Compressibility and Kolmogorov Complexity

Stephen Binns
Marie Nicholson
Department of Mathematics and Statistics
King Fahd University of Petroleum and Minerals
Dhahran 31261
DRAFT

January 21, 2011

Abstract

We continue the investigation of the path-connected geometry on the Cantor space and the related notions of dilution and compressibility described in [1]. These ideas are closely related to the notions of effective Hausdorff and packing dimensions of reals, and we argue that this geometry provides the natural context in which to study them.

In particular we show that every regular real can be maximally compressed - that is every regular real is a dilution of some real of maximum effective Hausdorff dimension.

1 Introduction

One of the fundamental objects of study in Information and Computability Theory is the set of all infinite binary sequences. It has a similar role in these subjects as the unit interval does in Analysis. This set, elements of which are called *reals*, when equipped with a standard metric,[1] is referred to as the *Cantor space* and we denote it $2^{\mathbb{N}}$. The standard metric provides the set of reals with a topology that is Hausdorff, complete, compact, and totally disconnected – that is any two elements of $2^{\mathbb{N}}$ can be separated with clopen sets. In particular (and important for our purposes) it is far from being path-connected; there is no way under this topology to conceive of one real transforming continuously into another.

[1]The distance between binary sequences X and Y is 2^{-n} where n is the length of the maximal initial segment the have in common. The resulting topology has a basis of clopen sets consisting of all sets of the form $U_\sigma = \{X \in 2^{\mathbb{N}} : X \supset \sigma\}$.

A natural path-connected topology can be given to $2^{\mathbb{N}}$ if every real is considered to be a binary expansion of a real number in the interval $[0, 1]$. This is not a one-to-one identification however as if σ is any finite binary string, then

$$0.\sigma 100000000000\cdots = 0.\sigma 011111111111111\ldots$$

represent the same real number. If this causes technical issues in practice, it is often dealt with by declaring any two such sequences equivalent and applying the topology of $[0, 1]$ onto the resulting set of equivalent classes in $2^{\mathbb{N}}$ (almost all of which will have only one element anyway). Alternatively, one could drop any expectation that the topology be Hausdorff and accept the fact that the two above sequences cannot be separated by disjoint open sets.

In [1] another metric was described that we argue is interesting and relevant to the study of Kolmogorov complexity. It also induces a non-Hausdorff, path-connected topology on $2^{\mathbb{N}}$, and is closely connected to the study of the effective dimensions of reals – specifically their effective Hausdorff and packing dimensions. These dimensions arise naturally as effectivisations of classical notions and have been studied in [5], [4], [6], [2], [7] and elsewhere. Effective dimensions have simple characterisations in terms of Kolmogorov complexity and we will take these characterisations as definitions. The Hausdorff and packing dimensions of a real X are respectively

$$\dim_{\mathcal{H}} X = \liminf_n \frac{C(X \upharpoonright n)}{n} \qquad \dim_p X = \limsup_n \frac{C(X \upharpoonright n)}{n}$$

where $C(X \upharpoonright n)$ is the plain Kolmogorov complexity of the first n bits of X.

The metric from [1] that we will be working with here is technically a pseudometric as it is possible for two distinct reals to be distance 0 from each other. We refer to it as the *d-metric* and introduce it first as a *directed pseudometric*[2]:

$$d(X \to Y) = \limsup_n \frac{C(Y \upharpoonright n \mid X \upharpoonright n)}{n},$$

with the pseudometric given by:

$$d(X, Y) = \max\{d(X \to Y), d(Y \to X)\}.$$

We will continue to stretch terminology somewhat by referring to d as a 'metric'. Two reals X, Y at distance 0 from one another will be considered equivalent – denoted $X \simeq_d Y$.

Along with the topological structure induced on $2^{\mathbb{N}}$ by d, we also have a certain amount of algebraic structure. This is produced by an associative scalar multiplication which was introduced in [1] and represents *dilutions* of the information in a real X. We

[2]The terminology for weakenings of the metric space axioms is inconsistent in the literature. d could also be described as a quasi-pseudometric, a term we avoid. All that is meant here is a function $d : 2^{\mathbb{N}} \times 2^{\mathbb{N}} \to \mathbb{R}^+ \cup \{0\}$ that obeys the triangle inequality in the direction of the arrow.

dilute the information in X by a factor $\alpha \in [0,1]$ by adding strings of 0s at defined positions in X. The result we denote by αX. The lower the value of α, the more X is diluted: if $\alpha = 0$, then $\alpha X = 000000 \cdots := \mathbf{0}$ and if $\alpha = 1$, then $\alpha X = X$. These dilutions cohere well with dimensional and metric properties as, as we show in Section 3,

$$\dim_{\mathcal{H}}(\alpha X) = \alpha \dim_{\mathcal{H}} X \text{ and } \dim_p(\alpha X) = \alpha \dim_p X,$$

and (more generally)

$$d(\alpha X \to \alpha Y) = \alpha d(X \to Y).$$

A distinguished subset of the reals – the so-called *regular reals* – were talked about in [1], [6] and [8] will also be considered in this paper. X is said to be regular if $\dim_{\mathcal{H}}(X) = \dim_p(X)$. In this case we simply refer to $\dim X$. It was proved in [1] that if X is regular, then

$$d(\alpha X, \beta X) = |\alpha - \beta| \dim X.$$

Regularity is preserved under scalar multiplication.

The relationship between scalar multiplication and the metric d allowed us in [1] to introduce and investigate geometric properties such as the *angle* between two reals and the *projection* of one real onto another.

In this paper we extend some of the results in [1] – relaxing requirements of regularity and proving that d induces a path-connected topology on $2^{\mathbb{N}}$. We also show that the Hausdorff and packing dimension functions

$$\dim_{\mathcal{H}} : 2^{\mathbb{N}} \to [0,1] \text{ and } \dim_p : 2^{\mathbb{N}} \to [0,1]$$

are continuous in the d topology.

The main result however is one of compressibility. A real X is an $1/\alpha$-compression of Y if $Y \simeq_d \alpha X$. Or in other words, if Y is d-equivalent to an α-dilution of X. We prove here that every regular real is maximally compressible. That is, for every regular real X there is a (unique) real Y (also regular) of dimension 1 such that $X \simeq_d (\dim X)Y$.

2 Definitions and Notation

We will follow [3] for the basic notation here and we outline this in this section. The Cantor space of reals (infinite binary sequences) is denoted $2^{\mathbb{N}}$, and $2^{<\mathbb{N}}$ is the set of finite binary strings. We conventionally denote elements of $2^{\mathbb{N}}$ by uppercase roman letters X, Y and Z, and elements of $2^{<\mathbb{N}}$ by lowercase greek letters σ, τ, μ, ν and so on. The length of a binary string is denoted $|\sigma|$. If $n \in \mathbb{N}$, then 0^n is the string of n 0s, and we will denote the infinite sequence of 0s by $\mathbf{0}$. If $n \in \mathbb{N}$, then we use $\log n$ to represent the length of the standard binary expansion of n.

$C(\sigma)$ denotes the plain Kolmogorov complexity of the string σ. This will be the only notion of complexity used in this paper, however all the following results regarding dimension apply equally to prefix-free complexity. $C(\sigma_1, \sigma_2, \dots \sigma_n)$ is the complexity of $\bigoplus_{i=1}^n \sigma_i$ and $C(\sigma|\tau)$ is the complexity of σ given τ.

