A Tight Upper Bound on Kolmogorov Complexity by Hausdorff Dimension and Uniformly Optimal Prediction

Ludwig Staiger*
RWTH Aachen
Lehrstuhl für Informatik II
Ahornstr. 55
D–52074 Aachen
Germany

Abstract

The present paper links the concepts of Kolmogorov complexity (in Complexity theory) and Hausdorff dimension (in Fractal geometry) for a class of recursive (computable) ω-languages.

It is shown that the complexity of an infinite string contained in a Σ_2-definable set of strings is upper bounded by the Hausdorff dimension of this set and that this upper bound is tight. Moreover, we show that there are computable gambling strategies guaranteeing a uniform prediction quality arbitrarily close to the optimal one estimated by Hausdorff dimension and Kolmogorov complexity provided the gambler's adversary plays according to a sequence chosen from a Σ_2-definable set of strings.

We provide also examples which give evidence that our results do not extend further in the Arithmetical hierarchy.

*electronic mail: `staiger@zeus.informatik.rwth-aachen.de`

The *Kolmogorov complexity* of a finite word w, $K(w)$, is, roughly speaking, the length of a shortest input (program) π for which a universal algorithm $\mathfrak{U}$ prints w (see e.g. [Ch87], [LV93]). For infinite words (ω-words) its (normalized) Kolmogorov complexity might be thought of as

$$\lim_{n\to\infty} \frac{K(\xi/n)}{n} \ , \tag{1}$$

where ξ/n denotes the prefix of length n of ξ.

Although this limit need not exist, Cai and Hartmanis [CH89] proved that the graph of the function $t : \xi \to [0,1]$ is a fractal regardless by which intermediate value

$$\underline{\kappa}(\xi) := \liminf_{n\to\infty} \frac{K(\xi/n)}{n} \leq t(\xi) \leq \kappa(\xi) := \limsup_{n\to\infty} \frac{K(\xi/n)}{n} \tag{2}$$

the possibly not existing limit in Eq. (1) is replaced.

Other evidence of the intimate relationship between fractals and Kolmogorov complexity was given in [Br82], [Ry86], [Ry93, 94], [St81] and [St93] where a different approach was pursued:

Given a (computable) set of infinite words (ω-language) F bound the maximum possible Kolmogorov complexity $\underline{\kappa}(\xi)$ and $\kappa(\xi)$ for $\xi \in F$.

Here Ryabko [Ry86] showed that for arbitrary ω-languages F the Hausdorff dimension, $\dim F$, is a lower bound to $\underline{\kappa}(F) := \sup\{\underline{\kappa}(\xi) : \xi \in F\}$, but as Example 3.18 of [St93] shows already for simple computable ω-languages Hausdorff dimension is not an upper bound to $\kappa(F) := \sup\{\kappa(\xi) : \xi \in F\}$. Instead the results of [Br82], [St81] and [St93] show that the upper Kolmogorov complexity of ω-words in F is closely related to (metric) entropy, an information size measure different from Hausdorff dimension (cf. [Fa90]).

In [St93], however, we showed that for restricted classes of computable ω-languages F (that is, ω-languages satisfying additionally certain combinatorial properties) its Hausdorff dimension is also an upper bound to $\underline{\kappa}(F)$, thus giving a partial completion to Ryabko's lower bound.

After having introduced some notation and preliminaries in Section 1, we show in the the second part that Hausdorff dimension is an upper bound to lower Kolmogorov complexity for arbitrary Σ_2-definable ω-languages. Moreover, we give evidence that, unless other structural properties of ω-languages are involved (e.g. as in [St93]), our upper bound does not extend further in the arithmetical hierarchy of ω-languages. This result tightens Ryabko's lower bound in the range of computable (recursive) ω-languages.

Then the third section considers the following prediction problem for infinite sequences (cf. [Ry93, 94]): A gambler plays a fair game against an adversary which draws its outcomes according to the symbols of an arbitrarily chosen infinite sequence $\xi \in X^\omega$. The prediction quality of the gambler's strategy is measured by the logarithmic increase of the gambler's capital. It was shown in [Ry93, 94] that this increase $\lambda(\xi)$ is bounded from above by $1 - \underline{\kappa}(\xi)$ provided the gambler plays according to a computable strategy.

In this paper we investigate the case when the adversary is bound to draw its outcomes from a sequence $\eta \in F$ (unknown to the gambler) where F is a Σ_2-definable ω-language.

We show that then there are computable strategies guaranteeing a prediction quality $\lambda(\eta)$ arbitrarily close to the value $(1-\dim F)$ uniformly on F, that is, regardless from which $\eta \in F$ the adversary draws its outcomes.

Finally, Section 4 gives some connections to previous work, in particular, upper bounds on Kolmogorov complexity via Hausdorff dimension for ω-languages satisfying additional combinatorial properties are considered.

1 Notation and Preliminary Results

By $\mathbb{N} = \{0, 1, 2, \ldots\}$ we denote the set of natural numbers. We consider the space X^ω of infinite strings (sequences, ω-words) on a finite alphabet of cardinality $r := \mathit{card}\, X \geq 2$. By X^* we denote the set (monoid) of finite strings (words) on X, including the *empty* word e. For $w \in X^*$ and $b \in X^* \cup X^\omega$ let $w \cdot b$ be their *concatenation*. This concatenation product extends in an obvious way to subsets $W \subseteq X^*$ and $B \subseteq X^* \cup X^\omega$. As usual we denote subsets of X^* as languages and subsets of X^ω as ω-languages.

