On the Hausdorff Dimension of Invariant Measures of Weakly Contracting on Average Measurable IFS

Joanna Jaroszewska · Michał Rams

Abstract We consider measures which are invariant under a measurable iterated function system with positive, place-dependent probabilities in a separable metric space. We provide an upper bound of the Hausdorff dimension of such a measure if it is ergodic. We also prove that it is ergodic iff the related skew product is.

1 Introduction and Statement of a Result

In this note we give a contribution to the study of the multifractal properties of measures which are invariant for iterated function systems. Recently this aspect of such measures has been widely investigated, e.g. some results concerning their Hausdorff dimension were obtained in [5–9] and [3]. For instance, in [9] and [6], the systems contracting on average and having Dini-continuous, separated from zero probabilities were considered and the upper bound of the Hausdorff dimension of the unique (in this case) invariant distribution was given. A certain class of contracting on average systems with constant probabilities was dealt with in [3]. (Note that "contracting on average" has different meanings in [3] and in [6, 9].) The system of that kind could have more than one probability invariant measure, however, in [3] the upper estimation of the Hausdorff dimension of any of them was provided. It seems that it was the first attempt to study iterated function systems without uniqueness of invariant distributions in this respect. Another such attempt was made in [5], where the exact dimension of ergodic invariant measures was calculated for a system which acts on a compact interval, is non-overlapping and has continuous probabilities.

J. Jaroszewska
Institute of Mathematics, Warsaw University, ul. Banacha 2, 02-097 Warszawa, Poland
e-mail: asia@venco.com.pl

M. Rams (✉)
Institute of Mathematics, Polish Academy of Sciences, ul. Śniadeckich 8, 00-950 Warszawa, Poland
e-mail: rams@impan.gov.pl

Here we would like to continue the line of research which we described above. Assume that (X, ϱ) is a separable metric space and $\{X, S_i, p_i\}, i \in I$ is a finite iterated function system with positive probabilities and with ergodic invariant measure μ. Given $N \in \mathbb{N}$, denote

$$h_N(\mu, \delta) = - \int_X \sum_{i_1 \ldots i_N \in I^N} p_{i_1 \ldots i_N}(x) \inf_{y \in B_N(x, i_1 \ldots i_N, \delta)} \log p_{i_1 \ldots i_N}(y) \, \mu(dx)$$

where $B_N(x, i_1 \ldots i_N, \delta)$ is the set of all points $y \in X$ such that for all $n \in [0, N]$

$$\varrho(S_{i_n} \circ \ldots \circ S_{i_1}(y), S_{i_n} \circ \ldots \circ S_{i_1}(x)) < \delta$$

holds. We will denote

$$h_N(\mu) = \lim_{\delta \searrow 0} h_N(\mu, \delta)$$

and

$$h(\mu) = \lim_{N \to \infty} \frac{1}{N} h_N(\mu)$$

Similarly, let

$$\lambda_N(\mu, \delta) = \int_X \sum_{i_1 \ldots i_N \in I^N} p_{i_1 \ldots i_N}(x)$$

$$\times \sup_{\substack{y \in B_N(x, i_1 \ldots i_N, \delta) \\ y \neq x}} \log \frac{\varrho(S_{i_N} \circ \ldots \circ S_{i_1}(x), S_{i_N} \circ \ldots \circ S_{i_1}(y))}{\varrho(x, y)} \mu(dx)$$

$$\lambda_N(\mu) = \lim_{\delta \searrow 0} \lambda_N(\mu, \delta)$$

and

$$\lambda(\mu) = \lim_{N \to \infty} \frac{1}{N} \lambda_N(\mu)$$

(and we accept $-\infty$ as value of $\lambda(\mu)$). As we will see, $h_N(\mu, \delta)$ and $\lambda_N(\mu, \delta)$ are monotone with respect to δ and subadditive with respect to N, hence for $h(\mu)$ and $\lambda(\mu)$ to exist it is enough to check that $h_1(\mu, \delta)$ and $\lambda_1(\mu, \delta)$ exist for some $\delta > 0$. Our main result is as follows.

Theorem 1 *Under the assumptions as above,*

$$\dim_H(\mu) \leq -\frac{h(\mu)}{\lambda(\mu)}, \tag{1.1}$$

provided the right-hand side is well defined and nonnegative.

We would like to emphasize that the assumptions of the above theorem are weaker than the ones that usually appear in the context of place-dependent iterated function systems. Even the existence of invariant measures is not assured here—it must be guaranteed by additional assumptions. However, it seems desirable to strengthen Theorem 1 so that the field of its applicability would contain the systems with probabilities positive only on certain parts of the space, recently considered by I. Werner (see e.g. [10]).

Note that both $h(\mu)$ and $\lambda(\mu)$ are well known in the case of continuous maps S_i and continuous probabilities p_i: $h(\mu)$ is the Kolmogorov-Sinai metric entropy of μ and $\lambda(\mu)$ is the greatest Lyapunov exponent of the system (with respect to the measure μ). In such a situation, the formula (1.1) is a generalization of the well known Hofbauer-Raith formula: ratio of entropy to the Lyapunov exponent.

To satisfy the assumptions of the above theorem, the system must have $\lambda_N(\mu)$ negative or equal $-\infty$ for some N (for otherwise the denominator limit would be from $[0^+, \infty]$ and the whole formula would be negative). This property, crucial for the proof, could be seen as a weak form of contractibility on average.

Let us present here an example of application of our result.

