

Article

A Catalog of Self-Affine Hierarchical Entropy Functions

John Kieffer

Department of Electrical & Computer Engineering, University of Minnesota Twin Cities,
200 Union Street SE, Minneapolis, MN 55455, USA; E-Mail: kieffer@umn.edu

Received: 23 September 2011; in revised form: 18 October 2011 / Accepted: 30 October 2011 /

Published: 1 November 2011

Abstract: For fixed $k \geq 2$ and fixed data alphabet of cardinality m , the hierarchical type class of a data string of length $n = k^j$ for some $j \geq 1$ is formed by permuting the string in all possible ways under permutations arising from the isomorphisms of the unique finite rooted tree of depth j which has n leaves and k children for each non-leaf vertex. Suppose the data strings in a hierarchical type class are losslessly encoded via binary codewords of minimal length. A hierarchical entropy function is a function on the set of m -dimensional probability distributions which describes the asymptotic compression rate performance of this lossless encoding scheme as the data length n is allowed to grow without bound. We determine infinitely many hierarchical entropy functions which are each self-affine. For each such function, an explicit iterated function system is found such that the graph of the function is the attractor of the system.

Keywords: types; type classes; lossless compression; hierarchical entropy; self-affine functions; iterated function systems

1. Introduction

A traditional type class consists of all permutations of a fixed finite-length data string. There is a well-developed data compression theory in which strings in a traditional type class are losslessly encoded into fixed-length binary codewords [1]. One can generalize the notion of traditional type class and the resulting data compression theory in the following natural way. Let T be a finite rooted tree; an isomorphism of T is a one-to-one mapping of the set of vertices of T onto itself which preserves the parent-child relation. Let n be the number of leaves of T , let $\mathcal{L}(T)$ be the set of leaves of T , and let σ be a one-to-one mapping of $\{1, 2, \dots, n\}$ onto $\mathcal{L}(T)$. Suppose $(X_1, X_2, \dots, X_n)$ is a data string of

length n . Define the T -type class of $(X_1, \dots, X_n)$ to consist of all strings $(Y_1, Y_2, \dots, Y_n)$ for which there exists an isomorphism ϕ of T such that

$$Y_i = X_{\sigma^{-1}(\phi(\sigma(i)))}, \quad i = 1, 2, \dots, n$$

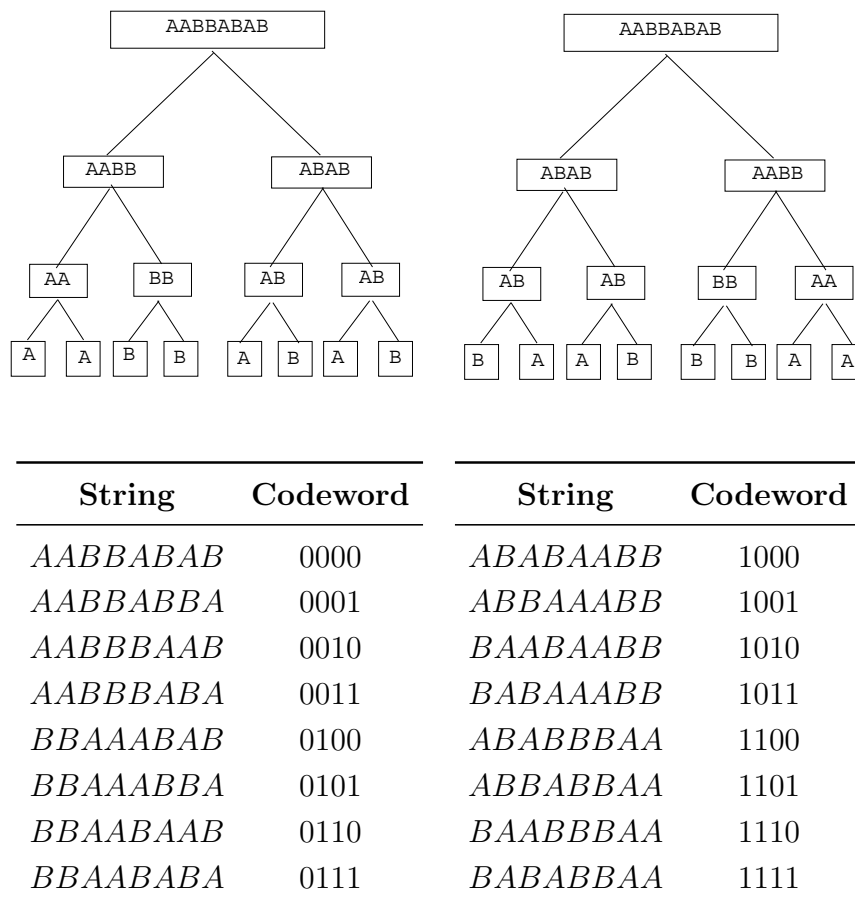
Consider the depth one tree $T = T_1(n)$ in which there are n children of the root, which are the leaves of the tree. Then, the notion of $T_1(n)$ -type class coincides with the notion of traditional type class. Now let $n = k^j$ for positive integer j and integer $k \geq 2$. Consider the depth j tree $T = T_j(k)$ with n leaves such that each non-leaf vertex has k children. Then, a $T_j(k)$ -type class is called a *hierarchical type class*, and k is called the *partitioning parameter* of the class. In the paper [2], we dealt with hierarchical type classes in which the partitioning parameter is $k = 2$. In the present paper, we deal with hierarchical type classes in which the partitioning parameter is an arbitrary $k \geq 2$.

Given a hierarchical type class S , there is a simple lossless coding algorithm which encodes each string in S into a fixed-length binary codeword of minimal length, and decodes the string from its codeword. This algorithm is particularly simple for the case when the partitioning parameter is $k = 2$, and we illustrate this case in Example 1 which follows; the case of general $k \geq 2$ is discussed in [3]. In Example 1 and subsequently, $x^1 * x^2 * \dots * x^k$ shall denote the data string obtained by concatenating together the finite-length data strings $x^1, x^2, \dots, x^k$ (left to right).

Example 1. Let $k = 2$, and let S be the hierarchical type class of data string $AABBABAB$. The 16 strings in S are illustrated in Figure 1. Each string $x \in S$ has a tree representation in which each vertex of tree $T_3(2)$ is assigned a label which is a substring of x . This assignment takes place as follows.

- The leaves of the tree, traversed left to right, are labeled with the respective left-to-right entries of the data string x .
- For each non-leaf vertex v , if the strings labeling the left and right children of v are x_L, x_R , respectively, then the string labeling v is $x_L * x_R$ if x_L precedes or is equal to x_R in the lexicographical order, and is $x_R * x_L$, otherwise.

In Figure 1, we have illustrated the tree representations of the strings $AABBABAB$ and $BAABBBAA$. The root label of all 16 tree representations will be the same string, namely, the first string in S in lexicographical order, which is the string $AABBABAB$ in this case. Each string in S is encoded by visiting, in depth-first order, the non-leaf vertices of its tree representation whose children have different labels. Each such vertex is assigned bit 0 if its label is $x_L * x_R$, where x_L, x_R are the labels of its left and right children, and is assigned bit 1 otherwise (meaning that the label is $x_R * x_L$). The resulting sequence of bits, in the order they are obtained, is the codeword of the string. Since both encoder and decoder will know what hierarchical type class is being encoded, the decoder will know what the root label of the tree representation should be, and then the successive bits of the codeword allow the decoder to grow the tree representation from the root downward.

Figure 1. Example 1 Tree Representations and Codeword Table.

Before discussing the nature of the results to be obtained in this paper, we need some definitions and notation. Fix integers $m, k \geq 2$, which serve as parameters in the subsequent development; k is the partitioning parameter already introduced, and m is called the “alphabet cardinality parameter” because we shall be dealing with an m -letter data alphabet, denoted $\mathcal{A}_m = \{a_1, a_2, \dots, a_m\}$. For each $j \geq 0$, we define a j -string x to be a string of length k^j over $\mathcal{A}_m$. Note that if $j \geq 1$, for each j -string x there is a unique k -tuple $(x^1, x^2, \dots, x^k)$ whose entries are $(j-1)$ -strings such that $x = x^1 * x^2 * \dots * x^k$; this k -tuple is called the k -partitioning of x . If $S_1, S_2, \dots, S_k$ are non-empty sets of j -strings, let $S_1 * S_2 * \dots * S_k$ be the set of all $(j+1)$ -strings of the form $x^1 * x^2 * \dots * x^k$, where x^i belongs to S_i for $i = 1, 2, \dots, k$. The 0-strings are the individual letters in $\mathcal{A}_m$.

We wish to formally define the family $\mathcal{S}_{m,k}$ of all hierarchical type classes in which the alphabet cardinality parameter is m and the partitioning parameter is k . Instead of using the tree isomorphism definition of hierarchical type class given at the beginning of the paper, we will use an equivalent inductive definition, which is more convenient in the subsequent development. First, we define the hierarchical type class of a 0-string to be the set consisting of the string itself. Given j -string x with $j \geq 1$, and assume hierarchical type classes of $(j-1)$ -strings have been defined. Let $(x^1, \dots, x^k)$ be the k -partitioning of x and let S_i be the hierarchical type class of x^i ($i = 1, \dots, k$). The hierarchical type class of x is then defined as

$$\bigcup_{\pi \in \Pi_k} [S_{\pi(1)} * S_{\pi(2)} * \dots * S_{\pi(k)}] \quad (1)$$

where, from now on, Π_k is the set of all permutations of $\{1, 2, \dots, k\}$. A set is called a hierarchical type class of order j if it is the hierarchical type class of some j -string. A set is called a hierarchical type class if it is a hierarchical type class of order j for some $j \geq 0$. The family $\mathcal{S}_{m,k}$ is then the set of all hierarchical type classes, of all orders.

We define the type of j -string x to be the vector $(n_1, \dots, n_m)$ whose i -th component n_i is the frequency of letter a_i in x . For each $j \geq 0$, let $\Lambda_j(m, k)$ be the set of all types of j -strings. Let $\Lambda(m, k)$ be the union of the $\Lambda_j(m, k)$'s for $j \geq 0$, and let $\Lambda^+(m, k)$ be the union of the $\Lambda_j(m, k)$'s for $j \geq 1$. A type in $\Lambda_j(m, k)$ will be said to be of order j . If $\lambda \in \Lambda(m, k)$, let $\|\lambda\|$ denote the sum of the components of λ . If λ is of order j , then $\|\lambda\| = k^j$. All strings in a hierarchical type class have the same type, because permuting a string does not change the type. This property is listed below, along with some other properties whose simple proofs are omitted.

- **Prop. 1:** All strings in a hierarchical type class have the same type.
- **Prop. 2:** For each $j \geq 0$, the distinct hierarchical type classes of order j form a partition of the set of all j -strings.
- **Prop. 3:** Let $\lambda \in \Lambda(m, k)$, and let $\mathcal{S}_{m,k}(\lambda)$ denote the set of all hierarchical type classes in $\mathcal{S}_{m,k}$ whose strings are of type λ . Then $\mathcal{S}_{m,k}(\lambda)$ forms a partition of the set of all strings of type λ .
- **Prop. 4:** Let $S \in \mathcal{S}_{m,k}$ be a hierarchical type class of order $j \geq 1$. Then there is a k -tuple $(S_1, S_2, \dots, S_k)$, unique up to permutation, such that each S_i is a hierarchical type class of order $j - 1$ and S is expressible as Equation (1).

Global Hierarchical Entropy Function. The *global hierarchical entropy function* is the function $H : \mathcal{S}_{m,k} \rightarrow [0, \infty)$ such that

$$H(S) \triangleq \log_2 |S|, \quad S \in \mathcal{S}_{m,k}$$

where, in this paper, if S is a finite set, $|S|$ shall denote the cardinality of S . $H(S)$ shall be called the entropy of S . Given a hierarchical type class S , its entropy $H(S)$ has the following interpretation. Suppose $H(S) > 0$, and we losslessly encode the strings in S into fixed-length binary codewords of minimal length (as discussed in Example 1 and in [3]). Then this minimal length is $\lceil H(S) \rceil$.

