim_{n\to\infty} Uf(w_n) = Uf(w).$$

As $w \in W$ has been arbitrarily chosen, we conclude that $Uf \in C\left(W\right)$.

Remark. **The assumption of the continuity of $u_x : W \to W$ for p-almost all $x \in X$ will be tacitly assumed throughout. As a rule, we shall only mention when it is not necessary**. $\square$

Clearly, U maps into itself the collection of $\mathcal{B}_W$-measurable extended real-valued functions f defined on W such that $Uf^+(w)$, and $Uf^-(w)$, $w \in W$, are not simultaneously equal to $+\infty$.

An important special case where U is well defined for possibly unbounded functions f is described below. Define

$$\ell(x) = \ell(x; d) = \mathrm{s}(u_x) := \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \frac{d(u_x(w'), u_x(w''))}{d(w', w'')}, \quad x \in X.$$

If the metric space W is assumed to be separable, then it is easy to see that the mapping $x \to \ell(x)$ of X into $\overline{\mathbf{R}}$ is $(\mathcal{X}, \mathcal{B}_{\overline{\mathbf{R}}})$-measurable. Assume that

$$\ell := \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \int_X \frac{d(u_x(w'), u_x(w''))}{d(w', w'')} p(\mathrm{d}x) < 1. \tag{2.4}$$

We clearly have

$$\ell \leq \int_X \ell(x) p(\mathrm{d}x).$$

Hence, if the integral in the inequality above is less that 1, then we also have $\ell < 1$, but the converse does not hold. Assume also that for some $w_0 \in W$ we have

$$\int_X d(w_0, u_x(w_0)) p(\mathrm{d}x) < \infty. \tag{2.5}$$

Under assumptions (2.4) and (2.5), the operator U takes $\mathrm{Lip}_1(W)$ into itself. See A1.2. For, (2.5) holds for any $w \in W$ in place of w_0 since

$$\begin{aligned} d(w, u_x(w)) &\leq d(w, w_0) + d(w_0, u_x(w_0)) + d(u_x(w_0), u_x(w)) \\ &\leq \left(\frac{d(u_x(w_0), u_x(w))}{d(w_0, w)} + 1 \right) d(w, w_0) + d(w_0, u_x(w_0)), \end{aligned}$$

which yields

$$\int_X d(w, u_x(w)) p(\mathrm{d}x) \leq (\ell+1) d(w_0, w) + \int_X d(w_0, u_x(w_0)) p(\mathrm{d}x) < \infty, \ w \in W. \tag{2.6}$$

Next, for any $f \in \mathrm{Lip}_1(W)$ we have

$$|f(u_x(w))| \leq |f(w)| + d(w, u_x(w)), \quad x \in X, \ w \in W,$$

hence

$$|Uf(w)| \leq \int_X |f(u_x(w))| p(\mathrm{d}x) < \infty, \quad w \in W,$$

while $\mathrm{s}(Uf) \le 1$ is an immediate consequence of (2.4).

Consider now another linear operator, closely related to U, defined on $\mathrm{pr}(\mathcal{B}_W)$ by

$$V\mu(A) = \int_W \mu(\mathrm{d}w)\, P(w, A), \quad A \in \mathcal{B}_W,$$

for any $\mu \in \mathrm{pr}(\mathcal{B}_W)$. Actually, this is a kind of adjoint of U, to mean that

$$(\mu, Uf) = (V\mu, f), \quad \mu \in \mathrm{pr}(\mathcal{B}_W), \; f \in B(W), \tag{2.7}$$

where (μ, f) is defined as the integral $\int_W f \mathrm{d}\mu$. Equation (2.7) is easily established by using Fubini's theorem. It is also easy to check that V can be expressed by means of an integral over X. We namely have

$$V\mu(A) = \int_X p(\mathrm{d}x)\, \mu u_x^{-1}(A), \quad A \in \mathcal{B}_W,$$

for any $\mu \in \mathrm{pr}(\mathcal{B}_W)$, where $\mu u_x^{-1}(A) := \mu\left(u_x^{-1}(A)\right)$, $x \in X$, $A \in \mathcal{B}_W$. Note that $V^n\mu(A) = \int_W \mu(\mathrm{d}w)\, P^n(w, A)$, $A \in \mathcal{B}_W$, or, alternatively,

$$V^n\mu(A) = \int_X \cdots \int_X p(\mathrm{d}x_1) \cdots p(\mathrm{d}x_n)\, \mu\left(u_{x_1} \circ \cdots \circ u_{x_n}\right)^{-1}(A), \quad A \in \mathcal{B}_W,$$

for any $n \in \mathbf{N}_+$ and $\mu \in \mathrm{pr}(\mathcal{B}_W)$. The probabilistic meaning of V^n is that $V^n\lambda(A) = \mathbb{P}_{\lambda,p}(\zeta_n \in A)$ for any $\lambda \in \mathrm{pr}(\mathcal{B}_W)$, $A \in \mathcal{B}_W$, and $n \in \mathbf{N}_+$. From the equation above we also have that

$$V^n\lambda(A) = \mathbb{P}_{\lambda,p}\left(u_{\xi_1} \circ \cdots \circ u_{\xi_n}(w_0) \in A\right) \tag{2.8}$$

for any $n \in \mathbf{N}_+$, $A \in \mathcal{B}_W$, and $\lambda \in \mathrm{pr}(\mathcal{B}_W)$, with $\mathbb{P}_{\lambda,p}(w_0 \in A) = \lambda(A)$.

The result below is well-known in the case where $f \in B(W)$, cf. (2.7). Its proof does not differ from that working when $f \in B(W)$, namely, Fubini's theorem.

Proposition 2.1. *If $\int_W Uf \mathrm{d}\mu$ exists for some real-valued $\mathcal{B}_W$-measurable function f and probability $\mu \in \mathrm{pr}(\mathcal{B}_W)$, then $\int_W f \mathrm{d}(V\mu)$ also exists and the two integrals are equal.*

In particular, Proposition 2.1 shows that in the case where U is a Feller operator the pair (U, V) is a Markov-Feller pair according to Zaharopol [50, p. 3].

The problem raised at the end of the preceding section, namely, the possibility of representing a given transition probability function

$$Q : W \times \mathcal{B}_W \to [0,1], \quad (w, A) \to Q(w, A),$$

as the transition probability function

$$P : W \times \mathcal{B}_W \to [0,1], \quad (w, A) \to P(w, A) = p(A_w),$$

of an IFS $\left((u_x)_{x \in X}, p\right)$, where $A_w = \{x \in X \mid u_x(w) \in A\}$, $w \in W$, can be also answered in the present more general case. Assume that (W, d) is a separable complete metric space. Then, with $X = (0,1)$ and with Λ the Lebesgue measure restricted to $\mathcal{B}_{(0,1)}$, there exists a $\left(\mathcal{B}_W \times \mathcal{B}_{(0,1)}, \mathcal{B}_W\right)$-measurable mapping $v : W \times (0,1) \to W$ such that

$$Q\,(w, A) = \Lambda\,(s \in (0,1) \mid v_s\,(w) \in A)\,,\ w \in W,\ A \in \mathcal{B}_W,$$

with $v_s(w) := v\,(w,s)$. In particular, if W is $\mathbf{R}$ (or a Borel subset of it), then one can take

$$v_s(w) = \inf\,\{y \in \mathbf{R} \mid Q(w, (-\infty, y] \geq s\}$$

for any $s \in (0,1)$, $w \in W$, and $A \in \mathcal{B}_W$. Explicit expressions for v do also exist in the general case $W \neq \mathbf{R}$. See Kifer [32, Theorem 1.1]. Earlier results can be found in Bergmann and Stoyan [8] and O'Brien [43]. See also Athreya and Stenflo [3], where it is shown that the condition on (W,d) to be a separable complete metric space can be replaced by that of being a standard Borel metric space, i.e., Borel measurably isomorphic to a Borel set on the real line. A still unsolved problem is whether nice solutions $(v_s)_{s\in X}$ do exist. For example, can the v_s, $s \in X$, be continuous or Lipschitz mappings? See Dubischar [16] for hints at this matter.

3. THE STATIONARY DISTRIBUTION: EXISTENCE AND UNIQUENESS

If U is a Feller operator and (W,d) is compact, then the Markov chain $(\zeta_n)_{\in\mathbf{N}}$ has invariant probabilities. See, e.g., Krengel [35, p. 178]. Nevertheless, it is important that the latter be approached in some sense by the n-step transition probabilities of the chain as $n \to \infty$. It is clear that such a convergence might only hold if extra conditions on the self-mappings u_x, $x \in X$, are imposed. Pursuing such an idea, we shall deal here with the asymptotic behaviour as $n \to \infty$ of the distribution of ζ_n under $\mathbb{P}_{\lambda,p}$. We shall see that, in our context, compactness of W is not necessary. In what follows the reader should refer to the Appendices A1 and A2 at the end.

The key result on which our approach is based is

Proposition 3.1. *Assume that* (2.4) *and* (2.5) *hold. Let* $\mu, \nu \in \mathrm{pr}\,(\mathcal{B}_W)$ *such that* $\rho_H\,(\mu,\nu) < \infty$. *Then*

$$\rho_H\,(V\mu, V\nu) \leq \ell\ \rho_H\,(\mu,\nu)\,.$$

Proof. We have already seen that under our assumptions the operator U takes $\mathrm{Lip}_1(W)$ into itself. By Proposition 2.1 we then have

$$\begin{aligned}(3.1)\qquad \rho_H\,(V\mu, V\nu) &= \sup\left\{\int_W f\mathrm{d}\,(V\mu) - \int_W f\mathrm{d}\,(V\nu)\ \middle|\ f \in \mathrm{Lip}_1(W)\right\}\\ &= \sup\left\{\int_W Uf\ \mathrm{d}\mu - \int_W Uf\mathrm{d}\nu\ \middle|\ f \in \mathrm{Lip}_1(W)\right\}.\end{aligned}$$

Consider the function $g = Uf/\ell$. Note that $g \in \mathrm{Lip}_1(W)$ since for any $w', w'' \in W,\ w' \neq w''$, by the very definition of ℓ we have

$$\begin{aligned}\frac{|g(w') - g(w'')|}{d(w', w'')} &= \frac{1}{\ell}\left|\int_X \frac{f(u_x(w')) - f(u_x(w''))}{d(w', w'')} p(\mathrm{d}x)\right| \\ &\leq \frac{1}{\ell}\int_X \frac{d(u_x(w'), u_x(w''))}{d(w', w'')} p(\mathrm{d}x) \leq 1.\end{aligned}$$

Then, by (3.1),

$$\begin{aligned}\rho_H(V\mu, V\nu) &= \ell\ \sup\left\{\int_W g\ \mathrm{d}\mu - \int_W g\ \mathrm{d}\nu \ \Big|\ g = \frac{Uf}{\ell},\ f \in \mathrm{Lip}_1(W)\right\} \\ &\leq \ell\ \sup\left\{\int_W f\mathrm{d}\mu - \int_W f\mathrm{d}\nu \ \Big|\ f \in \mathrm{Lip}_1(W)\right\} = \ell\ \rho_H(\mu, \nu),\end{aligned}$$

and the proof is complete. □

Clearly, A1.2 and the result just proved imply

COROLLARY 3.2. *Under the assumptions in Proposition* 3.1 *we have*

$$\rho_L(V^n\mu, V^n\nu) \leq \ell^n \rho_H(\mu, \nu)$$

for any $n \in \mathbf{N}_+$.

By only using contraction properties of the operators U and V we can now prove the important result below. Cf. Iosifescu [28].

THEOREM 3.3. *Let* (W, d) *be a separable complete metric space. Assume that* (2.4) *and* (2.5) *hold. Then the Markov chain* $(\zeta_n)_{n\in\mathbf{N}}$ *has a unique stationary distribution* π *and*

$$\rho_L(P^n(w, \cdot), \pi) \leq \frac{\ell^n}{1-\ell}\int_X d(w, u_x(w))\, p(\mathrm{d}x) \tag{3.2}$$

for any $n \in \mathbf{N}$ *and* $w \in W$. *On* $(\Omega, \mathcal{K}, \mathbb{P}_{\pi,p})$ *the sequence* $(\zeta_n)_{n\in\mathbf{N}}$ *is an ergodic strictly stationary process.*

Proof. Step 1. Let $\mu \in \mathrm{pr}(\mathcal{B}_W)$ such that $\rho_H(\mu, V\mu) < \infty$. For the existence of such a μ, see further Step 2. By Corollary 3.2, for any $m, n \in \mathbf{N}_+$ we can write

$$\begin{aligned}\rho_L\left(V^{n+m}\mu, V^n\mu\right) &\leq \sum_{k=0}^{m-1}\rho_L\left(V^{n+k}\mu, V^{n+k+1}\mu\right) \\ &\leq \sum_{k=0}^{m-1}\ell^{n+k}\rho_H(\mu, V\mu) \leq \frac{\ell^n}{1-\ell}\rho_H(\mu, V\mu).\end{aligned} \tag{3.3}$$

Since (W, d) is complete, so is $(\mathrm{pr}(\mathcal{B}_W),\ \rho_L)$, see A1.2. Hence the sequence $(V^n\mu)_{n\in\mathbf{N}}$ is convergent in $(\mathrm{pr}(\mathcal{B}_W),\ \rho_L)$ to some, say, $\pi \in \mathrm{pr}(\mathcal{B}_W)$.

Consider another $\nu \in \mathrm{pr}\,(\mathcal{B}_W)$ such that $\rho_H\,(\mu,\nu) < \infty$. Then, since

$$\begin{aligned}\rho_H\,(\nu, V\nu) &\leq \rho_H\,(\nu,\mu) + \rho_H(\mu, V\mu) + \rho_H\,(V\mu, V\nu)\\ &\leq (\ell+1)\,\rho_H\,(\mu,\nu) + \rho_H\,(\mu, V\mu)\,,\end{aligned}$$

we also have $\rho_H\,(\nu, V\nu) < \infty$. This allows to conclude that $(V^n\nu)_{n\in\mathbf{N}}$ converges to the same π as for any $n \in \mathbf{N}_+$ we have

$$\rho_L\,(V^n\nu,\pi) \leq \rho_L\,(V^n\mu,\pi) + \rho_L\,(V^n\mu, V^n\nu) \leq \rho_L\,(V^n\mu,\pi) + \ell^n\rho_H\,(\mu,\nu)\,.$$

To sum up, we have proved that if $\mu \in \mathrm{pr}\,(\mathcal{B}_W)$ satisfies the condition $\rho_H\,(\mu, V\mu) < \infty$, then there exists $\pi = \pi\,(\mu)$ such that

$$\rho_L\,(V^n\mu,\pi) \leq \frac{\ell^n}{1-\ell}\,\rho_H\,(\mu, V\mu)\,, \quad n \in \mathbf{N}_+. \tag{3.4}$$

[The last inequality follows at once from (3.3).] The same conclusion holds, with the same π, for any other $\nu \in \mathrm{pr}\,(\mathcal{B}_W)$ for which $\rho_H\,(\mu,\nu) < \infty$.

It is easy to prove that $\pi = V\pi$, that is, π is a stationary distribution for $(\zeta_n)_{n\in\mathbf{N}}$. We have $\rho_L\,(V\mu, V\nu) \leq \rho_L(\mu,\nu)$, $\mu,\nu \in \mathrm{pr}\,(\mathcal{B}_W)$, by the very definition of the distance ρ_L on account of Proposition 2.1. Then $\rho_L\left(V^{n+1}\mu, V\pi\right) \leq \rho_L\,(V^n\mu,\pi) \to 0$ as $n \to \infty$. Hence both $V\pi$ and π are equal to the limit in $(\mathrm{pr}\,(\mathcal{B}_W)\,,\rho_L)$ of the sequence $(V^n\mu)_{n\in\mathbf{N}}$, that is, $\pi = V\pi$.

Step 2. Clearly, δ_w (probability measure concentrated at $w \in W$) satisfies $\rho_H(\delta_w, V\delta_w) < \infty$ for any w since

$$\begin{aligned}\rho_H\,(\delta_w, V\delta_w) &= \sup\left\{ f\,(w) - \int_W f\mathrm{d}\,(V\delta_w)\;\middle|\; f \in \mathrm{Lip}_1(W)\right\}\\ &= \sup\left\{f(w) - Uf(w) \mid f \in \mathrm{Lip}_1(W)\right\} \quad \text{(by Proposition 2.1)}\\ &= \sup\left\{\int_X \left(f(w) - f\,(u_x(w))\right)p\,(\mathrm{d}x)\;\middle|\; f \in \mathrm{Lip}_1(W)\right\}\\ &\leq \int_X d\,(w, u_x(w))\,p\,(\mathrm{d}x) < \infty \quad \text{(by (2.5))}.\end{aligned}$$

It follows by Step 1 that the limiting $\pi\,(\delta_w) := \pi$ is the same for all $w \in W$ since

$$\rho_H\,(\delta_{w'},\delta_{w''}) \leq \sup\left\{f(w') - f\left(w''\right) \mid f \in \mathrm{Lip}_1(W)\right\} \leq d\left(w', w''\right) < \infty$$

for any $w', w'' \in W$.

Next, any finite linear combination $\overline{\mu} = \sum q_j\delta_{w_j}$ with positive rational coefficients such that $\sum q_j = 1$ satisfies the condition $\rho_H\,(\overline{\mu}, V\overline{\mu}) < \infty$ since, as is easy to see,

$$\rho_H\,(\overline{\mu}, V\overline{\mu}) \leq \sum q_j\rho_H\left(\delta_{w_j}, V\delta_{w_j}\right).$$

Moreover, $(\mathrm{pr}(\mathcal{B}_W),\rho_L)$ is separable since (W,d) was assumed to be, see A1.2, and it appears that the class of probability measures $\overline{\mu} = \sum q_j\delta_{w_j}$ just considered is dense in $(\mathrm{pr}(\mathcal{B}_W),\rho_L)$ if we start with a countable dense subset

$\{w_j \mid j \in \mathbf{N}_+\}$ in W. Cf. Hoffmann-Jørgensen [26, p. 83]. Let then $\lambda \in \mathrm{pr}\,(\mathcal{B}_W)$ be *arbitrary* and for any $\varepsilon > 0$ consider a probability measure $\overline{\mu}_\varepsilon$ from that class such that

$$\rho_L\,(\lambda, \overline{\mu}_\varepsilon) < \varepsilon.$$

Since $\lim_{n\to\infty} \rho_L\,(V^n\overline{\mu}_\varepsilon, \pi) = 0$ by Step 1 and

$$\begin{aligned}\rho_L\,(V^n\lambda,\ \pi) &\leq \rho_L\,(V^n\overline{\mu}_\varepsilon, \pi) + \rho_L\,(V^n\lambda,\ V^n\overline{\mu}_\varepsilon)\\ &\leq \rho_L\,(V^n\overline{\mu}_\varepsilon, \pi) + \rho_L\,(\lambda, \overline{\mu}_\varepsilon)\,, \quad n \in \mathbf{N}_+,\end{aligned}$$

we have

$$\limsup_{n\to\infty} \rho_L\,(V^n\lambda, \pi) \leq \varepsilon.$$

As $\varepsilon > 0$ is arbitrary, we conclude that *the sequence* $(V^n\lambda)_{n\in\mathbf{N}}$ *also converges to* π *in* $(\mathrm{pr}\,(\mathcal{B}_W)\,, \rho_L)$.

