A COMPACT FRAMEWORK FOR HIDDEN MARKOV CHAINS WITH APPLICATIONS TO FRACTAL GEOMETRY

VÍCTOR RUIZ,* *Universidad Complutense de Madrid*

Abstract

We introduce a class of stochastic processes in discrete time with finite state space by means of a simple matrix product. We show that this class coincides with that of the hidden Markov chains and provides a compact framework for it. We study a measure obtained by a projection on the real line of the uniform measure on the Sierpinski gasket, finding that the dimension of this measure fits with the Shannon entropy of an associated hidden Markov chain.

Keywords: Hidden Markov chain; Shannon entropy; self-similar measure; overlap; dimension; fractal

2000 Mathematics Subject Classification: Primary 60J10
Secondary 28A78; 28A80

1. Introduction

We introduce a simple and versatile way to express the distribution of hidden Markov chains that might be useful to address a longstanding open problem in fractal geometry: the determination of the Hausdorff dimension of self-similar measures with a significant overlap among its constituent cylinders. To make the ideas clearer, we choose the simplest case of the vertical projection μ on the real line of the natural or uniform probability measure on the Sierpinski gasket, i.e. that which gives equal weight to the three parts into which the triangles split at each step of the construction (see Section 3). We show that the measure μ can be understood in terms of a non-Markovian stationary and ergodic stochastic process in discrete time associated with a certain pair of matrices. This process turns out to be a hidden Markov chain and, in addition, every hidden Markov chain can be expressed in a similar way.

We devote Section 2 to introducing this framework for hidden Markov chains, and in Section 3 we show that the fractal dimension of the measure described above coincides with the Shannon entropy of the associated process, which provides two convergent sequences of estimates of the dimension of μ: one nonincreasing and the other nondecreasing. This allows us to obtain, with arbitrary accuracy, the dimension of μ. In particular, we obtain $\dim \mu = 0.988\,765\,871\,4$ ($\pm 0.5 \times 10^{-10}$). See Section 3 for some results on this issue by Ngai [12].

2. A framework for hidden Markov chains

2.1. Definition and remarks

Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$ be $N \times N$ dimensional and nonnegative matrices, and let $\mathbf{Z} = \mathbf{Z}_1 + \mathbf{Z}_2 + \cdots + \mathbf{Z}_n$. According to the Perron–Frobenius theorem (see [17, p. 16]) there exist a nonnegative

and maximal eigenvalue λ for $\boldsymbol{Z}$ with nonnegative left and right eigenvectors, which we shall call $\boldsymbol{u}$ and $\boldsymbol{w}^\top$, respectively. In order to avoid trivial cases, we shall assume that no row of $\boldsymbol{Z}$ is null. Then $\lambda > 0$, and $\boldsymbol{u}$ and $\boldsymbol{w}^\top$ are not null. By multiplying the $\boldsymbol{Z}_i$s by λ^{-1} if necessary, we may assume that $\lambda = 1$. Let

$$L_{\boldsymbol{w}}^+ := \{\boldsymbol{q} \in \mathbb{R}^N : \boldsymbol{q}\boldsymbol{w}^\top = 1, \ q_i \geq 0, \ 1 \leq i \leq N\}.$$

Definition 1. We define the stochastic process $V(\boldsymbol{q}, \boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n, \boldsymbol{w}^\top)$ associated to the matrices $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ as described above, the right eigenvector $\boldsymbol{w}^\top$, and to the row vector $\boldsymbol{q} \in L_{\boldsymbol{w}}^+$ as the sequence $\{V_j : j = 1, 2, \ldots\}$ of random variables on $J = \{1, 2, \ldots, n\}$ with finite-dimensional distributions Q_k given by

$$Q_k(\boldsymbol{i}) = \boldsymbol{q}\boldsymbol{Z}_{\boldsymbol{i}}\boldsymbol{w}^\top, \qquad \boldsymbol{i} \in J^k, \ k \in \mathbb{N}, \tag{1}$$

where $\boldsymbol{Z}_{\boldsymbol{i}} = \boldsymbol{Z}_{i_1} \cdots \boldsymbol{Z}_{i_k}$ for $\boldsymbol{i} = (i_1, \ldots, i_k)$.

The existence of the process can be easily checked, taking into account the fact that $(\sum_{j=1}^n \boldsymbol{Z}_j)\boldsymbol{w}^\top = \boldsymbol{Z}\boldsymbol{w}^\top = \boldsymbol{w}^\top$ and $\boldsymbol{q}\boldsymbol{w}^\top = 1$. Nevertheless, we show in Subsection 2.2 that (1) gives the finite-dimensional distributions of a hidden Markov chain.

Remark 1. Note that if $\boldsymbol{q} \in L_{\boldsymbol{w}}^+$ and $\alpha > 0$ then $\alpha^{-1}\boldsymbol{q} \in L_{\alpha\boldsymbol{w}}^+$ and $V(\alpha^{-1}\boldsymbol{q}, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n, \alpha\boldsymbol{w}^\top) = V(\boldsymbol{q}, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n, \boldsymbol{w}^\top)$.