As is standard σ^* refers to the shortest string that describes σ – the shortest string that will produce σ when input into a fixed universal machine. We introduce the new notation $(\sigma|\tau)^*$ to denote the shortest program that will produce σ when input into a fixed universal oracle machine given τ as oracle. Thus,

$$C(\sigma) = |\sigma^*| \text{ and } C(\sigma|\tau) = |(\sigma|\tau)^*|.$$

If σ and τ are two binary strings, then we denote their concatenation by $\sigma\tau$. We will often need to calculate the complexity of a sequence of concatenated strings, and issues of deciding where one string ends and the next starts will arise. To deal with this we let $\overline{\sigma}$ be the string σ with $2\log|\sigma|$ bits appended to the beginning to indicate the length of σ. More precisely, if $b_1 b_2 b_3 \dots b_n$ is the standard binary expansion of $|\sigma|$, then

$$\overline{\sigma} = b_1 0 b_2 0 b_3 0 \dots 0 b_n 1 \sigma.$$

The motivation here is that $C(\overline{\tau_1}\,\overline{\tau_2} \dots \overline{\tau_n}) = C(\tau_1, \tau_2, \dots, \tau_n) \pm \mathcal{O}(1)$.

Definition 2.1. Let X be any real. Then the *effective Hausdorff dimension of* X is

$$\dim_{\mathcal{H}} X = \liminf_n \frac{C(X\restriction n)}{n}.$$

The dual notion

$$\dim_p X = \limsup_n \frac{C(X\restriction n)}{n}$$

is the *effective packing dimension of* X. If these two dimensions are equal, that is, if $\lim_n C(X\restriction n)/n$ exists, then X is said to be a *regular real* and we refer to the *dimension of* X and denote this $\dim X$:

$$\dim X = \lim_n \frac{C(X\restriction n)}{n}.$$

Definition 2.2. Let $X, Y, Z \in 2^{\mathbb{N}}$, define

$$d(X \to Y) = \limsup_n \frac{C(Y\restriction n \mid X\restriction n)}{n}$$

where $C(Y\restriction n|X\restriction n)$ is the Kolmogorov complexity of $Y\restriction n$ given $X\restriction n$. The function d obeys the triangle inequality in the direction of the arrow, that is

$$d(X \to Y) + d(Y \to Z) \geqslant d(X \to Y).$$

See [1] for the proof of this fact and other details.

Notice that under this definition

$$\dim_p(X) = d(\mathbf{0} \to X).$$

Definition 2.3. A metric can be formed from d by defining

$$d(X, Y) = \max\{d(X \to Y), d(Y \to X)\}$$

and by identifying reals that are distance 0 from one another. We write $X \simeq_d Y$ if $d(X, Y) = 0$.

2.1 Symmetry of Information

The next lemma is well-known and will be used extensively. For a proof and historical reference see [3].

Lemma 2.4 (Symmetry of Information (Levin, Kolmogorov))**.** *If $\sigma, \tau \in 2^{<\mathbb{N}}$, then*

$$C(\sigma, \tau) = C(\sigma|\tau) + C(\tau) \pm \mathcal{O} \log C(\sigma, \tau).$$

We will also use the following two consequences:

Corollary 2.5.

$$\liminf_n \frac{C(X\restriction n \mid Y\restriction n)}{n} = \liminf_n \frac{C(X\restriction n, Y\restriction n) - C(Y\restriction n)}{n}$$

and similarly for limit supremum.

Corollary 2.6. *If $\tau_1, \tau_2 \dots, \tau_n \in 2^{<\mathbb{N}}$, then*

$$C(\tau_1, \tau_2 \dots, \tau_n) = C(\tau_1) + C(\tau_2|\tau_1) + C(\tau_3|\tau_1, \tau_2) + \dots \\ + C(\tau_n|\tau_1, \tau_2 \dots, \tau_{n-1}) \pm \mathcal{O}n \log C(\tau_1, \tau_2 \dots \tau_n).$$

Proof. The $\leqslant$ direction is derived by concatenating the required descriptions on the right-hand side along with enough bits to distinguish them. $\mathcal{O}n \log C(\tau_1, \tau_2 \dots \tau_n)$ bits are sufficient for this purpose.

For the $\geqslant$ direction, take a fixed constant k such that for all $\sigma, \tau \in 2^{<\mathbb{N}}$

$$C(\sigma, \tau) \geqslant C(\sigma|\tau) + C(\tau) - k \log C(\sigma, \tau),$$

and then repeatedly apply Lemma 2.4 starting with $C(\tau_n, \bigoplus_{i=1}^{n-1} \tau_i)$.

□

2.2 Scalar Multiplication

Another concept that we will use repeatedly is that of a *dilution*. It is a function from $\mathbb{R} \times 2^{\mathbb{N}}$ to $2^{\mathbb{N}}$ that consists of interpolating 0s into a real and consequently reducing its dimension. This we interpret as a scalar multiplication, and if $\alpha \in [0,1]$ and $X \in 2^{\mathbb{N}}$, we write αX for the dilution of X by a factor α. The effective dimensions of X are scaled by a factor of α as a result. We now give an exact definition.

Notation 2.7. Let $X \in 2^{\mathbb{N}}$, $\alpha \in [0,1]$ and $i \in \mathbb{N}^+$. Let $p_i(\alpha)$ be the least natural number k that minimises $\left| \alpha i - k \right|$. We then have that

$$\alpha i - 1/2 \leqslant p_i(\alpha) \leqslant \alpha i + 1/2$$

and that $\lim_i p_i(\alpha)/i = \alpha$.

Definition 2.8 (Scalar multiplication)**.** If $X \in 2^{\mathbb{N}}$ and $\alpha \in [0,1]$, then we let αX be the real

$$\sigma_1 0^{a_1} \sigma_2 0^{a_2} \sigma_3 0^{a_3} \dots \sigma_i 0^{a_i} \dots$$

where

1. $X = \sigma_1 \sigma_2 \sigma_3 \dots$
2. $|\sigma_i| = p_i(\alpha)$
3. $|\sigma_i 0^{a_i}| = i$.

Notation 2.9. Later we will be considering initial segments of αX, so we introduce some useful notation now. For every n, there exists a unique $m = m(n)$ and $\tau = \tau(n)$ such that

$$\alpha X \restriction n = \sigma_1 0^{a_1} \sigma_2 0^{a_2} \sigma_3 0^{a_3} \dots \sigma_m 0^{a_m} \tau$$

where m is the largest possible integer such that $\sum_{i=1}^{m} i = m(m+1)/2 \leqslant n$. To make the calculations more readable, we let

- $M := m(m+1)/2$ and
- $P_m(\alpha) := \sum_{i=1}^{m} p_i(\alpha)$,

both of which are implicitly functions of n. We will also refer to the string $\sigma_1 \sigma_2 \sigma_3 \dots \sigma_m$ above as *the bits of* X *in* $\alpha X \restriction M$, and to the added 0s as *the padding bits*. This notation will be used throughout the paper.

3 Basic Results

In this section we will prove some basic lemmas giving the relationship between the scalar multiplication and the metric d.

Lemma 3.1. *If $\alpha \in [0,1]$, $X \in 2^{\mathbb{N}}$ and $n \in \mathbb{N}$, and if m and $P_m(\alpha)$ are as in Notation 2.9, then $C(X \restriction P_m(\alpha)) = C(\alpha X \restriction n) \pm \mathcal{O}(m \log m)$.*

Proof. With all the notation as stated, $\alpha X \restriction n$ is of the form

$$\sigma_1 0^{a_1} \sigma_2 0^{a_2} \sigma_3 0^{a_3} \dots \sigma_m 0^{a_m} \tau$$

where $|\tau| < m+1$ and $\sigma_1 \sigma_2 \sigma_3 \dots \sigma_m \tau \subset X$.