Clearly, (3.2) follows from (3.4) with $\mu = \delta_w$, $w \in W$. For an arbitrary $\lambda \in \mathrm{pr}\,(\mathcal{B}_W)$, a similar upper bound for $\rho_L(V^n\lambda, \pi)$ holds if we assume that

$$\int_W \lambda\,(\mathrm{d}w) \int_X d\,(w, u_x\,(w))\,p\,(\mathrm{d}x) < \infty. \tag{3.5}$$

Step 3. The uniqueness of π as stationary measure, $\pi = V\pi$, follows now easily. If $\pi' \in \mathrm{pr}\,(\mathcal{B}_W)$ satisfies $\pi' = V\pi'$, then by Step 2 we have

$$\lim_{n\to\infty} \rho_L\left(V^n\pi', \pi\right) = 0$$

and, at the same time, $V^n\pi' = \pi',\ n \in \mathbf{N}_+$. Hence $\pi' = \pi$.

Next, the ergodicity of π, that is, $(\zeta_n)_{n\in\mathbf{N}}$ is an ergodic strictly stationary sequence on $(\Omega, \mathcal{K}, \mathbb{P}_{\pi,p})$, follows from the very uniqueness of π. See, e.g., Proposition 2.4.3 in Hernández-Lerma and Lasserre [24]. □

Corollary 3.4. *Under the assumptions in Theorem* 3.3*, for any real-valued bounded Lipschitz function* f *on* W *we have*

$$\left|U^n f(w) - \int_W f \mathrm{d}\pi\right| \leq \frac{\ell^n}{1-\ell} \int_X d((w, u_x(w))\,p\,(\mathrm{d}x) \max\,(\mathrm{osc} f, \mathrm{s}(f))\,,$$

$n \in \mathbf{N}_+$, $w \in W$, *with* $\mathrm{osc} f = \sup_{w\in W} f(w) - \inf_{w\in W} f(w)$.

Proof. Clearly, if f is constant there is nothing to prove. If $f \neq$ const. then it is enough to note that for the function

$$g := \frac{f - \inf_{w\in W} f(w)}{\max\,(\mathrm{osc} f, \mathrm{s}(f))} \in \mathrm{Lip}_1(W)$$

we have $0 \leq g \leq 1$, and to recall the definition of $\rho_L\,(V^n\delta_w, \pi)$. □

Remarks. **1.** Since $\lim_{n\to\infty} \rho_L(V^n\lambda, \pi) = 0$ for any $\lambda \in \mathrm{pr}(\mathcal{B}_W)$ (see Step 2 in the proof of Theorem 3.3), by equation (2.8) the backward process

$$\widetilde{\zeta_n^{w_0}} = u_{\xi_1} \circ \cdots \circ u_{\xi_n}(w_0), \quad n \in \mathbf{N}_+,$$

converges in distribution under $\mathbb{P}_{\lambda,p}$ as $n \to \infty$ to π, with $\lambda \in \mathrm{pr}(\mathcal{B}_W)$ and $\mathbb{P}_{\lambda,p}(w_0 \in A) = \lambda(A)$, $A \in \mathcal{B}_W$, that is,

$$\lim_{n\to\infty} \mathbb{P}_{\lambda,p}\big(\widetilde{\zeta_n^{w_0}} \in A\big) = \pi(A)$$

for any $A \in \mathcal{B}_W$ whose boundary is π-null. We shall show more, namely, that for any fixed $w \in W$ the sequence $(\widetilde{\zeta_n^{w}})_{n\in\mathbf{N}}$ converges $\mathbb{P}_p$-a.s. at a geometric rate as $n \to \infty$ to a W-valued random variable ζ_∞ such that $\mathbb{P}_p(\xi_\infty \in A) = \pi(A)$, $A \in \mathcal{B}_W$. See further Theorems 4.1 and 4.2.

2. As for the nature of the stationary distribution π, according to results of Dubins and Freedman [15] on Markov operators, it should be of pure type under appropriate assumptions. For example, if for some probability measure $m \in \mathrm{pr}(\mathcal{B}_W)$ either $mu_x^{-1} \ll m$ for any $x \in X$ or, when X is countable, $\nu \perp m$ implies $\nu u_x^{-1} \perp m$ for any $x \in X$ whatever $\nu \in \mathrm{pr}(\mathcal{B}_W)$, then π is either absolutely continuous or purely singular with respect to m. This also applies to similar further results as, e.g., Theorems 3.5 and 3.6 or Corollaries 3.7 and 3.8.

The type of π appears to be related to the so-called open set condition (OSC). The family $(u_x)_{x\in X}$ is said to satisfy the OSC if there is a non-empty bounded open set $V \subset W$ such that $u_x(V) \subset V$ for any $x \in X$ and $u_{x'}(V) \cap u_{x''}(V) = \emptyset$ for any $x', x'' \in X$, $x' \neq x''$. See, e.g., Lau and Ngai [36] and the references therein. $\square$

A more general version of Theorem 3.3 is obtained using the fact that d^α is still a metric in W for any $0 < \alpha \le 1$. [It is enough to note that if $a, b, c \ge 0$ and $c \le a + b$, then $c^\alpha \le (a+b)^\alpha \le a^\alpha + b^\alpha$.] Write then (see A1.2) $\rho_{L,\alpha}$ and $\mathrm{Lip}_1^\alpha(W)$ for the items associated with the metric space (W, d^α), which correspond for $\alpha = 1$ to ρ_L and $\mathrm{Lip}_1(W)$, respectively. (Remark that $\mathcal{B}_W$ is not altered when replacing d by d^α.) Clearly, $\ell(x; d^\alpha) = [\ell(x;\alpha)]^\alpha := \ell^\alpha(x)$, $x \in X$, and then the conditions corresponding to (2.4) and (2.5) are

$$\ell_\alpha := \sup_{\substack{w'\neq w''\\ w',w''\in W}} \int_X \frac{d^\alpha(u_x(w'), u_x(w''))}{d^\alpha(w', w'')} p(\mathrm{d}x) < 1 \tag{3.6}$$

and, respectively,

$$\int_X d^\alpha(w_0, u_x(w_0))\, p(\mathrm{d}x) < \infty \tag{3.7}$$

for some $w_0 \in W$, hence for all $w_0 \in W$.

We can now state

THEOREM 3.5. *Let (W,d) be a separable complete metric space. Assume that* (3.6) *and* (3.7) *hold. Then the Markov chain $(\zeta_n)_{n\in\mathbf{N}}$ has a unique stationary distribution π and*

$$\rho_L\left(P^n\left(w,\ \cdot\right),\pi\right) \le \frac{\ell_\alpha^n}{1-\ell_\alpha}\int_X d^\alpha\left(w,u_x\left(w\right)\right)p\left(\mathrm{d}x\right) \tag{3.8}$$

for any $n \in \mathbf{N}_+$ and $w \in W$. On $(\Omega,\mathcal{K},\mathbb{P}_{\pi,p})$ the sequence $(\zeta_n)_{n\in\mathbf{N}}$ is an ergodic strictly stationary process.

Proof. It follows from Theorem 3.3 that (3.8) holds with $\rho_{L,\alpha}$ in place of ρ_L. The validity of (3.8) will follow from the inequality $\rho_{L,\alpha} \ge \rho_L$ for any $0<\alpha<1$. We shall in fact prove that

$$\{f \mid f\in \mathrm{Lip}_1(W),\ 0\le f\le 1\} \subset \{f \mid f\in \mathrm{Lip}_1^\alpha(W),\ 0\le f\le 1\} \tag{3.9}$$

for any $0<\alpha\le 1$, which clearly implies $\rho_{L,\alpha}\ge\rho_L$.

To proceed, note that if $f\in\mathrm{Lip}_1(W)$ $\bigl(=\mathrm{Lip}_1^1(W)\bigr)$ and $0\le f\le 1$, then for any $0<\alpha\le 1$ we can write

$$\sup_{w'\neq w''}\frac{|f(w')-f(w'')|}{d^\alpha(w',w'')} =$$

$$= \max\left(\sup_{\substack{w'\neq w''\\ d(w',w'')\le 1}}\frac{|f(w')-f(w'')|}{d^\alpha(w',w'')},\ \sup_{d(w',w'')>1}\frac{|f(w')-f(w'')|}{d^\alpha(w',w'')}\right) \le$$

$$\le \max\left(\sup_{\substack{w'\neq w''\\ d(w',w'')\le 1}}\frac{|f(w')-f(w'')|}{d(w',w'')},\ \text{some quantity not exceeding 1}\right) \le$$

$$\le \max\left(\mathrm{s}(f),1\right)\le 1.$$

(We used the inequality $x^\alpha > x$ which holds for $0<\alpha,\ x<1$.) Hence $f\in \mathrm{Lip}_1^\alpha(W)$, showing that (3.9) holds. $\square$

Remarks. **1.** It is obvious that the assumptions in Theorem 3.5 are weaker than those in Theorem 3.3, so that the former is a real generalization of the latter. Also, the result corresponding to Corollary 3.4 under assumptions (2.4) and (2.5) also holds. Clearly, both Theorem 3.5 and the corresponding corollary have versions holding when (3.5) is replaced by the condition

$$\int_W \lambda\left(\mathrm{d}w\right)\int_X d^\alpha\left(w,u_x(w)\right)p\left(\mathrm{d}x\right) < \infty. \tag{3.10}$$

For example, if (3.6), (3.7), and (3.10) hold, then

$$\rho_L\left(V^n\lambda,\pi\right) \le \frac{\ell_\alpha^n}{1-\ell_\alpha}\int_W \lambda\left(\mathrm{d}w\right)\int_X d^\alpha\left(w,u_x(w)\right)p\left(\mathrm{d}x\right)$$

for all $n \in \mathbf{N}_+$.

2. To compare Theorem 3.5 and Theorem 5.1 in Diaconis and Freedman [13, pp. 58–59] let us first note (see, e.g., Hewitt and Stromberg [25, p. 201]) that the condition

$$L_\alpha := \int_X \ell^\alpha\left(x\right)p\left(\mathrm{d}x\right) < 1 \tag{3.11}$$

for some $0 < \alpha \le 1$, which is stronger than (3.6), implies the inequality

$$\int_X \log\left(\ell\left(x\right)\right)p\left(\mathrm{d}x\right) < 0. \tag{3.12}$$

Conversely, if $L_\beta := \int_X \ell^\beta\left(x\right)p\left(\mathrm{d}x\right) < \infty$ for some $\beta > 0$ and (3.12) holds, then there exists $\alpha > 0$ such that $L_\alpha < 1$.

The assumptions in Theorem 5.1 in Diaconis and Freedman (op.cit.) are (3.12) and a so-called "algebraic-tail" condition on ℓ and d which amounts to the existence of positive constants a and b such that

$$p\left(\{x|\ell\left(x\right) > y\}\right) < ay^{-b},\ p\left(\{x|d(w_0,\ u_x\left(w_0\right)) > y\}\right) < ay^{-b} \tag{3.13}$$

for $y > 0$ large enough and some $w_0 \in W$, hence for all $w_0 \in W$. We are going to prove that these assumptions are equivalent to (3.11) in conjunction with (3.7), so that they are stronger than those in Theorem 3.5.

First, on account of the equation

$$E\eta = \int_0^\infty P\left(\eta > y\right)\mathrm{d}y \tag{3.14}$$

which holds for any non-negative random variable η, it is clear that (3.11) and (3.7) imply both (3.12) and, via Markov's inequality, (3.13). Second, if (3.13) holds, then for any $\alpha > 0$ we have

$$p\left(\{x|\ell^\alpha\left(x\right) > y\}\right) < ay^{-b/\alpha},\quad p\left(\{x|d^\alpha\left(w_0,u_x\left(w_0\right)\right) > y\}\right) < ay^{-b/\alpha}$$

for $y > 0$ large enough. Choosing $\alpha < \min\left(b,1\right)$, it follows from (3.14) that both L_α and $\int_X d^\alpha\left(w_0,u_x\left(w_0\right)\right)\mathrm{d}x$ are finite. But $L_\alpha < \infty$ in conjunction with (3.12) implies the existence of $0 < \alpha' < \alpha$ such that $L_{\alpha'} < 1$, as has just been mentioned. The proof is complete.

3. The average contractibility condition (3.6) can be weakened to average contractibility after a given number of steps. To introduce it, for any $n \in \mathbf{N}_+$ and $x^{(n)} = (x_1,\ldots,x_n) \in X^n$ put

$$u_{x^{(n)}} = u_{x_n}\circ\cdots\circ u_{x_1}$$

and consider the IFS

$$\left(p_n, (u_{x^{(n)}})_{x^{(n)} \in X^n}\right),$$

where p_n denotes the nth product measure of p with itself. Clearly, for any fixed $n \in \mathbf{N}_+$ we have a new IFS for which condition (3.6) reads as

$$\ell_{\alpha,n} := \sup_{\substack{w' \neq w'' \\ w', w'' \in W}} \int_{X^n} \frac{d^\alpha\left(u_{x^{(n)}}(w'), u_{x^{(n)}}(w'')\right)}{d^\alpha(w', w'')} p_n\big(\mathrm{d}x^{(n)}\big) < 1. \tag{3.15}$$

It is not difficult to check that $(\ell_{\alpha,n})_{n \in \mathbf{N}_+}$ is a submultiplicative sequence, that is,

$$\ell_{\alpha,m+n} \leq \ell_{\alpha,m}\ \ell_{\alpha,n}, \quad m, n \in \mathbf{N}_+.$$

Hence, if $\ell_{\alpha,k} \leq 1$ for some $k \in \mathbf{N}_+$, then $(\ell_{\alpha,nk})_{n \in \mathbf{N}_+}$ is a non-increasing sequence. In particular, it follows that condition (3.15) for some $n \geq 2$ is weaker than the condition $\ell_{\alpha,1} < 1$, that is, (3.6).

It is easy to see that Theorem 3.5 carries over to an IFS satisfying condition (3.15) for some fixed $n = n_0$ together with the condition

$$\int_{X^{n_0}} d^\alpha\left(w_0, u_{x^{(n_0)}}(w_0)\right) p_{n_0}\big(\mathrm{d}x^{(n_0)}\big) < \infty. \tag{3.16}$$

for some $w_0 \in W$, hence for all $w_0 \in W$. The latter corresponds to condition (3.7) and reduces to it when $n_0 = 1$. More precisely, the following result holds.

THEOREM 3.6. *Let (W, d) be a separable complete metric space. Assume that* (3.15) *and* (3.16) *hold for some fixed $n_0 \in \mathbf{N}_+$. Then the Markov chain $(\zeta_{nn_0})_{n \in \mathbf{N}}$ has a unique stationary distribution π and*

$$\rho_L\left(P^{nn_0}(w\ \cdot), \pi\right) \leq \frac{\ell^n_{\alpha,n_0}}{1 - \ell_{\alpha,n_0}} \int_{X^{n_0}} d^\alpha\big(w, u_{x^{(n_0)}}(w)\big)\, p_{n_0}\big(\mathrm{d}x^{(n_0)}\big) \tag{3.17}$$

for any $n \in \mathbf{N}_+$ and $w \in W$. On $(\Omega, \mathcal{K}, \mathbb{P}_{\pi,p})$ the sequence $(\zeta_{nn_0})_{n \in \mathbf{N}}$ is an ergodic stationary process.

Note that this is just a transcription of Theorem 3.5 for the IFS $\big(p_{n_0}, \big(u_{x^{(n_0)}}\big)_{x^{(n_0)} \in X^{n_0}}\big)$. It does not yield a stationary distribution for the 'whole' Markov chain $(\zeta_n)_{n \in \mathbf{N}}$. To ensure that π occurring in the statement above is a stationary distribution for $(\zeta_n)_{n \in \mathbf{N}}$, more assumptions are to be made. We namely first have

COROLLARY 3.7. *Let $n_0 \geq 2$. Under the assumptions in Theorem* 3.6, *if for some $1 \leq r < n_0$ we have*

$$\int_{X^r} d^\alpha\left(w, u_{x^{(r)}}(w)\right) p_r\big(\mathrm{d}x^{(r)}\big) < \infty \tag{3.18}$$

for any $w \in W$, then π is the unique stationary distribution of the Markov chain $(\zeta_{nn_0+r})_{n\in\mathbf{N}}$ and

$$\rho_L\left(P^{nn_0+r}(w,\cdot),\pi\right) \leq \tag{3.19}$$

$$\leq \ell^n_{\alpha,n_0}\left(\frac{\int_{X^{n_0}} d^\alpha\big(w,u_{x^{(n_0)}}(w)\big)\,p_{n_0}\big(\mathrm{d}x^{(n_0)}\big)}{1-\ell_{\alpha,n_0}} + \int_{X^r} d^\alpha\left(w,u_{x^{(r)}}(w)\right)p_r\big(\mathrm{d}x^{(r)}\big)\right)$$

for any $n \in \mathbf{N}_+$ and $w \in W$.

Proof. We clearly have

$$\rho_L\left(P^{nn_0+r}(w,\cdot),\pi\right) \leq \tag{3.20}$$

$$\leq \rho_L\left(P^{nn_0+r}(w,\cdot),P^{nn_0}(w,\cdot)\right) + \rho_L\left(P^{nn_0}(w,\cdot)),\pi\right))$$

for any $w \in W$ and $n \in \mathbf{N}_+$.

Coming back to the proof of Theorem 3.5 and using Corollary 3.2 we can write

$$\rho_L\left(P^{nn_0+r}(w,\cdot),P^{nn_0}(w,\cdot)\right) = \rho_L\left(V^{nn_0+r}\delta_w,V^{nn_0}\delta_w\right) \leq \tag{3.21}$$

$$\leq \rho_{L,\alpha}\left(V^{nn_0+r}\delta_w,V^{nn_0}\delta_w\right) \leq \ell^n_{\alpha,n_0}\rho_{H,\alpha}\left(V^r\delta_w,\delta_w\right)$$

while

$$\rho_{H,\alpha}\left(V^r\delta_w,\delta_w\right) = \sup\left\{U^rf(w)-f(w) \mid f\in \mathrm{Lip}^\alpha_1(W)\right\} \leq \tag{3.22}$$

$$\leq \sup_{f\in\mathrm{Lip}^\alpha_1(W)}\int_{X^r} p\big(\mathrm{d}x^{(r)}\big)\left|f(w)-f\left(u_{x^{(r)}}(w)\right)\right| \leq$$

$$\leq \int_{X^r} d^\alpha\left(w,u_{x^{(r)}}(w)\right)p_r\big(\mathrm{d}x^{(r)}\big) < \infty$$

for any $w \in W$.

Now, (3.19) follows from (3.17), (3.20), (3.21), and (3.22). □

Let us note that as in the case $n_0 = 1$, condition (3.16) (for just one $w_0 \in W$) in conjunction with (3.15) implies that the former holds for *any* $w_0 \in W$. In the case of (3.18), assumed to hold for just one $w \in W$, a similar conclusion would follow when assuming in addition that

$$\sup_{w'\neq w''}\int_{X^r}\frac{d^\alpha\left(u_{x^{(r)}}(w'),u_{x^{(r)}}(w'')\right)}{d^\alpha\big(w',w''\big)}\,p_r\big(\mathrm{d}x^{(r)}\big) < \infty.$$

Clearly, such a condition is not implied by only (3.15), as simple examples show.