Furthermore $|w|$ is the *length* of the word $w \in X^*$, hence $X^n = \{w : w \in X^* \wedge |w| = n\}$. As introduced above by b/n we denote the *length n prefix* of a string $b \in X^* \cup X^\omega$ provided $|b| \geq n$ whenever $b \in X^*$, and $\mathbf{A}(b) := \{b/n : n \in \mathbb{N} \wedge n \leq |b|\}$ and $\mathbf{A}(B) := \bigcup_{b \in B} \mathbf{A}(b)$ is the set of all finite prefixes of $b \in X^* \cup X^\omega$ and $B \subseteq X^* \cup X^\omega$, respectively.

As usual we define Π_1-*definable ω-languages* $E \subseteq X^\omega$ as

$$E = \{\xi : \forall n (n \in \mathbb{N} \rightarrow \xi/n \notin W_E)\} \tag{3}$$

where $W_E \subseteq X^*$ is a recursive language, and Σ_2-*definable ω-languages* $F \subseteq X^\omega$ as

$$F = \{\xi : \exists i (i \in \mathbb{N} \wedge \forall n (n \in \mathbb{N} \rightarrow (i, \xi/n) \notin M_F))\} \tag{4}$$

where M_F is a recursive subset of $\mathbb{N} \times X^*$.

Define $L_i := \{w : (i, w) \in M_F\}$. Then Eqs. (3) and (4) may be alternatively written as

$$E = X^\omega \setminus W_E \cdot X^\omega , \quad \text{and} \tag{5}$$

$$F = \bigcup_{i \in \mathbb{N}} (X^\omega \setminus L_i \cdot X^\omega) = \bigcup_{k \in \mathbb{N}} (X^\omega \setminus (\bigcap_{i=0}^{k} L_i \cdot X^*) \cdot X^\omega) . \tag{6}$$

In particular, every Σ_2-definable ω-language is a countable union of Π_1-definable ω-languages.

From the definitions and Eqs. (5) and (6) it is clear that we may think of W_E and M_F as upwards closed, that is, $W_E = W_E \cdot X^*$ and M_F satisfies the condition

$$\begin{array}{lcl} (k, w) \in M_F & \Longrightarrow & \forall v (v \in X^* \rightarrow (k, w \cdot v) \in M_F) \ \text{ and} \\ (k, w) \in M_F & \Longrightarrow & \forall i (i \leq k \rightarrow (i, w) \in M_F) , \end{array} \tag{7}$$

that is, $L_i = L_i \cdot X^*$ and $L_i \supseteq L_{i+1}$.

In order to give a more detailed analysis we consider X^ω as a topological space (Cantor space), where the open subsets are defined as $W \cdot X^\omega$ ($W \subseteq X^*$). It should be mentioned that this space is compact, hence for every closed subset $F \subseteq X^\omega$ the inclusion $F \subseteq W \cdot X^\omega$ implies that there is already a finite subset $W' \subseteq W$ for which $F \subseteq W' \cdot X^\omega$.

It is known (e.g. [St86, Theorems 6.2]) that a subset $F \subseteq X^\omega$ is Π_1-definable iff F is closed and $X^* \setminus \mathbf{A}(F)$ is recursively enumerable, consequently, a Σ_2-definable subset $E \subseteq X^\omega$ is a countable union of closed sets (a so-called $\mathbb{F}_\sigma$*-set*).

The Hausdorff dimension of an ω-language $F \subseteq X^\omega$ may be defined using the entropy of languages. To this end we introduce the *structure function* of a language $L \subseteq X^*$, $\boldsymbol{s}_L : \mathbb{N} \to \mathbb{N}$, as

$$\boldsymbol{s}_L(n) := card\, L \cap X^n \ ,$$

and we define its *entropy*

$$H_L := \limsup_{n\to\infty} \frac{1 + \log_r \boldsymbol{s}_L(n)}{n} \ .$$

In other words, we have the following implications.

$$\textstyle\sum_{n\in\mathbb{N}} \boldsymbol{s}_L(n) \cdot r^{-\alpha\cdot n} = \sum_{v\in L} r^{-\alpha\cdot|v|} < \infty \quad \to \quad \alpha \geq H_L \tag{8}$$

$$\textstyle\sum_{n\in\mathbb{N}} \boldsymbol{s}_L(n) \cdot r^{-\alpha\cdot n} = \sum_{v\in L} r^{-\alpha\cdot|v|} = \infty \quad \to \quad \alpha \leq H_L \tag{9}$$

Moreover, let $V^\delta := \{\xi : \xi \in X^\omega \wedge \mathbf{A}(\xi) \cap V \text{ is infinite}\}$ be the δ*-limit* of the language $V \subseteq X^*$. Then according to Eq. (3.11) of [St93]

$$\dim F := \inf \{H_W : W \subseteq X^* \wedge F \subseteq W^\delta\} \tag{10}$$

is the *Hausdorff dimension* of $F \subseteq X^\omega$.

In particular, following Lemma 3.10 of [St93], we have

$$\alpha > \dim F \to \forall\varepsilon\Big(\varepsilon > 0 \to \exists W(W \subseteq X^* \wedge F \subseteq W \cdot X^\omega \wedge \sum_{w\in W} r^{-\alpha\cdot|w|} < \varepsilon)\Big) \ . \tag{11}$$

We mention still a relation between lower Kolmogorov complexity $\underline{\kappa}$ and the entropy of languages (cf. [St93, Proposition 2.14]).

Lemma 1 *If $L \subseteq X^*$ and L or $X^* \setminus L$ are recursively enumerable then*

$$\underline{\kappa}(\xi) \leq H_L \quad \textit{for all} \quad \xi \in L^\delta \ .$$

2 An Upper Bound for Kolmogorov Complexity

In the first section we derive the upper bound for $\underline{\kappa}(\xi)$ by the Hausdorff dimension of $F \subseteq X^\omega$, $\dim F$, provided $\xi \in F$ and F is Σ_2-definable.

