Instability of the Lempel-Ziv algorithm

October 12, 2017

1 Main Theorem

Theorem 1. *For any order function $\sigma : \mathbb{N} \to \mathbb{N}$ and any real $0 < \varepsilon < 1/4$, there is a stationary ergodic measure P computable with respect to σ with entropy $0 < H \leq \varepsilon$ such that for each universal code $\{\phi_n\}_{n \in \mathbb{N}}$, there is a sequence $x \in A^\infty$ such that*

1. *$d_P(x[1 \dots n]) \leq \sigma(n)$ - i.e. when σ is a slow-growing order function, x can be made fairly close to incompressible.*
2. *We have*

$$\limsup_{n\to\infty} \rho_{\phi_n}(x[1 \dots n]) \geq \frac{1}{4} \tag{1}$$

$$\liminf_{n\to\infty} \rho_{\phi_n}(x[1 \dots n]) \leq \varepsilon \tag{2}$$

The import of the theorem is as follows. We know that the Lempel-Ziv class of compressors is optimal with respect to stationary ergodic sources in the following sense :

$$P\{x \in A^\infty \mid \rho(x) = H\} = 1. \tag{3}$$

This statement says that for some set of infinite sequences which has probability 1, the Lempel-Ziv compressibility ratio attains the entropy rate of the stationary ergodic source.

We can refine the above statement by specifying one well-known probability one set for which the statement holds. This is the set of <u>Martin-Löf random</u>, or <u>Kolmogorov incompressible</u> sequences. We now introduce this class.

1.1 Kolmogorov incompressible sequences

We know that a Turing Machine can be specified using a 7-tuple consisting Input alphabet, Output alphabet, Blank Symbol, Transition Function, Start state, Accept state and Reject state. Since each of these can be encoded using a binary string, we can encode Turing machines using strings in A^*. Using such an encoding, we can rigorously talk about the <u>size</u> of a Turing Machine.

This allows us to say that a string is simple if there is a small Turing machine which outputs it, and incompressible when the smallest Turing machine outputting it is at least as long as the string itself.

In the study of randomness or incompressibility of infinite sequences, we need to extend the notion of computation to deal with infinite sequences. The definition of a Turing machine remains the same, but the notion of computation over infinite sequences is slightly different.

Definition 2. We say that a computable function $f : A^* \to A^*$ is monotone if given strings x and x', if x is a prefix of x', then $f(x)$ a prefix of $f(x')$.

We say that a string $y \leq z$ if y is a prefix of z. With the introduction of a $\leq$ ordering, we can now use the notion of a supremum.

Definition 3. For an infinite sequence $\alpha \in A^\infty$, the result of the computation of a computable monotone function $f : A^* \to A^*$ is defined by

$$f(\alpha) = \sup\{y \mid x \text{ is a finite prefix of } \alpha, \text{ and } f(x) = y\}. \tag{4}$$

We can imagine computable monotone functions being computed by Turing machines working on an infinite input, and outputting on a write-only tape. It can read an arbitrary finite number of input symbols before outputting another bit. However, once a bit has been output, the machine can never erase it. In that sense, it is a monotone computation - the output never shortens over time. (The output corresponding to an infinite input need not be itself infinite. However, the definition allows infinite outputs.)

Definition 4. The size of a monotone computable function $f : A^* \to A^*$ is the length of the encoding of the smallest Turing machine which computes f.

Definition 5. The monotone complexity of a finite string $x \in A^*$, denoted $Km(x)$, is the size of the smallest monotone computable function $f : A^* \to A^r*$ such that $f(\lambda) = x$.

Fix $c \in \mathbb{N}$.

We say that x is monotone incompressible if $Km(x) \geq n - c$. We say that x is monotone incompressible with respect to a probability measure P if $Km(x) \geq -\log P(x) - c$.

We can extend the notion of an incompressible finite string to that of infinite sequences in the natural way. We say that an infinite sequence is incompressible if all its finite prefixes are incompressible.

Definition 6. An infinite sequence $\alpha \in A^\omega$ is monotone incompressible with respect to a probability measure P if there is a constant c such that for all n, $Km(\alpha[1 \dots n]) \geq -\log P(\alpha[1 \dots n]) - c$.

