THE ENTROPY OF A CERTAIN INFINITELY CONVOLVED
BERNOULLI MEASURE

J. C. ALEXANDER AND DON ZAGIER

ABSTRACT

An entropy was introduced by A. Garsia to study certain infinitely convolved Bernoulli measures (ICBMs) \( \mu_\beta \) and showed it was strictly less than 1 for \( \beta \) the reciprocal of a Pisot–Vijayarghavan number. However, it is impossible to estimate values from Garsia’s work. The first author and J. A. Yorke have shown this entropy is closely related to the ‘information dimension’ of the attractors of fat baker transformations \( T_\beta \). When the entropy is strictly less than 1, the attractor is a type of strange attractor. In this paper, the entropy of \( \mu_\beta \) is estimated for the case when \( \beta = \phi^{-1} \), where \( \phi \) is the golden ratio. The estimate is fine enough to determine the entropy to several decimal places. The method of proof is totally unlike usual methods for determining dimensions of attractors; rather a relation with the Euclidean algorithm is exploited, and the proof has a number-theoretic flavour. It suggests that some interesting features of the Euclidean algorithm remain to be explored.

1. Introduction

Consider the discrete probability density on the real line with measure \( \frac{1}{2} \) at each of the two points \( \pm 1 \). This is called the Bernoulli measure. We denote it by \( b(x) \). Let \( a_1, a_2, \ldots, a_N > 0 \). We can form the convolution

\[
b(x/a_1) * b(x/a_2) * \ldots * b(x/a_N).
\]

This measure is supported on points

\[
\sum_{n=1}^{N} \pm a_n.
\]

The measure of each such point is \( 2^{-N} \) times the number of representations of the point as a sum (1.1). This measure is a finite probability measure and is called a finitely convolved Bernoulli measure. If \( a_1, a_2, \ldots \) is an infinite sequence of positive numbers decreasing to zero sufficiently rapidly, the infinitely convolved Bernoulli measure (ICBM)

\[
b(x/a_1) * b(x/a_2) * \ldots
\]

can be formed as a weak limit of finitely convolved Bernoulli measures.

These measures were studied extensively in the 1930s, largely because their characteristic functions have interesting asymptotic properties. The ICBMs with \( a_i = \beta^i \) for \( \beta < 1 \) fixed, were singled out. We denote this measure \( \mu_\beta \). For example, if \( \beta = \frac{1}{2} \), the measure \( \mu_\beta \) is uniform on the interval \([-1, 1]\). If \( \beta = \frac{1}{3} \), the measure \( \mu_\beta \) is the classical Cantor measure on \([-\frac{1}{3}, \frac{1}{3}]\). The following were proved. For \( \beta < 1 \),
the measures $\mu_\beta$ are well defined; that is, the limiting processes converge [8]. The $\mu_\beta$ are continuous and are either absolutely continuous on the interval $[-(1-\beta)^{-1}\beta, (1-\beta)^{-1}\beta]$ or totally singular [8]. For $\beta$ the $n$th root of $\frac{1}{2}$, $\mu_\beta$ is absolutely continuous (and indeed, progressively smoother as $n$ is increased) [12]. For $\beta < \frac{1}{2}$, it is easy to see that the support of $\mu_\beta$ is a Cantor set and hence $\mu_\beta$ is totally singular. However, for $\beta \geq \frac{1}{2}$, the support of $\mu_\beta$ is the interval $[-(1-\beta)^{-1}\beta, (1-\beta)^{-1}\beta]$, and it was generally supposed that these $\mu_\beta$ were absolutely continuous. Thus it was somewhat surprising when Paul Erdős proved that for $\beta = \phi^{-1}$, where $\phi = \frac{1}{2}(\sqrt{5}+1)$ is the golden ratio, the measure $\mu_\beta$ is totally singular [4]. A year later he proved a result in the opposite direction, namely there is $\gamma < 1$ such that for almost all $\beta > \gamma$, the measure $\mu_\beta$ is absolutely continuous [5]. Both of these results were proved by considering the asymptotics of the characteristic function. Erdős's method was shown to work whenever $\beta$ is the reciprocal of a Pisot–Vijayaraghavan (PV) number, that is, whenever $\beta$ is an algebraic integer and all of its conjugates lie outside the unit disk in the complex plane [11].

The question of which $\mu_\beta$ are totally singular for $\beta > \frac{1}{2}$ remains open. The only $\beta > \frac{1}{2}$ for which $\mu_\beta$ are known to be totally singular are the reciprocals of PV numbers, and the only explicit $\beta$ for which $\mu_\beta$ is known to be absolutely continuous are roots of $\frac{1}{2}$ and a countable family of similar algebraic numbers [6]. For other work and expositions, see [7, 9, 10, 11].

Garsia introduced a new concept in the study of the $\mu_\beta$ [7]. He considered the finite measure spaces $G^N_\beta$ of the finitely convolved Bernoulli measures

$$b(x/\beta) \ast b(x/\beta^2) \ast b(x/\beta^3) \ast \ldots \ast b(x/\beta^N)$$

for $N = 1, 2, \ldots$, and defined a limiting entropy as follows. Suppose that $x_1, x_2, \ldots, x_r$ are the points of $G^N_\beta$ with measures $p_1, p_2, \ldots, p_r$. The (base 2) entropy of $G^N_\beta$ is given by

$$H(G^N_\beta) = - \sum_{x_i \in G^N_\beta} p_i \log_2 p_i.$$ 

Then $H_\beta$ is defined as the limit

$$H_\beta = \lim_{N \to \infty} \frac{H(G^N_\beta)}{N \log_2 \beta},$$

which he showed exists (note that $H_\beta$ is independent of the choice of base of logarithms). He also proved that if $H_\beta < 1$, then $\mu_\beta$ is totally singular. He also showed that $H_\beta < 1$ if $\beta$ is the reciprocal of a PV number (this did not give a new proof that these $\mu_\beta$ are totally singular, since he used that fact in the argument). His argument is non-constructive and it is impossible to estimate $H_\beta$.

