Entropy and dyadic equivalence of random walks on a random scenery

Deborah Heicklen
Christopher Hoffman
Daniel Rudolph

October 9, 1999

Abstract

For any 1-1 measure preserving map $T$ of a probability space, consider the $[T, T^{-1}]$ endomorphism and the corresponding decreasing sequence of $\sigma$-algebras. We demonstrate that if the decreasing sequence of $\sigma$-algebras generated by $[T, T^{-1}]$ and $[S, S^{-1}]$ are isomorphic, then $T$ and $S$ must have equal entropies. As a consequence, if the $[T, T^{-1}]$ endomorphism is isomorphic to the $[S, S^{-1}]$ endomorphism, then the entropy of $T$ is equal to the entropy of $S$. Along the way a relationship is shown between Feldman’s $\tilde{f}$ metric [2] and Vershik’s $v$ metric [9].

1 Introduction

A decreasing sequence of $\sigma$-algebras is a measure space $(X, \mathcal{F}_0, \mu)$ and a sequence of $\sigma$-algebras $\mathcal{F}_0 \supset \mathcal{F}_1 \supset \mathcal{F}_2 \ldots$. A natural example of this arises from a sequence of independent identically distributed random variables, $\{X_i\}$. Namely, set $\mathcal{F}_i = \sigma(X_i, X_{i+1}, \ldots)$. If the $X_i$ take on the values 1 and -1 with probability 1/2, then this sequence has the property that $\mathcal{F}_i|\mathcal{F}_{i+1}$ has 2 point fibers of equal mass for every $i$. A decreasing sequence of $\sigma$-algebras with this property is called dyadic. This example also has the property that $\cap \mathcal{F}_i$ is trivial. Two decreasing sequences of $\sigma$-algebras are isomorphic if there exists a 1-1 measure preserving map between the two spaces that maps the $i$-th $\sigma$-algebras to each other. A dyadic decreasing sequence of $\sigma$-algebras is standard if it is isomorphic to the decreasing sequence previously mentioned.
Anatoly Vershik began the modern study of such decreasing sequences of \( \sigma \)-algebras in [10]. In [9] Vershik showed that there exist dyadic sequences of \( \sigma \)-algebras with trivial intersection that are not standard. In [9] Vershik also gave a necessary and sufficient condition for a dyadic decreasing sequence of \( \sigma \)-algebras to be standard. His criterion is more general than for just the dyadic case, but we only mention this case since this is all we use. For a thorough treatment of classification of decreasing sequences of \( \sigma \)-algebras, see [4]. An equivalent description of standardness for dyadic decreasing sequences of \( \sigma \)-algebras is for there to exist a sequence of partitions \( \{P_i\} \) of \( X \) into two sets, each of measure 1/2, such that

1. the partitions \( P_i \) are mutually independent and
2. for each \( i \), \( \mathcal{F}_i = \bigvee_{n=i}^{\infty} P_n \).

In this paper we will be working with a decreasing sequences of \( \sigma \)-algebras arising from a certain class of endomorphisms. These are known as \([T, T^{-1}]\) \textbf{endomorphisms} or random walks on a random scenery. Let \( T \), the scenery process, be any 1-1 measure preserving map on a probability space \((Y, \mathcal{C}, \nu)\) such that \( T^2 \) is ergodic. Let \( \sigma \) be the shift on \((X, \mathcal{B}, \mu)\) where \( X = \{-1, 1\}^\mathbb{N} \), \( \mathcal{B} \) is the Borel \( \sigma \)-algebra, and \( \mu \) is \((\frac{1}{2}, \frac{1}{2})\) product measure. Define \([T, T^{-1}]\) on \((X \times Y, \mathcal{F}, \mu \times \nu)\) where \( \mathcal{F} = \mathcal{B} \times \mathcal{C} \) by

\[
[T, T^{-1}](x, y) = (\sigma x, T^{x \cdot y} y).
\]

Let \( \mathcal{F}_n = [T, T^{-1}]^{-n} \mathcal{F} \). \([T, T^{-1}]\) is 2-1, since any point \((x, y)\) has the preimages \((-1x, Ty)\) and \((1x, T^{-1}y)\). Since each preimage has equal relative measure, the \([T, T^{-1}]\) endomorphism generates a dyadic decreasing sequence of \( \sigma \)-algebras. Furthermore, since \( T^2 \) is ergodic, \( \cap \mathcal{F}_n \) is trivial [8]. Notice that if \( T \) is the trivial 1 point transformation then \([T, T^{-1}]\) reduces to the shift on \((X, \mathcal{B}, \mu)\), and the corresponding sequence is the standard dyadic example.

The above construction, when carried out for \( X = \{-1, 1\}^\mathbb{Z} \), yields a 1-1 map we refer to as the \([T, T^{-1}]\) \textbf{automorphism}. This is also the natural two-sided extension of the \([T, T^{-1}]\) endomorphism. Kalikow proved in [7] that if \( T \) has positive entropy then the \([T, T^{-1}]\) automorphism is not isomorphic to a Bernoulli shift. Building on Kalikow’s techniques, Heicklen and Hoffman proved that if \( T \) has positive entropy then the decreasing sequence of \( \sigma \)-algebras generated by the \([T, T^{-1}]\) endomorphism is not standard [5]. When \( T \) has 0 entropy the picture appears to be significantly more complicated. Feldman and Rudolph proved that if \( T \) is rank 1 then the \([T, T^{-1}]\) endomorphism generates a standard decreasing
sequence of $\sigma$-algebras [3]. On the other hand Hoffman has given an example of a zero entropy $T$ such that the $[T, T^{-1}]$ endomorphism is not standard [6]. For most zero entropy transformations it is not known whether the $[T, T^{-1}]$ endomorphism generates a standard decreasing sequence of $\sigma$-algebras. Feldman and Rudolph’s result, combined with the work of Burton [1], provides an example of an endomorphism which is standard, but is not isomorphic to a Bernoulli shift.

The purpose of this paper is to demonstrate that if the decreasing sequence of $\sigma$-algebras generated by $[T, T^{-1}]$ and $[S, S^{-1}]$ are isomorphic, then the entropy of $T$ is equal to the entropy of $S$. Since two endomorphisms which are isomorphic produce isomorphic decreasing sequences of $\sigma$-algebras, this result implies that if the $[T, T^{-1}]$ endomorphism is isomorphic to the $[S, S^{-1}]$ endomorphism then the entropy of $T$ is equal to the entropy of $S$. It is not known if there exist two transformation $S$ and $T$ with different entropies such that the $[T, T^{-1}]$ automorphism is isomorphic to the $[S, S^{-1}]$ automorphism. Our methods rely heavily on the tree structure of an endomorphism, which is explained shortly. As an isomorphism of the $[T, T^{-1}]$ automorphism and the $[S, S^{-1}]$ automorphism need not preserve the tree structure of the endomorphism our methods do not apply to the invertible case.