If $\boldsymbol{Z}$ is irreducible then, according to the Perron–Frobenius theorem, $\lambda = 1$ is a simple eigenvalue and $\boldsymbol{u}$ and $\boldsymbol{w}^\top$ are strictly positive. Thus, in this irreducible case, $\boldsymbol{u}$ and $\boldsymbol{w}^\top$ are uniquely determined by the $\boldsymbol{Z}_i$ matrices up to a constant, and we define

$$V(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n) := V(\boldsymbol{u}, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n, \boldsymbol{w}^\top),$$

taking $\boldsymbol{u}$ and $\boldsymbol{w}^\top$ with $\boldsymbol{u}\boldsymbol{w}^\top = 1$.

Remark 2. It is easy to see that we can assume, without loss of generality, that $\boldsymbol{Z}$ is also a stochastic matrix.

Remark 3. Let $(\ldots, V_{-1}, V_0, V_1, \ldots)$ be the stationary extension of $V(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$ with indices in $\mathbb{Z}$. It is easy to check that the backward process $(\ldots, Y_{-1}, Y_0, Y_1, \ldots)$, given by $Y_k = V_{-k}$ for $k \in \mathbb{Z}$, is the stationary extension to $\mathbb{Z}$ of $V(\boldsymbol{Z}_1^\top, \ldots, \boldsymbol{Z}_n^\top)$.

2.2. Equivalence with hidden Markov chains

Let $\boldsymbol{P}$ be the transition matrix of a Markov chain $X = \{X_j : j = 1, 2, \ldots\}$ with state space $\{1, \ldots, N\}$. Let $\phi \colon \{1, \ldots, N\} \to \{1, \ldots, n\}$, $n \leq N$, and let $Y = \{Y_j : j = 1, 2, \ldots\}$ be the hidden Markov chain given by $Y_j = \phi(X_j)$. We can write $\boldsymbol{P} = \boldsymbol{M}_1 + \cdots + \boldsymbol{M}_n$, where each $\boldsymbol{M}_i$, $1 \leq i \leq n$, is the $N \times N$ matrix with zero entries except for the rows corresponding to indices in the subset of states $\phi^{-1}(i)$, which coincide with those of $\boldsymbol{P}$. It is easy to check that, if $\boldsymbol{q}$ is the probability vector with the distribution of X_1, we have

$$\boldsymbol{q}\boldsymbol{M}_{\boldsymbol{i}}\boldsymbol{w}^\top = P\{(Y_1, \ldots, Y_k) = \boldsymbol{i}\} \tag{2}$$

for $\boldsymbol{i} \in \{1, \ldots, n\}^k$, $k \in \mathbb{N}$, where $\boldsymbol{w}^\top = (1, \ldots, 1)^\top$. Hence, $V(\boldsymbol{q}, \boldsymbol{M}_1, \ldots, \boldsymbol{M}_n, \boldsymbol{w}^\top)$ is distributed as Y.

Thus, the class of hidden Markov chains is included in the class of processes $V(\boldsymbol{q}, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n, \boldsymbol{w}^\top)$. Next we show that both classes coincide.

Let $V = V(\boldsymbol{q}, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n, \boldsymbol{w}^\top)$. We assume that $\boldsymbol{Z}$ is stochastic and that $\boldsymbol{w}^\top = (1, \ldots, 1)^\top$, which can be done by Remark 2. For $i = 1, \ldots, n$, consider the following nN-dimensional matrices $\boldsymbol{M}_i$ defined by n rows of blocks, each one composed of n, N-dimensional blocks with zero entries, except for the ith row of blocks which is given by $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n$ (the same for all i):

$$\boldsymbol{M}_i = \begin{bmatrix} \boldsymbol{0}_{N(i-1)\times N} & \boldsymbol{0}_{N(i-1)\times N(n-2)} & \boldsymbol{0}_{N(i-1)\times N} \\ \boldsymbol{Z}_1 & \cdots & \boldsymbol{Z}_n \\ \boldsymbol{0}_{N(n-i)\times N} & \boldsymbol{0}_{N(n-i)\times N(n-2)} & \boldsymbol{0}_{N(n-i)\times N} \end{bmatrix}, \qquad i = 1, \ldots, n. \tag{3}$$

Let $\boldsymbol{q}_M = (\boldsymbol{q}\boldsymbol{Z}_1, \ldots, \boldsymbol{q}\boldsymbol{Z}_n)$, and let $\boldsymbol{w}_M^\top$ be an nN-dimensional column vector of 1s. It is easily seen that

$$\boldsymbol{q}_M \boldsymbol{M}_i \boldsymbol{w}_M^\top = \boldsymbol{q}\boldsymbol{Z}_i \boldsymbol{w}^\top, \tag{4}$$

and, therefore, $V(\boldsymbol{q}_M, \boldsymbol{M}_1, \ldots, \boldsymbol{M}_n, \boldsymbol{w}_M^\top) = V$. The $\boldsymbol{M}_i$ matrices have the same row structure as those in (2), $\boldsymbol{M} = \sum_{i=1}^n \boldsymbol{M}_i = \boldsymbol{P}$ is stochastic, and $\boldsymbol{w}_M^\top = (1, \ldots, 1)^\top$; hence, it follows, as in (2), that V is the hidden Markov chain function of the Markov chain with state space $\{1, \ldots, nN\}$ and transition matrix $\boldsymbol{P} = \boldsymbol{M}$ where $\phi(j)$ is the integer part of $(j-1)/N$ plus 1.