To this end we derive a recursive analogue to the definition of the Hausdorff dimension (cf. Eq. (10)).

Theorem 2 *If $F \subseteq X^\omega$ is a Σ_2-definable ω-language then*

$$\dim F = \inf \{ H_V : F \subseteq V^\delta \wedge V \subseteq X^* \text{ is recursive} \} .$$

Remark: It should be mentioned that an ω-language $F \subseteq X^\omega$ is Σ_2-definable if and only if there is a recursive language $W \subseteq X^*$ such that $F = X^\omega \setminus W^\delta$ (cf. [St86, Theorem 7.4]). Accordingly, an ω-language $E \subseteq X^\omega$ is Π_2-*definable* iff there is a recursive language $V \subseteq X^*$ such that $E = V^\delta$, but as we shall see below in Lemma 4 our recursive analogue to the definition of the Hausdorff dimension is not valid for Π_2-definable ω-languages.

Proof. We describe an algorithm which constructs for every $\alpha > \dim F$ a recursive language V_α such that $F \subseteq V_\alpha^\delta$. To this end let $(U_j)_{j\in\mathbb{N}}$ be an effective enumeration of all finite prefix codes in X^* such that $\sup\{|v| : v \in U_j\} \leq \sup\{|v| : v \in U_{j+1}\}$, and let $L_k \subseteq X^*$ be defined according to Eq. (6) as $L_k := (\bigcap_{i=0}^{k} L_i \cdot X^*)$.

Define $\mathbf{test}_\alpha(k,j,n) := \Big(U_j \cdot X^\omega \cup (L_k \cap X^n) \cdot X^\omega = X^\omega \ \wedge \ \sum_{v\in U_j} r^{-\alpha\cdot|v|} < 2^{-k} \Big)$.

Observe that $\mathbf{test}_\alpha(k,j,n)$ is effectively computable provided $r^{-\alpha}$ is rational and that if $\mathbf{test}_\alpha(k,j,n)$ is true then $F \subseteq U_j \cdot X^\omega$, because $L_k \cdot X^\omega \cap F = \emptyset$.

Now the following algorithm, when given M_F, computes a finite prefix code C_k satisfying $F \subseteq C_k \cdot X^\omega$ and $\sum_{w\in C_k} r^{-\alpha\cdot|w|} < 2^{-k}$:

Algorithm C_k

input	$k,\ r^\alpha$		
	$n = 0$		
	repeat	$j = -1$	
		repeat	$j = j+1$
		until	$\mathbf{test}_\alpha(k,j,n) \vee \big(\sup\{|v| : v \in U_j\} > n\big)$
		$n = n+1$	
	until	$\mathbf{test}_\alpha(k,j,n)$	
output	$C_k := U_j$		

Informally, our algorithm successively for every $n \geq 0$ searches for a U_j satisfying the condition $\mathbf{test}_\alpha(k,j,n)$, more precisely, it searches until such a U_j is found or else all U_j having $\sup\{|v| : v \in U_j\} \leq n$ fail to satisfy $\mathbf{test}_\alpha(k,j,n)$.

In the latter case the value of n is increased (thus resulting in allowing for a larger codeword length and in enlarging the complementary ω-language $(L_k \cap X^n) \cdot X^\omega$) and the search starts anew. Consequently, the algorithm terminates iff there is a finite prefix code U such that $\sum_{v\in U} r^{-\alpha\cdot|v|} < 2^{-k}$ and $U \cdot X^\omega \cup (L_k \cap X^n) \cdot X^\omega = X^\omega$ for some $n \in \mathbb{N}$.

We have still to verify that our algorithm always terminates provided $\alpha > \dim F$. To this end we observe that according to Eq. (11) for every $\varepsilon > 0$ there is a $W \subseteq X^*$ such that $F \subseteq W \cdot X^\omega$ and $\sum_{w\in W} r^{-\alpha\cdot|w|} < \varepsilon$.

Since $X^\omega \setminus L_k \cdot X^\omega$ is a closed subset of F, for $\varepsilon \leq 2^{-k}$ we find a finite subset $W_k \subseteq W$ such that $X^\omega \setminus L_k \cdot X^\omega \subseteq W_k \cdot X^\omega$. Consequently, there is also a finite prefix code U_j satisfying $(U_j \cup L_k) \cdot X^\omega = X^\omega$ and thus $(U_j \cup (L_k \cap X^n)) \cdot X^\omega = X^\omega$ for n large enough.

Now let $V_\alpha := \bigcup_{k\in\mathbb{N}} C_k$. Obviously, V_α is recursively enumerable if $r^{-\alpha}$ is rational. Note that V_α is even recursive: Since $\alpha > 0$ and $w \in C_k$ imply $r^{-\alpha\cdot|w|} < 2^{-k}$, we have $w \in V_\alpha$ iff $\exists k(k < \alpha \cdot |w| \wedge w \in C_k)$. The predicate $k < \alpha \cdot |w|$ is recursive and bounds the quantifier $\exists k$ from above.

Next, we have $\sum_{w\in V_\alpha} r^{-\alpha\cdot|w|} \le \sum_{k\in\mathbb{N}}\sum_{w\in C_k} r^{-\alpha\cdot|w|} \le \sum_{k\in\mathbb{N}} 2^{-k} = 2 < \infty$ whence $H_{V_\alpha} \le \alpha$.

It remains to show that $F \subseteq V_\alpha^\delta$. If $\xi \in F$ there is a k such that $\xi \in X^\omega \setminus L_i \cdot X^\omega$ for all $i \ge k$. Hence, ξ has a prefix $w_i \in C_i$ for $i \ge k$. As it was observed above, $|w_i| > i/\alpha$. Consequently, ξ has infinitely many prefixes in $V_\alpha = \bigcup_{i\in\mathbb{N}} C_i$. □

Utilizing Lemma 1 we obtain the announced upper bound on $\underline{\kappa}(F)$ by $\dim F$.