Corollary 3.8. *Let $n_0 \geq 2$. Under the assumptions in Theorem* 3.6 *in conjunction with* (3.18) *for any $1 \leq r < n_0$, we have*

$$\rho_L \left(P^n \left(w, \cdot\right), \pi\right) \leq \tag{3.23}$$

$$\ell_{\alpha,n_0}^{\lfloor \frac{n+1}{n_0} \rfloor - 1} \left(\frac{\int\limits_{X^{n_0}} d^{\alpha}\big(w, u_{x^{(n_0)}}(w)\big) p_{n_0}\big(\mathrm{d}x^{(n_0)}\big)}{1 - \ell_{\alpha,n_0}} + \max_{1 \leq r < n_0} \int\limits_{X^r} d^{\alpha}\big(w, u_{x^{(r)}}(w)\big) p_r\big(\mathrm{d}x^{(r)}\big) \right)$$

for any $n \geq 2n_0 - 1$ and $w \in W$. The Markov chain $(\zeta_n)_{n \in \mathbf{N}}$ has π as unique stationary distribution and is an ergodic strictly stationary process on $(\Omega, \mathcal{K}, \mathbb{P}_{\pi,p})$.

Proof. This follows from Theorem 3.6 and Corollary 3.7 taking into account that both (3.17) and (3.19) hold actually with $\rho_{L,\alpha}$ in place of ρ_L, see the proof of Theorem 3.5. Next, we have to note that $\rho_{L,\alpha}\left(P^n\left(w,\cdot\right)\right) = \rho_{L,\alpha}\left(V^n \delta_w, \pi\right)$, $w \in W$, $n \in \mathbf{N}_+$, and then follow the reasoning from Steps 2 and 3 in the proof of Theorem 3.4. $\square$

4. A natural and interesting question now arises. What does it happen when condition (3.15) does not hold for any $n \in \mathbf{N}_+$, that is, if $\ell_{\alpha,n} \geq 1$ for any $n \in \mathbf{N}_+$?

First, there is an interesting special case where $\ell_{\alpha,n} = 1$ for *any* $n \in \mathbf{N}_+$, namely, that of $W = X = \mathbf{R}_+$, $u_x(w) = |w - x|$, $w, x \in \mathbf{R}_+$, while the probability p on $\mathcal{B}_{\mathbf{R}_+}$ is such that $0 < \mathbb{E}_p(\xi_1) < \infty$. It can be proved that $\pi \in \mathrm{pr}\left(\mathcal{B}_{\mathbf{R}_+}\right)$ given by

$$\pi(A) = \int_A \frac{\mathbb{P}_p\left(\xi_1 > x\right) \mathrm{d}x}{\mathbb{E}_p\left(\xi_1\right)}, \quad A \in \mathcal{B}_{\mathbf{R}_+},$$

is the only stationary probability distribution for the Markov chain $(\zeta_n)_{n \in \mathbf{N}}$ when ξ_1 is not supported by a lattice. This case has been considered in Feller's 1971 classical book. The result above first appears in Knight [33] and Leguesdron [38]. A recent treatment based on the reversed sequence $(\widetilde{\zeta_n^{w_0}})_{n \in \mathbf{N}}$ has been given by Abrams *et al.* [2]. Very few is known in the lattice case and no rate of convergence (if any) to π in the non-lattice case is given.

Coming back to the case $\ell_{\alpha,n} \geq 1$ for all $n \in \mathbf{N}_+$, if $\ell_{\alpha,n} > 1$ for at least one $n \in \mathbf{N}_+$, then we should necessarily have $\ell_{\alpha,1} > 1$ by the submultiplicativity of the sequence $(\ell_{\alpha,n})_{n \in \mathbf{N}_+}$. Our guess is that if $\ell_{\alpha,n} > 1$ for infinitely many $n \in \mathbf{N}_+$, then there cannot exist a stationary probability π for $(\zeta_n)_{n \in \mathbf{N}}$.

5. It is possible to ensure the existence of the stationary distribution π for $(\zeta_n)_{n \in \mathbf{N}}$ without assuming global contraction and drift conditions. Instead, some local contraction conditions and appropriate drift conditions can be considered.

For example, Jarner and Tweedie [30] considered a separable complete metric space (W, d) with finite diameter, that is,

$$\sup_{w',w''\in W} d\left(w', w''\right) < \infty, \tag{3.24}$$

and assumed that

(i) the maps u_x, $x \in X$, are "non-separating on average", to mean that

$$\mathbb{E}_p\big(d(\zeta_1^{w'}, \zeta_1^{w''})\big) \le d\big(w', w''\big) \tag{3.25}$$

for all $w', w'' \in W$;

(ii) there exist a positive number $r < 1$ and a set $C \in \mathcal{B}_W$ such that contraction occurs after reaching C, to mean that

$$\mathbb{E}_p\left(d\left(\zeta^{w'}_{\tau_C(w')\vee\tau_C(w'')}, \zeta^{w''}_{\tau_C(w')\vee\tau_C(w'')}\right)\right) \le rd\left(w', w''\right) \tag{3.26}$$

for all $w', w'' \in W$, where

$$\tau_C(w) = \inf\left\{n \in \mathbf{N}_+ \mid \zeta_n^w \in C\right\}, \quad w \in W;$$

(iii) there exists a measurable function $L : W \to [1, \infty)$ such that $\sup\limits_{w\in C} L(w) < \infty$ and for some positive constants $q < 1$ and a the inequality

$$UL(w) = \int_X p\left(\mathrm{d}x\right) L\left(u_x\left(w\right)\right) \le qL(w) + aI_C(w) \tag{3.27}$$

holds for all $w \in W$.

These authors showed that under assumptions (i) through (iii) the conclusions of our Theorem 3.5 all hold with a convergence rate $O\big(b^n L^{\frac{1}{2}}(w)\big)$, $w \in W$, as $n \to \infty$, for some positive constant $b < 1$, with the constant implied in O independent of $n \in \mathbf{N}_+$ and $w \in W$.

It is clear that in the special case $C = W$ assumptions (i)–(ii) reduce to the only condition

$$\mathbb{E}_p\left(d\big(\zeta_1^{w'}, \zeta_1^{w''}\big)\right) \le rd(w', w''), \quad w', w'' \in W,$$

for some positive constant $r < 1$, that is, to condition (2.4) while (iii) is satisfied with $L \equiv 1$. As (3.24) implies (2.5), the case $C = W$ is covered by Theorem 3.5. On the other hand, a condition like (ii) seems quite difficult to be checked in the case where $C \ne W$. $\square$

4. ALMOST SURE CONVERGENCE PROPERTIES

We now come back to Remark 1 following Corollary 3.4 concerning the convergence in distribution of the backward process

$$\widetilde{\zeta_n^{w_0}} = u_{\xi_1} \circ \cdots \circ u_{\xi_n}\left(w_0\right), \quad n \in \mathbf{N}_+, \tag{4.1}$$

to the stationary distribution π under $\mathbb{P}_{\lambda,p}$, with $\lambda \in \mathrm{pr}(\mathcal{B}_W)$ and $\mathbb{P}_{\lambda,p}(w_0 \in A) = \lambda(A)$, $A \in \mathcal{B}_W$. We shall namely prove almost sure convergence properties of this process under the assumptions already made.

Before proceeding, we shall recall a result of Letac [39] that reads as follows.

PROPOSITION 4.1 (Letac's lemma). *If for p-almost all $x \in X$ the mapping $u_x : W \to W$ is continuous and if $\zeta_\infty := \lim_{n\to\infty} \widetilde{\zeta_n^{w_0}}$ exists $\mathbb{P}_p$-a.s. and does not depend on $w_0 \in W$, then the probability distribution $\mu = \mathbb{P}_p \zeta_\infty^{-1}$ of ζ_∞ under $\mathbb{P}_p$ is the only stationary distribution of $(\zeta_n^{w_0})_{n\in\mathbf{N}_+}$.*

The *proof* of this result is very simple. Let $\pi_n^{w_0}$ denote the probability distribution of both $\zeta_n^{w_0}$ and $\widetilde{\zeta_n^{w_0}}$, $n \in \mathbf{N}_+$. We clearly have $\pi_n^{w_0} = V\pi_{n-1}^{w_0}$, $n \geq 2$, with the operator V defined as in Section 2, where it has been shown that for any bounded continuous real-valued function g on W the function $Ug : w \in W \to Ug(w) = \int_X g(u_x(w))\, p(\mathrm{d}x)$ is bounded and continuous, too. According to equation (2.7), for any $n \geq 2$ we have

$$\int_W g(w)\pi_n^{w_0}(\mathrm{d}w) = \int_W g(w) V\pi_{n-1}^{w_0}(\mathrm{d}w) = \int_W Ug(w)\pi_{n-1}^{w_0}(\mathrm{d}w)\,.$$

Since $\widetilde{\zeta_n^{w_0}}$ converges $\mathbb{P}_p$-a.s., letting $n \to \infty$ we get

$$\int_W g(w)\mu(\mathrm{d}w) = \int_W Ug\,(w)\,\mu\,(\mathrm{d}w)\,,$$

showing that μ is a stationary distribution for $(\zeta_n^{w_0})_{n\in\mathbf{N}_+}$. If μ' is another stationary distribution for $(\zeta_n^{w_0})_{n\in\mathbf{N}_+}$, then it is the probability distribution of $\zeta_n^{w_0}$, hence of $\widetilde{\zeta_n^{w_0}}$, for any $n \in \mathbf{N}_+$. As the latter distribution should converge to μ, we have $\mu = \mu'$. □

Remarks. **1.** No assumption on the metric space (W, d) is needed in Proposition 4.1.

2. A weak variant of Proposition 4.1, that implies it under its stronger assumptions, see Athreya and Stenflo [3], is as follows. With (W, d) and $(u_x)_{x\in X}$ unrestricted, assume that for some $w_0 \in W$ there exists a random variable $\zeta_\infty^{w_0}$ to which $\widetilde{\zeta_n^{w_0}}$ converges in distribution under $\mathbb{P}_p$ as $n \to \infty$. Then (i) $\zeta_n^{w_0}$ also converges in distribution under $\mathbb{P}_p$ as $n \to \infty$ to $\zeta_\infty^{w_0}$, and (ii) if U is a Feller operator, the probability distribution $\mu^{w_0} = \mathbb{P}_p\,(\zeta_\infty^{w_0})^{-1}$ of $\zeta_\infty^{w_0}$ under $\mathbb{P}_p$ is a stationary distribution for the Markov chain $(\zeta_n^{w_0})_{n\in\mathbf{N}_+}$ while if μ^{w_0} does not depend on w_0, it is the *unique* stationary distribution. □

We start with

Theorem 4.2. *For any $w \in W$, under the assumptions of Theorem 3.5, the backward process*

$$\widetilde{\zeta_n^w} = u_{\xi_1} \circ \cdots \circ u_{\xi_n}(w), \quad n \in \mathbf{N}_+,$$

converges $\mathbb{P}_p$-a.s. at a geometric rate as $n \to \infty$ to a W-valued random variable ζ_∞. We have $\pi = \mathbb{P}_p \zeta_\infty^{-1}$, that is, $\pi(A) = \mathbb{P}_p(\zeta_\infty \in A)$, $A \in \mathcal{B}_W$.

Proof. By the very definition of ℓ_α we have

$$\int_X \left(\frac{d\left(u_x(w'),\ u_x(w'')\right)}{d(w', w'')} \right)^\alpha p(\mathrm{d}x) \le \ell_\alpha < 1$$

for any $x \in X$ and $w' \ne w''$, $w', w'' \in W$. This amounts to

$$\mathbb{E}_p\left(d^\alpha\left(u_{\xi_{n+1}}(w'), u_{\xi_{n+1}}(w'')\right)\right) \le \ell_\alpha d^\alpha(w', w'')$$

for any $n \in \mathbf{N}$ and $w', w'' \in W$. Since $\zeta_{n+1}^w = u_{\xi_{n+1}}(\zeta_n^w)$, for any $w \in W$ and $n \in \mathbf{N}_+$, the above inequality yields

$$\mathbb{E}_p\left(d^\alpha\left(\zeta_{n+1}^{w'}, \zeta_{n+1}^{w''}\right) \mid \xi_1, \dots, \xi_n\right) \le \ell_\alpha d^\alpha\left(\zeta_n^{w'}, \zeta_n^{w''}\right) \quad \mathbb{P}_p\text{-a.s.}$$

for any $n \in \mathbf{N}$ and $w', w'' \in W$. As

$$\mathbb{E}_p\left[\mathbb{E}_p\left(d^\alpha\left(\zeta_{n+1}^{w'}, \zeta_{n+1}^{w''}\right) \mid \xi_1, \dots, \xi_n\right)\right] = \mathbb{E}_p\left(d^\alpha\left(\zeta_{n+1}^{w'}, \zeta_{n+1}^{w''}\right)\right),$$

we deduce that

$$\mathbb{E}_p\left(d^\alpha\left(\zeta_{n+1}^{w'}, \zeta_{n+1}^{w''}\right)\right) \le \ell_\alpha \mathbb{E}_p\left(d^\alpha\left(\zeta_n^{w'}, \zeta_n^{w''}\right)\right)$$

for any $n \in \mathbf{N}_+$ and $w', w'' \in W$, which implies

$$\mathbb{E}_p\left(d^\alpha\left(\zeta_n^{w'}, \zeta_n^{w''}\right)\right) \le \ell_\alpha^n d^\alpha(w', w'')$$

for any $n \in \mathbf{N}_+$ and $w', w'' \in W$. Note now that for any $n \ge 1$ the nth product measure of p with itself is symmetrical, which allows us to also write

$$\mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^{w'}}, \widetilde{\zeta_n^{w''}}\right)\right) \le \ell_\alpha^n d^\alpha(w', w'') \tag{4.2}$$

for any $n \in \mathbf{N}_+$ and $w', w'' \in W$. Finally, since $\widetilde{\zeta_{n+1}^w} = \widetilde{\zeta_n^{u_{\xi_{n+1}}(w)}}$ for any $w \in W$ and $n \in \mathbf{N}_+$, we can write

$$\begin{aligned}
\mathbb{E}_p\left(d^\alpha(\widetilde{\zeta_n^w}, \widetilde{\zeta_{n+1}^w})\right) &= \mathbb{E}_p\left[\mathbb{E}_p\left(d^\alpha(\widetilde{\zeta_n^w}, \widetilde{\zeta_n^{u_{\xi_{n+1}}(w)}}) \;\Big|\; \xi_{n+1}\right)\right] \\
&\le \ell_n^\alpha \left[\mathbb{E}_p(d^\alpha\left(w, u_{\xi_{n+1}}(w)\right) \big| \, \xi_{n+1})\right] \quad \text{(by (4.2))} \\
&= \ell_n^\alpha \int_X d^\alpha\left(w, u_x(w)\right) p(\mathrm{d}x).
\end{aligned}$$

Therefore,

$$\mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^w},\widetilde{\zeta_{n+1}^w}\right)\right) \le \ell_\alpha^n \int_X d^\alpha\left(w,u_x(w)\right)p\left(\mathrm{d}x\right) \tag{4.3}$$

for any $n \in \mathbf{N}_+$ and $w \in W$, with $\int_X d^\alpha\left(w,u_x(w)\right)p\left(\mathrm{d}x\right) < \infty$ by (3.7). Hence for any $w \in W$ the series

$$\sum_{n\in\mathbf{N}_+} \mathbb{P}_p\left(d^\alpha\left(\widetilde{\zeta_n^w},\widetilde{\zeta_{n+1}^w}\right) > \ell_\alpha^{n/2}\right)$$

is convergent since by Markov's inequality we have

$$\begin{aligned}\mathbb{P}_p\left(d^\alpha\left(\widetilde{\zeta_n^w},\widetilde{\zeta_{n+1}^w}\right) > \ell_\alpha^{n/2}\right) &\le \frac{\mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^w},\widetilde{\zeta_{n+1}^w}\right)\right)}{\ell_\alpha^{n/2}} \\ &\le \ell_\alpha^{n/2}\int_X d^\alpha\left(w,u_x(w)\right)p\left(\mathrm{d}x\right)\end{aligned}$$

for any $n \in \mathbf{N}_+$ and $w \in W$. It follows from the Borel-Cantelli lemma that the inequality $d^\alpha\left(\widetilde{\zeta_n^w},\widetilde{\zeta_{n+1}^w}\right) \le \ell_\alpha^{n/2}$ holds $\mathbb{P}_p$-a.s. for all sufficiently large n. So, for any $w \in W$, $(\widetilde{\zeta_n^w})_{n\in\mathbf{N}_+}$ is $\mathbb{P}_p$-a.s. a Cauchy sequence in (W, d^α) that converges at a geometric rate to a W-valued random variable, say, $\zeta_{\infty,w}$. Note also that it follows from (4.3) that

$$\begin{aligned}\mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^w},\zeta_{\infty,w}\right)\right) &\le \sum_{j\in\mathbf{N}}\mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^w},\widetilde{\zeta_{n+j+1}^w}\right)\right) \\ &\le \frac{\ell_\alpha^n}{1-\ell_\alpha}\int_X d^\alpha\left(w,u_x(w)\right)p\left(\mathrm{d}x\right)\end{aligned} \tag{4.4}$$

for any $n \in \mathbf{N}_+$. Let us show that, actually, $\zeta_{\infty,w}$ does not depend on $w \in W$. Indeed, fix $w \in W$. For any $w' \in W$ and $n \in \mathbf{N}_+$ by (4.2) and (4.4) we have

$$\begin{aligned}\mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^{w'}},\zeta_{\infty,w}\right)\right) &\le \mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^{w'}},\widetilde{\zeta_n^w}\right)\right) + \mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^w},\zeta_{\infty,w}\right)\right) \\ &\le \ell_\alpha^n\left(d^\alpha\left(w,w'\right) + \int_X d^\alpha\left(w,u_x(w)\right)p\left(\mathrm{d}x\right)/\left(1-\ell_\alpha\right)\right),\end{aligned}$$

whence

$$\mathbb{P}_p\left(d^\alpha\left(\widetilde{\zeta_n^{w'}},\zeta_{\infty,w}\right) \ge \ell_\alpha^{n/2}\right) \le \ell_\alpha^{n/2}\left(d^\alpha\left(w,w'\right) + \int_X d^\alpha(w,u_x(w))p(\mathrm{d}x)/(1-\ell_\alpha)\right)$$

by Markov's inequality again. We thus conclude as before that

$$\lim_{n\to\infty}\widetilde{\zeta_n^{w'}} = \zeta_{\infty,w}\ \mathbb{P}_p\text{-a.s.},$$

and we are done.

Let us from now on write ζ_∞ for the unique $\mathbb{P}_p$-a.s. limiting random variable.