There is a tree structure associated with the preimages of any point in the $[T, T^{-1}]$ endomorphism. We introduce some notation for this tree structure. Fix a value $n$ for the height of the tree. An $n$ branch is an element $b \in \{-1, 1\}^n$. An $n$ tree is a binary tree of height $n$ consisting of $2^n$ branches. Given a point $(x, y)$, there is a map associating the $2^n$ branches of the $n$ tree with the $2^n$ elements of $[T, T^{-1}]^{-n}(x, y)$. This map takes the $n$ branch $b = (b_1, ..., b_n)$ to $(-b_1, ..., -b_n, x, T \Sigma b_j y)$. Based on this, if $P$ is a finite partition of $Y$ the labeled $n$ tree for a partition $P$ over a point $y \in Y$ assigns to each branch $b$ the label $P(T \Sigma b_j y)$. A node is an integer. A branch lands at a node $k$ if $\sum_1^n b_i = k$. A branch passes through a node $k$ at height $h$ if $\sum_{k+1}^n b_i = k$. This definition will be used repeatedly in Section 3. For $m \leq n$ define an $m$ subtree inside an $n$ tree to be a tree with $2^m$ branches such that the last $n - m$ coordinates of the branches all agree and the first $m$ coordinates vary over all possibilities. the $2^m$ branches of an $m$ tree inside an $n$ tree are associated with the $2^m$ elements of $[T, T^{-1}]^{-m}(x, y)$ which are mapped to the same point under $[T, T^{-1}]^m$. If a branch passes through a node $k$ at height $h$ then it is a branch in an $h$ subtree inside an $n$ tree, and all of the branches in this subtree pass through the node $k$ at height $h$. The distance between any two $m$ subtrees in an $n$ tree is $|\sum_{i=m+1}^n b_i - b'_i|$ for any branches $b$ and $b'$ passing through the two subtrees. Fix a partition $P$ and let $W$ and $W'$ be the labeled $n$ trees over $y$ and $y'$ respectively. The Hamming metric between two
labeled $n$ trees $W$ and $W'$ is given by

$$d_n(W, W') = \frac{\# \text{ of branches on which the labels of } W \text{ and } W' \text{ disagree}}{2^n}.$$ 

Let $A_n$ be the group of automorphisms of an $n$ tree. For an automorphism $a \in A_n$, let $f(a)$ be the largest integer such that $(a(b))_i = b_i$ for all $1 \leq i \leq f(a)$ and all branches $b$ if such an integer exists. If no such integer exists set $f(a) = 1$. Define $g(a) = 1/(1 + \log f(a))$ if $a \neq id$ and $g(a)=0$ if $a = id$.

Define

$$v_n^P(y, y') = \inf_{a \in A_n} (d_n(aW, W') + g(a)).$$

This is easily checked to be a pseudometric on points $y$ and $y'$ as it is a metric on the labeled $n$ trees $W$ and $W'$. If the partition $P$ is understood, it will be omitted from the formulation and we will write $v_n$.

In Section 2 we will briefly outline the path to the proof of our main result. In Sections 3 and 4 we prove that a tree automorphism on a labeled tree which produces a small minimum for distance $v_n(y, y')$ must have a certain form. This will establish a connection between $v_n$ and Feldman’s $f$ metric. In Section 5 we use this property in finite code approximations to $\Phi$ to prove that if the $[T, T^{-1}]$ and $[S, S^{-1}]$ endomorphisms generate isomorphic sequences of $\sigma$-algebras then $T$ and $S$ have the same entropy by showing that the exponential growth rate for $T$-names will bound that for $S$-names.

## 2 Outline of the Proof

On the way to proving our main result we answer the following two questions:

1. What kind of automorphism $a \in A_n$ will minimize the value of $v_n$ between the labeled trees over two points?

2. Given $T$ and a sufficiently small $\delta$, how does $\nu\{y' \mid v_n^P(y, y') < \delta\}$ behave as $n \to \infty$?

The answers to these questions will establish a connection between $v_n$ and Feldman’s $f$ metric between the $T, P, n$-names of $y$ and $y'$. We begin with the following result from [5].
Theorem 2.1 [5] If $T$ is a positive entropy transformation then there exists $\delta_0$ and a finite partition $P$ such that $(T, P)$ is an i.i.d. process and for any polynomial $p(n)$, $n$ sufficiently large and any point $y$

$$\nu\{y' \mid v_n^P(y, y') < \delta_0\} < \frac{1}{p(n)}.$$

In the case of $[T, T^{-1}]$ endomorphisms Vershik’s standardness criteria says that

$$\int v_n^P(y, y')d\nu \times \nu \to 0$$

for every finite $P$ iff $\{F_n\}$ is standard [9]. Thus Theorem 2.1 implies that if $T$ has positive entropy then the decreasing sequence of $\sigma$-algebras generated by the $[T, T^{-1}]$ endomorphism is not standard.

We strengthen this result to show that for $\delta$ small enough, $\nu\{y' \mid v_n^P(y, y') < \delta\}$ decays exponentially in $\sqrt{n}$. The proof of Theorem 2.1 in [5] does not give any indication of what type of tree automorphism is used to obtain $v_n^P(y, y') < \delta_0$. Building from the conclusion of this theorem we show that such an automorphism must take on a certain form. Namely, for most nodes $k$, it maps almost all of the branches that land at $k$ to branches that land at some single node $k'$. This gives a 1-1 map from most values $k$ to the value $A(k) = k'$. We next show that on a large subset of nodes, the map $A$ is monotone.

This map $A$ will establish a connection between $v_n$ and the $\tilde{f}$ of Feldman [2]. To remind the reader, the $\tilde{f}$ metric is defined as follows. For any $m, n$, and $w, w' \in \{0, ..., l\}^\mathbb{Z}$ let

$$\tilde{f}_{[m, n]}(w, w') = 1 - \frac{k}{n - m + 1},$$

where $k$ is the maximal integer for which there are subsequences of integers, $m \leq i_1 < i_2 < ... < i_k \leq n$ and $m \leq j_1 < j_2 < ... < j_k \leq n$ such that $w(i_r) = w'(j_r), 1 \leq r \leq k$. We also will use the $\tilde{d}$ metric on sequences. For any $m, n$, and $w, w' \in \{0, ..., l\}^\mathbb{Z}$ let

$$\tilde{d}_{[m, n]}(w, w') = 1 - \frac{k}{n - m + 1},$$

where $k$ is the number of $i, m \leq i \leq n$, such that $w(i) = w'(j)$.

There are some difficulties to expressing this connection simply. For one, if $n$ is even then $v_n(y, y')$ depends only on the even coordinates of $y$ and $y'$ between $-n$ and $n$, while $\tilde{f}$ depends on all the coordinates. Also, the number of branches in the $n$ tree over $y$ landing at node $k$ is given by the binomial distribution. Thus $v_n(y, y')$ depends much more heavily on
the values of \( y_i \) and \( y'_i \) for \( |i| < \sqrt{n} \) than for \( |i| > \sqrt{n} \). On the other hand \( \bar{f}_{[-n,n]} \) gives uniform weight to all coordinates in the interval \([-n, \ldots, n]\). Hence if \( v_n(y, y') \) is small, then we can only draw conclusions about \( \bar{f}_{[-c\sqrt{n},c\sqrt{n}]}(y, y') \). To overcome the first of these difficulties (and others as well) we assume the generating partition \( P \) has a certain form. Let \( Q \) be a partition which generates and is a refinement of a full entropy and i.i.d. partition for the action of \( T \). Now set \( P = Q \lor T(Q) \). Hence the sequence \( T^i(P) \) for \( i \) even determines \( T^i(Q) \) for all \( i \). We also will restrict ourselves to values \( n \) that are even and perfect squares. We obtain the following relationship between \( v_n \) and \( \bar{f}_{[-c\sqrt{n},c\sqrt{n}]} \).

**Theorem 2.2** Assume \( n \) is even and a perfect square and \( P \) is as described above. Given \( \epsilon > 0 \) and \( c > 1 \) there exists \( \delta > 0 \) and a good set \( G \) such that \( \mu(G) > 1 - \epsilon \) and for \( y, y' \in G \), if \( v_n(y, y') < \delta \) then

\[
\bar{f}_{[-c\sqrt{n},c\sqrt{n}]}(y, y') < \epsilon.
\]

The proof of this result is developed in Section 3 and completed in Section 4 except for certain technical material that appears in the last in Section 6. In Section 5 we use finite code approximations to the isomorphism of the decreasing sequence of \( \sigma \)-algebras and Theorem 2.2 to show that entropy is an invariant. Ignoring sets of small measure this is done as follows. Let \( \Phi \) be the isomorphism between \([T, T^{-1}]\) and \([S, S^{-1}]\) and suppose that \( \Phi(x, y) = (w, z) \) and \( \Phi(x', y) = (w', z') \). We will show that this implies \( z \) and \( z' \) are close in \( \bar{f} \). This implies that the number of names in the \( T \) process must grow at the same exponential rate as the number of names in the \( S \) process. Hence the entropies of the two processes must be equal.

