Falconer's Formula for the Hausdorff Dimension of a Self–Affine Set in $\mathbf{R}^2$

Irene Hueter* Purdue University

Steven P. Lalley† Purdue University

March 1993

Abstract

Simple sufficient conditions are given for the validity of a formula of Falconer [3] describing the Hausdorff dimension of a self-affine set. These conditions are natural (and easily checked) geometric restrictions on the actions of the affine mappings determining the self-affine set. It is also shown that under these hypotheses the self-affine set supports an invariant Gibbs measure whose Hausdorff dimension equals that of the set.

1 Introduction and Main Results

Let $A_1, A_2, \ldots, A_K$ be a finite set of contractive, affine, invertible self–mappings of $\mathbf{R}^2$. A compact subset Λ of $\mathbf{R}^2$ is said to be *self–affine* with *affinities* $A_1, A_2, \ldots, A_K$ if

$$\Lambda = \bigcup_{i=1}^{K} A_i(\Lambda). \tag{1}$$

It is known [8] that for any such set of contractive affine mappings there is a unique (compact) SA set with these affinities. When the affine mappings $A_1, A_2, \ldots, A_K$ are similarity transformations, the set Λ is said to be *self–similar*. Self–similar sets are well–understood,at least when the images $A_i(\Lambda)$ have "small" overlap: there is a simple and explicit formula for the Hausdorff and box dimensions [12, 10]; these are always equal; and the $\delta-$dimensional Hausdorff measure of such a set (where δ is the Hausdorff dimension) is always positive and finite.

Self–affine sets in general are not so well understood, however, and what is known makes clear that much more complex behavior is possible. The Hausdorff and box dimensions may be different [7, 11]; the $\delta-$dimensional Hausdorff measure need not be positive and finite [7, 13]; and for a smoothly parametrized family of self–affine sets the Hausdorff dimension need not vary continuously with the parameters [3, 7, 14]. On the other hand [3] there is reason to believe that "most" SA sets are not so badly behaved, and indeed that the various "bad" behaviors tend to occur together [7]. But this is all very much speculative:

*Supported by the Swiss National Science Foundation

†Supported by National Science Foundation Grant DMS-9307855

the values of the Hausdorff and box dimensions are exactly known in relatively few, and decidedly special cases. Formulas for the Hausdorff dimension, in particular, are known only for SA sets for which the matrix parts of the affinities are simultaneously diagonalizable.

We begin by describing the SA set with affinities $A_1, A_2, \ldots, A_K$. Let $\mathcal{A} = \{1, 2, \ldots, K\}$ and let $\mathcal{A}^{\mathbf{N}}$ be the set of all (one–sided) infinite sequences from the alphabet $\mathcal{A}$. There is a natural mapping $\pi : \mathcal{A}^{\mathbf{N}} \to \Lambda$ defined as follows: for any sequence $\mathbf{i} = (i_1 i_2 \ldots) \in \mathcal{A}^{\mathbf{N}}$,

$$\pi(\mathbf{i}) = \lim_{n\to\infty} A_{i_1} A_{i_2} \ldots A_{i_n} y, \tag{2}$$

where y is any point of $\mathbf{R}^2$. Since the mappings A_i are (strictly) contractive, the limit exists and is independent of y for every sequence $\mathbf{i}$ in $\mathcal{A}^{\mathbf{N}}$. Moreover, the mapping π is continuous. The SA sets to which our results apply will be totally disconnected, and the mapping π will be a homeomorphism. We may be somewhat cavalier about identifying points of $\mathcal{A}^{\mathbf{N}}$ with points of the SA set Λ, measures on $\mathcal{A}^{\mathbf{N}}$ with their projections on Λ, etc. Observe that there is a natural dynamical system on Λ suggested by the homeomorphism π: let $F : \Lambda \to \Lambda$ be defined by $F = \pi \circ \sigma \circ \pi^{-1}$. This is an expansive K–to–1 mapping of Λ (provided Λ is totally disconnected).

For a nonsingular linear transformation T of $\mathbf{R}^2$ the *singular values* $\alpha(T) \geq \beta(T)$ are defined to be half the lengths of the major and minor axes of the ellipse $T\mathbf{K}$, where $\mathbf{K}$ is the unit circle in $\mathbf{R}^2$. The *singular value function* $\phi^s(T)$ is defined by

$$\phi^s(T) = \begin{cases} \alpha(T)^s & \text{if } 0 \leq s \leq 1 \\ \alpha(T)\beta(T)^{s-1} & \text{if } 1 \leq s \leq 2. \end{cases} \tag{3}$$

For a given finite collection $\mathcal{T} = \{T_1, T_2, \ldots, T_K\}$ of nonsingular linear transformations, define

$$d = d(\mathcal{T}) = \inf\{s : \sum_{n=1}^{\infty} \sum_{\mathcal{A}^n} \phi^s(T_{i_1} T_{i_2} \ldots T_{i_n}) < \infty\} \tag{4}$$

where $\mathcal{A}^n$ is the set of all sequences of length n from the alphabet $\mathcal{A}$. It is easily established that for any such collection $\mathcal{T}$ of nonsingular linear transformations, d is positive and finite. We shall call $d(\mathcal{T})$ the "Falconer dimension" of the collection $\mathcal{T}$.

The main result of [3] is as follows. Let $\mathcal{T} = \{T_1, T_2, \ldots, T_K\}$ be a set of contractive, invertible linear transformations of $\mathbf{R}^2$, each of norm less than 1/3, and let $\mathbf{a} = (a_1, a_2, \ldots, a_K)$ be a vector of K points of $\mathbf{R}^2$. Define $\Lambda = \Lambda(\mathbf{a})$ to be the self–affine set with affinities $\{A_1, A_2, \ldots, A_K\}$, where

$$A_i x = T_i x + a_i. \tag{5}$$

Then the Hausdorff and box dimensions of Λ are bounded above by the Falconer dimension $d = d(T_1, T_2, \ldots, T_K)$; and for almost every $\mathbf{a} \in (\mathbf{R}^2)^K$ the Hausdorff and box dimensions of $\Lambda(\mathbf{a})$ are *equal* to d. Unfortunately, Falconer's proof does not give any information as to which $\mathbf{a}$ the formula applies.

The Falconer dimension d is known to equal the *box* dimension for a large class of *connected* SA sets: see, e.g., [5] and [2]. The SA sets considered in this paper are totally disconnected, so these results are inapplicable. The paper [5] also gives a *lower* bound for the Hausdorff dimension of a totally disconnected SA set. The lower bound is in general strictly less than d, however, and in particular, under the hypotheses we will state presently.

Our main result gives sufficient conditions for Falconer's formula to be valid. As above, let $\mathcal{T} = \{T_1, T_2, \ldots, T_K\}$ be a set of contractive, invertible linear transformations of $\mathbf{R}^2$. We make the following hypotheses abouvt these linear transformations:

Hypothesis 1 *(Contractivity) Each $T_i \in \mathcal{T}$ has matrix norm less than 1.*

Hypothesis 2 *(Distortion) Each $T_i \in \mathcal{T}$ satisfies the inequality $\alpha(T_i)^2 < \beta(T_i)$.*

Hypothesis 3 *(Separation) Let $\mathcal{Q}_2$ be the closed second quadrant of $\mathbf{R}^2$; then the sets $T_i^{-1}(\mathcal{Q}_2)$ are pairwise disjoint subsets of $Interior(\mathcal{Q}_2)$.*

Hypothesis 4 *(Orientation) Each $T_i \in \mathcal{T}$ has positive determinant.*

Our main results actually hold without the fourth hypothesis, but many of the arguments become quite cluttered without it. Observe that hypotheses 3-4 imply that each of the matrices T_i maps the closed *first* quadrant into its interior. Consequently, each T_i has (strictly) positive entries. It will be apparent that our results remain valid when the second quadrant is replaced by any other angular sector, because conjugation by an invertible linear transformation does not affect either Hausdorff or box dimension.

Let $\mathbf{a} = (a_1, a_2, \ldots, a_K)$ be a vector of K points of $\mathbf{R}^2$, and let A_i be the affine transformation of $\mathbf{R}^2$ defined by (5) above. We shall restrict attention to vectors $\mathbf{a}$ satisfying the

Hypothesis 5 *(Closed Set Condition) There exists a bounded open set $\mathcal{V}$ such that the images $\overline{A_1\mathcal{V}}, \overline{A_2\mathcal{V}}, \ldots, \overline{A_K\mathcal{V}}$ are pairwise disjoint closed subsets of $\mathcal{V}$.*

This is equivalent to assuming that the (compact) sets $A_i\Lambda$ are pairwise disjoint, by an elementary argument. This is also a hypothesis for Falconer's [5] lower bound for the Hausdorff dimension of Λ. It also implies that the projection π from sequence space onto Λ is a homeomorphism, and hence that Λ is a totally disconnected set. Observe that the set of vectors $\mathbf{a}$ for which the Closed Set Condition holds is an open subset of $(\mathbf{R}^2)^K$.

Theorem 1.1 *Let $\delta_H(\Lambda)$ and $\delta_B(\Lambda)$ be the Hausdorff and box dimensions of Λ. If Hypotheses 1-5 hold then*

$$\delta_H(\Lambda) = \delta_B(\Lambda) = d < 1. \tag{6}$$

Moreover, the $d-$dimensional Hausdorff measure of the set Λ satisfies

$$\mathcal{H}_d(\Lambda) < \infty. \tag{7}$$

For any Borel probability measure ν on the SA set Λ the Hausdorff dimension of ν is defined to be the supremum of the set of Hausdorff dimensions of Borel subsets of Λ with $\nu-$measure 1. It is not known in general whether a compact, invariant set Λ for an expansive mapping F must support an $F-$ invariant probability measure whose Hausdorff dimension equals the Hausdorff dimension of Λ. (For Axiom A mappings this is not true: one can easily construct counterexamples using a "horseshoe mapping".)

Theorem 1.2 *Under Hypotheses 1-5, there exists a unique ergodic $F-$invariant probability measure whose Hausdorff dimension is d. This measure is the image under the projection π of a Gibbs state on $\mathcal{A}^{\mathbf{N}}$.*

The Gibbs state μ will be described in the next section. It will play a central role in the proof of Theorem 1.1.

The key hypotheses in our theorems are the Distortion and Separation hypotheses 2 and 3. The rationale for such hypotheses may not be immediately clear, nor may it be apparent to which sets of matrices $\mathcal{T}$ they apply. Observe, however, that both are easily checked. To check the separation hypothesis one needs only compute the action of each matrix in the collection on the unit vectors $(0,1)$ and $(-1,0)$ and then compute the angles these images make with the positive x-axis. To check the distortion hypothesis one need only compute the eigenvalues of the matrices $T_i^t T_i$. Note that for any two by two matrix M there exists a constant $C > 0$ such that for all $0 < c < C$ the matrix cM satisfies the distortion hypothesis (this is because c multiplies both singular values of M). Note also that if a given collection of two by two matrices satisfies the separation hypothesis, then "neighboring" collections also satisfy it; thus, the set of collections $\mathcal{T}$ satisfying hypotheses 1-4 is open in the natural topology.