Therefore, to describe $\alpha X \restriction n$, it is sufficient to know $X \restriction P_m(\alpha)$, the values of $p_i(\alpha)$ for $i \leqslant m$ and the string τ. Therefore

$$C(\alpha X \restriction n) \leqslant C(X \restriction P_m(\alpha)) + \mathcal{O}(m \log m).$$

Similarily, to describe $X \restriction P_m(\alpha)$, it is sufficient to describe $\alpha X \restriction n$ and to distinguish in $\alpha X \restriction n$ the padding bits from the bits of X. To do this it is enough to know the values of $p_i(\alpha)$ for all $i \leqslant m$. Thus

$$C(X \restriction P_m(\alpha)) \leqslant C(\alpha X \restriction n) + \mathcal{O}(m \log m),$$

and consequently

$$C(X \restriction P_m(\alpha)) = C(\alpha X \restriction n) \pm \mathcal{O}(m \log m).$$

□

Lemma 3.2. *Again using the above notation, $\lim_n \frac{P_m(\alpha)}{n} = \alpha$.*

Proof. Recall that,

$$\alpha i - 1/2 \leqslant p_i(\alpha) \leqslant \alpha i + 1/2,$$

hence, by summing over all $i \leqslant m$,

$$\alpha M - \frac{m}{2} \leqslant P_m(\alpha) \leqslant \alpha M + \frac{m}{2}.$$

Now dividing by n we get

$$\frac{\alpha M}{n} - \frac{m}{2n} \leqslant \frac{P_m(\alpha)}{n} \leqslant \frac{\alpha M}{n} + \frac{m}{2n}.$$

As $\lim_n \frac{M}{n} = 1$ and $m = \mathcal{O}(\sqrt{n})$, the result follows. □

Lemma 3.3. *In the above notation*

$$\limsup_n \frac{C(X\upharpoonright P_m(\alpha))}{P_m(\alpha)} = \limsup_n \frac{C(X\upharpoonright n)}{n},$$

and similarly for $\liminf_n \frac{C(X\upharpoonright P_m(\alpha))}{P_m(\alpha)}$.

Proof. Let $[\cdot] : \mathbb{R} \to \mathbb{N}$ be the nearest integer function. The definition of $P_m(\alpha)$ suggests that $P_m(\alpha)$ should be close to $[\alpha n]$, and indeed it can be shown, using a similar argument to that in Lemma 3.2 that

$$[\alpha n] = P_m(\alpha) \pm \mathcal{O}(m).$$

The fact that $\alpha \leqslant 1$ means that $\{[\alpha n] : n \in \mathbb{N}\} = \mathbb{N}$ and therefore that,

$$\begin{aligned}
\limsup_n \frac{C(X\upharpoonright n)}{n} &= \limsup_n \frac{C(X\upharpoonright [\alpha n])}{[\alpha n]} &(1)\\
&= \limsup_n \frac{C(X\upharpoonright P_m(\alpha)) \pm \mathcal{O}(m)}{P_m(\alpha) \pm \mathcal{O}(m)} &(2)\\
&= \limsup_n \frac{C(X\upharpoonright P_m(\alpha))}{P_m(\alpha)} &(3)\\
& &(4)
\end{aligned}$$

as $P_m(\alpha) = \mathcal{O}(n)$ and $m = \mathcal{O}(\sqrt{n})$. The proof for $\liminf_n \frac{C(X\upharpoonright P_m(\alpha))}{P_m(\alpha)}$ is identical. □

Lemma 3.4. *For all* $X \in 2^{\mathbb{N}}$ *and* $\alpha \in [0, 1]$

$$\dim_{\mathcal{H}}(\alpha X) = \alpha \dim_{\mathcal{H}}(X) \text{ and } \dim_p(\alpha X) = \alpha \dim_p(X).$$

Proof. Let $n \in \mathbb{N}$ and consider $\alpha X\upharpoonright n$. Notice that, in the notation of 2.9, $m(m+1)/2 \leqslant n$ and therefore

$$\lim_n \frac{\mathcal{O}(m \log m)}{n} = 0. \qquad (5)$$

Now,

$$\begin{aligned}
\dim_p(\alpha X) &= \limsup_n \frac{C(\alpha X\upharpoonright n)}{n}\\
&= \limsup_n \frac{C(X\upharpoonright P_m(\alpha)) \pm \mathcal{O}(m \log m)}{n} &\text{by Lemma 3.1}\\
&= \limsup_n \frac{C(X\upharpoonright P_m(\alpha))}{n} &\text{by (5)}\\
&= \lim_n \frac{P_m(\alpha)}{n} \limsup_n \frac{C(X\upharpoonright P_m(\alpha))}{P_m(\alpha)}\\
&= \alpha \dim_p(X) &\text{by Lemmas 3.2 and 3.3.}
\end{aligned}$$

By a similar argument, $\dim_{\mathcal{H}}(\alpha X) = \alpha \dim_{\mathcal{H}}(X)$. Hence if X is regular, $\dim \alpha X = \alpha \dim X$. □

Lemma 3.5. *If $X, Y \in 2^{\mathbb{N}}$ are regular and $\dim(X) = \dim(Y)$, then $d(X \to Y) = d(Y \to X)$.*

Proof. Using the symmetry of information:

$$
\begin{aligned}
d(X \to Y) &= \limsup_n \frac{C(Y\restriction n \mid X\restriction n)}{n} \\
&= \limsup_n \frac{C(Y\restriction n, X\restriction n) - C(X\restriction n)}{n} \\
&= \limsup_n \frac{C(Y\restriction n, X\restriction n)}{n} - \lim_n \frac{C(X\restriction n)}{n} && \text{as X is regular} \\
&= \limsup_n \frac{C(X\restriction n, Y\restriction n)}{n} - \lim_n \frac{C(Y\restriction n)}{n} && \text{as } \dim X = \dim Y \\
&= \limsup_n \frac{C(X\restriction n, Y\restriction n) - C(Y\restriction n)}{n} && \text{as Y is regular} \\
&= \limsup_n \frac{C(Y\restriction n \mid X\restriction n)}{n} \\
&= d(Y \to X).
\end{aligned}
$$

□

In [1] we proved that for reals X, Y with strong regularity properties (namely being *mutually regular*) that $d(\alpha X \to \alpha Y) = \alpha d(X \to Y)$. We significantly improve this here by proving this for all pairs of reals. We need first the following Lemma.