2.3. Stationarity and ergodicity

Let $V = V(\boldsymbol{q}, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n, \boldsymbol{w}^\top) = V(\boldsymbol{q}_M, \boldsymbol{M}_1, \ldots, \boldsymbol{M}_n, \boldsymbol{w}_M^\top)$. Assume that $\boldsymbol{Z}$ is stochastic, $\boldsymbol{w}^\top = (1, \ldots, 1)^\top$, and that (3) and (4) hold. It is known that the Markov chain with transition matrix $\boldsymbol{M}$ and initial distribution $\boldsymbol{q}_M$ is stationary if and only if $\boldsymbol{q}_M \boldsymbol{M} = \boldsymbol{M}$. It is easy to check that $\boldsymbol{q}_M \boldsymbol{M} = \boldsymbol{M}$ if and only if $\boldsymbol{q}\boldsymbol{Z} = \boldsymbol{Z}$. Therefore, this is a sufficient condition for V to be stationary.

In particular, if $\boldsymbol{Z}$ is irreducible and stochastic then $V(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$ is stationary. Next we show that it is also ergodic.

If some columns of the $\boldsymbol{Z}_i$s are null then so are the corresponding columns of $\boldsymbol{M}$ (see (3)). It is easy to check that the process remains unchanged if we remove those null columns and the corresponding rows in the $\boldsymbol{M}_i$ matrices and those entries in $\boldsymbol{w}_M^\top$ and $\boldsymbol{q}_M$. Let $\boldsymbol{M}_i'$ be the modified matrices. The sum matrix $\boldsymbol{M}'$ is stochastic, as $\boldsymbol{M}$, and the $\boldsymbol{M}_i'$s have the same row structure as the $\boldsymbol{M}_i$s in (2). We show that if $\boldsymbol{Z}$ is irreducible then so is $\boldsymbol{M}'$, though $\boldsymbol{M}$ cannot be, which allows us to obtain the result about ergodicity.

Assume that $\boldsymbol{Z}$ is irreducible and stochastic and that $V = V(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n) = V(\boldsymbol{M}_1, \ldots, \boldsymbol{M}_n)$. The matrix $\boldsymbol{M}^k$ is formed by n identical rows of blocks, each one with the blocks $\boldsymbol{Z}^{k-1}\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}^{k-1}\boldsymbol{Z}_n$. It can be easily checked, for $i = 1, \ldots, nN$ and j out of the null columns of $\boldsymbol{M}$, that there exists a k such that the (i, j)th entry of $\boldsymbol{M}^k$ is positive, and from this, it follows that $\boldsymbol{M}'$ is irreducible. By the theory of Markov chains we know that the stationary Markov chain with irreducible transition matrix $\boldsymbol{M}'$ is ergodic, and, hence, so is $V(\boldsymbol{M}_1', \ldots, \boldsymbol{M}_n') = V$, since it is a function of it.

If $\boldsymbol{Z}$ is also aperiodic, it is easy to see that $\boldsymbol{M}'$ is too, and so V is strong mixing.

From this and Remark 2, we obtain the following result.

Proposition 1. *Let $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ be $N \times N$-dimensional, nonnegative matrices, and assume that $\boldsymbol{Z} = \boldsymbol{Z}_1 + \cdots + \boldsymbol{Z}_n$ is irreducible, with its greatest eigenvalue being unity. Then $V(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$ is a hidden Markov chain, stationary and ergodic. If $\boldsymbol{Z}$ is also aperiodic then $V(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$ is strong mixing.*

2.4. Suitability of this formulation

The matrix expression for the distribution of hidden Markov chains studied here provides an improvement in compactness in two ways.

1. The expression above for the process $V = V(\boldsymbol{q}, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n, \boldsymbol{w}^\top)$ as $V(\boldsymbol{q}_M, \boldsymbol{M}_1, \ldots, \boldsymbol{M}_n, \boldsymbol{w}_M^\top)$ was useful to show that V is a hidden Markov chain, but the original expression uses smaller matrices.

2. The finite-dimensional distributions have the following simple expression, which is easy to obtain:

$$\mathrm{P}\{V_{i_1} = j_1, \ldots, V_{i_k} = j_k\} = \boldsymbol{q}\,\boldsymbol{Z}^{(1}\boldsymbol{Z}^{(2}\cdots\boldsymbol{Z}^{(k}\boldsymbol{w}^\top,$$

where $\boldsymbol{Z}^{(l} = \boldsymbol{Z}^{i_l - i_{l-1} - 1}\boldsymbol{Z}_{j_l}$ and $i_0 = 0$. As an example of its advantages, this expression provides a simple alternative proof for Proposition 1, taking into account the fact that

$$\lim_{p \to \infty} \frac{1}{p}\sum_{k=1}^{p}\boldsymbol{Z}^k = \boldsymbol{w}^\top\boldsymbol{u}$$

if $\boldsymbol{Z}$ is stochastic and irreducible, where $\boldsymbol{u}\boldsymbol{Z} = \boldsymbol{u}$, $\boldsymbol{Z}\boldsymbol{w}^\top = \boldsymbol{w}^\top$, and $\boldsymbol{u}\boldsymbol{w}^\top = 1$, and the fact that $\lim_{k \to \infty}\boldsymbol{Z}^k = \boldsymbol{w}^\top\boldsymbol{u}$ if $\boldsymbol{Z}$ is also aperiodic.