Observation. Since a Lempel-Ziv compressor can be implemented by a monotone Turing machine, it follows that for all sufficiently large n,

$$Km(\alpha[1 \dots n]) \leq \mathcal{L}(\alpha[1 \dots n]),$$

where $\mathcal{L}(\alpha[1 \dots n])$ is the output of the Lempel-Ziv compression of the prefix of α. Hence if we pick monotone incompressible sequences with respect to P, then they are Lempel-Ziv incompressible.

Fact. By a standard counting argument, we can show that most strings of any given length n are incompressible. It is a well-known result in the theory of algorithmic randomness that the set of incompressible sequences with respect to a computable probability P has P-probability 1. We will assume this fact without proof.

1.2 Motivation for the result

It is possible to show that for every P-incompressible sequence α, $\rho(\alpha) = H$. This refines our earlier statement that $\rho = H$ with probability 1. We know that for a specific probability 1 set, namely the set of all P-incompressible sequence, $\rho = H$ for every element of this set.

Now, we show that if P-incompressibility is violated by a small amount, then we can construct an α which is compressibly to that extent, where $\rho(\alpha) \neq H$.

This is the manner in which Lempel-Ziv optimality is "non-robust" - a small violation in incompressibility is sufficient to alter the optimality of the compressor.[1]

2 An introduction to Cutting and Stacking

The sample space we work with is the set of infinite sequences from a finite alphabet A, denoted A^∞ (usually identified with the base-A expansion of numbers in $(0, 1)$). We study properties of a probabilistic system $(A^\infty, \mathcal{F}, P)$ under iterated applications of a measure-preserving (or ergodic) transformation $T : A^\infty \to A^\infty$. A single application of T can be seen as motion of the system in a single time-step. We are thus interested in the long-term evolution of a system under "motion". This is the origin of the term "symbolic dynamics" for this subject.

V'yugin's construction uses a technique called "cutting and stacking". This is a comparatively involved method to construct stationary probability distributions with a set of desired properties. With care, it is also possible to construct stationary ergodic probability distributions with desired properties.

In the method of cutting and stacking, we are provided initially with A^∞ and the σ-algebra $\mathcal{F}$. We have to construct a measure-preserving transformation $T : A^\infty \to A^\infty$ and an associated probability measure P preserved by T, in addition to certain set of desired properties.

2.1 Gadgets

We start with some basic terminology. Consider the uniform probability measure μ on the unit interval and a transformation $T : (0, 1) \to (0, 1)$. A partition $\Pi = (\pi_1, \ldots, \pi_k)$ is a sequence of pairwise disjoint subsets of $(0, 1)$ whose union is the interval.

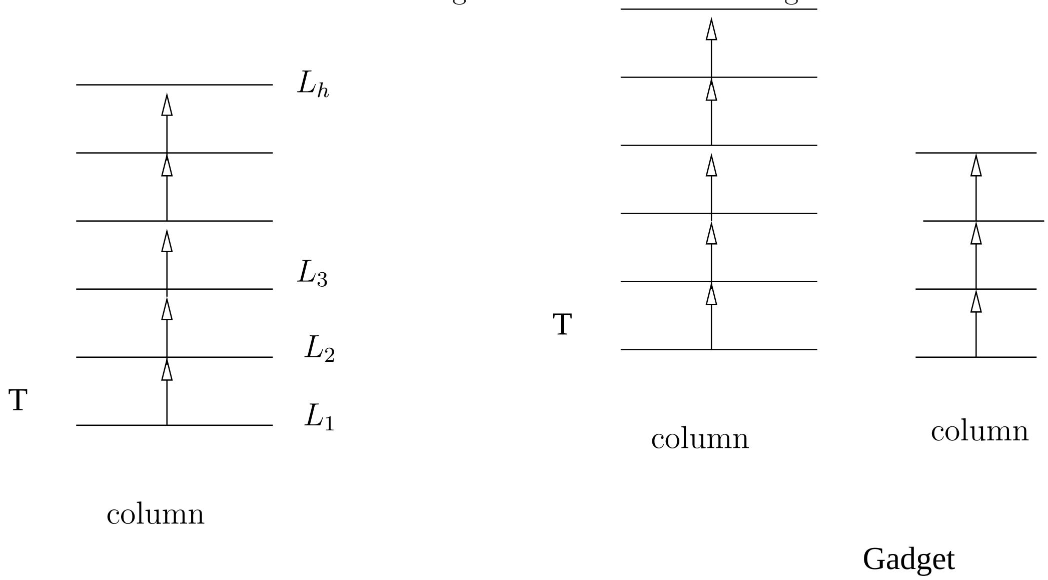
Let $A = \{1, \ldots, k\}$, an alphabet with the same cardinality as Π. A transformation $T : (0, 1) \to (0, 1)$ determines a measure on A^* and A^∞ as follows.