To prove the equation $\pi = \mathbb{P}_p\zeta_\infty^{-1}$, we notice that for any $f \in C(W)$ we have

$$\mathbb{E}_p\left(f\left(\zeta_n^w\right)\right) = \mathbb{E}_p\left(f\left(\widetilde{\zeta_n^w}\right)\right), \quad n \in \mathbf{N}_+,\ w \in W.$$

But, on one hand, by Theorem 3.5,

$$\lim_{n\to\infty} \mathbb{E}_p\left(f\left(\zeta_n^w\right)\right) = \int_W f\mathrm{d}\pi, \quad w \in W,$$

and, on the other hand, by bounded convergence,

$$\lim_{n\to\infty} \mathbb{E}_p\left(f\left(\widetilde{\zeta_n^w}\right)\right) = \int_W f\left(\zeta_\infty\right)\mathrm{d}\mathbb{P}_p = \int_W f\mathrm{d}\left(\mathbb{P}_p\zeta_\infty^{-1}\right), \quad w \in W.$$

Therefore, $\int_W f\mathrm{d}\pi = \int_W f\mathrm{d}\left(\mathbb{P}_p\zeta_\infty^{-1}\right)$ for any $f \in C(W)$. Hence $\pi = \mathbb{P}_p\zeta_\infty^{-1}$ by a well known result (see, e.g., Parthasarathy [44, Theorem 5.9]). $\square$

Remarks. **1.** The equation $\pi = \mathbb{P}_p\zeta_\infty^{-1}$ implies that the support $\operatorname{supp}\pi$ of π, that is, the smallest closed subset of W having π-measure 1, defined as

$$\begin{aligned}\operatorname{supp}\pi &= \{w \in W : \pi\left(\{v \in W \mid d\left(v,w\right) < \varepsilon\}\right) > 0 \text{ for all } \varepsilon > 0\} \\ &= \{w \in W \mid \mathbb{P}_p\left(d\left(\zeta_\infty, w\right) < \varepsilon\right) > 0 \text{ for all } \varepsilon > 0\}\end{aligned}$$

and the range of ζ_∞, that is, the set $\zeta_\infty\left(\Omega'\right)$, where Ω' is the random event $\{\omega \in \Omega : \zeta_\infty\left(\omega\right) \text{ exists}\}$, with $\mathbb{P}_p\left(\Omega'\right) = 1$, are strongly related. Indeed, we clearly have

$$1 = \mathbb{P}_p\left(\Omega'\right) = \mathbb{P}_p\left(\zeta_\infty^{-1}\left(\zeta_\infty\left(\Omega'\right)\right)\right) = \pi\left(\zeta_\infty\left(\Omega'\right)\right)$$

and

$$1 = \pi\left(\operatorname{supp}\pi\right) = \mathbb{P}_p\left(\zeta_\infty \in \operatorname{supp}\pi\right),$$

whence

$$\pi\left(\zeta_\infty\left(\Omega'\right)\Delta \operatorname{supp}\pi\right) = 0$$

and

$$\mathbb{P}_p\left(\Omega'\ \Delta\left(\zeta_\infty \in \operatorname{supp}\pi\right)\right) = 0,$$

where Δ stands for symmetric difference of sets.

2. A special case where $\operatorname{supp}\pi$ can be precisely described is further considered in Theorem 5.1. $\square$

Theorem 4.2 can be slightly generalized by considering the case of a random initial point $w_0 \in W$ with $\mathbb{P}_{\lambda,p}(w_0 \in A) = \lambda(A),\ A \in \mathcal{B}_W$, for any given $\lambda \in \operatorname{pr}(\mathcal{B}_W)$. We namely have

THEOREM 4.3. *Assume that* (3.6) *and* (3.7) *hold. Then the backward process* $(\widetilde{\zeta_n^{w_0}})_{n\in\mathbf{N}_+}$ *converges* $\mathbb{P}_{\lambda,p}$*-a.s. at a geometric rate as* $n \to \infty$ *to* *a W-valued random variable ζ_∞. We have $\pi = \mathbb{P}_{\lambda,p}\,\zeta_\infty^{-1}$, that is, $\pi(A) = \mathbb{P}_{\lambda,p}(\zeta_\infty \in A)$, $A \in \mathcal{B}_W$.*

Clearly, Remark 1 after Theorem 4.2 also has a version in which $w \in W$ is replaced by $\lambda \in \mathrm{pr}(\mathcal{B}_W)$.

We now consider assumptions more general than those we assumed before. These new assumptions involve a kind of local contractibility at one point only instead of (3.6).

We thus introduce Condition **C** below. Cf. Wu and Shao [49] and Herkenrath and Iosifescu [22].

Condition C. (i) There exist $w_0 \in W$, $\alpha \in (0,1]$, $r \in (0,1)$ and $\mathcal{W}$-measurable functions $\varphi_n : W \to [0,\infty)$, $n \in \mathbf{N}_+$, such that both

$$R{:=}\frac{1}{\limsup\limits_{n\to\infty} \delta_n^{1/n}} \quad \text{and} \quad R_w := \frac{1}{\limsup\limits_{n\to\infty} \varphi_n^{1/n}(w)}$$

strictly exceed r for any $w \in W$, where

$$\delta_n := \mathbb{E}_p\left(\varphi_n\left(\zeta_1^{w_0}\right) d^\alpha\left(w_0, \zeta_1^{w_0}\right)\right), \quad n \in \mathbf{N}_+.$$

(ii) For any $w \in W$ and $n \in \mathbf{N}_+$ one has

$$\mathbb{E}_p\left(d^\alpha\left(\zeta_n^w, \zeta_n^{w_0}\right)\right) \le r^n \varphi_n(w) d^\alpha(w, w_0).$$

It is clear that R and R_w, $w \in W$, are the convergence radii of the power series

$$\sum_{n\in\mathbf{N}_+} \delta_n t^n \quad \text{and} \quad \sum_{n\in\mathbf{N}_+} \varphi_n(w) t^n, \quad t \in \mathbb{R},$$

respectively. Also, (i) and (ii) generalize (3.6) and (3.7), respectively, where $\varphi_n \equiv 1$, $n \in \mathbf{N}_+$.

Theorem 4.4. *Assume Condition* **C** *holds.*

(j) *There exists a W-valued $\sigma(\xi_1, \xi_2, \dots)$-measurable random variable ζ_∞ to which $\widetilde{\zeta_n^w}$ converges $\mathbb{P}_p$-a.s. as $n \to \infty$ for any $w \in W$ and, moreover,*

$$\mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^w}, \zeta_\infty\right)\right) \le r^n \varphi_n(w) d^\alpha(w, w_0) + \sum_{m\ge n} r^m \delta_m, \quad n \in \mathbf{N}_+.$$

(jj) *Let π denote the probability distribution of ζ_∞ and let w_0', w_0'' be W-valued random variables independent of $(\xi_n)_{n\in\mathbf{N}_+}$ and with common distribution π. Then*

$$\mathbb{E}_{\pi,p}\left(d^\alpha\left(\zeta_n^{w_0'}, \zeta_n^{w_0''}\right)\right) \le 2\sum_{m\ge n} r^m \delta_m, \quad n \in \mathbf{N}_+.$$

Proof. Note first that $\widetilde{\zeta_{n+1}^{w_0}} = \zeta_n^{\widetilde{u_{\xi_{n+1}}(w_0)}}$, $n \in \mathbf{N}_+$. By Condition $\mathbf{C}$(ii) we then have

$$\begin{aligned}\mathbb{E}_p\left(d^\alpha\big(\widetilde{\zeta_n^{w_0}}, \widetilde{\zeta_{n+1}^{w_0}}\big)\right) &= \mathbb{E}_p\left[\mathbb{E}_p\Big(d^\alpha\big(\widetilde{\zeta_n^{w_0}}, \zeta_n^{\widetilde{u_{\xi_{n+1}(w_0)}}}\big) \;\Big|\; \xi_{n+1}\Big)\right]\\ &\le r^n \mathbb{E}_p\left(\varphi_n\left(u_{\xi_{n+1}}(w_0)\right) d^\alpha\left(w_0, u_{\xi_{n+1}}(w_0)\right)\right)\\ &= r^n \mathbb{E}_p\left(\varphi_n\left(u_{\xi_1}(w_0)\right) d^\alpha\left(w_0, u_{\xi_1}(w_0)\right)\right)\\ &= r^n \mathbb{E}_p\left(\varphi_n\left(\zeta_1^{w_0}\right) d^\alpha\left(w_0, \zeta_1^{w_0}\right)\right) = r^n \delta_n\end{aligned}$$

for any $n \in \mathbf{N}_+$. Fix $r_0 \in (0,1)$ with $\frac{r}{R} < r_0 < 1$. Then, by Markov's inequality,

$$\mathbb{P}_p\left(d^\alpha\left(\widetilde{\zeta_n^{w_0}}, \widetilde{\zeta_{n+1}^{w_0}}\right) \ge r_0^n\right) \le \left(\frac{r}{r_0}\right)^n \delta_n, \quad n \in \mathbf{N}_+.$$

As by Condition $\mathbf{C}$(i) the series with general term $(r/r_0)^n \delta_n$ is convergent, the Borel-Cantelli lemma implies that

$$\mathbb{P}_p\left(d^\alpha\big(\widetilde{\zeta_n^{w_0}}, \widetilde{\zeta_{n+1}^{w_0}}\big) \ge r_0^n \text{ infinitely often}\right) = 0,$$

hence $\widetilde{\zeta_n^{w_0}} \to \zeta_\infty$, say, $\mathbb{P}_p$-a.s. as $n \to \infty$ by the completeness of W and, obviously, ζ_∞ is $\sigma(\xi_1, \xi_2, \dots)$-measurable. Next, by the triangle inequality,

$$\mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^{w_0}}, \zeta_\infty\right)\right) \le \mathbb{E}_p\left(\sum_{j\in\mathbf{N}} d^\alpha\left(\widetilde{\zeta_{n+j}^{w_0}}, \widetilde{\zeta_{n+j+1}^{w_0}}\right)\right) \tag{4.5}$$

$$\sum_{j\in\mathbf{N}} \mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_{n+j}^{w_0}}, \widetilde{\zeta_{n+j+1}^{w_0}}\right)\right) \le \delta_n' := \sum_{m\ge n} r^m \delta_m$$

for any $n \in \mathbf{N}_+$. Finally, by Condition $\mathbf{C}$(ii),

$$\begin{aligned}\mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^{w}}, \zeta_\infty\right)\right) &\le \mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^{w}}, \widetilde{\zeta_n^{w_0}}\right)\right) + \mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^{w_0}}, \zeta_\infty\right)\right)\\ &\le r^n \varphi_n(w) d^\alpha(w, w_0) + \delta_n'\end{aligned}$$

for any $w \in W$, whence we conclude that $\widetilde{\zeta_n^{w}} \to \zeta_\infty$ $\mathbb{P}_p$-a.s. as $n \to \infty$, by invoking the argument already used to show that $\widetilde{\zeta_n^{w_0}} \to \zeta_\infty$ $\mathbb{P}_p$-a.s. as $n \to \infty$. We should now choose $r_0 \in (0,1)$ with $r/\min(R, R_w) < r_0 < 1$ and note that both series $\sum_{n\in\mathbf{N}_+} \varphi_n(w)(r/r_0)^n \delta_n$ and $\sum_{n\in\mathbf{N}_+}\sum_{m\ge n} (r/r_0)^m \delta_m = \sum_{n\in\mathbf{N}_+} n(r/r_0)^n \delta_n$ are convergent by Condition $\mathbf{C}$(i). The proof of (j) is thus complete.

(jj) For any $n, m \in \mathbf{N}_+$ and $w \in W$ consider the random variable

$$\widetilde{\zeta_{n,m}^{w}} = u_{\xi_n} \circ \cdots \circ u_{\xi_{n+m-1}}(w),$$

so that $\widetilde{\zeta_n^w} = \widetilde{\zeta_{1,n}^w}$. It follows from (j) that the limit

$$\widetilde{\zeta_{n,\infty}} := \nu_n = \lim_{m\to\infty} \widetilde{\zeta_{n,m}^w}, \quad n \in \mathbf{N}_+,$$

exists $\mathbb{P}_p$-a.s., does not depend on w, and has $\mathbb{P}_p$-distribution π. Clearly, $\widetilde{\zeta_n^{\nu_{n+1}}} = \zeta_\infty$ $\mathbb{P}_p$-a.s. for any $n \in \mathbf{N}_+$. Also, ν_{n+1} is independent of $(\xi_i)_{1\le i\le n}$ for any $n \in \mathbf{N}_+$. Then the triangle inequality and the above facts allow us to write

$$\mathbb{E}_{\pi,p}\left(d^\alpha\left(\zeta_n^{w_0'}, \zeta_n^{w_0''}\right)\right) \le \mathbb{E}_{\pi,p}\left(d^\alpha\left(\zeta_n^{w_0'}, \zeta_n^{w_0}\right)\right) + \mathbb{E}_{\pi,p}\left(d^\alpha\left(\zeta_n^{w_0}, \zeta_n^{w_0''}\right)\right)$$

$$= 2\mathbb{E}_{\pi,p}\left(d^\alpha\left(\widetilde{\zeta_n^{w_0'}}, \widetilde{\zeta_n^{w_0}}\right)\right) = 2\mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^{\nu_{n+1}}}, \widetilde{\zeta_n^{w_0}}\right)\right) = 2\mathbb{E}_p\left(d^\alpha\left(\zeta_\infty, \widetilde{\zeta_n^{w_0}}\right)\right)$$

for $n \in \mathbf{N}_+$, and the claim follows from (4.5). Note that π occurring in $\mathbb{E}_{\pi,p}$ in the relations above refers to w_0' and w_0'' while w_0 is the (fixed) point from Condition **C**. $\square$

Let us remind that by the Letac lemma, the distribution π of ζ_∞ under $\mathbb{P}_p$ is the *unique* stationary distribution of the Markov chain $(\zeta_n)_{n\in\mathbf{N}_+}$ if for p-almost all $x \in X$ the mapping $u_x : W \to W$ is continuous, an assumption we agreed about, see Section 2. Without such an assumption, the uniqueness of the stationary distribution cannot be asserted.

Theorem 4.4 allows one to estimate the rate of convergence of the n-step transition probability function Q^n of the Markov chain $(\zeta_n)_{n\in\mathbf{N}_+}$ to its stationary probability distribution π (cf. Theorem 3.5). We namely have

COROLLARY 4.5. *Assume Condition* **C** *holds. Then for any $w \in W$ and $n \in \mathbf{N}_+$ we have*

$$\rho_{L,\alpha}\left(Q^n\left(w,\cdot\right),\pi\right) \le r^n\varphi_n(w)d^\alpha\left(w,w_0\right) + \sum_{m\ge n} r^m\delta_m.$$

Proof. By the very definition of $\rho_{L,\alpha}$,

$$\rho_{L,\alpha}\left(Q^n\left(w,\cdot\right),\pi\right) = \sup\left\{\int_W f(v)\left(Q^n\left(w,\mathrm{d}v\right) - \pi\left(\mathrm{d}v\right)\right) \;\middle|\; 0 \le f \le 1,\right.$$
$$\left.\left|f\left(v'\right) - f\left(v''\right)\right| \le d^\alpha\left(v',v''\right),\ v',v'' \in W\right\}.$$

Since for any $f \in C(W)$ as in the definition above we have

$$\int_W f(v)\left(Q^n\left(w,\mathrm{d}v\right) - \pi\left(\mathrm{d}v\right)\right) = \mathbb{E}_p\left(f\left(\zeta_n^w\right) - f\left(\zeta_\infty\right)\right) = \mathbb{E}_p\left(f\left(\widetilde{\zeta_n^w}\right) - f\left(\zeta_\infty\right)\right)$$

and

$$\left|\mathbb{E}_p\left(f\left(\widetilde{\zeta_n^w}\right) - f\left(\zeta_\infty\right)\right)\right| \le \mathbb{E}_p\left(d^\alpha\left(\widetilde{\zeta_n^w},\zeta_\infty\right)\right),$$

the proof is complete by Theorem 4.4 (j). $\square$

Remark . It is interesting to compare the upper bound of $\rho_{L,\alpha}\big(Q^n(w,\cdot),\pi\big)$ in Theorem 3.5 and Corollary 4.5, under the assumptions of the former, when $\varphi_n \equiv 1$, $n \in \mathbf{N}_+$, and the part of r in Condition $\mathbf{C}$(ii) is played by ℓ. Then the two bounds are

$$\frac{\ell^n}{1-\ell}\int_X d^\alpha\left(w, u_x(w)\right) p\left(\mathrm{d}x\right)$$

and

$$\ell^n d^\alpha\left(w, w_0\right) + \frac{\ell^n \displaystyle\int_X d^\alpha\left(w_0, u_x\left(w_0\right)\right) p\left(\mathrm{d}x\right)}{1-\ell},$$

respectively, for any $w \in W$ and $n \in \mathbf{N}_+$. As

$$\int_X d^\alpha\left(w, u_x(w)\right) p\left(\mathrm{d}x\right) \le \left(\ell+1\right) d^\alpha\left(w, w_0\right) + \int_X d^\alpha\left(w_0, u_x\left(w_0\right)\right) p\left(\mathrm{d}x\right)$$

by (2.6), the second bound always exceeds the first one for any $w \in W$ and $n \in \mathbf{N}_+$ even if both of them are $O\left(\ell^n\right)$ as $n \to \infty$.

Theorem 4.4 allows to define W-valued random variables ζ_i, $i \le 0$, in such a way that the equation $\zeta_i = u_{\xi_i}\left(\zeta_{i-1}\right)$ that holds for $i \in \mathbf{N}_+$ still holds for $i \le 0$. Let $(\xi_i)_{i\in\mathbf{Z}}$ be a doubly infinite sequence of i.i.d. X-valued random variables with common distribution p. Similarly to the context in Theorem 4.4(j), the limit

$$\lim_{m\to\infty} u_{\xi_i} \circ u_{\xi_{i-1}} \circ \cdots \circ \xi_{i-m}(w) := \zeta_i^w, \quad w \in W,$$

exists $\mathbb{P}_p$-a.s. for any $i \le 0$, is a $\sigma\left(\ldots, \xi_{i-1}, \xi_i\right)$-measurable function that does not actually depend on w, has probability distribution π, and satisfy $\mathbb{P}_p$-a.s. the equation $\zeta_i = u_{\xi_i}\left(\zeta_{i-1}\right)$ for any $i \le 0$ if, as noted before in a different context, for p-almost all $x \in X$ the mapping $u_x : W \to W$ is continuous. Notice that with ζ_i, $i \le 0$, so defined, the doubly infinite sequence $(\zeta_i)_{i\in\mathbf{Z}}$ constructed by using the latter recurrence equation for *any* $i \in \mathbf{Z} := \{\ldots, -1, 0, 1, \ldots\}$ is a strictly stationary process, as well as a Markov process, on a suitable probability space, to be still denoted $\left(\Omega, \mathcal{K}, \mathbb{P}_{\pi,p}\right)$.

The idea of extending a strictly stationary process into the past allowing to consider it to have been going on forever, of which we have just seen an instance, is an old one. The possibility of always extending a real-valued strictly stationary process into the past is proved in Doob [14, pp. 456–458]. See also Elton [17]. Recall that a W-valued process $\mathbf{X} = (X_n)_{n\in\mathbf{N}}$ or $\mathbf{X} = (X_n)_{n\in\mathbf{Z}}$ on a probability space $(\Omega, \mathcal{K}, P)$ is said to be *strictly stationary* (*ergodic*) iff the left shift on $W^{\mathbf{N}}$ or $W^{\mathbf{Z}}$ is measure-preserving (ergodic) for the measure $P \circ \mathbf{X}^{-1}$.

Proposition 4.6 (Doob's lemma). *Assume W is a separable complete metric space. Given a W-valued strictly stationary process $\mathbf{X} = (X_n)_{n\in\mathbf{N}}$,* *there exists a strictly stationary process* $\widetilde{\mathbf{X}} = (\widetilde{X}_n)_{n\in\mathbf{Z}}$ *whose finite-dimensional distributions are identical with those of* $\mathbf{X}$. *Also,* $\widetilde{\mathbf{X}}$ *is ergodic iff* $\mathbf{X}$ *is ergodic.*

5. THE SUPPORT OF THE STATIONARY DISTRIBUTION

We have already noticed that there is a purely deterministic concept of an iterated function system. This is intensively investigated by people working in geometry of sets and measures (see, e.g., Mauldin and Urbański [41]).

