

Fat baker's transformations

J. C. ALEXANDER AND J. A. YORKE

Department of Mathematics, University of Maryland, College Park, MD 20742, USA;
Institute for Physical Science and Technology, University of Maryland, College Park,
MD 20742, USA

(Received 5 May 1982 and revised 2 December 1982)

Abstract. We investigate a variant of the baker's transformation in which the mapping is onto but is not one-to-one. The Bowen–Ruelle measure for this map is investigated.

1. Introduction

Baker's transformations. Consider the strip

$$\mathbb{R} \times [-1, 1] = \{(x, y): -\infty < x < \infty, -1 \leq y \leq 1\}$$

and for $0 < \beta < 1$, consider the transformation T_β of this strip:

$$T_\beta(x, y) = \begin{cases} (\beta x + (1 - \beta), 2y - 1) & \text{if } y \geq 0, \\ (\beta x - (1 - \beta), 2y + 1) & \text{if } y < 0. \end{cases}$$

Note that the square $[-1, 1] \times [-1, 1]$ is invariant and that the iterates of any (x, y) are either eventually in the square or (for a countable set of y) approach the square. Thus the attractor of T_β is contained in the square. Our investigations bear mostly on the structure of the attractor; we use the same notation for the restriction of T_β to the square. The x -axis is a set of measure zero and the definition of T_β for $y = 0$ can be altered without affecting any results. In other words, T_β vertically stretches the top (resp. bottom) half of the square to the full height of the square, shrinks it by a factor β in the horizontal direction, and shifts the resulting rectangle horizontally to the right (resp. left) as far as possible in the square.

If $\beta = \frac{1}{2}$, the transformation T_β is the classic baker's transformation. Accordingly we call T_β a *generalized baker's transformation*. A generalized baker's transformation is piecewise linear, with expansion by a factor of 2 in the vertical direction and compression by a factor β in the horizontal direction. If $\beta < \frac{1}{2}$ we call T_β a *skinny baker's transformation*; a skinny baker's transformation on the square is one-to-one and its structure is fairly easy to understand. In particular the attractor is the product of the vertical line $[-1, 1]$ with a Cantor set in the x -direction. On the other hand, if $\beta > \frac{1}{2}$, the transformation T_β is not one-to-one and its structure is considerably more interesting. This is the case on which we concentrate in this paper. We call T_β for $\beta > \frac{1}{2}$ a *fat baker's transformation*. The attractor of a fat baker's transformation is the whole square. We find there is (at least) a countable closed set of β , $\frac{1}{2} < \beta < 1$ for which every invariant measure is totally singular with respect to Lebesgue measure and for which the attractor has an information-theoretic dimension (defined below) strictly less than two.

This study is part of a larger programme to understand relations between the dynamics of transformations defined on manifolds and certain measure and information-theoretic aspects of their attractors. Generalized baker's transformations are simple enough that explicit computations are possible, yet they exhibit behaviour commonly seen in transformations with strange attractors.

Bowen–Ruelle measures. The stochastic behaviour of the transformation is captured by a measure first studied in [5] (see also [4]) for axiom A systems. Let M be a manifold, and $T: M \rightarrow M$ a continuous transformation. Suppose there is a set $M_0 \subset M$ of Lebesgue measure zero such that for all $x \in M \setminus M_0$ and for every continuous $f: M \rightarrow \mathbb{R}$ the limit

$$\bar{f} = \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=0}^N f(T^n(x))$$

exists and is independent of $x \in M \setminus M_0$. Then $f \rightarrow \bar{f}$ is a continuous linear functional on the space of bounded continuous functions with the sup-norm topology and thus determines a Borel measure μ such that

$$\bar{f} = \int_M f d\mu.$$

The measure μ is automatically invariant; we call it the Bowen–Ruelle measure. If T is also ergodic with respect to the Bowen–Ruelle measure, we say T is Bowen–Ruelle ergodic. Note that the Bowen–Ruelle measure is supported on the attractor of T . The Bowen–Ruelle measure is the one ‘seen’ on a computer simulation of T . If the computer selects a random point $x \in M$ and plots the iterates of x , the result is a study in shades of grey. The deepness of the grey indicates the concentration of μ .

In § 4 (theorem 1) we study Bowen–Ruelle measures of generalized baker's transformations. The method is to study the effect of T_β on the Fourier–Stieltjes transforms of measures. For the reader's convenience, the relevant probability theory is summarized in § 2. The Bowen–Ruelle measure of T_β is the product of the uniform measure in the y -direction and a measure in the x -direction we call an *infinitely convolved Bernoulli measure* (ICBM). These ICBM's have been studied for half a century, although their relation with dynamics has not been noted before (they were originally studied as interesting examples in harmonic analysis). A survey of these measures is given in § 3.

The Bowen–Ruelle measure describes the stochastic behaviour of a transformation T in the sense of time averages. One could also investigate ensemble averages. In § 5 (theorem 2) we prove a result of this form; we show that for a large class of measures μ_0 on $\mathbb{R} \times [-1, 1]$, the iterates $T_*^n \mu_0$ converge in distribution to scalar multiples of the Bowen–Ruelle measure. Again the machinery involves Fourier–Stieltjes transforms. A consequence of this is that generalized baker's transformations are (strongly) mixing. Hence they are Bowen–Ruelle ergodic. (By completely different methods, which we do not pursue in this paper, we can show that generalized baker's transformations are factors of Bernoulli shifts, hence k -endomorphisms.)

Fat baker's transformations are (area) expanding. The Bowen–Ruelle measures of one-dimensional piecewise smooth expanding transformations are absolutely continuous with respect to Lebesgue measure [6]. If T is piecewise analytic on the square in \mathbb{R}^2 and expands the length of every vector, the Bowen–Ruelle measure of each Lebesgue ergodic piece is absolutely continuous with respect to Lebesgue measure [19]. In contrast, there exists an infinite number of $\beta > \frac{1}{2}$ (e.g. $\beta = (-1 + \sqrt{5})/2$) such that the Bowen–Ruelle measure of T_β is totally singular with respect to Lebesgue measure. The Bowen–Ruelle measures of skinny baker's transformations are all products of Cantor measures with the uniform measure in the y -direction, and so are clearly totally singular.

Lyapunov dimension. The Lyapunov exponents (characteristic exponents) of a transformation, which we denote $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$, were defined in [21]. They depend on an ergodic measure. If T is Bowen–Ruelle ergodic and the measure in question is the Bowen–Ruelle measure, define the *Lyapunov dimension* of T , denoted $d_{\text{Lyp}}(T)$ as follows [16]. Let

$$k = \max \{i: \lambda_1 + \dots + \lambda_i \geq 0\}$$

(or 0 if $\lambda_1 < 0$). Then

$$d_{\text{Lyp}}(T) = \begin{cases} 0 & \text{if } k = 0, \\ k + \frac{\lambda_1 + \dots + \lambda_k}{|\lambda_{k+1}|} & \text{if } 1 \leq k < m, \\ m & \text{if } k = m. \end{cases}$$

All of this can be made to work for piecewise C^1 transformations, with a straightforward modification. Suppose the manifold M is divided into a finite number of regions R_1, \dots, R_k bounded by a finite number of codimension 1 submanifolds N_1, \dots, N_L . If T is C^1 on each region R_k , is Bowen–Ruelle ergodic, and if the Bowen–Ruelle measure of the inverse images at the boundaries $\bigcup_n T^{-n}(\bigcup N_i)$ is zero, then the appropriate Lyapunov exponents and the Lyapunov dimension can be defined.

In particular, since generalized baker's transformations are piecewise linear, it is easy to compute the Lyapunov exponents. We find that for T_β

$$\lambda_1 = \log 2 \quad \lambda_2 = \log \beta$$

and thus

$$d_{\text{Lyp}}(T_\beta) = \begin{cases} 1 + \frac{\log 2}{\log \beta^{-1}} & \beta < \frac{1}{2}, \\ 2 & \beta \geq \frac{1}{2}. \end{cases}$$

For a transformation T that is not Lebesgue ergodic, it may be useful on occasion to replace M by a small neighbourhood of a piece of the attractor on which T is Lebesgue ergodic.

Information dimension. The Lyapunov dimension is a property of the dynamics of a transformation T ; in particular it measures something about the expansion rates of T . We turn now to a dimension of a set with a measure essentially first defined

in [1]. In particular, when the measure is the Bowen–Ruelle measure of a transformation, we call the dimension the information dimension of T . It measures in some sense the infinitesimal complexity of the attractor of T . We then discuss a conjecture that relates the Lyapunov dimension and the information dimension.