Theorem 3 *If $F \subseteq X^\omega$ is a Σ_2-definable ω-language then $\underline{\kappa}(\xi) \le \dim F$ for all $\xi \in F$.*

Observe that, since Hausdorff dimension is countably stable, that is,

$$\dim \bigcup_{i\in\mathbb{N}} F_i = \sup \dim F_i \ , \tag{12}$$

this result immediately extends to arbitrary countable unions of Σ_2-definable (or, equivalently, to countable unions of Π_1-definable) ω-languages F_i.

The bound of Theorem 3 cannot be tightened to ensure that there is indeed a $\xi \in F$ such that $\underline{\kappa}(\xi) = \dim F$, as the following example shows.

Example 1 *Let $E := \{0\}^\omega \cup \bigcup_{n\in\mathbb{N}} 0^n \cdot (1 \cdot X^n)^\omega$. Then $\mathbf{A}(E)$ is a recursive language, hence E is Π_1-definable. In the same way one shows that the ω-languages $\{0\}^\omega$ and $0^n \cdot (1 \cdot X^n)^\omega$ are also Π_1-definable.*

Now, similar to [St93, Example 4.11] one computes $\dim\ 0^n \cdot (1 \cdot X^n)^\omega = \frac{n}{n+1}$, hence $\underline{\kappa}(\xi) \le \frac{n}{n+1}$ for $\xi \in 0^n \cdot (1 \cdot X^n)^\omega$, in contrast to the obvious identity $\dim E = 1$.

Next, we show that the upper bound of Theorem 3 cannot be extended further to higher classes of the Arithmetical hierarchy of ω-languages.

To this end we consider the class

$$\mathfrak{S} := \{E : E \subseteq X^\omega \wedge E \text{ is closed } \wedge \mathbf{A}(E) \text{ is recursively enumerable}\} \ ,$$

which is in some sense dual to the class of Π_1-definable ω-languages. Since $E \in \mathfrak{S}$ implies $E = \mathbf{A}(E)^\delta$, every $E \in \mathfrak{S}$ is Π_2-definable, but as the remark following Corollary 7.2 in [St86] shows, not every Π_2-definable and closed ω-language belongs to $\mathfrak{S}$.

In [St86] it was asked for a "nice" characterization of the class $\mathfrak{S}$. Already in Example 1.15 of [St93] it was explained that except for the above mentioned Π_2-definability no other relationship to the classes of the arithmetical hierarchy of ω-languages can be established. Here we give further evidence of this latter fact resulting in a proof that an analogue of Theorem 3 cannot hold for the class $\mathfrak{S}$. More precisely, we present a countable ω-language $E \in \mathfrak{S}$ which except for a single limit point $\xi \in E$ is Σ_2-definable but does not satisfy $\underline{\kappa}(E) \le \dim E$.

Lemma 4 *There is a countable ω-language $E \in \mathfrak{S} \setminus \Sigma_2$ such that E is the closure (in Cantor space) of a Σ_2-definable ω-language E' with $\underline{\kappa}(\xi) = 1$ for $\xi \in E \setminus E'$.*

Proof. P. Martin-Löf (cf. [LV93] or [Ca94]) proved that a random sequence $\zeta \in X^\omega$ has $\underline{\kappa}(\zeta) = 1$ and that the set of all random sequences $\mathfrak{R}_{\mathrm{rand}}$ is Σ_2-definable. Hence, $\mathfrak{R}_{\mathrm{rand}}$ contains a nonempty Π_1-definable subset $F = X^\omega \setminus W_F \cdot X^\omega$, where $W_F \subseteq X^*$ is a recursive language.

Fix the natural order on $X = \{0, 1, \ldots, r-1\}$ and extend it to a quasilexicographical order on X^*. Since F is closed in the Cantor topology of X^ω, it contains a leftmost sequence ζ_{left}.[1]

We show that there is a countable ω-language $E \in \mathfrak{S}$ such that E contains the leftmost sequence $\zeta_{\mathit{left}} \in F$.

Define for every $n \in \mathbb{N}$ the word v_n as the lexicographically first word in $X^n \setminus W_F \cdot X^*$. Obviously, the language $V := \{v_n : n \in \mathbb{N}\}$ is recursive, and $\zeta_n := v_n \cdot 0^\omega$ is the leftmost ω-word in $(X^n \setminus W_F \cdot X^*) \cdot X^\omega \supseteq F$.

Since $(X^n \setminus W_F \cdot X^*) \cdot X \supseteq X^{n+1} \setminus W_F \cdot X^*$, the family $\left((X^n \setminus W_F \cdot X^*) \cdot X^\omega\right)_{n \in \mathbb{N}}$ of closed subsets converges to $F = \bigcap_{n \in \mathbb{N}} (X^n \setminus W_F \cdot X^*) \cdot X^\omega$, and since ζ_n is the leftmost point in $(X^n \setminus W_F \cdot X^*) \cdot X^\omega \supseteq F$, we have also $\lim_{n \to \infty} \zeta_n = \zeta_{\mathit{left}}$ whence $\bigcup_{n \in \mathbb{N}} \mathbf{A}(\zeta_n) \supseteq \mathbf{A}(\zeta_{\mathit{left}})$.
Defining $E' := \{\zeta_n : n \in \mathbb{N}\} = V \cdot 0^\omega$ and $E := E' \cup \{\zeta_{\mathit{left}}\}$ yields that E is closed, $\mathbf{A}(E) = \mathbf{A}(V) \cup V \cdot 0^*$ and E' is Σ_2-definable. □

Since E is countable, $\dim E = 0$, and consequently, neither the analogue of Theorem 3 nor the recursive version of the definition of the Hausdorff dimension (Theorem 2) can hold for the class $\mathfrak{S}$. One can only show the following upper bound on $\underline{\kappa}(E)$ related to the structure function $\boldsymbol{s}_{\mathbf{A}(E)}$ (see [St93, Proposition 2.11]).