Lemma 1. Let S be a hierarchical type class of order $j \geq 1$. Let $(S_1, S_2, \dots, S_k)$ be the k -tuple of hierarchical type classes of order $j - 1$ associated with S according to Prop. 4, and let $N(S)$ be the number of distinct permutations of this k -tuple. Then,

$$H(S) = \left[\sum_{i=1}^k H(S_i) \right] + \log_2 N(S) \quad (2)$$

Proof. Represent S as the Expression (1). Formula (2) follows easily from this expression.

Remark. We see now how to inductively compute entropy values $H(S)$, as follows. If S is of order 0, then $|S| = 1$ and so $H(S) = 0$. If S is of order $j \geq 1$, assume all entropy values for hierarchical type classes of smaller order have been computed. Then Equation (2) is used to compute $H(S)$.

Discussion. Let $\{S_j : j \geq 1\}$ be a sequence of hierarchical type classes from $\mathcal{S}_{m,k}$ such that S_j is of order j ($j \geq 1$). Consider the sequence of normalized entropies $\{H(S_j)/k^j : j \geq 1\}$. As j becomes large, the normalized entropy $H(S_j)/k^j$ approximates more and more closely the compression

rate in bits per data sample that results from the compression scheme on S_j . It is therefore of interest to determine circumstances under which such a sequence of normalized entropies will have a limit that we can compute. We discuss our approach to this problem, which will be pursued in the rest of this paper. A *hierarchical source* is defined to be a family $\{S(\lambda) : \lambda \in \Lambda(m, k)\}$ in which each $S(\lambda)$ is a hierarchical type class selected from $\mathcal{S}_{m,k}(\lambda)$. (We will also impose a natural consistency condition on how these selections are made in our formal hierarchical source definition to be given in the next section.) Let $\mathbb{R}$ denote the real line, and let $\mathbb{P}_m$ be the subset of $\mathbb{R}^m$ consisting of all m -dimensional probability vectors. We consider $\mathbb{P}_m$ to be a metric space with the Euclidean metric. For each $\lambda \in \Lambda(m, k)$, let p_λ be the probability vector $\lambda/\|\lambda\|$ in $\mathbb{P}_m$. Suppose there exists a (necessarily unique) continuous function $h : \mathbb{P}_m \rightarrow [0, \infty)$ such that for each $p \in \mathbb{P}_m$, and each sequence $\{\lambda_j : j \geq 0\}$ for which $\lambda_j \in \Lambda_j(m, k)$ ($j \geq 0$) and $\lim_{j \rightarrow \infty} p_{\lambda_j} = p$, the limit property

$$h(p) = \lim_{j \rightarrow \infty} H(S(\lambda_j))/k^j$$

holds. Then we call the function h the *hierarchical entropy function induced by the source* $\{S(\lambda) : \lambda \in \Lambda(m, k)\}$. A *hierarchical entropy function* is defined to be any function on $\mathbb{P}_m$ which is the hierarchical entropy function induced by some hierarchical source. One of the goals of hierarchical data compression theory is to identify hierarchical entropy functions and to learn about their properties. In the paper [2], two hierarchical entropy functions were introduced. In the present paper, we go further by identifying infinitely many hierarchical entropy functions which are each self-affine, and for each one of these entropy functions, we exhibit an explicit iterated function system whose attractor is the graph of the entropy function.

2. Hierarchical Sources

This section is devoted to the discussion of hierarchical sources. The concept of hierarchical source was informally described in the Introduction. In Section 2.1., we make this concept formal. In Section 2.2., we define the entropy-stable hierarchical sources, which are the hierarchical sources that induce hierarchical entropy functions. In Section 2.3., we introduce a particular type of entropy-stable hierarchical source called finitary hierarchical source. The finitary hierarchical sources induce the hierarchical entropy functions that are the subject of this paper.

2.1. Formal Definition of Hierarchical Source

Let $\mathcal{S} = \{S(\lambda) : \lambda \in \Lambda(m, k)\}$ be a family of hierarchical type classes in which each class $S(\lambda)$ belongs to the set of classes $\mathcal{S}_{m,k}(\lambda)$. Then $\mathcal{S}$ is defined to be a $(\Lambda(m, k)$ -indexed) hierarchical source if the following additional condition is satisfied.

- **Consistency Condition:** For each $S \in \mathcal{S}$ of order > 0 , each term in the k -tuple $(S_1, S_2, \dots, S_k)$ associated with S in Prop. 4 also belongs to $\mathcal{S}$.

We discuss how the Consistency Condition gives us a way to describe every possible hierarchical source. Let $\Lambda(m, k)^k$ be the set of all k -tuples whose entries come from $\Lambda(m, k)$. Let $\Phi(m, k)$ be the set of all mappings $\phi : \Lambda(m, k)^+ \rightarrow \Lambda(m, k)^k$ such that whenever $\phi(\lambda) = (\lambda_1, \lambda_2, \dots, \lambda_k)$, we have

- $\lambda = \sum_{i=1}^k \lambda_i$.
- If λ is of order j , then each entry λ_i of $\phi(\lambda)$ is of order $j - 1$.

Each $\phi \in \Phi(m, k)$ gives rise to a $\Lambda(m, k)$ -indexed hierarchical source $\mathcal{S}^\phi = \{S^\phi(\lambda) : \lambda \in \Lambda(m, k)\}$, defined inductively as follows.

- If $\lambda \in \Lambda(m, k)$ is of order 0, define class $S^\phi(\lambda)$ to be the set $\{a_i\}$, where a_i is the unique letter in $\mathcal{A}_m$ whose type is λ .
- If $\lambda \in \Lambda(m, k)^+$, assume class $S^\phi(\lambda^*)$ has been defined for all types λ^* of order less than the order of λ . Letting $\phi(\lambda) = (\lambda_1, \lambda_2, \dots, \lambda_k)$, define

$$S^\phi(\lambda) \triangleq \bigcup_{\pi \in \Pi_k} [S^\phi(\lambda_{\pi(1)}) * S^\phi(\lambda_{\pi(2)}) * \dots * S^\phi(\lambda_{\pi(k)})]$$

From the Consistency Condition, all possible hierarchical sources arise in this way, that is, given any $\Lambda(m, k)$ -indexed hierarchical source $\mathcal{S}$, there exists $\phi \in \Phi(m, k)$ such that $\mathcal{S} = \mathcal{S}^\phi$.

Another advantage of the Consistency Condition is that it allows the entropies of the classes in a hierarchical source to be recursively computed. To see this, let $\mathcal{S} = \{S(\lambda) : \lambda \in \Lambda(m, k)\}$ be a $\Lambda(m, k)$ -indexed hierarchical source and choose $\phi \in \Phi(m, k)$ such that $\mathcal{S} = \mathcal{S}^\phi$. Define $H_\phi : \Lambda(m, k) \rightarrow [0, \infty)$ to be the function which takes the value zero on $\Lambda_0(m, k)$, and for each $\lambda \in \Lambda^+(m, k)$,

$$H_\phi(\lambda) = \left[\sum_{i=1}^k H_\phi(\lambda_i) \right] + \log_2 N(\lambda) \quad (3)$$

where $(\lambda_1, \dots, \lambda_k)$ is the k -tuple $\phi(\lambda)$ and $N(\lambda)$ is the number of distinct permutations of this k -tuple. By the Consistency Condition and Lemma 1,

$$H_\phi(\lambda) = H(S(\lambda)), \quad \lambda \in \Lambda(m, k)$$

2.2. Entropy-Stable Hierarchical Sources

The concept of entropy-stable source discussed in this section allows us to formally define the concept of hierarchical entropy function.

For each $j \geq 0$, define the finite set of probability vectors

$$\mathbb{P}_m(j) \triangleq \{p_\lambda : \lambda \in \Lambda_j(m, k)\}$$

where the reader will recall that $p_\lambda = \lambda / \|\lambda\|$. Note that the sets $\{\mathbb{P}_m(j) : j \geq 0\}$ are increasing in the sense that

$$\mathbb{P}_m(j) \subset \mathbb{P}_m(j+1), \quad j \geq 0 \quad (4)$$

Let $\mathbb{P}_m^*$ be the countably infinite set of probability vectors which is the union of the $\mathbb{P}_m(j)$'s.

Suppose we have a hierarchical source $\mathcal{S} = \{S(\lambda) : \lambda \in \Lambda(m, k)\}$. For each $j \geq 0$, let $h_j : \mathbb{P}_m(j) \rightarrow [0, \infty)$ be the unique function for which

$$h_j(p_\lambda) = H(S(\lambda)) / \|\lambda\|, \quad \lambda \in \Lambda_j(m, k)$$

Suppose $p \in \mathbb{P}_m^*$. Because of the increasing sets property Equation (4), p is a member of the set $\mathbb{P}_m(j)$ for j sufficiently large. Consequently, $h_j(p)$ is defined for j sufficiently large, and so it makes sense to

talk about the limit of the sequence $\{h_j(p) : j \geq 0\}$, if this limit exists. We define the source $\mathcal{S}$ to be *entropy-stable* if there exists a continuous function $h : \mathbb{P}_m \rightarrow [0, \infty)$ such that

$$h(p) = \lim_{j \rightarrow \infty} h_j(p), \quad p \in \mathbb{P}_m^*$$

and the function h (which is unique since $\mathbb{P}_m^*$ is dense in $\mathbb{P}_m$) is called the hierarchical entropy function induced by $\mathcal{S}$. Henceforth, the terminology “hierarchical entropy function” denotes a function which is the hierarchical entropy function induced by some entropy-stable hierarchical source.

2.3. Finitary Hierarchical Sources

If $\lambda = (n_1, n_2, \dots, n_m)$ is a type in $\Lambda(m, k)^+$, define

$$r(\lambda) \triangleq (\text{mod}(n_1, k), \text{mod}(n_2, k), \dots, \text{mod}(n_m, k))$$

where $\text{mod}(n, k) \in \{0, 1, \dots, k-1\}$ is the remainder upon division of n by k . Each entry of $r(\lambda)$ belongs to the set $\{0, 1, \dots, k-1\}$ and the sum of the entries of $r(\lambda)$ is an integer multiple of k .

Definitions.

- $\mathcal{R}(m, k)$ is defined to be the set of all m -tuples whose entries come from $\{0, 1, \dots, k-1\}$ and sum to an integer multiple of k .
- $\Psi(m, k)$ is defined to be the set of all mappings ψ from $\mathcal{R}(m, k)$ to the set of binary $k \times m$ matrices such that if $r = (r_1, \dots, r_m)$ belongs to $\mathcal{R}(m, k)$, then $\psi(r)$ has left-to-right column sums $r_1, r_2, \dots, r_m$ and row sums all equal to $(r_1 + r_2 + \dots + r_m)/k$. The set $\Psi(m, k)$ is nonempty for each choice of parameters $m, k \geq 2$ [4,5].
- If $\psi \in \Psi(m, k)$, define ψ^* to be the unique mapping in $\Phi(m, k)$ which does the following. If $\lambda = (n_1, n_2, \dots, n_m)$ belongs to $\Lambda(m, k)^+$, let $A = \psi(r(\lambda))$. Then $\psi^*(\lambda) = (\lambda_1, \lambda_2, \dots, \lambda_k)$, where

$$\lambda_i = (\lfloor n_1/k \rfloor, \lfloor n_2/k \rfloor, \dots, \lfloor n_m/k \rfloor) + A(i, 1 : m), \quad i = 1, 2, \dots, k$$

with $A(i, 1 : m)$ denoting the i -th row of A .