Example 2 Let $S_1(x) = x/3$ and $S_2(x) = (x+2)/3$, both maps acting on $X = [0, 1]$. Let $A \subset [0, 1]$ be a set with (at most) countable boundary and let $p \in (0, 1/2)$. Set $p_1(x) = p$ for $x \in A$ and $p_1(x) = 1 - p$ otherwise. Set $p_2 = 1 - p_1$. This iterated function system has at least one ergodic invariant measure and every its ergodic invariant measure μ satisfies

$$\dim_H(\mu) \leq -\frac{p \log p + (1-p) \log(1-p)}{\log 3} \tag{1.2}$$

The paper is divided as follows. In the Sect. 2 we introduce the notation and give introductory information about iterated function systems and Markov operators. We finish the section with discussion of Example 2. We also give there an important result on the relationship between ergodicity of iterated function systems and ergodicity of the corresponding skew product. This allows us to finish the proof of our main result, which we give in the Sect. 3.

2 Preliminaries

Let (X, ϱ) be a fixed nonempty separable metric space and let I be a finite set of cardinality at least 2. The following notation will be used through this paper. The set of natural numbers will not contain 0, i.e. $\mathbb{N} = \{1, \ldots\}$. To count elements of covers of X, which are needed to estimate the Hausdorff dimension of a measure, we use the space $\Sigma = I^{\mathbb{N}}$. We endow it with the product topology of I taken with the discrete metric. For a sequence $\omega \in \Sigma$, the n-th term of ω is denoted by ω_n, whereas by ω^n—the concatenation of the first n terms of ω (i.e. a finite sequence $(\omega_1, \ldots, \omega_n) \in I^n$). Such an ω^n determines the *cylinder* $C_{\omega^n} = \{\xi \in \Sigma : \xi^n = \omega^n\}$. We denote by σ the left shift map acting on Σ, i.e. a map such that $(\sigma\omega)_n = \omega_{n+1}$. Finally, $|E|$ stands for the diameter of $E \subset X$, whereas $B(x, r)$ denotes the closed ball in X with center at x and radius $r > 0$.

Suppose we are given Borel measurable maps $S_i : X \to X, i \in I$, and Borel measurable functions $p_i : X \to [0, 1], i \in I$, such that $\sum_I p_i \equiv 1$. Then we call the triple $\{X, S_i, p_i\}$ a *(measurable) iterated function system*. The functions p_i are called *probabilities*.

Iterated function systems are usually studied by means of the corresponding Markov chains. Generally, if we want to define a discrete-time Markov chain, we can start with fixing a *transition probability function* (t.p.f.) $P : X \times \mathcal{B}(X) \to [0, 1]$, i.e. a function such that $P(x, \cdot)$ is a probability measure for each $x \in X$ and $P(\cdot, A)$ is a Borel measurable function for each $A \in \mathcal{B}(X)$ (by $\mathcal{B}(Y)$ we denote the family of all Borel subsets of a metric space Y). For example, for a given iterated function system $\{X, S_i, p_i\}$, let us consider the

function

$$P : X \times \mathcal{B}(X) \ni (x, A) \mapsto \sum_I p_i(x) 1_A(S_i(x)).$$

Clearly it is a t.p.f.—we say that it *corresponds* to $\{X, S_i, p_i\}$.

Having given a t.p.f. P, we can look at it "dynamically": to a fixed $x \in X$ we assign another point, choosing it randomly—according to the distribution $P(x, \cdot)$. To better understand this action, it is convenient to think not about individual points, but about their distributions. This leads to the definition of a *Markov operator* corresponding to P, acting on the set $\mathcal{M}$ of all finite Borel measures on X via the formula

$$\mu P(A) = \int_X P(x, A) \, \mu(dx) \quad \text{for } A \in \mathcal{B}(X), \mu \in \mathcal{M}.$$

This operator transforms the set $\mathcal{M}_1 = \{\mu \in \mathcal{M} : \|\mu\| = 1\}$ of distributions into itself ($\| \cdot \|$ denotes here the total variation norm).

If the action defined above has no influence on a measure $\mu \in \mathcal{M}$, i.e. if $\mu P = \mu$, then μ is said to be *invariant* under P. If additionally $\mu \in \mathcal{M}_1$ is an extremal point (in $\mathcal{M}$) of the set of distributions invariant under P then it is called *ergodic*. Later we use a convenient characterization of ergodic measures in terms of invariant sets—we call a set $A \in \mathcal{B}(X)$ μ-*invariant* provided $P(\cdot, A) = 1_A$ μ-a.e., where μ is a given invariant measure.

Let us now come back to the situation when P corresponds to an iterated function system $\{X, S_i, p_i\}$. Obviously $\mu P = \sum_I S_{i*}(p_i \mu)$, where $T_*(fv)(A) = \int f \cdot 1_A \circ T \, dv$ for $v \in \mathcal{M}$, $f \in \mathcal{B}$, $A \in \mathcal{B}(X)$ and $T : X \to X$ is a Borel measurable map. (By $\mathcal{B}$ we denote the space of all bounded Borel measurable functions on X.) We will say that a measure is ergodic or a measure/a set is invariant under $\{X, S_i, p_i\}$ if P has the appropriate property.