Theorem 5 *If $E \subseteq X^\omega$ is an ω-language such that $\mathbf{A}(E)$ or the complement $X^* \setminus \mathbf{A}(E)$ is recursively enumerable then*

$$\underline{\kappa}(\xi) \leq \liminf_{n \to \infty} \frac{\log_r \boldsymbol{s}_{\mathbf{A}(E)}(n)}{n} \quad \text{for all} \quad \xi \in E \ .$$

As a further consequence of the proof of Lemma 4 one obtains the following result on ω-languages in the arithmetical hierarchy.

Corollary 6 *There is an ω-language $E \in \mathfrak{S}$ which is not representable as a countable union of Σ_2-definable ω-languages.*

3 Computable Martingales and Prediction Quality

Quite recently Ryabko [Ry93, 94] proved another kind of relationship between Hausdorff dimension and Kolmogorov complexity. He considered a problem which is related

[1]This corresponds to the obvious interpretation of ω-words $\zeta \in X^\omega$ as r-ary expansions of real numbers $\iota(\zeta) := 0.\zeta \in [0,1]$. The mapping $\iota : X^\omega \to [0,1]$ is continuous and is one-to-one on $X^\omega \setminus X^* \cdot \{0^\omega, (r-1)^\omega\} \supseteq \mathfrak{R}_{\mathrm{rand}}$.

to the prediction problem of infinite sequences. This problem, addressed to in the introduction, can be easily described using Martingales, Ville's formalism of the concept of gambling strategy.

Here we consider Martingales as functions $\mathcal{V} : X^* \to [0, \infty)$ which satisfy

$$\mathcal{V}(e) > 0 \quad \text{and} \quad \mathcal{V}(w) = \frac{\sum_{x \in X} \mathcal{V}(wx)}{card\, X} \;,$$

and we denote by $\Lambda_{\mathcal{V}}(w) := \log_{card\, X} \mathcal{V}(w)$ the *average logarithmic increase (pay-off)* of $\mathcal{V}$.

Let $\mathcal{V}$ be a computable Martingale (or as in [Ry93, 94]: a Turing prediction strategy). Then in [Ry93, 94] it is proved that for all $\xi \in X^\omega$

$$\lambda_{\mathcal{V}}(\xi) := \limsup_{w \to \xi} \frac{\Lambda_{\mathcal{V}}(w)}{|w|} \leq 1 - \underline{\kappa}(\xi) \quad \text{, and} \tag{13}$$

$$\dim\Big\{\xi : \sup\{\lambda_{\mathcal{V}}(\xi) : \mathcal{V} \text{ is computable}\} \geq 1 - \alpha\Big\} = \alpha \;. \tag{14}$$

But it is also mentioned that there is no optimal prediction method, that is, there is no computable Martingale $\mathcal{V}_{\rm opt}$ such that $\lambda_{\mathcal{V}_{\rm opt}}(\xi) \geq \lambda_{\mathcal{V}}(\xi)$ for all $\xi \in X^\omega$ and all computable Martingales $\mathcal{V}$.

Our aim is to prove that for every Σ_2-definable ω-language F and every $\alpha > \dim F$ there is a computable Martingale $\mathcal{V}$ such that $\lambda_{\mathcal{V}}(\xi) \geq 1 - \alpha$ for all $\xi \in F$, thus being nearly uniformly optimal for the ω-language F.

In order to achieve our goal we introduce families of covering codes as follows.

For a prefix code $C \subseteq X^*$ we define its *minimal complementary code* as

$$\widehat{C} := (X \cup \mathbf{A}(C) \cdot X) \setminus \mathbf{A}(C) \;. \tag{15}$$

If $C = \emptyset$ we have $\widehat{C} = X$, and if $C \neq \emptyset$ the set $\widehat{C}$ consists of all words $w \cdot x \notin \mathbf{A}(C)$ where $w \in \mathbf{A}(C)$ and $x \in X$. It is readily seen that $C \cup \widehat{C}$ is a maximal prefix code and that $C \cap \widehat{C} = \emptyset$, and $\mathbf{A}(C \cup \widehat{C}) = \{e\} \cup \mathbf{A}(C) \cup \widehat{C}$.

We call $\mathfrak{C} := (C_w)_{w \in X^*}$ a *family of covering codes* provided each C_w is a finite prefix code. Since then every $C_w \cup \widehat{C}_w$ is a finite maximal prefix code, every word $u \in X^*$ has a uniquely specified $\mathfrak{C}$-factorization $u = u_1 \cdots u_n \cdot u'$ where $u_{i+1} \in C_{u_1 \cdots u_i} \cup \widehat{C}_{u_1 \cdots u_i}$ for $i = 0, \ldots, n-1$ ($u_1 \cdots u_i = e$, if $i = 0$) and $u' \in \mathbf{A}(C_{u_1 \cdots u_n} \cup \widehat{C}_{u_1 \cdots u_n})$. Analogously, every $\xi \in X^\omega$ has a uniquely specified $\mathfrak{C}$-factorization $\xi = u_1 \cdots u_i \cdots$ where $u_{i+1} \in C_{u_1 \cdots u_i} \cup \widehat{C}_{u_1 \cdots u_i}$ for $i = 1, \ldots$.

In the sequel we use Martingales derived from prefix codes in the following manner.