- Suppose $\psi \in \Psi(m, k)$ and let $\phi = \psi^*$. The $\Lambda(m, k)$ -indexed hierarchical source $\{S^\phi : \lambda \in \Lambda(m, k)\}$ defines a *finitary source*. For each choice of parameters $m, k \geq 2$, since $\Psi(m, k)$ is nonempty, there is at least one finitary $\Lambda(m, k)$ -indexed hierarchical source. The word “finitary” is used to describe these sources because they are each definable in finite terms by the specification of $mk|\mathcal{R}(m, k)|$ bits (the elements of a number of $k \times m$ binary matrices).

Example 2. Note that (1122) belongs to $\mathcal{R}(4, 3)$. Suppose

$$\psi(1122) = \begin{bmatrix} 1100 \\ 0011 \\ 0011 \end{bmatrix}$$

Note that $(7758) \in \Lambda^+(4, 3)$, and that

$$r(7758) = \text{mod}((7758), 3) = (1122)$$

Since $\lfloor (7758)/3 \rfloor = (2212)$, we see that $\psi^*(7758) = (\lambda_1, \lambda_2, \lambda_3)$, where

$$\lambda_1 = (2212) + (1100) = (3312)$$

$$\lambda_2 = (2212) + (0011) = (2223)$$

$$\lambda_3 = (2212) + (0011) = (2223)$$

Note that the splitting up of (7758) into the three types (3312), (2223), (2223) indeed does make sense because these latter three types sum to (7758) and are of order 2, one less than the order of (7758).

Example 3. Fix the alphabet cardinality parameter to be 2, and fix the partitioning parameter k to be any integer ≥ 2 . Let (r_1, r_2) belong to $\mathcal{R}(2, k)$. Then either (a) $(r_1, r_2) = (0, 0)$ or (b) $r_1 + r_2 = k$. In case (a), we define $\psi(r_1, r_2)$ to be the $k \times 2$ zero matrix. In case (b), we define $\psi(r_1, r_2)$ to be the $k \times 2$ matrix whose first r_1 rows are (1, 0) and whose last r_2 rows are (0, 1). Letting $\phi = \psi^*$, we obtain finitary $\Lambda(2, k)$ -indexed hierarchical source $\mathcal{S}^\phi$.

Example 4. Now fix the alphabet cardinality parameter to be 3, and fix the partitioning parameter k to be any integer ≥ 2 . Let (r_1, r_2, r_3) belong to $\mathcal{R}(3, k)$. Then either (a) $(r_1, r_2, r_3) = (0, 0, 0)$; (b) $r_1 + r_2 + r_3 = k$; or (c) $r_1 + r_2 + r_3 = 2k$. In case (a), we define $\psi(r_1, r_2, r_3)$ to be the $k \times 3$ zero matrix. In case (b), we define $\psi(r_1, r_2, r_3)$ to be the $k \times 3$ matrix whose first r_1 rows are (100), whose next r_2 rows are (010), and whose last r_3 rows are (001). In case (c), we define $\psi(r_1, r_2, r_3)$ to be the $k \times 3$ matrix whose first $k - r_1$ rows are (011), whose next $k - r_2$ rows are (101), and whose last $k - r_3$ rows are (110). Letting $\phi = \psi^*$, we obtain finitary $\Lambda(3, k)$ -indexed hierarchical source $\mathcal{S}^\phi$.

Remarks. For each fixed $k \geq 2$,

- The source defined in Example 3 is the unique finitary $\Lambda(2, k)$ -indexed hierarchical source.
- The source defined in Example 4 is the unique finitary $\Lambda(3, k)$ -indexed hierarchical source.

This is because the matrices employed in these examples are unique up to row permutation.

Theorem 1. Let $m, k \geq 2$ be arbitrary, and let $\{S(\lambda) : \lambda \in \Lambda(m, k)\}$ be any finitary $\Lambda(m, k)$ -indexed hierarchical source. Then the source is entropy-stable and the hierarchical entropy function induced by the source can be characterized as the unique continuous function $h : \mathbb{P}_m \rightarrow [0, \infty)$ such that

$$h(\lambda/\|\lambda\|) = H(S(\lambda))/\|\lambda\|, \quad \lambda \in \Lambda(m, k)$$

Theme 1 is proved in Appendix A.

Notations and Remarks.

- Fix k to be an arbitrary integer ≥ 2 . Let $\{S(\lambda) : \lambda \in \Lambda(2, k)\}$ be the unique finitary $\Lambda(2, k)$ -indexed hierarchical source. $H_{2,k} : \Lambda(2, k) \rightarrow [0, \infty)$ shall denote the entropy function

$$H_{2,k}(\lambda) = H(S(\lambda)), \quad \lambda \in \Lambda(2, k)$$

For later use, we remark that

$$H_{2,k}(n_1, n_2) = \log_2 \left[\frac{k!}{n_1! n_2!} \right] = \log_2 \binom{k}{n_1}, \quad (n_1, n_2) \in \Lambda_1(2, k) \quad (5)$$

The hierarchical entropy function induced by this source maps $\mathbb{P}_2$ into $[0, \infty)$ and shall be denoted $h_{2,k}$. The relationship between functions $H_{2,k}$ and $h_{2,k}$ is

$$h_{2,k}(\lambda/\|\lambda\|) = H_{2,k}(\lambda)/\|\lambda\|, \quad \lambda \in \Lambda(2, k) \quad (6)$$

- Fix k to be an arbitrary integer ≥ 2 . Let $\{S(\lambda) : \lambda \in \Lambda(3, k)\}$ be the unique finitary $\Lambda(3, k)$ -indexed hierarchical source. $H_{3,k} : \Lambda(3, k) \rightarrow [0, \infty)$ shall denote the entropy function

$$H_{3,k}(\lambda) = H(S(\lambda)), \quad \lambda \in \Lambda(3, k)$$

For later use, we remark that

$$H_{3,k}(n_1, n_2, n_3) = \log_2 \left[\frac{k!}{n_1! n_2! n_3!} \right], \quad (n_1, n_2, n_3) \in \Lambda_1(3, k) \quad (7)$$

The hierarchical entropy function induced by this source maps $\mathbb{P}_3$ into $[0, \infty)$ and shall be denoted $h_{3,k}$. The relationship between functions $H_{3,k}$ and $h_{3,k}$ is

$$h_{3,k}(\lambda/\|\lambda\|) = H_{3,k}(\lambda)/\|\lambda\|, \quad \lambda \in \Lambda(3, k) \quad (8)$$

In Section 3, we show that hierarchical entropy function $h_{2,k}$ is self-affine for each $k \geq 2$, and in Section 4, we show that hierarchical entropy function $h_{3,k}$ is self-affine for each $k \geq 2$.

3. $h_{2,k}$ Is Self-Affine

An iterated function system (IFS) on a closed nonempty subset Ω of a finite-dimensional Euclidean space is a finite nonempty set of mappings which map Ω into itself and are each contraction mappings. Given an IFS $\mathcal{T}$ on Ω , there exists ([6], Theme 9.1) a unique nonempty compact set $Q \subset \Omega$ such that

$$Q = \bigcup_{T \in \mathcal{T}} T(Q)$$

Q is called the attractor of the IFS $\mathcal{T}$.

Suppose $h : \mathbb{P}_m \rightarrow [0, \infty)$ is the hierarchical entropy function induced by an entropy-stable $\Lambda(m, k)$ -indexed hierarchical source. Let $\Omega_m = \mathbb{P}_m \times \mathbb{R}$, regarded as a metric space with the Euclidean metric that it inherits from being regarded as a closed convex subset of $\mathbb{R}^{m+1}$. We define h to be self-affine if there is an IFS $\mathcal{T}$ on Ω_m such that

- Each mapping in $\mathcal{T}$ is an affine mapping.
- The attractor of $\mathcal{T}$ is $\{(p, h(p)) : p \in \mathbb{P}_m\}$, the graph of h .

For the rest of this section, $k \geq 2$ is fixed. Our goal is to show that the function $h_{2,k} : \mathbb{P}_2 \rightarrow [0, \infty)$ is self-affine, where $h_{2,k}$ is the hierarchical entropy function induced by the unique finitary $\Lambda(2, k)$ -indexed hierarchical source.

For each $i = 0, 1, \dots, k-1$,

- Define the matrix

$$M_i \triangleq \begin{bmatrix} i+1 & k-i-1 \\ i & k-i \end{bmatrix}$$

- Define $T_i^* : \mathbb{P}_2 \rightarrow \mathbb{P}_2$ to be the mapping

$$T_i^*(p) \triangleq k^{-1} p M_i, \quad p \in \mathbb{P}_2 \quad (9)$$

- Define the vector

$$v_i \triangleq \left(\log_2 \binom{k}{i+1}, \log_2 \binom{k}{i} \right)$$

- Define $T_i : \Omega_2 \rightarrow \Omega_2$ to be the mapping

$$T_i(p, y) \triangleq (T_i^*(p), k^{-1}y + k^{-1}p \cdot v_i), \quad (p, y) \in \Omega_2 \quad (10)$$

where $p \cdot v_i$ denotes the usual dot product.

Remarks. It is clear that the set of mappings $\{T_i^* : i = 0, 1, \dots, k-1\}$ is an IFS on $\mathbb{P}_2$. This fact allows one to prove (Lemma B.3 of Appendix B) that the related set of mappings $\{T_i : i = 0, 1, \dots, k-1\}$ is an IFS on Ω_2 . This result is the first part of the following theorem.

Theorem 2. Let $k \geq 2$ be arbitrary. The following statements hold:

- **(a):** $\{T_0, T_1, \dots, T_{k-1}\}$ is an IFS on Ω_2 .
- **(b):** $h_{2,k}$ is self-affine and its graph is the attractor of the IFS in (a).
- **(c):** For each $i = 0, 1, \dots, k-1$,

$$T_i(p, h_{2,k}(p)) = (T_i^*(p), h_{2,k}(T_i^*(p))), \quad p \in \mathbb{P}_2 \quad (11)$$

Our proof of Theme 2 requires the following lemma.

Lemma 2. Let $\phi \in \Phi(2, k)$ be the function in Example 3 such that $\mathcal{S}^\phi$ is the unique finitary $\Lambda(2, k)$ -indexed hierarchical source. For each $i = 0, 1, \dots, k-1$,

- **(a.1):** If $\lambda \in \Lambda(2, k)$, then $\lambda M_i \in \Lambda(2, k)$ and $\|\lambda M_i\| = k\|\lambda\|$;
- **(a.2):** If $\lambda \in \Lambda(2, k)^+$ and $\phi(\lambda) = (\lambda_1, \lambda_2, \dots, \lambda_k)$, then

$$\phi(\lambda M_i) = (\lambda_1 M_i, \lambda_2 M_i, \dots, \lambda_k M_i)$$

Proof. Property (a.1), whose proof we omit, is a simple consequence of the fact that M_i has row sums equal to k . Fix a type λ from $\Lambda(2, k)^+$. Letting $\phi(\lambda) = (\lambda_1, \lambda_2, \dots, \lambda_k)$ and letting $\phi(\lambda M_i) = (\mu_1, \mu_2, \dots, \mu_k)$, we show $\mu_s = \lambda_s M_i$ ($s = 1, \dots, k$), which will establish Property (a.2). Write λ in the form

$$\lambda = (kq_1 + r_1, kq_2 + r_2)$$

where $r = (r_1, r_2)$. As remarked in Example 3, either $r_1 = r_2 = 0$, or $r_1 + r_2 = k$. Let us first handle the case $r_1 + r_2 = k$. Then

$$\begin{aligned} \lambda_s &= (q_1, q_2) + (1, 0), & 1 \leq s \leq r_1 \\ \lambda_s &= (q_1, q_2) + (0, 1), & r_1 + 1 \leq s \leq k \end{aligned}$$