Lemma 3.6. *Let $X, Y \in 2^{\mathbb{N}}$ and $\alpha \in (0, 1]$. If $m = m(n)$ and $P_m(\alpha)$ are as above in 2.9, then*

$$\limsup_n \frac{C(Y\restriction P_m(\alpha) \mid X\restriction P_m(\alpha))}{P_m(\alpha)} = \limsup_k \frac{C(Y\restriction k \mid X\restriction k)}{k}.$$

Proof. Just from the definition of the limit supremum, we get

$$\limsup_n \frac{C(Y\restriction P_m(\alpha) \mid X\restriction P_m(\alpha))}{P_m(\alpha)} \leqslant \limsup_k \frac{C(Y\restriction k \mid X\restriction k)}{k}.$$

For the other direction, temporarily fix $k \in \mathbb{N}$ and let n be the largest positive integer such that $P_m(\alpha) = P_{m(n)}(\alpha)$ does not exceed k. Thus $k < P_{m(n+1)}(\alpha) \leqslant P_{m(n)}(\alpha) + m + 1$ and $k \leqslant P_m(\alpha) + m$. Symmetry of information then gives

$$
\begin{aligned}
C(Y\restriction P_m(\alpha) \mid X\restriction P_m(\alpha)) = C(Y\restriction P_m(\alpha), X\restriction P_m(\alpha)) - C(X\restriction P_m(\alpha)) \\
\pm \mathcal{O} \log C(Y\restriction P_m(\alpha), X\restriction P_m(\alpha)).
\end{aligned}
$$

But because $k \leqslant P_m(\alpha) + m$,

$$C(Y \upharpoonright P_m(\alpha), X \upharpoonright P_m(\alpha)) + \mathcal{O}(m) \geqslant C(Y \upharpoonright k, X \upharpoonright k)$$

and

$$C(X \upharpoonright P_m(\alpha)) \leqslant C(X \upharpoonright k) + \mathcal{O} \log m.$$

Therefore

$$\begin{aligned} C(Y \upharpoonright P_m(\alpha), X \upharpoonright P_m(\alpha)) - C(X \upharpoonright P_m(\alpha)) + \mathcal{O} \log C(Y \upharpoonright P_m(\alpha), X \upharpoonright P_m(\alpha)) \geqslant \\ C(Y \upharpoonright k, X \upharpoonright k) - \mathcal{O}(m) - C(X \upharpoonright k) - \mathcal{O} \log m. \end{aligned}$$

Giving

$$C(Y \upharpoonright P_m(\alpha), X \upharpoonright P_m(\alpha)) - C(X \upharpoonright P_m(\alpha)) \geqslant C(Y \upharpoonright k, X \upharpoonright k) - C(X \upharpoonright k) - \mathcal{O}(m).$$

From here we can get

$$C(Y \upharpoonright P_m(\alpha) \bigm| X \upharpoonright P_m(\alpha)) \geqslant C(Y \upharpoonright k \bigm| X \upharpoonright k) - \mathcal{O}(m),$$

and the result follows. □

Theorem 3.7. *If $X, Y \in 2^{\mathbb{N}}$ and $\alpha \in [0,1]$, then $d(\alpha X \to \alpha Y) = \alpha d(X \to Y)$.*

Proof. If $\alpha = 0$, then the result is immediate. Assume $\alpha > 0$. Again using the symmetry of information:

$$\begin{aligned} d(\alpha X \to \alpha Y) &= \limsup_n \frac{C(\alpha Y \upharpoonright n \bigm| \alpha X \upharpoonright n)}{n} \\ &= \limsup_n \frac{C(\alpha Y \upharpoonright n, \alpha X \upharpoonright n) - C(\alpha X \upharpoonright n)}{n} \\ &= \limsup_n \frac{C(Y \upharpoonright P_m(\alpha), X \upharpoonright P_m(\alpha)) - C(X \upharpoonright P_m(\alpha)) \pm \mathcal{O}(m \log m)}{n}, \end{aligned}$$

using an argument similar to Lemma 3.1. Since $\lim_n \frac{\mathcal{O}(m \log m)}{n} = 0$, and using Lemma 2.5,

$$\begin{aligned} d(\alpha X \to \alpha Y) &= \limsup_n \frac{C(Y \upharpoonright P_m(\alpha) \bigm| X \upharpoonright P_m(\alpha))}{n} \\ &= \limsup_n \frac{P_m(\alpha)}{n} \cdot \frac{C(Y \upharpoonright P_m(\alpha) \bigm| X \upharpoonright P_m(\alpha))}{P_m(\alpha)} \\ &= \alpha \limsup_n \frac{C(Y \upharpoonright P_m(\alpha) \bigm| X \upharpoonright P_m(\alpha))}{P_m(\alpha)} \\ &= \alpha d(X \to Y) && \text{by Lemma 3.6} \end{aligned}$$

□

4 Topological Results

4.1 Path connectedness

In [1] it was shown that for any regular real X, $d(\alpha X \to \beta X) = |\alpha - \beta| \dim X$, and hence that the map $\alpha \mapsto \alpha X$ is continuous. Thus every regular real is path connected to the point $\mathbf{0}$ and, by concatenation of paths, to each other. We extend this to the set of all reals – proving not that $d(\alpha X \to \beta X) = |\alpha - \beta| \dim X$ for nonregular X, but merely that $d(\alpha X \to \beta X) \leqslant |\beta - \alpha|$. This is still sufficient to imply continuity but does not require that X be regular.

Theorem 4.1. *$2^{\mathbb{N}}$ is path connected under the metric d.*

Proof. Let X be any (possibly irregular) element of $2^{\mathbb{N}}$ and let $\alpha < \beta$ be elements of $[0, 1]$. Consider $d(\beta X \to \alpha X)$. For any $n \in \mathbb{N}$, $\beta X \restriction n$ will contain at least as many bits of X as $\alpha X \restriction n$ because $\alpha < \beta$. So to describe $\alpha X \restriction n$ given $\beta X \restriction n$ it is sufficient to know the values of $p_i(\alpha)$ and $p_i(\beta)$ for all $i \leqslant m$ as in Notation 2.9. As before this requires at most $\mathcal{O}(m \log m)$ bits. This term disappears in the limit, so $d(\beta X \to \alpha X) = 0$.

Now, for the other direction, let $X[n, m] = \langle X(n), X(n + 1), \ldots, X(m - 1)\rangle$ and consider $d(\alpha X \to \beta X)$. Then

$$\begin{aligned} d(\alpha X \to \beta X) &= \limsup_n \frac{C(\beta X \restriction n \mid \alpha X \restriction n)}{n} \\ &\leqslant \limsup_n \frac{C(X[P_m(\alpha), P_m(\beta)]) + \mathcal{O}m \log m}{n} \end{aligned}$$

because to describe $\beta X \restriction n$ it is sufficient to describe $X \restriction P_m(\alpha)$, $X[P_m(\alpha), P_m(\beta)]$ and strings of padding bits given by a description of length $\mathcal{O}m \log m$. But

$$C(X[P_m(\alpha), P_m(\beta)]) \leqslant P_m(\beta) - P_m(\alpha) + \mathcal{O}(1)$$

so

$$\begin{aligned} d(\alpha X \to \beta X) &\leqslant \limsup_n \frac{P_m(\beta) - P_m(\alpha) + \mathcal{O}m \log m}{n} \\ &= \beta - \alpha, \quad \text{by Lemma 3.2.} \end{aligned}$$

□

4.2 Continuity Theorems

Theorem 4.2. *The functions*

$$\dim_p : 2^{\mathbb{N}} \longrightarrow [0, 1]$$

and

$$\dim_{\mathcal{H}} : 2^{\mathbb{N}} \longrightarrow [0,1]$$

are continuous under the d metric.

Proof. For the first, let $X, Y \in 2^{\mathbb{N}}$. Then

$$C(Y \restriction n) \leqslant C(Y \restriction n \bigm| X \restriction n) + C(X \restriction n) + \mathcal{O}(\log n).$$

And so

$$\limsup_n \frac{C(Y \restriction n)}{n} \leqslant \limsup_n \frac{C(Y \restriction n \bigm| X \restriction n)}{n} + \limsup_n \frac{C(X \restriction n)}{n},$$

which means

$$\dim_p(Y) - \dim_p(X) \leqslant d(X,Y)$$

from which the continuity of $\dim_p$ follows.