Let (W, d) be a complete metric space, I a countable set with at least two elements, and $(u_i)_{i\in I}$ a collection of contracting self-mappings of W, that is, such that

$$\mathrm{s}\,(u_i) := s_i < 1, \quad i \in I. \tag{5.1}$$

For any $n \in \mathbf{N}_+$ and $i^{(n)} := (i_1, \ldots, i_n) \in I^n$ put $v_{i^{(n)}} = u_{i_1} \circ \cdots \circ u_{i_n}$ and note that $v_i \equiv u_i,\ i \in I$. Define a *scaling* operator $\mathcal{S} : \mathcal{P}\,(W) \to \mathcal{P}(W)$ by

$$\mathcal{S}(E) = \bigcup_{i\in I} u_i(E), \quad E \subset W,$$

so that

$$\mathcal{S}^n(E) = \bigcup_{i^{(n)}\in I^n} v_{i^{(n)}}(E), \quad n \in \mathbf{N}_+,\ E \subset W.$$

A set $E \subset W$ is said to be *subinvariant* (*invariant*) under scaling iff $\mathcal{S}(E) \subset E\ (\mathcal{S}(E) = E)$. The main concern is the limit set associated with $(u_i)_{i\in I}$, which is an invariant set under scaling, which we are going to define.

Assume first that I is finite, hence $s := \max_{i\in I} s_i < 1$. Then there exists a unique compact subset K of W that is invariant under scaling. To show this, consider (see A2.2) the collection $\mathrm{bcl}\,W$ of non-empty bounded closed subsets of W, which under the Hausdorff metric d_H is a complete metric space. Consider also the collection $\mathrm{c}W \subset \mathrm{bcl}\,W$ of the compact subsets of W. Note that $\mathrm{c}W$ is a closed subset of $\mathrm{bcl}\,W$. Clearly, $\mathcal{S}$ takes $\mathrm{bcl}\,W$ and $\mathrm{c}W$ into themselves. For any $A, B \in \mathrm{bcl}\,W$ (or $\mathrm{c}W$) we have

$$d_H\left(\mathcal{S}(A), \mathcal{S}(B)\right) = d_H\left(\bigcup_{i\in I} u_i(A), \bigcup_{i\in I} u_i(B)\right)$$
$$\le \max_{i\in I} d_H\left(u_i(A), u_i(B)\right) \le \left(\max_{i\in I} s_i\right) d_H(A,B) = s d_H(A,B).$$

That is, $\mathcal{S}$ is a contraction map on both $\mathrm{bcl}\,W$ and $\mathrm{c}W$ in the Hausdorff metric d_H. As $(\mathrm{bcl}\,W, d_H)$ is a complete metric space, the existence of some $K \in \mathrm{bcl}\,W$ satisfying $K = \mathcal{S}(K)$ follows from the contraction mapping principle. Actually, we have $\lim_{n\to\infty} \mathcal{S}^n(A) = K$ in $(\mathrm{bcl}\,W, d_H)$ for any $A \in \mathrm{bcl}\,W$, hence any $A \in \mathrm{c}W$.

As cW is closed in $(\operatorname{bcl} W, d_H)$ and $\mathcal{S}$ takes cW into itself, we should have $K \in \mathrm{c}W$, as asserted.

Remark. It follows in particular that $v_{i^{(\infty)}} := \lim_{n\to\infty} v_{i^{(n)}}(w)$ exists and is an element of K for any $w \in W$ and $i^{(\infty)} = (i_1, i_2, \dots) \in I^{\mathbf{N}_+}$. Clearly, $v_{i^{(\infty)}}$ does not depend on w as $d\left(v_{i^{(n)}}(w'), v_{i^{(n)}}(w'')\right) \le \mathrm{s}\left(v_{i^{(n)}}\right) d(w', w'') \le s^n d(w', w'') \to 0$ as $n \to \infty$ for any $w', w'' \in W$. It is for this reason that K is also called the *attractor* of $(u_i)_{i\in I}$. $\square$

Theorem 5.1 (Hutchinson [27]). *The set K is the intersection of the sets $\mathcal{S}^n(W)$, $n \in \mathbf{N}_+$, that is,*

$$K = \bigcap_{n\in\mathbf{N}_+} \bigcup_{i^{(n)}\in I^n} v_{i^{(n)}}(W)$$

(*even if, possibly,* $W \notin \operatorname{bcl} W$). *It also is the closure of the set*

$$\bigcup_{n\in\mathbf{N}_+} \bigcup_{i^{(n)}\in I^n} \left(\theta_{i^{(n)}} : v_{i^{(n)}}\left(\theta_{i^{(n)}}\right)\right) = \theta_{i^{(n)}}$$

of fixed points of all $v_{i^{(n)}}$, $i^{(n)} \in I^n$, $n \in \mathbf{N}_+$.

For any $(i_1, i_2, \dots) \in I^{\mathbf{N}_+}$ *the limit* $\lim_{n\to\infty} \theta_{i^{(n)}} := k_{i^{(\infty)}}$ *exists and clearly belongs to* K *while the mapping* $\theta : I^{\mathbf{N}_+} \to K$ *defined by* $\theta\left(i^{(\infty)}\right) = k_{i^{(\infty)}}$, $i^{(\infty)} = (i_1, i_2, \dots) \in I^{\mathbf{N}_+}$, *is continuous onto* K.

The compact set K supports probability measures in a natural way. Let $p_i \notin 0, 1$, $i \in I$, with $\sum_{i\in I} p_i = 1$. It then follows from Theorem 5.1 and Remark 1 following Theorem 4.2 that in the case where X is the finite set I and $\mathrm{s}(u_i) < 1$ for any $i \in I$, the support of the invariant probability measure π of the IFS $((u_i)_{i\in I}, (p)_{i\in I})$ is precisely K. It should be stressed that K is only and uniquely determined by the mappings u_i, $i \in I$. Assigning them probabilities yields just a probability distribution over K while two different assignments $p' = (p'_i)_{i\in I}$ and $p'' = (p''_i)_{i\in I}$ of probabilities yield in general different probability distributions over the same K.

Remark. Hata [21] proved the existence of a unique nonempty compact set $K \subset W$ which is invariant under scaling ($K = \mathcal{S}(K)$) by only assuming that the u_i, $i \in I$, are weakly contractive. This means that

$$\alpha_i(t) := \sup_{d(y,z)\le t} d\left(u_i(y), u_i(z)\right) < t, \quad i \in I,$$

for all $t > 0$. Next, for any given $i^{(\infty)} = (i_1, \ldots, i_n, i_{n+1}, \cdots) = \left(i^{(n)}, i_{n+1}, \cdots\right)$, the Lebesgue measure of the set $v_{i^{(n)}}(K)$ converges to 0 as $n \to \infty$. Also,

$$K = \bigcap_{n \in \mathbf{N}_+} \bigcup_{i^{(n)} \in I^n} v_{i^{(n)}}(K)$$

(compare with Theorem 5.1). Assuming that W is a compact subset of $\mathbf{R}^m$ for some $m \in \mathbf{N}_+$ with d the Euclidean distance in $\mathbf{R}^m$, Lau and Ye [37] showed that these properties still hold if the u_i, $i \in I$, are continuous and at least one of them is weakly contractive. Here, unicity means that of a smallest such nonempty compact K. $\square$

When I is infinite, it is usual to assume that the metric space (W, d) is compact and, instead of (5.1), that

$$\lim_{n \to \infty} \sup_{i^{(n)} \in I^n} \operatorname{diam}\left(v_{i^{(n)}}(W)\right) = 0, \tag{5.2}$$

a condition which clearly holds if $\sup_{i \in I} s_i < 1$, when

$$\operatorname{diam}\left(v_{i^{(n)}}(W)\right) \leq \left(\sup_{i \in I} s_i\right)^n \operatorname{diam}(W), \quad i^{(n)} \in I^n,\ n \in \mathbf{N}_+.$$

Since for any given $i^{(\infty)} = (i_k)_{k \in \mathbf{N}_+} \in I^{\mathbf{N}_+}$ the closed sets $v_{i^{(n)}}(W)$, $n \in \mathbf{N}_+$, are decreasing and by (5.2) their diameters converge to zero uniformly with respect to $i^{(n)} \in I^n$ as $n \to \infty$, the set

$$\Pi\left(i^{(\infty)}\right) = \bigcap_{n \in \mathbf{N}_+} v_{i^{(n)}}(W)$$

is a singleton and the equation above defines a map $\Pi : I^{\mathbf{N}_+} \to W$ which is continuous when $I^{\mathbf{N}_+}$ is given the product topology induced from the discrete topology on each factor (thus making it a compact set if I is finite). Then the limit set associated with $(u_i)_{i \in I}$ is defined as

$$J = \bigcup_{i^{(\infty)} \in I^{\mathbf{N}_+}} \bigcap_{n \in \mathbf{N}_+} v_{i^{(n)}}(W). \tag{5.3}$$

It is also called the *fractal set* determined by $(u_i)_{i \in I}$. As clearly

$$u_i \circ \Pi\left((i_k)_{k \in \mathbf{N}_+}\right) = \Pi\left(i, i_1, i_2, \ldots\right)$$

for any $i \in I$ and $(i_k)_{k \in \mathbf{N}_+} \in I^{\mathbf{N}_+}$, we see that $J = \bigcup_{i \in I} u_i(J)$, that is, *$J$ is invariant under scaling.*

Now, assume that any element of W belongs to at most finitely many $u_i(W)$, $i \in I$, which clearly implies that, whatever $n \in \mathbf{N}_+$, any element of W

belongs to at most finitely many $v_{i^{(n)}}(W),\ i^{(n)} \in I^n$. It is immediate that we can then write

$$J = \bigcap_{n\in\mathbf{N}_+} \bigcup_{i^{(n)}\in I^n} v_{i^{(n)}}(W), \tag{5.4}$$

a conclusion which has also been reached in the case where I is finite. The difference is that if I is finite, then J is compact and coincides with K as described in Theorem 5.1, when $\max_{i\in I} s_i < 1$. In the present instance, for an infinite I, all what can be asserted about J given by (5.4) is that it is a $F_{\sigma\delta}$ set (that is, an intersection of countable unions of closed sets). Finally, when the assumption leading to (5.4) does not hold, the limit set (5.3) may have a much more complicated descriptive set-theoretic structure. See, e.g., Mauldin and Urbański [40, Section 5].

Example 5.2. Let $W = [0, 1]$ with the Euclidean distance, $I = \{0, 1\}$, and $u_0(w) = aw,\ u_1(w) = aw + 1 - a,\ w \in W$, with $0 < a < 1$. In this case the composition $u_{i_1} \circ \cdots \circ u_{i_n} = v_{i^{(n)}}$ can be expressed in closed terms as

$$v_{i^{(n)}}(w) = \left(a^{-1} - 1\right) \sum_{k=1}^{n} i_k a^k + a^n w \tag{5.5}$$

for any $i_1, \ldots, i_n \in I,\ n \in \mathbf{N}_+$, and $w \in W$. This can be easily seen by induction: (5.5) clearly holds for $n = 1$ and, assuming that

$$u_{i_2} \circ \cdots \circ u_{i_n}(w) = \left(a^{-1} - 1\right) \sum_{k=2}^{n} i_k a^{k-1} + a^{n-1} w$$

for $n \geq 2$, we have

$$\begin{aligned}
v_{i^{(n)}}(w) &= u_{i_1} \circ \cdots \circ u_{i_n}(w) = u_{i_1}\left(\left(a^{-1} - 1\right) \sum_{k=2}^{n} i_k a^{k-1} + a^{n-1} w\right) \\
&= \begin{cases} a\left(a^{-1} - 1\right) \sum\limits_{k=2}^{n} i_k a^{k-1} + a^n w & \text{if } i_1 = 0 \\ a\left(a^{-1} - 1\right) \sum\limits_{k=2}^{n} i_k a^{k-1} + a^n w + 1 - a & \text{if } i_1 = 1 \end{cases} \\
&= \left(a^{-1} - 1\right) \sum_{k=1}^{n} i_k a^k + a^n w.
\end{aligned}$$

Hence, for any $i^{(\infty)} = (i_1, i_2, \ldots) \in I^{\mathbf{N}_+}$, the so-called *a-expansion*, namely,

$$\lim_{n\to\infty} \left(a^{-1} - 1\right) \sum_{k=1}^{n} i_k a^k = \left(a^{-1} - 1\right) \sum_{k\in\mathbf{N}_+} i_k a^k$$

exists and is an element of $K = K_a$. Conversely, as we have seen, any element of K_a is such an a-expansion. Concerning K_a, we shall distinguish two cases: $a < 1/2$ and $a \geq 1/2$.

Case $a < 1/2$. We have $\mathcal{S}(W) = u_0(W) \cup u_1(W) = [0, a] \cup [1-a, 1]$, so that $W \setminus \mathcal{S}(W) = (a, 1-a)$ is an interval of length $1 - 2a$. In general, $W \setminus \mathcal{S}^n(W)$ is a union of $2^n - 1$ pairwise disjoint intervals consisting of 2^{k-1} intervals, each of them of length $(1-2a)(a)^{k-1}$, for all $k = 1, \ldots, n$. So, the (Lebesgue) measure of $W \setminus \mathcal{S}^n(W)$ is equal to $(1-2a) \sum_{k=1}^{n} (2a)^{k-1}$, which converges to 1 as $n \to \infty$. Hence K_a has Lebesgue measure 0, so its interior is empty for any $a < 1/2$. For example, $K_{1/3}$ is the Cantor set.

Case $a \geq 1/2$. We have $\mathcal{S}(W) = u_0(W) \cup u_1(W) = [0, a] \cup [1-a, 1] = W$, as we have a proper overlap $u_0(W) \cap u_1(W) = [1-a, a]$. Hence $\mathcal{S}^n(W) = W$ for any $n \in \mathbf{N}_+$ and so, $K_a = W = [0, 1]$ for any $a \geq 1/2$. In particular, a-expansion is just the usual binary expansion for $a = 1/2$.

In the present case, a lot of work has been done (cf. Sidorov [47] and the references therein) about the possibility that a typical $w \in (0, 1)$ have infinitely many distinct a-expansions. Set

$$\mathcal{R}_a(w) = \left\{ i^{(\infty)} = (i_1, i_2, \ldots) \in I^{\mathbf{N}_+} : w = \left(a^{-1} - 1\right) \sum_{k \in \mathbf{N}_+} i_k a^k \right\}, \ w \in (0, 1),$$

and

$$\mathcal{U}_a = \Bigg\{ w \in (0, 1) : \text{there is only one } i^{(\infty)} \in I^{\mathbf{N}_+} \text{ such that } w = \left(a^{-1} - 1\right) \sum_{k \in \mathbf{N}_+} i_k a^k \Bigg\}.$$

Concerning $\mathcal{R}_a(w)$ it is known that

– if $1 > a > g = \left(\sqrt{5} - 1\right)/2 = 0.618\ldots$ then $\mathcal{R}_a(w)$ has the cardinality of the continuum for any $w \in (0, 1)$;

– if $a \in (1/2, g]$ then $\mathcal{R}_a(w)$ has the cardinality of the continuum for (Lebesgue)-a.e. $w \in (0, 1)$.

As for $\mathcal{U}_a$, it is known that it is

– a set of positive Hausdorff dimension for $a \in (1/2, a_*)$;

– an uncountable set of zero Hausdorff dimension for $a = a_*$;

– a countable set for $a \in (a_*, g)$;

– void for $a > g$.

Here, $a_* = 0.559525\ldots$ is the Komornik-Loreti constant (see Komornik and Loreti [34]), the unique positive solution of the equation $1 = \sum_{k \in \mathbf{N}_+} \delta_k a^k$, where the sequence $(\delta_1, \delta_2, \ldots) \in I^{\mathbf{N}_+}$ is defined recursively by setting $\delta_1 = 1$, and if $\delta_1, \ldots, \delta_{2^n}$, $n \in \mathbf{N}$, are already defined, then $\delta_{2^n+k} = 1-\delta_k$, $1 \le k < 2^n$, and $\delta_{2^{n+1}} = 1$.

Now, let us now discuss the probability framework where the $i^{(\infty)} = (i_1, i_2, \ldots) \in I^{\mathbf{N}_+}$ appear as the trajectories of a sequence $(\xi_n)_{n\in\mathbf{N}_+}$ of i.i.d. I-valued random variables with $P(\xi_n = 0) = 1 - P(\xi_n = 1) = p$, $0 < p < 1$, $n \in \mathbf{N}_+$. For this, take $\Omega = I^{\mathbf{N}_+}$ and define $P = \mathbb{P}_{(p,1-p)} = \mathbb{P}_p$ as the product measure on Ω with equal multipliers $(p, 1-p)$. Then $\xi_n(\omega) = \mathrm{pr}_n\omega$, $n \in \mathbf{N}_+$, $\omega = (i_1, i_2, \ldots) \in \Omega$, $i_k \in I$, $k \in \mathbf{N}_+$. With the notation from previous sections, we are in the case where $u_x(w) = aw + (1-a)x$, $x \in I$, $w \in [0,1]$ and, by (5.5),

$$\widetilde{\zeta_n^w} = u_{\xi_1} \circ \cdots \circ u_{\xi_n}(w) = \left(a^{-1} - 1\right) \sum_{j=1}^{n} \xi_j a^j + a^n w$$

and

$$\zeta_n^w = u_{\xi_n} \circ \cdots \circ u_{\xi_1}(w) = \left(a^{-1} - 1\right) \sum_{j=1}^{n} \xi_j a^{n+1-j} + a^n w$$

for any $n \in \mathbf{N}_+$ and $w \in [0,1]$. Hence

$$\zeta_\infty := \lim_{n\to\infty} \widetilde{\zeta_n^w} = \left(a^{-1} - 1\right) \sum_{j\in\mathbf{N}_+} \xi_j a^j$$

while ζ_n^w converges in distribution to ζ_∞ as $n \to \infty$ for any $w \in W$. So, we retrive in this very special case the conclusions we have reached in the general case. Clearly, according to the general theory above, the support of the probability distribution $\mathbb{P}_p\zeta_\infty^{-1}$ of ζ_∞ is K_a we have just described, that is, either a perfect set of Lebesgue measure 0 contained in the unit interval or the whole unit interval, according as $a < 1/2$ or $a \ge 1/2$. By the above, $\mathbb{P}_p\zeta_\infty^{-1}$ is singular with respect to Lebesgue measure Λ on $[0,1]$ for any $a < 1/2$ and any $0 < p < 1$. Also, for $a = 1/2$ we clearly have $\mathbb{P}_{1/2}\zeta_\infty^{-1} = \Lambda$. According to Remark 2 following Theorem 3.3, the probability distribution $\mathbb{P}_p\zeta_\infty^{-1}$ should be of pure type, that is, either singular or absolutely continuous (with respect to Λ), for any $1/2 < a < 1$ and $0 < p < 1$. It seems natural to conjecture that in the latter case $\mathbb{P}_p\zeta_\infty^{-1}$ is absolutely continuous, but this does not hold. Erdös [18, 19] showed that $\mathbb{P}_{1/2}\zeta_\infty^{-1}$ is singular for $a > 1/2$ when a^{-1} is a PV-number, that is, an algebraic integer whose Galois conjugates are strictly less than 1 in modulus, and the golden ratio $a = \left(\sqrt{5} - 1\right)/2$ is an example, the only known to date. Since 1940s this problem, known as that of Bernoulli convolutions, has attracted considerable attention, but with no end in sight. A real breakthrough was made by Solomyak [48] who proved that $\mathbb{P}_{1/2}\zeta_\infty^{-1}$ is absolutely continuous for Λ-almost every a in the interval $(1/2, 1)$. See Peres, Schlag and Solomyak [45] for a recent review on Bernoulli convolutions. $\square$

Example 5.3. This is somewhat a multidimensional analogue of Example 5.2. Let now $W = [0,1]^d$, $d \geq 2$, with the Euclidean distance, $I_m = \{0, \dots, m-1\}$, $m \geq 3$, and $u_i(w) = aw + (1-a)\,x_i$, $i \in I_m$, $w \in W$, with $0 < a < 1$, where $x_0, \dots, x_{m-1}$ are distinct points in $\mathbf{R}^d$ such that the convex hull $C\,(x_0, \dots, x_{m-1})$ of them has dimension d. It is obvious that the support $K = K_a$ satisfies

$$K_a \subset C\,(x_0, \dots, x_{m-1})\,.$$

In this case, too, the composition $u_{i_1} \circ \dots \circ u_{i_n} = v_{i^{(n)}}$ can be expressed in closed terms as

$$v_{i^{(n)}}(w) = \left(a^{-1} - 1\right) \sum_{k=1}^{n} x_{i_k} a^k + a^n w$$

for any $i_1, \dots, i_n \in I_m$, $n \in \mathbf{N}_+$ and $w \in W$. Hence, for any $i^{(\infty)} = (i_1, i_2, \dots) \in I^{\mathbf{N}_+}$, the so-called *$d$-dimensional a-expansion*

$$\lim_{n \to \infty} \left(a^{-1} - 1\right) \sum_{k=1}^{n} x_{i_k} a^k = \left(a^{-1} - 1\right) \sum_{k \in \mathbf{N}_+} x_{i_k} a^k$$

exists and is an element of K_a. Conversely, any element of K_a is such an a-expansion.