It is easy to show that

$$\lambda M_i = (kq'_1 + r_1, kq'_2 + r_2)$$

where

$$q'_1 = (i+1)q_1 + iq_2 + i, \quad q'_2 = (k-i-1)q_1 + (k-i)q_2 + k-i-1$$

It follows that

$$\mu_s = (q'_1, q'_2) + (1, 0), \quad 1 \leq s \leq r_1$$

$$\mu_s = (q'_1, q'_2) + (0, 1), \quad r_1 + 1 \leq s \leq k$$

For $1 \leq s \leq r_1$, we have

$$\lambda_s M_i = (q_1 + 1, q_2) M_i = (q_1(i + 1) + q_2 i + i + 1, q_1(k - i - 1) + q_2(k - i) + k - i - 1) = (q'_1 + 1, q'_2) = \mu_s$$

For $r_1 + 1 \leq s \leq k$, we have

$$\lambda_s M_i = (q_1, q_2 + 1) M_i = (q_1(i + 1) + q_2 i + i, q_1(k - i - 1) + q_2(k - i) + k - i) = (q'_1, q'_2 + 1) = \mu_s$$

The remaining case $r_1 = r_2 = 0$ is much easier. We have

$$\begin{aligned} \lambda &= (q_1 k, q_2 k) \\ \lambda_s &= (q_1, q_2), \quad 1 \leq s \leq k \\ \lambda M_i &= (q_1 k(i + 1) + q_2 k i, q_1 k(k - i - 1) + q_2 k(k - i)) \\ \mu_s &= (q_1(i + 1) + q_2 i, q_1(k - i - 1) + q_2(k - i)) = \lambda_s M_i, \quad 1 \leq s \leq k \end{aligned}$$

Proof of Theme 2. We first derive part(c) and then part(b) (part(a) is already taken care of, as remarked previously). We derive part(c) by establishing Equation (11) for a fixed $i \in \{0, 1, \dots, k - 1\}$. Let $\phi \in \Phi(2, k)$ be the function given in Example 3 and recall that $H_{2,k}$ denotes the entropy function H_ϕ on $\Lambda(2, k)$. Referring to the definition of T_i^* in Equation (9) and T_i in Equation (10), we see that proving Equation (11) is equivalent to proving

$$h_{2,k}(k^{-1} p M_i) = k^{-1} h_{2,k}(p) + k^{-1} p \cdot v_i, \quad p \in \mathbb{P}_2 \quad (12)$$

We first show that

$$H_{2,k}(\lambda M_i) = H_{2,k}(\lambda) + \lambda \cdot v_i, \quad \lambda \in \Lambda(2, k) \quad (13)$$

Our proof of Equation (13) is by induction on $\|\lambda\|$. We first must verify Equation (13) for $\|\lambda\| = 1$, which is the two cases $\lambda = (1, 0)$ and $\lambda = (0, 1)$. For $\lambda = (1, 0)$, the left side of Equation (13) is the entropy of the first row of M_i , which by Equation (5) is

$$H_{2,k}(i + 1, k - i - 1) = \log_2 \binom{k}{i + 1}$$

and the right side is

$$H_{2,k}(1, 0) + (1, 0) \cdot v_i = 0 + \log_2 \binom{k}{i + 1}$$

Similarly, if $\lambda = (0, 1)$, both sides of Equation (13) are equal to $\log_2 \binom{k}{i}$. Fix $\lambda^* \in \Lambda(2, k)$ for which $\|\lambda^*\| > 1$, and for the induction hypothesis assume that Equation (13) holds when $\|\lambda\|$ is smaller than $\|\lambda^*\|$. The proof by induction is then completed by showing that Equation (13) holds for $\lambda = \lambda^*$. Let

$$\phi(\lambda^*) = (\lambda_1, \lambda_2, \dots, \lambda_k)$$

By the induction hypothesis,

$$H_{2,k}(\lambda_s M_i) = H_{2,k}(\lambda_s) + \lambda_s \cdot v_i, \quad s = 1, 2, \dots, k$$

Adding,

$$\sum_{s=1}^k H_{2,k}(\lambda_s M_i) = \left[\sum_{s=1}^k H_{2,k}(\lambda_s) \right] + \lambda^* \cdot v_i \quad (14)$$

By Lemma 2,

$$\phi(\lambda^* M_i) = (\lambda_1 M_i, \lambda_2 M_i, \dots, \lambda_k M_i)$$

Appealing to Equation (3), we then have

$$\sum_{s=1}^k H_{2,k}(\lambda_s M_i) = H_{2,k}(\lambda^* M_i) - \log_2 N$$

where N is the number of permutations of the k -tuple $(\lambda_1 M_i, \dots, \lambda_k M_i)$. Similarly,

$$\sum_{s=1}^k H_{2,k}(\lambda_s) = H_{2,k}(\lambda^*) - \log_2 N_2$$

where N_2 is the number of permutations of the k -tuple $(\lambda_1, \dots, \lambda_k)$. Since M_i is nonsingular (its determinant is k), we have $N = N_2$. Substituting the right hand sides of the previous two equations into Equation (14), we obtain Equation (13) for $\lambda = \lambda^*$, completing the proof by induction. Dividing both sides of Equation (13) by $\|\lambda\|$, and using the fact that $\|\lambda M_i\| = k\|\lambda\|$, we see that

$$kH_{2,k}(\lambda M_i)/\|\lambda M_i\| = (H_{2,k}(\lambda)/\|\lambda\|) + p_\lambda \cdot v_i$$

which by Equation (6) becomes

$$kh_{2,k}((\lambda M_i)/\|\lambda M_i\|) = h_{2,k}(p_\lambda) + p_\lambda \cdot v_i$$

It is easy to see that

$$(\lambda M_i)/\|\lambda M_i\| = k^{-1}p_\lambda M_i$$

Therefore,

$$h_{2,k}(k^{-1}p_\lambda M_i) = k^{-1}h_{2,k}(p_\lambda) + k^{-1}(p_\lambda \cdot v_i)$$

Equation (12) then follows since the set $\mathbb{P}_2^* = \{p_\lambda : \lambda \in \Lambda(2, k)\}$ is dense in $\mathbb{P}_2$ and $h_{2,k}$ is a continuous function on $\mathbb{P}_2$, completing the derivation of part(c) of Theme 2. All that remains is to prove part(b) of Theme 2. Let $G = \{(p, h_{2,k}(p)) : p \in \mathbb{P}_2\}$ be the graph of $h_{2,k}$. Part(c) is equivalent to the property that

$$T_i(G) \subset G, \quad i = 0, 1, \dots, k-1$$

This property, together with the fact that $\{T_i^* : i = 0, 1, \dots, k-1\}$ is an IFS on $\mathbb{P}_2$ with attractor $\mathbb{P}_2$, allows us to conclude that G is the attractor of the IFS $\{T_0, \dots, T_{i-1}\}$ (Lemma B.1 of Appendix B), and $h_{2,k}$ is self-affine because the T_i 's are affine. Theme 2(b) is therefore true.

Generating Hierarchical Entropy Function Plots. For each $k \geq 2$, let $h_{2,k}^* : [0, 1] \rightarrow \mathbb{R}$ be the function

$$h_{2,k}^*(x) = h_{2,k}(x, 1-x), \quad 0 \leq x \leq 1$$

We can obtain k^n points on the plot of $h_{2,k}^*$ as follows. Let $\{T_i : i = 0, 1, \dots, k-1\}$ be the IFS on Ω_2 given in Theme 2, such that the attractor of this IFS is the graph of $h_{2,k}$. Let $S_0(k) = \{(0, 1, 0)\}$, and generate subsets $S_1(k), S_2(k), \dots, S_n(k)$ of $\mathbb{R}^3$ by the recursion

$$S_j(k) = \bigcup_{i=0}^{k-1} T_i(S_{j-1}(k)), \quad j = 1, 2, \dots, n$$

Then $S_n(k)$ consists of k^n points of the form $(x, 1-x, h_{2,k}(x, 1-x))$. Projecting according to

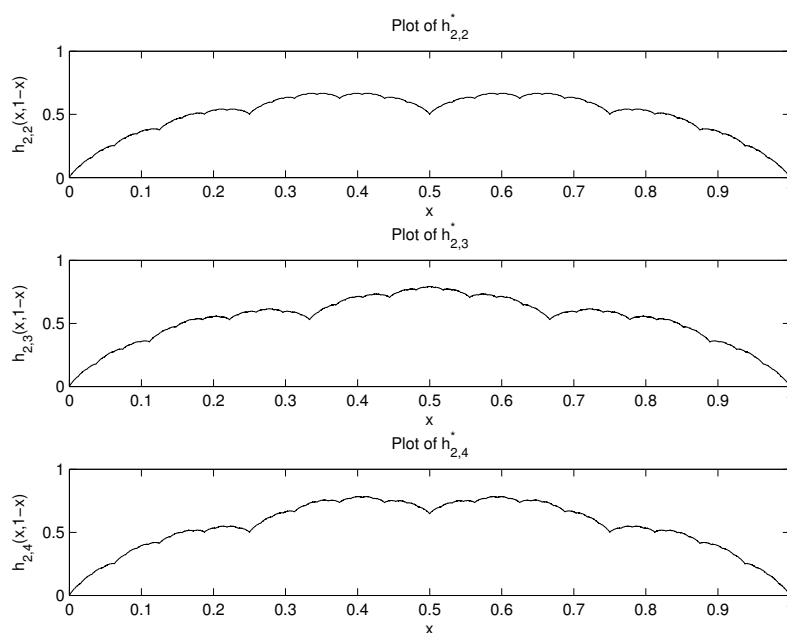
$$(x, 1-x, h_{2,k}(x, 1-x)) \rightarrow (x, h_{2,k}(x, 1-x)) = (x, h_{2,k}^*(x))$$

we obtain k^n points on the plot of $h_{2,k}^*$. Using a Dell Latitude D620 laptop, we did $S_n(k)$ computations to obtain the plots in Figure 2, as follows.

- The plot of $h_{2,2}^*$ used the set $S_{24}(2)$, consisting of $2^{24} = 16777216$ points, computed in 4.2 seconds.
- The plot of $h_{2,3}^*$ used the set $S_{15}(3)$ consisting of $3^{15} = 14348907$ points, computed in 3.3 seconds.
- The plot of $h_{2,4}^*$ used the set $S_{12}(4)$ consisting of $4^{12} = 16777216$ points, computed in 3.5 seconds.

We point out that the functions $h_{2,2}^*$ and $h_{2,4}^*$, although their plots look similar, are not the same. For example, $h_{2,2}^*(1/2) = 1/2$, whereas $h_{2,4}^*(1/2) = \log_2(6)/4 \approx 0.646$.

Figure 2. Plots of $h_{2,k}^*(x) = h_{2,k}(x, 1-x)$ for $k = 2, 3, 4$.