The purpose here is to determine an explicit value for $H_\beta$ for one $\beta$, namely the one Erdős first considered, $\beta = \phi^{-1}$. There are three reasons why such a computation might be of interest.

1. It gives a direct proof that $H_\beta < 1$ for this $\beta$. It may be possible to generalize to other $\beta$ (even perhaps $\beta$ which are not the reciprocals of PV numbers?). The general approach is valid for any $\beta$ which satisfies a polynomial equation with all coefficients 0 or 1 (which is the case for PV numbers and their reciprocals), and it seems likely the particulars will extend to the sequence of $\beta$ which satisfy the equations

$$\beta^n + \beta^{n-1} + \beta^{n-2} + \ldots + \beta^2 + \beta - 1 = 0$$

(which are the reciprocals of PV numbers).
2. James Yorke and the first author introduced dynamical systems \( T_\beta \) on a square which are generalizations of the classical baker's transformation \([1]\). These transformations have natural invariant measures which are products of the \( \mu_\beta \) with a uniform measure. We considered a certain dimension of the attractor of \( T_\beta \) and showed that for \( \beta \) the reciprocal of a PV number, this dimension is \( 1 + H_\beta \). In particular, since the dimension is not an integer, the attractor is some kind of strange attractor. The motivation for the present paper was to give an explicit computation of this dimension. In \([1]\), an estimate for the dimension was given, but was not rigorous. In the present paper, a rigorous estimate is given. The present method is completely different from standard methods of estimating metric dimensions of strange attractors. Indeed, one of the conclusions of \([1]\) is that standard methods cannot work very well for this particular map.

3. Since \( \beta = \phi^{-1} \), the structure of the \( G_\beta^N \) is related to representations of integers in terms of Fibonacci numbers, and through them to the Euclidean algorithm. The computation involves generating functions related to the Euclidean algorithm. In the details, it becomes clear that there are some interesting new features of the Euclidean algorithm to be explored. We do not pursue more than we need in this paper.

For the remainder of the paper, we fix \( \beta = \phi^{-1} \) and let \( \Lambda = 1/\log_2 \phi \). We next briefly state the final results. Consider the \textit{simple Euclidean algorithm}, that is, the Euclidean algorithm with a sequence of subtractions replacing the usual division. Given two positive integers \( k \) and \( i \), the \textit{length} \( e(k, i) \) of the pair \( \{k, i\} \) is the number of steps in the simple Euclidean algorithm applied to \( k \) and \( i \) (formally: \( e(i, i) = 0, e(i+k, i) = e(i+k, k) = e(i, k) + 1 \)).

\textbf{Theorem 1.} Let

\begin{equation}
\kappa_n = \sum_{\substack{k > i > 0 \gcd(k, i) = 1, e(k, i) = n}} k \log_2 k,
\end{equation}

and

\begin{equation}
\mathcal{G}(x) = 1 - \frac{1}{2} \left( \frac{1 - 3x}{1 - x} \right)^2 \sum_{n=1}^{\infty} \kappa_n x^n.
\end{equation}

The series (1.3) converges for \( |x| < \frac{1}{2} \) and

\[ H_\beta = \Lambda \mathcal{G}(\frac{1}{4}) = \Lambda \left( 1 - \frac{1}{18} \sum_{\substack{k > i > 0 \gcd(k, i) = 1}} \frac{k \log_2 k}{4e(k, i)} \right). \]  

By determining the growth rate of the \( \kappa_n \), and making rearrangements in \( \mathcal{G}(x) \), we can estimate truncation errors. In particular, by truncating one rearranged series at the 20th term, we obtain the estimates

\[ 0.995570 < H_{\phi^{-1}} < 0.995736. \]  

Another rearrangement gives sharper estimates; however, we have not been able to justify them rigorously. This series leads to the point estimate 0.995713126686.

We also show that (1.4) is equivalent to the following, somewhat different-looking, formula for \( H_\beta \).
THEOREM 1'. There is a unique bounded function $g(t)$ on $[0, 1]$ which satisfies the functional equation

$$g(t) = (1 + t) \log_2(1 + t) + (2 - t) \log_2(2 - t) + \frac{1 + t}{4} g\left(\frac{1}{1 + t}\right) + \frac{2 - t}{4} g\left(\frac{1}{2 - t}\right).$$

(1.6)

This function is analytic on $(0, 1)$ and, moreover,

$$H_{\beta} = \Lambda(\beta - \frac{1}{3} g(\frac{1}{3})) = \Lambda(1 - \frac{1}{3} g(0)).$$

(1.7)

It is possible to compute with (1.6), (1.7); the computations are identical to those coming from (1.3).