Concerning K_a, things are not completely analogous to the case from Example 5.2, where $m = 2$, $d = 1$, and $x_0 = 0$, $x_1 = 1$. We shall present them while for the proofs the reader is referred to Sidorov (op. cit.).

If $a \geq d/(d+1)$ then $K_a = C\,(x_0, \dots, x_{m-1})$. There exists $a_0 \geq d/(d+1)$ such that one has $K_a = C\,(x_0, \dots, x_{m-1})$ for any $a \in (a_0, 1)$ while any point in $C\,(x_0, \dots, x_{m-1})$ except for its vertices has $2^{\aleph_0}$ distinct a-expansions.

Let $\mathcal{U}_a$ denote the set of $w \in K_a$ that have a unique a-expansion and $\mathcal{R}_a(w)$ the set of all a-expansions of a given $w \in K_a$. Define

$$\mathcal{V}_a = \left(w \in K_a : \operatorname{card} R_a(w) < 2^{\aleph_0}\right).$$

The Hausdorff dimensions of $\mathcal{V}_a$ and $\mathcal{U}_a$ are equal. If $K_a = C\,(x_0, \dots, x_{m-1})$ and there exist $i, j \in I_m$ such that a vertex of $u_i(C(x_0, \dots, x_{m-1}))$ belongs to the interior of $u_j(C(x_0, \dots, x_{m-1}))$, then Lebesgue-almost every $w \in C(x_0, \dots, x_{m-1})$ has $2^{\aleph_0}$ distinct a-expansions while the sets $\mathcal{U}_a$ and $\mathcal{V}_a$ have Hausdorff dimension strictly less than d.

For the triangular case $m = 3$ and $d = 2$, there is an explicit analogue of the results from Example 5.2. Here, $a_0 \approx 0.68233$ is the unique positive root of the equation $x^3 + x = 1$. Then for $a < a_0$ the set of uniqueness $\mathcal{U}_a$ is nonvoid while for $a \in (a_0, 1)$ any $w \in C\,(x_0, x_1, x_2) \setminus (x_0, x_1, x_2)$ has $2^{\aleph_0}$ distinct a-expansions.

A probability framework similar to that in Example 5.2 can be set up by considering a sequence $(\xi_n)_{n \in \mathbf{N}_+}$ of i.i.d. I_m-valued random variables with

$P(\xi_n = i) = p_i > 0,\ i \in I_m,\ n \in \mathbf{N}_+,\ \sum_{i\in I_m} p_i = 1$. This time we take $\Omega = I_m^{\mathbf{N}_+}$ and define $P = \mathbb{P}_{(p_0,\dots,p_{m-1})}$ as the product measure on Ω with equal multipliers $(p_0, \dots, p_{m-1})$. Then $\xi_n(\omega) = \mathrm{pr}_n\omega,\ n \in \mathbf{N}_+,\ \omega = (x_{i_1}, x_{i_2}, \dots) \in \Omega,\ i_k \in I_m,\ k \in \mathbf{N}_+$. Clearly, relations generalizing those from Example 5.2 hold, too. $\square$

6. THE STRICTLY STATIONARY CASE

We shall now consider the more general case where $(\xi_n)_{n\in\mathbf{N}_+}$ is an X-valued strictly stationary process. In particular, this means that the distribution of the random vector

$$(\xi_m, \dots, \xi_{m+k})$$

where $m \in \mathbf{N}_+$ and $k \in \mathbf{N}$, does not depend on m. It is frequently convenient to assume that we are given a doubly infinite strictly stationary process. By Doob's lemma (Proposition 4.6), there always exists a strictly stationary process $(\overline{\xi}_k)_{k\in\mathbf{Z}}$ with the same distribution as $(\xi_n)_{n\in\mathbf{N}_+}$. Still, $(\overline{\xi}_k)_{k\in\mathbf{Z}}$ is ergodic iff $(\xi_n)_{n\in\mathbf{N}_+}$ is ergodic.

On account of these facts, we'll suppose from the very beginning that we are dealing with an X-valued strictly stationary process $(\xi_k)_{k\in\mathbf{Z}}$, where X is a separable complete metric space. Note also that there exists an invertible transformation $\tau : \Omega \to \Omega$ which preserves the probability P such that $\xi_k(\omega) = \xi_0\left(\tau^k(\omega)\right),\ k \in \mathbf{Z},\ \omega \in \Omega$. Here, $(\Omega, \mathcal{K}, P)$ is the probability space that supports the random variables considered while preserving of P by τ amounts to the equations $P\tau^{-1}(A) := P\left(\tau^{-1}(A)\right) = P(A)$ for any $A \in \mathcal{K}$.

Under our new assumptions, we have again the random variables

$$\zeta_n^w = u_{\xi_n} \circ \cdots \circ u_{\xi_1}(w), \quad w \in W,$$

for any $n \in \mathbf{N}_+$. Now, differing from the case of independent random variables $\xi_n,\ n \in \mathbf{N}_+$, the random sequence $(\zeta_n^w)_{n\in\mathbf{N}_+}$ with $\zeta_0^w = w$, is *no longer* a Markov process. Instead, there are interesting results as n takes negative values. (We agreed that ξ_n is defined for $n \in \mathbf{Z}$.)

In what follows the self-mappings $u_x : W \to W,\ x \in X$, will be assumed to be Lipschitz. We should mention that unlike Elton [17], whose approach we are following here, we go on with considering a 'parametric' identification of the self-mappings $u_x : W \to W,\ x \in X$, which are assumed as before to be $(\mathcal{B}_W \otimes \mathcal{X}, \mathcal{B}_W)$-measurable. (Recall that $u_x(w) := u(w, x),\ w \in W,\ x \in X$). This allows us to preserve the assumption of a separable complete metric space W while Elton (op.cit.), dealing with a strictly stationary process whose values are Lipschitz self-mappings of W, is bound to assume that W is a separable complete locally compact metric space. A similar remark should be made in connection with a paper by Kellerer [31], where W is assumed to be an ordered topological space that is locally compact and second countable while i.i.d. processes whose values are order-preserving continuous self-mappings of W are considered.

PROPOSITION 6.1. *Assume that the u_x, $x \in X$, are Lipschitz functions with*

$$\int_\Omega \log^+ \mathrm{s}\,(u_{\xi_0})\,\mathrm{d}P < \infty.$$

(i) *There exists an invariant function $\chi : \Omega \to \mathbf{R} \cup \{-\infty\}$, called the Lyapunov exponent, with $\chi^+ \in L^1(P)$ and such that*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{s}\,(\zeta_n^\bullet) = \chi \ P\text{-a.s.}$$

and

$$\lim_{n\to\infty} \int_\Omega \frac{1}{n} \log \mathrm{s}\,(\zeta_n^\bullet)\,\mathrm{d}P = \inf_{n\in\mathbf{N}_+} \frac{1}{n} \int_\Omega \log \mathrm{s}\,(\zeta_n^\bullet)\,\mathrm{d}P = \int_\Omega \chi \mathrm{d}P.$$

(ii) *For any $k \in \mathbf{Z}$ we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{s}\left(u_{\xi_{k-1}} \circ \cdots \circ u_{\xi_{k-n}}\right) = \chi \ P\text{-a.s.}$$

Proof. (i) This follows from Kingman's subadditive ergodic theorem. Cf. Krengel [35, p. 40] for a special case (linear maps u_x, $x \in X$);

(ii) As noticed before, we have $\xi_k = \xi_0 \circ \tau^k$, $k \in \mathbf{Z}$, where τ is an invertible measure-preserving transformation of Ω. Then

$$\frac{1}{n} \log \mathrm{s}\left(u_{\xi_{k-1}} \circ \cdots \circ u_{\xi_{k-n}}\right) = \frac{1}{n} \log \mathrm{s}\left(u\left(\bullet, \xi_0 \circ \tau^{k-1}\right) \circ \cdots \circ u\left(\bullet, \xi_0 \circ \tau^{k-n}\right)\right)$$

converges P-a.s. as $n \to \infty$ to an invariant function, say, $\widetilde{\chi}$, by the same argument as in (i) above.

Next, assume for a while that L^1-convergence to $\widetilde{\chi}$ holds as well. Then, since τ is measure-preserving and $\widetilde{\chi}$ is invariant, we can write

$$\int_\Omega \left| \frac{1}{n} \log \mathrm{s}\left(u(\bullet, \xi_0 \circ \tau^{k-1}) \circ \cdots \circ u(\bullet, \xi_0 \circ \tau^{k-n})\right) - \widetilde{\chi} \right| \mathrm{d}P$$
$$= \int_\Omega \left| \frac{1}{n} \log \mathrm{s}\left(u\left(\bullet, \xi_0 \circ \tau^{n}\right) \circ \cdots \circ u\left(\bullet, \xi_0 \circ \tau\right)\right) - \widetilde{\chi} \right| \mathrm{d}P$$
$$= \int_\Omega \left| \frac{1}{n} \log \mathrm{s}\left(u_{\xi_n} \circ \cdots \circ u_{\xi_1}\right) - \widetilde{\chi} \right| \mathrm{d}P = \int_\Omega \left| \frac{1}{n} \log \mathrm{s}\,(\zeta_n^\bullet) - \widetilde{\chi} \right| \mathrm{d}P,$$

and the uniqueness of L^1-limit implies that $\widetilde{\chi} = \chi$ a.s. The general case can be reduced to the assumption of L^1-convergence by a truncation argument. □

Remark. If the process $(\xi_k)_{k\in\mathbf{Z}}$ is ergodic, then χ is a constant a.s. and one has $\chi < 0$ iff $\int_\Omega \log \mathrm{s}\,(\zeta_n^\bullet)\,\mathrm{d}P < 0$ for some $n \in \mathbf{N}_+$. This is the usual practical way to establish that $\chi < 0$ in the ergodic case. □

THEOREM 6.2. *Assume that the* u_x, $x \in X$, *are Lipschitz functions with*

$$\int_\Omega \log^+ \mathrm{s}\,(u_{\xi_0})\,\mathrm{d}P < \infty$$

and that

$$\int_\Omega \log^+ d\,(w_0, u_{\xi_0}(w))\,\mathrm{d}P < \infty$$

for some $w_0 \in W$. *Assume also that* $\chi < 0$ *P-a.s.* (cf. *Proposition* 6.1).

(i) *The limit*

$$\eta_k = \lim_{m\to\infty} u_{\xi_k} \circ \cdots \circ u_{\xi_{k-m+1}}(w)$$

exists for all $w \in W$ *and* $k \in \mathbf{Z}$, *and does not depend on* w. *The process* $(\eta_k)_{k\in\mathbf{Z}}$ *is a strictly stationary* W*-valued process.* (*As* $\eta_n = u_{\xi_n} \circ \cdots \circ u_{\xi_1} \circ \eta_0$, *P-a.s.*, $n \in \mathbf{N}_+$, *it appears that* η_0 *is a* W*-valued random variable of starting values making the process* $(\zeta_n^w)_{n\in\mathbf{N}}$ *a strictly stationary one.*)

(ii) *For any* $w \in W$ *the random sequence* $\left(\zeta_n^w,\ \zeta_{n+1}^w, \dots\right)$ *converges in distribution to* $(\eta_0, \eta_1, \dots)$ *as* $n \to \infty$. *In particular,* ζ_n^w *converges in distribution to* η_0 *as* $n \to \infty$.

(iii) *If the process* $(\xi_k)_{k\in\mathbf{Z}}$ *is ergodic, then so is* $(\eta_k)_{k\in\mathbf{Z}}$ *while the latter is a factor of the former in the sense of ergodic theory.* (Cf., *e.g., Cornfeld et al.* [11, p. 230]).

(iv) *For any* $w \in W$ *one has*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f\,(\zeta_i^w) = E\,(f\,(\eta_0)\,|\,\mathcal{I})\ \ P\text{-}a.s.,$$

where $f : W \to \mathbf{R}$ *is any real-valued bounded continuous function and* $\mathcal{I}$ *is the invariant tail* σ*-algebra of* $(\xi_k)_{k\in\mathbf{Z}}$. *In particular, if the process* $(\xi_k)_{k\in\mathbf{Z}}$ *is ergodic, then the empirical distribution of the trajectories of* $(\zeta_n^w)_{n\in\mathbf{N}_+}$ *converges weakly to the probability measure* $P\eta_0^{-1}$, *P-a.s.*

(v) *Let* S *denote the support of* $P\eta_0^{-1}$ *in* W. *Let* $w \in W$ *and* $\varepsilon > 0$ *be arbitrarily chosen. Then for P-almost all* $\omega \in \Omega$ *there exists* $n_0\,(\omega)$ *such that* $d\,(\zeta_n^w, S) \le \varepsilon$ *for any* $n \ge n_0\,(\omega)$. *So, in the ergodic case,* S *can be characterized as the set of points* $v \in W$ *such that for any* $w \in W$ *and any neighborhood* G *of* v, *the points* ζ_n^w, $n \in \mathbf{N}_+$, *visit* G *infinitely often P-a.s.*

Proof. (i) Using the triangle inequality and the inequality

$$\log^+(a+b) \le \max\left(\log^+ a,\ \log^+ b\right) + \log 2$$

valid for any positive a and b, it is easy to show that the hypotheses imply that

$$E(\log^+ d\,(w, u_{\xi_0}(w)) = \int_\Omega \log^+ d\,(w, u_{\xi_0}(w))\,\mathrm{d}P < \infty \tag{6.1}$$

for *any* $w \in W$.

For any fixed $k \in \mathbf{Z}$ let

$$\zeta_{k,m} := u_{\xi_k} \circ u_{\xi_{k-1}} \circ \cdots \circ u_{\xi_{k-m+1}}, \quad m \in \mathbf{N}_+$$

and note that

$$d\left(\zeta^w_{k,m}, \zeta^w_{k,m+1}\right) \leq \mathrm{s}\left(\zeta^\bullet_{k,m}\right) d\left(w, u_{\xi_{k-m}}(w)\right)$$

for any $w \in W$. Let $\Omega_j = \{\chi < -1/j\}$, $j \in \mathbf{N}_+$, so that $P\left(\bigcup_{j\in\mathbf{N}_+} \Omega_j\right) = 1$ since $\chi < 0$, as assumed. By Proposition 6.1 (ii) we have

$$\frac{1}{m} \log \mathrm{s}\left(\zeta^\bullet_{k,m}\right) < -1/j,$$

hence

$$\mathrm{s}\left(\zeta^\bullet_{k,m}\right) < \exp(-m/j) \tag{6.2}$$

for P-almost all $\omega \in \Omega_j$ and for all sufficiently large m (depending on ω). Next, by strict stationarity, for any $j \in \mathbf{N}_+$ and $w \in W$ we can write

$$\sum_{m\in\mathbf{N}_+} P\left(\log^+ d(w, u_{\xi_{k-m}}(w) > m/2j\right) =$$

$$= \sum_{m\in\mathbf{N}_+} P(\log^+ d(w, u_{\xi_0}(w)) > m/2j) \leq 2jE(\log^+ d(w, u_{\xi_0}(w)),$$

since

$$E(\log^+ d(w, u_{\xi_0}(w)) = \int_0^\infty P\left(\log^+ d(w, u_{\xi_0}(w)) > y\right) \mathrm{d}y =$$

$$= \sum_{m\in\mathbf{N}_+} \int_{(m-1)/2j}^{m/2j} P\left(\log^+ d(w, u_{\xi_0}(w)) > y\right) \mathrm{d}y \geq$$

$$\geq \frac{1}{2j} \sum_{m\in\mathbf{N}_+} P\left(\log^+ d(w, u_{\xi_0}(w)) > m/2j\right).$$

As $E\left(\log^+ d(w, u_{\xi_0}(w))\right)$ is finite by (6.1), the Borel-Cantelli lemma implies that

$$d(w, u_{\xi_{k-m}}(w)) \leq \exp(m/2j)$$

eventually for almost all $\omega \in \Omega_j$. In conjunction with (6.2) this yields

$$d\left(\zeta^w_{k,m}, \zeta^w_{k,m+1}\right) \leq \exp(-m/2j), \quad w \in W,$$

for m large enough and almost all $\omega \in \Omega_j$, so that $(\zeta_{k,m})_{m\in\mathbf{N}_+}$ is a Cauchy sequence in Ω_j and thus converges to a random variable, say, η_k. Finally, we have

$$d\left(\zeta^{w'}_{k,m}, \zeta^{w''}_{k,m}\right) \leq \mathrm{s}\left(\zeta^\bullet_{k,m}\right) d\left(w', w''\right)$$

for any $w', w'' \in W$, showing that η_k does not depend on w, as asserted. The strict stationarity of $(\eta_k)_{k\in\mathbf{Z}}$ follows at once from that of $(\xi_k)_{k\in\mathbf{Z}}$.

(ii) We have $\zeta_n^w = \zeta_{n,n}^w$, $n \in \mathbf{N}_+$. Then, by stationarity, $\left(\zeta_n^w, \zeta_{n+1}^w, \dots\right)$ has the same distribution as $\left(\zeta_{0,n}^w, \zeta_{1,n+1}^w, \dots\right)$, and, by (i), the latter converges in distribution to $(\eta_0, \eta_1, \dots)$ as $n \to \infty$.

(iii) As in the proof of Proposition 6.1(ii), we can assume that $\xi_k = \xi_0 \circ \tau^k$, $k \in \mathbf{Z}$, where τ is an ergodic invertible measure-preserving transformation of Ω. Therefore,

$$\eta_k = \lim_{n\to\infty} u_{\xi_0 \circ \tau^k} \circ \dots \circ u_{\xi_0 \circ \tau^{k-n+1}}(w), \quad w \in W,$$

P-a.s. for any $k \in \mathbf{Z}$. Hence $\eta_k = \eta_0 \circ \tau^k$, $k \in \mathbf{Z}$, P-a.s., so that the strictly stationary process $(\eta_k)_{k\in\mathbf{Z}}$ is ergodic.