For $\dim_{\mathcal{H}}$ the situation is only slightly more complicated. Fix $X \in 2^{\mathbb{N}}$ and $\epsilon > 0$. We show there is a δ such that for all $Y \in 2^{\mathbb{N}}$

$$d(X,Y) < \delta \rightarrow |\dim_{\mathcal{H}} X - \dim_{\mathcal{H}} Y| < \epsilon,$$

namely $\delta = \epsilon/4$. If Y is as above and if $\dim_{\mathcal{H}} X = \alpha$ and $\dim_{\mathcal{H}} Y = \beta$, then we can find an $N \in \mathbb{N}$ such that for all $n \geqslant N$

1. $\dfrac{C(X \restriction n \bigm| Y \restriction n)}{n} < \dfrac{\epsilon}{4}$
2. $\dfrac{C(X \restriction n)}{n} > \alpha - \dfrac{\epsilon}{4}$
3. $\dfrac{3 \log n}{n} < \dfrac{\epsilon}{4}$

as $d(Y \to X) = \limsup_n \frac{C(X \restriction n | Y \restriction n)}{n} < \delta = \epsilon/4$ and $\dim_{\mathcal{H}} X = \liminf_n \frac{C(X \restriction n)}{n} = \alpha$. We can also find an $m \geqslant N$ such that

$$\frac{C(Y \restriction m)}{m} < \beta + \frac{\epsilon}{4}$$

as $\liminf_n \frac{C(Y \restriction n)}{n} = \dim_{\mathcal{H}} Y = \beta$. Fix such an m and now using basic theory:

$$C(X \restriction m) \leqslant C(X \restriction m \bigm| Y \restriction m) + C(Y \restriction m) + 3 \log m$$

and so

$$\frac{C(X \restriction m)}{m} \leqslant \frac{C(X \restriction m \bigm| Y \restriction m)}{m} + \frac{C(Y \restriction m)}{m} + \frac{3 \log m}{m},$$

giving, using the bounds above, that $\alpha - \beta < \epsilon$. A symmetrical argument shows that $\beta - \alpha < \epsilon$ and the result follows. $\square$

Corollary 4.3. *The set of regular reals is closed in the topology induced by the d metric.*

Proof. The irregularity function irreg : $2^{\mathbb{N}} \to [0,1]$ defined by $\text{irreg}(X) = \dim_p X - \dim_{\mathcal{H}} X$ is continuous by Theorem 4.2 and thus the set of regular reals $\{X \in 2^{\mathbb{N}} : \text{irreg(X)} = 0\}$ is closed. □

Corollary 4.4. *If $X \simeq_d Y$, then $\dim_{\mathcal{H}} X = \dim_{\mathcal{H}} Y$ and $\dim_p X = \dim_p Y$.*

Proof. If $d(X,Y) < \delta$ for all $\delta > 0$, then $\dim_p X - \dim_p Y < \epsilon$ for all $\epsilon > 0$ and similarly for $\dim_{\mathcal{H}} X - \dim_{\mathcal{H}} Y$. □

5 Compressibility of Reals

Theorem 5.1. *Let $X \in 2^{\mathbb{N}}$ be regular and of dimension α. Then there exists a regular $Y \in 2^{\mathbb{N}}$ of dimension 1 such that $X \simeq_d \alpha Y$. Furthermore, if $\alpha > 0$, then Y is unique up to d-equivalence.*

Proof. If $\alpha = 0$, the result is immediate, so we suppose that $\alpha > 0$.

Let $X \in 2^{\mathbb{N}}$ be regular. We will divide X into finite blocks of strings:

$$X = \tau_1\tau_2\tau_3\dots$$

where $|\tau_i| = i$ for all i.

Let $\gamma_1 = \tau_1^*$ and $\gamma_{i+1} = (\tau_{i+1}|\tau_1, \tau_2, \dots \tau_i)^*$ where, as before, $(\tau_{i+1}|\tau_1, \tau_2, \dots \tau_i)^*$ is the shortest program that will output τ_{i+1} using $\tau_1, \tau_2, \dots, \tau_i$ as oracles. We now define Y to be

$$Y = \overline{\gamma_1}\,\overline{\gamma_2}\,\overline{\gamma_3}\dots$$

We will find it convenient to use the notations

$$X_i = \tau_1\tau_2\dots\tau_i$$

and

$$Y_i = \overline{\gamma_1}\,\overline{\gamma_2}\dots\overline{\gamma_i}$$

Notice that $C(X_i) = C(\tau_1, \tau_2, \dots, \tau_i) \pm \mathcal{O}(1)$ and we can recover the strings $\tau_1, \tau_2, \dots, \tau_i$ uniformly from the string Y_i. So

$$C(X_i \,\big|\, Y_i) = \mathcal{O}(1). \tag{6}$$

We will now establish some technical lemmas. First, we note that we can bound $|Y_i|$ in terms of the complexity of X_i.

Lemma 5.2. $|Y_i| = C(X_i) \pm \mathcal{O}i \log C(X_i)$.

Proof. Using Corollary 2.6:

$$\begin{aligned}
C(X_i) &= C(\tau_1, \tau_2, \ldots, \tau_i) \\
&\geqslant C(\tau_1) + C(\tau_2 \mid \tau_1) + C(\tau_3 \mid \tau_1, \tau_2) + \cdots + C(\tau_i \mid \tau_1, \tau_2, \ldots, \tau_{i-1}) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad - \mathcal{O} i \log C(\tau_1, \tau_2, \ldots, \tau_i) \\
&\geqslant |\gamma_1| + |\gamma_2| + \cdots + |\gamma_i| - \mathcal{O} i \log C(\tau_1, \tau_2, \ldots, \tau_i) \\
&= |Y_i| - \mathcal{O} \sum_{j=1}^{i} \log |\gamma_j| - \mathcal{O} i \log C(\tau_1, \tau_2, \ldots, \tau_i) \\
&\geqslant |Y_i| - \mathcal{O} i \log C(\tau_1, \tau_2, \ldots, \tau_i) \\
&= |Y_i| - \mathcal{O} i \log C(X_i)
\end{aligned}$$

we obtain the inequality

$$|Y_i| \leqslant C(X_i) + \mathcal{O} i \log C(X_i).$$

A similar argument will give us a lower bound on $|Y_i|$:

$$|Y_i| \geqslant C(X_i) - \mathcal{O} i \log C(X_i).$$

□

Next, we show that if X is regular, then $\lim_i \frac{|Y_i|}{|X_i|} = \dim(X)$. We note here that Lemma 5.3, as stated, also applies to irregular X. This extra generalisation will be useful later.

Lemma 5.3. *Let* $X \in 2^{\mathbb{N}}$, *then* $\limsup_i \frac{|Y_i|}{|X_i|} \leqslant \dim_p(X)$ *and* $\liminf_i \frac{|Y_i|}{|X_i|} \geqslant \dim_{\mathcal{H}}(X)$.