Now we assume that $\{X, S_i, p_i\}$ is the iterated function system with an invariant measure μ. We are going to construct a measure-preserving transformation which corresponds to the initial system and has similar properties. For any $x \in X, \omega^n \in I^n$, let

$$p_{\omega^n}(x) = p_{\omega_1}(x) \cdot p_{\omega_2}(S_{\omega_1}(x)) \cdot \ldots \cdot p_{\omega_n}(S_{\omega^{n-1}}(x))$$

and

$$S_{\omega^n} = S_{\omega_n} \circ \ldots \circ S_{\omega_1}$$

(we also put $S_{\omega^0} \equiv S_{\omega_0} \equiv \text{id}_X$, $p_{\omega_0} \equiv 1$, treating ω_0 as an empty sequence). Moreover let p_x be a family of probability measures on Σ, defined on cylinders in the following way

$$p_x(C_{\omega^n}) = p_{\omega^n}(x) \quad \text{for } \omega^n \in I^n, x \in X.$$

Measures p_x can be, in turn, used to define the probability measure v on $X \times \Sigma$ by the formula

$$v(d(x, \omega)) = p_x(d\omega)\mu(dx). \tag{2.1}$$

This last measure, v, is invariant under the skew product S acting on $X \times \Sigma$ as follows

$$S : (x, \omega) \mapsto (S_{\omega_1}(x), \sigma\omega).$$

It is clear that the properties of the aforementioned invariant measures are related. Later we show equivalence of their ergodicity—to do this we need some simple though useful facts.

Let us fix a t.p.f. P.

Lemma 3 *Assume that $\mu, \mu_1 \in \mathcal{M}_1$ are invariant under P, $\mu_1 \ll \mu$ and the following condition holds*

$$\mu(A) = 0 \quad or \quad \mu(A) = 1 \text{ for any } A \in \mathcal{B}(X) \text{ } \mu\text{-invariant under } P. \qquad (2.2)$$

Then $\mu_1 = \mu$.

Proof Suppose the lemma is false and set $\tau = \mu - \mu_1 \neq 0$. The minimum property of the Jordan decomposition $\tau = \tau^+ - \tau^-$ into nonnegative measures τ^+, τ^- implies that $\tau^+ \leq \tau^+ P$ and $\tau^- \leq \tau^- P$. Hence and because P preserves the total variation norm, the measures τ^+, τ^- are invariant under P. Moreover, according to the Hahn decomposition theorem, there exist disjoint sets $X^+, X^- \in \mathcal{B}(X)$, $X^+ \cup X^- = X$, on which the measures τ^+, τ^- (respectively) are concentrated. Now put $X_0 = X^+$, $X_m = \{x \in X_{m-1} : P1_{X_{m-1}}(x) = 1\}$ for $m \in \mathbb{N}$ and $A = \bigcap_m X_m$. It is easy to see that $1_A \leq P(\cdot, A)$—therefore A is μ-invariant. Furthermore, $\tau^+(X) = \tau(X^+)$, $\tau^-(X) = -\tau(X^-)$. As $\tau \neq 0$, both these numbers are positive and so are $\mu(A), \mu(X \setminus A)$—the latter is true since $\mu_1 \ll \mu$. This contradicts (2.2). □

Corollary 4 *If $\mu \in \mathcal{M}_1$ is invariant under P, then μ is ergodic iff the condition (2.2) holds.*

Proof Sufficiency of (2.2) follows from Lemma 3 whereas necessity is implied by Lemma 1 from [2]. □

Assume now that $\{X, S_i, p_i\}$ is an iterated function system with an invariant distribution μ and the distribution ν is defined by (2.1).

Lemma 5 *The measure μ is ergodic w.r.t. $\{S_i, p_i, X\}$ iff the measure ν is ergodic for $(X \times \Sigma, S)$.*

Proof By Corollary 4 it suffices to prove that μ satisfies (2.2) iff ν is ergodic. Sufficiency is obvious, so we turn to necessity. Suppose ν is not ergodic, i.e. there exists a set $D \in \mathcal{B}(X \times \Sigma)$ which is S-invariant and such that

$$\nu(D) > 0, \qquad \nu(D^c) > 0. \qquad (2.3)$$

For any $x \in X$, $A \in \mathcal{B}(X \times \Sigma)$ let

$$A_x = \{\omega \in \Sigma : (x, \omega) \in A\}, \qquad L_A(x) = p_x(A_x),$$
$$A^X = \{y \in X : L_A(y) > 0\}.$$

Clearly $A_x \in \mathcal{B}(\Sigma)$, $A^X \in \mathcal{B}(X)$; the latter is true because $L_A \in \mathcal{B}$. Indeed, if we set $\mathcal{L} = \{E \in \mathcal{B}(X \times \Sigma) : L_E \in \mathcal{B}\}$, then $\mathcal{L}$ is a λ-system containing the π-system $\mathcal{P}$ of Borel measurable rectangles, subsets of $X \times \Sigma$. By the Sierpiński-Dynkin theorem on π-λ-systems, $\mathcal{L} \supset \sigma(\mathcal{P}) = \mathcal{B}(X \times \Sigma)$.

From the S-invariance of D it follows that

$$\sum_I p_i(x) L_D(S_i(x)) = L_D(x) \quad \text{for every } x \in X, \qquad (2.4)$$

which, in turn, implies that $P(\cdot, D^X) \leq 1_{D^X}$. Thus $\mu(D^X) = 1$; similarly $\mu((D^c)^X) = 1$. Therefore there exists a set $\tilde{X} \in \mathcal{B}(X)$ of full measure μ such that $(D^c)_x = (D_x)^c \neq \emptyset$ for

every $x \in \tilde{X}$. Consequently, $L_D + L_{D^c}$ is μ-a.e. equal to 1. We are going to examine properties of L_D more precisely.