Example: The matrices $M_1^{-1}, M_2^{-1}, M_3^{-1}$ given by

$$\begin{pmatrix} 2 & -1 \\ -4 & 4 \end{pmatrix}, \qquad \begin{pmatrix} 4 & -3 \\ -4 & 4 \end{pmatrix}, \qquad \begin{pmatrix} 4 & -4 \\ -2 & 3 \end{pmatrix},$$

are such that M_1, M_2, M_3 satisfy the separation hypothesis. Hence, any constant multiples of these matrices satisfy the separation hypothesis. Multiplication by $1/30$ is sufficient to force the distortion hypothesis. The resulting collection of matrices T_1, T_2, T_3 given by

$$\begin{pmatrix} \frac{1}{30} & \frac{1}{120} \\ \frac{1}{30} & \frac{1}{60} \end{pmatrix}, \qquad \begin{pmatrix} \frac{1}{30} & \frac{1}{40} \\ \frac{1}{30} & \frac{1}{30} \end{pmatrix}, \qquad \begin{pmatrix} \frac{1}{40} & \frac{1}{30} \\ \frac{1}{60} & \frac{1}{30} \end{pmatrix}$$

satisfies hypotheses 1-4.

2 Background: Thermodynamic Formalism

The proof of the main result will rely on standard results from the theory of Gibbs states and thermodynamic formalism, as developed in [1, 15]. In this section we review some of the salient features of this theory.

Define

$$\begin{aligned}
\mathcal{A} &= \{1, 2, \ldots, K\}; \\
\mathcal{A}^* &= \cup_{n=0}^{\infty} \mathcal{A}^n; \\
\mathcal{A}^{\mathbf{N}} &= \{\text{1-sided infinite sequences from } \mathcal{A}\}; \\
\mathcal{A}^{\mathbf{Z}} &= \{\text{2-sided infinite sequences from } \mathcal{A}\}; \\
\overline{\mathcal{A}^*} &= \mathcal{A}^* \cup \mathcal{A}^{\mathbf{N}}; \\
\sigma &= \text{shift on } \overline{\mathcal{A}^*} \text{ or } \mathcal{A}^{\mathbf{Z}}
\end{aligned}$$

For any two sequences $\mathbf{i},\mathbf{j} \in \overline{\mathcal{A}^*}$ define the distance $d(\mathbf{i},\mathbf{j})$ to be 2^{-m}, where m is the index of the first entry where $\mathbf{i}$ and $\mathbf{j}$ differ. A function f with domain $\mathcal{A}^{\mathbf{N}}$ or $\overline{\mathcal{A}^*}$ is said to be *Hölder continuous* if it is Hölder continuous for some exponent with respect to this metric. There is an analogous metric on $\mathcal{A}^{\mathbf{Z}}$, and a corresponding notion of Hölder continuity. Observe that any function on $\mathcal{A}^{\mathbf{N}}$ may be considered also a function on $\mathcal{A}^{\mathbf{Z}}$, and that Hölder continuity on $\mathcal{A}^{\mathbf{N}}$ implies Hölder continuity on $\mathcal{A}^{\mathbf{Z}}$. Moreover, any Hölder continuous function f on $\mathcal{A}^{\mathbf{N}}$ can be extended to a Hölder continuous function on $\overline{\mathcal{A}^*}$ with the same sup and Hölder norms: e.g., if f is real–valued, define, for any finite sequence $i_1 i_2 \dots i_n$,

$$f(\mathbf{i}) = \sup\{f(\mathbf{i}') : \, i'_j = i_j \, \forall j \leq n\}.$$

We will call a Hölder continuous extension of $f : \mathcal{A}^{\mathbf{N}} \to \mathbf{R}$ to $f : \overline{\mathcal{A}^*} \to \mathbf{R}$ a *completion* of f. Note that there are many completions of any Hölder continuous function on $\mathcal{A}^{\mathbf{N}}$.

Given a real-valued function f with domain $\mathcal{A}^{\mathbf{N}}$, $\overline{\mathcal{A}^*}$, or $\mathcal{A}^{\mathbf{Z}}$ define

$$S_n f = f + f \circ \sigma + f \circ \sigma^2 + \ldots + f \circ \sigma^{n-1}.$$

Observe that if $d(\mathbf{i},\mathbf{j}) \leq 2^{-n}$ then $|S_n f(\mathbf{i}) - S_n f(\mathbf{j})| \leq C_f$ for a constant C_f independent of $n, \mathbf{i}, \mathbf{j}$. Two Hölder continuous functions f and g are said to be *cohomologous* if there exists a Hölder continuous function h such that $f - g = h - h \circ \sigma$. If f, g are cohomologous then there exists a constant $C < \infty$ such that $|S_n f - S_n g| \leq C$ for all $n \geq 1$; and conversely, if there exists such a constant, then f and g are cohomologous (see [1], Th. 1.28).

For any sequence $\mathbf{i} \in \mathcal{A}^*$ or $\mathcal{A}^{\mathbf{Z}}$, let $\mathbf{i}^*$ denote the reversed sequence. For any function f with domain $\mathcal{A}^*$ or $\mathcal{A}^{\mathbf{Z}}$, let f^* be the "reverse" function, i.e., $f^*(\mathbf{i}) = f(\mathbf{i}^*)$. Observe that the operation $*$ is an isometric involution of the space of Hölder continuous functions on $\mathcal{A}^{\mathbf{Z}}$.

Say that a finite sequence $\mathbf{i}$ is a *prefix* of another sequence $\mathbf{i}'$ if the length n of $\mathbf{i}$ is no greater than that of $\mathbf{i}'$ and $i_j = i'_j$ for all $j \leq n$. We will also say that $\mathbf{i}'$ is an *extension* of $\mathbf{i}$, and write $\mathbf{i} \preceq \mathbf{i}'$. Note that if φ is a strictly negative function on $\mathcal{A}^*$ and $\mathbf{i} \preceq \mathbf{i}'$ then $S_n\varphi(\mathbf{i}) > S_{n'}\varphi(\mathbf{i}')$ (here n and n' denote the lengths of the sequences $\mathbf{i}$ and $\mathbf{i}'$, respectively).

Given two sequences f_n, g_n of nonnegative functions on $\mathcal{A}^{\mathbf{N}}$ or, more generally, on any domain, write

$$f_n \asymp g_n$$

if there exist constants $0 < c_1 < c_2 < \infty$ such that $c_1 f_n(\mathbf{i}) \leq g_n(\mathbf{i}) \leq c_2 f_n(\mathbf{i})$ for all arguments $\mathbf{i}$ and all positive integers n. Similar notation will be used for functions parametrized by positive numbers ρ: e.g., $f_\rho \asymp g_\rho$ for $\rho > 0$. In general, when the notation is used it should be understood that the implied constants are independent of any arguments or parameters on which the functions might depend.

Gibbs States and Pressure

For any Hölder continuous function f on $\mathcal{A}^{\mathbf{Z}}$ there exists a constant $P(f)$ and a unique $\sigma-$invariant probability measure μ_f on the Borel sets of $\mathcal{A}^{\mathbf{Z}}$ such that for each $\mathbf{i} \in \mathcal{A}^{\mathbf{Z}}$,

$$\mu_f(\Gamma(\mathbf{i})) \asymp \exp\{S_n f(\mathbf{i}) - nP(f)\} \tag{8}$$

where n is the length of $\mathbf{i}$ and $\Gamma(\mathbf{i})$ is the cylinder set $\Gamma(\mathbf{i}) = \{\mathbf{j} \in \mathcal{A}^{\mathbf{N}} : \mathbf{i} \preceq \mathbf{j}\}$ (see [1]). The measure μ_f is called the *Gibbs state* with potential function f, and the constant $P(f)$ is called the *pressure*.

If two functions are cohomologous then they have the same pressure and the same Gibbs state. The pressure functional is monotone and continuous: if $f < g$ then $P(f) < P(g)$, and for any Hölder continuous f the function $P(af)$ is a continuous function of the scalar a (see [15]). The pressure functional commutes with the involution $*$, i.e., $P(f) = P(f^*)$. For every Hölder continuous f and for every integer $n \geq 1$ the pressure functional satisfies $P(S_n f) = nP(f)$. Of key importance to us is that if $f < 0$ then there exists a unique constant $\delta > 0$ such that

$$P(\delta f) = 0 \tag{9}$$

(see [9]).

The pressure and the entropy of the Gibbs state are related to each other by the *Variational Principle* (see [1], Th. 1.22). This implies that

$$h(\mu_{\delta f}) = P(\delta f) + h(\mu_{\delta f}) = -\int \delta f \, d\mu_{\delta f} \tag{10}$$

where $h(\mu_{\delta f})$ denotes the entropy of the measure $\mu_{\delta f}$.

Counting Problems and Thermodynamic Formalism

Let $f : \overline{\mathcal{A}^*} \to (-\infty, 0)$ be a Hölder continuous function such that, for some integer $n \geq 1$,

$$S_n f < 0. \tag{11}$$

Note that this property is actually determined by the restriction $f|\mathcal{A}^{\mathbf{N}}$ of f to $\mathcal{A}^{\mathbf{N}}$, in particular, if $S_n f < 0$ on $\mathcal{A}^{\mathbf{N}}$, then for any completion of $f|\mathcal{A}^{\mathbf{N}}$, there exists a positive integer k such that $S_{kn} f < 0$ on $\overline{\mathcal{A}^*}$. For any function f satisfying (11) there exists a unique $\delta > 0$ such that $P(\delta f) = 0$. For $0 < \rho < 1$, define

$$\mathcal{A}^*(\rho) = \bigcup_{n=1}^{\infty} \{\mathbf{i} \in \mathcal{A}^n : S_n f(\mathbf{i}) \leq \log \rho \text{ and } S_k f(\mathbf{i}) > \log \rho \; \forall k < n\}. \tag{12}$$

Thus, $\mathcal{A}^*(\rho)$ consists of finite sequences of possibly different lengths n such that $S_n f$ takes a value just below $\log \rho$. Note that since f is bounded, for any $\mathbf{i} \in \mathcal{A}^*(\rho)$, $S_n f(\mathbf{i})$ differs from $\log \rho$ by at most $\|f\|_\infty < \infty$. For every $\mathbf{i} \in \mathcal{A}^{\mathbf{N}}$ there exists a unique n such that the *finite* sequence $i_1 i_2 \ldots i_n$ is an element of $\mathcal{A}^*(\rho)$.

Proposition 2.1 *Let δ be the unique positive number such that $P(\delta f) = 0$. Then, as $\rho \to 0$,*

$$\#\mathcal{A}^*(\rho) \asymp \rho^{-\delta}. \tag{13}$$

Observe that the set $\mathcal{A}^*(\rho)$ is defined solely in terms of the function $f|\mathcal{A}^*$, but that the asymptotic behavior of the cardinality is determined by $f|\mathcal{A}^{\mathbf{N}}$. Thus, relation (13) is valid for *every* completion of $f|\mathcal{A}^{\mathbf{N}}$. Stronger statements than (13) are proved in [9]: see Th. 1 and Th. 3. However, a much simpler and more direct proof can be given.