First we discuss the heuristics of the information dimension. Consider again the computer plot of a transformation T discussed above. Ideally, it is a picture of the attractor of T together with the Bowen–Ruelle measure on the attractor. We study this picture with some degree of visual acuity. The information dimension measures the asymptotic rate of growth of the information we can obtain about (alternately, needed to specify) the plot as our acuity becomes arbitrarily fine.

More precisely, let X be a metric space with a Borel probability measure μ . If K is a finite partition of X , let the diameter of K , denoted $\text{diam } K$, be the maximum of the diameters of the elements of K . Let

$$\begin{aligned} \psi(x) &= x \log x^{-1} & 0 < x \leq 1, \\ \psi(0) &= 0 \end{aligned}$$

be the information function. Define

$$h(\mu, K) = \sum_{A \in K} \psi(\mu(A)),$$

the absolute value of the logarithm of the geometric mean of the measures of the $A \in K$. For $\varepsilon > 0$, let

$$h(\mu, \varepsilon) = \inf \{h(\mu, K) : \text{diam } K \leq \varepsilon\}.$$

Then we define upper and lower *information dimensions*:

$$\begin{aligned} \bar{d}_{\text{info}}(\mu) &= \limsup_{\varepsilon \rightarrow 0} \frac{h(\mu, \varepsilon)}{\log \varepsilon^{-1}} \\ \underline{d}_{\text{info}}(\mu) &= \liminf_{\varepsilon \rightarrow 0} \frac{h(\mu, \varepsilon)}{\log \varepsilon^{-1}} \end{aligned}$$

and if they are equal $d_{\text{info}}(\mu)$ denotes the common value. These quantities do not involve any dynamics, only a measure. If μ is the Bowen–Ruelle measure of a transformation T , we call these quantities the (upper, lower) *information dimensions* of (the attractor of) T and denote them $\bar{d}_{\text{info}}(T)$, $\underline{d}_{\text{info}}(T)$, $d_{\text{info}}(T)$, respectively.

Main conjecture. The main conjecture states that the equality

$$d_{\text{info}}(T) = d_{\text{Lyap}}(T) \tag{*}$$

holds most of the time. More precisely, consider the space of C^2 maps of a manifold M to itself with compact attractors. The conjecture is that *the set of Bowen–Ruelle ergodic transformations which do not satisfy (*) is of infinite codimension in this space.*

A conjecture of this form was first stated in [16] and refined in [10]. The present form is a further refinement. In [20] it is shown the right-hand side of (*) always dominates the left-hand side. The value of (*) is that $d_{\text{Lyap}}(T)$ can be computed in terms of the dynamics of T ; indeed, it can be estimated by computer calculations

[2]. On the other hand, $d_{\text{info}}(T)$ gives geometric measure-theoretic information about the complexity of the attractor of T . The conjecture can be extended in the obvious way for piecewise C^2 transformations. There are a variety of other ways of defining information-type dimensions; it seems they should be equal for 'most' Bowen–Ruelle ergodic transformations. See [27].

We remark that in order to define the terms in (*), T must be Bowen–Ruelle ergodic. The conjecture is non-trivial in that the set of Bowen–Ruelle ergodic transformations contains open sets. Two transformations T, T' are topologically conjugate if there is a homeomorphism h such that $T' = h^{-1}Th$. If T and T' are topologically conjugate and T is Bowen–Ruelle ergodic, so is T' . (If h is C^1 , the Lyapunov and information dimensions of T and T' are equal.) There exist structurally stable Bowen–Ruelle ergodic transformations (e.g. ergodic axiom A systems [5], diffeomorphisms of surfaces [27], and the examples of [17]).

Further examples and evidence for the conjecture may be found in [16], [10], [9] and the references therein. More recently, L-S. Young has shown that (*) holds for diffeomorphisms of two dimensional manifolds provided the Lyapunov exponents $\lambda_1, \lambda_2 \neq 0$ and the measure theoretic entropy equals λ_1 [27]. (This proviso is true for generalized baker's transformations; they can be shown to have entropy $\log 2$.) A class of transformations of a three-dimensional manifold is investigated in [17] for which the attractor is a two-dimensional topological torus, but $d_{\text{Lyap}} > 2$. Here capacity rather than information dimension is studied. The formula corresponding to (*) is shown to hold except for a set of transformations of infinite codimension (and in these exceptions the torus is smooth and so has capacity 2).

The present paper is an investigation of another type of example. The computation of the Lyapunov dimension is trivial. In § 6 we consider the information dimension. The equality (*) is valid for many fat baker's transformations, in particular all those with absolutely continuous Bowen–Ruelle measures. On the other hand, there are an infinite number of fat baker's transformations for which (*) is false. In [12] a type of entropy is defined for ICBM's and shown to be strictly less than 1 for a countable set of $\beta > \frac{1}{2}$. In § 8 we show that for these β , $d_{\text{info}}(T_\beta)$ equals Garsia's entropy invariant plus 1.

One of these exceptional values is $\beta = (-1 + \sqrt{5})/2$. In § 7, we exhibit numerical calculations to estimate $d_{\text{info}}(T_\beta)$ for this β . The calculations have only an heuristic foundation, but they seem to indicate that for this β ,

$$d_{\text{info}}(T_\beta) \approx 1.99571312 \pm 0.00000005.$$

Acknowledgement. The authors are indebted to a number of their colleagues for suggestions and encouragement. In particular they would like to thank E. Slud for assisting several times with probability theory. The research of J. C. Alexander was supported by NSF Grants MSC 79-03462, MCS 81-17967. The research of J. A. Yorke was sponsored by the Air Force Office of Scientific Research under grant AFOSR-81-0217. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon.

2. *Fourier transforms*

We recall some basic standard facts about Fourier transforms of measures, and develop a convergence result we need. Let μ be a measure on the real line and $F_\mu(x) = \mu(-\infty, x)$ be its distribution function. We define the *Fourier–Stieltjes transform (characteristic function)* $\phi(\omega) = \phi_\mu(\omega)$ of μ by

$$\phi(\omega) = \int_{-\infty}^{\infty} \exp(i\omega x) dF_\mu(x). \tag{2.1}$$

Every measure is determined by its characteristic function. If μ is supported on the finite interval $[-a, a]$, we write the formal *Fourier–Stieltjes expansion* of dF_μ :

$$dF_\mu(x) \sim \sum_{m=-\infty}^{\infty} \phi(m) \exp(imx). \tag{2.2}$$

Here the $\phi(m)$ are the *Fourier–Stieltjes coefficients* of μ (or dF_μ) defined by

$$\phi(m) = \frac{1}{2a} \int_{-a}^a \exp \frac{-im\pi x}{a} dF_\mu(x). \tag{2.3}$$

Every measure is determined by its Fourier–Stieltjes expansion. A sequence of measures μ_n (with distribution functions F_n) converges *weakly (or in distribution)* to μ (with distribution function F) if:

$$\int f(x) dF_n(x) \rightarrow \int f(x) dF(x)$$

for each continuous f . Such convergence is known to be equivalent to any of the following:

- (i) convergence of $F_n(x) \rightarrow F(x)$ at every continuity point x of F ;
- (ii) convergence of $\mu_n(E) \rightarrow \mu(E)$ for every Borel set E with $\mu(\partial E) = 0$;
- (iii) pointwise convergence of $\phi_n(\omega) \rightarrow \phi(\omega)$, where ϕ_n (resp. ϕ) are the characteristic functions of μ_n (resp. μ);
- (iv) convergence of each $\phi_n(m) \rightarrow \phi(m)$, where $\phi_n(m)$ (resp. $\phi(m)$) are the Fourier–Stieltjes coefficients of μ_n (resp. μ), if the coefficients are defined over a finite interval which supports all the μ_n, μ .

If the measure μ is defined on higher-dimensional Euclidean space (say $\mathbb{R}^2 = \{(x, y)\}$), everything is formally the same. The distribution function is

$$F_\mu(x, y) = \mu\{(-\infty, x) \times (-\infty, y)\}.$$

It is convenient for us to use the following (mixed) *Fourier–Stieltjes transform* of μ . In the x -direction we form the Fourier–Stieltjes transform; in the y -direction we calculate the Fourier–Stieltjes coefficients (with respect to the interval $[-1, 1]$).

Thus we let

$$\phi(\omega, n) = \frac{1}{2} \int_{x=-\infty}^{\infty} \int_{y=-1}^1 \exp(i\omega x - in\pi y) dF_\mu(x, y). \tag{2.4}$$

where ω ranges over the reals and n over the integers.

Then weak convergence is defined as above and is equivalent to (i), (ii) above. We combine (iii), (iv) into the following:

- (v) convergence of $\phi_n(\omega, m) \rightarrow \phi(\omega, m)$ for each ω, m .