Claim 1. L_D is μ-a.e. constant.

Set $X_l = L_D^{-1}(l, \infty)$, where $l > 0$ is such that $\mu(X_l) > 0$. Consider the measure $\tilde{\mu}$ define by the formula: $\tilde{\mu}(E) = \mu(E \cap X_l)$ for $E \in \mathcal{B}(X)$. We are going to prove that $\tilde{\mu}$ is invariant. Since μ is so, $\tilde{\mu}P \ll \mu$. Furthermore, putting

$$g = \frac{d\tilde{\mu}}{d\mu} - \frac{d\tilde{\mu}P}{d\mu}$$

we get $(L_D - l)g \geq 0$ μ-a.e. and $\int g \, d\mu = 0$. Hence

$$\int_X L_D g \, d\mu \geq 0$$

with equality iff g μ-a.e. equal to 0. But

$$\int_X L_D g \, d\mu = \int_X L_D \, d\tilde{\mu} - \int_X L_D \, d\tilde{\mu}P = 0,$$

where the last equality is a consequence of (2.4). Thus $\tilde{\mu}$ is an invariant measure absolutely continuous w.r.t. μ. Lemma 3 gives $\mu = \tilde{\mu}$, which proves our claim.

To finish the proof of Lemma 5 it suffices to justify

Claim 2. The following disjunction holds

$$L_D = 1 \; \mu\text{-a.e.} \quad \text{or} \quad L_{D^c} = 1 \; \mu\text{-a.e.} \tag{2.5}$$

Assume the contrary and fix an $\varepsilon > 0$ so small that the sets $\{L_D > 1 - \varepsilon\}$ and $\{L_{D^c} > 1 - \varepsilon\}$ have both zero measure μ. Put $\Lambda(A, \omega^n) = \{x \in X : p_x(A_x \cap C_{\omega^n}) > (1 - \varepsilon)p_x(C_{\omega^n})\}$ for any $n \in \mathbb{N}$, $\omega^n \in I^n$ and $A \subset X \times \Sigma$. We are going to show that

$$\tilde{X} \subset \bigcup_{Z \in \mathcal{F}} Z, \tag{2.6}$$

where $\mathcal{F} = \{\Lambda(A, \omega^n) : n \in \mathbb{N}, \omega^n \in I^n, A \in \{D, D^c\}\}$.

Pick any $x \in \tilde{X}$ then. Two cases may occur: either p_x is nonatomic or it has at least one atom. In the second case that atom, let us call it ω, may be a member of D_x (then we conclude that $x \in \Lambda(D, \omega^n)$ for some $n \in \mathbb{N}$) or it may happen that $\omega \in (D^c)_x$ (in this case there exists $n \in \mathbb{N}$ such that $x \in \Lambda(D^c, \omega^n)$).

Now suppose $x \in \tilde{X}$ is such that p_x has no atoms. By means of the "Cantor function"-type construction it is easy to build a measure-preserving homeomorphism between the spaces $(\Sigma, \mathcal{B}(\Sigma), p_x)$ and $([0, 1], \mathcal{B}([0, 1]), \lambda)$, which transforms cylinders into some intervals. This and the Lebesgue theorem on density points applied to D_x imply the existence of $\omega^n \in I^n$ such that $x \in \Lambda(D, \omega^n)$. The proof of (2.6) is finished.

To make use of (2.6) we notice that the family $\mathcal{F} \subset \mathcal{B}(X)$ is countable and each member of $\mathcal{F}$ has zero measure μ. It is so since, by the definition of p_x and S-invariance of sets D, D^c,

$$\Lambda(A, \omega^n) \subset S_{\omega^n}^{-1}(\{L_A > 1 - \varepsilon\}) \quad \text{for } A \in \{D, D^c\}, n \in \mathbb{N}, \omega^n \in I^n.$$

But this implies the equality $\mu(\tilde{X}) = 0$ which contradicts the way we chose the set $\tilde{X}$. $\qquad\square$

On account of Lemma 5, the Egorov theorem and the Birkhoff ergodic theorem, we have an immediate corollary:

Corollary 6 *Let μ be ergodic. Then for every $\varepsilon > 0$ and for any family of Borel measurable functions $\{h_i : X \to \mathbb{R}\}_I$ satisfying the inequalities*

$$-\infty < \int_X \sum_I p_i h_i \, d\mu < 0$$

there exist $K > 0$, $A_K \in \mathcal{B}(X \times \Sigma)$ such that $\nu(A_K) < \varepsilon$ and

$$\sum_{j=1}^n h_{\omega_j}(S_{\omega^{j-1}}(x)) < K$$

for all $(x, \omega) \in A_K{}^c$ and all $n \in \mathbb{N}$.

Let us go back to Example 2. Let μ be any invariant measure for (X, S_i, p_i). As $\mu P^n = \mu$, $\mu([k \cdot 3^{-n}, (k + 1) \cdot 3^{-n})) \leq (1 - p)^n \searrow 0$, μ cannot have atoms. It implies that $\mu(B_\delta(\partial A)) \searrow 0$ as $\delta \searrow 0$.

The function integrated in the definition of $h_N(\mu, \delta)$ equals $-N \sum_{i=1}^2 p_i(x) \log p_i(x) \times \mu(dx)$ for all x whose all trajectories avoid $B_\delta(\partial A)$ for time N and is bounded by $-N \log p$ everywhere. Hence, by Lebesgue majorized convergence theorem

$$\frac{1}{N} h_N(\mu) = -\int_X \sum_{i=1}^2 p_i(x) \log p_i(x) \mu(dx) = -p \log p - (1 - p) \log(1 - p)$$

At the same time, $\lambda_N(\mu, \delta) = -N \log 3$ for all N and δ and (1.2) follows.