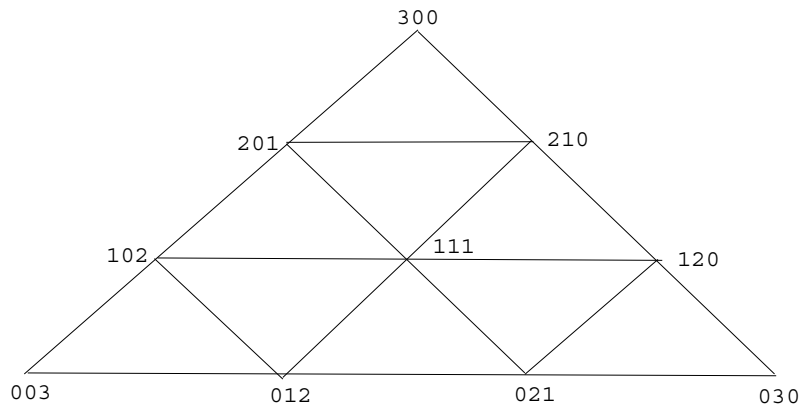


4. $h_{3,k}$ Is Self-Affine

Fix $k \geq 2$. It is the purpose of this section to study $h_{3,k} : \mathbb{P}_3 \rightarrow [0, \infty)$, the hierarchical entropy function induced by the unique finitary $\Lambda(3, k)$ -indexed hierarchical source. In $\mathbb{R}^3$, let Q_k be the convex hull of the set $\{(k, 0, 0), (0, k, 0), (0, 0, k)\}$. Then Q_k is an equilateral triangle whose three vertices are $(k, 0, 0), (0, k, 0), (0, 0, k)$. We employ the well-known *quadratic partition* [7] of Q_k into k^2 congruent equilateral triangles, formed as follows. Partition each of the three sides of Q_k into k line segments of equal length by laying down $k-1$ interior points along the side. For each vertex of Q_k , draw a line

segment connecting the first interior points reached going out from the vertex along its two sides, then draw a line segment connecting the second interior points reached, and so forth until $k - 1$ line segments have been drawn. Doing this for each of the three vertices, you will have drawn a total of $3(k - 1)$ line segments, which subdivide Q_k into the k^2 congruent equilateral triangles of the quadratic partition. See Figure 3, which illustrates the quadratic partition of triangle Q_3 into nine sub-triangles.

Figure 3. Quadratic Partition Of Triangle Q_3 .



Let V_1 be the set of all points (a, b, c) in Q_k such that a is a positive integer and b, c are non-negative integers. There are $k(k + 1)/2$ points in V_1 . For each $v = (a, b, c)$ in V_1 , let $M_{1,v}$ be the 3×3 matrix

$$M_{1,v} = \begin{bmatrix} a & b & c \\ a - 1 & b + 1 & c \\ a - 1 & b & c + 1 \end{bmatrix} \quad (15)$$

For each $v \in V_1$, the convex hull of the rows of $M_{1,v}$ is one of the sub-triangles in the quadratic partition of Q_k , and these sub-triangles are distinct as v varies through V_1 . This gives us a total of $k(k + 1)/2$ of the sub-triangles in the quadratic partition of Q_k , and we call these the V_1 sub-triangles of the partition. Let V_2 be the set of all (a, b, c) in Q_k such that a is a non-negative integer and b, c are positive integers. There are $k(k - 1)/2$ points in V_2 . For each $v = (a, b, c)$ in V_2 , let $M_{2,v}$ be the 3×3 matrix

$$M_{2,v} = \begin{bmatrix} a & b & c \\ a + 1 & b - 1 & c \\ a + 1 & b & c - 1 \end{bmatrix} \quad (16)$$

For each $v \in V_2$, the convex hull of the rows of $M_{2,v}$ is one of the sub-triangles in the quadratic partition of Q_k , and these sub-triangles are distinct as v varies through V_2 . This gives us a total of $k(k - 1)/2$ of the sub-triangles in the quadratic partition of Q_k , and we call these the V_2 sub-triangles of the partition. The V_1 sub-triangles are all translations of each other; the V_2 sub-triangles are all translations of each other and each one can be obtained by rotating a V_1 sub-triangle about its center 180 degrees, followed by a translation. Together, the $k(k + 1)/2$ V_1 sub-triangles and the $k(k - 1)/2$ V_2 sub-triangles constitute all k^2 sub-triangles in the quadratic partition of Q_k .

We define $\mathcal{M}(k)$ to be the set of k^2 matrices

$$\mathcal{M}(k) \triangleq \{M_{1,v} : v \in V_1\} \cup \{M_{2,v} : v \in V_2\}$$

Each row sum of each matrix in $\mathcal{M}(k)$ is equal to k . Because of this property, we can define for each $M \in \mathcal{M}(k)$ the mapping $T_M^* : \mathbb{P}_3 \rightarrow \mathbb{P}_3$ in which

$$T_M^*(p) \triangleq k^{-1}pM, \quad p \in \mathbb{P}_3 \quad (17)$$

and we can also define the mapping $T_M : \Omega_3 \rightarrow \Omega_3$ in which

$$T_M(p, y) \triangleq (T_M^*(p), k^{-1}y + k^{-1}p \cdot v_M), \quad (p, y) \in \Omega_3 \quad (18)$$

where

$$v_M = (H_{3,k}(M(1, 1 : 3)), H_{3,k}(M(2, 1 : 3)), H_{3,k}(M(3, 1 : 3))) \quad (19)$$

Remarks. It is clear that the set of k^2 mappings $\{T_M^* : M \in \mathcal{M}(k)\}$ is an IFS on $\mathbb{P}_3$. This fact allows one to prove (Lemma B.4 of Appendix B) that the related set of k^2 mappings $\{T_M : M \in \mathcal{M}(k)\}$ is an IFS on Ω_3 . In the following example, we exhibit this IFS in a special case.

Example 5. Let $k = 3$. Referring to Figure 3, we see that the 9 matrices in $\mathcal{M}(3)$ are

$$\begin{aligned} M_1 &= \begin{bmatrix} 300 \\ 210 \\ 201 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 201 \\ 111 \\ 102 \end{bmatrix}, \quad M_3 = \begin{bmatrix} 210 \\ 120 \\ 111 \end{bmatrix} \\ M_4 &= \begin{bmatrix} 102 \\ 012 \\ 003 \end{bmatrix}, \quad M_5 = \begin{bmatrix} 111 \\ 021 \\ 012 \end{bmatrix}, \quad M_6 = \begin{bmatrix} 120 \\ 030 \\ 021 \end{bmatrix} \\ M_7 &= \begin{bmatrix} 111 \\ 201 \\ 210 \end{bmatrix}, \quad M_8 = \begin{bmatrix} 012 \\ 102 \\ 111 \end{bmatrix}, \quad M_9 = \begin{bmatrix} 021 \\ 111 \\ 120 \end{bmatrix} \end{aligned}$$

Following Equation (19), let $v_i \in \mathbb{R}^3$ be the vector whose components are the $H_{3,3}$ entropies of the rows of M_i . Letting $\alpha = \log_2 3$ and $\beta = \log_2 6$, Formula (7) is used to obtain

$$\begin{aligned} v_1 &= (0, \alpha, \alpha), \quad v_2 = (\alpha, \beta, \alpha), \quad v_3 = (\alpha, \alpha, \beta) \\ v_4 &= (\alpha, \alpha, 0), \quad v_5 = (\beta, \alpha, \alpha), \quad v_6 = (\alpha, 0, \alpha) \\ v_7 &= (\beta, \alpha, \alpha), \quad v_8 = (\alpha, \alpha, \beta), \quad v_9 = (\alpha, \beta, \alpha) \end{aligned}$$

Following Equation (18), for each $i = 1, 2, \dots, 9$, let $T_i : \Omega_3 \rightarrow \Omega_3$ be the mapping defined by

$$T_i(p, y) \triangleq (pM_i, y + p \cdot v_i)/3, \quad (p, y) \in \Omega_3$$

Theme 3 which follows will tell us that the graph of $h_{3,3}$ is the attractor of the IFS $\{T_i\}_{i=1}^9$.

Theorem 3. Let $k \geq 2$ be arbitrary. The following statements hold:

- **(a):** $\{T_M : M \in \mathcal{M}(k)\}$ is an IFS on Ω_3 .
- **(b):** $h_{3,k}$ is self-affine and its graph is the attractor of the IFS in (a).
- **(c):** For each $M \in \mathcal{M}(k)$,

$$T_M(p, h_{3,k}(p)) = (T_M^*(p), h_{3,k}(T_M^*(p))), \quad p \in \mathbb{P}_3 \quad (20)$$

Our proof of Theme 3 requires a couple of lemmas, which follow.

Lemma 3. Let $\lambda \in \Lambda(3, k)^+$, let $\phi \in \Phi(3, k)$ be the function given in Example 4, and let $\phi(\lambda) = (\lambda_1, \lambda_2, \dots, \lambda_k)$. Suppose we write

$$\lambda = (kq_1 + r_1, kq_2 + r_2, kq_3 + r_3)$$

where the q_i 's are non-negative, the r_i 's belong to the set $\{1, 2, \dots, k\}$, and $r_1 + r_2 + r_3 = 2k$. Then

$$\text{card}\{1 \leq i \leq k : \lambda_i = (q_1, q_2 + 1, q_3 + 1)\} = k - r_1 \quad (21)$$

$$\text{card}\{1 \leq i \leq k : \lambda_i = (q_1 + 1, q_2, q_3 + 1)\} = k - r_2 \quad (22)$$

$$\text{card}\{1 \leq i \leq k : \lambda_i = (q_1 + 1, q_2 + 1, q_3)\} = k - r_3 \quad (23)$$

Proof. If each $r_i < k$, then by definition of $\phi(\lambda)$ in Example 4, the properties Equations (21)–(23) are true. Now suppose at least one $r_i = k$. Then exactly one $r_i = k$ (since otherwise some $r_i = 0$, which is not allowed). By symmetry, we may suppose that $r_1 = k$. We may now express λ as

$$\lambda = ((q_1 + 1)k, q_2k + r_2, q_3k + r_3)$$

Since $r_2, r_3 \in \{1, 2, \dots, k - 1\}$, and $r_2 + r_3 = k$, the definition of $\phi(\lambda)$ tells us that

$$\text{card}\{1 \leq i \leq k : \lambda_i = (q_1 + 1, q_2, q_3) + (0, 1, 0)\} = r_2 \quad (24)$$

$$\text{card}\{1 \leq i \leq k : \lambda_i = (q_1 + 1, q_2, q_3) + (0, 0, 1)\} = r_3 \quad (25)$$

Equation (24) yields Equation (23), Equation (25) yields Equation (22), and Equation (21) is vacuously true because $k - r_1 = 0$.

Lemma 4. Let $\phi \in \Phi(3, k)$ be the function given in Example 4. Properties (a.1)–(a.2) below are true for any matrix M in the set $\mathcal{M}(k)$.

- **(a.1):** If λ is a type in $\Lambda(3, k)$, then $\lambda M \in \Lambda(3, k)$ and $\|\lambda M\| = k\|\lambda\|$;
- **(a.2):** If λ is a type in $\Lambda(3, k)^+$, and $\phi(\lambda) = (\lambda_1, \lambda_2, \dots, \lambda_k)$, then $\phi(\lambda M)$ is some permutation of $(\lambda_1 M, \lambda_2 M, \dots, \lambda_k M)$.

Proof. Property (a.1), whose proof we omit, is a simple consequence of the fact that each matrix in $\mathcal{M}(k)$ has row sums equal to k . Fix $\lambda \in \Lambda(3, k)^+$ and fix $M \in \mathcal{M}(k)$. Let $r(\lambda) = (r_1, r_2, r_3)$, and let

$$\begin{aligned} \lambda &= (kq_1 + r_1, kq_2 + r_2, kq_3 + r_3) \\ \phi(\lambda) &= (\lambda_1, \lambda_2, \dots, \lambda_k) \\ \phi(\lambda M) &= (\mu_1, \mu_2, \dots, \mu_k) \end{aligned}$$

M is either of the form Equation (15) (Case 1) or of the form Equation (16) (Case 2). Throughout the rest of the proof, we employ the parameter $\beta = (r_1 + r_2 + r_3)/k$. As remarked in Example 4, $\beta \in \{0, 1, 2\}$.