2. The combinatorics

The spaces $G^N_\beta$ consist of sums

$$\sum_{n=1}^{N} \pm \beta^n.$$ 

(2.1)

Since $\beta^2 + \beta - 1 = 0$, there are multiple sums representing one point and of course this increases the measure at that point. For example $\beta^2 + \beta - 1$ and $-\beta^2 - \beta + 1$ both

![Fig. 1. The Fibonacci graph](image-url)
represent 0. The whole structure of the $G^N_\beta$ can be represented graphically as in Figure 1. The space $G^N_\beta$ is represented by the nodes of the graph at level $N$, constructed from the top down. At level $N$, for each node, we consider the sequence of $\pm$ signs in (2.1) corresponding to that point. From each node, beginning with the node corresponding to the empty sequence at level 0, there are two descending edges. The left (respectively right) edge descends to a node corresponding to appending a $-$ (respectively $+$) sign to the sequence. Because of the relation $\beta^2 + \beta - 1 = 0$, the graph is not a tree. Above each node through level 3, we list the sequence(s) corresponding to each node. Below the nodes through level 5, we list the number of sequences corresponding to that node. We call this number the frequency. It is equal to the number of descending paths from the top node to the node in question. Moreover, at level $N$, $2^{-N}$ times the frequency is the measure of the point in $G^N_\beta$. Thus the entropies $H(G^N_\beta)$ can be computed from the frequencies. Let $F_n(k)$ denote the number of nodes at level $N$ with frequency $k$. Then

$$H(G^N_\beta) = - \sum_k F_n(k) 2^{-N} \log_2 2^{-N}k = N - \sum_k F_n(k) 2^{-N} \log_2 k. \quad (2.2)$$

In fact, there is another simple interpretation of this graph and the entropies we consider. Let $f_i$ be the $i$th element of the Fibonacci sequence 1, 2, 3, 5, 8,….. A Fibonacci expansion of length $N$ of a non-negative integer $n$ is a representation $n = \sum_{i=1}^{N} a_i f_i$, $a_i = 0, 1$. The graph of Figure 1 encodes Fibonacci representations of the non-negative integers. If the nodes at level $N$ are labelled from 0 to $f_{N+2} - 2$ from left to right, the frequency associated to a node is the number of Fibonacci expansions of length $N$ of the label. For this reason, we call the graph of Figure 1 the Fibonacci graph. These frequencies have been studied; see for example [2]. The entropy $H(G^N_\beta)$ measures the average redundancy of the Fibonacci representations of length $N$. There is obvious redundancy since $2^N$ words on symbols zero and one represent $f_{N+2} - 1$ numbers. If, asymptotically, the redundancy were more or less evenly distributed, the entropies $H(G^N_\beta)$ would be asymptotic to $N \log_2 \beta + o(N)$. The fact that $H_\beta$ is strictly less than one implies that the redundancies in the representation are not evenly distributed in some strong sense.

There is considerable and repeated structure in the Fibonacci graph. In particular, suppose we erase from that graph the node at level 0 and all the nodes with frequency 1 and the edges descending from those nodes. The remainder is the disjoint union of an infinite number of isomorphic graphs. Each of these has a ‘top’ node with frequency 2 in the lowest level, two nodes with frequency 2 on the next level, two with frequency 2 and two with frequency 3 two levels beyond, etc. There is one such top at level 3, and two such tops at each succeeding level. Moreover, these subgraphs themselves break apart into more basic subgraphs. In this section, we expose this decomposition and develop the structure of the most basic subgraphs. To this end we proceed through a succession of technical definitions and observations, some labelled for future reference. The reader may find it convenient to illustrate the statements with Figure 1, concentrating on the central node at level 3.

Any edge connects two nodes, one at level $N$ and one at level $N + 1$. We call the one at level $N$ the parent of the one at level $N + 1$ and the one at level $N + 1$ the child of the one at level $N$. We extend the terminology to speak of grandchildren, etc. The descendents of a node are its children, their children, etc. The ancestors of a node are its parents, grandparents, etc.

The graph divides the plane into bounded regions we call diamonds. The diamonds
are consequences of the relation \(-1 + \beta + \beta^2 = 1 - \beta - \beta^2\) (or \(-+++=--\) in terms of plus and minus signs). Each diamond is bounded by six edges and six nodes. The nodes are at levels \(N, N+1, N+2,\) and \(N+3\) for some \(N\). There is one top node at level \(N\). There are two nodes at level \(N+1\), two at level \(N+2\), and a bottom node at level \(N+3\). The bottom node we call a relational node, since its position at the bottom is a consequence of the relation above. The node at the top is called the relational parent of the node at the bottom and the node at the bottom is called the relational child of the node at the top. A relational node has two parents; others have one.

In addition to a frequency \(d\), a relational node \(X\) has a frequency pair. A relational node has two parents with frequencies \(a\) and \(b\). The frequency pair of \(X\) is the unordered pair \(\{a, b\}\). Clearly the frequency of \(X\) is \(a + b\). Finally, for purposes of setting up inductions, we assign a depth to each node. This is done inductively. The node at level 0 and all nodes with frequency 1 are assigned depth 0. If a node is not relational, it is assigned the depth of its parent. If it is relational, it is assigned depth 1 plus the depth of its relational parent.

There are four edges emanating from a relational node; hence a relational node is in the boundary of four diamonds—one above, one below, one left, and one right. (a) If the top node of a diamond is at level \(N\), neither of the boundary nodes at level \(N+2\) is relational. (b) Moreover, at most one of the nodes at level \(N+1\) is relational, since there is not enough room for two diamonds above two relational nodes in those positions. On the other hand, a non-relational node which is not at the extreme left or right of a level is in the boundary of three diamonds. It is the top of one diamond, at level 1 below the top of another diamond, and at level 2 below the top of a third. (c) Accordingly one of the children of a non-relational node is relational and one is non-relational; the non-relational child has the same frequency and same depth as its parent.