This is almost a standard fact (it is standard for the higher-dimensional characteristic functions). One uses the Cramér–Wald argument [3, pp. 334–335] to reduce it to the one-dimensional case. More precisely, let θ be the polar angle in the (x, y) -plane. Consider the half-space measure at angle θ , $0 \leq \theta < \pi$. If $\theta \neq 0$, $\phi_n(m \cot \theta, m)$, $\phi(m \cot \theta, m)$ are the Fourier–Stieltjes coefficients of the half-space measures determined by μ_n, μ , respectively. Similarly for $\theta = 0$, $\phi_n(\omega, 0)$, $\phi(\omega, 0)$ are the characteristic functions of the horizontal half-space measures. If (v) is true, each of these, with fixed θ , converges; hence the distribution functions

$$F_n(y \cot \theta, y) \rightarrow F(y \cot \theta, y), \quad F_n(x, 0) \rightarrow F(x, 0),$$

at each point of continuity of $F(y \cot \theta, y)$, $F(x, 0)$ respectively. Hence $F_n(x, y) \rightarrow F(x, y)$ at each point of continuity of F , so $\mu_n \rightarrow \mu$.

3. The component measures

The Bowen–Ruelle measure of a fat baker's transformation is the product of measures in the x and y -directions. In this section we discuss the component measures.

To exploit the symmetries in the transformation, it is convenient to use *signed dyadic expansions* on $[-1, 1]$. Any real number y , $0 \leq y \leq 1$ has a dyadic expansion

$$y = \sum_{r=1}^{\infty} (y)^{(r)} 2^{-r}, \quad (y)^{(r)} = 0 \text{ or } 1.$$

The $(y)^{(r)}$ are the dyadic digits. They are uniquely defined except when y is a dyadic rational, in which case a choice must be made between the terminating and non-terminating expansion.

Similarly any real number y , $-1 \leq y \leq 1$ has a signed dyadic expansion

$$y = \sum_{r=1}^{\infty} (y)_r 2^{-r}, \quad (y)_r = -1 \text{ or } +1. \tag{3.1}$$

The coefficients $(y)_r$ are called the *signed dyadic digits* of y . They can be determined by forming the dyadic expansion of $y' = \frac{1}{2}(y + 1)$ and letting $(y)_r = 1 - 2(y')^{(r)}$. Again a choice must be made for dyadic rationals. The dyadic rationals form a set of measure zero and any choice is immaterial to our discussion of measure. For definiteness, we choose a tail of -1 's over a tail of $+1$'s, except for $y = +1$ which is expanded

$$y = \sum_{r=1}^{\infty} (+1) 2^{-r}.$$

(Another, more symmetric, possibility is to allow tails of 0's for dyadic rationals, except for $y = \pm 1$. Then the $(y)_i$ are the Rademacher functions of $\frac{1}{2}(y + 1)$ if $-1 < y < 1$.)

Let $\tau: [-1, 1] \rightarrow [-1, 1]$ denote the doubling map:

$$\tau(y) = \begin{cases} 2y - 1, & y > 0 \\ 2y + 1, & y \leq 0, \end{cases} \tag{3.2}$$

or

$$\tau(y) = 2y - (y)_1.$$

The transformation τ obviously preserves Lebesgue measure. Also $(\tau y)_r = (y)_{r+1}$; this represents τ as a one-sided Bernoulli shift.

We now consider the component measures on the x -axis. We describe them in several ways. Fix β , $0 < \beta < 1$. We denote the measure by ν_β .

Let a_n , $n = 0, 1, 2, \dots$, be a sequence of independent random variables, each with the values ± 1 with probability $\frac{1}{2}$. Let F_β be the distribution function of the random variable

$$X_\beta = \sum_{n=0}^{\infty} a_n(1 - \beta)\beta^n. \tag{3.3}$$

Such a random variable β is a type of Rademacher series. The measure ν_β is the measure with distribution function F_β .

In other words, ν_β measures the density of points of the form $\sum_{n=0}^{\infty} \pm(1 - \beta)\beta^n$. For any interval (a, b) and any integer $N > 0$, let $\nu_{\beta,N}(a, b)$ denote the proportion of points of the form $\sum_{n=0}^{N-1} \pm(1 - \beta)\beta^n$ that lie in (a, b) . That is

$$\nu_{\beta,N}(a, b) = 2^{-N} \# \left\{ x: x = \sum_{n=0}^{N-1} \pm(1 - \beta)\beta^n, \quad a < x < b \right\} \tag{3.4}$$

where $\#$ denotes counting with multiplicity. Then

$$\nu_\beta(a, b) = \lim_{N \rightarrow \infty} \omega_{\beta,N}(a, b). \tag{3.5}$$

Equivalently ν_β can be expressed as an infinite convolution. Let $b(\xi)$ denote the measure with two atoms, each of weight $\frac{1}{2}$, at the points $\pm \xi$. Such a measure is called a *Bernoulli measure*. The measure ν_β can be expressed as the infinite convolution

$$\nu_\beta = b(1 - \beta) * b((1 - \beta)\beta) * \dots * b((1 - \beta)\beta^n) * \dots$$

These measures have been studied for more than 40 years, and they are not yet completely understood. They have been called by a variety of names; we call them *infinitely convolved Bernoulli measures* (ICBM's). Consider two examples. If $\beta = \frac{1}{2}$, the expressions $\sum_{n=0}^{N-1} \pm(\frac{1}{2})^{n+1}$ are signed binary expansions. Thus

$$\nu_{\beta,N}(a, b) \rightarrow \frac{1}{2}(b - a)$$

for $-1 \leq a \leq b \leq 1$, and ν_β is the uniform probability measure. If $\beta = \frac{1}{3}$, let $z = \frac{1}{2}(x + 1)$. If

$$x = \sum_{n=0}^{\infty} a_n \cdot \frac{2}{3}(\frac{1}{3})^n,$$

with each $a_n = \pm 1$, then

$$z = \sum_{n=0}^{\infty} (a_n + 1)(\frac{1}{3})^{n+1}.$$

As the $\{a_n\}$ range over all sequences of ± 1 , the z range over all numbers without 1's in their triadic expansions. Thus ν_β is (an affine copy of) the Cantor measure. More generally, if $\beta < \frac{1}{2}$, the support of ν_β is a Cantor set and ν_β is totally singular. If $\beta \geq \frac{1}{2}$ every open set has positive measure.

Consider now the characteristic function of ν_β . The characteristic function of the Bernoulli measure $b(\xi)$ is $\cos \xi\omega$. Since the characteristic function of a convolution is the product of the characteristic functions of the components, the characteristic function ϕ_β of ν_β is

$$\phi_\beta(\omega) = \prod_{r=0}^{\infty} \cos [(1 - \beta)\beta^r \omega]. \tag{3.6}$$

Convolved Bernoulli measures have been studied since the 1930's, especially for $\beta < \frac{1}{2}$, in no small measure because they are interesting examples of phenomena in harmonic analysis. For our purposes, we recall the following results. The measures and their characteristic functions are well-defined; i.e. all the limiting processes above converge [14]. The ν_β are continuous and are either absolutely continuous or totally singular [14]. For β the n 'th root of $\frac{1}{2}$, ν_β is absolutely continuous (and indeed, progressively smoother as n is increased) [26].

Thus it was suspected that all ν_β , $\beta \geq \frac{1}{2}$ might be absolutely continuous. However, P. Erdős showed there were values of β , e.g. $\beta = (-1 + \sqrt{5})/2$, such that ν_β is totally singular [7]. The key fact he used is that β^{-1} is an algebraic integer, all of whose conjugates lie inside the unit circle in the complex plane. This allowed him to estimate the characteristic function $\phi_\beta(\omega)$ for large ω and show it did not approach zero as $\omega \rightarrow \infty$. Thus by the Riemann–Lebesgue lemma, ν_β cannot be absolutely continuous. A year later, again by using estimates on the asymptotics of the characteristic function, he showed there is a $\gamma < 1$ such that for almost all β with $\gamma < \beta < 1$, the measure ν_β is absolutely continuous [8].

Real numbers which are algebraic integers whose conjugates lie within the unit circle were evidently originally mentioned by Hardy. They were studied by his student Vijayarhavan, by Pisot, by R. Salem and others and have come to be known as PV (Pisot–Vijayarhavan) numbers. When γ is a PV number, γ^n is nearly an integer; if $d(\gamma^n)$ denotes the distance from γ^n to the nearest integer, then $\sum_{n=0}^{\infty} d(\gamma^n)$ converges. There is an extensive literature concerning them and their generalizations, mostly in the context of algebraic number theory. In particular, Salem showed the set of PV numbers is a closed (obviously countable) subset of the reals and that 1 is an isolated element [22]. He also showed that β^{-1} being a PV number is necessary and sufficient for $\phi_\beta(\omega)$ not to approach zero as $\omega \rightarrow \infty$. C. L. Siegel showed that the smallest PV number larger than 1 is the positive root $\theta_1 \sim 1.324718$, ($\theta_1^{-1} \sim .7548777$), of $x^3 - x - 1$, that the next in order is the root $\theta_2 \sim 1.3802777$, ($\theta_2^{-1} \sim .7244918$), of $x^4 - x^3 - 1$, that both of these PV numbers are isolated, and that all other PV numbers are larger than $\sqrt{2}$ [24].