The one thing remaining to check is that (X, S_i, p_i) from Example 2 has any invariant measures at all. Let μ_0 be any probabilistic measure on X and define

$$\mu_n = \frac{1}{n} \sum_{m=0}^{n-1} \mu_0 P^m$$

As μ_n form a sequence of probabilistic measures on a compact space, they have a subsequence μ_{n_k} convergent in law to some measure μ. Let us fix $\varepsilon > 0$. As $\mu_0 P^n([k \cdot 3^{-m}, (k + 1) \cdot 3^{-m})) \leq (1 - p)^m$ for all $n > m$, the same is true for μ. It follows that μ cannot have atoms, hence $\mu(B_r(\partial A)) \leq \frac{3}{13}\varepsilon$ for r small enough. As μ_{n_k} converges to μ,

$$\mu_{n_k}(B_r(\partial A) \leq \frac{3}{13}\varepsilon \tag{2.7}$$

and

$$\sum_{D_i} |\mu_{n_k}(D_i) - \mu(D_i)| \leq \frac{3}{13}\varepsilon \tag{2.8}$$

for k big enough (where the sum is taken over the components of the complement of $B_r(\partial A)$). We will prove that $\mu P = \mu$.

Let d_{FM} be the Fortet-Mourier metric [4] on the space of finite measures, defined as

$$d_{FM}(\mu, \nu) = \sup \left| \int_X f d\mu - \int_X f d\nu \right|$$

where the supremum is taken over Lipschitz functions with Lipschitz constant 1 and with absolute value bounded by 1. It is well known that the topology defined by Fortet-Mourier metric is equivalent to the usual weak* topology, see [1].

As μ_{n_k} converge to μ, we have

$$d_{FM}(\mu_{n_k}, \mu) \le \frac{3}{13} \varepsilon \qquad (2.9)$$

for k big enough.

By the definition of μ_{n_k} we have

$$d_{FM}(\mu_{n_k}, \mu_{n_k} P) = \frac{1}{n_k} d_{FM}(\mu_0, \mu_0 P^{n_k}) \le \frac{1}{n_k} \le \frac{3}{13} \varepsilon \qquad (2.10)$$

for k big enough.

We can write

$$d_{FM}(\mu_{n_k} P, \mu P) \le d_{FM}\left((\chi_{B_r(\partial A)} \mu_{n_k}) P, (\chi_{B_r(\partial A)} \mu) P \right) + \sum_{D_i} d_{FM}\left((\chi_{D_i} \mu_{n_k}) P, (\chi_{D_i} \mu) P \right)$$

As the Fortet-Mourier distance of two measures cannot be greater than the sum of their L^1-norms, by (2.7)

$$\max(\mu_{n_k}(B_r(\partial A)), \mu(B_r(\partial A))) \le \frac{3}{13} \varepsilon$$

To estimate the following sum note that on each D_i the iterated function system acts as two linear contracting maps with contraction coefficients $1/3$, one chosen with fixed probability p and the other $(1 - p)$. Hence, the Fortet-Mourier distance of images of two measures is bounded from above by $1/3$ of the Fortet-Mourier distance of the original measures plus the difference of L^1-norms of the original measures. Summing over D_i and applying (2.8) we get

$$\sum_{D_i} d_{FM}((\chi_{D_i} \mu_{n_k}) P, (\chi_{D_i} \mu) P) \le \frac{1}{3} d_{FM}(\mu_{n_k}, \mu) + \sum_{D_i} |\mu_{n_k}(D_i) - \mu(D_i)| \le \frac{4}{13} \varepsilon$$

Hence,

$$d_{FM}(\mu_{n_k} P, \mu P) \le \frac{7}{13} \varepsilon$$

Applying (2.9) and (2.10) we get

$$d_{FM}(\mu, \mu P) \le \varepsilon$$

As ε was arbitrary, $\mu = \mu P$.

3 The Proof of Theorem 1

First we would like to clear up some simple case. Namely, notice that w.l.o.g. we may assume that X do not contain any isolated points. Indeed, every isolated point is either of zero measure μ and therefore can be removed from the space (without changing $\dim_H(\mu)$) or has it positive. In the second case μ is concentrated on a finite set and consequently $\dim_H(\mu) = 0$.

Lemma 7 *If $h_N(\mu, \delta)$ and $\lambda_N(\mu, \delta)$ exist, they are subadditive in N.*

Proof Both $h_N(\mu, \delta)$ and $\lambda_N(\mu, \delta)$ can be written in the form

$$\int_X \sum_{\omega^N \in I^N} p_{\omega^N}(x) \sup_{y \in B_N(x, \omega^N, \delta)} \phi(x, y, \omega^N) \, \mu(dx),$$

where

$$\phi(x, y, \omega^N) = \sum_{n=1}^N \phi(S_{\omega^{n-1}}(x), S_{\omega^{n-1}}(y), \omega_n)$$

is some real-valued function. Given N_1 and N_2, for any fixed δ we have $y \in B_{N_1+N_2}(x, \omega^{N_1+N_2}, \delta)$ if and only if $y \in B_{N_1}(x, \omega^{N_1}, \delta)$ and

$$S_{\omega^{N_1}}(y) \in B_{N_1}(S_{\omega^{N_1}}(x), \omega_{N_1+1} \ldots \omega_{N_1+N_2}, \delta).$$