$$P(a[1 \ldots n]) \quad = \quad \lambda(\omega \mid \omega \in [0, 1),\ T^i(\omega) \in \pi_{a[i]},\ 1 \leq i \leq n\}, \tag{5}$$

i.e. The probability of a finite string $a[1 \ldots n]$ is the uniform probability of the set of all $\omega \in [0, 1)$ such that $T^1(\omega) \in \pi_{a[1]}$, $T^2(\omega) \in \pi_{a[2]}$, $\ldots$, and $T^n(\omega) \in \pi_{a[n]}$. For example, $P(1211)$ is the probability of the set of all $\omega \in [0, 1)$ such that $T^1(\omega) \in \pi_1$, $T^2(\omega) \in \pi_2$, and so on.

[1] Note that with $o(n)$ changes, H of the new sequence is the same as before, since for a stationary measure this is the finite-state block entropy of the original sequence. The point here is that the Lempel-Ziv compressibility changes. Thus with $o(n)$ changes, a finite-state incompressible sequence and Lempel-Ziv somewhat-compressible sequence is converted into another finite-state incompressible sequence which is now very highly Lempel-Ziv compressible.

Figure 1: Columns and Gadgets



This roughly corresponds to the "base-k" encoding if Π consists of the natural ordering of subintervals of the unit interval, each of length $1/k$.

We now introduce the main notions and properties of the "cutting and stacking" method. A column is a sequence $E = (L_1, \ldots, L_h)$ of pairwise disjoint subintervals of equal width, with L_1 being the base of the column and L_h the top, the width of the column $w(E)$ is the length of L_1, $\hat{E} = \cup_{i=1}^{h} L_i$ the support of the column, and $\lambda(\hat{E})$ the measure of E. The idea of a column is that $T(L_i) = L_{i+1}$ for $1 \leq i \leq h-1$. The inverse image of L_1 and the image of L_h under T are not specified by the column. Any point $\omega \in L_j$ defines a finite trajectory

$$\omega, T^{\omega}, T^2\omega, \ldots, T^{h-j}\omega.$$

A partition Π is *compatible* with a column if for each i, there is a j such that $L_i \subseteq \pi_j$, *i.e.* each L_i is inside a unique partition, and does not overlap two partitions. The number i is called the name of the interval L_j. The sequence of names of all intervals is the name of the column E. For a point $\omega \in \hat{E}$, the E-name of the trajectory

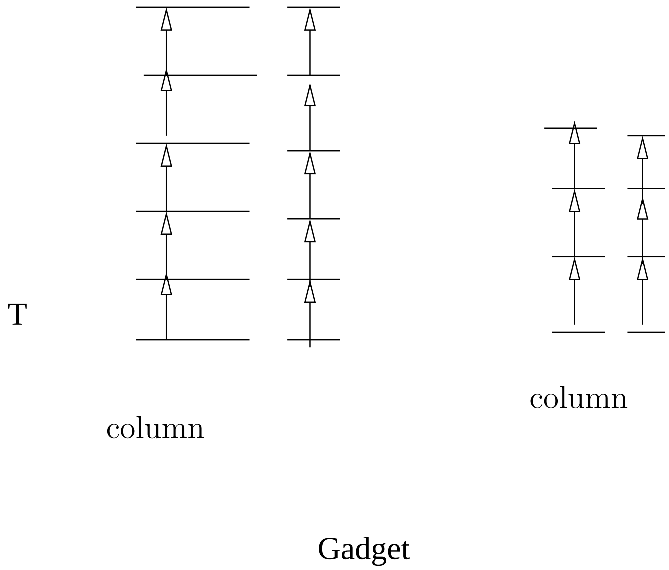
$$\omega, T^{\omega}, T^2\omega, \ldots, T^{h-j}\omega$$

is the sequence of the names of the intervals L_j, $\ldots$, L_h from the column E. The length of the E-name of ω is $h - j + 1$.