Proof: Each element $\mathbf{i}$ of the set $\mathcal{A}^*(\rho)$ determines a "cylinder set" $\Gamma(\mathbf{i})$ of $\mathcal{A}^{\mathbf{N}}$, to wit, $\Gamma(\mathbf{i}) = \{\mathbf{j} \in \mathcal{A}^{\mathbf{N}} : \mathbf{i} \preceq \mathbf{j}\}$. The cylinder sets $\{\Gamma(\mathbf{i}) : \mathbf{i} \in \mathcal{A}^*(\rho)\}$ are pairwise disjoint and their union is the entire sequence space $\mathcal{A}^{\mathbf{N}}$. Hence,

$$\sum_{\mathbf{i} \in \mathcal{A}^*(\rho)} \mu_{\delta f}(\Gamma(\mathbf{i})) = 1.$$

By the defining property of a Gibbs state, there are constants $0 < C_1 \leq C_2 < \infty$ such that for every $\mathbf{i} \in \mathcal{A}^*(\rho)$ the measure of the cylinder set $\Gamma(\mathbf{i})$ satisfies

$$C_1 \exp\{\delta S_n f(\mathbf{i})\} \leq \mu_{\delta f}(\Gamma(\mathbf{i})) \leq C_2 \exp\{\delta S_n f(\mathbf{i})\}$$

where n is the length of $\mathbf{i}$. But by the defining property of $\mathcal{A}^*(\rho)$, there exists a constant $0 < C_3 \leq 1$ such that for every $\mathbf{i} \in \mathcal{A}^*(\rho)$,

$$C_3 \rho \leq \exp\{S_n f(\mathbf{i})\} \leq \rho.$$

Combining the last three displayed formulas gives

$$C_4 \sum_{\mathbf{i} \in \mathcal{A}^*(\rho)} \rho^\delta \leq 1 \leq C_2 \sum_{\mathbf{i} \in \mathcal{A}^*(\rho)} \rho^\delta,$$

for a suitable constant C_4; this proves the proposition. ///

Proposition 2.2 *"Most" sequences in $\mathcal{A}^*(\rho)$ are approximately $\mu_{\delta f}-$distributed. More precisely, for every Hölder continuous function $g : \overline{\mathcal{A}^*} \to \mathbf{R}$, every $0 < t \leq 1$, and every $\varepsilon > 0$, there exists $0 < \eta < \delta$ such that*

$$\#\{\mathbf{i} = i_1 i_2 \ldots i_n \in \mathcal{A}^*(\rho) : \max_{0 \leq t \leq 1} \left| \frac{S_{[nt]} g(\mathbf{i})}{n} - t \int g \, d\mu_{\delta f} \right| > \varepsilon\} = O(\rho^{-\eta}). \tag{14}$$

The proof is accomplished by the same techniques as used in the proof of Th. 6 in [9]. (There only the case t=1 is proved.) We shall not give the details. For proving that the box dimension equals the Falconer dimension d only the weaker estimate $o(\rho^{-\delta})$ is needed; but for the proof that the Hausdorff dimension equals d the stronger exponential estimates are needed. (It is not difficult to derive the weaker estimate $o(\rho^{-\delta})$ from Birkhoff's ergodic theorem for the measure $\mu_{\delta f}$. However, the "large deviations" type exponential bounds (14) seem to require more of the thermodynamic formalism, specifically, properties of the Ruelle operators.)

Let

$$n_\rho = \frac{\log \rho}{\int f \, d\mu_{\delta f}}. \tag{15}$$

Corollary 2.3 *"Most" sequences in $\mathcal{A}^*(\rho)$ have lengths between $n_\rho(1 - \varepsilon)$ and $n_\rho(1 + \varepsilon)$.*

Corollary 2.3 and Proposition 2.2 suggest that the set $\mathcal{A}^*(\rho)$ is in some appropriate sense "close" to the set of sequences of length n_ρ that are approximately "generic" for the measure $\mu_{\delta f}$. By the Shannon-McMillan-Breiman theorem, the cardinality of the latter set is $\approx e^{h n_\rho}$, where h is the entropy of the measure $\mu_{\delta f}$; the cardinality of the former is given by Proposition 2.1 . This is consistent with the variational principle (10), which implies that

$$\rho^{-\delta} = e^{n_\rho h}.$$

3 Products of Positive Matrices

Let $\mathcal{T} = \{T_1, T_2, \ldots, T_K\}$ be a set of invertible, strictly contractive 2×2 matrices with strictly positive entries. Any finite product of matrices taken from $\mathcal{T}$ is again a matrix with strictly positive entries. In this section we will present some properties of such products. There is a large literature devoted to the theory of *random* matrix products, beginning with [6]; however, our needs require that we relate the "Lyapunov exponents" of such products to the thermodynamic formalism discussed in the previous section, so we must begin from scratch.

Any invertible 2×2 matrix T induces a mapping $\hat{T}$ of projective space $\mathcal{P}$, the space of lines through the origin in $\mathbf{R}^2$. If T has (strictly) positive entries then $\hat{T}(\mathcal{P}_+) \subset \mathcal{P}_+$ and $\hat{T}^{-1}(\mathcal{P}_-) \subset \mathcal{P}_-$, where $\mathcal{P}_+$ and $\mathcal{P}_-$ are the sets of lines with positive and negative slopes, respectively. (The lines of slope 0 and slope ∞ are included in both $\mathcal{P}_+$ and $\mathcal{P}_-$, so they are both closed subsets of $\mathcal{P}$.) Moreover, $\hat{T}|\mathcal{P}_+$ and $\hat{T}^{-1}|\mathcal{P}_-$ are strictly contractive relative to the natural metric on $\mathcal{P}$ (the distance between two lines being the smaller angle between them).

For any sequence $\mathbf{i} \in \overline{\mathcal{A}^*}$ of length $|\mathbf{i}| \geq n$ define the matrix products

$$\begin{aligned} \Phi_n(\mathbf{i}) &= T_{i_1} T_{i_2} \ldots T_{i_n}; && (16) \\ \Psi_n(\mathbf{i}) &= T_{i_1}^{-1} T_{i_2}^{-1} \ldots T_{i_n}^{-1}. && (17) \end{aligned}$$

Observe that $\Psi_n(\mathbf{i}^*) = \Phi_n(\mathbf{i})^{-1}$. Let $\hat{\Phi}_n$ and $\hat{\Psi}_n$ be the corresponding mappings of projective space.

Proposition 3.1 *There exist constants $C > 0$ and $0 < r < 1$ and Hölder continuous functions $V : \overline{\mathcal{A}^*} \to \mathcal{P}_+$ and $W : \overline{\mathcal{A}^*} \to \mathcal{P}_-$ such that for every $\mathbf{i} \in \overline{\mathcal{A}^*}$,*

$$\begin{aligned} \mathit{diameter}(\hat{\Phi}_n(\mathbf{i})(\mathcal{P}_+)) &\leq Cr^n; && (18) \\ \mathit{diameter}(\hat{\Psi}_n(\mathbf{i})(\mathcal{P}_-)) &\leq Cr^n; && (19) \\ \lim_{n\to\infty} \hat{\Phi}_n(\mathbf{i})(\mathcal{P}_+) &= \{V(\mathbf{i})\}; && (20) \\ \lim_{n\to\infty} \hat{\Psi}_n(\mathbf{i})(\mathcal{P}_-) &= \{W(\mathbf{i})\}. && (21) \end{aligned}$$

Proof: This follows immediately from the fact that the induced operators $\hat{T}_i$ and $\hat{T}_i^{-1}$ are strictly contractive on $\mathcal{P}_+$ and $\mathcal{P}_-$, respectively. ///

Let $\mathcal{Q}_+$ and $\mathcal{Q}_-$ be the sets of unit vectors in the closed first and second quadrants, respectively. These vectors serve as representatives of the positive and negative arcs $\mathcal{P}_+$ and $\mathcal{P}_-$. In the following we will not always be careful to distinguish between elements of $\mathcal{Q}_\pm$ and $\mathcal{P}_\pm$. In particular,we will let $V(\mathbf{i})$ and $W(\mathbf{i})$ also denote the unit vectors in the first and second quadrants representing the lines $V(\mathbf{i})$ and $W(\mathbf{i})$; the meaning should be clear from context. It follows from the preceding proposition that for every $\mathbf{i} \in \mathcal{A}^{\mathbf{N}}$ the vector $T_{i_1}V(\sigma\mathbf{i})$ is a scalar multiple of $V(\mathbf{i})$; consequently,we may define a function $\varphi : \mathcal{A}^{\mathbf{N}} \to \mathbf{R}$ by

$$T_{i_1}V(\sigma\mathbf{i}) = e^{\varphi(\mathbf{i})}V(\mathbf{i}). \tag{22}$$

By the Hölder continuity of the function V, φ is also Hölder continuous. Moreover, iterating the defining relation gives

$$\Phi_n(\mathbf{i})V(\sigma^n\mathbf{i}) = e^{S_n\varphi(\mathbf{i})}V(\mathbf{i}). \tag{23}$$

Proposition 3.2 *The function φ is strictly negative on $\mathcal{A}^{\mathbf{N}}$. Therefore, there exists a unique $\delta > 0$ such that $P(\delta\varphi) = 0$.*

Proof: Since each T_i has matrix norm less than 1, and since each $V(\mathbf{i})$ is a unit vector, $T_{i_1}V(\sigma\mathbf{i})$ has norm strictly smaller than that of $V(\mathbf{i})$. Consequently, $\varphi(\mathbf{i}) < 0$. The existence and uniqueness of δ now follow from the general considerations of the previous section. ///

For all $\mathbf{i} \in \mathcal{A}^{\mathbf{N}}$ and integers $n \geq 1$ define

$$\begin{aligned}
\Phi'_n(\mathbf{i}) &= \min\{|\Phi_n(\mathbf{i})u| : u \in \mathcal{Q}_+\}; && (24)\\
\Phi^*_n(\mathbf{i}) &= \min\{|\Phi_n(\mathbf{i})^{-1}u| : u \in \mathcal{Q}_-\}; && (25)\\
\Psi'_n(\mathbf{i}) &= \min\{|\Psi_n(\mathbf{i})u| : u \in \mathcal{Q}_-\}; && (26)
\end{aligned}$$

and define Φ''_n, Φ^{**}_n, and Ψ''_n to be the corresponding functions with *min* replaced by *max*. Observe that $\Phi'_n(\mathbf{i}), \Phi''_n(\mathbf{i})$, etc. , depend only on the first n entries of $\mathbf{i}$. Also, $\Phi^*_n(\mathbf{i}) = \Psi'_n(\mathbf{i}^*)$ and $\Phi^{**}_n(\mathbf{i}) = \Psi''_n(\mathbf{i})$ for $\mathbf{i} \in \mathcal{A}^*$.