Hence,

$$\sup_{y \in B_{N_1+N_2}(x, \omega^{N_1+N_2}, \delta)} \phi(x, y, \omega^{N_1+N_2})$$

$$\leq \sup_{y \in B_{N_1}(x, \omega^{N_1}, \delta)} \phi(x, y, \omega^{N_1})$$

$$+ \sup_{y \in B_{N_2}(S_{\omega^{N_1}}(x), \omega_{N_1+1} \ldots \omega_{N_1+N_2}, \delta)} \phi(S_{\omega^{N_1}}(x), y, \omega_{N_1+1} \ldots \omega_{N_1+N_2})$$

As $\mu = \mu P^{N_1}$, $h_{N_1+N_2}(\mu, \delta) \leq h_{N_1}(\mu, \delta) + h_{N_2}(\mu, \delta)$ for every $\delta > 0$ (and analogously for $\lambda_{N_1+N_2}(\mu, \delta)$). $\qquad \square$

Now let us present the idea of the proof of Theorem 1. We need to prove that $\dim_H(\mu) \leq h_N(\mu, \delta)/\lambda_N(\mu, \delta)$. We will only give the detailed proof for $N = 1$, for higher N one works with P^N instead of P and the proof is almost identical.

We are going to analyze a family of measures $\{\mu_{j, \omega^n}\}$, which sum up to μ and are associated with a certain finite partition $\{E_j\}$ of X. Every measure μ_{j, ω^n} is concentrated on $B(S_{\omega^n}(e_j), |S_{\omega^n}(E_j)|)$ (where $e_j \in E_j$)—typically a set of small diameter. We will choose some of the pairs (j, ω^n) in such a way that the union of sets $B(S_{\omega^n}(e_j), |S_{\omega^n}(E_j)|)$ corresponding to the chosen pairs will be both of big measure μ and geometrically small (see Lemma 8). We will use these balls to estimate the Hausdorff dimension of μ.

We need some additional notations: for $i \in I, x \in X, \vartheta < 0, \delta > 0, m \in \mathbb{N}, \omega \in \Sigma$ we will write

$$L_i^{\delta,\vartheta}(x) = \max\left\{\sup_{\substack{y \in B(x,\delta) \\ y \neq x}} \log \frac{\varrho(S_i(x), S_i(y))}{\varrho(x, y)}, \vartheta\right\}$$

$$H_i^{\delta}(x) = \inf_{y \in B(x,\delta)} \log p_i(y),$$

$$L_m^{\delta,\vartheta}(x, \omega) = \frac{1}{m} \sum_{k=0}^{m-1} L_{\omega_{k+1}}^{\delta,\vartheta}(S_{\omega^k}(x)),$$

$$H_m^{\delta}(x, \omega) = \frac{1}{m} \sum_{k=0}^{m-1} H_{\omega_{k+1}}^{\delta}(S_{\omega^k}(x)),$$

$$I_L^{\delta,\vartheta} = \int_X \sum_I p_i L_i^{\delta,\vartheta} \, d\mu,$$

$$I_H^{\delta} = \int_X \sum_I p_i H_i^{\delta} \, d\mu,$$

$$s(\delta, \vartheta) = I_H^{\delta} / I_L^{\delta,\vartheta}.$$

Note that inequality (1.1) holds if $I_H^{\delta} = -\infty$ for each $\delta > 0$. So, since I_H^{δ} is a nonincreasing function of δ, w.l.o.g. we can assume that there is $\Delta \in (0, 1)$ such that for every $\delta \in (0, \Delta)$ the integral I_H^{δ} is finite and, at the same time, the corresponding integral from the denominator of the right-hand side of (1.1) is negative (see comments in the first section) or equal to $-\infty$. In the latter case there exists a number $\Theta < 0$ such that $I_L^{\Delta,\vartheta} \in (-\infty, 0)$ for all $\vartheta \in (-\infty, \Theta]$. In the former one we set $\Theta = -\infty$—that constant would play no role in the proof then.

Obviously s (considered on $(0, \Delta) \times [-\infty, \Theta]$) is a nondecreasing function of every variable with another one fixed. Moreover, the limit of s, taken as $(\delta, \vartheta) \to (0, -\infty)$, is equal to the right-hand side of (1.1).

After making these introductory remarks we can start the proof. First we choose (δ, ϑ) from the domain of s. Next we fix $s > s(\delta, \vartheta)$ and $\iota \in \mathbb{N}$ in an arbitrary way. We also pick $\varepsilon \in (0, 1/\iota)$ such that $s > s(\delta, \vartheta)(1 + \varepsilon)/(1 - \varepsilon)$ and then we apply Corollary 6 for ε and $\{L_i^{\delta,\vartheta}\}_I$—we are allowed since $I_L^{\delta,\vartheta} \in (-\infty, 0)$. As a result we obtain K, A_K.