(iv) Since W is a separable metric space, there exists a countable set $\mathcal{F}$ of real-valued bounded uniformly continuous functions defined on W such that any sequence $(\mu_n)_{n\in\mathbf{N}_+}$ of probability measures on $\mathcal{B}_W$ converges weakly to a probability measure μ on $\mathcal{B}_W$ if an only if

$$\lim_{n\to\infty} \int_W f \mathrm{d}\mu_n = \int_W f \mathrm{d}\mu$$

for any $f \in \mathcal{F}$ (cf. Parthasarathy [44, Theorems II 6.1 and II 6.6]).

First, by the pointwise ergodic theorem, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f\left(\zeta_i^{\eta_0}\right) = E\left(f\left(\eta_0\right) | \mathcal{I}\right) P\text{-a.s.} \tag{6.3}$$

and for any $f \in F$ let $\Omega(f)$ denote the set of P-probability 1 where equation (6.3) does hold.

Next, note that

$$d\left(\zeta_n^{\eta_0}, \zeta_n^w\right) \le \mathrm{s}\left(\zeta_n^{\bullet}\right) d\left(\eta_0, w\right) \to 0 \ P\text{-a.s.} \tag{6.4}$$

as $n \to \infty$ by Proposition 6.1(i). Let Ω_0 denote the intersection of the subset of Ω where (6.4) holds and of all the sets $\Omega(f)$, $f \in \mathcal{F}$. Clearly, $P\left(\Omega_0\right) = 1$.

Finally, consider the (random) empirical measures μ_n, $n \in \mathbf{N}_+$, of $(\zeta_n^{\eta_0})_{n\in\mathbf{N}_+}$ defined by

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^{n} I_A(\zeta_n^{\eta_0}), \quad A \in \mathcal{B}_W.$$

We clearly have

$$\int_W f \mathrm{d}\mu_n = \frac{1}{n} \sum_{i=1}^{n} f\left(\zeta_n^{\eta_0}\right), \quad n \in \mathbf{N}_+,$$

for any $f \in C(W)$. So that, by (6.3), μ_n converges weakly to the probability measure μ defined as

$$\mu(A) = P\left(\eta_0 \in A | \mathcal{I}\right), \quad A \in \mathcal{B}_W,$$

for any fixed $\omega \in \Omega_0$, μ_n while, by (6.4) and the uniform continuity of the functions in $\mathcal{F}$ we have

$$\lim_{n\to\infty} \frac{1}{n} \left(\sum_{i=1}^{n} \left(f\left(\zeta_i^w\right) - f\left(\zeta_i^{\eta_0}\right)\right) \right) = 0,$$

hence

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f\left(\zeta_i^w\right) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f\left(\zeta_i^{\eta_0}\right) = \int_W f \mathrm{d}\mu = E\left(f\left(\eta_0\right) \mid \mathcal{I}\right)$$

for any $\omega \in \Omega_0$ and $f \in \mathcal{F}$.

(v) Let $v \in S$ be arbitrarily fixed. By Proposition 6.1(i), for P-almost all $\omega \in \Omega$ there exists $n_0 = n_0(\omega)$ such that $\mathrm{s}\left(\zeta_n^\bullet\right) \le \varepsilon/2\left(1 + d\left(v, w\right)\right)$ for all $n \ge n_0$. Next, since

$$\eta_k = \lim_{m\to\infty} \zeta_{k,m}^v \in S \ P\text{-a.s.}$$

for all $k \in \mathbf{Z}$, the set of points

$$\lim_{m\to\infty} u_{\xi_0(\omega)} \circ \cdots \circ u_{\xi_0 \circ \tau^{-m}(\omega)}(v)$$

is dense in S for any set of points ω of P-probability 1. Hence for each $n \ge n_0$ there exists $m > n$ such that $d\left(\zeta_{n,m}^v, S\right) < \varepsilon/2$ and $d\left(v, \zeta_{0,n-m}^v\right) \le 1$. Therefore,

$$\begin{aligned} d\left(\zeta_n^w, S\right) &\le d\left(\zeta_n^w, \zeta_n^v\right) + d\left(\zeta_n^v, \zeta_{n,m}^v\right) + d\left(\zeta_{n,m}^v, S\right) \le \\ &\le \mathrm{s}\left(\zeta_n^\bullet\right) d\left(w, v\right) + \mathrm{s}\left(\zeta_n^\bullet\right) d\left(v, \zeta_{0,n-m}^v\right) + \frac{\varepsilon}{2} \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

for any $n \ge n_0$. □

Remark. Relation (6.4) and Proposition 6.1(i) yield

$$\limsup_{n\to\infty} \frac{1}{n} \log d\left(\zeta_n^{\eta_0}, \zeta_n^w\right) \le \chi \ P\text{-a.s.}$$

with w either random or nonrandom. Hence

$$d\left(\zeta_n^{\eta_0}, \zeta_n^w\right) \le \exp\left(n\left(\chi + a\right)\right) \ P\text{-a.s.}$$

for any $a < |\chi|$ and all n greater than some positive random integer depending on a. Recall that χ was assumed to be negative. □

The results just proved can be immediately extended to 'forward' compositions

$$\widetilde{\zeta_{k,m}^w} = u_{\xi_k} \circ \cdots \circ u_{\xi_{k+m-1}}(w), \quad w \in W,$$

where $k \in \mathbf{Z}$ and $m \in \mathbf{N}_+$. This can be done by applying Proposition 6.1 and Theorem 6.2 to the time-reversed process $\left(\xi_{-k}\right)_{k\in Z}$ that is again a strictly stationary process. We state all this as

Corollary 6.3. *Assume that the u_x, $x \in X$, are Lipschitz functions with*

$$\int_\Omega \log^+ \mathrm{s}\,(u_{\xi_0})\,\mathrm{d}P < \infty.$$

(j) *There exists an invariant function $\widetilde{\chi} : \Omega \to \mathbf{R} \cup \{-\infty\}$, called the (forward) Lyapunov exponent, with $\widetilde{\chi}^+ \in L^1(P)$ and such that*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{s}\left(\widetilde{\zeta^{\bullet}_{-n,n}}\right) = \widetilde{\chi}\ P\text{-a.s.}$$

and

$$\lim_{n\to\infty} \int_\Omega \frac{1}{n} \log \mathrm{s}\left(\widetilde{\zeta^{\bullet}_{-n,n}}\right) \mathrm{d}P = \inf_{n\in\mathbf{N}_+} \int_\Omega \frac{1}{n} \log \mathrm{s}\left(\widetilde{\zeta^{\bullet}_{-n,n}}\right) \mathrm{d}P = \int_\Omega \widetilde{\chi}\mathrm{d}P.$$

(jj) *For any $k \in \mathbf{Z}$ we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{s}\left(u_{\xi_k} \circ \cdots \circ u_{\xi_{k+n-1}}\right) = \widetilde{\chi}\ P\text{-a.s.}$$

If, in addition,

$$\int_\Omega \log^+ d\,(w_0, u_{\xi_0}(w_0))\,\mathrm{d}P < \infty$$

for some $w_0 \in W$ and $\widetilde{\chi} < 0$ P-a.s., then the following assertions hold.

$(\widetilde{\mathrm{i}})$ *The limit*

$$\widetilde{\eta}_k = \lim_{m\to\infty} u_{\xi_{-k}} \circ \cdots \circ u_{\xi_{m-k+1}}(w)$$

exists for all $w \in W$ and $k \in \mathbf{Z}$, and does not depend on w. The process $(\widetilde{\eta}_k)_{k\in\mathbf{Z}}$ is a strictly stationary W-valued process. (As $\widetilde{\eta}_k = u_{\xi_{-k}} \circ \cdots \circ u_{\xi_{-1}} \circ \widetilde{\eta}_0$, P-a.s., $k < 0$, it appears that $\widetilde{\eta}_0$ is a W-valued random variable of starting values making the process $(\widetilde{\zeta^w_{-n,n}})_{n\in\mathbf{N}}$ a strictly stationary one.)

$(\widetilde{\mathrm{ii}})$ *For any $w \in W$ the random sequence $\left(\widetilde{\zeta^w_{-n,n}},\ \widetilde{\zeta^w_{-(n+1),(n+1)}}, \dots\right)$ converges in distribution to $(\widetilde{\eta}_0, \widetilde{\eta}_1, \dots)$* as $n \to \infty$. *In particular, $\widetilde{\zeta^w_{-n,n}}$ converges in distribution to $\widetilde{\eta}_0$ as $n \to \infty$.*

$(\widetilde{\mathrm{iii}})$ *If the process $(\xi_k)_{k\in\mathbf{Z}}$ is ergodic, then so is $(\widetilde{\eta}_k)_{k\in\mathbf{Z}}$ while the latter is a factor of the former in the sense of ergodic theory.*

$(\widetilde{\mathrm{iv}})$ *For any $w \in W$ one has*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f\left(\widetilde{\zeta^w_{-i,i}}\right) = E\left(f\,(\widetilde{\eta}_0)|\,\widetilde{\mathcal{I}}\right)\ P\text{-a.s.},$$

where $f : W \to \mathbf{R}$ is any real-valued bounded continuous function and $\widetilde{\mathcal{I}}$ is the invariant tail σ-algebra of $(\xi_{-k})_{k\in\mathbf{Z}}$. In particular, if the process $(\xi_k)_{k\in\mathbf{Z}}$ is ergodic, then the empirical distribution of the trajectories of $(\widetilde{\zeta^w_{-n,n}})_{n\in\mathbf{N}}$ converges weakly to the probability measure $P\widetilde{\eta}_0^{-1}$, P-a.s.

($\widetilde{\mathrm{v}}$) *Let S denote the support of $P\widetilde{\eta}_0^{-1}$ in W. Let $w \in W$ and $\varepsilon > 0$ be arbitrarily chosen. Then for P-almost all $\omega \in \Omega$ there exists $n_0(\omega)$ such that $d\left(\widetilde{\zeta_{-n,n}^{w}}, S\right) \leq \varepsilon$ for any $n \geq n_0(\omega)$. So, in the ergodic case, S can be characterized as the set of points $v \in W$ such that for any $w \in W$ and any neighbourhood G of v, the points $\widetilde{\zeta_{-n,n}^{w}}$, $n \in \mathbf{N}_+$, visit G infinitely often P-a.s.* $\square$

Example 6.4. Let us take up again Example 5.2 where instead of a sequence $(\xi_n)_{n\in\mathbf{N}_+}$ of i.i.d. I-valued random variables we consider an I-valued strictly stationary process $(\xi_k)_{k\in\mathbf{Z}}$. In this case, by (5.5) again, we have

$$\widetilde{\zeta_{k,m}^{w}} := u_{\xi_k} \circ \cdots \circ u_{\xi_{k+m-1}}(w) = \left(a^{-1} - 1\right) \sum_{j=k}^{k+m-1} \xi_j a^{j-k+1} + a^m w,$$

$$\zeta_{k,m}^{w} := u_{\xi_k} \circ \cdots \circ u_{\xi_{k-m+1}}(w) = \left(a^{-1} - 1\right) \sum_{j=k-m+1}^{k} \xi_j a^{k-j+1} + a^m w,$$

for any $k \in \mathbf{Z}$, $m \in \mathbf{N}_+$, and $w \in [0,1]$, so that $\chi = \widetilde{\chi} = \log a < 0$. Hence

$$\widetilde{\eta}_k = \lim_{m\to\infty} \widetilde{\zeta_{-k,m}^{w}} = \left(a^{-1} - 1\right) \sum_{j\geq -k} \xi_j a^{j+k+1}$$

and

$$\eta_k = \lim_{m\to\infty} \zeta_{k,m}^{w} = \left(a^{-1} - 1\right) \sum_{j\leq k} \xi_j a^{k-j+1}$$

for all $k \in \mathbf{Z}$ and $w \in [0,1]$. However, in this case the support of the probability distribution of $\widetilde{\eta}_0$, that plays the part of ζ_∞ from Sections 4 and 5, actually, of *all* $\widetilde{\eta}_k$, $k \in \mathbf{Z}$, might be different from K_a defined in Example 5.2. We can assert that this support is K_a if, e.g., all the events $(\xi_1 = i_1, \ldots, \xi_{i_n} = i_n)$, $i_1, \ldots, i_n \in I$, $n \in \mathbf{N}_+$, have non-zero probabilities. Cf. Theorem 6.5 below. $\square$

In special cases, more can be said about the support S of $P\eta_0^{-1}$. Such a case is that when the strictly stationary process $(\xi_k)_{k\in\mathbf{Z}}$ is a strictly stationary irreducible X-valued Markov chain with $X = I := \{1, 2, \ldots, m\}$, $m \geq 2$, and transition matrix $(p_{ij})_{1\leq i,j\leq m}$.

Theorem 6.5 (Barnsley, Elton and Hardin [7]). *Assume W is a compact metric space and u_i, $i \in I = \{1, \ldots, m\}$, are contractions with $\max_{i\in I} \in \mathrm{s}(u_i) < 1$. Then*

$$\eta\left(i^{(\infty)}\right) = \lim_{n\to\infty} u_{i_1} \circ \cdots \circ u_{i_n}(w)$$

exists for every sequence $i^{(\infty)} = (i_1, i_2, \ldots)$, $i_j \in I$, $j \in \mathbf{N}_+$, and does not depend on w. The values $\eta\left(i^{(\infty)}\right)$ as $i^{(\infty)} = (i_1, i_2, \ldots)$ ranges over **allowable** *sequences, i.e., those for which* $p_{i_j i_{j+1}} > 0$, $j \in \mathbf{N}_+$, *correspond to the points of the support* S *of* $P\eta_0^{-1}$ *in Theorem* 6.2(v). *Further,* S *can be uniquely written as a union of non-empty compact sets* $S_1, \ldots S_m$ *that satisfy the invariance relation*

$$S_j = \bigcup_{\{i: p_{ij} > 0\}} u_j(S_i) \in I, \quad j \in I. \tag{6.5}$$

For the *proof* see Barnsley, Elton and Hardin (op.cit.) where results on the fractal dimension of S are also given. Clearly, (6.5) generalizes Hutchinson's equation $K = \bigcup_{i \in I} u_i(K)$, see Theorem 5.1. □

THEOREM 6.6 (Elton [17]). *Assume the hypotheses of Theorem* 6.2 *under our Markov assumptions on* $(\xi_k)_{k \in \mathbf{Z}}$. *There exist non-empty closed sets* S_i, $i \in I = \{1, \ldots, m\}$, *such that* $S = \bigcup_{i \in 1} S_i$, *where* S *is the support of* $P\eta_0^{-1}$ *in Theorem* 6.2(v) *and*

$$\bigcup_{\{i: p_{ij} > 0\}} u_j(S_i) \subset S_j \subset \bigcup_{\{i: p_{i,j} > 0\}} \overline{u_j(S_i)}, \quad j \in I.$$

The S_j, $j \in I$, *are minimal, to mean that if* S_i', $i \in I$, *are non-empty closed sets satisfying*

$$\bigcup_{\{i: p_{ij} > 0\}} u_j\left(S_i'\right) \subset S_j', \quad j \in I,$$

then $S_j \subset S_j'$, $j \in I$. *In particular, if the* u_i, $i \in I$, *are closed mappings* (*e.g., non-singular affine mappings*), *then*

$$\bigcup_{\{i: p_{ij} > 0\}} u_j(S_i) = S_j, \quad j \in I.$$

For the *proof* see Elton [17, pp. 45–46]. □

Remark. If $(\xi_k)_{k \in \mathbf{Z}}$ is a strictly stationary Markov process, then, in general, $(\eta_k)_{k \in \mathbf{Z}}$ *is not* Markov. However, $(\eta_k, \xi_k)_{k \in \mathbf{Z}}$ *is* a $W \times X$-valued strictly stationary Markov process with transition function

$$p((w, x), A) = \int_X q(x, \mathrm{d}y) I_A(u_y(w), y), \quad w \in W, \ x \in X, \ A \in \mathcal{B}_{W \times X}.$$

Here, q is the transition function of $(\xi_k)_{k \in \mathbf{Z}}$. □

We conclude this section by particularizing the results proved here for the case of a process $(\xi_k)_{k \in \mathbf{Z}}$ with i.i.d. random variables ξ_k, $k \in \mathbf{Z}$. It is interesting to compare the results thus obtained with those given in Section 3.

Such a process $(\xi_k)_{k\in\mathbf{Z}}$ with $P(\xi_k \in A) = p(A),\ k \in \mathbf{Z},\ A \in \mathcal{X}$, is clearly ergodic, so that

$$\chi = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{s}\,(u_{\xi_n} \circ \cdots \circ u_{\xi_1})$$

exists P-a.s. and is a constant. Moreover, we have

$$\chi = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{s}\left(u_{\xi_{k-1}} \circ \cdots \circ u_{\xi_{k-n}}\right)\ P\text{-a.s.}$$

for any $k \in \mathbf{Z}$. Also, $\chi < 0$ is equivalent to

$$\text{(6.6)}\quad \int_\Omega \log \mathrm{s}(u_{\xi_n} \circ \cdots \circ u_{\xi_1}) \mathrm{d}P = \int_{X^n} \log \mathrm{s}(u_{x_n} \circ \cdots \circ u_{x_1}) p(\mathrm{d}x_1) \ldots p(\mathrm{d}x_n) < 0$$

for some $n \in \mathbf{N}_+$. In particular, (6.6) holds if (3.12) holds, that is, if

$$\int_X \log \mathrm{s}\,(u_x)\, p\,(\mathrm{d}x) < 0.$$

Note also that by Proposition 6.1(i), Corollary 6.3$(\widetilde{\mathrm{i}})$, and (6.6) we have $\chi = \widetilde{\chi}$ P-a.s.

On account of Proposition 6.1 and Theorem 6.2 we can state

THEOREM 6.7. *Consider the special context above. Assume that* (6.6) *holds and that* $\int_X \log^+ \mathrm{s}\,(u_x)\, p\,(\mathrm{d}x) < \infty$ *and* $\int_X \log^+ d\,(w_0, u_x\,(w_0))\, p\,(\mathrm{d}x) < \infty$ *for some* $w_0 \in W$. *Then the limit*

$$\eta_k = \lim_{m\to\infty} u_{\xi_k} \circ \cdots \circ u_{\xi_{k-m+1}}(w)$$

exists for all $w \in W$ *and* $k \in \mathbf{Z}$, *and does not depend on* w. *The process* $(\eta_k)_{k\in\mathbf{Z}}$ *is strictly stationary and ergodic. For any* $w \in W$ *the random sequence* $\left(\zeta_n^w, \zeta_{n+1}^w, \ldots\right)$ *converges in distribution to* $(\eta_0, \eta_1, \ldots)$ *as* $n \to \infty$. *In particular,* ζ_n^w *converges in distribution to* η_0 *as* $n \to \infty$, *and the probability measure* $P\eta_0^{-1}$ *is a stationary distribution for the Markov chain* $(\zeta_n^w)_{n\in\mathbf{N}}$.

Remark. Similar results can be stated for the random variables $\widetilde{\eta_k},\ k \in \mathbf{Z}$. See Corollary 6.3. □

It should be noted that Theorem 6.7 provides what can might be seen as minimal conditions for the existence of a stationary probability for the Markov process $(\zeta_n^w)_{n\in\mathbf{N}_+}$ in the case of an i.i.d. sequence $(\xi_n)_{n\in\mathbf{N}_+}$.The logarithmic-moment assumptions involved here should be compared with those in Section 3. The price paid for such minimal assumptions is the inavailability of an estimate of the rate of convergence to the stationary distribution as in Theorems 3.3, 3.5, 3.6, and Corollary 3.8.