The reciprocals of PV numbers are the only $\beta \geq \frac{1}{2}$ for which ν_β is known to be totally singular. It is possible (but thought unlikely) that ν_β is absolutely continuous for all other $\beta \geq \frac{1}{2}$. The only quadratic β of this type is $\beta = (-1 + \sqrt{5})/2$, the root of $x^2 + x - 1$. The only such β satisfying a cubic equation are the real roots of the following four polynomials: $x^3 + x^2 + x - 1$, $x^3 + x^2 - 1$, $x^3 + x - 1$, $x^3 - x^2 + 2x - 1$, with $\beta \approx .7548777$, $.5436898$, $.6823278$, $.5698403$, resp. We also note that there is a sequence of such β approaching $\frac{1}{2}$ from above, namely roots of the polynomials

$$x^n + x^{n-1} + \dots + x - 1.$$

A. Garsia discovered a family of algebraic β for which ν_β is absolutely continuous [11]. These and the roots of $\frac{1}{2}$ are the only explicit β for which ν_β is known to be absolutely continuous, although by Erdős's result there are uncountably many such β .

For more expository treatments of some of the above material, see [23], [15], [18].

4. Bowen–Ruelle measure

Let $\sigma = \sigma_\beta$ be the measure on $\mathbb{R} \times [-1, 1]$ which is the product of the infinitely convolved Bernoulli measure ν_β on the x -axis and the uniform probability measure (with density $\frac{1}{2}$) on the y -axis. The purpose of this section is to prove that σ_β is the Bowen–Ruelle measure for the generalized baker's transformation T_β . As indicated in the introduction, we need to show that for almost all points (x, y) (with respect to Lebesgue measure), the measures

$$\frac{1}{M+1} \sum_{N=0}^M T_*^N \delta_{(x,y)}$$

converge weakly to σ_β , where $\delta_{(x,y)}$ is the atomic probability measure at (x, y) . We prove this by studying the effect of T_β on the Fourier transform of $\delta_{(x,y)}$.

Using signed dyadic notation (3.1), we write $T = T_\beta$ as

$$T(x, y) = (\beta x + (y)_i(1 - \beta), 2y - (y)_i) \tag{4.1}$$

for $x \in \mathbb{R}, -1 \leq y \leq 1$. For any measure γ on $\mathbb{R} \times [-1, 1]$, let $\phi_\gamma(\omega, n)$ be the mixed Fourier transform (2.4) of γ . Then

$$\begin{aligned} T_* \phi_\gamma(\omega, n) &= \frac{1}{2} \int_{x=-\infty}^{\infty} \int_{y=-1}^1 \exp(i\omega x - in\pi y) dF_{T_*\gamma}(x, y) \\ &= \frac{1}{2} \int_{x=-\infty}^{\infty} \int_{y=-1}^1 \exp(i\omega\beta x) \exp(-2in\pi y) \\ &\quad \cdot \exp(i(y)_i(1 - \beta)\omega) \exp(i(y)_i n\pi y) dF_\gamma(x, y) \\ &= \frac{(-1)^n}{2} \int_{x=-\infty}^{\infty} \int_{y=-1}^1 \exp(i\omega\beta x) \exp(-2in\pi y) \\ &\quad \cdot \exp(i(y)_i(1 - \beta)\omega) dF_\gamma(x, y). \end{aligned} \tag{4.2}$$

Thus

$$\begin{aligned} T_*^N \phi_\gamma(\omega, n) &= \frac{(-1)^n}{2} \int_{x=-\infty}^{\infty} \int_{y=-1}^1 \exp(i\omega\beta^N x) \exp(-2^N in\pi y) \\ &\quad \cdot \prod_{r=0}^{N-1} \exp(i(y)_{N-r}(1 - \beta)\beta^r \omega) dF_\gamma(x, y). \end{aligned} \tag{4.3}$$

The Fourier transform of $\sigma = \sigma_\beta$ is clearly

$$\phi_\sigma(\omega, n) = \begin{cases} \frac{1}{2} \prod_{r=0}^{\infty} \cos(1 - \beta)\beta^r \omega & n = 0, \\ 0 & n \neq 0. \end{cases} \tag{4.4}$$

Using (4.2) (and integrating first in the y -direction), we can easily check that ϕ_σ and thus σ is invariant under the action of T .

THEOREM 1. For almost all y in $[-1, 1]$,

$$\frac{1}{M+1} \sum_{N=0}^M T_*^N \delta_{(x,y)}$$

converges weakly to σ for all $x \in \mathbb{R}$.

Proof. Let $\delta = \delta_{(x,y)}$. We can assume $|x| \leq 1$. We show

$$\frac{1}{M+1} \sum_{N=0}^M T_*^N \phi_\delta(\omega, n) \rightarrow \phi_\sigma(\omega, n) \tag{4.5}$$

pointwise for almost all y . Fix ω, n and $\varepsilon > 0$. The left-hand side of (4.5) is, by (4.3),

$$\frac{(-1)^n}{2} \frac{1}{M+1} \sum_{N=0}^M \exp(i\omega\beta^N x) \exp(-2^N i\pi n y) \prod_{r=0}^{N-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega). \tag{4.6}$$

Choose M_1 so that if $M > M_1$,

$$|1 - \exp(i\omega\beta^M)| < \varepsilon \quad \text{and} \quad \left| 1 - \prod_{r=M_1+1}^\infty \exp(i(1-\beta)\beta^r \omega) \right| < \varepsilon.$$

Choose $M_2 > M_1$ so that

$$\left| \frac{M - M_1}{M + 1} - 1 \right| < \varepsilon$$

if $M > M_2$. Write (4.6) as

$$\begin{aligned} & \frac{1}{2} \frac{(-1)^n}{M+1} \sum_{N=0}^{M_1} \exp(i\omega\beta^N x) \exp(-2^N i\pi n y) \prod_{r=0}^{N-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) \\ & + \frac{1}{2} \frac{(-1)^n}{M+1} \sum_{N=M_1+1}^M (\exp(i\omega\beta^N x) - 1) \exp(-2^N i\pi n y) \prod_{r=0}^{N-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) \\ & + \frac{1}{2} \frac{(-1)^n}{M+1} \sum_{N=M_1+1}^M \exp(-2^N i\pi n y) \prod_{r=0}^{M_1-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) \\ & \quad \cdot \left[\prod_{r=M_1}^{N-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) - 1 \right] \\ & + \frac{1}{2} \left(\frac{M - M_1}{M + 1} - 1 \right) \frac{(-1)^n}{M - M_1} \sum_{N=M_1+1}^M \exp(-2^N i\pi n y) \prod_{r=0}^{M_1-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) \\ & + \frac{1}{2} \frac{(-1)^n}{M - M_1} \sum_{N=M_1+1}^M \exp(-2^N i\pi n y) \prod_{r=0}^{M_1-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega). \end{aligned} \tag{4.7}$$

Each of the first four terms is less than $\varepsilon/2$ if $M > M_2$. We consider the last term. Note that $\exp(-2^N i\pi n y)$ does not depend on the first N signed dyadic digits $(y)_1, \dots, (y)_N$ of y . Indeed, using the doubling operator τ of (3.2),

$$\exp(-2^N i\pi n y) = \exp(-i\pi n(\tau^N y)) = \exp(-i\pi n \tau^{N-M_1}(\tau^{M_1} y)).$$

Thus if we let

$$g(y) = \frac{1}{2} \exp(-2^{M_1} i\pi n y) \prod_{r=0}^{M_1-1} \exp(i(y)_{M_1-r}(1-\beta)\beta^r \omega),$$

the last term of (4.7) can be written

$$\frac{(-1)^n}{M - M_1} \sum_{N=M_1+1}^M g(\tau^{N-M_1}y).$$

Now τ preserves the uniform measure and is ergodic. By Birkhoff’s ergodic theorem [25], for almost all y ,

$$\begin{aligned} \frac{1}{M - M_1} \sum_{N=M_1+1}^M g(\tau^{N-M_1}y) &\rightarrow \int_{-1}^1 g(y) dy \\ &= \frac{1}{2} \int_{-1}^1 \exp(-2^{M_1}i\pi ny) \prod_{r=0}^{M_1-1} \exp(i(y)_{M_1-r}(1-\beta)\beta^r\omega) dy, \end{aligned}$$

as $M \rightarrow \infty$. Consider the subintervals

$$I_k = [-1 + (k-1)2^{-M_1+1}, -1 + k \cdot 2^{-M_1+1}), \quad k = 1, \dots, 2^{n_1}.$$