Lemma 7 *Let $0 \leq \alpha \leq 1$ and $\emptyset \neq C \subseteq X^*$ be a prefix code satisfying $\sum_{v \in C} r^{-\alpha|v|} < \infty$. Then there is a Martingale $\mathcal{V}_C^{(\alpha)} : X^* \to \mathbb{R}_+$ such that*

$$\mathcal{V}_C^{(\alpha)}(w) = \begin{cases} \dfrac{r^{(1-\alpha)|w|}}{\sum_{v \in C} r^{-\alpha|v|} + \sum_{u \in \widehat{C}} r^{-|u|}} & \text{, for } w \in C \text{, and} \\[2ex] \dfrac{1}{\sum_{v \in C} r^{-\alpha|v|} + \sum_{u \in \widehat{C}} r^{-|u|}} & \text{, for } w \in \widehat{C} \;. \end{cases} \tag{16}$$

Proof. Set $\Gamma := \sum_{v\in C} r^{-\alpha|v|} + \sum_{u\in\widehat{C}} r^{-|u|}$, and define

$$\begin{aligned}
\mathcal{V}_C^{(\alpha)}(u) &:= \frac{r^{|u|}}{\Gamma}\cdot\left(\sum_{u\cdot w\in C} r^{-\alpha|u\cdot w|} + \sum_{u\cdot w\in\widehat{C}} r^{-|u\cdot w|}\right) &&, \text{if } u\in\mathbf{A}(C\cup\widehat{C})\ , \text{and}\\
\mathcal{V}_C^{(\alpha)}(w\cdot v) &:= \mathcal{V}_C^{(\alpha)}(w) &&, \text{for } w\in C\cup\widehat{C}\ , v\in X^*\ .
\end{aligned}$$

Then $\mathcal{V}_C^{(\alpha)}$ fulfills Eq. (16). We have still to show that $\mathcal{V}_C^{(\alpha)}(u) = \frac{1}{r}\sum_{x\in X}\mathcal{V}_C^{(\alpha)}(ux)$. This identity is obvious if $u\notin\mathbf{A}(C\cup\widehat{C})$.
Now, let $u\in\mathbf{A}(C\cup\widehat{C})$. Then

$$\begin{aligned}
\frac{1}{r}\sum_{x\in X}\mathcal{V}_C^{(\alpha)}(ux) &= \frac{1}{r}\cdot\sum_{x\in X}\frac{r^{|ux|}}{\Gamma}\cdot\left(\sum_{uxw\in C} r^{-\alpha|uxw|} + \sum_{uxw\in\widehat{C}} r^{-|uxw|}\right)\\
&= \frac{r^{|u|}}{\Gamma}\cdot\sum_{x\in X}\left(\sum_{uxw\in C} r^{-\alpha|uxw|} + \sum_{uxw\in\widehat{C}} r^{-|uxw|}\right)\ ,
\end{aligned}$$

because for $u\in\mathbf{A}(C\cup\widehat{C})\setminus(C\cup\widehat{C})$ the set $\{w : w\in C\cup\widehat{C}\wedge u\sqsubseteq w\}$ partitions into the sets $\{w : w\in C\cup\widehat{C}\wedge ux\sqsubseteq w\}$ $(x\in X)$, and the required equation follows. □

Remark. If C is a finite prefix code and $r^{-\alpha}$ is a rational number then $\mathcal{V}_C^{(\alpha)}$ is a computable (recursive) Martingale.

Now let be given a family of covering codes $\mathfrak{C} := (C_w)_{w\in X^*}$ such that $\sum_{v\in C_w} r^{-\alpha|v|} < \infty$, and let $\mathcal{V}_{C_w}^{(\alpha)}$ be as defined above. Then we define our Martingale $\mathcal{V}$ as follows:
For $u\in X^*$ consider the $\mathfrak{C}$-factorization $u_1\cdots u_n\cdot u'$.
Then

$$\mathcal{V}_{\mathfrak{C}}^{(\alpha)}(u) := \left(\prod_{i=0}^{n-1}\mathcal{V}_{C_{u_1\cdots u_i}}^{(\alpha)}(u_{i+1})\right)\cdot\mathcal{V}_{C_{u_1\cdots u_n}}^{(\alpha)}(u')\ , \tag{17}$$

that is, $\mathcal{V}_{\mathfrak{C}}^{(\alpha)}$ is in some sense the concatenation of the Martingales $\mathcal{V}_{C_w}^{(\alpha)}$. Observe that $\mathcal{V}_{\mathfrak{C}}^{(\alpha)}$ is computable if only $r^{-\alpha}$ is rational and the function which assigns to every w the corresponding code C_w is computable.
Then we have the following.

Lemma 8 *Let $\mathfrak{C} = (C_w)_{w\in X^*}$ be a family of covering codes such that $\sum_{v\in C_w} r^{-\alpha|v|} < 2^{-|w|}$ for some $\alpha\in(0,1)$, and let $\xi\in X^\omega$ have a $\mathfrak{C}$-factorization $\xi = u_1\cdots u_i\cdots$ such that all factors u_{i+1} belong to $C_{u_1\cdots u_i}$ whenever $i\geq n_\xi$. Then it holds*

$$\mathcal{V}_{\mathfrak{C}}(u_1\cdots u_i)\geq c_\xi\cdot r^{(1-\alpha)\cdot|u_1\cdots u_i|}$$

where $c_\xi > 0$ is a constant not depending on i.