We next consider the structure of the graph below a relational node \(X\). Say \(X\) has frequency pair \(\{a, b\}\), level \(N\), and depth \(r\). There are two children at level \(N+1\), neither of which is relational by statement (a) above. Accordingly, these two nodes have frequency \(a + b\) and depth \(r\). There are four grandchildren at level \(N+2\). Two of these are relational, being the relational children of the parents of \(X\), and two are not. The two which are not have frequency \(a + b\) and depth \(r\). The two which are relational have frequency pairs \(\{a, a+b\}\) and \(\{b, a+b\}\) and depth \(r\), since their relational parents are the parents of \(X\). Finally the relational child of \(X\) has frequency pair \(\{a+b, a+b\}\) and depth \(r+1\). In the Fibonacci graph, all nodes through level 6 have depth 0 or 1 except the central node at level 6 which has depth 2. At level 7 there are four nodes of depth 2.

The preceding paragraph describes the local structure near a relational node. We turn to the global structure. (d) First note that from every relational node, two edge paths descend; every node on these two paths has the same depth and the same frequency as the parent relational node. This follows from (a) and (c) above. Consider a relational node \(X\) of depth \(r\) both of whose parents have depth less than \(r\). Call such a node a primary relational node. In the Fibonacci graph, there is one primary relational node of depth 1 at level 3, and two at each succeeding level. There is a primary relational node of depth 2 at level 6. There are two at level 7 and four at level 8. Consider the set \(D(X)\) of all descendants of \(X\) which are relational and have depth \(r\). Since each relational node has two relational grandchildren, if we insert edges in
$D(X)$ between each node and its two relational grandchildren, we make $D(X)$ into a binary tree. Note that if $X$ and $Y$ are both relational nodes, $D(X) \cap D(Y) = \emptyset$. This follows from statements (b) and (c) above. We next note that the frequency pair of $X$ has the form $(d, d)$, where $d = 1$ if the depth of $X$ is 1, or $d$ is the frequency of the relational parent of $X$ if the depth of $X$ is greater than 1. The frequency pair of $X$ determines the frequency pairs of all members of $D(X)$, as we study next.

The simple Euclidean algorithm is the Euclidean algorithm without division. We can describe one step of the algorithm as follows. For any pair of non-negative integers $(k, i)$ (possibly equal), let $T(k, i) = \{(k - i, \min(k, i))\}$. The length of $(k, i)$ is the greatest $n$ such that $0 \notin T^n(k, i)$. For example, if we use $\leftarrow$ to denote the effect of $T$, 

\[
\{11, 3\} \leftarrow \{8, 3\} \leftarrow \{5, 3\} \leftarrow \{3, 2\} \leftarrow \{2, 1\} \leftarrow \{1, 1\} \leftarrow \{1, 0\}
\]

(length 5). The length is denoted by $e(k, i)$. This definition is clearly equivalent to the one in the introduction.

Conversely, we can make a binary tree labelled with pairs of integers as follows. Start with one node at level 0 labelled with the pair $(1, 1)$ and one node at level 1 labelled with the pair $(2, 1)$. Inductively, given a node at level $n$ labelled with the pair $(a, b)$, there are two descending edges (left and right) top nodes at level $n + 1$ labelled with pairs $(a + b, a)$ and $(a + b, b)$. At level $n$, there are $2^{n-1}$ nodes. For example, at level 3, the pairs are $(5, 3), (5, 2), (4, 3), (4, 1)$. Start with any pair $(a, b)$ at level $n$. The labels on the nodes of the unique path up the tree to the node at level 1 are precisely those of the simple Euclidean algorithm of $(a, b)$, up to the last step. The length of the pair is $n$. Conversely, given the expansion of the simple Euclidean algorithm for any pair of coprime integers $(a, b)$, it is routine to locate the pair in the tree. Thus each pair of coprime integers $(a, b)$ appears exactly once in this tree at level $e(a, b)$. We call this tree the Euclidean tree. This tree has appeared in other contexts, see for example [3, 13]. We trivially form the $d$-Euclidean tree by multiplying all the labels involved in the Euclidean tree by $d$.

We next observe that for any primary relational node $X$ in the Fibonacci graph with frequency pair $(d, d)$, the labelled tree structure of $D(X)$ is a $d$-Euclidean tree. This is immediate from the way frequency pairs of relational descendants are determined. Thus the Euclidean tree is the irreducible kernel of the Fibonacci graph and, in the end, all computations will be done in terms of the Euclidean graph. Note that if a node at level $L \geq 1$ is labelled with $(a, b)$ with $a > b$ in the Euclidean tree, then $a$ is the sum of the two labels of the parent of the node. Thus it is essentially equivalent to consider the maxima of the two integers of a label or the sum of the labels. For minor technical reasons, we work below with the former.

We can reverse the above process and, starting from the Euclidean tree, build the nodes of the Fibonacci graph. We construct a set of nodes; each node is assigned a level, a frequency, and a depth. We can speak of appending a $d$-Euclidean tree of depth $r$, starting at level $N$ to an existing set. By this we mean making a disjoint union of the existing set and the nodes of levels $L \geq 1$ of a copy of a $d$-Euclidean tree, so that if a node in the $d$-Euclidean tree has length $L$, it is assigned level $N + 2L$ and depth $r$ in the set. If the node in the Euclidean tree is labelled with integers $(a, b)$, with $a > b$, it is given frequency $a$ in the set.

The discussion above proves the following lemma.
**Lemma.** The following process constructs a set of points (nodes) labelled with levels, frequencies and depths, which is isomorphic to the set of nodes of the Fibonacci graph in levels greater than 1.

1. Begin with two nodes of frequency 1 and depth 0 at each level \( N \geq 1 \).
2. Inductively append Euclidean trees to obtain relational nodes.
   a. Append one 1-Euclidean tree starting at level 3 and two 1-Euclidean trees starting at each level \( N > 3 \).
   b. Proceed by induction on depth. Assume that nodes up through depth \( r \) have been assigned. For each node of depth \( r \) which has, say, level \( N \) and frequency \( d \), append a 2d-Euclidean tree of depth \( r+1 \) starting at level \( N+3 \).
3. Append non-relational nodes. For each node of depth \( r \), frequency \( d \), and level \( N \), append two nodes at each level greater than \( N \) of depth \( r \) and frequency \( d \).