### 3 Tree automorphisms

In this section we will show that for most \( y \) and \( y' \), if \( v_n(y, y') \) is small enough then an automorphism \( a \) which achieves this small value induces a bijection \( A \) from most of \([-n, \ldots, n]\) to most of \([-n, \ldots, n]\) with the following property. For most of the branches which land at node \( k \) in the tree over \( y \), the image of those branches under \( A \) land at the node \( A(k) \) in the tree over \( y' \).

First we define the small set of points \( y \) that are **degenerate**. We will deal only with points that are not degenerate. Then we define what it means for a branch in a tree to be **very good**. The very good branches are the large set of branches mentioned in the previous
paragraph. Finally, in Lemmas 3.5 and 3.6, we will show how the automorphism \( a \) induces the function \( A \).

For the rest of this section and the next we will set and then work with two parameters, \( \delta_0 \) and \( \delta \). At this point we assume \( \delta_0 \) to be a value satisfying Theorem 2.1. After Lemma 3.1 we will fix \( \delta_0 \). After Lemma 3.2 we will fix \( \delta \). Given a value for \( \delta \), fix \( n_0 = \lfloor 2^{1/\delta} \rfloor \), where \( \lfloor x \rfloor \) means the greatest integer less than or equal to \( x \). Remember we have fixed a partition \( P \) of the space \( Y \). The metric \( v_k \) depends on \( P \) but we suppress this dependence in the notation. Now we restrict our attention to a certain (large) class of \( y \) and \( y' \). For \( y \in \{0, ..., l\}^2 \), define

\[ \gamma_k = \{ y \mid \exists i \ 2k < |i| < k^5 \text{ such that } v_k(y, T^i(y)) < \delta_0 \}. \]

Also define

\[ B(\delta, \delta_0) = \{ y \mid \exists i 0 < |i| < n_0 \text{ such that } d([-n_0, n_0], y, T^i(y)) < \delta_0/4 \}. \]

Given a point \( y \) we say a node \( l \) is colored if

1. there exists \( j \) and \( k \) such that \( T^j(y) \in \gamma_k \) and \( l \in [j - k, j + k] \) or

2. there exists \( j \) such that \( T^j y \in B(\delta, \delta_0) \) and \( l \in [j - n_0, j + n_0] \).

We will use the term the \( n \) word of \( y \) to refer to the string of symbols \( y_{-n}, ..., y_n \), where \( y_i = P(T^i(y)) \). The colored nodes are bad for our arguments and our immediate goal is to show that they are scarce. The translation interval \( [2k, k^5] \) in the definition of \( \gamma_k \) is set as such to ensure that the measure of \( \gamma_k \) is small. A lower bound on the amount of translation is needed because any word can be matched pretty well in the \( v_n \) metric to a small translate of itself. See [3] for how this can be achieved. An upper bound is needed since if we allow a significantly larger translate (i.e. exponential in \( k \)), then any word is bound to eventually reoccur. Thus any word will have a close \( d \) matching (and a close \( v_n \) matching) to some translate of itself if the window size is large enough.

**Definition 3.1** The \( n \) word of \( y \) is \((\delta, \delta_0)\) degenerate if the fraction of branches in the \( n \) tree over \( y \) that land at colored nodes is more than \( \frac{\delta}{2} \).

The simplest example of a degenerate word is one with all symbols equal. We want to rule out this type of word, since applying any automorphism to this word will not change it and such an arbitrary automorphism does not have the property described earlier of acting essentially as a permutation on the nodes.
Lemma 3.1 Depending only on the process \((T, P)\), there exists \(\delta_0\) such that for all \(\delta\) sufficiently small, \(\mu(B(\delta, \delta_0)) < \delta^2/4\).

Proof: Suppose \(\{y_i\}\) are i.i.d. random variables on symbols \(\{1, \ldots, \ell\}\) with probability distribution \(\{p(1), \ldots, p(\ell)\}\). Assume \(p(1)\) to be the largest term of the probability vector. For any fixed \(i\), in order to get a \(d\) matching to within \(\delta \ell/4\), at least a fraction \(1 - \delta \ell/4\) of the coordinates must match. There are \(\binom{2n_0 + 1}{(2n_0 + 1)\delta \ell/4}\) different choices of subsets of indices to match. The probability that any particular pair matches is less than \(p(1)^{(1 - \delta \ell/4)(2n_0 + 1)}\). Hence

\[
\mu(\{y|d_{[-n_0, n_0]}(y, T^i(y)) < \delta_0/4\}) \leq \left( \frac{2n_0 + 1}{(2n_0 + 1)\delta \ell/4} \right) p(1)^{(1 - \delta \ell/4)(2n_0 + 1)}.
\]

Using an estimate on the binomial coefficients and summing over all possible \(i\) yields

\[
\mu(B(\delta, \delta_0)) \leq 2n_02^{h(\delta \ell/4)(2n_0 + 1)} p(1)^{(1 - \delta \ell/4)(2n_0 + 1)}.
\]

In our case \(P\) refines a fixed i.i.d. process and so we obtain this bound on \(\mu(B(\delta, \delta_0))\) where \(p(1)\) is the probability of the largest symbol in the i.i.d. process. We assume that \(\delta_0\) has been chosen small enough that

\[
p(1)^{(1 - \delta \ell/4)} < 2^{-2h(\delta \ell/4)}
\]

and hence

\[
\mu(B(\delta, \delta_0)) \leq 2n_02^{-h(\delta \ell/4)(2n_0 + 1)}.
\]

Since \(n_0 > 2^{1/\ell} - 1\), \(\mu(B(\delta, \delta_0))\) decays superexponentially in \(\delta\). Hence, for \(\delta\) sufficiently small \(\mu(B(\delta, \delta_0)) < \delta^2/4\).

Now we fix \(\delta_0\) such that the previous lemma is satisfied.

Lemma 3.2 For any sufficiently small \(\delta\), the set of all \(y\) which are degenerate for a given \(n > n_0 = \lfloor 2^{1/\ell} \rfloor\) has measure less than \(\delta\).

Proof: Applying Theorem 2.1 with \(p(k) = k^{11}\) shows that if \(\delta\) is small enough then there is an \(n_0\) with \(\nu(\gamma_k) < \frac{1}{k^2}\) for \(k > n_0\). Thus if \(\delta\) is small enough then \(\sum_{k \geq n_0} 2k\nu(\gamma_k) < \delta^2/4\). Using Lemma 3.1 we can choose \(\delta\) small enough so that \(\mu(B(\delta, \delta_0)) < \delta^2/4\). Now Chebychev’s lemma implies that the set of all \(y\) which are degenerate has measure less than \(\delta\).
Now we fix $\delta$ so that the previous two lemmas are satisfied and so that $n_0 = \lfloor 2^{1/\delta} \rfloor > N_0$, where $N_0$ is defined in Lemma 6.4. Theorem 2.2 and the first paragraph in this section apply to all points with nondegenerate $n$ words. In particular, we can choose the good set $G$ in this theorem to be the set of points whose $n$ words are nondegenerate. For the rest of this section and the next section we will fix, in addition to $\delta$ and $\delta_0$, $n$, two points $y$ and $y'$ whose $n$ words are not degenerate and $v_n(y, y') < \delta$, and an automorphism $a \in A_n$ that minimizes the quantity in the definition of $v_n(y, y')$.