Proposition 3.3

$$\Phi'_n \asymp \Phi''_n \asymp e^{S_n\varphi}. \tag{27}$$

Proof: This is a routine consequence of (18) and (20) in Proposition 3.1. ///

There are a number of useful completions of the function φ. One completion is defined in terms of the first singular value α as follows:

$$\begin{aligned}
\varphi(\mathbf{i}) &= \log\alpha(\Phi_n(\mathbf{i})) - \log\alpha(\Phi_{n-1}(\sigma\mathbf{i})) && \text{if } |\mathbf{i}| = n > 1\\
&= \log\alpha(\Phi_1(\mathbf{i})) && \text{if } |\mathbf{i}| = 1\\
&= 0 && \text{if } |\mathbf{i}| = 0.
\end{aligned} \tag{28}$$

Notice that for every $\mathbf{i} \in \mathcal{A}^*$ and every $n \geq 1$,

$$\alpha(\Phi_n(\mathbf{i})) = e^{S_n\varphi(\mathbf{i})}. \tag{29}$$

Also, since the matrices T_i are strictly contractive, the function φ defined by (28) is strictly negative on $\mathcal{A}^*$ and hence, by Proposition 3.2, on $\overline{\mathcal{A}^*}$.

Proposition 3.4 *The function φ is Hölder continuous on $\overline{\mathcal{A}^*}$.*

Proof: Let $\mathbf{i} \in \mathcal{A}^*$ be a finite sequence of length n, and let $\mathbf{i}' \in \mathcal{A}^{\mathbf{N}}$ be an infinite sequence such that $\mathbf{i}'$ is an extension of $\mathbf{i}$. We will show that the difference between $\varphi(\mathbf{i})$ and $\varphi(\mathbf{i}')$ is less than Cs^n for some constants $C < \infty$ and $0 < s < 1$ not depending on n or $\mathbf{i}, \mathbf{i}'$.

If M is a matrix with positive entries, then the major axis of the ellipse $M\mathbf{K}$ ends at the point Mu where u is the unit vector that maximizes $\|Mu\|$. Since M has positive entries so must the vector u. It follows that the major axis of $\Phi_n(\mathbf{i})\mathbf{K}$ ends at a point $T_{i_1}\Phi_{n-1}(\sigma\mathbf{i})u$ where u is a unit vector in $\mathcal{Q}_+$. Similarly, the major axis of $\Phi_{n-1}(\sigma\mathbf{i})\mathbf{K}$ ends at a point $\Phi_{n-1}(\sigma\mathbf{i})u'$ where u' is another unit vector in $\mathcal{Q}_+$. Now the vectors $\Phi_{n-1}(\sigma\mathbf{i})u$ and $\Phi_{n-1}(\sigma\mathbf{i})u'$ are both in $\Phi_{n-1}(\sigma\mathbf{i})\mathcal{Q}_+$, which is an angular sector of aperture smaller than Cr^{n-1}, by Proposition 3.1. This angular sector also contains the vector $V(\sigma\mathbf{i}')$.

Let v and v' be the unit vectors in the directions $\Phi_{n-1}(\sigma\mathbf{i})u$ and $\Phi_{n-1}(\sigma\mathbf{i})u'$, respectively. Then v, v' and $V(\sigma\mathbf{i}')$ are all unit vectors contained in an arc of the unit circle of length smaller than Cr^{n-1}. Moreover,

$$\begin{aligned} e^{\varphi(\mathbf{i}')} &= |T_{i_1}V(\sigma\mathbf{i}')|; \\ e^{\varphi(\mathbf{i})} &= |T_{i_1}v| \left| \frac{\Phi_{n-1}(\sigma\mathbf{i})u}{\Phi_{n-1}(\sigma\mathbf{i})u'} \right|. \end{aligned}$$

Since the unit vectors $v, V(\sigma\mathbf{i}')$ are at distance less than Cr^{n-1} it follows from the Lipschitz character of T_i that $|T_{i_1}V(\sigma\mathbf{i}')|$ and $|T_{i_1}v|$ differ by less than $C'r^{n-1}$ for a suitable constant C'. Thus, to complete the proof we must show that $|\Phi_{n-1}(\sigma\mathbf{i})u|/|\Phi_{n-1}(\sigma\mathbf{i})u'|$ differs from 1 by less than $C''s^n$.

Observe first that the ratio is less than one, because the vector $\Phi_{n-1}(\sigma\mathbf{i})u'$ is the endpoint of the major axis of the ellipse $\Phi_{n-1}(\sigma\mathbf{i})\mathbf{K}$. Recall that the *directions* of the vectors $\Phi_{n-1}(\sigma\mathbf{i})u'$ and $\Phi_{n-1}(\sigma\mathbf{i})u$ differ by less than Cr^{n-1}. But if the ratio of the lengths were greater than $C'''r^{n-1}$,for sufficiently large C''', then by the Lipschitz continuity of T_i the length of $T_i\Phi_{n-1}(\sigma\mathbf{i})u'$ would be greater than that of $T_i\Phi_{n-1}(\sigma\mathbf{i})u$, contradicting the fact that $T_i\Phi_{n-1}(\sigma\mathbf{i})u$ is the major axis of $\Phi_n(\mathbf{i})\mathbf{K}$.

///

Corollary 3.5

$$\alpha(\Phi_n) \asymp e^{S_n\varphi} \ \textit{on}\ \mathcal{A}^{\mathbf{N}}. \tag{30}$$

Proof: This follows routinely from the exact equality in (29) and the fact that φ is Hölder continuous on $\overline{\mathcal{A}^*}$. ///

Another completion of φ will be useful in sections 5 and 6 below. Let U be an arbitrary bounded open subset of $\mathbf{R}^2$, and for $\mathbf{i} \in \mathcal{A}^*$ define

$$\begin{aligned} \varphi(\mathbf{i}) &= \log(\text{diameter}\Phi_n(\mathbf{i})U) - \log(\text{diameter}\Phi_{n-1}(\sigma\mathbf{i})U) && \text{if } |\mathbf{i}| = n > 1 \\ &= \log(\text{diameter}\Phi_n(\mathbf{i})U) && \text{if } |\mathbf{i}| = 1 \\ &= 0 && \text{if } |\mathbf{i}| = 0. \end{aligned} \tag{31}$$

For every $\mathbf{i} \in \mathcal{A}^*$ and every $n \geq 1$,

$$\text{diameter}(\Phi_n(\mathbf{i})U) = e^{S_n\varphi(\mathbf{i})}. \tag{32}$$

Proposition 3.6 *The function $\varphi : \mathcal{A}^* \to \mathbf{R}$ defined by (31) is a (Hölder continuous) completion of $\varphi : \mathcal{A}^{\mathbf{N}} \to \mathbf{R}$.*

The proof is similar to that of Proposition 3.4, and uses the fact that there are concentric discs Δ_1 and Δ_2 such that $\Delta_1 \subseteq U \subseteq \Delta_2$. The condition that U be a bounded open set is essential to the validity of the proposition: if, for instance, U were a line segment, then the function φ defined by (31) would no longer necessarily be a completion.

For this and the next section, φ will denote the completion defined by (28).

Define another function $\psi : \overline{\mathcal{A}^*} \to \mathbf{R}$ as follows:

$$\psi(\mathbf{i}) = \log\det\Phi_1(\mathbf{i}) = \log\det T_{i_1} \tag{33}$$

if $|\mathbf{i}| \geq 1$, and $\psi(\mathbf{i}) = 0$ if $|\mathbf{i}| = 0$. Observe that ψ is a function only of the first entry of $\mathbf{i}$, hence is Hölder continuous. Moreover, since each of the matrices T_i is strictly contractive, the determinants are less than one, so $\psi < 0$. Note that for every sequence $\mathbf{i}$ of length at least n,

$$\det\Phi_n(\mathbf{i}) = e^{S_n\psi(\mathbf{i})}, \tag{34}$$

$$\det\Psi_n(\mathbf{i}) = e^{-S_n\psi(\mathbf{i})}. \tag{35}$$

Recall that for a 2×2 matrix M the second singular value is denoted by $\beta(M)$. It is the length of the *minor* axis of the ellipse $M\mathbf{K}$.

Proposition 3.7

$$\beta(\Phi_n) \asymp \exp\{S_n\psi - S_n\varphi\}. \tag{36}$$

Proof: The area of the ellipse $\Phi_n(\mathbf{i})\mathbf{K}$ is

$$\pi\det\Phi_n(\mathbf{i}) = \pi e^{S_n\psi(\mathbf{i})} = \pi\alpha(\Phi_n(\mathbf{i}))\beta(\Phi_n(\mathbf{i})).$$

The result therefore follows from Corollary 3.5. ///

Proposition 3.8

$$\Psi_n' \asymp \Psi_n'' \asymp \exp\{S_n\varphi^* - S_n\psi\}. \tag{37}$$

Proof: Recall that φ^* is the "reverse" function to φ, also that $\Psi_n'(\mathbf{i}) = \Phi_n^*(\mathbf{i}^*)$ and $\Psi_n''(\mathbf{i}) = \Phi_n^{**}(\mathbf{i}^*)$. Thus, it suffices to show that $\Phi_n^*(\mathbf{i})$ and $\Phi_n^{**}(\mathbf{i})$ are comparable ($\asymp$) to $\exp\{S_n\varphi(\mathbf{i}) - S_n\psi(\mathbf{i})\}$. Note that by (19) of Proposition 3.1, $\Phi_n^*(\mathbf{i}) \asymp \Phi_n^{**}(\mathbf{i})$; consequently, it suffices to show that for *some* unit vector $u \in \mathcal{Q}_-$, $|\Phi_n(\mathbf{i})^{-1}u|$ is comparable to $\exp\{S_n\varphi(\mathbf{i}) - S_n\psi(\mathbf{i})\}$. But this follows from the previous proposition: just let u be the unit vector in the direction of the minor axis of the ellipse $\Phi_n(\mathbf{i})\mathbf{K}$ (observe that the major axis has positive slope, since the matrices $\Phi_n(\mathbf{i})$ have positive entries, so the vector u does indeed lie in the second quadrant).

///

4 Consequences of Hypotheses 1–4

Assume henceforth that the matrices $T_1, T_2, \ldots, T_K$ satisfy Hypotheses 1–4 of section 1. Recall that Hypotheses 3 and 4 guarantee that the matrices all have positive entries. Hypothesis 1 states that the matrices are all contractive. Consequently, all results of the previous section are applicable. We will continue to use the notation established there.

Let $\mathbf{i} = i_1 i_2 \ldots i_n$ be an element of $\mathcal{A}^*$ of length n. Define

$$\mathcal{P}_-(\mathbf{i}) = \hat{\Psi}_n(\mathbf{i})\mathcal{P}_- = \hat{T}_{i_1}^{-1}\hat{T}_{i_2}^{-1}\cdots\hat{T}_{i_n}^{-1}\mathcal{P}_- \ . \tag{38}$$

Under Hypothesis 3 these are, for a fixed n, pairwise disjoint closed subsets of $\mathcal{P}_-$. Moreover, they are naturally nested: if $\mathbf{i}'$ is an extension of $\mathbf{i}$ then $\mathcal{P}_-(\mathbf{i}') \subseteq \mathcal{P}_-(\mathbf{i})$. Notice that

$$\mathcal{W} = \{W(\mathbf{i}) : \mathbf{i} \in \mathcal{A}^{\mathbf{N}}\} = \bigcap_{n=1}^{\infty} \bigcup_{\mathbf{i} \in \mathcal{A}^n} \mathcal{P}_-(\mathbf{i}) \tag{39}$$

is a Cantor subset of $\mathcal{P}_-$, hence has Hausdorff and box dimensions ≤ 1.