Proof. By Lemma 5.2, $|Y_i| = C(X_i) \pm \mathcal{O} i \log C(X_i)$. Dividing both sides by $|X_i|$ and taking the limit supremum gives

$$\begin{aligned}
\limsup_i \frac{|Y_i|}{|X_i|} &\leqslant \limsup_i \left(\frac{C(X_i)}{|X_i|} + \frac{\mathcal{O} i \log C(X_i)}{|X_i|} \right) \\
&\leqslant \limsup_i \frac{C(X_i)}{|X_i|} + \limsup_i \frac{\mathcal{O} i \log C(X_i)}{|X_i|} \\
&= \limsup_i \frac{C(X_i)}{|X_i|} \qquad\qquad (\text{as } |X_i| = \mathcal{O}(i^2)) \\
&\leqslant \dim_p(X).
\end{aligned}$$

Similarily,

$$\begin{aligned}\liminf_i \frac{|Y_i|}{|X_i|} &\geqslant \liminf_i \left(\frac{C(X_i)}{|X_i|} - \frac{\mathcal{O} i \log C(X_i)}{|X_i|}\right) \\ &\geqslant \liminf_i \frac{C(X_i)}{|X_i|} + \liminf_i -\frac{\mathcal{O} i \log C(X_i)}{|X_i|} \\ &= \liminf_i \frac{C(X_i)}{|X_i|} \\ &\geqslant \dim_{\mathcal{H}}(X).\end{aligned}$$

□

Every initial segment of Y is of the form $Y_i\mu$ for some unique i and μ an initial segment of $\overline{\gamma_{i+1}}$. In the next two lemmas we will calculate a lower bound on the complexity of $Y_i\mu$.

Lemma 5.4. $C(\overline{\gamma_{i+1}}|Y_i) \geqslant |\overline{\gamma_{i+1}}| - \mathcal{O}\log|\gamma_{i+1}| - C(Y_i) + C(X_i) - \mathcal{O}\log C(Y_i)$.

Proof. To calculate a bound on $C(\overline{\gamma_{i+1}}|Y_i)$, we begin by noting that

$$\begin{aligned}C(\overline{\gamma_{i+1}} \mid Y_i) + C(Y_i \mid X_i) + \mathcal{O}\log C(Y_i \mid X_i) &\geqslant C(\tau_{i+1} \mid \tau_1, \tau_2, \ldots, \tau_i) \\ &= |\gamma_{i+1}|\end{aligned}$$

as, from a machine that computes $\overline{\gamma_{i+1}}$ given Y_i and a machine that computes Y_i given X_i (and enough bits to tell them apart), we can easily construct a machine that, given $\tau_1, \tau_2, \ldots, \tau_i$, computes Y_{i+1} and hence τ_{i+1}. Thus

$$C(\overline{\gamma_{i+1}} \mid Y_i) \geqslant |\gamma_{i+1}| - C(Y_i \mid X_i) - \mathcal{O}\log C(Y_i \mid X_i).$$

But we can also put a bound on the second term on the right-hand side, using Symmetry of Information and the fact that $C(Y_i, X_i) = C(Y_i) \pm \mathcal{O}(1)$.

$$\begin{aligned}C(Y_i \mid X_i) &= C(Y_i, X_i) - C(X_i) \pm \mathcal{O}\log C(Y_i, X_i) \\ &= C(Y_i) - C(X_i) \pm \mathcal{O}\log C(Y_i).\end{aligned}$$

And combining this with the previous equation gives

$$C(\overline{\gamma_{i+1}} \mid Y_i) \geqslant |\gamma_{i+1}| - C(Y_i) + C(X_i) - \mathcal{O}\log C(Y_i)$$

or equivalently,

$$C(\overline{\gamma_{i+1}} \mid Y_i) \geqslant |\overline{\gamma_{i+1}}| - \mathcal{O}\log|\gamma_{i+1}| - C(Y_i) + C(X_i) - \mathcal{O}\log C(Y_i),$$

as required. □

Lemma 5.5. $C(Y_i\mu) \geqslant |Y_i\mu| - \mathcal{O}\log|\gamma_{i+1}| - \mathcal{O}i\log C(X_i) - \mathcal{O}\log C(Y_i\mu)$

Proof. Let $\overline{\gamma_{i+1}} = \mu\nu$ then,

$$\begin{aligned} C(\overline{\gamma_{i+1}} \mid Y_i) &= C(\mu\nu \mid Y_i) \\ &\leqslant |\overline{(\mu|Y_i)^*}\nu| + \mathcal{O}(1) \\ &= C(\mu \mid Y_i) + \mathcal{O}\log C(\mu \mid Y_i) + |\nu|. \end{aligned}$$

Therefore,

$$\begin{aligned} C(\mu \mid Y_i) &\geqslant C(\overline{\gamma_{i+1}} \mid Y_i) - |\nu| - \mathcal{O}\log C(\mu \mid Y_i) \\ &\geqslant |\overline{\gamma_{i+1}}| - \mathcal{O}\log|\gamma_{i+1}| - C(Y_i) + C(X_i) \\ &\qquad - \mathcal{O}\log C(Y_i) - |\nu| - \mathcal{O}\log C(\mu \mid Y_i) \qquad \text{from Lemma (5.4)} \\ &= |\mu| - \mathcal{O}\log|\gamma_{i+1}| - C(Y_i) + C(X_i) - \mathcal{O}\log C(Y_i) - \mathcal{O}\log C(\mu|Y_i) \end{aligned}$$

But then the Symmetry of Information again gives us

$$\begin{aligned} C(Y_i\mu) &= C(Y_i, \mu) \\ &\geqslant C(\mu \mid Y_i) + C(Y_i) - \mathcal{O}\log C(Y_i\mu) \\ &\geqslant |\mu| - \mathcal{O}\log|\gamma_{i+1}| - C(Y_i) + C(X_i) \\ &\qquad - \mathcal{O}\log C(Y_i) - \mathcal{O}\log C(\mu \mid Y_i) + C(Y_i) - \mathcal{O}\log C(Y_i\mu) \\ &= |\mu| - \mathcal{O}\log|\gamma_{i+1}| + C(X_i) - \mathcal{O}\log C(Y_i\mu) \\ &\qquad \text{(using the fact that } C(\mu \mid Y_i), C(Y_i) \leqslant C(Y_i\mu) + \mathcal{O}(1)) \\ &\geqslant |\mu| - \mathcal{O}\log|\gamma_{i+1}| + |Y_i| - \mathcal{O}i\log C(X_i) - \mathcal{O}\log C(Y_i\mu) \qquad \text{from Lemma (5.2)} \\ &\geqslant |Y_i\mu| - \mathcal{O}\log|\gamma_{i+1}| - \mathcal{O}i\log C(X_i) - \mathcal{O}\log C(Y_i\mu) \end{aligned}$$

□

Now that we have completed these preparatory calculations, we can complete the proof. We begin by proving that $\dim Y = 1$. To do this it is enough to prove that $\liminf_n \frac{C(Y\upharpoonright n)}{n} = 1$. Every initial segment of Y is of the form $Y_i\mu$ for some unique i and μ an initial segment of $\overline{\gamma_{i+1}}$. Accordingly, we will calculate a lower bound of $\frac{C(Y_i\mu)}{|Y_i\mu|}$ in terms of i and show that this lower bound approaches 1 as i goes to infinity.