A gadget Υ is a finite collection of disjoint columns. The width of the gadget $w(\Upsilon)$ is the sum of the width of its columns. If $\Upsilon = \cup_i \Upsilon_i$ is a union of gadgets with disjoint supports, then its support is the union of the supports of all its constitutent gadgets. A transformation $T(\Upsilon)$ on a gadget Υ is the union of transformations defined on all its columns. The Υ-name of a trajectory belonging to a column E is its E-name.

A gadget Υ extends a column F if the support of Υ extends the support of F, $T(\Upsilon)$ extends $T(F)$ and the partition of Υ extends the partition of F.

Figure 2: Cutting a gadget according to the ratio $(\frac{3}{4}, \frac{1}{4})$.



2.2 Cutting and Stacking

The distribution of a gadget Υ with columns E_1, E_2, …, E_n is a vector of probabilities

$$\left(\frac{w(E_1)}{w(\Upsilon)}, \frac{w(E_2)}{w(\Upsilon)}, \ldots, \frac{w(E_n)}{w(\Upsilon)}\right).$$

A gadget Υ is a copy of a gadget Λ if they have the ame distribution and the corresponding columns have the same names.

A gadget Υ can be cut into M copies Υ_1, Υ_2, …, Υ_M according to a given probability vector $(\gamma_1, \ldots, \gamma_M)$ by cutting each column $E_i = (L_{i,j} \mid 1 \leq j \leq h(E_i))$ into disjoint subcolumns

$$E_{i,m} = (L_{i,j,m} \mid 1 \leq j \leq h(E_i)), \tag{6}$$

where $w(E_{i,m}) = w(L_{i,j,m}) = \gamma_m w(L_{i,j})$, for $1 \leq m \leq M$. The piece $\Upsilon_m = \{E_{i,m} \mid 1 \leq i \leq L\}$ is called the *copy* of the gadget Υ of width γ_m.

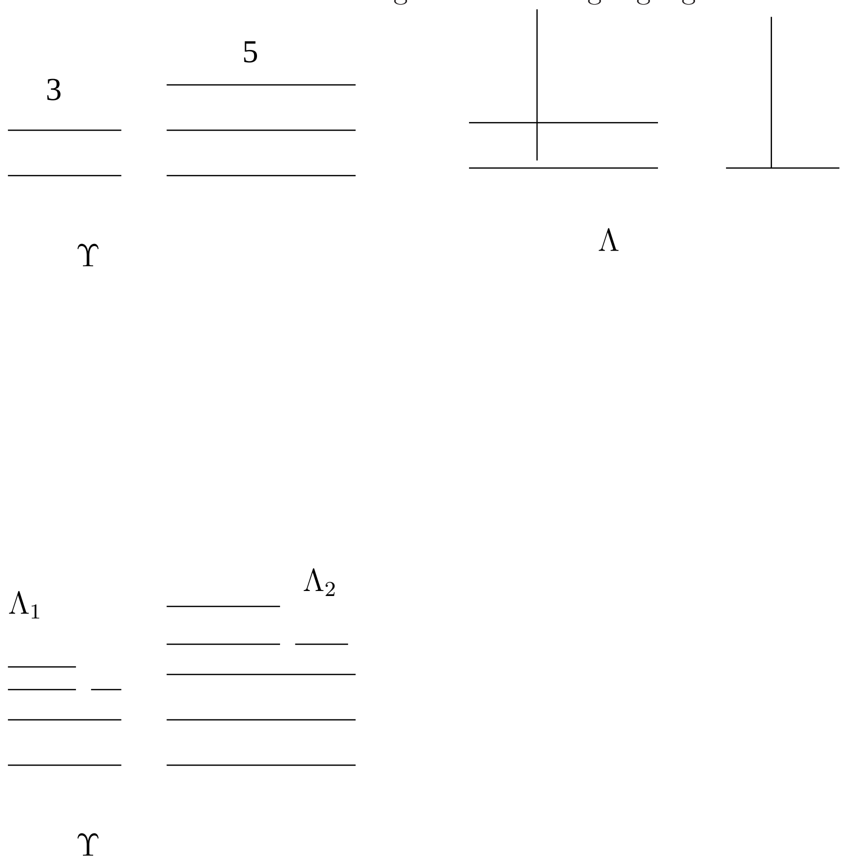
Another basic operation is stacking gadgets onto gadgets. We define this by first looking at the elementary operation of stacking columns onto other columns of equal width.