On each I_k , $\prod_{r=0}^{M_1-1} \exp(i(y)_{M_1-r}(1-\beta)\beta^r\omega)$ is a constant C_k . Indeed as k runs over the 2^{M_1} subintervals, the sequences $(y)_{M_1-r}$, $r = 0, \dots, M_1-1$, run over all 2^{M_1} possibilities. On each I_k ,

$$\frac{1}{2} \int_{I_k} \exp(-2^{M_1}i\pi ny) dy = \begin{cases} 2^{-M_1} & n = 0, \\ 0 & n \neq 0. \end{cases}$$

Therefore

$$\int_{-1}^1 g(y) dy = \prod_{r=0}^{M_1-1} \cos(1-\beta)\beta^r\omega.$$

With our choice of M_1 ,

$$\left| 1 - \prod_{r=M_1}^{\infty} \cos(1-\beta)\beta^r\omega \right| < \varepsilon.$$

Therefore, for almost all y , $\lim_{M \rightarrow \infty} (1/M+1) \sum_{N=0}^M T_*^N \phi_\delta(\omega, n)$ is within 3ε of $\phi_\delta(\omega, n)$. This proves the theorem. \square

Remark. By the converse of the ergodic theorem, theorem 1 proves that σ is ergodic. In the next section, we prove a different result which implies σ is strongly mixing.

5. Ensemble averages

In this section we show that for many probability measures μ on $\mathbb{R} \times [-1, 1]$ (in particular, any absolutely continuous measure), the weak limit of $T_*^N \mu$ is σ . No averaging over N is needed.

THEOREM 2. *Let μ be any measure whose projection $\pi\mu$ on the y -axis is absolutely continuous with respect to Lebesgue measure. Then*

$$T_*^N \mu \rightarrow |\mu| \sigma$$

where $|\mu| = \mu(\mathbb{R} \times [-1, 1])$ is the total mass of μ .

Proof. Suppose $|\mu| = 1$. Note that T_* is an isometry on measures. We can approximate μ by measures with compact support with absolutely continuous projections to the y -axis. Thus we can suppose without loss of generality that μ is supported on

$[-K, K] \times [-1, 1]$. We show

$$T_*^N \phi_\mu(\omega, n) \rightarrow \phi_\sigma(\omega, n)$$

pointwise. Fix ω, n and $\varepsilon > 0$. Choose L so that

$$|1 - \exp(i\omega K\beta^N)| < \varepsilon \tag{5.1}$$

for $N \geq L$. Suppose $\int_{x=-K}^K dF_\mu(x, y) = \int_{-1}^1 f(y) dy$, where $f(y) \in L^1[-1, 1]$. From (4.3) and (5.1), for $N > L$, $T_*^N \phi_\mu(\omega, n)$ is within $\frac{1}{2}\varepsilon$ of

$$\frac{(-1)^n}{2} \int_{y=-1}^1 \exp(-2^N in\pi y) \prod_{r=0}^{N-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) f(y) dy. \tag{5.2}$$

Choose M so large that

$$\left| 1 - \prod_{r=M}^\infty \exp(i(1-\beta)\beta^r \omega) \right| < \varepsilon. \tag{5.3}$$

Let $\hat{f}(y)$ (obtained by chopping off the peaks of f) be a non-negative bounded function so that

$$\int_{-1}^1 |f(y) - \hat{f}(y)| dy < \varepsilon \tag{5.4}$$

and

$$\int_{-1}^1 \hat{f}(y) dy \leq 1. \tag{5.5}$$

Write (5.2) as

$$\begin{aligned} & \frac{(-1)^n}{2} \int_{y=-1}^1 \exp(-2^N in\pi y) \prod_{r=0}^{M-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) (f(y) - \hat{f}(y)) dy \\ & + \frac{(-1)^n}{2} \int_{y=-1}^1 \exp(-2^N in\pi y) \prod_{r=0}^{M-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) \hat{f}(y) dy \\ & + \frac{(-1)^n}{2} \int_{y=-1}^1 \exp(-2^N in\pi y) \prod_{r=0}^{M-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) \\ & \quad \cdot \left[\prod_{r=M}^{N-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) - 1 \right] \hat{f}(y) dy. \end{aligned} \tag{5.6}$$

By (5.4) and (5.3), the first and third terms of (5.6) are bounded by $\frac{1}{2}\varepsilon$. Given P , consider the 2^P subintervals I_k of $[-1, 1]$;

$$I_k = [-1 + (k-1)2^{-P+1}, -1 + k \cdot 2^{-P+1}), \quad k = 1, \dots, 2^P.$$

Define

$$f_P(y) = \begin{cases} C_k = 2^{P-1} \int_{I_k} \hat{f}(y) dy & y \in I_k, \\ 0 & y = 1. \end{cases}$$

Then $f_P(y) \rightarrow \hat{f}(y)$ in $L^1[-1, 1]$, so choose P so that

$$\int_{y=-1}^1 |\hat{f}(y) - f_{P-M+1}(y)| dy < \varepsilon.$$

From (5.4), (5.5) we see

$$1 - \varepsilon < \sum_{k/2^{P-M}} C_k \leq 1. \tag{5.7}$$

By the choice of P , we see that $T_*^N \phi_\mu(\omega, n)$ is within 2ε of

$$\begin{aligned} & \frac{(-1)^n}{2} \int_{y=-1}^1 \exp(-2^N i\pi n y) \prod_{r=0}^{M-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) f_{P-M+1}(y) dy \\ &= \frac{(-1)^n}{2} \sum_k \int_{I_k} \exp(-2^N i\pi n y) \prod_{r=0}^{M-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) f_{P-M+1}(y) dy \\ &= \frac{(-1)^n}{2} \sum_k C_k \int_{I_k} \exp(-2^N i\pi n y) \prod_{r=0}^{M-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) dy. \end{aligned} \tag{5.8}$$

Let $N > P$. In I_k consider subintervals $I_{k,l}$ of length 2^{-N+1} . On each $I_{k,l}$

$$\prod_{r=0}^{M-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega)$$

is constant. Therefore

$$\begin{aligned} & \int_{I_k} \exp(-2^N i\pi n y) \prod_{r=0}^{M-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) dy \\ &= \begin{cases} \frac{1}{2^{P-M}} \prod_{r=0}^{M-1} \cos(1-\beta)\beta^r \omega & \text{if } n = 0, \\ 0 & \text{if } n \neq 0. \end{cases} \end{aligned}$$

Therefore

$$\begin{aligned} & \frac{(-1)^n}{2} \int_{y=-1}^1 \exp(-2^N i\pi n y) \prod_{r=0}^{M-1} \exp(i(y)_{N-r}(1-\beta)\beta^r \omega) f_{P-M+1}(y) dy \\ &= \begin{cases} \frac{1}{2} \frac{C_k}{2^{P-M}} \prod_{r=0}^{M-1} \cos(1-\beta)\beta^r \omega & \text{if } n = 0, \\ 0 & \text{if } n \neq 0. \end{cases} \end{aligned}$$

Note that

$$\left| 1 - \prod_{r=M}^{\infty} \cos(1-\beta)\beta^r \omega \right| < \varepsilon.$$

Therefore for $N > P$, $T_*^N \phi_\mu(\omega, n)$ is within 3ε of $\phi_\sigma(\omega, n)$. This completes the proof. \square

Recall [25] that a measure-preserving transformation T on a measure space (X, μ) is called *strongly mixing* if for every measurable $A, B \subset X$,

$$\lim_{N \rightarrow \infty} \mu(T^{-N}A \cap B) = \mu(A)\mu(B). \tag{5.9}$$

This is a stronger condition than ergodicity. Property (5.9) need be checked only for A, B in an algebra generating the measure algebra of X (in our case, on rectangles).

COROLLARY. T_β , operating on $\mathbb{R} \times [-1, 1]$ with measure σ_β , is strongly mixing.