By Lemma 5.5 we have a lower bound on $C(Y_i\mu)$. Dividing both sides by $|Y_i\mu|$ gives

$$\frac{C(Y_i\mu)}{|Y_i\mu|} \geqslant 1 - \frac{\mathcal{O}\log|\gamma_{i+1}|}{|Y_i\mu|} - \frac{\mathcal{O}i\log C(X_i)}{|Y_i\mu|} - \frac{\mathcal{O}\log C(Y_i\mu)}{|Y_i\mu|}.$$

Hence,

$$\liminf_n \frac{C(Y_i\mu)}{|Y_i\mu|} \geqslant 1 - \limsup_n \frac{\mathcal{O}\log|\gamma_{i+1}|}{|Y_i\mu|} - \limsup_n \frac{\mathcal{O}i\log C(X_i)}{|Y_i\mu|} - \limsup_n \frac{\mathcal{O}\log C(Y_i\mu)}{|Y_i\mu|}.$$

We now show that each term, except the first, on the righthand side of the inequality vanishes as n (and hence i) goes to infinity. This implies that the dimension of Y exists and is equal to 1. This is easy to see in the case of the last term, as $C(Y_i\mu) \leqslant |Y_i\mu| + \mathcal{O}(1)$. In the case of the second term on the right hand side, notice that

$$|\gamma_{i+1}| = C(\tau_{i+1} \mid \tau_1, \tau_2, \ldots, \tau_i) \leqslant C(\tau_{i+1}) \leqslant |\tau_{i+1}| + \mathcal{O}(1) = i + 1 + \mathcal{O}(1).$$

This means

$$\frac{\log|\gamma_{i+1}|}{|Y_i\mu|} \leqslant \frac{\log(i+\mathcal{O}(1))}{|Y_i|} \leqslant \frac{\log(i+\mathcal{O}(1))}{i} \longrightarrow 0 \text{ as } i \longrightarrow \infty.$$

Lastly,

$$\begin{aligned} \frac{i\log C(X_i)}{|Y_i\mu|} &\leqslant \frac{i\log C(X_i)}{|Y_i|} \\ &\leqslant \frac{i\log C(X_i)}{C(X_i) - \mathcal{O}i\log C(X_i)} \quad \text{from Lemma (5.2)} \end{aligned}$$

So, it is enough to show that $\limsup_i \frac{i\log C(X_i)}{C(X_i)} = 0$. Now, because $\dim_{\mathcal{H}} X = \alpha$, there is a function ϵ bounded by 1, such that $\limsup_n \epsilon(n) = 0$ and for all $n \in \mathbb{N}$

$$\frac{C(X\restriction n)}{n} = \alpha(1-\epsilon(n)). \tag{7}$$

Hence

$$\begin{aligned} \limsup_i \frac{i\log C(X_i)}{C(X_i)} &= \limsup_i \frac{i\log\alpha(1-\epsilon(|X_i|))|X_i|}{\alpha(1-\epsilon(|X_i|))|X_i|} \\ &\leqslant \limsup_i \frac{i\log\alpha(1-\epsilon(|X_i|))}{\alpha(1-\epsilon(|X_i|))|X_i|} + \limsup_i \frac{i\log|X_i|}{\alpha(1-\epsilon(|X_i|))|X_i|}. \end{aligned}$$

But the first term on the righthand side gives:

$$\begin{aligned} \limsup_i \frac{i\log\alpha(1-\epsilon(|X_i|))}{\alpha(1-\epsilon(|X_i|))|X_i|} &\leqslant \limsup_i \frac{i}{\alpha|X_i|} \cdot \limsup_i \frac{\log\alpha(1-\epsilon(|X_i|)}{1-\epsilon(|X_i|)} \\ &= 0, \end{aligned}$$

and similarly for the second term:

$$\limsup_i \frac{i\log|X_i|}{(1-\epsilon(|X_i|))|X_i|} \leqslant \limsup_i \frac{i\log|X_i|}{|X_i|} \cdot \limsup_i \frac{1}{(1-\epsilon(|X_i|))}$$

$$= 0.$$

(Using the fact that $|X_i| = \mathcal{O}(i^2)$ and $\limsup_n \epsilon(n) = 0$.)

We now show that $X \simeq_d \alpha Y$. Notice that up to this point in the proof we have not used the fact that X is regular, and we have only assumed in (7) that $\dim_{\mathcal{H}} X = \alpha$. This allows us to easily generalise the result in the next section. We must now however make use of regularity. We will first show that $d(\alpha Y \to X) \leqslant \dim_p X - \alpha$ and then that Lemma 3.5 implies $X \simeq_d \alpha Y$ as X and αY are both regular.

For every n there exists a unique m and μ such that $X \restriction n = X_m \mu$ where, as before, m is the largest possible integer such that $\frac{m(m+1)}{2} \leqslant n$. Thus

$$\begin{aligned}
d(\alpha Y \to X) &= \limsup_n \frac{C(X\restriction n \mid \alpha Y \restriction n)}{n} \\
&\leqslant \limsup_n \frac{C(X_m \mu \mid Y \restriction P_m(\alpha)) + \mathcal{O}(m\log m)}{n} \\
&\leqslant \limsup_n \frac{C(X_m \mu \mid Y_m) + C(Y_m \mid Y \restriction P_m(\alpha)) + \mathcal{O}(m\log m)}{n} \\
&\quad \text{(using } \mathcal{O}(\log m) \text{ bits to distinguish between the programs for the first two terms)} \\
&\leqslant \limsup_n \frac{C(X_m \mid Y_m) + |\mu| + C(Y_m \mid Y \restriction P_m(\alpha)) + \mathcal{O}(m\log m)}{n} \\
&= \limsup_n \frac{C(X_m \mid Y_m) + C(Y_m \mid Y \restriction P_m(\alpha)) + \mathcal{O}(m\log m)}{n}, \text{ since } |\mu| \leqslant m \\
&\leqslant \limsup_n \frac{C(X_m \mid Y_m)}{n} + \limsup_n \frac{C(Y_m \mid Y \restriction P_m(\alpha))}{n} + \limsup_n \frac{\mathcal{O}(m\log m)}{n}.
\end{aligned}$$

Now, since

$$\lim_n \frac{C(X_m \mid Y_m)}{n} = 0 \quad \text{by (6)}$$

and

$$\lim_n \frac{\mathcal{O}(m\log m)}{n} = 0,$$

it remains to show that

$$\limsup_n \frac{C(Y_m \mid Y \restriction P_m(\alpha))}{n} = 0.$$

Now if m is such that $P_m(\alpha) \geqslant |Y_m|$, then

$$C(Y_m \mid Y \upharpoonright P_m(\alpha)) = \mathcal{O}(\log(P_m(\alpha) - |Y_m|)).$$

From the second half of Lemma 5.3,

$$\liminf_m \frac{|Y_m|}{|X_m|} \geqslant \alpha,$$

but $|X_m| = \frac{m(m+1)}{2}$ so for all $\epsilon > 0$ and sufficiently large m

$$|Y_m| \geqslant (\alpha - \epsilon)\frac{m(m+1)}{2}.$$

From the proof of Lemma 3.2

$$P_m(\alpha) \leqslant \alpha M + \frac{m}{2} = \alpha\frac{m(m+1)}{2} + \frac{m}{2}.$$

So

$$P_m(\alpha) - |Y_m| \leqslant \epsilon\frac{m(m+1)}{2} + \frac{m}{2},$$

and therefore

$$C(Y_m \mid Y \upharpoonright P_m(\alpha)) \leqslant \mathcal{O}(\log(m^2)) = \mathcal{O}(\log m).$$

Therefore the limit over all such m of $\frac{C(Y_m|Y\upharpoonright P_m(\alpha))}{n}$ is 0.

