The Hausdorff Measure of Regular ω-languages is Computable*

Ludwig Staiger
Martin-Luther-Universität Halle-Wittenberg
Institut für Informatik
Kurt-Mothes-Str. 1, **D–06120** Halle,
Germany
Electronic mail: `staiger@cantor.informatik.uni-halle.de`

*This paper was written during my visit to the Centre of Discrete Mathematics and Theoretical Computer Science, Auckland, August 1998

Abstract

We show that there is an algorithm which computes Hausdorff dimension and Hausdorff measure of arbitrary regular ω-languages. Our algorithm is a generalization of the one given in [MS94] which was designed to compute Hausdorff dimension and Hausdorff measure of regular ω-languages closed in the Cantor topology.

Our new algorithm is based on a decomposition lemma for regular ω-languages and a relationship between the Hausdorff measure of regular ω-languages of a special shape and their closures.

In several previous papers we have shown how to calculate Hausdorff dimension and measure for certain classes of regular ω-languages (cf. [MS94], [St89], and [St93]). In this note we show that the results obtained in the papers [MS94] and [St93] can be used to give an effective procedure for the calculation of the Hausdorff measure for arbitrary regular ω-languages.

To this end we derive a decomposition lemma for regular ω-languages which extends in some sense decompositions presented by A. Arnold [Ar83], K. Wagner [Wa79] and L. Staiger and K. Wagner [SW74].

We assume the reader to be familiar with the basic facts of the theory of regular languages. Let X be a finite alphabet of cardinality $r := \mathrm{card}\, X \geq 2$, and let X^* and X^ω be the sets of (finite) words and ω-words over X, respectively. Concatenation is denoted by "$\cdot$" and the prefix relation by "$\sqsubseteq$". As usual, we consider X^ω as a topological space (Cantor space). The *closure* of a subset $F \subseteq X^\omega$, $\mathcal{C}(F)$, is described as $\mathcal{C}(F) := \{\xi : \mathbf{A}(\{\xi\}) \subseteq \mathbf{A}(F)\}$, where $\mathbf{A}(E)$ is the set of all finite prefixes of ω-words $\eta \in E$.

We postpone the definition of regularity for ω-languages to Section 2.

For more details on ω-languages and regular ω-languages see the survey papers [St97] and [Th90].

1 Hausdorff Dimension and Hausdorff Measure

First, we shall describe briefly the basic formulae needed for the definition of Hausdorff dimension and Hausdorff measure. For more background and motivation see Section 1 of [MS94].

We define, for $\alpha \in [0, \infty)$, $F \subseteq X^\omega$ and $V \subseteq X^*$,

$$\begin{array}{lcl} \mathbb{L}_\alpha(F;V) & := & \sum_{v\in V} r^{-\alpha\cdot|v|}, \text{ and} \\ \mathbb{L}_\alpha(F) & := & \liminf_{n\to\infty} \left\{ \mathbb{L}_\alpha(F;V) : V\cdot X^\omega \supseteq F \ \wedge\ \underline{\ell}(V) \geq n \right\}, \end{array} \tag{1}$$

where $\underline{\ell}(V) := \inf\{|v| : v \in V\}$.

Now consider $\mathbb{L}_\alpha(F)$ as a function of α. Then there is an $\alpha(F) \in [0, \infty)$ such that

$$\mathbb{L}_\alpha(F) = \begin{cases} \infty, & \text{if } \alpha < \alpha(F), \\ 0, & \text{if } \alpha > \alpha(F). \end{cases} \tag{2}$$

This number $\alpha(F)$ is called the *Hausdorff dimension* of F, $\dim F$, that is, the Hausdorff dimension of F is given by

$$\dim F = \sup\{\alpha : \alpha = 0 \vee \mathbb{L}_\alpha(F) = \infty\} = \inf\{\alpha : \mathbb{L}_\alpha(F) = 0\}.$$