APPENDIX 1. METRICS AND DISTANCES

A1.1. Let W be an arbitrary non-empty set. A pair (W, d) is said to be a *metric* (*distance*) *space* with *metric* (*distance*) d if d maps $W \times W$ into $\mathbf{R}_+$ ($\mathbf{R}_+ \cup \{\infty\}$) and has the following properties for any $x, y, z \in W$:

(i) Identity: $d(x,y) = 0 \Leftrightarrow x = y$;

(ii) Symmetry: $d(x,y) = d(y,x)$;

(iii) Triangle inequality: $d(x,z) \le d(x,y) + d(y,z)$.

It has been usual to only consider functions d taking on finite values. We shall distinguish between the case when d takes on finite values and the case when the value ∞ for d is allowed, by using the terms metric, respectively, distance, as indicated. Cf. Rachev [46, p. 9]. There is a very simple relationship between a metric and a distance: a distance space splits canonically into subspaces that carry (finite) metrics and are separated from one another by infinite distances. More precisely, the relation $d(x,y) \ne \infty$, $x, y \in W$, is an equivalence relation; any of its equivalence classes together with the corresponding restriction of d, is a metric space. [In general, if (W, d) is a metric (distance) space and A is a subset of W, then a metric (distance) on A is obtained by simply restricting d to $A \times A$, that is, the metric (distance) between points of A is equal to the metric (distance) between these points in (W, d).]

A pair (W, d) is said to be a *semi-metric* (*semi-distance*) *space* with *semi-metric* (*semi-distance*) d if just properties (ii) and (iii) above are satisfied. In other words zero distance between distinct points is allowed. It is easy to see that identifying points with zero distance in a semi-metric (semi-distance) space leads to a metric (distance) space.

Formally, the relation $d(x,y) = 0$, $x, y \in W$, is an equivalence relation. The semi-metric (semi-distance) d induces a metric (distance) in the set of its equivalence classes since we have $d(v,w) = d(v_1,w_1)$ if $d(v,v_1) = d(w,w_1) = 0$.

A1.2. Given a metric space (W, d) with Borel σ-algebra $\mathcal{B}_W$, let us denote by $\mathrm{pr}(\mathcal{B}_W)$ the collection of all probability measures on $\mathcal{B}_W$. In $\mathrm{pr}(\mathcal{B}_W)$ a distance ρ_H is defined by

$$\rho_H(\mu,\nu) = \sup\left\{ \int_W f \mathrm{d}\mu - \int_W f \mathrm{d}\nu \,\middle|\, f \in \mathrm{Lip}_1(W) \right\}$$

for any $\mu, \nu \in \mathrm{pr}(\mathcal{B}_W)$, where $\mathrm{Lip}_1(W) = \{f : W \to \mathbf{R} \mid \mathrm{s}(f) \le 1\}$ with

$$\mathrm{s}(f) = \mathrm{s}(f,d) := \sup_{\substack{w' \ne w'' \\ w', w'' \in W}} \frac{|f(w') - f(w'')|}{d(w', w'')}.$$

It is possible that $\rho_H(\mu,\nu) = \infty$ for some $\mu, \nu \in \mathrm{pr}(\mathcal{B}_W)$. However, we have $\rho_H(\mu,\nu) < \infty$ when, for instance, both μ and ν have bounded supports. Cf. Hutchinson [27, p. 732].

A genuine well-known metric in $\mathrm{pr}(\mathcal{B}_W)$ is the Lipschitz metric ρ_L which is defined by

$$\rho_L(\mu,\nu) = \sup\left\{\int_W f\mathrm{d}\mu - \int_W f\mathrm{d}\nu \,\middle|\, f \in \mathrm{Lip}_1(W),\ 0 \le f \le 1\right\}$$

for any $\mu, \nu \in \mathrm{pr}(\mathcal{B}_W)$. If (W,d) is a separable (complete) metric space, then $(\mathrm{pr}(\mathcal{B}_W), \rho_L)$ is a separable (complete) metric space, too. Another usual metric in $\mathrm{pr}(\mathcal{B}_W)$ is the Prokhorov metric ρ_p which is defined by

$$\rho_p(\mu,\nu) = \inf\{\varepsilon > 0 \mid \mu(A) \le \varepsilon + \nu(A^\varepsilon),\ A \in \mathcal{B}_W\}$$

for any $\mu, \nu \in \mathrm{pr}(\mathcal{B}_W)$, where $A^\varepsilon = \left\{w \mid d(w,A) = \inf_{a\in A}(w,a) < \varepsilon\right\}$. We have

$$\frac{1}{2}\,\rho_L(\mu,\nu) \le \rho_p(\mu,\nu) \le \rho_L^{1/2}(\mu,\nu)$$

for any $\mu, \nu \in \mathrm{pr}(\mathcal{B}_W)$. Cf. Hoffmann-Jørgensen [26, pp. 80–82].

Clearly, $\rho_L(\mu,\nu) \le \rho_H(\mu,\nu)$ and $\rho_p(\mu,\nu) \le \rho_H^{1/2}(\mu,\nu)$ for any $\mu, \nu \in \mathrm{pr}(\mathcal{B}_W)$.

APPENDIX 2. THE HAUSDORFF DISTANCE AND METRIC

A2.1. Let A and B be subsets of a metric (distance) space (W,d). Define the *Hausdorff distance* $d_H(A,B)$ between A and B by

$$d_H(A,B) = \inf\{\varepsilon > 0 \mid A^\varepsilon \supset B,\ B^\varepsilon \subset A\},$$

where

$$A^\varepsilon = \{w \in W \mid d(w,A) < \varepsilon\}.$$

Convenient reformulations of the definition above are as follows:

(i) $d_H(A,B) = \max\left(\sup_{x\in A} d(x,B), \sup_{y\in B} d(y,A)\right)$;

(ii) $d_H(A,B) \le \varepsilon$ iff $d(x,B) \le \varepsilon$ for any $x \in A$ and $d(y,A) \le \varepsilon$ for any $y \in B$ (this statement fails when '$\le$' is replaced by '$<$').

A2.2. The following properties can be easily proved:

(j) d_H is a semi-metric or a semi-distance in $\mathcal{P}(W)$;

(jj) $d_H(A, \mathrm{cl}\,A) = 0$ for any $A \in \mathcal{P}(W)$;

(jjj) if A and B are closed subsets of W and $d_H(A,B) = 0$, then $A = B$.

Let $\mathrm{cl}\,W$ denote the collection of all closed non-empty subsets of W. By properties (j) and (jjj) above, $(\mathrm{cl}\,W, d_H)$ is a metric or a distance space. (Note that by (jj) we also have $d_H(A,B) = d_H(\mathrm{cl}(A), \mathrm{cl}(B))$ for any $A, B \in \mathcal{P}(W)$.) Clearly, d_H may take on in $\mathrm{cl}\,W$ both finite and infinite values while if (W,d) is a bounded metric space, then d_H is finite and $(\mathrm{cl}\,W, d_H)$ is a metric space.

More generally, letting $\mathrm{bcl}\, W$ denote the collection of all bounded closed non-empty subsets of W, $(\mathrm{bcl}\, W, d_H)$ always is a metric space.

A2.3. Assume a sequence $(A_n)_{n\in\mathbf{N}_+}$ of closed non-empty subsets of W converges in $(\mathrm{cl}\, W, d_H)$ to some closed subset $A \subset W$: $\lim_{n\to\infty} d_H(A_n, A) = 0$. Then (i) A is the set of limits in (W, d) of all convergent sequences $(a_n)_{n\in\mathbf{N}_+}$ with $a_n \in A_n$, $n \in \mathbf{N}_+$, and (ii) $A = \bigcap_{n\in\mathbf{N}_+} \mathrm{cl}\left(\bigcup_{m\geq n} A_m\right)$. In particular, if (W, d) is compact and for non-empty compact subsets A_n, $n \in \mathbf{N}_+$, of W one has $A_{n+1} \subset A_n$ (resp. $A_n \subset A_{n+1}$), $n \in \mathbf{N}_+$, then $\lim_{n\to\infty} d_H(A_n, \bigcap_{m\in N_+} A_m) = 0$, resp. $\lim_{n\to\infty} d_H\Big(A_n, \mathrm{cl}\big(\bigcup_{m\in N_+} A_m\big)\Big) = 0$). Also, if $A_n \to A$ as $n \to \infty$ in $(\mathrm{cl}\, \mathbf{R}^m, d_H)$ for some $m \in N_+$ and the A_n are convex, then A is convex, too.

A2.4. If a metric (distance) space (W, d) is separable [resp. complete; resp. totally bounded (precompact); resp. compact], then both $(\mathrm{cl}\, W, d_H)$ and $(\mathrm{bcl}\, W, d_H)$ are separable [resp. complete; resp. totally bounded (precompact); resp. compact].

In particular, the collection of all closed convex sets contained in any fixed closed ball S in $\mathbf{R}^m$ for some $m \geq 1$, is compact in $(\mathrm{cl}\, S, d_H)$ (Blaschke's theorem).

A2.5. Let $\mathrm{c}W\, (\subset \mathrm{bcl}\, W \subset \mathrm{cl}\, W)$ denote the collection of all compact non-empty subsets of W. Clearly, $(\mathrm{c}W, d_H)$ is a metric space. In this restricted framework, the Hausdorff metric can be also defined as

$$d_H(A, B) = \sup_{w\in W} |d(w, A) - d(w, B)|$$

for any $A, B \in \mathrm{c}W$. Cf. Hoffmann-Jørgensen [26, p. 72]. All properties stated in **A2.4** for $(\mathrm{cl}W, d_H)$ and $(\mathrm{bcl}\, W, d_H)$ still hold for $(\mathrm{c}W, d_H)$.

A2.6. Let $A_i, B_i \in \mathrm{cl}\, W,\ 1 \leq i \leq n,\ n \geq 2$. Then

$$d_H\left(\bigcup_{i=1}^n A_i,\ \bigcup_{i=1}^n B_i\right) \leq \min{}_{(j_1,\ldots,j_n)\in\prod_n} \max\left(d_H\left(A_1, B_{j_1}\right), \ldots, d_H\left(A_n, B_{j_n}\right)\right),$$

where $\prod_n$ stands for the collection of the $n!$ permutations of $1, \ldots, n$. In particular,

$$d_H\left(\bigcup_{i=1}^n A_i,\ \bigcup_{i=1}^n B_i\right) \leq \max\left(d_H\left(A_1, B_1\right), \ldots, d_H\left(A_n, B_n\right)\right).$$

A2.7. Let $F : W \to W$ be a Lipschitz self-mapping of W. Then

$$d_H\left(F(A), F(B)\right) \leq \mathrm{s}(F)\ d_H(A, B)$$

for any $A, B \in \mathrm{cl}\, W$.

Acknowledgements. The author gratefully acknowledges support from the Deutsche Forschungsgemeinschaft under Grant 936 RUM 113/21/0-2. He also expresses his gratitude towards Prof. Dr. Ulrich Herkenrath for warm hospitality and unlimited kindness during many years of fruitful collaboration in both Germany and Romania. Thanks are also due to him for his very careful critical reading of several preliminary versions of this paper. The last of them has appeared as Preprint SM-DU-692/2009 with the Mathematical Institute of Duisburg – Essen University, Campus Duisburg.

REFERENCES

[1] R. Abraham and Y. Ueda (Eds.), *The Chaos Avant-garde. Memories of the Early Days of Chaos Theory.* World Sci. Publ. Co., Inc., River Edge, NJ, 2000.

[2] A. Abrams, H. Landau, Z. Landau, J. Pommersheim and E. Zaslow, *An iterated random function with Lipschitz number one.* Teor. Veroyatnost. i Primenen. **47** (2002), 286–300.

[3] K.A. Athreya and Ö. Stenflo, *Perfect sampling for Doeblin chains.* Sankhyā **65** (2003), *4*, 1–15.

[4] M.F. Barnsley, *Fractals Everywhere.* Academic Press, Boston, MA, 1988.

[5] M.F. Barnsley, *Superfractals.* Cambridge Univ. Press, Cambridge, 2006.

[6] M.F. Barnsley and J.H. Elton, *A new class of Markov processes for image encoding.* Adv. in Appl. Probab. **20** (1988), 14–32.

[7] M.F. Barnsley, J. Elton and D. Hardin, *Recurrent iterated function systems.* Constructive Approx. **5** (1989), 3–31.

[8] R. Bergmann and D. Stoyan, *Monotonicity properties of second order characteristics of stochastically monotone Markov chains.* Math. Nachr. **85** (1978), 99–102.

[9] P.C. Bressloff and J. Stark, *Neural networks, learning automata and iterated function systems.* pp. 145–164. In Crilly, Earnshaw and Jones (Eds.) [12].

[10] J.-F. Chamayou and G. Letac, *Explicit stationary distributions for compositions of random functions and products of random matrices.* J. Theoret. Probab. **4** (1991),*1*, 3–36.

[11] I.P. Cornfeld, S.V. Fomin and Ya.G. Sinaĭ, *Ergodic Theory.* Springer-Verlag, New York, 1982.

[12] A.J. Crilly, R.A. Earnshaw and J. Jones (Eds.), *Fractals and Chaos.* Springer-Verlag, New York, 1991.

[13] P. Diaconis and D. Freedman, *Iterated random functions.* SIAM Rev. **41** (1999), 45–76 & 77–82.

[14] J.L. Doob, *Stochastic Processes.* Wiley, New York, 1953.

[15] L.E. Dubins and D.A. Freedman, *Invariant probabilities for certain Markov processes.* Ann. Math. Statist. **32** (1966), 837–848.

[16] P. Dubischar, *The Representation of Transition Probabilities by Random Maps.* Ph. D. Dissertation, Fachbereich 3 (Mathematik & Informatik), Universität Bremen, 1999.

[17] J.H. Elton, *A multiplicative ergodic theorem for Lipschitz maps.* Stochastic Process. Appl. **34** (1990), 39–47.

[18] P. Erdös, *On a family of symmetric Bernoulli convolutions.* Amer. J. Math. **61** (1939), 974–975.

[19] P. Erdös, *On the smoothness properties of Bernoulli convolutions.* Amer. J. Math. **62** (1940), 180–186.

[20] J. Gleick, *Chaos. Making a New Science.* Penguin Books, New York, 1987.

[21] M. Hata, *On the structure of self-similar sets.* Japan J. Appl. Math. **2** (1985), 381–414.

[22] U. Herkenrath and M. Iosifescu, *On a contractibility condition for iterated random functions.* Rev. Roumaine Math. Pures Appl. **52** (2007), 563–571.

[23] U. Herkenrath, M. Iosifescu and A. Rudolph, *Random systems with complete connections and iterated function systems.* Math. Rep. (Bucur.) **5**(**55**) (2003), 127–140.

[24] O. Hernández-Lerma and J.B. Lassere, *Markov Chains and Invariant Probabilities.* Birkhäuser, Basel–Boston–Berlin, 2003.

[25] E. Hewitt and K. Stromberg, *Real and Abstract Analysis.* Springer-Verlag, New York, 1965.

[26] J. Hoffmann-Jørgensen, *Probability with a Wiew towards Statistics*, Vol. 2. Chapman & Hall, New York–London, 1994.

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

[28] M. Iosifescu, *A simple proof of a basic theorem on iterated random functions.* Proc. Rom. Acad. Ser. A Math. Phys. Tech. Sci. Inf. Sci. **4** (2003), 167–174.

[29] M. Iosifescu and S. Grigorescu, *Dependence with Complete Connections and its Applications.* Cambridge Tracts in Math. **96**. Cambridge Univ. Press, Cambridge, 1990. (Corrected paperback edition, 2009)

[30] S.F. Jarner and R.L. Tweedie, *Locally contracting iterated functions and stability of Markov chains.* J. Appl. Probab. **38** (2001), 494–507.

[31] H.G. Kellerer, Random dynamical systems on ordered topological spaces. *Stoch. Dyn.* **6** (2006), 255–300.

[32] Y. Kifer, *Ergodic Theory of Random Transformations.* Birkhäuser, Boston, 1986.

[33] F.B. Knight, *On the absolute difference chains.* Z. Wahrch. Verw. Gebiete **43** (1978), 57–63.

[34] V. Komornik and P. Loreti, *Unique developments in non-integer bases.* Amer. Math. Monthly **105** (1998), 636–639.

[35] U. Krengel, *Ergodic Theorems.* With a supplement by Antoine Brunel. Walter de Gruyter, Berlin, 1985.

[36] K.-S. Lau and S.-M. Ngai, *A generalized finite type condition for iterated function systems.* Adv. Math. **208** (2007), 647–671.

[37] K.-S. Lau and Y.-L. Ye, *Ruelle operator with nonexpansive IFS.* Studia Math. **148** (2001), 143–169.

[38] J.P. Leguesdron, *Marche aléatoire sur le semi-groupe des contractions de* $\mathbf{R}^d$. *Cas de la marche aléatoire sur* $\mathbf{R}_+$ *avec choc élastique en zéro.* Ann. Inst. H. Poincaré Probab. Statist. **25** (1989), 483–502.

[39] G. Letac, *A contraction principle for certain Markov chains and its applications.* In: *Random Matrices and their Applications* (*Brunswick, Maine*, 1984), pp. 263–273. Contemp. Math. **50**. Amer. Math. Soc., Providence, RI, 1986.

[40] R.D. Mauldin and M. Urbánski, *Dimensions and measures in infinite iterated function systems.* Proc. London Math. Soc. (3) **73** (1996), 105–154.

[41] R.D. Mauldin and M. Urbánski, *Graph Directed Markov Systems: Geometry and Dynamics of Limit Sets.* Cambridge Univ. Press, Cambridge, 2003.

[42] 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), 715–730.

[43] G.L. O'Brien, *The comparison method for stochastic processes.* Ann. Probab. **3** (1975), 80–88.

[44] K.R. Parthasarathy, *Probability Measures on Metric Spaces.* Academic Press, New York–London, 1967.

[45] Y. Peres, W. Schlag and B. Solomyak, *Sixty years of Bernoulli convolutions.* In: *Fractal Geometry and Stochastics*, II (*Greifswald/Koserow*, 1998), pp. 39–65. Progress in Probability **46**. Birkhäuser, Basel, 2000.
[46] S.T. Rachev, *Probabilty Metrics and the Stability of Stochastic Models.* Wiley, Chichester, 1991.
[47] N. Sidorov, *Combinatorics of linear iterated function systems with overlap.* Nonlinearity **20** (2007), 1299–1312.
[48] B. Solomyak, *On the random series* $\sum \pm\lambda^n$ (*an Erdös problem*). Ann. of Math. (2) **142** (1995), 611–625.
[49] W.B. Wu and X. Shao, *Limit theorems for iterated random functions.* J. Appl. Probab. **41** (2004), 425–436.
[50] R. Zaharopol, *Invariant Probabilities of Markov-Feller Operators and their Supports.* Birkhäuser Verlag, Basel, 2005.

Received 12 March 2009

Romanian Academy
"Gheorghe Mihoc-Caius Iacob" Institute
of Mathematical Statistics and Applied Mathematics
Casa Academiei Române
Calea 13 Septembrie nr. 13
050711 Bucharest 5, Romania
miosifes@acad.ro