Proof. For any Borel set B , let $\mu_B(A) = \sigma(A \cap B)$. Then

$$(T_*^N \mu_B)(A) = \mu_B(T^{-N}A) = \sigma(T^{-N}A \cap B).$$

Since μ_B is absolutely continuous with respect to σ , its projection $\pi\mu_B$ is absolutely continuous on the y -axis. Thus for any rectangle A

$$\lim_{N \rightarrow \infty} \sigma(T^{-N}A \cap B) = \lim_{N \rightarrow \infty} T_*^N \mu_B(A) = |\mu_B| \sigma(A) = \sigma(B)\sigma(A)$$

by (ii) of § 2 and the fact that σ is continuous. □

COROLLARY. *Suppose ν_β is totally singular. Then any measure invariant under T is totally singular.*

Proof. Let μ be an invariant measure. Write

$$\mu = \mu_{\text{sing}} + \mu_{\text{abs}}$$

where μ_{sing} and μ_{abs} are the singular and absolutely continuous parts of μ . Apply T_*^N and let $N \rightarrow \infty$; we find $\mu_{\text{abs}} = 0$. □

6. Dimensions of fat baker's transformations

Recall from the introduction that the Lyapunov dimension of a fat baker's transformation T_β is $d_{\text{Lyap}}(T_\beta) = 2$. In this section we compute in some sense the information dimension of fat baker's transformations. In particular we can show that (*) is not true for β the reciprocal of a PV number. For such β the information dimension is strictly less than 2. We also note that for many $\beta \geq \frac{1}{2}$, the value $d_{\text{info}}(T_\beta) = 2$. In particular, by [7] for almost all β near 1, the Bowen–Ruelle measure is absolutely continuous. We also have the following result (see also [1]).

PROPOSITION. *If μ is an absolutely continuous measure on a compact domain in d -dimensional Euclidean space, then*

$$d_{\text{info}}(\mu) = d.$$

Proof. We use a result of Young [27, prop. 4.3] which states that if there is a constant α_0 such that for almost all x (with respect to μ),

$$\liminf_{\rho \rightarrow 0} \frac{\log \mu(B_\rho(x))}{\log \rho} \geq \alpha_0,$$

then

$$d_{\text{info}}(\mu) \geq \alpha_0.$$

Here $B_\rho(x)$ denotes the ball of radius ρ centred at x . If μ is absolutely continuous, it has a density, say f . Almost every point x is a regular point of f ; thus for almost all x ,

$$\lim_{\rho \rightarrow 0} \frac{\mu(B_\rho(x))}{\rho^d} = f(x).$$

That is for each such x , given $\eta > 0$

$$\left| \frac{\mu(B_\rho(x))}{\rho^d} - f(x) \right| < \eta$$

for ρ small enough. Thus

$$\mu(B_\rho(x)) = \rho^d(\theta\eta + f(x))$$

where $|\theta| < 1$. For almost all x (with respect to μ), $f(x) > 0$. For such x , $\log \mu(B_\rho(x)) = d \log \rho + Q$, where Q is bounded independently of ρ , so that

$$\frac{\log \mu(B_\rho(x))}{\log \rho} = d + \frac{Q}{\log \rho} \rightarrow d,$$

as $\rho \rightarrow 0$. Setting $\alpha_0 = d$, the proof is complete. □

Our computation of information dimensions is in terms of an entropy invariant introduced by Garsia in [12]. He considered the entropies of the distributions of the random variables (3.3) and related them to the singularity or continuity of the ICBM's ν_β . Recall

$$\begin{aligned} \psi(x) &= x \log x^{-1} & 0 < x \leq 1 \\ \psi(0) &= 0 \end{aligned} \tag{6.1}$$

is the information function. Garsia considers points of the form

$$x = \sum_{r=0}^N \pm(1 - \beta)^r \tag{6.2}$$

and weights them according to multiplicity. Thus he has a discrete probability space $P_N = P_N(\beta)$ with measure ν_N ; let

$$H_N = \sum_{x \in P_N} \psi(\nu_N(x)) \tag{6.3}$$

denote its entropy. He then considers the limit

$$G = G_\beta = \lim_{N \rightarrow \infty} \frac{H_N}{N + 1} \tag{6.4}$$

which he shows exists and is the infimum of the $H_N/N + 1$. If there are no coincidences among the points x , then $H_N = N + 1$ and $G_\beta = 1$. If on the other hand there are coincidences, then $G_\beta < 1$. Thus G_β is a measure of how fast the points are piling up. In particular he shows that if

$$\frac{G_\beta}{\log \beta^{-1}} < 1 \tag{6.5}$$

then ν_β is totally singular, and in particular, if β is the reciprocal of a PV number, then (6.5) is true [12, theorem I.2]. We might note that our results in this section furnish another, completely different proof, that ν_β is totally singular if (6.5) holds. We can now state our result.

THEOREM 3. *For any β , $0 < \beta < 1$,*

$$\bar{d}_{\text{info}}(T_\beta) \leq \frac{G_\beta}{\log \beta^{-1}} + 1. \tag{6.6}$$

If β is the reciprocal of a PV number, then

$$d_{\text{info}}(T_\beta) = \frac{G_\beta}{\log \beta^{-1}} + 1. \tag{6.7}$$

Proof. The addition of 1 in the formulae is from the uniform measure in the y -direction. Accordingly we show that for any β , $0 < \beta < 1$,

$$\bar{d}_{\text{info}}(\nu_\beta) \leq \frac{G_\beta}{\log \beta^{-1}}, \tag{6.8}$$

and for β the reciprocal of a PV number,

$$d_{\text{info}}(\nu_\beta) = \frac{G_\beta}{\log \beta^{-1}}. \tag{6.9}$$

Given a Borel subset E of the real line, and constants $\gamma > 0$, x , let $\gamma E + x$ denote $\{y = \gamma z + x; z \in E\}$. Fix β . For any Borel measure μ on the real line, define

$$(T_*\mu)(E) = \frac{1}{2}[\mu(\beta^{-1}E + (1 - \beta^{-1})) + \mu(\beta^{-1}E - (1 - \beta^{-1}))].$$

For any μ , let $J(\mu)$ denote its *support interval*; i.e. the smallest closed interval which supports μ . Let $|\mu| = \mu(J(\mu))$. We consider $T_*\mu$ the sum of two measures with support intervals $\beta J(\mu) \pm (1 - \beta)$. Iterating we consider $T_*^n\mu$ as the sum of 2^n measures which are translates of each other, each a small copy of μ ; call them the components of $T_*^n\mu$.

Suppose now that μ is a probability measure with $J(\mu) = [-1, 1]$. Then the centres of the support intervals of the components of $T_*^n\mu$ are the points

$$x = \sum_{r=0}^{n-1} \pm(1 - \beta)\beta^r;$$

their lengths are $2\beta^n$. If some of the points x are equal, some of the components of $T_*^n\mu$ are equal. Amalgamate such equal components by adding; denote the resulting measures μ_1, \dots, μ_k . Since $|\mu| = 1$,

$$H_{n-1} = \sum_{i=1}^k \psi(|\mu_i|). \tag{6.10}$$

We say that two support intervals overlap if they contain a common subinterval. Many of the support intervals $J(\mu_i)$, $i = 1, \dots, k$ may overlap. We cut them up and redistribute the μ_i to eliminate overlaps. Suppose $J(\mu_i), J(\mu_j)$ overlap. If one support interval is contained in the other assume $J(\mu_i) \supset J(\mu_j)$, otherwise assume $|\mu_i| \geq |\mu_j|$. Construct two measures μ'_i, μ'_j by defining for any Borel set E :

$$\begin{aligned} \mu'_i(E) &= (\mu_i + \mu_j)(E \cap J(\mu_i)), \\ \mu'_j(E) &= \mu_j(E \setminus J(\mu_i)). \end{aligned}$$

Then $J(\mu'_i), J(\mu'_j)$ do not overlap, so we have eliminated at least one overlap. Moreover the following properties are satisfied;

$$\mu'_i + \mu'_j = \mu_i + \mu_j, \tag{6.11}$$

$$J(\mu'_i) = J(\mu_i), \quad J(\mu'_j) \subset J(\mu_j), \tag{6.12}$$

$$|\mu'_j| \leq |\mu_k| \leq |\mu'_i|, \quad k = i, j. \tag{6.13}$$

Since ψ is concave, we obtain from (6.13) that

$$\psi(|\mu'_i|) + \psi(|\mu'_j|) \leq \psi(|\mu_i|) + \psi(|\mu_j|). \tag{6.14}$$

Iterating this process, we eventually obtain a finite number of measures ξ_1, \dots, ξ_k with the properties

$$\sum_{i=1}^k \xi_i = \sum_{i=1}^k \mu_i = T_*^n \mu;$$

$J(\xi_i)$ and $J(\xi_j)$ do not overlap if $i \neq j$;

$$\sum_{i=1}^k \psi(|\xi_i|) \leq H_{n-1};$$

length $J(\xi_i) \leq 2\beta^n$ for each i .