The Hausdorff dimension for regular ω-languages has been proved to be computable (cf. [Ba89], [MW88] or [St89]). The aim of this note is to show how one can compute the value $\mathbb{L}_{\dim F}(F)$ (the Hausdorff measure of F) for an arbitrary regular ω-language.[1]

In [MS94] we presented an algorithm which computes, simultaneously, the dimension $\dim F$ and the value $\mathbb{L}_{\dim F}(F)$ for regular ω-languages closed[2] in the Cantor topology of X^ω. Our new algorithm will be based on this procedure. To this end we derive some properties of the function $\mathbb{L}_\alpha$. From the definition (1) one has immediately

$$\mathbb{L}_\alpha(w \cdot F) = r^{-\alpha\cdot|w|} \cdot \mathbb{L}_\alpha(F) \tag{3}$$

Since regular ω-languages are Borel sets in the Cantor space (cf. [St97], [Th90]), they are measurable with respect to $\mathbb{L}_\alpha$. Thus we have the following (cf. [Fa85]).

---

[1] Observe that $\mathbb{L}_{\dim F}(F)$ is not specified by (2).

[2] A set $F \subseteq X^\omega$ is *closed* if $F = \mathcal{C}(F)$.

Proposition 1 *Assume $(F_i)_{i=0}^{\infty}$ is a family of mutually disjoint ω-languages measurable with respect to $I\!L_\alpha$. Then*

$$I\!L_\alpha(\bigcup_{i=0}^{\infty} F_i) = \sum_{i=0}^{\infty} I\!L_\alpha(F_i) \ .$$

Finally, we quote Theorem 6 of [MS94] (see also Lemma 4.5 of [St93]).

Proposition 2 *Let $V \subseteq X^*$ be regular and prefix-free. Then*

$$I\!L_\alpha(V^\omega) = I\!L_\alpha(\mathcal{C}(V^\omega)) \ .$$

2 Decomposition of Regular ω-languages

An ω-language $F \subseteq X^\omega$ is called *regular* provided there are a finite automaton $\mathfrak{A} = (X, S, s_0, \Delta)$ and a table $\mathcal{T} \subseteq \{S' : S' \subseteq S\}$ such that $\xi \in F$ if and only if $Inf(\mathfrak{A}, \xi) \in \mathcal{T}$ where $Inf(\mathfrak{A}, \xi)$ is the set of all states $s \in S$ through which the automaton $\mathfrak{A}$ runs infinitely often when reading the input ξ.

Observe that the ω-language $F = \{\xi : Inf(\mathfrak{A}, \xi) \in \mathcal{T}\}$ is the disjoint union of all sets $F_{S'} = \{\xi : Inf(\mathfrak{A}, \xi) = S'\}$ where $S' \in \mathcal{T}$.

We are going to split F into smaller mutually disjoint parts. Let $\mathfrak{A} = (X, S, s_0, \Delta)$ be fixed. We refer to a word $v \in X^*$ as (s, S')-*loop completing* if and only if

1. v is not the empty word,
2. $\Delta(s, v) = s$ and $\{\Delta(s, v') : v' \sqsubseteq v\} = S'$, and
3. $\{\Delta(s, v') : v' \sqsubseteq v''\} \subset S'$ for all proper prefixes $v'' \sqsubset v$ with $\Delta(s, v'') = s$,

and we call a word $w \in X^*$ (s, S')-*loop entering* provided

1. $\Delta(s_0, w) = s$, and
2. if $w = w' \cdot x$ for some $x \in X$ then $\Delta(s_0, w') \notin S'$.

Denote by $V_{(s,S')}$ the set of all (s, S')-loop completing words and by $W_{(s,S')}$ the set of all (s, S')-loop entering words. Both languages are regular and constructible from the finite automaton $\mathfrak{A} = (X, S, s_0, \Delta)$. Moreover, $V_{(s,S')}$ is prefix-free, whereas $W_{(s,S')}$ need not be so.