Proposition 1 in turn provides an alternative proof for the known result that an irreducible and stationary Markov chain with finite state space is ergodic, and strong mixing if it is also aperiodic. To show this, we take $\boldsymbol{Z} = \boldsymbol{P}$, the transition matrix, and $\boldsymbol{Z}_i$ with zero entries except for the ith row which coincides with that of $\boldsymbol{P}$.

We now comment on how this expression for the distribution of hidden Markov chains offers hidden Markov chains as a natural generalization for sequences of independent, identically distributed, discrete random variables. Commutative products of real numbers $(p_1, \ldots, p_n)$ are substituted by noncommutative products of matrices $(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$. The numbers and the matrices are nonnegative; the sum of the p_is is unity and the sum of the $\boldsymbol{Z}_i$s has its dominant eigenvalue equal to unity. The latter case also involves the elements $\boldsymbol{q}$, $\boldsymbol{u}$, and $\boldsymbol{w}^\top$, which are equal to unity in the former case. In this framework, Markov chains lie in an intermediate place.

The contents of this section can be easily generalized to the case of $\mathbb{N} \times \mathbb{N}$-dimensional, countable many matrices $\boldsymbol{Z}_i$, thus, characterizing the hidden Markov chains with countable state space.

3. An application: computation of the Hausdorff dimension of a self-similar measure with overlaps

The characterization of the hidden Markov chains presented in Section 2 originated in the study of a fractal geometry problem, namely, the dimension of self-similar measures with overlaps. In this section we will consider a simple case (the self-similar measure μ to be defined below) that has been analyzed in [12].

An important question in the multifractal analysis of a measure is the dimension of the measure. Ngai [12] obtained a function τ, given as a lower limit, whose Legendre transform, which was not calculated, determines the multifractal spectrum of μ and particularly its dimension. Ngai [11] demonstrated that the dimension coincides with $\tau'(1)$, and with Lau in [7] he obtained formulae (as limits) for the dimension of some related cases. In [12] it was shown that $\tau(2) < 1$, and from [8, Proposition 4.1] it follows that μ is a singular measure.
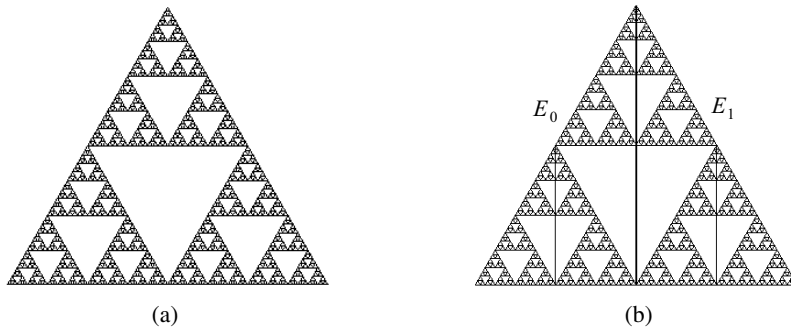
FIGURE 1.

Here we introduce some improvements in the analysis of μ and a procedure, new as far as the author is aware, that permits more complete results in the study of the dimension. We obtain a means of calculating the dimension by proving that it coincides with the Shannon entropy of a hidden Markov chain.

This method is likely to be successful with other measures under study.

We analyze the problem of determining the Hausdorff dimension of the vertical projection on the horizontal axis of the natural measure on the Sierpinski gasket. Recall that the Sierpinski gasket E is the limit set of the following iterative process (see Figure 1(a)).

  (i) Divide an equilateral triangle T, whose basis is the horizontal unit interval I, into four equilateral triangles determined by the middle points of the sides of T.

 (ii) Delete the central triangle.

(iii) Repeat steps (i) and (ii) in the remaining equilateral triangles.

The natural measure ν on E is obtained as the weak limit of the sequence of probability measures ν_k resulting from the uniform distribution of unit mass on the 3^k equilateral triangles that made up the prefractal at its kth stage of construction. It is well known that the measure ν is singular with respect to the Lebesgue bidimensional measure. In fact, ν is a fractal measure, i.e. a measure concentrated on the set E of fractional dimension $\dim E = \ln 3/\ln 2$ and such that any subset F of dimension strictly smaller than $\dim E$ is a ν-null set. For these reasons, the measure ν is said to have Hausdorff dimension $\dim \nu = \ln 3/\ln 2$. Now consider the vertical projection $\mathrm{pr}\colon \mathbb{R}^2 \to \mathbb{R}$ and the induced measure $\mu = \nu \circ \mathrm{pr}^{-1}$. Because $\dim \nu > 1$, the measures obtained by the projection of ν on the real line have Hausdorff dimension equal to 1 for almost every direction of projection (considering the unidimensional Lebesgue measure on the set of such directions; this is obtained as a consequence of [9, Corollary 9.8, p. 131] and of the definition of dimension of a measure). In this section we show how a hidden Markov chain can be used to determine in a simple manner $\dim \mu$, and we shall show that indeed $\dim \mu < 1$, and, therefore, the vertical projection belongs to the set of exceptional directions. This also provides an alternative way to show the singularity of μ.

3.1. Some notions of fractal geometry

See [5, Chapter 6] for an exposition of the following notions and results.