Proof for Case 1: We have $\lambda M = (kq'_1 + r_1, kq'_2 + r_2, kq'_3 + r_3)$, where

$$\begin{aligned} q'_1 &\triangleq q_1a + q_2(a - 1) + q_3(a - 1) + \beta(a - 1) \\ q'_2 &\triangleq q_1b + q_2(b + 1) + q_3b + \beta b \\ q'_3 &\triangleq q_1c + q_2c + q_3(c + 1) + \beta c \end{aligned}$$

Note that

$$(q_1, q_2, q_3)M = (q'_1, q'_2, q'_3) - \beta(a - 1, b, c) \quad (26)$$

If $\beta = 0$, then

$$\lambda_i = (q_1, q_2, q_3), \quad \mu_i = (q'_1, q'_2, q'_3), \quad i = 1, 2, \dots, k$$

From Equation (26), we have $(q_1, q_2, q_3)M = (q'_1, q'_2, q'_3)$, and therefore Property (a.2) follows. If $\beta = 1$, by definition of $\phi(\lambda)$ and $\phi(\lambda M)$ in Example 4,

$$\text{card}\{1 \leq i \leq k : \lambda_i = (q_1 + 1, q_2, q_3)\} = r_1 \quad (27)$$

$$\text{card}\{1 \leq i \leq k : \lambda_i = (q_1, q_2 + 1, q_3)\} = r_2 \quad (28)$$

$$\text{card}\{1 \leq i \leq k : \lambda_i = (q_1, q_2, q_3 + 1)\} = r_3 \quad (29)$$

$$\text{card}\{1 \leq i \leq k : \mu_i = (q'_1 + 1, q'_2, q'_3)\} = r_1$$

$$\text{card}\{1 \leq i \leq k : \mu_i = (q'_1, q'_2 + 1, q'_3)\} = r_2$$

$$\text{card}\{1 \leq i \leq k : \mu_i = (q'_1, q'_2, q'_3 + 1)\} = r_3$$

Property (a.2) then follows if the equations

$$(q_1 + 1, q_2, q_3)M = (q'_1 + 1, q'_2, q'_3)$$

$$(q_1, q_2 + 1, q_3)M = (q'_1, q'_2 + 1, q'_3)$$

$$(q_1, q_2, q_3 + 1)M = (q'_1, q'_2, q'_3 + 1)$$

are valid. These three equations can be seen to hold using the fact from Equation (26) that

$$(q_1, q_2, q_3)M = (q'_1, q'_2, q'_3) - (a - 1, b, c)$$

Finally, if $\beta = 2$,

$$\text{card}\{1 \leq i \leq k : \lambda_i = (q_1, q_2 + 1, q_3 + 1)\} = k - r_1 \quad (30)$$

$$\text{card}\{1 \leq i \leq k : \lambda_i = (q_1 + 1, q_2, q_3 + 1)\} = k - r_2 \quad (31)$$

$$\text{card}\{1 \leq i \leq k : \lambda_i = (q_1 + 1, q_2 + 1, q_3)\} = k - r_3 \quad (32)$$

$$\text{card}\{1 \leq i \leq k : \mu_i = (q'_1, q'_2 + 1, q'_3 + 1)\} = k - r_1$$

$$\text{card}\{1 \leq i \leq k : \mu_i = (q'_1 + 1, q'_2, q'_3 + 1)\} = k - r_2$$

$$\text{card}\{1 \leq i \leq k : \mu_i = (q'_1 + 1, q'_2 + 1, q'_3)\} = k - r_3$$

Property (a.2) then follows if the equations

$$(q_1, q_2 + 1, q_3 + 1)M = (q'_1, q'_2 + 1, q'_3 + 1)$$

$$(q_1 + 1, q_2, q_3 + 1)M = (q'_1 + 1, q'_2, q'_3 + 1)$$

$$(q_1 + 1, q_2 + 1, q_3)M = (q'_1 + 1, q'_2 + 1, q'_3)$$

are valid. These equations can be seen to hold using the fact from Equation (26) that

$$(q_1, q_2, q_3)M = (q'_1, q'_2, q'_3) - 2(a - 1, b, c)$$

Proof for Case 2: We have

$$\lambda M = (kq'_1 + r'_1, kq'_2 + r'_2, kq'_3 + r'_3)$$

where

$$\begin{aligned} q'_1 &= q_1a + q_2(a+1) + q_3(a+1) + \beta(a+1) - 1 \\ q'_2 &= q_1b + q_2(b-1) + q_3b + \beta b - 1 \\ q'_3 &= q_1c + q_2c + q_3(c-1) + \beta c - 1 \\ r'_i &= k - r_i, \quad i = 1, 2, 3 \end{aligned}$$

Note that

$$(q_1, q_2, q_3)M = (q'_1, q'_2, q'_3) - \beta(a+1, b, c) + (1, 1, 1) \quad (33)$$

If $\beta = 0$, then

$$\lambda_i = (q_1, q_2, q_3), \quad \mu_i = (q'_1 + 1, q'_2 + 1, q'_3 + 1), \quad i = 1, 2, \dots, k$$

From Equation (33), we have

$$(q_1, q_2, q_3)M = (q'_1 + 1, q'_2 + 1, q'_3 + 1)$$

and therefore Property (a.2) follows. Now suppose $\beta = 1$. The entries of (r'_1, r'_2, r'_3) belong to $\{1, 2, \dots, k\}$ and their sum is $2k$. By Lemma 3,

$$\begin{aligned} \text{card}\{1 \leq s \leq k : \mu_s = (q'_1, q'_2 + 1, q'_3 + 1)\} &= k - r'_1 = r_1 \\ \text{card}\{1 \leq s \leq k : \mu_s = (q'_1 + 1, q'_2, q'_3 + 1)\} &= k - r'_2 = r_2 \\ \text{card}\{1 \leq s \leq k : \mu_s = (q'_1 + 1, q'_2 + 1, q'_3)\} &= k - r'_3 = r_3 \end{aligned}$$

In view of the fact that Equations (27–29) also hold, Property (a.2) then follows if the equations

$$\begin{aligned} (q_1 + 1, q_2, q_3)M &= (q'_1, q'_2 + 1, q'_3 + 1) \\ (q_1, q_2 + 1, q_3)M &= (q'_1 + 1, q'_2, q'_3 + 1) \\ (q_1, q_2, q_3 + 1)M &= (q'_1 + 1, q'_2 + 1, q'_3) \end{aligned}$$

are valid. These equations can be seen to hold using the fact from Equation (33) that

$$(q_1, q_2, q_3)M = (q'_1, q'_2, q'_3) - (a, b - 1, c - 1)$$

Thus, Property (a.2) holds. Finally, suppose that $\beta = 2$. The entries of (r'_1, r'_2, r'_3) belong to $\{1, 2, \dots, k\}$ and their sum is k . Under these conditions, no entry of (r'_1, r'_2, r'_3) can be equal to k , and so all entries belong to the set $\{1, 2, \dots, k - 1\}$. By definition of $\phi(\lambda M)$ in Example 4,

$$\begin{aligned} \text{card}\{1 \leq s \leq k : \mu_s = (q'_1 + 1, q'_2, q'_3)\} &= r'_1 = k - r_1 \\ \text{card}\{1 \leq s \leq k : \mu_s = (q'_1, q'_2 + 1, q'_3)\} &= r'_2 = k - r_2 \\ \text{card}\{1 \leq s \leq k : \mu_s = (q'_1, q'_2, q'_3 + 1)\} &= r'_3 = k - r_3 \end{aligned}$$

In view of the fact that Equation (30–32) also hold, Property (a.2) then follows if the equations

$$\begin{aligned}(q_1, q_2 + 1, q_3 + 1)M &= (q'_1 + 1, q'_2, q'_3) \\ (q_1 + 1, q_2, q_3 + 1)M &= (q'_1, q'_2 + 1, q'_3) \\ (q_1 + 1, q_2 + 1, q_3)M &= (q'_1, q'_2, q'_3 + 1)\end{aligned}$$

are valid. These equations can be seen to hold using the fact from Equation (33) that

$$(q_1, q_2, q_3)M = (q'_1, q'_2, q'_3) - (2a + 1, 2b - 1, 2c - 1)$$

Thus, Property (a.2) holds.

Proof of Theme 3. We first derive part(c) and then part(b) (part(a) is already taken care of, as remarked previously). We derive part(c) by establishing Equation (20) for a fixed $M \in \mathcal{M}(k)$. Let $\phi \in \Phi(3, k)$ be the function given in Example 4 and recall that $H_{3,k}$ denotes the entropy function H_ϕ on $\Lambda(3, k)$. Referring to the definition of T_M^* in Equation (17) and T_M in Equation (18), we see that proving Equation (20) is equivalent to proving

$$h_{3,k}(k^{-1}pM) = k^{-1}h_{3,k}(p) + k^{-1}p \cdot v_M, \quad p \in \mathbb{P}_3 \quad (34)$$

We first show that

$$H_{3,k}(\lambda M) = H_{3,k}(\lambda) + \lambda \cdot v_M, \quad \lambda \in \Lambda(3, k) \quad (35)$$

The proof is by induction on $\|\lambda\|$. Equation (35) holds for $\|\lambda\| = 1$, which is the three cases $\lambda = (1, 0, 0)$, $\lambda = (0, 1, 0)$, $\lambda = (0, 0, 1)$. Fix $\lambda^* \in \Lambda(3, k)$ for which $\|\lambda^*\| > 1$, and for the induction hypothesis assume that Equation (35) holds when $\|\lambda\|$ is smaller than $\|\lambda^*\|$. The proof by induction is then completed by showing that Equation (35) holds for $\lambda = \lambda^*$. Let $\phi(\lambda^*) = (\lambda_1, \lambda_2, \dots, \lambda_k)$. By the induction hypothesis,

$$H_{3,k}(\lambda_i M) = H_{3,k}(\lambda_i) + \lambda_i \cdot v_M, \quad i = 1, 2, \dots, k$$

Adding,

$$\sum_{i=1}^k H_{3,k}(\lambda_i M) = \left[\sum_{i=1}^k H_{3,k}(\lambda_i) \right] + \lambda^* \cdot v_M \quad (36)$$

By Lemma 4, $\phi(\lambda^* M)$ is a permutation of $(\lambda_1 M, \lambda_2 M, \dots, \lambda_k M)$, and so by Equation (3),

$$\sum_{i=1}^k H_{3,k}(\lambda_i M) = H_{3,k}(\lambda^* M) - \log_2 N$$

where N is the number of permutations of the k -tuple $(\lambda_1 M, \dots, \lambda_k M)$. Similarly,

$$\sum_{i=1}^k H_{3,k}(\lambda_i) = H_{3,k}(\lambda^*) - \log_2 N_2$$

where N_2 is the number of permutations of the k -tuple $(\lambda_1, \dots, \lambda_k)$. Since M is nonsingular (its determinant is k), we must have $N = N_2$. Substituting the right hand sides of the previous two equations

into Equation (36), we obtain Equation (35) for $\lambda = \lambda^*$, completing the proof by induction. Dividing both sides of Equation (35) by $\|\lambda\|$, and using the fact that $\|\lambda M\| = k\|\lambda\|$, we see that

$$kH_{3,k}(\lambda M)/\|\lambda M\| = (H_{3,k}(\lambda)/\|\lambda\|) + p_\lambda \cdot v_M$$

which, using Equation (8), becomes

$$kh_{3,k}((\lambda M)/\|\lambda M\|) = h_{3,k}(p_\lambda) + p_\lambda \cdot v_M$$

It is easy to see that

$$(\lambda M)/\|\lambda M\| = k^{-1}p_\lambda M$$

Therefore,

$$h_{3,k}(k^{-1}p_\lambda M) = k^{-1}h_{3,k}(p_\lambda) + k^{-1}(p_\lambda \cdot v_M)$$

Equation (34) then follows since the set $\mathbb{P}_3^* = \{p_\lambda : \lambda \in \Lambda(3, k)\}$ is dense in $\mathbb{P}_3$ and $h_{3,k}$ is a continuous function on $\mathbb{P}_3$, completing the derivation of part(c) of Theme 3. All that remains is to prove part(b) of Theme 3. Letting $G = \{(p, h_{3,k}(p)) : p \in \mathbb{P}_3\}$ be the graph of $h_{3,k}$, part(c) is equivalent to the property that

$$T_M(G) \subset G, \quad M \in \mathcal{M}(k) \quad (37)$$

Note that

$$\bigcup \{T_M^*(\mathbb{P}_3) : M \in \mathcal{M}(k)\} = \mathbb{P}_3$$

since the sets in the union form the quadratic partition of $\mathbb{P}_3$, and so $\mathbb{P}_3$ must be the attractor of the IFS $\{T_M^* : M \in \mathcal{M}(k)\}$. This fact, together with Equation (37), allows us to conclude (via Lemma B.1 of Appendix B) that G is the attractor of the IFS $\{T_M : M \in \mathcal{M}(k)\}$, and $h_{3,k}$ is self-affine because the T_M 's are affine. Theme 3(b) is therefore true.