Let $\{E_j\}_J$ be a finite family of nonempty disjoint Borel subsets of X, of diameter at most $\delta e^{-K}/4$ and such that

$$\mu\left(\bigcup_J E_j\right) > 1 - \varepsilon,$$

where $J \subset \mathbb{N}$. We add to this family $E_0 = X \setminus \bigcup_J E_j$ to form a partition of X. Then we pick $n \in \mathbb{N}$ big enough ($n > n_0$) to apply Lemma 8 (see below) and such that

$$\sharp J (\delta e^{-K})^s \exp\{n(s(1 - \varepsilon)I_L^{\delta,\vartheta} - (1 + \varepsilon)I_H^{\delta})\} < 2^{-\iota}. \tag{3.1}$$

Before we formulate this Lemma, we notice that

$$\mu = \sum_{\{J \cup \{0\}\} \times I^m} \mu_{j,\omega^m},$$

where

$$\mu_{j,\omega^m} = S_{\omega^m *}(p_{\omega^m}\mu|_{E_j})$$

and $m \in \mathbb{N} \cup \{0\}$. If, for any $(j, \omega^m) \in \{J \cup \{0\}\} \times I^m, i \in I$, we put

$$p_i(j, \omega^m) = \frac{1}{\|\mu_{j,\omega^m}\|} \int_X p_i \, d\mu_{j,\omega^m}$$

then

$$\|\mu_{j,\omega^m i}\| = p_i(j, \omega^m)\|\mu_{j,\omega^m}\|;$$

moreover $p_i(j, \omega^m) > 0$, as it is an average of p_i under some measure.

Lemma 8 *There exists $n_0 \in \mathbb{N}$ such that for every $n > n_0$ from the set $\{J \cup \{0\}\} \times I^n$ one can choose a subset $Z(\iota)$ satisfying the following conditions*:

  i) *for $(j, \omega^n) \in Z(\iota)$ and $m \leq n$, $|S_{\omega^m}(E_j)| \leq \delta$,*
  ii) *$|S_{\omega^n}(E_j)| \leq \delta e^{-K+n(1-\varepsilon)I_L^{\delta,\vartheta}}/2$ for $(j, \omega^n) \in Z(\iota)$,*
  iii) *$Z(\iota)$ has at most $\sharp J e^{-n(1+\varepsilon)I_H^{\delta}}$ elements,*
  iv) *$\mu(\bigcup_{Z(\iota)} B(S_{\omega^n}(e_j), |S_{\omega^n}(E_j)|)) \geq 1 - 4\varepsilon$ provided $e_j \in E_j$ for all $j \in J$.*

Proof Fix points $e_j \in E_j$. Since μ is invariant, we get

$$\mu\left(\bigcup_{Z(\iota)} B\big(S_{\omega^n}(e_j), \big|S_{\omega^n}(E_j)\big|\big)\right) \geq 1 - \sum_{Z(\iota)^c} \big\|\mu_{j,\omega^n}\big\|, \tag{3.2}$$

regardless of n and $Z(\iota) \subset \{J \cup \{0\}\} \times I^n$. Hence, we only need to estimate the sum of $\|\mu_{j,\omega^n}\|$ over pairs (j, ω^n) for which i), ii) or iii) does not hold.

First we are going to establish how big n_0 should be. Lemma 5 and the Birkhoff theorem show that $L_m^{\delta,\vartheta}(x, \omega)$ converges to $I_L^{\delta,\vartheta}$ and $H_m^{\delta}(x, \omega)$ converges to I_H^{δ} for μ-almost every $x \in X$ and p_x-almost every $\omega \in \Sigma$. By the Egorov theorem we can choose a big (w.r.t. μ) subset $X_0 \in \mathcal{B}(X)$ for which those convergences are uniform for a big set of ω's. Specifying, we have $n_0 \in \mathbb{N}$, $X_0 \in \mathcal{B}(X)$ such that

$$\mu(X_0) \geq 1 - \varepsilon \tag{3.3}$$

and for every $x \in X_0$, the set of ω which does not satisfy

$$|L_n^{\delta,\vartheta}(x, \omega) - I_L^{\delta,\vartheta}| + |H_n^{\delta}(x, \omega) - I_H^{\delta}| \leq -\varepsilon \max\{I_L^{\delta,\vartheta}, I_H^{\delta}\} \tag{3.4}$$

for some $n > n_0$, has p_x-measure not greater than ε.

Now we are going to construct $Z(\iota)$. We fix $n > n_0$. Notice that $Z(\iota)$ cannot contain any element of the form $(0, \omega^n)$. Let us denote the set of all such pairs by A_1. We obtain

$$\sum_{A_1} \|\mu_{j,\omega^n}\| = \mu(E_0) < \varepsilon. \tag{3.5}$$

Next, we want to exclude from $Z(\iota)$ the set A_2 of all the pairs $(j, \omega^n) \in A_1^c$ for which i) does not hold. So we take $(j, \omega^n) \in A_2$, i.e. such that

$$|S_{\omega^m}(E_j)| > \delta$$

for some $m \leq n$, but

$$|S_{\omega^u}(E_j)| \leq \delta$$

for all $u < m$. We have

$$\delta < \sup_{y \in E_j} \varrho(S_{\omega^m}(x), S_{\omega^m}(y))$$

for some $x \in E_j$. Furthermore

$$\sup_{y \in E_j} \varrho(S_{\omega^m}(x), S_{\omega^m}(y)) \leq |E_j| \prod_{u=0}^{m-1} \sup_{y \in S_{\omega^u}^{-1}(S_{\omega^u}\{x\})^c \cap E_j} \frac{\varrho(S_{\omega^{u+1}}(x), S_{\omega^{u+1}}(y))}{\varrho(S_{\omega^u}(x), S_{\omega^u}(y))}$$

$$\leq |E_j| \exp(m L_m^{\delta, \vartheta}(x, \omega)),$$

where ω is any continuation of the finite sequence ω^n. Thus $E_j \times C_{\omega^n} \subset A_K$, and so