As a consequence of this result, many of the results about Fibonacci representations [2] can be deduced from known results about the Euclidean algorithm. Particulars are left to the interested reader.

3. Generating functions

In this section we translate the construction of the previous lemma into the language of generating functions. The object of the present section is to prove the following result.

**Proposition.** Let \( \mathcal{H}(x) = \sum_{N=0}^{\infty} H(G^N_f) x^N \) be the generating function for the entropies of the finite spaces \( G^N_f \). Then

\[
\mathcal{H}(x) = \frac{x}{(1-x)^2} \mathcal{T} \left( \frac{x^2}{4} \right)
\]

with \( \mathcal{T} \) as in (1.3).

**Proof.** Let \( F_N(k) \) be equal to the number of nodes at level \( N \) with frequency \( k \). Note that

\[
\sum_{k=1}^{\infty} F_N(k) = f_{N+2} - 1, \quad \sum_{k=1}^{\infty} k F_N(k) = 2^N, \quad \sum_{k=1}^{\infty} \frac{k}{2^N} F_N(k) \log_2 \frac{2^N}{k} = H(G^N_f).
\]

Let \( f_k(x) = \sum_{N=1}^{\infty} F_N(k) x^N \) (this and all further generating functions are to be considered formal power series in \( x \), although all in fact converge for \( |x| \) small enough). Let \( \delta_k(N) \) be the number of times the integer \( k \) appears as the larger of the two integers in the pairs associated to nodes of length \( N \) in the Euclidean tree, and let \( \delta_k(x) = \sum_{N=1}^{\infty} \delta_k(N) x^N \). From the description of the Euclidean tree in the previous section, it is apparent that

\[
\delta_k(x) = \sum_{1 \leq \ell \leq k \atop \gcd(k,\ell) = 1} x^{e(k, \ell)}.
\]

When the nodes of Euclidean trees are embedded in the nodes of the Fibonacci graph,
the levels of the nodes come from the lengths, up to a shift and a stretch; this leads us to define
\[ \alpha_k(x) = \sum_{i=1}^{\infty} x^{1+2e(k,i)}. \]

Starting from \( l_1(x) = 1 \), inductively define
\[ l_k(x) = \sum_{d|k, d \neq 1} \alpha_d(x) l_{k/d}(x) \]
(a short list of \( \alpha_k(x) \) and \( l_k(x) \) is given in Table 1). Next, define functions of two variables \( x \) and \( s \):
\[ \mathcal{L}(x;s) = 1 + \sum_{k=1}^{\infty} k^4 l_k(x), \]
\[ \mathcal{A}(x;s) = 1 - \sum_{k=2}^{\infty} k^4 \alpha_k(x), \]
\[ \Phi(x;s) = 1 + \sum_{k=2}^{\infty} k^4 f_k(x). \]

The formulae above imply that
\[ \mathcal{L}(x;s) = \mathcal{A}(x;s)^{-1}. \]

Note also that
\[ \Phi(x;0) = \sum_{N=0}^{\infty} (f_{N+2} - 1) x^N = \frac{1}{(1-x)(1-x-x^2)}, \]
\[ \Phi(x;1) = \sum_{N=0}^{\infty} 2^N x^N = \frac{1}{1-2x}, \]
\[ \frac{\partial \Phi(x;s)}{\partial s} \bigg|_{s=1} = \sum_{k \geq 1} k F_N(k) \ln k x^N, \]
so that
\[ \mathcal{H}(x) = \sum_{N=0}^{\infty} H(G_{f}^N) x^N = \frac{x}{(1-x)^2} \left( 1 - \ln 2 \frac{\partial \Phi(x/2;s)}{\partial s} \right) \bigg|_{s=1}. \]

By induction, it is seen that the sum of the maxima of the pairs associated to the nodes of length \( N \) in the Euclidean tree is \( 2 \cdot 3^{N-1} \). Thus
\[ \mathcal{A}(x;1) = 1 - \sum_{k>1 \text{ even, } \gcd(k,0) = 1} k 3^{N-1} x^N = \frac{(1+x)^3(1-2x)}{1-3x^2}. \]
The lemma of the previous section leads to the expression
\[ f_k(x) = \begin{cases} 
2x/(1-x)^2 & \text{if } k = 1, \\
(1+x)^2/1-x & \text{if } k > 1.
\end{cases} \]

To see this, note that \( l_k(x) \) is the generating function for the set constructed in the induction (ii)(b) of the lemma, starting from a single 1-Euclidean tree starting at level 0. By (ii)(a), the generating function for the set constructed in (ii) is
\[ (1 + 2x + 2x^2 + 2x^3 + \ldots) l_k(x) = 1 + \frac{x}{1-x} l_k(x). \]

By part (iii) of that lemma, to obtain the generating function for \( k > 1 \), this expression is multiplied by another \((1 + x)/(1-x)\). The count for \( k = 1 \) is handled separately, giving the above expression. Hence
\[ \Phi(x; s) = \frac{1+x^2}{(1-x)^2} + \left( \frac{1+x^2}{1-x} \right) \sum_{k=2}^{\infty} k^k l_k(x) = \left( \frac{1+x^2}{1-x} \right)^2 \mathcal{L}(x; s) - \frac{2x}{(1-x)^2}. \]