Now we want to define the large set of very good branches on which the automorphism acts well. Before we can do this we need one intermediate definition.

**Definition 3.2** A branch $b$ is good for the automorphism $a$ if the following hold:

1. The label assigned to $b$ in the $n$ tree over $y$ is the same as the label assigned to $a(b)$ in the $n$ tree over $y'$,
2. $b$ and $a(b)$ do not land on a colored node, and
3. $b$ and $a(b)$ are $n_0$ regular.

We defer defining the large set of “$n_0$ regular” branches until Section 6. The property of two $n_0$ regular branches that we will use is if $b$ and $b'$ are $n_0$ regular branches and $g \geq n_0$, then there exists an $h \in (g^{1/4}, g^{1/2})$ such that $|\sum_1^h b_i|, |\sum_1^h b'_i| < 101 \sqrt{h}$. The difficulty arises in finding a height $h$ that satisfies this for both $b$ and $b'$ simultaneously. This property of $n_0$ regular branches is proved in Lemma 6.5.

**Lemma 3.3** If the $n$ words of $y$ and $y'$ are not degenerate and $v_n(y, y') < \delta$ then the fraction of good branches is greater than $1 - 10\delta$.

**Proof:** Because $v_n(y, y') < \delta$ the fraction of branches satisfying the first property is greater than $1 - \delta$. Since $y$ and $y'$ are not degenerate the fraction of branches satisfying the second property is greater than $1 - \delta$. In Lemma 6.4 we will show that the fraction of branches satisfying the third property is more than $1 - 8\delta$. 

**Definition 3.3** A branch $b$ is very good for the automorphism $a$ if for every $j \leq n$ the $j$ subtree through which the branch $b$ passes has at least $2^j(1 - C_1)$ good branches, where $C_1 = \left(\frac{5}{16}\int_{202}^{\infty} e^{-\frac{x^2}{2}} dx\right)$.
Lemma 3.4 If the $n$ words of $y$ and $y'$ are not degenerate, and $v_n(y, y') < \delta$, then the fraction of very good branches is more than $1 - \frac{10}{1\gamma} \delta$.

Proof: For each branch $b$ which is not good, consider the maximal $j$ such that the $j$ subtree containing $b$ has less than $2^j (1 - C_1)$ good branches. For this subtree, the fraction of branches in this $j$ subtree which are not good is greater than $C_1$. Hence
\[
\frac{\text{# of branches in this subtree}}{\text{# of not good branches in this subtree}} \leq \frac{1}{C_1}.
\]
The union of all these subtrees is the set of all not very good branches. This is because we chose $j$ to be maximal. Hence any branch which is not contained in one of the previously mentioned subtrees always passes through nodes with more than $1 - C_1$ fraction of good branches. Thus the fraction of branches which are not very good is less than $\frac{10\delta}{C_1}$. \[\]

We will show that the automorphism $a$ maps all of the very good branches in the tree over $y$ that land at one node to branches in the tree over $y'$ that land at a single node. In fact, the image of two very good branches that pass through the same node at height $h$ will be separated by no more than a distance $2h$. Before stating and proving the result formally, we give an idea of the proof.

Since the two branches pass through the same node at height $h$, the images of the two $h$ subtrees containing the two very good branches can be matched well in the Hamming distance after some tree automorphism. Thus the $y'$ names in the intervals that the trees lie over are close in $v_n$. This implies that either the intervals overlap, or they are separated by at least $h^5$. If they are far apart then, at some height $h'$, the images of the two very good branches must have been separated by some distance between $2h'$ and $(h')^5$. We can use the regularity condition to show that this implies there must be some good branches landing on colored nodes, which is a contradiction.

Lemma 3.5 Suppose the $n$ words of $y$ and $y'$ are not degenerate, and $v_n(y, y') < \delta$. Given $b$ and $b'$, any two very good branches for $a$, and $h \geq n_0$,

if $\sum_{i=h+1}^{n} b_i - b'_i = 0$ then $\sum_{i=h+1}^{n} a(b)_i - a(b')_i \leq 2h$,

Proof: The proof is by contradiction. Suppose there exists two very good branches, $b$ and $b'$, and height $h \geq n_0$ such that $b$ and $b'$ pass through the same node at height $h$, but $a(b)$ and $a(b')$ are in $h$ subtrees that are at least $2h$ apart. Call these subtrees $t_{a(b),h}$ and $t_{a(b'),h}$.
Since \( b \) and \( b' \) are very good they do not land on colored nodes, and \( t_{a(b),k} \) and \( t_{a(b'),k} \) must be a distance \( g \geq h^5 \) apart. By Lemma 6.5 it is possible to find a height \( h' \) with \( g^2 < h' < g^2 \) such that \( b \) and \( b' \) are in \( h' \) subtrees, \( t_{b,h'} \) and \( t_{b',h'} \), that are less than \( 202\sqrt{h'} \) apart.

Define the overlap between \( t_{b,h'} \) and \( t_{b',h'} \) at a node \( j \) at height \( h'/4 \) to be the minimum of the fraction of branches in \( t_{b,h'} \) that pass through node \( j \) at height \( h'/4 \) and the fraction of branches in \( t_{b',h'} \) that pass through node \( j \) at height \( h'/4 \). The overlap summed over all nodes at height \( h'/4 \) between the two subtrees is greater than \( \frac{1}{4}(\int_{\frac{202}{2}}^{\infty} e^{-\frac{x}{2}} dx) \). Since the fraction of branches in \( t_{b,h'} \) and \( t_{b',h'} \) that are not good is small compared with this number, there exists a choice of node \( J \) and two \( \frac{h}{4} \) subtrees, \( t_{b,h'/4} \) and \( t_{b',h'/4} \), in \( t_{b,h'} \) and \( t_{b',h'} \) respectively, which pass through the node \( J \) at height \( h'/4 \) and have at least \( 1 - \frac{\delta}{2} \) fraction of good branches. Moreover the distance between the images of these two subtrees, \( t_{a(b),h'/4} \) and \( t_{a(b'),h'/4} \), is between \( \frac{h}{2} \) and \( (\frac{h}{4})^6 \). Now the labeled trees corresponding to \( t_{b,h'/4} \) and \( t_{b',h'/4} \) are the same. The fraction of good branches in \( t_{b,h'/4} \) and \( t_{b',h'/4} \) is large enough so that

\[
v_{h'/4}(t_{b,h'/4}, t_{a(b),h'/4}) \leq \delta_0/2 \text{ and } v_{h'/4}(t_{b,h'/4}, t_{a(b'),h'/4}) \leq \delta_0/2.
\]

Thus

\[
v_{h'/4}(t_{a(b),h'/4}, t_{a(b'),h'/4}) \leq \delta_0.
\]

This implies that all of the branches in \( t_{a(b),h'/4} \) and \( t_{a(b'),h'/4} \) land on colored nodes. This is a contradiction with the fact that most of these branches are good.

\[ \end{proof} \]

**Lemma 3.6** If the \( n \) words around \( y \) and \( y' \) are not degenerate, and \( v_n(y, y') < \delta \) then given \( k \) there exists \( k' \) with the following property. For any very good branch \( b \) for the automorphism \( a \) with \( \sum_{i=1}^{n} b_i = k \), it follows that \( \sum_{i=1}^{n} a(b)_i = k' \).