Define the *diameter function* Δ_n on $\mathcal{A}^{\mathbf{N}}$ as follows: for $\mathbf{i} = i_1 i_2 \ldots \in \mathcal{A}^{\mathbf{N}}$, set

$$\Delta_n(\mathbf{i}) = \text{diameter}(\mathcal{P}_-(i_1 i_2 \ldots i_n)). \tag{40}$$

Proposition 4.1

$$\Delta_n \asymp \exp\{S_n\psi - 2S_n\varphi^*\}. \tag{41}$$

Proof: Let e_1 and e_2 be the unit vectors (0,1) and (-1,0), respectively. By Proposition 3.8, the matrix $\Psi_n(\mathbf{i})$ maps e_1 and e_2 to vectors v_1 and v_2 of lengths comparable ($\asymp$) to $\exp\{S_n\varphi^*(\mathbf{i}) - S_n\psi(\mathbf{i})\}$. In addition, the matrix $\Psi_n(\mathbf{i})$ maps the set of unit vectors $\mathcal{Q}_-$ to an arc of an ellipse; the endpoints of this arc are v_1 and v_2. The angular sector bounded by this arc and the two segments connecting the origin to v_1 and v_2 has area

$$(\pi/4)|\det \Psi_n(\mathbf{i})| = (\pi/4)e^{-S_n\psi(\mathbf{i})}.$$

It follows that the distance between v_1 and v_2 is comparable to $\exp\{-S_n\varphi^*(\mathbf{i})\}$ (because it is comparable to area/length). Projecting the line segment connecting v_1 and v_2 onto the unit circle $\mathcal{Q}_-$ gives the set $\mathcal{P}_-(i_1 i_2 \ldots i_n)$; since v_1 and v_2 both have lengths comparable to $\exp\{S_n\varphi^*(\mathbf{i}) - S_n\psi(\mathbf{i})\}$, the length of the projection $\mathcal{P}_-(i_1 i_2 \ldots i_n)$ is comparable to $\exp\{S_n\psi(\mathbf{i}) - 2S_n\varphi^*(\mathbf{i})\}$. ///

Proposition 4.2 *There exists an integer $n \geq 1$ sufficiently large that*

$$S_n\psi - 2S_n\varphi < 0 \ \textit{on} \ \mathcal{A}^{\mathbf{N}} \tag{42}$$

and

$$S_n\psi - 3S_n\varphi > 0 \ \textit{on} \ \mathcal{A}^{\mathbf{N}}. \tag{43}$$

Proof: (a) First observe that as $n \to \infty$ the ratio $\alpha(\Phi_n(\mathbf{i}))/\beta(\Phi_n(\mathbf{i}))$ converges to ∞ uniformly on $\mathcal{A}^{\mathbf{N}}$. This is a consequence of Propositions 3.1 and 3.3. Proposition 3.3 guarantees that the image of the quarter circle $\mathcal{Q}_+$ under the mapping $\Phi_n(\mathbf{i})$ consists of points all with Euclidean norms comparable to $\exp\{S_n\varphi(\mathbf{i})\}$, and Proposition 3.1, equation (18) guarantees that this image is an arc of an ellipse of length $\leq Cr^n e^{S_n\varphi(\mathbf{i})}$ for some $r < 1$. Consequently, the area of the angular sector bounded by the arc $\Phi_n(\mathbf{i})\mathcal{Q}_+$ and the two line segments joining the endpoints of this arc to the origin is bounded above by $C'r^n e^{2S_n\varphi(\mathbf{i})}$ for a suitable constant $C' < \infty$ independent of n and $\mathbf{i}$. But the area of this angular sector is also given by $(\pi/4)\exp\{S_n\psi(\mathbf{i})\}$, by the determinant formula. Therefore,

$$\exp\{S_n\psi(\mathbf{i}) - 2S_n\varphi(\mathbf{i})\} \leq C''r^n.$$

This implies the first statement of the proposition.

(b) This is where the "Distortion Hypothesis" 2 is used. This hypothesis implies that for a certain constant $0 < s < 1$ the singular values of the matrices T_i satisfy $\alpha(T_i)^2/\beta(T_i) < s$. Hence, for every sequence $\mathbf{i} \in \mathcal{A}^{\mathbf{N}}$ and every integer $n \geq 1$ the singular values of the matrix $\Phi_n(\mathbf{i})$ satisfy $\alpha(\Phi_n(\mathbf{i}))^2/\beta(\Phi_n(\mathbf{i})) \leq s^n$. The estimates for $\alpha(\Phi_n(\mathbf{i}))$ and $\beta(\Phi_n(\mathbf{i}))$ given in Corollary 3.5 and Proposition 3.7 now yield the second statement of the proposition.

///

Corollary 4.3 *If μ is any $\sigma-$invariant probability measure on $\mathcal{A}^{\mathbf{N}}$ then*

$$3\int \varphi\, d\mu < \int \psi\, d\mu < 2\int \varphi\, d\mu < 0.$$

Proof: The first two inequalities follow immediately from the shift invariance of the probability measure μ and the result of the preceding proposition. The last of the three inequalities holds because $\varphi < 0$ on the space $\mathcal{A}^{\mathbf{N}}$ of *infinite* sequences. ///

Corollary 4.4 *Let $\delta > 0$ and $\delta_* > 0$ be the unique real numbers satisfying $P(\delta\varphi) = 0$ and $P(\delta_*(\psi - 2\varphi)) = 0$. Then*

$$0 < \delta < \delta_*. \tag{44}$$

Proof: The existence , uniqueness, and positivity of δ have already been established (see Proposition 3.2). The existence, uniqueness, and positivity of δ_* are proved as follows. By the preceding proposition there exists $n \geq 1$ such that $S_n(\psi - 2\varphi) < 0$ on $\mathcal{A}^{\mathbf{N}}$. Thus, there exists a unique $\delta_* > 0$ such that $P(\delta_* S_n(\psi - 2\varphi)/n) = 0$. But for any θ, $P(\theta S_n(\psi - 2\varphi)/n) = P(\theta(\psi - 2\varphi))$.

To show that $\delta < \delta_*$, we use the second statement of the last proposition together with the monotonicity of the pressure functional. Take n sufficiently large that $S_n\psi - 2S_n\varphi > S_n\varphi$ on $\mathcal{A}^{\mathbf{N}}$: the monotonicity of the pressure implies that

$$P(\delta_* S_n\varphi/n) < P(\delta_*(S_n\psi - 2S_n\varphi)/n) = 0.$$

But $P(\theta S_n\varphi/n)$ is a continuous, nonincreasing function of θ; therefore, the unique value δ of θ such that $P(\theta S_n\varphi/n) = 0$ must be smaller than δ_*. ///

Proposition 4.5

$$0 < \delta < 1. \tag{45}$$

Proof: By the preceding corollary, $\delta < \delta_*$, where δ_* is the unique real number satisfying $P(\delta_*(\psi - 2\varphi)) = 0$. Hence it suffices to show that $\delta_* \leq 1$. We will accomplish this by showing that δ_* is no larger than the box dimension of $\mathcal{W}$; since $\mathcal{W}$ is contained in $\mathcal{P}$ its box dimension cannot be larger than 1.

To show that δ_* is less or equal to the box dimension of $\mathcal{W}$ it suffices to show that for any small ρ a covering of $\mathcal{W}$ by $\rho-$balls (intervals) must have at least $O(\rho^{-\delta_*})$ elements. Consider the collection of arcs $\mathcal{C} = \{\mathcal{P}_-(\mathbf{i}) : \mathbf{i} \in \mathcal{A}^*(\varepsilon\rho)\}$, where $\mathcal{A}^*(\rho)$ is as defined by (12), with $f = \psi - 2\varphi_*$, and where ε is a constant whose value will be specified shortly. For any $\rho > 0, \varepsilon > 0$, $\mathcal{C}$ is a covering of $\mathcal{W}$ by *pairwise disjoint* arcs of lengths comparable ($\asymp$) to $\exp\{S_n\psi - 2S_n\varphi^*\}$ (by Proposition 4.1), each having nonempty intersection with $\mathcal{W}$. By the definition of the set $\mathcal{A}^*(\varepsilon\rho)$,

$$S_n\psi - 2S_n\varphi^* \geq \log(\varepsilon\rho);$$

hence, if ε is chosen sufficiently large, each arc in the collection will have length $\geq \rho$. Hence, *any* covering $\mathcal{C}'$ of $\mathcal{W}$ by arcs of length ρ must have cardinality at least $\frac{1}{3}$ that of $\mathcal{C}$, because any arc in $\mathcal{C}'$ can intersect at most 3 of the arcs in $\mathcal{C}$.

By Proposition 2.1, the cardinality of $\mathcal{C}$ is comparable to $\rho^{-\delta_{**}}$, where δ_{**} is the unique real number satisfying $P(\delta_{**}(\psi - 2\varphi_*)) = 0$. Now recall that the pressure function commutes with the involution $*$, and $\psi = \psi^*$ because $\psi(\mathbf{i})$ depends only on the initial entry of the sequence $\mathbf{i}$; consequently, $\delta_* = \delta_{**}$.

///

Proposition 4.6

$$\delta = d = d(T_1, T_2, \ldots, T_K). \tag{46}$$

Proof: Recall that $d = d(T_1, T_2, \ldots, T_K)$ is defined to be the infimum of the set of all $s > 0$ such that $\sum_{\mathcal{A}^*} \alpha(\Phi_n(\mathbf{i}))^s < \infty$ *provided* this infimum is ≤ 1. We will show that the inf is in fact equal to δ. By the preceding proposition, $0 < \delta < 1$.

Let $\mathcal{A}^*(\rho)$ be defined by (12) with $f = \varphi$ and $/zv$ as defined in (28). For any $\mathbf{i} \in \mathcal{A}^*(\rho)$ of length n, the value of $\alpha(\Phi_n(\mathbf{i}))^s$ is comparable to ρ^s, by Corollary 3.5 and the definition of $\mathcal{A}^*(\rho)$. For any $\rho > 0$, $\sum_{\mathcal{A}^*} \alpha(\Phi_n(\mathbf{i}))^s \geq \sum_{\mathcal{A}^*(\rho)} \alpha(\Phi_n(\mathbf{i}))^s$. Moreover, by Proposition 2.1, the cardinality of $\mathcal{A}^*(\rho)$ is comparable to $\rho^{-\delta}$. Thus, choosing ρ small, one finds that $\sum_{\mathcal{A}^*} \alpha(\Phi_n(\mathbf{i}))^s \geq O(\rho^{s-\delta})$. Letting $\rho \to 0$ shows that if $s < \delta$ then

$$\sum_{\mathcal{A}^*} \alpha(\Phi_n(\mathbf{i}))^s = \infty.$$

To complete the proof we must show that for every $s > \delta$, $\sum_{\mathcal{A}^*} \alpha(\Phi_n(\mathbf{i}))^s < \infty$. For this we will argue that for some constant $C < \infty$,

$$\sum_{\mathcal{A}^*} \alpha(\Phi_n(\mathbf{i}))^s < C \sum_{n=1}^{\infty} \sum_{\mathcal{A}^*(2^{-n})} 2^{-ns}.$$

This last sum is finite, because the cardinality of $\mathcal{A}^*(2^{-n})$ is comparable to $2^{n\delta}$.