Thus
\[ \frac{\partial \Phi(x; s)}{\partial s} \bigg|_{s=1} = \left( \frac{1+x^2}{1-x} \right)^2 \frac{\partial \mathcal{L}(x; s)}{\partial s} \bigg|_{s=1} = -\left( \frac{1+x}{1-x} \right)^2 \frac{\partial \mathcal{A}(x; s)}{\partial s} \bigg|_{s=1} \]
\[ = -\frac{(1-3x^2)^2}{(1-x)^2(1-x^2)^2} \cdot \frac{\partial \mathcal{A}(x; s)}{\partial s} \bigg|_{s=1}. \]

On the other hand
\[ \frac{\partial \mathcal{A}(x; s)}{\partial s} \bigg|_{s=1} = -\sum_{k>0 \atop \gcd(k, i)=1} x^{2\gcd(k, i)/i^2} \ln k = \sum_{N=1}^{\infty} \left( \sum_{k>0 \atop \gcd(k, i)=1, \gcd(k, i)=N} k \ln k \right) x^{2N+1}. \]

Hence
\[ \mathcal{H}(x) = \frac{x}{(1-x)^2} \left( \frac{4 - 3x^2)^2}{(4 - x^2)^2(1-x)^2} \right) \sum_{N=1}^{\infty} \left( \sum_{k>0 \atop \gcd(k, i)=1, \gcd(k, i)=N} k \log_2 k \right) \left( \frac{x}{2} \right)^{2N+1} \]
\[ = \frac{x}{(1-x)^2} \mathcal{F} \left( \frac{x^2}{4} \right). \]

This completes the proof of the proposition.

The series (1.3), as written, converges too slowly at \( t = \frac{1}{4} \) for useful estimates. It is not hard to see that \( 2 \cdot 3N+1 \log_2(N+1) < \kappa_n < 2 \cdot 3N+1/N \log_2 \phi \); that is, the \( \kappa_n \) grow faster than geometrically. It is possible to rearrange the series to obtain coefficients which grow geometrically. Accordingly we define coefficients \( \mu_N \) and \( \lambda_N \) by the formulae
\[ \sum_{N=1}^{\infty} \mu_N x^N = \frac{1}{2} (1-3x) \sum_{N=1}^{\infty} \kappa_N x^N, \]  
\[ \sum_{N=1}^{\infty} \lambda_N x^N = \frac{1}{2} \left( \frac{1-3x}{1-x} \right)^2 \sum_{N=1}^{\infty} \kappa_N x^N = 1 - \mathcal{F}(x). \]
We can put useful rigorous bounds on the $\mu_N$; the $\lambda_N$ are used to make sharper, but non-rigorous, estimates.

**Lemma.**

$$0.585 \ldots = \log_2 1.5 < \frac{\mu_N}{3^{N-1}} < \frac{2}{3}$$  \hspace{1cm} (3.4)

(*in particular the $\mu_N$ are positive and grow geometrically*).

**Proof.** Consider a node of length $N$ in the Euclidean tree associated to the pair $\{a, b\}$, $a > b$. It has a 'sibling' pair $\{a, a-b\}$ (both descending from the pair $\{a-b, a\}$). These two spawn pairs labelled

$$\{a+b,a\}, \ \{a+b,b\}, \ \{2a-b,a\}, \ \{2a-b,a-b\}$$

of length $N+1$. Accordingly $\mu_{N+1} = \frac{1}{2}(\kappa_{N+1} - 3\kappa_N)$ can be written

$$\sum_{a > b > 0 \atop (a, b) \in \{a, b\}, e(a, b) = N} \frac{1}{2}((a+b) \log_2(a+b) + (2a-b) \log_2(2a-b) - 3a \log_2 a)$$

$$= \sum_{a > b > 0 \atop (a, b) \in \{a, b\}, e(a, b) = N} \frac{1}{2}a \left[ \frac{a+b}{a} \log_2 \left( \frac{a+b}{a} \right) + \frac{2a-b}{a} \log_2 \left( \frac{2a-b}{a} \right) \right].$$

By the convexity of the function $x \mapsto x \log_2 x$ for $x > 0$, this expression is greater than

$$\sum_{a > b > 0 \atop (a, b) \in \{a, b\}, e(a, b) = N} \frac{3a}{2} \log_2 1.5 = 3^N \log_2 1.5.$$

This proves the first inequality of (3.4). On the other hand, the function

$$x \mapsto x \log_2 x + (3-x) \log_2(3-x)$$

is convex on the interval $[1,2]$ and thus takes its maximum at one (actually both) of its end points. Letting $x = (a+b)/a$, we find that $\mu_{N+1}$ is bounded by

$$\sum_{a > b > 0 \atop (a, b) \in \{a, b\}, e(a, b) = N} a \log_2 2 = 2 \cdot 3^{N-1}.$$

This proves the second inequality of (3.4) and completes the proof of the lemma.

The behaviour of $\mathcal{H}(x)$ as a meromorphic function depends on the rate of growth of the coefficients. From (3.4) we see that $\sum_{N=1}^{\infty} H_N x^N$ is dominated by $2x/3(1-3x)$, so that (3.2) converges for $|x| < \frac{1}{3}$. Consequently, $\mathcal{H}(x)$ is holomorphic for $|x| < \sqrt{3}$/2 except for a double pole at $x = 1$. This is consistent with the known rate of growth of $\mathcal{H}(G^N_\beta)$. Thus

$$\mathcal{H}(x) = H_{-2} \frac{x}{(1-x)^2} + H_{-1} \frac{1}{1-x} + O(1), \quad |x| < \sqrt{3}/2$$  \hspace{1cm} (3.5)

for some numbers $H_{-2}$ and $H_{-1}$. Equivalently, as $N \to \infty$,

$$H(G^N_\beta) = NH_{-2} + H_{-1} + O((\beta)^{N/2}).$$  \hspace{1cm} (3.6)
Note, then, that
\[ H_\beta = \Lambda H_{-2}. \] (3.7)

4. Numerical computations

In this section we prove Theorem 1 and make numerical computations. We present some tables and derive the estimates (1.5). Recall that \( \Lambda = 1/\log_2 \phi \).