Now we specialize to $\nu = \nu_\beta$. Since σ_β is invariant under $(T_\beta)_*$, we find $T_* \nu_\beta = \nu_\beta$. Thus $T_*^n \nu_\beta = \nu_\beta$ and $\nu_\beta(J(\xi_i)) = |\xi_i|$. In other words, the $J(\xi_i)$ are a partition of $[-1, 1]$ of diameter $\leq 2\beta^n$. Therefore

$$h(\nu_\beta, 2\beta^n) \leq \sum_{i=1}^k \psi(|\xi_i|) \leq H_{n-1}.$$

For $2\beta^{n+1} \leq \varepsilon \leq 2\beta^n$, $h(\nu_\beta, \varepsilon) \leq h(\nu_\beta, 2\beta^n)$; therefore

$$\limsup_{\varepsilon \rightarrow 0} \frac{h(\nu_\beta, \varepsilon)}{\log \varepsilon^{-1}} = \limsup_{n \rightarrow \infty} \frac{h(\nu_\beta, 2\beta^n)}{n \log \beta^{-1}} \leq \lim_{n \rightarrow \infty} \frac{H_{n-1}}{n} = G_\beta.$$

This proves (6.8).

To prove (6.9) we need more precise information about the function ψ of (6.1) and information about the spacing of the points (6.2) when β is the reciprocal of a PV number.

Consider numbers $p_j \geq 0$ for $j = 1, \dots, J$. Suppose

$$p_0 = \sum_{j=1}^J p_j \leq 1. \tag{6.15}$$

Then

$$\sum_{j=1}^J \psi(p_j) - p_0 \log J \leq \psi(p_0) \leq \sum_{i=1}^J \psi(p_i). \tag{6.16}$$

This is shown by using Lagrange multipliers to find the maximum of $\sum_{i=1}^J \psi(p_i)$ subject to the constraint (6.14). The maximum occurs when all $p_j = p_0/J$. With this, and induction on J , the minimum occurs when all but one p_j are zero. This gives the second inequality in (6.15). If all $p_j = p_0/J$,

$$\sum_{j=1}^J \psi(p_j) = J\psi(p_0/J) = \psi(p_0) + p_0 \log J,$$

which gives the first inequality. Let $p_i, i = 1, \dots, I, p_{ij}, i = 1, \dots, I, j = 1, \dots, J(i)$ be the two finite sets of non-negative numbers with

$$\sum_{i=1}^I p_i = 1, \tag{6.17}$$

$$\sum_{j=1}^{J(i)} p_{ij} = p_i \quad \text{for each } i.$$

Suppose each $J(i) \leq g$ for some g . Then (6.15) implies

$$\sum_{i,j} \psi(p_{ij}) - \log g \leq \sum_i \psi(p_i) \leq \sum_{i,j} \psi(p_{ij}). \tag{6.18}$$

If K and K' are two finite partitions of $[-1, 1]$, K' is a *refinement* of K if each interval in K' is a subinterval of one of the intervals of K . If each interval of K contains at most g intervals of K' , we call K' a *g-moderate* refinement of K . Recalling the definition in the introduction of $h(\mu, K)$ for μ a probability measure, and using (6.18), we find if K' is a g -moderate refinement of K , then

$$h(\mu, K) \leq h(\mu, K') \leq h(\mu, K) + \log g. \tag{6.19}$$

Let $-1 = \kappa_0 < \kappa_1 < \dots < \kappa_{k+1} = 1$ be the end-points of the intervals in a partition K . If $\kappa_{i+1} - \kappa_i \leq \varepsilon$ for each i , then K is an ε -partition; $\text{diam } K \leq \varepsilon$. If also $\kappa_{i+2} - \kappa_i > \varepsilon$ for all i , we say K is a *moderate* ε -partition. For any partition K' , points κ'_i can be deleted to form a moderate ε -partition K . Therefore the infimum in the definition of $h(\mu, \varepsilon)$ can be taken over all moderate ε -partitions.

Suppose K, K' are two moderate ε -partitions. Then each interval in K intersects at most three intervals of K' , and vice-versa. Let $K \cup K'$ be the partition whose intervals are intersections at those of K and those of K' . Then $K \cup K'$ is a 3-moderate refinement of both K, K' . Thus from (6.19)

$$\begin{aligned} h(\mu, K \cup K') - \log 3 &\leq h(\mu, K) \leq h(\mu, K \cup K'), \\ h(\mu, K \cup K') - \log 3 &\leq h(\mu, K') \leq h(\mu, K \cup K'), \end{aligned}$$

and so

$$|h(\mu, K) - h(\mu, K')| \leq \log 3. \tag{6.20}$$

We turn now to a property of β , the reciprocal of a PV number. Let

$$A(\beta) = \sum_{r=0}^N a_r \beta^r \tag{6.21}$$

be a polynomial in β with each $a_r = 0$ or ± 1 . There exists $\gamma > 0$ (depending on β but independent of N) such that if $A(\beta) \neq 0$, then

$$|A(\beta)| \geq \frac{1}{2} \gamma \beta^N.$$

This is the statement of [12, lemma 1.51] for this case. Now consider two numbers x, x' of the form (6.2). Then $x - x' = 2A(\beta)$ for A of the form (6.21). Thus if $x - x' \neq 0$,

$$|x - x'| \geq \gamma \beta^N. \tag{6.22}$$

Let g be the smallest integer $\geq \gamma^{-1}$. Let K be the partition defined by

$$-1 = x_0 < x_1 < \dots < x_k < x_{k+1} = 1,$$

where the x_i are the points of (6.2). Our key claim is that for H_n as in (6.3)

$$|h(\nu_\beta, K) - H_{n-1}| \leq \log 2g. \tag{6.23}$$

To prove (6.23) we let q_i be the measure of x_i in the space P_n . Let $\nu_i^{(n)}$ be one of the components of $T_*^n \nu_\beta$ centred at x_i . By the invariance of ν_β under T_* , we see

$$\nu_\beta = \sum_{i=1}^k q_i \nu_i^{(n)}. \tag{6.24}$$

Let

$$p_{ij} = q_i \nu_i^{(n)}[x_j, x_{j+1}). \tag{6.25}$$

The following are immediate:

$$\sum_j p_{ij} = q_i \quad \text{for each } i, \tag{6.26}$$

$$\sum_i p_{ij} = \nu_\beta[x_j, x_{j+1}) \quad \text{for each } j, \tag{6.27}$$

$$p_{ij} = 0 \quad \text{unless } i - g \leq j \leq i + g - 1. \tag{6.28}$$

Therefore, from (6.18)

$$\begin{aligned} \sum_{i,j} \psi(p_{ij}) - \log 2g &\leq \sum_i q_i \leq \sum_{i,j} \psi(p_{ij}), \\ \sum_{i,j} \psi(p_{ij}) - \log 2g &\leq \sum_j \psi(\nu_\beta[x_j, x_{j+1})) \leq \sum_{i,j} \psi(p_{ij}). \end{aligned}$$

Since $\sum_i q_i = H_{n-1}$ and $\sum_j \psi(\nu_\beta[x_j, x_{j+1})) = h(\nu_\beta, K)$, we obtain (6.23).

For $\beta \geq \frac{1}{2}$ any open set has positive ν_β measure. The length of the support interval of any $\mu_j^{(n)}$ is $2\beta^n$. Therefore

$$x_{j+1} - x_j \leq 2(1 - \beta)\beta^n \leq \beta^n.$$

Thus K is an ε -partition. By deleting some of the x_j , it can be coarsened to a moderate β^n partition K' . Moreover K is a g -moderate refinement of K' . For any other moderate β^n -partition K' , we combine (6.18), (6.20) and (6.23) to obtain

$$|h(\nu_\beta, \beta^n) - H_{n-1}| \leq \log 6g^2.$$

Thus

$$\left| \frac{h(\nu_\beta, \beta^n)}{\log \beta^{-n}} - \frac{H_{n-1}}{n \log \beta^{-1}} \right| \leq \frac{\log 6g^2}{n \log \beta^{-1}}. \tag{6.29}$$

As before we can take the limit over β^n instead of ε and we find from (6.29) that

$$|d_{\text{info}} - G_\beta| = 0.$$

This proves (6.9). □

7. Numerical calculations

In the last section we have seen that one plus a certain limit of entropies defined by Garsia equals the information dimension of T_β when β is the reciprocal of a PV number. Moreover Garsia showed this limit is strictly less than one; thus the information dimension is strictly less than 2. Numerical calculations were made to determine this dimension in the particular case $\beta = (-1 + \sqrt{5})/2 \approx 0.61803399$. In this case our best estimate is

$$d_{\text{info}}(T_\beta) \approx 1.99571312 \pm 5 \times 10^{-8}.$$

Calculations for other such β were not so definitive. However there is some indication that for β the root of $x^3 + x^2 + x - 1$ ($\beta \approx 0.54368981$), the value of $d_{\text{info}}(T_\beta)$ is about 1.98.