Choose any $\mathbf{i} \in \mathcal{A}^*$; there exist a unique integer $n \geq 1$ and unique sequences $\mathbf{i}' \in \mathcal{A}^*(2^{-n})$ and $\mathbf{i}'' \in \mathcal{A}^*(2^{-n-1})$ such that $\mathbf{i}' \preceq \mathbf{i} \preceq \mathbf{i}''$, where $\mathbf{i} \preceq \mathbf{i}'$ indicates that $\mathbf{i}'$ is an extension of $\mathbf{i}$. Moreover, the difference $|\mathbf{i}''| - |\mathbf{i}'|$ in lengths is bounded by a constant C' independent of $\mathbf{i}$, because by Proposition 3.2 φ is strictly negative on $\overline{\mathcal{A}^*}$ and so the partial sums $S_n\varphi$ decrease by a definite negative amount with each increment of n . Consequently, for a given pair $\mathbf{i}' \in \mathcal{A}^*(2^{-n}), \mathbf{i}'' \in \mathcal{A}^*(2^{-n-1})$ there are at most $K^{C'}$ sequences $\mathbf{i}$ such that $\mathbf{i}' \preceq \mathbf{i} \preceq \mathbf{i}''$. Finally, because $\varphi < 0$ on $\mathcal{A}^*$ and $\alpha(\Phi_n)^s = \exp\{sS_n\varphi\}$ on $\mathcal{A}^*$, for any pair $\mathbf{i}, \mathbf{i}'$ such that $\mathbf{i}' \in \mathcal{A}^*(2^{-n})$ and $\mathbf{i}' \preceq \mathbf{i}$,

$$\alpha(\Phi_{|\mathbf{i}|}(\mathbf{i})) \leq 2^{-n}.$$

Therefore,

$$\sum_{\mathcal{A}^*} \alpha(\Phi_n(\mathbf{i}))^s \leq K^{C'} \sum_{n=1}^{\infty} \sum_{\mathcal{A}^*(2^{-n})} 2^{-ns} < \infty.$$

///

Proposition 4.7 *Let u be a unit vector in $\mathcal{Q}_-$ such that*

$$u \in \mathcal{P}_-(i_m i_{m-1} \dots i_{k+1}); \tag{47}$$
$$u \notin \mathcal{P}_-(i_m i_{m-1} \dots i_k) \tag{48}$$

for some $\mathbf{i} = i_1 i_2 \ldots \in \mathcal{A}^{\mathbf{N}}$, where $1 \leq k < m$, or such that (47) holds with $k = 0$ and (48) holds with $k = m$. Then

$$|\Phi_m(\mathbf{i})u| \asymp \exp\{2S_k\varphi(\mathbf{i}) - S_m\varphi(\mathbf{i}) + S_m\psi(\mathbf{i}) - S_k\psi(\mathbf{i})\}. \tag{49}$$

Here $\asymp$ indicates that the implied constants are independent of $\mathbf{i}, u, m,$ and k.

Proof: This is a consequence of Propositions 3.3 and 3.7 of the preceding section. The second hypothesis implies that $\hat{\Phi}_{m-k}(\sigma^k\mathbf{i})u = \hat{T}_{i_{k+1}}\hat{T}_{i_{k+2}} \dots \hat{T}_{i_m}u$ is in $\mathcal{P}_+$. Consequently, by Proposition 3.3,

$$|\Phi_m(\mathbf{i})u| = |\Phi_k(\mathbf{i})\Phi_{m-k}(\sigma^k\mathbf{i})u| \asymp \exp\{S_k\varphi(\mathbf{i})\}|\Phi_{m-k}(\sigma^k\mathbf{i})u|.$$

On the other hand, the first hypothesis implies that there exists a unit vector v in $\mathcal{P}_-$ such that

$$u = \hat{T}_{i_m}^{-1}\hat{T}_{i_{m-1}}^{-1} \dots \hat{T}_{i_{k+1}}^{-1}v = \hat{\Phi}_{m-k}(\sigma^k\mathbf{i})^{-1}v;$$

hence, by Proposition 3.7

$$|u| = 1 \asymp \exp\{-S_m\varphi(\mathbf{i}) + S_k\varphi(\mathbf{i}) + S_m\psi(\mathbf{i}) - S_k\psi(\mathbf{i})\}|T_{i_m}^{-1}T_{i_{m-1}}^{-1} \dots T_{i_{k+1}}^{-1}v|,$$

which, along with the representation of $\Phi_m(\mathbf{i})u$ given above, proves the proposition.

///

5 Efficient Coverings of Λ

Let Λ be the self–affine set with affinities $A_1, A_2, \dots, A_K$ where $A_i x = T_i x + a_i$. We assume that the linear transformations T_i satisfy the hypotheses 1-4; thus, the results of the preceding sections are applicable. We assume also that the vectors $a_1, a_2, \dots, a_K$ are such that the closed set condition is satisfied, i.e., that there exists a bounded open set U such that $\overline{A_1 U}, \overline{A_2 U}, \dots, \overline{A_K U}$ are pairwise disjoint compact subsets of U. These assumptions have the following consequences:

$$\begin{aligned} D_{\min} &= \min_{i \neq j} \operatorname{distance}(\overline{A_i U}, \overline{A_j U}) > 0; & (50)\\ \Lambda &= \bigcap_{n=1}^{\infty} \bigcup_{\mathcal{A}^n} \overline{A_{i_1} A_{i_2} \dots A_{i_n} U}. & (51) \end{aligned}$$

The closed set condition implies that $\{\cup_{\mathcal{A}^n} \overline{A_{i_1} \dots A_{i_n} U}\}_{n \geq 1}$ is a nested sequence of nonempty compact sets, so the intersection is a nonempty compact set. It is clear that the intersection satisfies (1), so it must be Λ, by the uniqueness of self–affine sets with given affinities. Observe that each of the sets $A_{i_1} A_{i_2} \dots A_{i_n} U$ contains a point of Λ.

Let x_* be a distinguished point of U. For sequences $\mathbf{i}, \mathbf{i}' \in \overline{\mathcal{A}^*}$ of length greater than or equal to n define

$$\begin{aligned} U_n(\mathbf{i}) &= A_{i_1} A_{i_2} \dots A_{i_n} U; \\ x_n(\mathbf{i}) &= A_{i_1} A_{i_2} \dots A_{i_n} x_*; \\ U(\mathbf{i}) &= U_n(\mathbf{i}) \ \text{ if } n = |\mathbf{i}|; \\ x(\mathbf{i}) &= x_n(\mathbf{i}) \ \text{ if } n = |\mathbf{i}|; \\ D(\mathbf{i}, \mathbf{i}') &= \operatorname{distance}(U(\mathbf{i}), U(\mathbf{i}')), \ \mathbf{i}, \mathbf{i}' \in \mathcal{A}^*. \end{aligned}$$

In this section let $\varphi : \mathcal{A}^* \to \mathbf{R}$ be the completion of $\varphi : \mathcal{A}^{\mathbf{N}} \to \mathbf{R}$ defined by (31). Then since translations have no effect on diameters, for any sequence $\mathbf{i}$ of length n,

$$\operatorname{diameter}(U(\mathbf{i})) = e^{S_n \varphi(\mathbf{i})}. \tag{52}$$

Observe that the sets $U(\mathbf{i})$ are nested: if $\mathbf{i} \preceq \mathbf{i}'$ then $U(\mathbf{i}') \subseteq U(\mathbf{i})$. For any collection $\mathcal{C}$ of finite sequences with the property that every infinite sequence has a prefix in $\mathcal{C}$, the collection $\{U(\mathbf{i}) : \ \mathbf{i} \in \mathcal{C}\}$ is an open covering of Λ.

Let $\mathcal{A}^*(\rho)$ be as defined in (12), with $f = \varphi$ and φ the completion defined by (31); thus, $\mathcal{A}^*(\rho)$ consists of all finite sequences $\mathbf{i}$ such that $\operatorname{diameter}(U_n(\mathbf{i})) \leq \rho$ but $\operatorname{diameter}(U_k(\mathbf{i})) > \rho$ for all $1 \leq k < n$, where n denotes the length of $\mathbf{i}$. Then $\mathcal{A}^*(\rho)$ has the covering property described above, so for every $\rho > 0$ the collection $\{U(\mathbf{i}) : \ \mathbf{i} \in \mathcal{A}^*(\rho)\}$ is a covering of the SA set Λ by sets of diameters no larger than ρ. The main objective of this section is to show that these are *efficient* coverings of Λ.

For $\mathbf{i} \in \mathcal{A}^*(\rho)$ define

$$F_\rho(\mathbf{i}) = \{\mathbf{i}' \in \mathcal{A}^*(\rho) : \ D(\mathbf{i}, \mathbf{i}') \leq \rho\}. \tag{53}$$

Proposition 5.1 *For each $\gamma > 0$ there exists $\eta = \eta(\gamma) < \delta$ such that*

$$\#\{\mathbf{i} \in \mathcal{A}^*(\rho) : \ \#F_\rho(\mathbf{i}) \geq \rho^{-\gamma}\} = o(\rho^{-\eta}). \tag{54}$$

The remainder of this section will be devoted to the proof of this proposition.

Henceforth, let $\mu = \mu_{\delta\varphi}$, and for any real–valued continuous function f defined on $\mathcal{A}^{\mathbf{N}}$ let $\overline{f} = \int f\, d\mu$. We will adopt the following notational convention: for any finite sequences $\mathbf{i}, \mathbf{i}'$ the lengths of $\mathbf{i}, \mathbf{i}'$ will be denoted by n, n'. For any $1 > \rho > 0$, $n_\rho = (\log \rho)/\overline{\varphi}$.

Lemma 5.2 *To prove (54) it suffices to prove that for any $\varepsilon > 0$ statement (54) is valid when $\mathcal{A}^*(\rho)$ is replaced by*

$$\mathcal{B}_\varepsilon(\rho) = \{\mathbf{i} \in \mathcal{A}^*(\rho) : \forall\ f = \varphi, \psi \ \textit{and} \ \max_{t\in[0,1]} \left| S_{[tn]} f(\mathbf{i}) - tn\overline{f} \right| \leq n_\rho \varepsilon\} \tag{55}$$

This follows immediately from Proposition 2.2.