$$\sum_{A_2} \|\mu_{j,\omega^n}\| < \varepsilon. \tag{3.6}$$

The last set to exclude is the one containing the elements for which we cannot make use of the choice of n_0. We distinguish the set A_3 of pairs (j, ω^n) with j such that E_j does not intersect X_0; here, by (3.3), we have

$$\sum_{A_3} \|\mu_{j,\omega^n}\| \leq \varepsilon. \tag{3.7}$$

Next, for every $(j, \omega^n) \in A_3^c$ let us choose a point $x_j \in E_j \cap X_0$. Denote by A_4 the set of pairs $(j, \omega^n) \in A_3^c$ such that (3.4) does not hold for $(x, \omega) \in \{x_j\} \times C_{\omega^n}$ and n. Again,

$$\sum_{A_4} \|\mu_{j,\omega^n}\| \leq \varepsilon. \tag{3.8}$$

Let $Z(\iota) = \bigcap_{t=1}^{4} A_t^c$. Since i) and (3.4) hold for $(j, \omega^n) \in Z(\iota)$, we obtain

$$\prod_{k=0}^{n-1} p_{\omega_{k+1}}(j; \omega^k) \geq \prod_{k=0}^{n-1} \inf_{y \in B(S_{\omega^k}(x_j), |S_{\omega^k}(E_j)|)} p_{\omega_{k+1}}(y) \geq e^{n(1+\varepsilon)I_H^\delta} \tag{3.9}$$

and

$$\left| S_{\omega^n}(E_j) \right| \leq 2|E_j| \prod_{k=0}^{n-1} \sup_{\substack{y \in B(S_{\omega^k}(x_j), |S_{\omega^k}(E_j)|) \\ y \neq S_{\omega^k}(x_j)}} \frac{\varrho(S_{\omega_{k+1}}(x_j), S_{\omega_{k+1}}(y))}{\varrho(S_{\omega^k}(x_j), y)}$$

which implies that

$$\left| S_{\omega^n}(E_j) \right| \leq \delta e^{-K + n(1-\varepsilon)I_L^{\delta, \vartheta}}/2. \tag{3.10}$$

By (3.10), ii) is satisfied. Similarly, as $\sum_{\omega^n} \prod_{k=0}^{n-1} p_{\omega_{k+1}}(j; \omega^k) = 1$ for every $j \in J$, (3.9) implies iii). The assertion iv) follows from (3.5), (3.6), (3.7), (3.8) and (3.2). We are done. $\square$

The rest of the proof is standard. Let $e_j \in E_j$ be any points. The set

$$Y_\iota = \bigcup_{Z(\iota)} B\left(S_{\omega^n}(e_j), \left|S_{\omega^n}(E_j)\right|\right)$$

has measure μ at least $1 - 4\varepsilon$. At the same time it can be covered with a family

$$\left\{B\left(S_{\omega^n}(e_j), |S_{\omega^n}(E_j)|\right)\right\}_{(j,\omega^n)\in Z(\iota)} \tag{3.11}$$

of at most $\sharp J e^{-n(1+\varepsilon)I_H^\delta}$ sets of diameter less than $\delta e^{-K+n(1-\varepsilon)I_L^{\delta,\vartheta}}$.

The set

$$Y = \bigcap_\iota \bigcup_{\kappa > \iota} Y_\kappa$$

has full measure μ. At the same time Y has zero s-dimensional Hausdorff measure, because $Y \subset \bigcup_{\kappa > \iota} Y_\kappa$ for every ι, the diameter of covers (3.11) converges to 0 as $\iota \to \infty$ and (3.1) yields the inequality

$$\sum_{\kappa > \iota} \sum_{(j,\omega^n) \in Z(\kappa)} \left|B\left(S_{\omega^n}(e_j), \left|S_{\omega^n}(E_j)\right|\right)\right|^s < 2^{-\iota}.$$

Since $s > s(\delta, \vartheta)$ was arbitrary, the Hausdorff dimension of μ is not greater than $s(\delta, \vartheta)$. As $-\vartheta > 0$ can be chosen arbitrarily big, we get the assertion.

References

1. Dudley, R.M.: Probabilities and Metrics. Aarhus Universitet, Aarhus (1976)
2. Elton, J.H.: An ergodic theorem for iterated maps. Ergod. Theory Dyn. Syst. **7**, 481–488 (1987)
3. Fan, A.H., Simon, K., Toth, H.R.: Contracting on average random IFS with repelling fixpoint. J. Stat. Phys. **122**, 169–193 (2006)
4. Fortet, R., Mourier, B.: Convergence de la répartition empirique vers la répartition théorétique. Ann. Sci. Éc. Norm. Super. **70**, 267–285 (1953)
5. Johansson, A., Öberg, A.: Exact dimension of Cantor type measures generated by iterated function systems. Preprint
6. Myjak, J., Szarek, T.: On Hausdorff dimension of invariant measures arising from non-contractive iterated function systems. Ann. Mat. Pura Appl. **181**, 223–237 (2002)
7. Nicol, M., Sidorov, N., Broomhead, D.: On the fine structure of stationary measures in systems which contract-on-average. J. Theor. Probab. **15**, 715–730 (2002)
8. Rams, M.: Dimension estimates for invariant measures of contract ing-on-average iterated function systems. Preprint
9. Szarek, T.: The dimension of self-similar measures. Bull. Pol. Acad. Sci. Math. **48**, 293–302 (2000)
10. Werner, I.: Contractive Markov systems. J. Lond. Math. Soc. **71**(2), 236–258 (2005)