The Hausdorff dimension of a measure μ on $\mathbb{R}^p$ is defined as

$$\dim \mu = \inf\{\dim A \colon \mu A > 0\},$$

where dim A is the Hausdorff dimension of the set A. The computation of dim μ can be achieved by means of the *local dimension of* μ, defined as

$$\theta(\mu, x) = \lim_{r \to 0} \frac{\log \mu B(x, r)}{\log r},$$

where $B(x, r)$ denotes the closed ball centered at x and with radius r. If such a limit exists and takes, for μ-almost every (μ-a.e.) x, a constant value α, then dim $\mu = \alpha$ (see [4, Theorem 3.3.14(b), p. 127]).

Associated to any finite system (set) $S = \{f_1, \ldots, f_q\}$ of contracting similarities of $\mathbb{R}^p$ and a positive probability vector $(p_1, \ldots, p_q)$ there exists a unique probability measure ν, called a self-similar measure, that satisfies $\nu = \sum_{i=1}^q p_i \nu \circ f_i^{-1}$.

The system S is said to satisfy the open set condition if there exists an open set $O \subset \mathbb{R}^p$ such that $\bigcup_{i=1}^q f_i O \subset O$ and the copies $f_i O$ of O are disjoint.

For the study of the dimension of a self-similar measure with the open set condition, an associated Bernoulli process (see [4, Theorem 5.2.5, p. 212]) can be used. In the absence of the open set condition (i.e. for systems of contracting similarities with significant overlapping) such analysis cannot proceed and few things are known about the structure of self-similar sets and measures, and, in particular, their Hausdorff dimension remains unknown. The research on fractal sets and measures with overlapping is one of the most active areas in the field of fractal geometry; see, for instance, [6], [8], [13], [14], [15], and [16].

The problem that we analyze in this section may be regarded as a problem of an overlapping system of similarities. This is because it can be easily seen that if we take the base of the initial equilateral triangle to be the interval $[0, 1]$, in the construction of the Sierpinski gasket described above, then μ is the self-similar measure associated to the overlapping system $\{f_1, f_2, f_3\}$ with

$$f_i(x) = \frac{1}{2}x + \frac{i}{4}, \qquad i = 0, 1, 2,$$

and probability weights $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. In order to make more transparent the application of the processes described in Section 2, we do not adopt this viewpoint, instead we carry out the analysis on the Sierpinski gasket.

3.2. Dimension of μ

The main result to be proved in this section is the following.

Theorem 1. *Let H be the Shannon entropy of the process $V(\mathbf{Z}_0, \mathbf{Z}_1)$, where*

$$\mathbf{Z}_0 = \frac{1}{3} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad and \quad \mathbf{Z}_1 = \frac{1}{3} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Then dim $\mu = H$.

Remark 4. The value of H with ten decimal digits of approximation is $0.988\,765\,871\,4$. To obtain this, we use two known results on entropy of hidden Markov chains. The sequence of conditional entropies $H(V_k/V_1, \ldots, V_{k-1})$ is nonincreasing and converges to H (see [1, Theorem 11.2, p. 116]). From [2] we know that if $V = V(\mathbf{Z}_0, \mathbf{Z}_1)$ is a function of an irreducible Markov chain X then the sequence $H(V_k/X_1, V_2, \ldots, V_{k-1}) = H(X_1, V_2, \ldots, V_k) - H(X_1, V_2, \ldots, V_{k-1})$ is nondecreasing and converges to H. To use the latter result, we take

$$M_0 = \begin{bmatrix} \mathbf{Z}_0 & \mathbf{Z}_1 \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \end{bmatrix}, \qquad M_1 = \begin{bmatrix} \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{Z}_0 & \mathbf{Z}_1 \end{bmatrix},$$

and M_i^* for $i = 1, 2, 3, 4$ with zero entries except for the ith row, which coincides with that of $M = M_0 + M_1$. As in Subsection 2.2, we have $V(M_0, M_1) = V(Z_0, Z_1) = V$. Let X be the stationary Markov chain with transition matrix M (it is irreducible). Consider V as the function of X given by $\phi(1) = \phi(2) = 0$ and $\phi(3) = \phi(4) = 1$. It is easy to check that

$$P\{X_1 = i_1, (V_2, \ldots, V_k) = \mathbf{i}\} = u_M M_{i_1}^* M_{\mathbf{i}} w_M^\top,$$

with $u_M = (uZ_0, uZ_1) = \frac{1}{6}(2, 1, 1, 2)$ and $w_M^\top = (1, 1, 1, 1)^\top$, which can be used to calculate $H(X_1, V_2, \ldots, V_k)$. Actually, we have $u_M M_{i_1}^* M_{\mathbf{i}} w_M^\top = uZ_{i_1}^* Z_{\mathbf{i}} w^\top$ with $Z_1^* = Z_0 A$, $Z_2^* = Z_0 B$, $Z_3^* = Z_1 A$, $Z_4^* = Z_1 B$,

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

For the value of $\dim \mu = H$ given above, we have calculated the term with $k = 17$ for each sequence.

In order to prove Theorem 1 we need two lemmas.