§a. **Columns onto Columns.** Let $E_1 = (L_{1,j} \mid 1 \leq j \leq h(E_1))$ and $E_2 = (L_{2,j} \mid 1 \leq j \leq h(E_2))$ be two columns of equal width. The stacked column $E_1 * E_2$ is formed by first stacking the columns of E_1 in the same order, and placing the columns of E_2 above them, in the same order as in E_2. Thus $E_1 * E_2$ has height $h(E_1) + h(E_2)$.

§b. **Columns onto Gadgets.** Now we stack columns onto gadgets of equal width. Let a gadget $\Upsilon = (U_1, \ldots, U_m)$ and a column E have the same width. Then cut E into m copies $(E_1, \ldots, E_m)$ where $w(E_i) = w(U_i)$. Now, $E * \Upsilon = (E_1 * U_1, \ldots, E_m * U_m)$.

§c. **Gadgets onto Gadgets.** Let Υ and Λ be two gadgets of equal width, though not necessarily having the same number of columns or columns of the same width. The gadget $\Upsilon * \Lambda$ is defined

Figure 3: Stacking a gadget onto another



as follows. Let the columns of Υ be $(U_1, \ldots, U_k)$. Cut Λ into copies Λ_i such that $w(\Lambda_i) = E_i$ for all i. Then, for each i, stack Λ_i onto E_i. Then

$$\Upsilon * \Lambda = (U_1 * \Lambda_1, U_2 * \Lambda_2, \ldots, U_k * \Lambda_k). \tag{7}$$

The number of columns in the resultant is the product of the number of columns in each gadget.

The M-fold cutting and stacking of the gadget Υ onto itself is defined in a natural way.

§d. **Limit of the construction.** A sequence of gadgets $\langle \Upsilon_n \rangle_{n \in \mathbb{N}}$ is complete if

1. $\lim_{n\to\infty} w(\Upsilon_n) = 0$.
2. $\lim_{n\to\infty} \lambda(\hat{\Upsilon}_m) = 1$.
3. Υ_{n+1} extends Υ_n for $n \in \mathbb{N}$.

Any such complete gadget determines a transformation $T = T\langle \Upsilon_s \rangle$ defined almost everywhere on $[0, 1)$.

§e. **Ergodic Properties of the construction.** The transformation is measure-preserving by construction, if the sequence of gadgets is complete. Sufficient conditions for ergodicity are given below.

Let Υ be constructed from Λ by cutting and stacking. Let E be column from Υ and D from Λ. Then $\hat{E} \cap \hat{D}$ is defined as the union of subcolumns from D of width $w(E)$ used in the construction of E.

Let $0 < \epsilon < 1$. A gadget Λ is $(1 - \epsilon)$ well-distributed in Υ if

$$\sum_{D \in \Lambda} \sum_{E \in \Upsilon} |\lambda(\hat{E} \cap \hat{D}) - \lambda(\hat{E})\lambda(\hat{D})| < \epsilon. \tag{8}$$

Note that if $\hat{E}$ is independent of $\hat{D}$, then the above term will be 0. The expression therefore implies that the subcolumns of D of width $w(E)$ used in the construction of E are almost independent of the columns in E.

The concept of well-distribution gives us a condition that we can ensure in each step of a sequence of gadgets, in order to ensure that in the limit, the transformation we define is ergodic. We have the following consequences if gadgets are well-distributed.