On the other hand, if m is such that $|Y_m| \geqslant P_m(\alpha)$,

$$\begin{aligned}
\limsup_n \frac{C(Y_m \mid Y \upharpoonright P_m(\alpha))}{n} &\leqslant \limsup_n \frac{|Y_m| - P_m(\alpha) + \mathcal{O}(1)}{n} \\
&= \limsup_n \frac{|Y_m|}{n} - \alpha \text{ by Lemma 3.2} \\
&\leqslant \limsup_n \frac{|Y_m|}{|X_m|} \cdot \limsup \frac{|X_m|}{n} - \alpha \\
&\leqslant \dim_p X - \alpha. \qquad (8)
\end{aligned}$$

The final line following from Lemma 5.3 and the fact that $|X_m| \leqslant n$. The result now follows as X is regular.

To complete the proof we must show that Y is unique up to d-equivalence. Suppose there exist Y_1, $Y_2 \in 2^{\mathbb{N}}$, both of dimension 1, such that $X \simeq_d \alpha Y_1$ and $X \simeq_d \alpha Y_2$. Then $d(\alpha Y_1, \alpha Y_2) \leqslant d(\alpha Y_1, X) + d(X, \alpha Y_2) = 0$ and hence $\alpha Y_1 \simeq_d \alpha Y_2$. We know, $d(\alpha Y_1, \alpha Y_2) = \alpha d(Y_1, Y_2)$ by Lemma 3.7, and so $d(Y_1, Y_2) = 0$ whenever $d(\alpha Y_1, \alpha Y_2) = 0$. So we have $Y_1 \simeq_d Y_2$. Therefore Y is unique up to d-equivalence. □

5.1 Generalisation to Irregular Reals

Theorem 5.6. *If $X \in 2^{\mathbb{N}}$ and Y is as defined in Theorem 5.1, then if $\dim_{\mathcal{H}} X = \alpha \leqslant \beta = \dim_p X$,*

$$d(X, \alpha Y) = \beta - \alpha = \mathrm{irreg}(X).$$

Furthermore, if Z is any regular real of dimension α, then

$$d(X, Z) \geqslant d(X, \alpha Y).$$

Proof. First note that from (8) in the proof of Theorem 5.1

$$d(\alpha Y \to X) \leqslant \beta - \alpha.$$

But suppose now that $d(\alpha Y \to X) < \beta - \alpha$. Let $\gamma = \frac{(\beta-\alpha)-d(\alpha Y \to X)}{2}$. Now

$$\frac{C(X \restriction n)}{n} \leqslant \frac{C(X \restriction n \mid \alpha Y \restriction n)}{n} + \frac{C(\alpha Y \restriction n)}{n} + \frac{3 \log n}{n}. \tag{9}$$

Similarly to Theorem 4.2, we can choose an m such that

1. $\dfrac{C(X \restriction m)}{m} \geqslant \beta - \dfrac{\gamma}{3}$ (as $\dim_p X \geqslant \beta$)
2. $\dfrac{C(X \restriction m | \alpha Y \restriction m)}{m} < \beta - \alpha - \gamma$ (as $d(\alpha Y \to X) < \beta - \alpha - \gamma$)
3. $\dfrac{3 \log m}{m} \leqslant \dfrac{\gamma}{3}$
4. $\dfrac{C(\alpha Y \restriction m)}{m} \leqslant \alpha + \dfrac{\gamma}{3}$ (as $\dim_p(\alpha Y) \leqslant \alpha$)

Then substituting into (9) we get

$$\beta - \frac{\gamma}{3} < (\beta - \alpha - \gamma) + (\alpha + \frac{\gamma}{3}) + \frac{\gamma}{3},$$

which gives a contradiction. Thus $d(\alpha Y \to X) = \mathrm{irreg}(X)$. But

$$\begin{aligned} d(X \to \alpha Y) &= \limsup_n \frac{C(\alpha Y \restriction n \mid X \restriction n)}{n} \\ &= \limsup_n \frac{C(X \restriction n \mid \alpha Y \restriction n)}{n} + \frac{C(\alpha Y \restriction n)}{n} - \frac{C(X \restriction n)}{n} \\ &\leqslant d(\alpha Y \to X) + \dim_p \alpha Y - \dim_{\mathcal{H}} X \\ &\leqslant \beta - \alpha. \end{aligned}$$

So $d(X, \alpha Y) = \max\{d(X \to \alpha Y), d(\alpha Y \to X)\} = \text{irreg}(X)$.

Finally let Z be regular and of dimension α. If we suppose $d(Z \to X) < \beta - \alpha$, then we can repeat the proof above with Z replacing αY to get a contradiction as the only assumption we made on αY was that $\dim_p \alpha Y \leqslant \alpha$. Therefore $d(Z \to X) \geqslant \beta - \alpha$ and $d(X, Z) \geqslant \beta - \alpha = d(X, \alpha Y)$.

$\square$

6 Open Questions and Directions

This paper is part of an ongoing project to investigate the properties of $2^{\mathbb{N}}$ under the d-metric. These properties are generally geometric in nature – arising from the interplay between the scalar multiplication and the metric. The ultimate goal is to get a thorough picture of the geometric structure of this space. We give a list here of some open questions that will inform the general direction of our work.

Of course virtually any question one could ask about a general topological space, one could ask about the d-topology. We limit questions here to ones we consider most interesting and tractable.

1. Is $2^{\mathbb{N}}$ complete under the d-metric, i.e. does every Cauchy sequence converge?

2. Between any two points X, Y in $2^{\mathbb{N}}$, does there exist a path between X and Y whose length is exactly $d(X, Y)$? Is there a simple description of such a path if it exists?

3. Given and two regular reals X and Y of the same dimension α, does there exist a path from X to Y all of whose points have dimension α?

4. Is $2^{\mathbb{N}}$ locally compact? Or perhaps more to the point: Does there exist any point in $2^{\mathbb{N}}$ with a compact neighbourhood?

5. What is the fundamental group and topological dimension of $2^{\mathbb{N}}$?

   There are also interesting computability theoretic questions that can be asked. In particular what role if any do the Martin-Löf random reals play in this metric space?

6. Does every d-equivalence class of dimension 1 regular reals contain a random?

References

[1] Stephen Binns. Relative Kolmogorov Complexity and Geometry. *Submitted*.

[2] Rod Downey and Noam Greenberg. Turing degrees of reals of positive effective packing dimension. *Information Processing Letters*, 108:198–203, 2008.

[3] Rodney G Downey and Denis R Hirschfeldt. *Algorithmic Randomness and Complexity*. Springer, 2010. xxviii+855 pages.

[4] Jack H. Lutz. Dimension in complexity classes. *Proceeedings of the Fifteenth Annual IEEE Conference on Computational Complexity, IEEE Computer Society*, pages 158–169, 2000.

[5] Elvira Mayordomo. A kolmogorov complexity characterization of constructive hausdorff dimension. *Information Processing Letters*, 84, 2002.

[6] Krishna B. Athreya, John M. Hitchcock, Jack H. Lutz, Elvira Mayordomo. Effective strong dimension, algorithmic information, and computationl complexity. *arXiv:cs/0211025v3*, 2003.

[7] Jan Reimann. *Computability and Fractal dimension*. PhD thesis, Ruprecht-Karls-Universität, Heidelberg, 2004. 130 pages.

[8] C. Tricot. Two definitions of fractional dimensions. *Mathematical Proceedings of the Cambridge Philosophical Society*, 91:57–74, 1982.