To obtain (1.4), we multiply equation (3.1) by \((1 - x)^2\) to obtain
\[ (1 - x)^2 \mathcal{H}(x) = x \mathcal{F}(\frac{1}{2} x^2) = H_{-2} + (1 - x) H_{-1} + (1 - x)^2 O(x^{-n}). \]
This is convergent at \( x = 1 \), so setting \( x = 1 \), we obtain from (3.7) that
\[ H_\beta = \Lambda \mathcal{F}(\frac{1}{2}). \] (4.1)

Note that since \( \mathcal{F} \) converges for \( x < \sqrt{\frac{1}{2}} \), this series converges. We can evaluate it in several ways, depending on how we expand \( \mathcal{F} \). Thus we obtain
\[ H_\beta = \Lambda \left(1 - \frac{1}{18} \sum_{n=1}^{\infty} \frac{\kappa_n}{4^n}\right), \] (4.2)
\[ \quad = \Lambda \left(1 - \frac{4}{9} \sum_{n=1}^{\infty} \frac{\mu_n}{4^n}\right), \] (4.3)
\[ \quad = \Lambda \left(1 - \sum_{n=1}^{\infty} \frac{\lambda_n}{4^n}\right). \] (4.4)

The series (4.2) converges too slowly for effective computation. We use series (4.3). The partial sums of the series are exhibited in Table 2. Since each \( \mu_n > 0 \), the values

<table>
<thead>
<tr>
<th>( N )</th>
<th>( \kappa_N )</th>
<th>( \frac{4^{-N} \mu_N}{4} )</th>
<th>( \Lambda \left[1 - \frac{1}{4} \sum_{n=1}^{N} 4^{-n} \mu_n\right] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.00000000000</td>
<td>0.111111111</td>
<td>1.2803734137</td>
</tr>
<tr>
<td>2</td>
<td>9.50977500433</td>
<td>0.0487468751</td>
<td>1.210574355</td>
</tr>
<tr>
<td>3</td>
<td>39.2192809489</td>
<td>0.0371179026</td>
<td>1.1566920630</td>
</tr>
<tr>
<td>4</td>
<td>149.825218767</td>
<td>0.0279230694</td>
<td>1.1164711129</td>
</tr>
<tr>
<td>5</td>
<td>546.040935867</td>
<td>0.0209560069</td>
<td>1.0862856596</td>
</tr>
<tr>
<td>6</td>
<td>1927.86131506</td>
<td>0.0157193201</td>
<td>1.063642351</td>
</tr>
<tr>
<td>7</td>
<td>6652.82922633</td>
<td>0.0117898937</td>
<td>1.046608354</td>
</tr>
<tr>
<td>8</td>
<td>22566.2448165</td>
<td>0.0084249292</td>
<td>1.0339239316</td>
</tr>
<tr>
<td>9</td>
<td>75522.0214340</td>
<td>0.0063188250</td>
<td>1.0243712348</td>
</tr>
<tr>
<td>10</td>
<td>250035.936861</td>
<td>0.0049791944</td>
<td>1.0172067086</td>
</tr>
<tr>
<td>11</td>
<td>820517.437055</td>
<td>0.0037304362</td>
<td>1.0118333133</td>
</tr>
<tr>
<td>12</td>
<td>2672781.19736</td>
<td>0.0029798727</td>
<td>1.0078032667</td>
</tr>
<tr>
<td>13</td>
<td>8652030.25592</td>
<td>0.0020983705</td>
<td>1.0047807317</td>
</tr>
<tr>
<td>14</td>
<td>27857150.7635</td>
<td>0.0015737779</td>
<td>1.0025138304</td>
</tr>
<tr>
<td>15</td>
<td>89274632.8111</td>
<td>0.0011803313</td>
<td>1.0008136545</td>
</tr>
<tr>
<td>16</td>
<td>284933436.818</td>
<td>0.0008852500</td>
<td>0.9995385225</td>
</tr>
<tr>
<td>17</td>
<td>906128930.379</td>
<td>0.0006639375</td>
<td>0.9985821736</td>
</tr>
<tr>
<td>18</td>
<td>2872372650.92</td>
<td>0.0004979532</td>
<td>0.9978649119</td>
</tr>
<tr>
<td>19</td>
<td>907907532.09</td>
<td>0.0003734649</td>
<td>0.9973269656</td>
</tr>
<tr>
<td>20</td>
<td>28623099334.3</td>
<td>0.0002800986</td>
<td>0.9969235058</td>
</tr>
</tbody>
</table>
in the third column are upper bounds for $H_\beta$; hence $H_\beta$ is clearly seen to be less than 1. However, we can bound truncation errors in $H_\beta$ and thus compute. The following result is immediate from (3.4).

**LEMMA.** If (4.3) is truncated at term $N$, the error $E_N$ is bounded by

$$\left(\frac{3}{4}\right)^{N-1} \Lambda \log 2 \cdot 1.5 < E_N < \frac{2}{3}\left(\frac{3}{4}\right)^{N-1} \Lambda. \quad (4.5)$$

These bounds combined with the computations of Table 3 lead to the bounds (1.5).