Lemma 7. *Let $\langle \Upsilon_n \rangle_{n \in \mathbb{N}}$ be a complete sequence of gadgets, and, for each n, let Υ_n be well-distributed in Υ_{n+1}. Then $\langle \Upsilon_n \rangle$ defines an ergodic process.*

Lemma 8. *For any $\epsilon > 0$ and any gadget Υ, tere is a number M such that for all $m \geq M$, the gadget is $(1-\epsilon)$-well-distributed in the gadget $\Upsilon^{*(M)}$.*

The above lemmas give an easy way to ensure that a process defined by cutting and stacking is ergodic as well as stationary.

3 Nonrobustness of the property of the universal data compression scheme.

Consider the following analogue of the main theorem, which holds for Kolmogorov Complexity as opposed to Lempel-Ziv compressibility.

Theorem 9. *For any order function $\sigma : \mathbb{N} \to \mathbb{N}$ and any real $0 < \varepsilon < 1/4$, there is a stationary ergodic measure P computable with respect to σ with entropy $0 < H \leq \varepsilon$ such that for each universal code $\{\phi_n\}_{n \in \mathbb{N}}$, there is a sequence $x \in A^\infty$ such that*

1. *$d_P(x[1 \dots n]) \leq \sigma(n)$ - i.e. when σ is a slow-growing order function, x can be made fairly close to incompressible.*
2. *We have*

$$\limsup_{n \to \infty} \frac{K(x[1 \dots n])}{n} \geq \frac{1}{4} \tag{9}$$

$$\liminf_{n \to \infty} \frac{K(x[1 \dots n])}{n} \leq \varepsilon \tag{10}$$

Proof. Let $r > 0$ be a sufficiently small rational number to be determined later. Consider the partition

$$\pi_0 = \left[0, \frac{1}{2}\right) \cup \left(\frac{1}{2} + r, 1\right]$$
$$\pi_1 = \left[\frac{1}{2}, \frac{1}{2} + r\right].$$

We use cutting and stacking to define an ergodic transformation $T : [0, 1) \to [0, 1)$ defining a stationary ergodic measure P on the set A^∞. The definition of P is given by

$$P(a[1 \dots n]) \quad = \quad \lambda\{\omega \mid \omega \in [0, 1), \;\; T^i(\omega) \in \pi_{a[i]}, \;\; 1 \le i \le n\}. \tag{11}$$

Thus the definition of P on A^∞ depends on the uniform measure on $[0, 1)$ and the partition. The measure P can be extended in a natural way to all Borel sets of A^∞.

The ergodic transformation T will be defined by a sequence of gadgets $\langle \Delta_s \rangle_{s \in \mathbb{N}}$ and $\langle \Pi_s \rangle_{s \in \mathbb{N}}$. Let $\Phi_s = \Delta_s \cup \Pi_s$ for $s \in \mathbb{N}$.

The purpose of the construction is to ensure conditions such that there is an infinite trajectory α satisfying the conditions in the theorem.

Since the function σ is non-decreasing and unbounded,a sequence of positive integers exists such that

$$0 < h_{-2} < h_{-1} < h_0 < h_1 < \dots$$

with

$$\sigma(h_{i-1}) - \sigma_(h_{i-2}) \;\; > \;\; -\log r + i + 13.$$

If σ is a slow-growing function, then the h sequence grows rapidly.

§1. **Base Case of the Construction.** Δ_0 is obtained by cutting the interval $[1/2 - r, 1/2 + r)$ into $2h_0$ equal parts and stacking them. Π_0 is obtained by cutting $[0, 1/2 - r) \cup (1/2 - r, 1]$ into $2h_0$ equal parts and stacking them. Note that the measures of the gadgets are $\lambda(\Delta_0) = 2r$ and $\lambda(\Pi_0) = 1 - 2r$.

§2. **Inductive Stages of the Construction.** At step $s - 1$, assume that the gadgets Δ_{s-1} and Π_{s-1} have been defined. Then, cut Δ into two copies Δ' and Δ'' of equal width.