5. Properties of Hierarchical Entropy Functions

We conclude the paper with a discussion of some properties of the self-affine hierarchical entropy functions $h_{2,k}$ and $h_{3,k}$. For each $m \in \{2, 3\}$ and each $k \geq 2$, hierarchical entropy function $h_{m,k}$ obeys the following properties.

- **P1:** $h_{m,k}$ is a continuous function on $\mathbb{P}_m$.
- **P2:** If two probability vectors p_1, p_2 in $\mathbb{P}_m$ are permutations of each other, then

$$h_{m,k}(p_1) = h_{m,k}(p_2)$$

- **P3:** If $p \in \mathbb{P}_m$ is degenerate (meaning that it is a permutation of the vector $(1, 0, 0, \dots, 0)$), then $h_{m,k}(p) = 0$.
- **P4:** For each $p \in \mathbb{P}_m$,

$$0 \leq h_{m,k}(p) \leq \log_2 m$$

Properties P1-P4 are simple consequences of what has gone before. For example, to see why the symmetry property P2 is true, first observe that $H_{m,k}(\lambda_1) = H_{m,k}(\lambda_2)$ if λ_1, λ_2 are types which are

permutations of each other; this symmetry property for entropy on types then extends to $\mathbb{P}_m$ using the fact that the finitary source which induces $h_{m,k}$ is entropy-stable.

The well-known Shannon entropy function h_m on $\mathbb{P}_m$ is defined by

$$h_m(p_1, p_2, \dots, p_m) \triangleq \sum_{i=1}^m -p_i \log_2 p_i$$

where $p_i \log_2 p_i$ is taken to be zero if $p_i = 0$. We point out that h_m also satisfies properties P1-P4. In addition, h_m satisfies the property that it attains its maximum value at the equiprobable distribution $(1/m, 1/m, \dots, 1/m)$. This property fails in general for the $h_{m,k}$ functions, although it is true for some of them; for example, referring to Figure 2, we see that $h_{2,2}$ and $h_{2,4}$ do not reach their maximum at $(1/2, 1/2)$, but $h_{2,3}$ does. It is an open problem to determine the maximum value of each $h_{2,k}$ and $h_{3,k}$ and to see where the maximum is attained.

The inequality

$$h_{m,k}(p) \leq h_m(p), \quad p \in \mathbb{P}_m, \quad m \in \{2, 3\}, \quad k \geq 2$$

gives us a relationship between hierarchical entropy and Shannon entropy; it follows from the fact that every string in a hierarchical type class is of the same type. It is an open problem whether this inequality is strict at every non-degenerate $p \in \mathbb{P}_m$; we have proved this strict inequality property in some special cases (for example, $m = k = 2$).

Appendix A

In this Appendix, we prove Theme 1. In the following, the infinity norm $\|x\|_\infty$ of a vector $x = (x_1, x_2, \dots, x_m) \in \mathbb{R}^m$ is defined as $\max_i |x_i|$.

Lemma A.1. Let $f : \mathbb{P}_m^* \rightarrow \mathbb{R}$ be a function, and let

$$\epsilon_j = \max\{|f(p_{\lambda_1}) - f(p_{\lambda_2})| : \|\lambda_1 - \lambda_2\|_\infty \leq 1, \quad \lambda_1, \lambda_2 \in \Lambda_j(m, k)\}, \quad j \geq 0 \quad (38)$$

If $\sum_{j=0}^{\infty} \epsilon_j < \infty$, then f is uniformly continuous on $\mathbb{P}_m^*$.

Proof. We show there exists $B > 0$ such that

$$\sup\{|f(q_1) - f(q_2)| : \|q_1 - q_2\|_\infty \leq k^{-J}, \quad q_1, q_2 \in \mathbb{P}_m^*\} \leq B \sum_{j=J}^{\infty} \epsilon_j, \quad J \geq 0$$

from which the uniform continuity follows. It can be shown that the following two properties hold.

- **(p.1):** For each $j \geq 0$ and each pair of distinct types $\lambda_0, \lambda \in \Lambda_j(m, k)$, the following is true. Letting $I = m\|\lambda_0 - \lambda\|_\infty$, there exist types $\lambda_1, \lambda_2, \dots, \lambda_I$ in $\Lambda_j(m, k)$ such that $\lambda_I = \lambda$ and

$$\|\lambda_i - \lambda_{i-1}\|_\infty \leq 1, \quad i = 1, 2, \dots, I$$

(In other words, we can travel from λ_0 to λ via a path in $\Lambda_j(m, k)$ consisting of I terms, with successive terms no more than distance 1 apart in the infinity norm.)

- **(p.2):** There is a positive integer M for which the following is true. For each $j \geq 1$ and each $\lambda_0 \in \Lambda_j(m, k)$, there exist types $\lambda_1, \lambda_2, \dots, \lambda_M$ in $\Lambda_j(m, k)$ such that $\lambda_M/k \in \Lambda_{j-1}(m, k)$ and

$$\|\lambda_i - \lambda_{i-1}\|_\infty \leq 1, \quad i = 1, 2, \dots, M$$

(In other words, we can travel in $\Lambda_j(m, k)$ from any type to a type divisible by k via a path consisting of M terms, with successive terms no more than distance 1 apart in the infinity norm.)

Let $J \geq 0$. Suppose q_1, q_2 belong to $\mathbb{P}_m^*$ and $\|q_1 - q_2\|_\infty \leq k^{-J}$. Fix $J' > J$ and types λ_1, λ_2 in $\Lambda_{J'}(m, k)$ such that $q_1 = p_{\lambda_1}$ and $q_2 = p_{\lambda_2}$. Starting at λ_1 and applying property (p.2) repeatedly (that is, for each j going backwards from $j = J'$ to $j = J + 1$), we obtain $\lambda'_1 \in \Lambda_J(m, k)$ such that

$$\|q_1 - p_{\lambda'_1}\|_\infty \leq M \sum_{j=J+1}^{J'} k^{-j} \leq Mk^{-J}$$

$$|f(q_1) - f(p_{\lambda'_1})| \leq M \sum_{j=J+1}^{J'} \epsilon_j \leq M \sum_{j=J+1}^{\infty} \epsilon_j$$

Similarly, we find $\lambda'_2 \in \Lambda_J(m, k)$ such that

$$\|q_2 - p_{\lambda'_2}\|_\infty \leq Mk^{-J}$$

$$|f(q_2) - f(p_{\lambda'_2})| \leq M \sum_{j=J+1}^{\infty} \epsilon_j$$

By the triangle inequality, we have

$$\|p_{\lambda'_1} - p_{\lambda'_2}\|_\infty \leq \|q_1 - q_2\|_\infty + 2Mk^{-J} \leq (2M + 1)k^{-J}$$

and then

$$\|\lambda'_1 - \lambda'_2\|_\infty \leq 2M + 1$$

Applying property (p.1),

$$|f(p_{\lambda'_1}) - f(p_{\lambda'_2})| \leq m(2M + 1)\epsilon_J$$

and then using the triangle inequality again,

$$|f(q_1) - f(q_2)| \leq m(2M + 1)\epsilon_J + 2M \sum_{j=J+1}^{\infty} \epsilon_j \leq B \sum_{j=J}^{\infty} \epsilon_j$$

where $B = m(2M + 1)$.

Proof of Thmeme 1. Let $\mathcal{S} = \{S(\lambda) : \lambda \in \Lambda(m, k)\}$ be a finitary hierarchical source. For every $\lambda \in \Lambda(m, k)$, we have $H(S(k\lambda)) = kH(S(\lambda))$ and hence the normalized entropies $H(S(k\lambda))/\|k\lambda\|$ and $H(S(\lambda))/\|\lambda\|$ coincide. It follows that there exists a unique function $f : \mathbb{P}_m^* \rightarrow [0, \infty)$ such that

$$f(p_\lambda) = H(S(\lambda))/\|\lambda\|, \quad \lambda \in \Lambda(m, k)$$

It is easily seen that $\mathcal{S}$ is entropy-stable by the definition in Section 2.2 if f can be extended to a continuous function on $\mathbb{P}_m$ (which will be the hierarchical entropy function induced by $\mathcal{S}$). This extension will be possible if f is uniformly continuous on $\mathbb{P}_m^*$, and we establish this by showing that $\sum_j \epsilon_j < \infty$, where $\{\epsilon_j\}$ is the sequence Equation (38). Let $\phi \in \Phi(m, k)$ be such that $\mathcal{S}^\phi = \mathcal{S}$. Let $j \geq 1$ and let λ, μ be types in $\Lambda_j(m, k)$ for which $\|\lambda - \mu\|_\infty \leq 1$. Letting

$$\phi(\lambda) = (\lambda_1, \dots, \lambda_k), \quad \phi(\mu) = (\mu_1, \dots, \mu_k)$$

it follows that

$$\|\lambda_i - \mu_i\|_\infty \leq 1, \quad i = 1, 2, \dots, k \quad (39)$$

and by Lemma 1 we have

$$H(S(\lambda)) = \sum_{i=1}^k H(S(\lambda_i)) + \log_2 N_1$$

$$H(S(\mu)) = \sum_{i=1}^k H(S(\mu_i)) + \log_2 N_2$$

where N_1, N_2 are positive integers $\leq k!$. The latter two equations imply

$$|f(p_\lambda) - f(p_\mu)| \leq k^{-1} \sum_{i=1}^k |f(p_{\lambda_i}) - f(p_{\mu_i})| + \log_2(k!)/k^j$$

from which, using Equation (39),

$$|f(p_\lambda) - f(p_\mu)| \leq \epsilon_{j-1} + \log_2(k!)/k^j$$

We conclude that

$$\epsilon_j \leq \epsilon_{j-1} + \log_2(k!)/k^j, \quad j \geq 1$$

from which it follows that $\sum_{j=0}^{\infty} \epsilon_j < \infty$. Applying Lemma A.1, we can now say that f is uniformly continuous on $\mathbb{P}_m^*$.

Appendix B

This Appendix proves some auxiliary results useful for proving Themes 2–3. Henceforth, $\|x\|_2$ shall denote the Euclidean norm of a vector x in a finite-dimensional Euclidean space.

Lemma B.1. Let $\mathcal{T}$ be an IFS of contraction mappings on Ω_m . Let π be the projection mapping $(p, y) \rightarrow p$ from Ω_m onto $\mathbb{P}_m$. Suppose for each $T \in \mathcal{T}$, there is a contraction mapping T^* on $\mathbb{P}_m$ such that $T^*(p) = \pi(T(p, y))$ for every (p, y) in Ω_m , and suppose $\mathbb{P}_m$ is the attractor of the IFS $\{T^* : T \in \mathcal{T}\}$. Suppose $h : \mathbb{P}_m \rightarrow \mathbb{R}$ is a continuous mapping whose graph $G_h = \{(p, h(p)) : p \in \mathbb{P}_m\}$ satisfies the property

$$T(G_h) \subset G_h, \quad T \in \mathcal{T}$$

Then G_h is the attractor of $\mathcal{T}$.