**Proof:** Since \( f(a) \geq n_0 \) it suffices to show that if \( b \) and \( b' \) are two very good branches pass through the same node \( k \) at height \( n_0 \) then there exists \( k' \) such that \( a(b) \) and \( a(b') \) pass through the node \( k' \) at height \( n_0 \). In other words if \( b \) and \( b' \) are two very good branches and \( \sum_{n_0+1}^{h} b_i = \sum_{n_0+1}^{h} b'_i = k \) then \( \sum_{n_0+1}^{h} a(b)_i = \sum_{n_0+1}^{h} a(b')_i = k' \). By the previous lemma

\[
| \sum_{n_0+1}^{h} a(b)_i - \sum_{n_0+1}^{h} a(b')_i | \leq 2n_0
\]

As \( f(a) \geq n_0 \) we have \( y'_n \sum_{n_0+1}^{h} a(b)_i = n_0 \cdots, y'_n \sum_{n_0+1}^{h} a(b)_i + n_0 \) and \( y'_n \sum_{n_0+1}^{h} a(b')_i - n_0 \cdots, y'_n \sum_{n_0+1}^{h} a(b')_i + n_0 \) are within \( \delta_0/4 \) in \( d \). If

\[
| \sum_{n_0+1}^{h} a(b)_i - \sum_{n_0+1}^{h} a(b')_i | \neq 0
\]
then $a(b)$ and $a(b')$ land on colored nodes. This is a contradiction.

If we define $A(k) = k'$ for any node $k$ on which a very good branch lands, then $A$ satisfies the condition described at the beginning of this section.

4 The relation between $v_n$ and $\tilde{f}_{[-c\sqrt{n},c\sqrt{n}]}$

In this section we complete the proof of Theorem 2.2 based on the conclusions of the previous section. We still have fixed the same $y, y', a, \delta$ and $A$ from the previous section. We also fix $c > 0$. We want to show that $A$ is monotone on a large part of the interval $[-c\sqrt{n}, c\sqrt{n}]$. This will tell us that $y$ and $y'$ are $\tilde{f}$ close on this interval.

First we define the set of good nodes. We prove that the density of good nodes in the interval $[-c\sqrt{n}, c\sqrt{n}]$ is close to 1. In lemma 4.2 we show that if $j$ and $k$ are good nodes and $|j - k|$ is small then the ratio of $|j - k|$ to $|A(j) - A(k)|$ is close to one. These two facts allow us to define a set of nodes of slightly smaller density on which $A$ is monotone. Since $A$ is monotone on a set of large density $\tilde{f}(y, y')$ is small.

**Definition 4.1** A node $k$ is good for the automorphism $a$ if a fraction at least $1 - C_2$ of the branches with $\sum b_i = k$ are very good, where $C_2 = \frac{1}{4}(\int_{\frac{1}{2}}^{1} e^{-\frac{x}{2}} dx)(\frac{1}{2} e^{-1})$.

**Lemma 4.1** There exists a function $F(c)$ such that at least $1 - F(c)\delta$ fraction of the nodes in $[-c\sqrt{n}, c\sqrt{n}]$ are good. Moreover, the function $F(c)$ is independent of all other variables.

**Proof:** Each of these nodes has at least $\frac{1}{4}(\frac{2n}{\sqrt{n}}) e^{-\delta^2/2}$ branches landing at this node. Thus for each bad node in this region, there are at least $(C_2)\frac{1}{4}(\frac{2n}{\sqrt{n}}) e^{-\delta^2/2}$ branches that are not very good. Thus, by Lemma 3.4, there are at most

$$\frac{2n(10\delta)/C_1}{(C_2)(2n e^{-\delta^2/2})/A\sqrt{n}} < \frac{40e^{\delta^2/2}}{C_1C_2} \delta \sqrt{n}$$

nodes which are not good. Thus $F(c) = \frac{40e^{\delta^2/2}}{C_1C_2}$. 

We will now use this lemma to prove that $A$ does not distort the distances between two good nodes which are close together by too much.

**Lemma 4.2** Given $c$ and nodes $k$ and $j$ that satisfy the following:

1. $k$ and $j$ are good,
2. \(|k - j| < .1\sqrt{n}\) and
3. \(|k|, |j| < c\sqrt{n}\),

then \(.5|k - j| \leq |A(k) - A(j)| \leq 2|k - j|\).

**Proof:** The local central limit theorem tells us that for a fixed \(c\) there exists an \(N\) such that for any \(n \geq N\), \(|k - j| < .1\sqrt{n}\), and \(|k|, |j| < c\sqrt{n}\), the minimum of the two distributions

\[
P\left(\frac{(j-k)^2}{\sum_{i=1}^{n} b_i} \leq \frac{1}{n} \frac{(j-k)^2}{\sum_{i=1}^{n} b_i} = k\right) \text{ and } P\left(\frac{(j-k)^2}{\sum_{i=1}^{n} b_i} \leq \frac{1}{n} \frac{(j-k)^2}{\sum_{i=1}^{n} b_i} = j\right)
\]

has an amount of mass that is bounded below by \(\frac{1}{4}(\int_{\frac{1}{2}e^{-\frac{x^2}{2}}} dx)\).

Now, in each one of these subtrees in the overlap, if we have a subtree of height \(j - k\) centered over \(c\), \(k \leq c \leq j\), and \(|A(k) - A(j)| \geq 2|k - j|\), then we have two possibilities. Either \(A(j)\) is at least \(\frac{j-k}{2}\) further away from the center of the image of the subtree over \(c\) than \(j\) was from the center of the subtree over \(c\), or the same statement applies with \(A(k)\) and \(k\) replacing \(A(j)\) and \(j\). Without loss of generality assume that it applies to \(j\). Then the ratio of the number of branches in the image of the subtree that land at \(A(j)\) to the number of branches in the subtree that land at \(j\) is at most \(1 - \frac{1}{2}e^{-\frac{j-k}{2}}\). Thus the fraction of branches in the subtree that landed at \(j\) but whose image is not at \(A(j)\) is greater than \(\frac{1}{2}e^{-\frac{j-k}{2}}\). Now a fraction at least \(\frac{1}{4}(\int_{\frac{1}{2}e^{-\frac{x^2}{2}}} dx)(\frac{1}{2}e^{-\frac{j-k}{2}})\) of the branches that land at \(j\) do not land at \(A(j)\). Since this applies to all subtrees centered between \(j\) and \(k\), this contradicts the way \(C_2\) was chosen.

Now we will restrict to a smaller set of nodes on which \(A\) is monotone.

**Definition 4.2** Let \(K\) be the set of all nodes \(k\) such that

1. \(k \in [-c\sqrt{n}, c\sqrt{n}]\),
2. \(k\) is even
3. there does not exist an interval \([i, j], -c\sqrt{n} \leq i \leq k \leq j \leq c\sqrt{n}\), where at least \(\frac{1}{4}\) of the even nodes in \([i, j]\) are not good.

**Lemma 4.3** If \(v_n(y, y') < \frac{1}{160F(c)}\) then \(A\) is monotone on \(K\).
\textbf{Proof:} By lemma 4.1 a fraction at least \( 1 - F(c)\delta \) of the nodes in the interval are good. Restricted to this interval \( K^c \) can be written as the union of all intervals \([i, j] \subset [-c\sqrt{n}, c\sqrt{n}]\) such that more than \( \frac{1}{4} \) of \((i, j)\) is bad. There exist a disjoint collection of such intervals that covers at least half of \( K^c \). Thus \( |K^c| \leq 8F(c)\frac{1}{100e^2(c)}\sqrt{n} \leq \frac{1}{20}\sqrt{n} \).

It suffices to show that \( A \) is monotone on the intersection of \( K \) and any interval of length \( .1\sqrt{n} \) between \(-c\sqrt{n} \) and \( c\sqrt{n} \). This is true because the previous paragraph assures us that in each one of these intervals there are at least two nodes in \( K \). Choose any of these intervals. Take \( k \in K \) in this interval. Let \( k' \) be the minimal good node inside this interval such that \( k' > k \) and \( A(k') < A(k) \). Find a good node \( j \) such that \( j \in (k, k') \) and \((k' - k) > 4(k' - j)\). Then

\[ A(j) - A(k') = A(j) - A(k) + A(k) - A(k') > A(k) - A(k') \geq .5(k' - k) > 2(k' - j). \]

This contradicts Lemma 4.2.