Given $x \in I := [0, 1]$ with binary expansion $x = 0.i_1 i_2 i_3 \ldots$, $i_j \in M := \{0, 1\}$, let $D_0(x) = I$ and, for $k \geq 1$, let $D_k(x)$ denote the dyadic interval $[\sum_{j=1}^k 2^{-j} i_j, \sum_{j=1}^k 2^{-j} i_j + 2^{-k}]$. Note that, since $\mathrm{pr}^{-1}(x)$ is a one-dimensional affine subspace and $\dim \nu = \ln 3 / \ln 2 > 1$, we have $\mu(x) = \nu(\mathrm{pr}^{-1}(x)) = 0$, and since the points in I with double binary expansion are a countable subset of I, $D_k(x)$ is uniquely defined for μ-a.e. $x \in I$. We also denote by $\langle i_1, i_2, \ldots, i_k \rangle$ the dyadic interval $[\sum_{j=1}^k 2^{-j} i_j, \sum_{j=1}^k 2^{-j} i_j + 2^{-k}]$ for $i_1, i_2, \ldots, i_k \in M$.

Lemma 1. *We have*

$$\mu\langle i_1, i_2, \ldots, i_k \rangle = e_{i_1} Z_{i_2} \cdots Z_{i_k} p^\top \quad \text{for all } k \text{ and } i_1, i_2, \ldots, i_k \in M,$$

where $e_0 = (1, 0)$, $e_1 = (0, 1)$, and $p^\top = (2^{-1}, 2^{-1})^\top$.

Proof. $\mu\langle i_1, \ldots, i_k \rangle = \nu(\mathrm{pr}^{-1}(\langle i_1, \ldots, i_k \rangle))$ is given by the ν-measure of the vertical band built upon $\langle i_1, \ldots, i_k \rangle$. For the dyadic intervals of the first generation, we have $\mu\langle 0 \rangle = \mu[0, \frac{1}{2}] = \frac{1}{2}$ and $\mu\langle 1 \rangle = \frac{1}{2}$, because the parts of E placed upon these intervals, which we respectively denote by E_0 and E_1, are symmetric, as shown in Figure 1(b). Each vertical band built upon a dyadic interval J of the kth generation, $k \geq 1$, intersects E on a collection $E_i^k(J)$ of sets that are copies under homotheties of contraction ratio $1/2^{k-1}$ of E_i, $i = 0, 1$; in particular, for $k = 1$, we have $E_0^1\langle 0 \rangle = \{E_0\}$, $E_1^1\langle 0 \rangle = \varnothing$, $E_0^1\langle 1 \rangle = \varnothing$, and $E_1^1\langle 1 \rangle = \{E_1\}$. Since the μ-measure of each set in $E_i^k(J)$ is $2^{-1} 3^{-(k-1)}$, the task of computing μJ reduces to computing the cardinality of $E_i^k(J)$, $i = 0, 1$. We shall do it recursively. For each dyadic interval $J = \langle i_1, \ldots, i_k \rangle$ of the kth generation, $k \geq 1$, let $u^k(J) = (\#E_0^k(J), \#E_1^k(J))$. The left half of each set in E_0^k is a homothetic copy of E_0, whereas the right half of each set in E_0^k consists of one copy of E_0 and one copy of E_1, as happens with the set E_0 (see Figure 1(b)). Similarly, the left half of each set in E_1^k consists of one copy of E_0 plus one copy of E_1, whereas the right half of each set in E_1^k is a single copy of E_1. The interval J splits into two intervals of the $(k+1)$th generation, $J_i := \langle i_1, \ldots, i_k, i \rangle$, $i = 0, 1$. Therefore, the sets in $E_0^{k+1}(J_0)$ come one from each set in $E_0^k(J)$ and one from each set in $E_1^k(J)$, whereas the sets in $E_1^{k+1}(J_0)$ come one from each set in $E_1^k(J)$, giving

$$u^{k+1}(J_0) = u^k(J) \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{5}$$

Similarly,

$$u^{k+1}(J_1) = u^k(J) \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}. \tag{6}$$

Let $\boldsymbol{L}_0$ and $\boldsymbol{L}_1$ be the matrices in expressions (5) and (6), respectively. Hence, taking into account the fact that $u^1\langle i_1 \rangle = \boldsymbol{e}_{i_1}$, we have

$$u^k \langle i_1, i_2, \ldots, i_k \rangle = \boldsymbol{e}_{i_1} \boldsymbol{L}_{i_2} \cdots \boldsymbol{L}_{i_k},$$

and, thus,

$$\mu \langle i_1, i_2, \ldots, i_k \rangle = u^k \langle i_1, i_2, \ldots, i_k \rangle \begin{bmatrix} 2^{-1}3^{-(k-1)} \\ 2^{-1}3^{-(k-1)} \end{bmatrix} = \boldsymbol{e}_{i_1} \boldsymbol{Z}_{i_2} \cdots \boldsymbol{Z}_{i_k} \boldsymbol{p}^\top.$$