It is very interesting to consider the computations from (4.4). This series seems to converge much more rapidly than (3.2); the results are exhibited in Table 3. There are

**Table 3. Estimates using series (4.4)**

<table>
<thead>
<tr>
<th>$N$</th>
<th>$\lambda_n$</th>
<th>$\Lambda[1 - \sum_{n=1}^{N} 4^{-n}\lambda_n]$</th>
<th>$N$</th>
<th>$\lambda_n$</th>
<th>$\Lambda[1 - \sum_{n=1}^{N} 4^{-n}\lambda_n]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00000000000000</td>
<td>1.08031506780942</td>
<td>11</td>
<td>0.1671310135</td>
<td>0.99571314346575</td>
</tr>
<tr>
<td>2</td>
<td>0.75488750216347</td>
<td>1.01235537255207</td>
<td>12</td>
<td>0.1508973530</td>
<td>0.99571313051034</td>
</tr>
<tr>
<td>3</td>
<td>0.59009046578294</td>
<td>0.99907446376965</td>
<td>13</td>
<td>0.137306234</td>
<td>0.99571312756321</td>
</tr>
<tr>
<td>4</td>
<td>0.47404748549602</td>
<td>0.99640716876216</td>
<td>14</td>
<td>0.125804500</td>
<td>0.99571312688815</td>
</tr>
<tr>
<td>5</td>
<td>0.3895804093515</td>
<td>0.99585916148827</td>
<td>15</td>
<td>0.11597459</td>
<td>0.99571312673272</td>
</tr>
<tr>
<td>6</td>
<td>0.3264477118497</td>
<td>0.99574436123372</td>
<td>16</td>
<td>0.10749731</td>
<td>0.99571312669652</td>
</tr>
<tr>
<td>7</td>
<td>0.27819431272</td>
<td>0.99571990342182</td>
<td>17</td>
<td>0.1001257</td>
<td>0.99571312668812</td>
</tr>
<tr>
<td>8</td>
<td>0.240588130815</td>
<td>0.99571461551934</td>
<td>18</td>
<td>0.0936666</td>
<td>0.99571312668616</td>
</tr>
<tr>
<td>9</td>
<td>0.210767979308</td>
<td>0.9957145739951</td>
<td>19</td>
<td>0.087967</td>
<td>0.99571312668750</td>
</tr>
<tr>
<td>10</td>
<td>0.18675016890</td>
<td>0.99571320086237</td>
<td>20</td>
<td>0.082907</td>
<td>0.99571312668559</td>
</tr>
</tbody>
</table>

a couple of surprises in this table. The $\lambda_n$ are obtained by adding and subtracting large numbers. There is no reason to expect them to be positive and decreasing to zero, although the table indicates that they are. We can prove only that $\lambda_n$ is $O((3/4)^{n/2})$. There is evidently something more subtle going on in the Euclidean algorithm. If the $\lambda_n$ are bounded, the data in Table 3 converge at a rate no worse than $O(4^{-N})$. From the tabulations, the data are converging faster than $O(4^{-N})$. Equivalently, $\sum \lambda_n x^n$ converges for $|x| < C$ for some $C \sim 1.1$. This is surprising since, from the form of (3.3), $\sum \lambda_n x^n$ could have a pole at $x = 1$. We leave the investigation of this series to the interested reader. Other calculations (higher precision) indicate that the exhibited digits of Table 3 are uncontaminated by machine roundoff error. If the indicated convergence is valid, the value of $H_\beta$ is most likely between 0.995713126685 and 0.995713126686.

5. A functional relation

In this section we prove Theorem 1'. Define a sequence of functions $g_n(t)$ ($n = 1, 2, \ldots$) by

$$g_1(t) = (1 + t) \log_2 (1 + t) + (2 - t) \log_2 (2 - t),$$

$$g_{n+1}(t) = 3^n g_1(t) + (1 + t) g_n \left( \frac{1}{1 + t} \right) + (2 - t) g_n \left( \frac{1}{2 - t} \right) \quad (n \geq 1).$$

An easy induction shows that

$$g_n(t) = \sum_{k, l \geq 0, \gcd(k, l) = 1, t(k, l) = n} [kt + i(1 - t)] \log_2 [kt + i(1 - t)].$$
Hence each $g_n$ is analytic on the set $T = \mathbb{C} \setminus ([1, \infty) \cup (-\infty, 0])$, and the identity $\sum_{k,t} (k+i) = 2 \cdot 3^n$ implies that the series

$$g(t, x) = \sum_{n=1}^{\infty} g_n(t) x^n$$

converges absolutely and locally uniformly in $T \times \{x: |x| < \tfrac{1}{3}\}$ and therefore defines a jointly analytic function of $t$ and $x$ there. The recurrence defining the $g_n$ translates into the functional equation

$$g(t, x) = \frac{x}{1 - 3x} g_1(t) + x(1+t) g\left(\frac{1}{1+t}, x\right) + x(2-t) g\left(\frac{1}{2-t}, x\right).$$

Conversely, it is easily seen by the estimate just used that this functional equation has a unique solution whenever $|x| < \tfrac{1}{3}$ (the crucial point is that then the sum of the factors $x(1+t)$ and $x(2-t)$ is uniformly bounded by a constant less than 1). For $x = \tfrac{1}{3}$, the functional equation for $g(t, x)$ reduces to (1.6), so the uniqueness implies that $g(t) = g(t, \tfrac{1}{3})$. On the other hand, the closed formula for $g_n(t)$ implies that $g_n(\tfrac{1}{3}) = \tfrac{1}{3} \kappa_{n+1} - 3^n$ for every $n \geq 1$ with $\kappa_n$ as in (1.2). Thus (1.7) is indeed equivalent to (1.4), as asserted.

References


Department of Mathematics
University of Maryland
College Park
Maryland 20742
USA