Cut the remaining half Δ' into R_s equal copies, and stack them independently. Call the resulting gadget Δ_s. We choose R_s so that this gadget has a height of $2h_s$. The measure of Δ_s is

$$\lambda(\Delta_s) = \frac{r}{2^{s+1}}.$$

Form the intermediate gadget $\Pi_{s-1} \cup \Delta''$. Cut the intermediate gadget into R_s copies, and stack them independently. The resulting gadget also has height $2h_s$. Call this resulting gadget as Π_s. The measure of Π_s is

$$\lambda(\Pi_s) = 1 - \frac{r}{2^{s+1}}$$

§3. **Properties of the Construction** The consruction ensures the following.

1. The transformation T is defined on a set of probability 1.

2. The measure P is stationary.

3. The measure P is ergodic since

   (a) $\langle \Pi_s \rangle_{s \in \mathbb{N}}$ is complete.

   (b) $\Pi_{s-1} \cup \Delta''$ and hence Π_{s-1} are $(1 - \frac{1}{s})$ well-distributed in Π_s.

§4. **Construction of α** We define prefixes of α in each stage s, dentoted $\alpha(s)$, such that

$$\alpha(0) \prec \alpha(1) \prec \ldots.$$

For all sufficiently large k, $\alpha(k)$ will be a compressible sequence for odd k and incompressible for even k.

Define $\alpha(0)$ be the Π_0-name of some trajectory of length $\geq h_0$ such that $d_P(\alpha(0)) \leq 2$. Such a trajectory exists since incompressible sequences are abundant.[2]

§§4.1 **Inductive Construction of $\alpha(k)$** Assume that $\alpha(0) \prec \ldots \alpha(k-1)$ have been constructed. For some stel $s(k-1)$ in the construction, the word $\alpha(k-1)$ is the $\Pi_{s(k-1)}$ name of a trajectory of some point from the support of the gadget $\Pi_{s(k-1)}$, having length $\geq h_{s(k-1)}$.

Further, if k is odd, then $\alpha(k-1)$ must be incompressible, so we assume that $d_P(\alpha(k-1)) \leq \sigma(h_{s(k-2)}) - 4$. If k is even, the $\alpha(k-1)$ must be compressible, then $d_P(\alpha(k-1)) \leq \sigma(h_{s(k-2)})$. [3] If k is even, then we also assume that

$$P^{s(k-1)}(\alpha(k-1)) > \frac{1}{8} P(\alpha(k-1)).$$ [4]

§§4.2 **Construction of $\alpha(k)$, k odd (compressible prefixes).** Consider all intrervals of Π_{s-1} such that for <u>any</u> trajectory starting in this interval with Π_{s-1}-name extending $\alpha(k-1)$, the frequency of visiting partition π_1 is $\leq 2r$. Then by an estimate using binomial coefficients, we know that for any such trajectory γ,

$$\frac{K(\gamma)}{|\gamma|} \quad \leq \quad -3r \log r \quad \leq \quad \varepsilon$$

if r is sufficiently small. By the Ergodic Theorem, for all sufficiently large s, the total length of all intervals in this set is greater than or equal to $\frac{1}{2} P(\alpha(k-1))$. [5]

Consider an arbitrary column from Π_s. Divide the set of intervals into the lower and upper halves. Consider only intervals in the lower part. Any trajectory starting here has length $\geq h_s$. Fix some s as above and let $s(k) = s$.

Let $U_s(\alpha(k-1))$ be the set of all intervals from the lower half such that trajectories γ starting in them and having Π_s names extending $\alpha(k-1)$ satisfy

$$\frac{K(\gamma)}{|\gamma|} \quad \leq \quad -3r \log r \quad \leq \quad \varepsilon.$$

[2] We will not go into quantitative estimates establishing this statement here, but will assume this.

[3] By construction, $\sigma(h_{s(k-2)})$ is approximately $-\log r + s(k-2) + 13$, hence very small compared to the length of $\alpha(k-1)$.

[4] This is satisfied, for instance when $P(\alpha(k-1))$ is very close to 1.

[5] Because Δ'' is well-distributed in Π_{s-1}?