Proof. Since $\widehat{C}$ is a code, we have $\sum_{v\in\widehat{C}_w} r^{-|v|}\leq 1$, and from the assumption we obtain

$$\sum_{v\in C_w} r^{-\alpha|v|} + \sum_{v\in\widehat{C}_w} r^{-|v|}\leq 2^{-|w|} + 1\ .$$

Now $|u_i| \geq 1$ and the above Eq. (16) yield

$$\mathcal{V}^{(\alpha)}_{C_{u_1\cdots u_i}}(u_{i+1}) \geq \begin{cases} \dfrac{1}{2^{-i}+1} & , \text{ if } i \leq n_\xi \ , \text{ and} \\ \dfrac{r^{(1-s)\cdot|u_{i+1}|}}{2^{-i}+1} & , \text{ if } i > n_\xi \ . \end{cases} \tag{18}$$

Put

$$c_\xi := \prod_{i=0}^{\infty} \frac{1}{2^{-i}+1} \cdot \prod_{i=0}^{n_\xi} r^{-(1-s)\cdot|u_{i+1}|} \ .$$

Clearly, $c_\xi > 0$, and using Eq. (18) by induction on i the assertion is easily verified. □

Now we prove the announced result.

Theorem 9 *For every Σ_2-definable ω-language $F \subseteq X^\omega$ and every $\alpha > \dim F$ there is a recursive Martingale $\mathcal{V}$ such that*

$$\lambda_{\mathcal{V}}(\xi) \geq 1 - \alpha \quad \textit{for all } \xi \in F \ .$$

Proof. If $\dim F = 1$ the assertion is obvious. Let $\dim F < 1$. We have to construct for every α such that $1 > \alpha > \dim F$ and r^α is rational a computable Martingale $\mathcal{V}$ satisfying the assertion.

In virtue of Lemma 8 it suffices to construct a computable family of covering codes $\mathfrak{C} = (C_w)_{w\in X^*}$ such that the function which assigns to every w the corresponding finite prefix code C_w is computable.

To this end we modify the predicate **test** introduced in the proof of Theorem 3 as follows.

$$\mathbf{test}'_\alpha(w,k,j,n) := \Big(w \cdot U_j \cdot X^\omega \cup (L_k \cap X^{k+n}) \cdot X^\omega \supseteq w \cdot X^\omega \ \wedge \sum_{v\in U_j} r^{-\alpha\cdot|v|} < 2^{-k} \Big) \ .$$

In the same way we modify the algorithm presented there.

Algorithm C_w

input	$w,\ r^\alpha$	
	$n = 0,\ k = \lvert w \rvert$	
	repeat	$j = -1$
		repeat $j = j + 1$
		until $\mathbf{test}'_\alpha(w,k,j,n) \vee \Big(\sup\{\lvert v \rvert : v \in U_j\} > n\Big)$
		$n = n + 1$
	until	$\mathbf{test}'_\alpha(w,k,j,n)$
output	$C_k := U_j$	

This algorithm computes a prefix code C_w which satisfies $\sum_{v\in C_w} r^{-\alpha\cdot|v|} < 2^{-|w|}$ and $w\cdot C_w\cdot X^\omega \supseteq w\cdot X^\omega \setminus L_{|w|}\cdot X^\omega$, and it terminates for $\alpha > \dim F$, because $\dim(F \cap w\cdot X^\omega) \leq \dim F < \alpha$.

It remains to show that every $\xi \in F$ has a $\mathfrak{C}$-factorization $\xi = u_1 \cdots u_i \cdots$ such that almost all factors u_{i+1} belong to the corresponding $C_{u_1 \cdots u_i}$.

Let $\xi \in F$. Then there is a $k \in \mathbb{N}$ such that $\xi \in X^\omega \setminus L_i \cdot X^\omega$ for all $i \geq k$. Consequently, $w \in \mathbf{A}(\xi)$ implies $w \notin L_k$, and according to the definition of $\mathfrak{C}$ there is a $u \in C_w$ such that $w \cdot u \in \mathbf{A}(\xi)$ whenever $|w| \geq k$. □

We cannot improve Theorem 9 much further, because on the one hand Theorem 3 and Eq. (13) show that it is not possible to construct a computable Martingale $\mathcal{V}$ such that $\forall \xi (\xi \in F \to \lambda_{\mathcal{V}}(\xi) \geq 1 - \alpha)$ when $\alpha < \dim F$, and on the other hand again the ω-language E defined in the proof of Lemma 4 shows that the result of Theorem 9 cannot be extended to further classes of the Arithmetical hierarchy.

4 Relation to previous results

In the introduction we mentioned that under certain additional structural constraints the result of Theorem 3 extends beyond the range of Σ_2-definable ω-languages. In this section we recall some of the results of [St93] giving evidence of this fact, but showing also that the constraints involved are not recursion-theoretic ones.

The first two classes of ω-languages satisfying properties analogous to Theorems 2 and 3 are the classes of recursive ω-power languages and of δ-limits of recursive submonoids.

An *ω-power language* is an ω-language of the form $W^\omega := \{\xi : \xi \in X^\omega \wedge \exists (w_i)_{i \in \mathbb{N}} (w_i \in W \wedge w_i \neq e \wedge \xi = w_0 \cdots w_i \cdots)\}$, where $W \subseteq X^*$, and the *submonoid* of X^* generated by W, W^*, is the language $W^* := \{w_1 \cdots w_n : n \in \mathbb{N} \wedge w_i \in W\}$.

Now Eq. 6.2 of [St93] is an analogue to Theorem 2 for recursive languages $W \subseteq X^*$.

$$\dim W^\omega = \dim (W^*)^\delta = H_{W^*} \tag{19}$$

Then Ryabko's lower bound $\dim W^\omega \leq \underline{\kappa}(W^\omega)$ in connection with Lemma 1 shows the following.

Proposition 10 *If $W \subseteq X^*$ or its complement $X^* \setminus W$ are recursively enumerable languages then $\underline{\kappa}(W^\omega) = \underline{\kappa}((W^*)^\delta) = dim W^\omega$.*

According to the discussion in Section 6 of [St93], the class $\{W^\omega : W \subseteq X^* \wedge W \text{ is recursive}\}$ is incomparable to the class of all Σ_2-definable subsets of X^ω.