We are now ready to prove that the map \( A \) is an \( \tilde{f} \) matching of \( y \) and \( y' \).

\textbf{Theorem 4.1} Given \( c \) and nondegenerate \( y \) and \( y' \), if \( v_n(y, y') < \frac{1}{100e^2(c)} \), then

\[ \tilde{f}_{[-c\sqrt{n}, c\sqrt{n}]}(y, y') < \frac{10}{c} \]

\textbf{Proof:} For most \( k \in K \), \( A(k) \in [-c\sqrt{n}, ..., c\sqrt{n}] \). This is true because each node in \([-c\sqrt{n}, ..., c\sqrt{n}] \) has at least \( \frac{1}{4}e^{-c^2/2}(\frac{2n}{\sqrt{n}}) \) branches landing at that node and there are and most \( \frac{2}{c}e^{-c^2/2}(2^n) \) branches landing outside \([-c\sqrt{n}, ..., c\sqrt{n}] \). Thus for at most \( \frac{8\sqrt{n}}{c} \) nodes \( k \in K \), \( A(k) \notin [-c\sqrt{n}, ..., c\sqrt{n}] \). Combining this with Lemma 4.3 and the fact that for any good \( k \), \( y_k = y'_{A(k)} \) proves the lemma is true.

Theorem 2.2 is a corollary of this theorem.

\section{Invariance of Entropy}

In this section we prove our main result, that if the \([T, T^{-1}] \) endomorphism and the \([S, S^{-1}] \) endomorphism produce isomorphic decreasing sequences of \( \sigma \)-algebras then the entropy of \( T \) is equal to the entropy of \( S \). The proof is by contradiction. We assume that \( h(S) < h(T) \). Thus in particular \( S^2 \) is of finite entropy and has a finite generating partition. Hence we can take \( S \) to be the shift map on some \( Z = \{1, 2, \ldots, q\}^\mathbb{Z} \) and the points \( s \in Z \) are of the

\[ \text{A}[\text{B}^2] \text{C} \]
form \( \{ s_i \}_{i \in \mathbb{Z}} \), \( s_i \in \{ 1, \ldots, q \} \) for all \( i \) and just the even terms of the sequence almost surely determines the full sequence, just as we did for \( T \).

We now explain how to construct a **finite code approximation** to the isomorphism \( \Phi \). Fix a \( c > 0 \) and let \( I_n \) consist of the even values in \( [-c\sqrt{n}, c\sqrt{n}] \). Assuming \( n \) to be an even perfect square, consider the increasing sequence of finite algebras \( C_n \) generated by sets of the form

\[
\{(x, z) \mid (x, z, \ldots, z_{-c\sqrt{n}}, \ldots, z_{c\sqrt{n}}) \in \{ -1, 1 \}^{[0,n]} \times \{ 1, \ldots, q \} \}
\]

partitioning \( X \times Z \). As these partitions refine to points, the associated algebras will increase to the full algebra of measurable sets.

For a natural number \( n' \) and for each \( (w, s) \in X \times Z \) consider the \( n' \) tree over \( (w, s) \). The map \( \Phi \) takes this tree to the \( n' \) tree over \( \Phi(w, s) \) by some tree automorphism \( a_{n'}(w, s) \). As there are only finitely many automorphisms of a binary tree of height \( n' \), this gives a finite partition of \( X \times Z \). Namely, partition \( X \times Z \) into sets based on which automorphism of the \( n' \) tree \( \phi \) uses. Call this partition \( R = R(n', \Phi) \). For any choice \( \epsilon > 0 \) there will be an \( n \geq n' \) so that \( R \) can be approximated to within \( \epsilon \) by a \( C_n \)-measurable partition. That is to say, we can choose an automorphism \( a_n^0(w, s) \) of the \( n' \) tree over \( (w, s) \) depending solely on the value of the first \( n \) coordinates of \( w \) and the values \( s_i, i \in I_n \) and \( a_n^0(w, s) \) will agree with \( a_n(w, s) \) on all but \( \epsilon \) in measure of \( X \times Z \).

Now define a new isomorphism \( \Phi_n \) as follows. Setting \( \Phi(w, s) = (u, t) \) we have \( \{u_1, \ldots, u_n\} = a_n(w, s)(\{w_1, \ldots, w_n\}) \). Define the new image point \( (u', t') = \Phi_n(w, s) \) by setting \( \{u'_1, \ldots, u'_{n'}\} = a_n^0(w, s)(\{w_1, \ldots, w_{n'}\}) \) and for \( i > n' \), \( u'_i = u_i \). That is to say just replace the \( n' \) initial terms of \( u \) by those terms obtained from the action of \( a_n^0(w, s) \).

To define \( t' \) let \( j = a_n^0(w, s)(\{w_1, \ldots, w_{n'}\}) - a_n^0(w, s)(\{w_1, \ldots, w_n\}) \) This is the difference between where the original image and the new image branches land. Set \( t' = T'(t) \). It is easy to see that \( \Phi_n \) is again an isomorphism between the two sequences of \( \sigma \)-algebras. We refer to such a \( \Phi_n \) as a **finite code approximation** to \( \Phi \). Notice that for all but \( \epsilon \) of \( X \times Z \), \( \Phi \) and \( \Phi_n \) agree.

As the partitions \( C_n \) increase to the whole \( \sigma \)-algebra we can also construct a \( C_n \)-measurable approximation to the partition \( \Phi^{-1}(P) \). We call this approximation \( \tilde{P} \). This means we can assign a name to a point \( (w, s) \), depending only on the initial \( n \) terms of \( w \) and the values \( s_i, i \in I_n \) and have this name agree with the \( P \)-name of \( \Phi(w, s) \) on all but \( \epsilon \) of \( X \times Z \). Note that \( \Phi \) and \( \Phi_n \) may not take \( \tilde{P} \) to a partition of \( X \times Z \) that depends only on the \( Z \) coordinate. If \( \Phi \) and \( \Phi_n \) agree at \( (w, s) \) and the \( P \)-name of \( \Phi(w, s) \) agrees with its finite
approximation from \( C_n \) then we say \((w, s)\) codes well relative to these approximations.

To sketch the path to the conclusion, consider any two points \((w, s)\) and \((w', s')\) for the \([S, S^{-1}]\) endomorphism such that \( s_i = s'_i \) for all \( i \) in \( I_n \). Suppose \( \Phi(w, s) = (u, t) \), and \( \Phi(w', s') = (u', t') \). Using a finite code approximation of \( \Phi \) we will show that there is a natural automorphism \( a \in A_n \) which makes \( v_n(t, t') \) small. Theorem 4.1 now implies that \( \bar{f}^{-c\sqrt{n}, c\sqrt{n}}(t, t') \) is small. This shows that the exponential number of names in the \( S \) process must be at least as big as the exponential number of names in the \( T \) process, which is a contradiction.

To make this sketch precise, let \( T \) and \( S \) be 1-1 measure preserving transformations as described above. For \( w \in \{ -1, 1 \}^n \) define \( \bar{w} = \sum_{i=1}^{n} w_i \). Let \( \Phi \) be an isomorphism of the decreasing sequences of \( \sigma \)-algebras generated by \([T, T^{-1}]\) and \([S, S^{-1}]\). For any given \((w, s), c, \) and \( n \) consider the set

\[
E = E_{(w, s), c, n} = \{(w', s')| s_{i-\bar{w}} = s'_{i-\bar{w}} \text{ for all even } i \in [-c\sqrt{n}, c\sqrt{n}] \}.
\]

By the Shannon-McMillan theorem for large \( n \) we can cover all but \( \epsilon \) of the points \((w, s)\) with \( 2^{(k(S)+c)(2c\sqrt{n}+1)} \) of these sets as we assume we have chosen a generator for \( S^2 \).