The inequaluty $P^s(U_s(\alpha(k-1))) > \frac{1}{4}P(\alpha(k-1))$ holds. [6]

Define

$$D = \cup_{\gamma \in U_s(\alpha(k-1))} C_\gamma.$$

Then it is easy to prove that a set $S \subseteq U_s(\alpha(k-1))$ exists[7] such that $P(S) > \frac{1}{8}P(D)$ and $P^s(b) > \frac{1}{8}P(b)$ for all $b \in S$.

Then there is an element $b \in S$ such that $d_P(b \upharpoonright j) \le d_P(a) + 4$ if $|\alpha(k-1)| \le j \le |b|$. Define $\alpha(k) = b$.

By the induction hypothesis,

$$d_P(\alpha(k-1)) \le \sigma(h(s_{k-2})) - 4$$

holds, and $|\alpha(k-1)| \ge h_{s(k-1)} > h(s(k-2))$. Hence,

$$d_P(b \upharpoonright j) \quad \le \quad \sigma(h_{s(k-2)}) \quad \le \quad \sigma(|\alpha(k-1)|) \quad \le \quad \sigma(j)$$

holds for $|\alpha(k-1)| \le j \le |b|$.

§§4.3 **Inductive Construction of α, even k (incompressible)**

First, consider an arbitrary column from the gadget Δ_{s-1}.

The uniform measure of the lower half is $\frac{1}{2}\lambda(\Delta_{s-1})$. Let $L = 2h_{s-1}$ be the height of the gadget Δ_{s-1}. Consider the names $x^{L/2}$ of initial fragments of all trajectories starting in the lower half of Δ_{s-1}. By the incompressibility of sequences, the Bernoulli probability of all strings of length $L/2$ satisfying

$$\frac{K(x\left[0 \ldots \frac{L}{2} - 1\right])}{L/2} \;<\; 1 - \frac{2}{h_{s-2}}$$

is at most

$$\left(2^{\frac{2}{h_{s-2}}}\right)^{L/2} = 2^{-L/h_{s-2}}.$$

Any other string in the same set satisfies

$$\frac{K(x\left[0 \ldots \frac{L}{2} - 1\right])}{L/2} \;\ge\; 1 - \frac{2}{h_{s-2}}. \tag{12}$$

We now argue that this set has sufficient probability.

For any step s of the construction, and any trajectory x, $P_{s-1}(x) = 2^{-|x|}\lambda(\hat{\Delta}_{s-1})$. Hence this set has probability at least $\frac{1}{4}\lambda(\hat{\Delta}_{s-1})$.

Now we look at the proportionate probability of extensions of this trajectory in the next stage s. We have,

$$\gamma = \frac{\lambda(\Delta'')}{\lambda(\Pi_{s-1})} = \frac{\lambda(\Delta)}{2\lambda(\Pi_{s-1})} = \frac{2^{-s+1}r}{1 - 2^{-s+2}} \ge 2^{\sigma(h_{s-1}) - \sigma(h_{s-2}) + 12}.$$

Consider the R_s-fold independent cutting and stacking of the intermediate gadget $\Pi_{s-1} \cup \Delta''$. When we stack, the portion of trajectories of any column from Π_{s-1} which go into the column Δ'' is equal to

[6] Why?
[7] Why?

$$\frac{\lambda(\Delta'')}{\lambda(\Pi_{s-1}) + \lambda(\hat{\Delta}'')} = \frac{\gamma}{1+\gamma}.$$

In stage s, consider the lower half of all subintervals in which trajectories with Π_{s-1}-names extend $\alpha(k-1)$. The length of each such trajectory in stage s is at least h_s.

1. The measure of all the remaining subintervals will reduce by a factor of 2. *(*is this because we are neglecting the upper half?)
2. Consider a subset of these subintervals such that the trajectories starting there go into columns of Δ''. This is $\frac{\gamma}{1+\gamma}$ fraction of the subintervals which remain after step 1.
3. Consider a subset of these which generate trajectories which satisfy (12) What remains has probability at least $\frac{1}{4}$ of that in step 4.

Let D be the set of all Π_s names of all trajectories that remain after these filtering operations. Then

$$P_s(D) \geq \frac{\gamma}{8(1+\gamma)} P_{s-1}(\alpha(k-1)).$$

□