Lemma 2. *Let $x \in I$ with $D_k(x)$ uniquely defined, and let $D_k'(x)$ be any of the two dyadic intervals of the kth generation adjacent to $D_k(x)$. Then, for large enough k,*

$$\frac{1}{2} \le \frac{\mu(D_k'(x))}{\mu(D_k(x))} \le 2.$$

Proof. Assume that x has a single binary expansion and that k is large enough so that the right endpoint of $D_k(x)$ is neither $\frac{1}{2}$ nor 1, that is, $D_k(x) \ne \langle 0, 1, 1, \ldots, 1 \rangle$ and $D_k(x) \ne \langle 1, 1, \ldots, 1 \rangle$. If $D_k(x) = \langle i, \ldots, i_k \rangle$, the dyadic interval $D_k'(x)$ of the kth generation adjacent to the right of $D_k(x)$ can be obtained in the following way. Let $l = \max\{t \in \{2, 3, \ldots, k\} : i_t = 0\}$. If $l = k$ then $D_k'(x) = \langle i, \ldots, i_{k-1}, 1 \rangle$, and if $l < k$ then $D_k'(x) = \langle i, \ldots, i_{l-1}, 100, \ldots, 0 \rangle$. Hence, by the previous lemma,

$$\mu(D_k(x)) = \boldsymbol{e}_{i_1} \boldsymbol{Z}_{i_2} \cdots \boldsymbol{Z}_{i_{l-1}} \boldsymbol{Z}_0 \boldsymbol{Z}_1^{k-l} \boldsymbol{p}^\top,$$
$$\mu(D_k'(x)) = \boldsymbol{e}_{i_1} \boldsymbol{Z}_{i_2} \cdots \boldsymbol{Z}_{i_{l-1}} \boldsymbol{Z}_1 \boldsymbol{Z}_0^{k-l} \boldsymbol{p}^\top.$$

Forming the products of the matrices, we obtain

$$\boldsymbol{Z}_0 \boldsymbol{Z}_1^q \boldsymbol{p}^\top = \frac{1}{2(3^{q+1})}(q+1, q+2)^\top, \qquad \boldsymbol{Z}_1 \boldsymbol{Z}_0^q \boldsymbol{p}^\top = \frac{1}{2(3^{q+1})}(q+2, q+1)^\top,$$

$$\frac{\mu(D_k'(x))}{\mu(D_k(x))} = \frac{(q+2)u_1 + (q+1)u_2}{(q+1)u_1 + (q+2)u_2},$$

where $q = k - l \in \{0, \ldots, k-2\}$ and $(u_1, u_2) = \boldsymbol{e}_{i_1} \boldsymbol{Z}_{i_2} \cdots \boldsymbol{Z}_{i_{l-1}} \ge 0$. From this, it follows that

$$\frac{1}{2} \le \frac{\mu(D_k'(x))}{\mu(D_k(x))} \le 2.$$

If $D_k'(x)$ is adjacent to the left of $D_k(x)$ then the above result can be applied, interchanging the roles of $D_k'(x)$ and $D_k(x)$.

Now we can proceed to prove Theorem 1.

Proof of Theorem 1. The matrix $\boldsymbol{Z}_0 + \boldsymbol{Z}_1$ is irreducible with right and left eigenvectors $\boldsymbol{p}^\top = (2^{-1}, 2^{-1})^\top$ and $\boldsymbol{e} = (1, 1)$, respectively. Let Q be the distribution of the process $V(\boldsymbol{Z}_0, \boldsymbol{Z}_1)$, which is stationary and ergodic by Proposition 1. Consider the restrictions of μ, $\mu_0 = \mu|_{\langle 0 \rangle}$ and $\mu_1 = \mu|_{\langle 1 \rangle}$, and the auxiliary probability measure supported on the interval $\langle 0 \rangle = [0, \frac{1}{2}]$ given by

$$\mu^* := \mu_0 + \mu_1 \circ t,$$

where t is the translation $t(x) = x + \frac{1}{2}$. Let $\pi : \mathbb{M} \to [0, \frac{1}{2}]$, with $\mathbb{M} := \{0, 1\}^\infty$, given by $\pi(\boldsymbol{i}) = \sum_{j=1}^\infty 2^{-(j+1)} i_j$. Note that $\pi[i_1, \ldots, i_k] = \langle 0, i_1, \ldots, i_k \rangle$, where $[i_1, \ldots, i_k]$ is the cylinder in $\mathbb{M}$ with base $(i_1, \ldots, i_k)$. From Lemma 1, it follows that

$$\mu^* \langle 0, i_1, \ldots, i_k \rangle = Q[i_1, \ldots, i_k],$$

and, thus, $Q = \mu^* \circ \pi$. Since π^{-1} is uniquely defined μ^*-a.e. $x \in \langle 0 \rangle$, we have $\mu^* = Q \circ \pi^{-1}$. By the Shannon–McMillan theorem (see [1, Theorem 13.1, p. 129]), there exists a set $G \subset \mathbb{M}$ with $Q(G) = 1$ and

$$\lim_{k \to \infty} -\frac{1}{k} \log_2 Q[i_1, \ldots, i_k] = H \quad \text{for all } \boldsymbol{i} = (i_1, i_2, \ldots) \in G.$$

For all $x \in \pi G$ such that $\boldsymbol{i} = \pi^{-1}(x)$ is uniquely defined, with $\boldsymbol{i} = (i_1, i_2, \ldots) \in G$, we have $Q(\pi^{-1} D_k(x)) = Q[i_1, \ldots, i_k]$. Thus,

$$\lim_{k \to \infty} -\frac{1}{k} \log_2 \mu^* D_k(x) = \lim_{k \to \infty} -\frac{1}{k} \log_2 Q(\pi^{-1} D_k(x)) = H$$

for almost all $x \in \pi G$ and so for μ^*-a.e. $x \in \langle 0 \rangle$, since $\mu^*(\pi G) = Q(G) = 1$. Now let $2^{-(k+1)} \leq r \leq 2^{-k}$. Then

$$D_{k+1}(x) \subset [x - r, x + r] \subset D_k(x) \cup D_k'(x) \cup D_k''(x),$$

where $D_k'(x)$ and $D_k''(x)$ are the dyadic intervals of the kth generation adjacent to $D_k(x)$. This and Lemma 2 gives $\mu D_{k+1}(x) \leq \mu[x - r, x + r] \leq 5\mu D_k(x)$ for μ-a.e. $x \in [0, 1]$ and

$$\mu^* D_{k+1}(x) \leq \mu^*[x - r, x + r] \leq 5\mu^* D_k(x) \quad \text{for } \mu^*\text{-a.e. } x \in \langle 0 \rangle.$$