A second class which exhibits a similar property is the class of so-called *balanced* ω-languages introduced in [St89], that is, subsets F of X^ω satisfying the property that there is a function $f : \mathbb{N} \to \mathbb{N}$ such that $f(n) = o(2^{c \cdot n})$ for arbitrary $c > 0$ for which the inequality

$$\boldsymbol{s}_F(w, |w| + n) \leq f(|w|) \cdot \frac{\boldsymbol{s}_{\mathbf{A}(F)}(|w| + n)}{\boldsymbol{s}_{\mathbf{A}(F)}(|w|)} \tag{20}$$

holds for arbitrary $w \in X^*$ and $n \in \mathbb{N}$. Here $\boldsymbol{s}_F(w, \cdot)$ is a shorthand for the structure function of the language $U(F, w) := \mathbf{A}(F) \cap w \cdot X^*$, $\boldsymbol{s}_{U(F,w)}$.

Roughly speaking, our Eq. (20) means that for arbitrary $w \in X^*$ the function $\boldsymbol{s}_F(w, \cdot)$

does not grow much faster than the average taken over all functions $\boldsymbol{s}_F(v,\cdot)$ such that $|v| = |w|$ and $v \in \mathbf{A}(F)$.
Then Theorem 4 of [St89] proves that for closed and balanced $F \subseteq X^\omega$ the following identity holds true.

$$\dim F = \liminf_{n\to\infty} \frac{\log_r \boldsymbol{s}_{\mathbf{A}(F)}(n)}{n} \tag{21}$$

Consequently, Theorem 5 implies an upper bound on $\underline{\kappa}(F)$ for balanced ω-languages in $\mathfrak{S}$.

Lemma 11 *If $F \in \mathfrak{S}$ is a balanced ω-language then $\underline{\kappa}(F) \leq \dim F$.*

We conclude this section with mentioning Corollary 3.17 of [St93] which proves that for closed and balanced ω-languages F there is always a $\xi \in F$ such that $\underline{\kappa}(\xi) \geq \dim F$, thus yielding, in contrast to Example 1 above, that for balanced closed[2] Σ_2-definable ω-languages F as well as for balanced ω-languages $F \in \mathfrak{S}$ there is indeed a $\xi \in F$ such that $\underline{\kappa}(\xi) = \underline{\kappa}(F) = \dim F$.

Concluding Remark

Our Theorems 2, 3 and 9 in connection with previous results of Ryabko [Ry86,93,94] and this author [St93] gave evidence that there is a strong coincidence between the concepts of Kolmogorov complexity (in Complexity theory) and Hausdorff dimension (in Fractal geometry) for a class of recursive (computable) ω-languages. The results of this paper show a borderline in the Arithmetical hierarchy up to which this coincidence holds true, and our examples give evidence that it does not extend much further.

References

[Br82] A. A. Brudno, Entropy and complexity of trajectories of dynamic systems. *Trudy Mosk. Mat. Obshchestva* **44** (1982), 124 – 149 (in Russian; English translation: *Trans. Mosc. Math. Soc.* **44** (1983) 127 – 151).

[CH89] J.-Y. Cai and J. Hartmanis, The Kolmogorov complexity of the real line is a fractal, in *"Proc. 4th IEEE Structure in Complexity Conference,"*, 138 – 146, IEEE Computer Science Press, Washington DC, 1989.

[Ca94] C. Calude, *Information and Randomness. An Algorithmic Perspective.* Springer-Verlag, Berlin, 1994.

[Ch87] G. J. Chaitin, *Information, Randomness, & Incompleteness. Papers on Algorithmic Information Theory.* World Scientific, Singapore, 1987.

[Fa90] K.J. Falconer, *Fractal Geometry.* Wiley, Chichester, 1990.

[2]This property, however, does not hold in general for non-closed balanced Σ_2-definable ω-languages, as one easily verifies. E.g. it does not hold for $E' := \bigcup_{n\in\mathbb{N}}(1\cdot X^n)^\omega$.

[LV93] M. Li and P.M.B. Vitányi, *An Introduction to Kolmogorov Complexity and its Applications*. Springer–Verlag, New York, 1993.

[Ry86] B. Ya. Ryabko, Noiseless Coding of Combinatorial Sources, Hausdorff Dimension, and Kolmogorov Complexity. *Problemy Peredachi Informatsii* **22** (1986), No. 3, 16–26 (in Russian; English translation: *Problems of Information Transmission* **22** (1986), No. 3, 170–179).

[Ry93] B. Ya. Ryabko, An Algorithmic Approach to Prediction Problems. *Problemy Peredachi Informatsii* **29** (1993), No. 2, 96–103 (in Russian).

[Ry94] B. Ya. Ryabko, The Complexity and Effectiveness of Prediction Algorithms. *J. of Complexity* **10** (1994), 281–295.

[St81] L. Staiger, Complexity and entropy, in J. Gruska and M. Chytil, (Eds.) *Mathematical Foundations of Computer Science*, Lecture Notes in Comput. Sci. **118**, 508 – 514, Springer–Verlag, Berlin, 1981.

[St86] L. Staiger, Hierarchies of recursive ω-languages. *J. Inform. Process. Cybernet. EIK*, **22** (1986), No. 5/6, 219 – 241.

[St89] L. Staiger, Combinatorial properties of the Hausdorff dimension. *J. Statist. Plann. Inference* **23** (1989), No. 1, 95 – 100.

[St93] L. Staiger, Kolmogorov Complexity and Hausdorff Dimension, *Inform. and Comput.* **102** (1993), No. 2, 159 – 194.