For $\beta = -1 + \sqrt{5}/2$, the finite probability space P_n has a straightforward combinatorial structure. Computer calculations were done up through $n = 30(2^{30} > 10^9)$. The results are exhibited in tabular form so the convergence may be observed.

The first column is n . The second column is the number of distinct points in the finite probability space P_{n-1} . Note that if one is added to each of these entries, the

$$\beta = \frac{-1 + \sqrt{5}}{2}$$

n	Number of points in P_{n-1}	$\frac{H_{n-1}}{n \log \beta^{-1}}$	Estimate of $G_\beta = d_{\text{info}}(\nu_\beta)$
1	2	1.4404200	1.4404200
2	4	1.4404200	1.4404200
3	7	1.3203850	1.0803150
4	12	1.2603675	1.0803151
5	20	1.2107650	1.0123551
6	33	1.17769684	1.01235605
7	54	1.15217936	0.99907446
8	88	1.13304124	0.99907447
9	143	1.11785968	0.99640716
10	232	1.10571443	0.99640716
11	376	1.09572759	0.99585917
12	609	1.08740522	0.99585916
13	986	1.08035438	0.99574435
14	1 596	1.07431081	0.99574436
15	2 583	1.06907141	0.99571990
16	4 180	1.06448695	0.99571990
17	6 764	1.06044151	0.99571462
18	10 945	1.05684558	0.99571461
19	17 710	1.05362809	0.99571345
20	28 656	1.05073236	0.99571346
21	46 367	1.04811240	0.99571320
22	75 024	1.04573062	0.99571320
23	121 392	1.04355595	0.99571315
24	196 417	1.04156250	0.99571314
25	317 810	1.03972852	0.99571313
26	514 228	1.03803562	0.99571313
27	832 039	1.03646812	0.99571313
28	1 346 268	1.03501259	0.99571313
29	2 178 308	1.03365743	0.99571312
30	3 524 577	1.03239262	0.99571312

Fibonacci sequence obtains. That this is so is straightforward from the combinatorial structure of P_{n-1} (and is related to the fact that β is the golden ratio). The values in this column were computed on the computer and serve as an error check on the program. For $n \geq 6$ the values in the third and fourth column were computed in double precision and truncated to 8 decimal places. The value $H_{n-1}/n \log \beta^{-1}$ in the third column is exact up to round-off error (which should not affect the exhibited digits). The value in the fourth column is an extrapolation to $n = \infty$. If the entries in this column are denoted G_{n-1} , the formula is

$$G_{n-1} = \frac{H_{n-1}}{\log \beta^{-1}} - \frac{H_{n-2}}{\log \beta^{-1}} \tag{7.1}$$

The heuristic justification is the following. We know that

$$\frac{H_{n-1}}{\log \beta^{-1}} = nG_\beta + o(n).$$

Assume that the error term is the same at the n th and $(n-1)$ st row. That is

$$\begin{aligned} \frac{H_{n-2}}{\log \beta^{-1}} &= (n-1)G_\beta + e, \\ \frac{H_{n-1}}{\log \beta^{-1}} &= nG_\beta + e. \end{aligned} \tag{7.2}$$

Solving for G_β , we obtain (7.1). As a check on the consistency of this procedure, we note that if e is calculated from (7.2), its value varies slowly from 1.100305+ at $n = 12$ to 1.10038504 at $n = 29$ and 1.10038499 at $n = 30$. These last two values differ by 5×10^{-8} and it is not unreasonable to accept this difference as an estimate of the error.

For similar numerical calculations to estimate capacities of sets, see [13].

At the rate of convergence in the table, the value of $H_{n-1}/n \log \beta^{-1}$ drops below 1 at the 257th iteration (which requires, in the heuristics of the introduction, an acuity of about 4 parts in 10^{54} in discern).

REFERENCES

- [1] J. Balatoni & A. Rényi. Remarks on entropy, (in Hungarian with English and Russian summaries). *Publ. of Math. Inst. Hungarian Acad. Sci.* (1) (1956), 9-40. (See also A. Rényi, Dimension, entropy and information, *Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, (1957), 545-556; see also A. Rényi, On the dimension and entropy of probability distributions, *Acta Math. Hung.* **10** (1959), 193-215.)
- [2] G. Benettin, L. Galgani, A. Giorgilli & J-M. Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems; a method for computing all of them, Part 1: Theory. *Meccanica* **15** (1980), 9-20; Part 2: Numerical application, *ibid.*, 21-29.
- [3] P. Billingsley. *Probability and Measure*. J. Wiley, 1979.
- [4] R. Bowen. A model for Couette flow data. In *Turbulence Seminar*, Springer-Verlag Lecture Notes in Mathematics #615, (1977), 117-134.
- [5] R. Bowen & D. Ruelle. The ergodic theory of axiom-A flows. *Invent. Math.* **29** (1975), 181-202.
- [6] A. Boyarsky. On the significance of absolutely continuous invariant measures. Preprint: Concordia Univ., 1982.
- [7] P. Erdős. On a family of symmetric Bernoulli convolutions. *Amer. J. Math.* **61** (1939), 974-976.
- [8] P. Erdős. On the smoothness properties of a family of Bernoulli convolutions. *Amer. J. Math.* **62** (1940), 180-186.
- [9] D. Farmer. Chaotic attractors of an infinite dimensional dynamical system. *Phys. D*, **4** (1982), to appear.
- [10] P. Frederickson, J. L. Kaplan, E. D. Yorke & J. A. Yorke. The Lyapunov dimension of strange attractors. *J. Differential Equations*, to appear, 1982.
- [11] A. M. Garsia. Arithmetic properties of Bernoulli convolutions. *Trans. Amer. Math. Soc.* **102** (1962), 409-432.
- [12] A. M. Garsia. Entropy and singularity of infinite convolutions. *Pacific J. Math.* **13** (1963), 1159-1169.
- [13] J. E. Hanson, E. Ott & D. A. Russell. Dimensionality and Lyapunov numbers of strange attractors. *Phys. Rev. Lett.* **45** (1980), 1175-1178.
- [14] B. Jessen & A. Wintner. Distribution functions and the Riemann zeta function. *Trans. Amer. Math. Soc.* **38** (1938), 48-88.

- [15] J. P. Kahane. Sur la distribution de certaines series aleatoires. *Bull. Soc. Math. France* **25** (1971), 119–122.
- [16] J. L. Kaplan & J. A. Yorke. Chaotic behaviour of multidimensional difference equations. In *Functional Differential Equations and Approximation of Fixed Points*, (H-O. Peitgen and H-O. Walther, eds.). Springer-Verlag Lecture Notes in Mathematics #730 (1979), 223–237.
- [17] J. L. Kaplan, J. Mallet-Paret & J. A. Yorke. The Lyapunov dimension of a nowhere differentiable torus. *Ergod. Th. & Dynam. Sys.* To appear.
- [18] T. Kawata. *Fourier Analysis in Probability Theory*. Academic Press, 1972.
- [19] G. Keller. Ergodicité et mesures invariantes pour les transformations dilantes par morceaux d'une région bornée du plan. *C.R. Acad. Sc. Paris*, **289** (1979), 625–627.
- [20] F. Ledrappier. Some relations between dimension and Lyapunov exponents. *Comm. Math. Phys.* **81** (1981), 229–238.
- [21] V. I. Oseledec. A multiplicative ergodic theorem, Lyapunov characteristic numbers for dynamical systems. *Trudy Moskov. Math. Obsch.* **19** (1968), 179–210. (English translation: *Trans. Mosc. Math. Soc.* **19** (1968), 197–231.)
- [22] R. Salem. A remarkable class of algebraic integers, proof of a conjecture of Vijayaraghavan. *Duke Math. J.* **11** (1944), 103–108.
- [23] R. Salem. *Algebraic Numbers and Fourier Analysis*. Heath., 1963.
- [24] C. L. Siegel. Algebraic integers whose conjugates lie in the unit circle. *Duke Math. J.* **11** (1944), 597–602.
- [25] P. Walters. *Ergodic Theory*. Springer-Verlag Lecture Notes in Mathematics # 458, Springer-Verlag: Berlin–Heidelberg–New York, (1975).
- [26] A. Wintner. On convergent Poisson convolutions. *Amer. J. Math.* **57** (1935), 827–838.
- [27] L-S. Young. Dimension, entropy and Lyapunov exponents. *Ergod. Th. & Dynam. Sys.* **2** (1982), 109–124.