From this, it follows that

$$\theta(\mu^*, x) = \lim_{r \to 0} \frac{\log \mu^*[x - r, x + r]}{\log r} = \lim_{k \to \infty} -\frac{\log_2 \mu^* D_k(x)}{k} = H$$

for a.e. x and, thus, $\dim \mu^* = H$. Since $\mu_0 \ll \mu^*$ and $\mu_1 \circ t \ll \mu^*$, it follows from Lemma 2.3 of [10] that $\dim \mu_0 = \dim(\mu_1 \circ t) = \dim \mu^* = H$. Also, since translations are bilipschitz maps, we have $\dim \mu_1 = \dim(\mu_1 \circ t) = H$, and it follows that $\dim \mu = H$.

We have $\dim \mu < 1$, since

$$\dim \mu = H \leq \frac{1}{2} H(V_1, V_2) = -\frac{1}{2} \sum_{i_1, i_2 = 0, 1} (e Z_{i_1} Z_{i_2} \boldsymbol{p}^\top) \log_2 (e Z_{i_1} Z_{i_2} \boldsymbol{p}^\top) < 0.996,$$

or by Remark 4. This provides an alternative way to prove the following known result (see Section 3).

Corollary 1. *The measure μ is singular.*

4. Conclusion

In this paper we have introduced a new expression for the representation of hidden Markov chains which can significantly simplify the practical work with this relevant probabilistic tool.

As an example, we have solved an open problem in fractal geometry, that of the determination of the dimension of the vertical projection μ of the natural measure on the Sierpinski gasket.

The consideration of the associated hidden Markov chain allows us to obtain the dimension of the measure μ as a Shannon entropy, as in the known formula for the dimension of a self-similar measure without overlaps (see [3] or [4, Theorem 5.2.5, p. 212]). This is interesting for the understanding of the notion of dimension and for computational purposes.

Work in progress allows us to be optimistic in regard to the role that this expression for hidden Markov chains can play in the research on problems of fractal constructions with overlapping.

Acknowledgements

References

[1] BILLINGSLEY, P. (1965). *Ergodic Theory and Information.* John Wiley, New York.

[2] BIRCH, J. J. (1962). Approximations for the entropy for functions of Markov chains. *Ann. Math. Statist.* **33,** 930–938.

[3] DELIU, A., GERONIMO, J. S., SHONKWILER, R. AND HARDIN, D. (1991). Dimensions associated with recurrent self-similar sets. *Math. Proc. Camb. Philos. Soc.* **110,** 327–336.

[4] EDGAR, G. A. (1998). *Integral, Probability and Fractal Measures.* Springer, New York.

[5] FALCONER, K. (1997). *Techniques in Fractal Geometry.* John Wiley, New York.

[6] KENION, R. W. (1997). Projecting the one-dimensional Sierpiński gasket. *Israel J. Math.* **97,** 221–238.

[7] LAU, K. S. AND NGAI, S. M. (2000). Second-order self-similar identities and multifractal decompositions. *Indiana Univ. Math. J.* **49,** 925–972.

[8] LAU, K. S., NGAI, S. M. AND RAO, H. (2001). Iterated function systems with overlaps and self-similar measures. *J. London Math. Soc.* **63,** 99–116.

[9] MATTILA, P. (1995). *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability.* Cambridge University Press.

[10] MATTILA, P., MORÁN, M. AND REY, J.-M. (2000). Dimension of a measure. *Studia Math.* **142,** 219–253.

[11] NGAI, S. M. (1997). A dimension result arising from the L^q-spectrum of a measure. *Proc. Amer. Math. Soc.* **125,** 2943–2951.

[12] NGAI, S. M. (1997). Multifractal decomposition for a family of overlapping self-similar measures. In *Fractal Frontiers*, eds M. M. Novak and T. G. Dewey, World Scientific Publishing, River Edge, NJ, pp. 151–161.

[13] NGUYEN, N. (2001). Iterated function systems of finite type and the weak separation property. *Proc. Amer. Math. Soc.* **130,** 483–487.

[14] PERES, Y. AND SOLOMYAK, B. (1996). Absolute continuity of Bernoulli convolutions, a simple proof. *Math. Res. Lett.* **3,** 231–239.

[15] SCHIEF, A. (1994). Separation properties for self-similar sets. *Proc. Amer. Math. Soc.* **122,** 111–115.

[16] TESTUD, B. (2006). Measures quasi-Bernoulli au sens faible: résultats et exemples. *Ann. Inst. H. Poincaré Prob. Statist.* **42,** 1–35.

[17] WALTERS, P. (1982). *An Introduction to Ergodic Theory.* Springer, New York.