For any set \( A \) define \( \Phi^*(A) \) to be the projection of \( \Phi(A) \) onto its second coordinate. Now we show that for most of the sets \( E \), \( \Phi^*(E) \) is contained in a small \( f \) neighborhood. We do this by showing that \( \Phi^*(E) \) is contained in a small \( v_n \) neighborhood and applying then Theorem 4.1.

**Lemma 5.1** For any \( \epsilon > 0 \), there exists a good set \( G \), with \( \mu(G) > 1 - \epsilon \), \( c_0 \), and \( N_0 \) such that for any \((w, s) \in G, c > c_0, \) and \( n > N_0 \), the following property holds. If \( E \) is the set associated to \((w, s), c \) and \( n \) and if \((u, t), (u', t') \in \Phi(G \cap E) \), then

\[
\bar{f}^{-c\sqrt{n}, c\sqrt{n}}(t, t') < \epsilon.
\]

**Proof:** Suppose \( \Phi(w, s) = (u, t) \) and \( \Phi(w', s') = (u', t') \). There exists a natural automorphism \( a \) between the tree over \( T^2 t \) and the tree over \( T^2 t' \). This is \( a = a_1(Id)(a_2)^{-1} \), where \( a_1 \) is the restriction of \( \Phi \) to the tree of the \( 2^n \) preimages of \([T, T^{-1}]^n(w, s)\) and \( a_2 \) is the restriction of \( \Phi \) to the tree of the \( 2^n \) preimages of \([T, T^{-1}]^n(w', s')\). \( Id \) is the map from the tree of the \( 2^n \) preimages of \([T, T^{-1}]^n(w', s')\) to the tree of the \( 2^n \) preimages of \([T, T^{-1}]^n(w, s)\) that acts as the identity automorphism on the tree.

Let \( n' = n_0 \) from the previous section. Choose a finite approximation \( \Phi_N \) to \( \Phi \) so that the set where the first \( n_0 \) coordinates of \( \Phi_N(w, t) \) and \( \Phi(w, t) \) do not agree is small.
The branches \( b \) and \( a(b) \) have the same label if \( \Phi_N \) coded well to both \( b \) and \( a(b) \) and if 
\[
| \sum_{i=1}^{n'} (a_2^{-1}b_i) | < c \sqrt{n} \quad \text{for all } n' \leq n.
\]
For any one of these branches \( b \), \( a(b)_j = b_j \) for all \( j < n_0 \). Thus it causes no loss of generality to modify \( a \) so that for all branches \( a(b)_j = b_j \) for all \( j < n_0 \).

The fraction of branches that ever go outside of the interval \([-c \sqrt{n}, c \sqrt{n}]\) is less than 
\[
\frac{4}{c^2} = e^{-c^2/2}.
\]
For any value \( c \) large enough, by first choosing \( n_0 \) large enough and then \( N \) large enough we will obtain 
\[
v_n(T^a_t, T^{a'}_t) < \frac{4}{c} e^{-c^2/2} < \frac{C_1 C_2}{12,800} e^{-c^2/2} = \frac{1}{160 c F(c)}.
\]

The second inequality is by definition of \( C_2 \) and the equality is by definition of \( F(c) \). Finally, applying Theorem 4.1, we conclude 
\[
\tilde{f}_{[-c \sqrt{n}, c \sqrt{n}]}(T^a_t, T^{a'}_t) < \frac{10}{c}.
\]

Choose \( b \) so that 
\[
\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx < \epsilon / 100
\]
and \( c \) so that \( c > \frac{20b}{c} \). Set \( G \) to be the set of all points \((w, s)\) such that 

1. at least \( 2^n (1 - 1/160cF(c)) \) of the \( 2^n \) preimages of \([S, S^{-1}]^n (w, s)\) code well,

2. the scenery of \( \Phi([S, S^{-1}]^n (w, s)) \) is a nondegenerate name, and

3. \( \bar{u} < b \sqrt{n} \).

Consider \((w, s)\) and \((w', s')\) in \( G \cap E \). Since \( c > \frac{20b}{c} \) we have that 
\[
\tilde{f}_{[-c \sqrt{n}, c \sqrt{n}]}(T^a_t, T^{a'}_t) < 10 \frac{c}{20b} < c/2.
\]
Since \( \bar{u}, \bar{u}' < b \sqrt{n} \) we have that \( \tilde{f}_{[-c \sqrt{n}, c \sqrt{n}]}(t, t') < \epsilon \).

**Theorem 5.1** If the decreasing sequence of \( \sigma \)-algebras generated by the \([T, T^{-1}] \) endomorphism is isomorphic to the decreasing sequence of \( \sigma \)-algebras generated by the \([S, S^{-1}] \) endomorphism then \( h(T) = h(S) \).

**Proof:** Suppose there was an isomorphism \( \Phi \) and \( h(S) < h(T) \). Consider the generator \( P \) described in section. Excluding a set of \( P \) names of small measure, any \( \epsilon \) neighborhood in \( \tilde{f} \) can have measure at most 
\[
\left( \frac{c \sqrt{n}}{\epsilon c \sqrt{n}} \right)^2 |P|^{-c \sqrt{n} - (1 - 2\epsilon) h(T) - c \sqrt{n}}.
\]
Thus any set $\Phi^*(G \cap E)$ can have measure at most

$$4n^2 \left( \frac{c}{\epsilon} \frac{\sqrt{n}}{n} \right)^2 |P|^{c \epsilon \sqrt{n} 2 - (1-2\epsilon)h(T) \sqrt{n}}.$$ 

By the comment above there are at most $2^{(h(S)+\epsilon)\sqrt{n}}$ of these neighborhoods and they cover all but $2\epsilon$ of the space. Because $\epsilon$ and $n$ are arbitrary and $h(S) < h(T)$ this is a contradiction.

**Corollary 5.1** If the $[T, T^{-1}]$ endomorphism is isomorphic to the $[S, S^{-1}]$ endomorphism then $h(T) = h(S)$.

**Proof:** If two endomorphisms are isomorphic then they generate isomorphic decreasing sequences of $\sigma$-algebras. Thus Theorem 5.1 implies the corollary.

### 6 Regularity Condition

There remains a gap in the proof of Theorem 5.1. That is, to define what it means for a branch to be $n_0$ regular and to develop its properties. To begin we say that a branch $b$ behaves well at height $h$ if $|\sum_{i=1}^{h} b_i| < 101 \sqrt{h}$. The definition of $n_0$ regular will imply that if $b$ and $b'$ are $n_0$ regular branches then, for any $h \geq n_0$, there exists an $h' \in (h^{1/4}, h^{1/2})$ such that both $b$ and $b'$ behave well at height $h'$.

We will define a sequence $\{h_j\}$ and show not only that there exists such an $h'$, but that it can be chosen to be an element of the sequence $h_j$. The elements of the sequence $h_j$ will be chosen far enough apart so that we will be able to use the law of the iterated logarithms to prove that the set of branches that behave well at $h_j$ and the set of branches that behave well at $h_{j+1}$ are almost independent. We will then apply the exponential rate of convergence for the weak law of large numbers to show that a large proportion of the branches behave well at more than half of the $h_j \in (h^{1/4}, h^{1/2})$, for all $h$ sufficiently large. Thus, for any two of those branches, we will be able to find at least one height in the interval $(h^{1/4}, h^{1/2})$ where both branches behave well.