For any two sequences $\mathbf{i}, \mathbf{i}'$ define

$$m(\mathbf{i}, \mathbf{i}') = \max\{j : \ i_j = i'_j\}. \tag{56}$$

Suppose that $\mathbf{i} \in \mathcal{B}_\varepsilon(\rho)$ and $\mathbf{i}' \in F_\rho(\mathbf{i})$. By the definition of $m = m(\mathbf{i}, \mathbf{i}')$ the $(m+1)$th coordinates of $\mathbf{i}$ and $\mathbf{i}'$ differ; hence, the points $x(\sigma^m \mathbf{i})$ and $x(\sigma^m \mathbf{i}')$ are in different "first generation" images $A_i U$ of U. Thus, their distance satisfies

$$D_{\min} \leq \text{distance}(x(\sigma^m \mathbf{i}), x(\sigma^m \mathbf{i}')) \leq \text{diameter}(U). \tag{57}$$

Recall (50) that $D_{\min} > 0$. On the other hand, the points $x(\mathbf{i})$ and $x(\mathbf{i}')$ are in the sets $U(\mathbf{i})$ and $U(\mathbf{i}')$, respectively, which are at distance $\leq \rho$, since $\mathbf{i}' \in F_\rho(\mathbf{i})$. Moreover, these sets have diameters no larger than ρ, because $\mathbf{i} \in \mathcal{B}_\varepsilon(\rho)$ and $\mathbf{i}' \in F_\rho(\mathbf{i})$. Consequently, by the triangle inequality,

$$\text{distance}(x(\mathbf{i}), x(\mathbf{i}')) \leq 3\rho. \tag{58}$$

Keep in mind that the points $x(\mathbf{i})$ and $x(\mathbf{i}')$ are the images under the affine mapping $A_{i_1} A_{i_2} \dots A_{i_m}$ of the points $x(\sigma^m \mathbf{i})$ and $x(\sigma^m \mathbf{i}')$, respectively. The effect on distance is a function only of the matrix part of the affine mapping; therefore, if $y = x(\sigma^m \mathbf{i}) - x(\sigma^m \mathbf{i}')$, then by (58), $|\Phi_m(\mathbf{i}) y| \leq 3\rho$. If we now set $u = y/|y|$ then by (57) u is a unit vector satisfying

$$|\Phi_m(\mathbf{i}) u| \leq \kappa\rho \tag{59}$$

where $\kappa = 3/D_{\min}$.

Define $k = k(\mathbf{i}, \mathbf{i}')$ to be the unique integer satisfying $0 \leq k \leq m$ such that (47)-(48) hold (or just (47) if $k = 0$).

Lemma 5.3 *There exist constants $C < \infty$ and $\varepsilon_* > 0$ sufficiently small that for all $0 < \varepsilon < \varepsilon_*$ the following is true. For all $\rho > 0$ sufficiently small, if $\mathbf{i} \in \mathcal{B}_\varepsilon(\rho)$ and $\mathbf{i}' \in F_\rho(\mathbf{i})$ then*

$$2m(\mathbf{i}, \mathbf{i}') - k(\mathbf{i}, \mathbf{i}') \geq n_\rho (1 - C\varepsilon). \tag{60}$$

Proof: The integer k is defined so that equations (47)-(48) hold. Proposition 4.7 gives an estimate for the magnitude of $|\Phi_m(\mathbf{i}) u|$ in terms of k : together with (59) this estimate implies that for an appropriate constant $0 < c < \infty$ independent of $\rho > 0$,

$$\log \rho \geq 2S_k \varphi(\mathbf{i}) - S_m \varphi(\mathbf{i}) + S_m \psi(\mathbf{i}) - S_k \psi(\mathbf{i}) + c.$$

(Note that Proposition 4.7 was proved for another completion φ than the one defined earlier in this section; but since the result stated in the lemma depends only on properties of $\varphi|\mathcal{A}^{\mathbf{N}}$, we may use any completion here.) We now use the fact that $\mathbf{i} \in \mathcal{B}_\varepsilon(\rho)$ to obtain lower bounds for the terms $2S_k\varphi(\mathbf{i}), S_m\varphi(\mathbf{i})$, etc., in terms of the expectations $\overline{\varphi}, \overline{\psi}$: this gives

$$\log \rho \geq (2k - m)\overline{\varphi} + (m - k)\overline{\psi} - c'\varepsilon n_\rho + c$$

for another constant $c' > 0$ ($c' = 16$ should suffice).

Recall (Corollary 4.3) that $\overline{\varphi} < 0$ and that $\overline{\psi}/\overline{\varphi} < 3$. Thus, dividing both sides of the last displayed inequality by $\overline{\varphi}$ changes its direction, giving $n_\rho \leq (2k-m) + (m-k)(\overline{\psi}/\overline{\varphi}) + n_\rho C'\varepsilon + C''$ where $C' = -c'/\overline{\varphi} > 0$ and $C'' = c/\overline{\varphi}$. Since $m - k \geq 0$ and $(\overline{\psi}/\overline{\varphi}) > 0$ the inequality remains valid when the term $(m-k)(\overline{\psi}/\overline{\varphi})$ is replaced by $3(m-k)$. This yields

$$n_\rho \leq 2m - k + n_\rho C'\varepsilon + C''.$$

Since C'' does not depend on ε or ρ, it may be absorbed into the term $n_\rho C\varepsilon$, with $C = 2C'$, provided ρ is sufficiently small.

///

Lemma 5.4 *There exists a constant $C^* < \infty$ such that for all sufficiently small $\rho > 0$ and $\varepsilon > 0$,*

$$\#\{(\mathbf{i}, \mathbf{i}') : \mathbf{i} \in \mathcal{B}_\varepsilon(\rho) \text{ and } \mathbf{i}' \in F_\rho(\mathbf{i})\} \leq \rho^{-\delta - C^*\varepsilon}. \tag{61}$$

Proof: Let $m = m(\mathbf{i}, \mathbf{i}')$. Since $\mathbf{i} \in \mathcal{B}_\varepsilon(\rho)$, the values of $S_m\varphi(\mathbf{i})$ and $S_m\varphi(\mathbf{i}')$ must both be within $n_\rho\varepsilon$ of $m\overline{\varphi}$. Since $\mathbf{i}$ and $\mathbf{i}'$ are both elements of $\mathcal{A}^*(\rho)$, the values of $S_n\varphi(\mathbf{i})$ and $S_{n'}\varphi(\mathbf{i}')$ must both be within a constant of $\log \rho$ (here n and n' denote the lengths of $\mathbf{i}$ and $\mathbf{i}'$). Thus, $S_{n-m}\varphi(\sigma^m \mathbf{i})$ and $S_{n'-m}\varphi(\sigma^m \mathbf{i}')$ are both greater than $(n_\rho - m)\overline{\varphi} - 2n_\rho\varepsilon$ for all sufficiently small $\rho > 0$. It follows that $\sigma^m\mathbf{i}$ and $\sigma^m\mathbf{i}'$ are prefixes of distinct sequences $\mathbf{j}$ and $\mathbf{j}'$ in $\mathcal{A}^*(\rho^{1-(m/n_\rho)+2\varepsilon})$. Consequently, by Proposition 2.1, for any given m the number of admissible pairs $(\mathbf{j}, \mathbf{j}')$ is no larger than $\rho^{-2\delta(1-(m/n_\rho)+4\varepsilon)}$, for all sufficiently small ρ.

Given an admissible pair $(\sigma^m\mathbf{i}, \sigma^m\mathbf{i}')$, consider the possible prefixes $i_1 i_2 \dots i_m$. By the preceding lemma, there is only one allowable string $i_{k+1}i_{k+2}\dots i_m$ for the $(k+1)$th through the mth coordinates, where $k = [2m - n_\rho(1 - C\varepsilon)] + 1$ and C is chosen as in (60). Moreover, the string $i_1 i_2 \dots i_k$ is constrained by the requirement that $\mathbf{i} \in \mathcal{B}_\varepsilon(\rho)$: the sum $S_k\varphi(\mathbf{i})$ must be within $n_\rho\varepsilon$ of $k\overline{\varphi}$. Consequently, $i_1 i_2 \dots i_k$ is a prefix to an element of $\mathcal{A}^*(\rho^{k/n_\rho+\varepsilon})$ (with $f = \varphi$: see (12)). Proposition 2.1 guarantees that the number of allowable strings $i_1 i_2 \dots i_k$ is no greater than $\rho^{-\delta(k/n_\rho+2\varepsilon)}$, for all sufficiently small $\rho > 0$.

Combining the results of the last two paragraphs shows that for all sufficiently small ρ and any integer m satisfying $1 \leq m \leq n$ the number of pairs $(\mathbf{i}, \mathbf{i}')$ such that $\mathbf{i} \in \mathcal{B}_\varepsilon(\rho), \mathbf{i}' \in F_\rho(\mathbf{i})$, and $m(\mathbf{i}, \mathbf{i}') = m$ is less than

$$\exp\{-\delta(\log\rho)(2(1 - m/n_\rho) + k/n_\rho + 6\varepsilon)\} \leq 2\exp\{-\delta(\log\rho)(1 + (C+6)\varepsilon)\}.$$

(The second inequality follows by substituting $k = [2m - n_\rho(1 - C\varepsilon)] + 1$.) The lemma now follows, because the number of integers m between 1 and m is $O(\log\rho)$, which is smaller than $\rho^{-C\varepsilon}$ for sufficiently small ρ.

///

Proof *of Proposition 5.1:* By Lemma 5.2, it suffices to show that for some $\varepsilon > 0$ the cardinality of $\{\mathbf{i} \in \mathcal{B}_\varepsilon(\rho) : \#F_\rho(\mathbf{i}) \geq \rho^{-\gamma}\}$ is $o(\rho^{-\eta})$ for some $\eta = \eta(\gamma) < \delta$. Lemma 5.4 shows that there is a constant C such that for all sufficiently small $\rho, \varepsilon > 0$,

$$\#\{(\mathbf{i}, \mathbf{i}') : \mathbf{i} \in \mathcal{B}_\varepsilon(\rho) \text{ and } \mathbf{i}' \in F_\rho(\mathbf{i})\} = O(\rho^{-\delta - C\varepsilon}).$$

Choose ε sufficiently small that $C\varepsilon < \gamma/2$; then since $\#\mathcal{B}_\varepsilon(\rho) \approx \rho^{-\delta}$, the desired inequality follows, with (say) $\eta = \delta - \gamma/3$.

///

Remark: The key to the preceding argument is Lemma 5.3, which leads to Lemma 5.4. The details of the proofs obscure the roles of the various hypotheses; however, it should be noted that the distortion hypothesis is used in an essential way here. Without it the estimate $\overline{\psi}/\overline{\varphi} < 3$ need not be valid, and without this Lemma 5.3 may fail to hold.

6 Proofs of the Main Results

Proof *of Theorem 1.1:* It suffices to show that $\delta_B(\Lambda) = d$, that the $d-$ dimensional Hausdorff measure of Λ is finite, and that the Hausdorff dimension of Λ is at least δ.

Proof *$\delta_B(\Lambda) = d$:* As noted in the preceding section, for every $\rho > 0$ the collection $\mathcal{U}(\rho) = \{U(\mathbf{i}) : \mathbf{i} \in \mathcal{A}^*(\rho)\}$ is a covering of Λ by sets of diameter no larger than ρ. By Proposition 2.1, the cardinality of $\mathcal{U}(\rho)$ satisfies

$$\#\mathcal{U}(\rho) \asymp \rho^{-\delta}. \tag{62}$$

Consequently, the box dimension of Λ is no larger than δ. By Proposition 4.5, $\delta = d$.