Proof. Let Q be the attractor of $\mathcal{T}$. Since each mapping in $\mathcal{T}$ maps the compact set G_h into itself, $Q \subset G_h$ by uniqueness of the attractor. The proof is completed by showing the reverse inclusion $G_h \subset Q$. Since $\pi(Q)$ is the attractor of the IFS $\{T^* : T \in \mathcal{T}\}$, we must have $\pi(Q) = \mathbb{P}_m$ by assumption. Let $(p, h(p))$ be an arbitrary element of G_h . Since $\pi(Q) = \mathbb{P}_m$, there exists a point in Q of the form (p, y) . But (p, y) and $(p, h(p))$ both belong to G_h , so $y = h(p)$. We conclude $(p, h(p))$ belongs to Q , and therefore $G_h \subset Q$.

Lemma B.2. Let $T^* : \mathbb{P}_m \rightarrow \mathbb{P}_m$ be a contraction mapping with contraction coefficient $\sigma \in (0, 1)$, meaning that

$$\|T^*(p_1) - T^*(p_2)\|_2 \leq \sigma \|p_1 - p_2\|_2, \quad p_1, p_2 \in \mathbb{P}_m$$

Let $c = (c_1, c_2, \dots, c_m)$ be a vector in $\mathbb{R}^m$ and define its variance by

$$V(c) \triangleq \sum_{i=1}^m (c_i - \bar{c})^2 \quad (40)$$

where $\bar{c}$ is the average of the entries of c . Let $T : \Omega_m \rightarrow \Omega_m$ be the mapping

$$T(p, y) \triangleq (T^*(p), \sigma(y + p \cdot c)), \quad p \in \mathbb{P}_m, y \in \mathbb{R}$$

Then T is a contraction mapping if $V(c) < \sigma^{-2}(1 - \sigma^2)^2$.

Proof. By the intermediate value theorem, there is a real number λ in the interval $[\sigma, 1)$ such that

$$V(c) = \lambda^{-2} \sigma^{-2} (\lambda^2 - \sigma^2)^2 \quad (41)$$

Then T is a contraction if we show that

$$\|T(p, u) - T(q, v)\|_2^2 \leq \lambda^2 \|(p, u) - (q, v)\|_2^2 \quad (42)$$

for p, q in $\mathbb{P}_m$ and $u, v \in \mathbb{R}$. The left hand side of Equation (42) is less than or equal to

$$\sigma^2 \|p - q\|_2^2 + [\sigma(p - q) \cdot c + \sigma(u - v)]^2$$

The right hand side of Equation (42) is equal to

$$\lambda^2 \|p - q\|_2^2 + \lambda^2 (u - v)^2$$

Therefore, we will be done if we can show that

$$(\lambda^2 - \sigma^2) \|p - q\|_2^2 + \lambda^2 t^2 - [\sigma(p - q) \cdot c + \sigma t]^2 \geq 0 \quad (43)$$

for all p, q in $\mathbb{P}_m$ and all real numbers t . If $V(c) = 0$, then we are done because the left side of Equation (43) is identically zero (this is because $\lambda = \sigma$ and because $(p - q) \cdot c = 0$ due to the fact that the components of c are constant). We assume $V(c) > 0$ and therefore $\lambda > \sigma$. Letting $Q_{p,q}(t)$ for fixed p, q be the quadratic polynomial

$$Q_{p,q}(t) = \lambda^2 t^2 - [\sigma(p - q) \cdot c + \sigma t]^2$$

the plot of $Q_{p,q}(t)$ is a parabola opening upward because the coefficient of t^2 is the positive number $\lambda^2 - \sigma^2$. Therefore, $Q_{p,q}(t)$ possesses a unique global minimum over t and it is easy to compute

$$\min_t Q_{p,q}(t) = \frac{\lambda^2 \sigma^2 [(p - q) \cdot c]^2}{\sigma^2 - \lambda^2}$$

It follows that Equation (43) will be true for all p, q, u, v if

$$(\lambda^2 - \sigma^2)^2 \|p - q\|_2^2 \geq \lambda^2 \sigma^2 [(p - q) \cdot c]^2$$

holds for all p, q in $\mathbb{P}_m$, which in turn will be true if we can show that

$$x \cdot c \leq \lambda^{-1} \sigma^{-1} (\lambda^2 - \sigma^2)$$

holds for all $x = (x_1, x_2, \dots, x_m) \in \mathbb{R}^m$ for which

$$\sum_{i=1}^m x_i^2 = 1, \quad \sum_{i=1}^m x_i = 0 \quad (44)$$

It is a simple exercise in Lagrange multipliers, which we omit, to show that the vector $x = (x_1, \dots, x_m)$ satisfying the constraints Equation (44) which maximizes the dot product $x \cdot c$ is the vector for which

$$x_i = (c_i - \bar{c}) / \sqrt{V(c)}, \quad i = 1, 2, \dots, m$$

For this choice of x , $x \cdot c$ can be seen to be $\sqrt{V(c)}$. Therefore, we will be done if

$$V(c) \leq \lambda^{-2} \sigma^{-2} (\lambda^2 - \sigma^2)^2$$

But this is true with equality, by Equation (41).

Lemma B.3. Let $k \geq 2$ be arbitrary. Then, for each $i = 0, 1, \dots, k-1$, the mapping $T_i : \Omega_2 \rightarrow \Omega_2$ defined in Section 3 is a contraction.

Proof. Fix i in $\{0, 1, \dots, i-1\}$. The mapping $T_i^* : \mathbb{P}_2 \rightarrow \mathbb{P}_2$ is a contraction mapping with contraction coefficient k^{-1} . Applying Lemma B.2 with $\sigma = k^{-1}$, T_i will be a contraction mapping if we can show that

$$V\left(\log_2 \binom{k}{i+1}, \log_2 \binom{k}{i}\right) < (k - k^{-1})^2, \quad k \geq 2 \quad (45)$$

It is easy to compute that

$$V\left(\log_2 \binom{k}{i+1}, \log_2 \binom{k}{i}\right) = V(\log_2 a_1, \log_2 a_2)$$

where $a_1 = i+1$, $a_2 = k-i$. For any constant γ satisfying $0 < \gamma < 1$, we have

$$V(\log_2 a_1, \log_2 a_2) \leq \sum_{j=1}^2 \left[\log_2 \left(\frac{\gamma^k}{a_j} \right) \right]^2 \quad (46)$$

Using the fact that

$$\gamma \leq \frac{\gamma^k}{a_j} \leq \gamma^k, \quad j = 1, 2$$

the right side of Equation (46) is upper bounded by

$$2 \max [(\log_2 \gamma)^2, (\log_2 \{\gamma^k\})^2]$$

Choosing the smallest value of γ for which

$$(\log_2 \{\gamma^k\})^2 \geq (\log_2 \gamma)^2$$

holds for every $k \geq 2$, we obtain $\gamma = 1/\sqrt{2}$. We have thus proved the variance bound

$$V(\log_2 a_1, \log_2 a_2) \leq 2 \left[\log_2 \left(\frac{k}{\sqrt{2}} \right) \right]^2, \quad k \geq 2$$

Using calculus, it is easy to show that

$$\sqrt{2} \log_2 \left(\frac{k}{\sqrt{2}} \right) < k - 0.5, \quad k \geq 2$$

Thus, Equation (45) holds, and our proof is complete.

Lemma B.4. Let $k \geq 2$ be arbitrary. Then, for each matrix M in the set of matrices $\mathcal{M}(k)$, the mapping $T_M : \Omega_3 \rightarrow \Omega_3$ defined in Section 4 is a contraction.

Proof. The mapping $T_M^* : \mathbb{P}_3 \rightarrow \mathbb{P}_3$ is a contraction with contraction coefficient k^{-1} . Applying Lemma B.2 with $\sigma = k^{-1}$, we have to show that various variances are all less than $(k - k^{-1})^2$. Specifically, for each $(a, b, c) \in V_1$ we wish to show

$$V(H_{3,k}(a, b, c), H_{3,k}(a - 1, b + 1, c), H_{3,k}(a - 1, b, c + 1)) < (k - k^{-1})^2 \quad (47)$$

and for each $(a, b, c) \in V_2$ we wish to show

$$V(H_{3,k}(a, b, c), H_{3,k}(a + 1, b - 1, c), H_{3,k}(a + 1, b, c - 1)) < (k - k^{-1})^2 \quad (48)$$

Using formula Equation (7), the variance on the left side of Equation (47) is equal to

$$V(\log_2 a, \log_2(b + 1), \log_2(c + 1))$$

Let $a_1 = a$, $a_2 = b + 1$, $a_3 = c + 1$. For any constant γ satisfying $0 < \gamma < 1$, we have

$$V(\log_2 a, \log_2(b + 1), \log_2(c + 1)) \leq \sum_{i=1}^3 \left[\log_2 \left(\frac{\gamma(k + 1)}{a_i} \right) \right]^2 \quad (49)$$

Using the fact that

$$\gamma \leq \frac{\gamma(k + 1)}{a_i} \leq \gamma(k + 1), \quad i = 1, 2, 3$$

the right side of Equation (49) is upper bounded by

$$3 \max [(\log_2 \gamma)^2, (\log_2 \{\gamma(k + 1)\})^2]$$

Choosing the smallest value of γ for which

$$(\log_2 \{\gamma(k + 1)\})^2 \geq (\log_2 \gamma)^2$$

holds for every $k \geq 2$, we obtain $\gamma = 1/\sqrt{3}$. We have thus proved the variance bound

$$V(\log_2 a, \log_2(b + 1), \log_2(c + 1)) \leq 3 \left[\log_2 \left(\frac{k + 1}{\sqrt{3}} \right) \right]^2, \quad (a, b, c) \in V_1$$

Similarly, the variance on the left side of Equation (48) is $V(\log_2(a + 1), \log_2 b, \log_2 c)$, and

$$V(\log_2(a + 1), \log_2 b, \log_2 c) \leq 3 \left[\log_2 \left(\frac{k + 1}{\sqrt{3}} \right) \right]^2, \quad (a, b, c) \in V_2$$

Using calculus, it is easy to show that

$$\sqrt{3} \log_2 \left(\frac{k + 1}{\sqrt{3}} \right) < k - 0.5, \quad k \geq 2$$

Thus, for each $(a, b, c) \in V_1$, we have the desired inequality

$$V(\log_2 a, \log_2(b + 1), \log_2(c + 1)) < (k - k^{-1})^2$$

and for each $(a, b, c) \in V_2$, we have the desired inequality

$$V(\log_2(a + 1), \log_2 b, \log_2 c) < (k - k^{-1})^2$$

Acknowledgement

The work of the author was supported in part by National Science Foundation Grant CCF-0830457.

References

1. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2011.
2. Kieffer, J. Hierarchical Type Classes and Their Entropy Functions. In *Proceedings of the 1st International Conference on Data Compression, Communication and Processing*, Palinuro, Campania, Italy, 21–24 June 2011; pp. 246–254.
3. Oh, S.-Y. Information Theory of Random Trees Induced by Stochastic Grammars. Ph.D. Thesis, University of Minnesota Twin Cities, Department of Electrical & Computer Engineering, Minneapolis, MN, USA, 2011.
4. Brualdi, R. Algorithms for constructing $(0,1)$ -matrices with prescribed row and column sum vectors. *Discret. Math.* **2006**, *306*, 3054–3062.
5. Fonseca, C.; Mamede, R. On $(0,1)$ -matrices with prescribed row and column sum vectors. *Discret. Math.* **2009**, *309*, 2519–2527.
6. Falconer, K. *Fractal Geometry*; John Wiley & Sons: Hoboken, NJ, USA, 2003.
7. Soifer, A. *How Does One Cut a Triangle?* Springer-Verlag: Berlin, Heidelberg, Germany, 2010.

© 2011 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>.)