Define $h_1 = 100$ and $h_{j+1} = [h_j(1 + 2 \log(\log(h_j)))^2]$. Now we show that the set of branches that behave well at $h_j$ and the set of branches that behave well at $h_{j+1}$ are almost independent.
Lemma 6.1  For any $b_1, ..., b_{h_j}$ such that $|\sum_{i=1}^{h_j} b_i| < 2\sqrt{h_j} (\log(\log h_j))$

$$P\left( \left| \sum_{i=1}^{h_j+1} b_i \right| < 101 \sqrt{h_{j+1}} \mid b_1, ..., b_{h_j} \right) > .9999.$$  

Proof: If both $|\sum_{i=1}^{h_j+1} b_i| < 100 \sqrt{h_{j+1}}$ and $|\sum_{i=1}^{h_j} b_i| < 2\sqrt{h_j} (\log(\log h_j)) < \sqrt{h_{j+1}}$ are true, then $|\sum_{i=1}^{h_j+1} b_i| < 100 \sqrt{h_{j+1}}$. Since the first event is independent of $b_1, ..., b_{h_j}$, the conditional probability is at least the probability of the first event. By Chebychev’s inequality this is at least .9999. 

Lemma 6.2 Between $2^I$ and $2^{I+1}$ there exist at least $2^I/4I$ elements of the subsequence $h_j$.

Proof: The ratio $h_{j+1}/h_j$ for any $h_j$ in the interval $[2^I, 2^{I+1}]$ is at most $(1+2\log(\log(2^{I+1})))^2$. Thus there are at least $k$ such $h_j$, where $k$ is the greatest integer such that the inequality

$$2^I (1+2\log(\log(2^{I+1})))^{2k} < 2^{I+1}$$

is still true. Now the following calculation proves the lemma.

$$2^I (1+2\log(\log(2^{I+1})))^{2k+2} > 2^{I+1}$$

$$(1+2(1+1))^{2k+2} > 2^I$$

$$(2k+2) \log(1+2(1+1)) > 2^I$$

$$k > 2^I/4I.$$ 

Now we define

$$G_{h,p} = \left\{ b \mid \frac{|\{h_j|h_j^{1/4} < h_j < h_j^{1/2} \text{ and } \sum_{i=1}^{h_j} b_i < 101 \sqrt{h_j}\}|}{|\{h_j|h_j^{1/4} < h_j < h_j^{1/2}\}|} > p \right\}.$$ 

This is the set of branches that behave well on a fraction at least $p$ of the $h_j$ between $h_j^{1/4}$ and $h_j^{1/2}$. Now we show that the fraction of branches that follow the law of the iterated logarithms, but are not in $G_{2^I, p}$, is decreasing faster than exponentially in $I$.

Lemma 6.3 There exists $C < 1$ such that for all $I$ and all $I' > I$

$$\mu\left( G_{2^I, p} \mid \left| \sum_{i=1}^{h_j} b_i < 2\sqrt{h_j} (\log(\log h_j)) \text{ and } h_j > 2^I \right| \forall h_j > 2^I \right) < 2C^{2^{I'/4I'}}.$$ 

19
**Proof:** For every branch \( b \) there exists a subset \( S \) of the elements of \( h_j \) between \( 2^{2^I} \) and \( 2^{2^I+1} \) such that \( b \) does not behave well on every \( h_j \in S \) and \( b \) does behave well on every \( h_j \notin S \). If \( b \notin G_{2^{2^I+2} \cdot 9} \) then the cardinality of \( S \) is at least \( m/10 \), where \( m \) is the number of \( h_j \) between \( 2^{2^I} \) and \( 2^{2^I+1} \). If \( S \) has \( k \) elements then Lemma 6.1 implies that the fraction of branches that don’t behave well on \( S \) is less than \((.0001)^k\). Since the number of subsets with \( k \) elements is \( \binom{m}{k} < 2^m \), the probability that \( b \notin G_{2^{2^I+2} \cdot 9} \) is bounded by

\[
\sum_{k=m/10}^{m} 2^m(.0001)^k < 2^m(.0001)^{m/10} \sum_{0}^{9m} (.0001)^k < 2(1(.0001)^{1})^m < 2C^m.
\]

The previous lemma says that \( m \geq 2^I / 4I' \) so the lemma is true.

**Definition 6.1** A branch \( b \) is \( n_0 \) regular if

1. \( |\sum_{i} h_{j} b_{i}| < 2 \sqrt{h_{j} (\log(\log h_{j}))} \) for all \( h_{j} \geq n_0 \) and
2. \( b \in G_{h_{j} \cdot 9} \) for all \( h \geq n_0 \).

**Lemma 6.4** There exists an \( N_0 \) such that for any \( n_0 > N_0 \) all but \( 4/\log n_0 \) of the branches are \( n_0 \) regular.

**Proof:** First assume \( n_0 = 2^{2^{I+2}} \) for some \( I \). The rate of convergence to the law of iterated logarithms shows that the fraction of branches that don’t satisfy

\[
|\sum_{i} h_{j} b_{i}| < 2 \sqrt{h_{j} (\log(\log h_{j}))} \text{ for all } h_{j} \geq 2^{2^I}
\]

is less than \( 1/6(2^I) \) for \( I \) sufficiently large. Lemma 6.3 implies that conditioning on Line 1 holding, the fraction of the branches that do not satisfy

\[
b \in G_{2^{2^I+2} \cdot 9} \text{ for all } I' \geq I
\]

is less than \( 1/6(2^I) \) for \( I \) sufficiently large. Thus Lines 1 and 2 are satisfied for all but \( 1/3(2^I) < 2/\log n_0 \) of the branches.

If \( b \) satisfies Line 1 then \( b \) satisfies the first condition in the definition of \( n_0 \) regular. Now we show that if \( b \) satisfies Line 2 then \( b \) satisfies the second condition in the definition of \( n_0 \) regular. If \( 2^{2^I+1} > h_{j}^{1/4} \geq 2^{2^I} \) there are at most three times as many \( h_{j} \) between \( 2^{2^I} \) and \( 2^{2^I+2} \) as there are between \( h_{j}^{1/4} \) and \( h_{j}^{1/2} \). If \( b \in G_{2^{2^I+2} \cdot 9} \) and \( b \in G_{2^{2^I+3} \cdot 9} \), then \( b \) behaves well on at least 90% of the \( h_{j} \) between \( 2^{2^I} \) and \( 2^{2^I+2} \). Thus \( b \) behaves well on at least 70%
of the $h_j$ between $h^{1/4}$ and $h^{1/2}$. So if a branch $b \in G_{2^{2l'+2}}$ for all $l' \geq I$ then $b \in G_{h,I}$ for all $h$ such that $h^{1/4} \geq 2^{2l}$ and $h \geq (2^{2l})^4 = 2^{2l+2} = n_0$. Thus the fraction of branches that are not $n_0$ regular is less than $2/ \log n_0$. Note that if $b$ is $n_0$ regular then it is $n$ regular for all $n \geq n_0$. Thus for arbitrary $n_0$ the fraction of branches that are not $n_0$ regular is less than $4/ \log n_0$.

\textbf{Lemma 6.5} If $b$ and $b'$ are $n_0$ regular branches and $h \geq n_0$, then there exists an $h' \in (h^{1/4}, h^{1/2})$ such that $|\sum_{i=1}^{h'} b_i|, |\sum_{i=1}^{h'} b'_i| < 101 \sqrt{h'}$.

\textbf{Proof:} This follows from the second condition of the definition of $n_0$ regular.

\textbf{References}


[6] C. Hoffman. A zero entropy $T$ such that the $(T, \text{Id})$ endomorphism is not standard. to appear in \textit{Proceeding of the AMS.}