To prove the reverse inequality, we use Proposition 5.1. This result implies that for any $\gamma > 0$ the set of $U(\mathbf{i})$ in $\mathcal{U}(\rho)$ such that there are more than $\rho^{-\gamma}$ other elements of $\mathcal{U}(\rho)$ at distance less than ρ from $U(\mathbf{i})$ has cardinality on the order $o(\#\mathcal{U}(\rho))$. Thus, for any collection of $\rho-$balls whose union contains at least one point of every $U(\mathbf{i})$ in $\mathcal{U}(\rho)$, most of the elements of $\mathcal{U}(\rho)$ have points covered by balls intersecting no more than $\rho^{-\gamma}$ other elements of $\mathcal{U}(\rho)$. Consequently, the cardinality of such a collection must be at least $O(\rho^\gamma \#\mathcal{U}(\rho)) = O(\rho^{-\delta+\gamma})$, by (62). Now (51) and the nesting property of the sets $U(\mathbf{i})$ implies that any covering of Λ must contain a point from every one of the sets in $\mathcal{U}(\rho)$. Therefore, the cardinality of any covering of Λ by $\rho-$balls has cardinality at least $O(\rho^{-\delta+\gamma})$. This proves that the box dimension of Λ must be at least $\delta - \gamma$. Since $\gamma > 0$ is arbitrary, it now follows that the box dimension is at least $\delta = d$.

///

Proof *$\mathcal{H}_d(\Lambda) < \infty$:* The same coverings $\mathcal{U}(\rho)$ can be used. Each $U(\mathbf{i}) \in \mathcal{U}(\rho)$ has diameter less than or equal to ρ, and by (62) the cardinality of the collection is $\asymp \rho^{-\delta}$. Consequently,

$$\sum_{\mathcal{U}(\rho)} \text{diameter}(U(\mathbf{i}))^{\delta} \asymp 1.$$

Since $\rho > 0$ is arbitrary, the definition of outer Hausdorff measure implies that $\mathcal{H}_d(\Lambda) < \infty$.
///

Proof *Proof that $\delta_H(\Lambda) \geq d$:* Recall that the Hausdorff dimension $\delta_H(\nu)$ of a Borel probability measure supported by the set Λ is defined to be the infimum of the set of Hausdorff dimensions of Borel subsets of Λ with $\nu-$measure 1. It is obvious that any probability measure supported by Λ has dimension no larger than $\delta_H(\Lambda)$. Consequently, to show that the Hausdorff dimension of Λ is at least δ it suffices to exhibit a probability measure supported by Λ whose dimension is at least δ. Consider the projection $\pi\mu$ to Λ of the measure $\mu = \mu_{\delta\varphi}$. We will show that the Hausdorff dimension of $\pi\mu$ is δ. For this we quote the following well-known

Lemma 6.1 *If ν is a Borel probability measure on a metric space such that*

$$\liminf_{\rho \to 0} \frac{\log \nu(B_\rho(x))}{\log \rho} = d \qquad \nu - a.e.\ x, \tag{63}$$

then the Hausdorff dimension of ν is at least d.

Here $B_\rho(x)$ denotes the ball of radius ρ centered at the point x. Observe that the limit may be replaced by the limit as $\rho \to 0$ through the inverse powers of 2: i.e., $\rho = 2^{-1}, 2^{-2}, \ldots$. For a proof of the lemma see [16] or [4],Ch. 1, exercise 1.8.

Choose any sequence $\mathbf{i} \in \mathcal{A}^{\mathbf{N}}$, and let $\pi\mathbf{i}$ be the corresponding point of the set Λ. For any $\rho > 0$ there is a unique finite sequence $\mathbf{i}(\rho) \in \mathcal{A}^*(\rho)$ such that $\mathbf{i}$ is an extension of $\mathbf{i}(\rho)$; it is clearly the case that $\pi\mathbf{i} \in U(\mathbf{i}(\rho))$. The diameter of the set $U(\mathbf{i}(\rho))$ is, by construction, $\asymp \rho$. Moreover, since $\mu = \mu_{\delta\varphi}$ is the Gibbs state with potential $\delta\varphi$,

$$\pi\mu(U(\mathbf{i}(\rho))) \asymp \rho^{\delta}. \tag{64}$$

Consequently,

$$\pi\mu(B_\rho(\pi\mathbf{i})) \leq C\rho^{\delta} \# F_\rho(\mathbf{i}(\rho)), \tag{65}$$

where $F_\rho(\mathbf{i}(\rho))$ is the set defined by (53).

We will now argue that for sequences $\mathbf{i} \in \mathcal{A}^{\mathbf{N}}$ "generated" by the probability measure μ, eventually $\# F_\rho(\mathbf{i}(2^{-n}))$ is less than $2^{n\gamma}$, for any $\gamma > 0$. Specifically, we will show that for any $\gamma > 0$,

$$\mu\{\mathbf{i} \in \mathcal{A}^{\mathbf{N}} : \# F_\rho(\mathbf{i}(2^{-n})) \geq 2^{n\gamma}\ i.o.\} = 0. \tag{66}$$

Notice that for any n the cylinder sets $\Gamma(\mathbf{j})$, $\mathbf{j} \in \mathcal{A}^*(2^{-n})$, (see the proof of Proposition 2.1) partition the sequence space $\mathcal{A}^{\mathbf{N}}$ and all have (approximately) equal probabilities ($\asymp 2^{-n\delta}$), by (64). There are approximately $2^{n\delta}$ of these cylinder sets. By Proposition 5.1, the number of these cylinder sets for which $\# F_\rho(\mathbf{i}(2^{-n})) \geq 2^{n\gamma}$ is $o(2^{n\eta})$ for some $\eta < \delta$. Consequently,

$\mu\{\mathbf{i} \in \mathcal{A}^{\mathbf{N}} : \#F_\rho(\mathbf{i}(2^{-n})) \geq 2^{n\gamma}\} = o(2^{-n(\delta-\eta)})$ as $n \to \infty$. Since $\sum_n 2^{-n(\delta-\eta)} < \infty$, (66) follows from the Borel–Cantelli lemma.

It now follows from (65) and (66) that for any $\gamma > 0$,

$$\liminf_{n\to\infty} \frac{\log \pi\mu(B_{2^{-n}}(x))}{-n\log 2} \geq \delta - \gamma \quad \pi\mu - \text{a.e. } x.$$

Since $\gamma > 0$ is arbitrary, this proves that the Hausdorff dimension of $\pi\mu$ is at least δ.

///

Proof *of Theorem 1.2:* We have already shown that the probability measure $\pi\mu$ has Hausdorff dimension at least δ; since we have also shown that the set Λ has Hausdorff dimension no greater than δ it follows that $\delta_H(\pi\mu) = \delta$.

Since $\pi : \mathcal{A}^{\mathbf{N}} \to \Lambda$ is a homeomorphism conjugating F with the shift σ the $F-$invariant probability measures on Λ all have the form $\pi\nu$, where ν is a shift-invariant probability measure on $\mathcal{A}^{\mathbf{N}}$. Moreover, $\pi\nu$ is ergodic iff ν is ergodic. Now if ν is an ergodic shift-invariant probability measure distinct from μ then there exists a Hölder continuous function $g : \mathcal{A}^{\mathbf{N}} \to \mathbf{R}$ such that $\int g\, d\nu \neq \int g\, d\mu$. Define

$$\Lambda_\nu = \{\pi\mathbf{i} : \mathbf{i} \in \mathcal{A}^{\mathbf{N}} \text{ such that } \lim_{n\to\infty} \frac{S_n g(\mathbf{i})}{n} = \int g\, d\nu\}.$$

By Birkhoff's ergodic theorem, Λ_ν is a support set for ν. To show that the Hausdorff dimension of $\pi\nu$ is less than δ it suffices to show that the Hausdorff dimension of Λ_ν is less than δ.

Let $2\varepsilon = |\int g\, d\nu - \int g\, d\mu| > 0$. By Proposition 2.2 there exists a constant $\eta < \delta$ such that (14) holds. For each positive integer k define a covering $\mathcal{V}_k$ of Λ_ν by sets of diameters no larger than 2^{-k} as follows:

$$\mathcal{V}_k = \bigcup_{m=k}^{\infty} V_m$$

where

$$V_m = \{U(\mathbf{i}) : \mathbf{i} = i_1 i_2 \dots i_n \in \mathcal{A}^*(2^{-m}) \text{ such that } \left|\frac{S_n g(\mathbf{i})}{n} - \int g\, d\mu\right| > \varepsilon\}.$$

That each $\mathcal{V}_k$ is a covering of Λ_ν follows from the ergodic theorem and the definition of ε. Furthermore, by definition of $\mathcal{A}^*(\rho)$ and (52), together with the estimate (14) for the cardinality of V_m, for any τ satisfying $\eta < \tau < \delta$,

$$\sum_{\mathcal{V}_k} \text{diameter} U(\mathbf{i})^\tau = O(\sum_{m=k}^{\infty} 2^{-m\tau} \# V_m) = O(\sum_{m=k}^{\infty} 2^{-m\tau} 2^{m\eta}).$$

The implied constant is independent of k. This shows that the outer $\tau-$ dimensional Hausdorff measure of Λ_ν is finite.

///

References

[1] R. Bowen (1975). *Equilibrium States and the ErgodicTheorem of Anosov Diffeomorphisms*. Springer Lecture Notes **470**

[2] T. Bedford and M. Urbanski: The box and Hausdorff dimension of self-affine sets. *Ergodic Th. Dynamical Systems* **10** (1990), 627-644.

[3] K.J. Falconer: The Hausdorff dimension of self-affine fractals, *Math. Proc. Camb. Phil. Soc.* **103** (1988), 339–350.

[4] K.J. Falconer (1985) *The Geometry of Fractal Sets*. Cambridge University Press.

[5] K.J. Falconer: The dimension of self-affine fractals II, *Math. Proc. Camb. Phil. Soc.* **111** (1992), 169-179.

[6] H. Furstenberg and H. Kesten: Products of random matrices. *Annals of Mathematical Statistics* **31** (1960) 457-469.

[7] D. Gatzouras and S.P. Lalley: Hausdorff and box dimensions of certain self-affine fractals, *Indiana Univ. Math. J.* **41** (1992), 533–568.

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

[9] S. Lalley: Renewal theorems in symbolic dynamics, with applications to geodesic flows, noneuclidean tessellations, and their fractal limits. *Acta. Math.* **163** (1989), 1–55.

[10] B.B. Mandelbrot (1982) *The Fractal Geometry of Nature* Freeman & Co.

[11] C. McMullen: The Hausdorff dimension of general Sierpinski Carpets, *Nagoya Math J.* **96** (1984), 1–9.

[12] P.A.P. Moran: Additive functions of intervals and Hausdorff measure. *Proc. Cambridge Phil. Soc.* **42** (1946), 15–23.

[13] Y. Peres: The self–affine carpets of McMullen and Bedford have infinite Hausdorff measure. Preprint, 1992.

[14] F. Przytycki and M. Urbanski: On the Hausdorff dimension of some fractal sets. *Studia Math.* **93** (1989), 155-186.

[15] D. Ruelle (1978) *Thermodynamic Formalism*, Addison-Wesley.

[16] L.-S. Young: Dimension, entropy, and Lyapunov exponents. *Ergodic Theory Dynamical Systems* **2** (1982), 109–124.