Nevertheless, every $\xi \in F_{S'}$ has a unique representation $\xi = w{\cdot}v_1 \cdots v_i \cdots$ where $w \in W_{(s,S')}$ and $v_i \in V_{(s,S')}$. Here the state $s \in S'$ is uniquely determined as the state succeeding the last state $\hat{s} \notin S'$ in the sequence $(\Delta(s_0, u))_{u \sqsubset \xi}$. Thus we obtain the following.

Lemma 3 (Decomposition Lemma) *Let $\mathfrak{A} = (X, S, s_0, \Delta)$ be a finite automaton and let $S' \subseteq S$. Then*

$$F_{S'} = \bigcup_{s \in S'} \bigcup_{w \in W_{(s,S')}} w \cdot V^{\omega}_{(s,S')} \ , \tag{4}$$

and the sets $w \cdot V^{\omega}_{(s,S')}$ are pairwise disjoint.

3 The Algorithm

Finally we derive the proof of our main result.

Theorem 4 *There is an algorithm which computes the Hausdorff measure $I\!\!L_{\dim F}(F)$ for every regular ω-language.*

From the decomposition in Lemma 3 we obtain via (3) and Proposition 1 a formula for the Hausdorff measure of $F_{S'}$:

$$I\!\!L_{\alpha}(F_{S'}) = \sum_{s \in S'} \Big(\sum_{w \in W_{(s,S')}} r^{-\alpha \cdot |w|} \Big) \cdot I\!\!L_{\alpha}(V^{\omega}_{(s,S')}) \ . \tag{5}$$

Since for regular languages $L \subseteq X^*$ the structure generating function of L, $\mathfrak{s}_L(t) := \sum_{w \in L} t^{|w|}$, is rational with integer coefficients and computable from L (cf. [KS86] or [SS78]), the sum $\sum_{w \in W_{(s,S')}} r^{-\alpha \cdot |w|}$ is computable, provided α is computable.

Proposition 2 shows that $I\!\!L_{\alpha}(V^{\omega}_{(s,S')}) = I\!\!L_{\alpha}(\mathcal{C}(V^{\omega}_{(s,S')}))$, because the language $V_{(s,S')}$ is regular and prefix-free.

Thus we obtain

$$\mathbb{L}_\alpha(F_{S'}) = \sum_{s\in S'} \mathfrak{s}_{W_{(s,S')}}(r^{-\alpha}) \cdot \mathbb{L}_\alpha(\mathcal{C}(V^\omega_{(s,S')})) \ . \tag{6}$$

Now the simultaneous computation of Hausdorff dimension and Hausdorff measure of a regular ω-language $F \subseteq X^\omega$ given by some finite automaton $\mathfrak{A} = (X, S, s_0, \Delta)$ and a table $\mathcal{T} \subseteq \{S' : S' \subseteq S\}$ proceeds as follows. Details should be carried out analogously to the algorithm described in Section 3 of [MS94].

1. For every $S' \in \mathcal{T}$ and every $s \in S'$ estimate the regular languages $V_{(s,S')}$ and $W_{(s,S')}$.
2. For every $S' \in \mathcal{T}$ estimate the adjacency matrix $\mathcal{A}_{S'}$ of $\mathcal{C}(V^\omega_{(s,S')})$.[3]
3. Calculate an eigenvalue $\lambda_{S'}$ of $\mathcal{A}_{S'}$ of maximum modulus.[4]
4. $\lambda_{\max} := \max\{|\lambda_{S'}| : S' \in \mathcal{T}\}$.
5. $\dim F := \log_r \lambda_{\max}$.
6. If $|\lambda_{S'}| < \lambda_{\max}$ then $\mathbb{L}_{\dim F}(\mathcal{C}(V^\omega_{(s,S')})) := 0$.
7. If $|\lambda_{S'}| = \lambda_{\max}$ then compute
   (a) $\mathbb{L}_{\dim F}(\mathcal{C}(V^\omega_{(s,S')}))$ according to Section 3 of [MS94], and
   (b) $\mathfrak{s}_{W_{(s,S')}}(\lambda^{-1}_{\max})$.
8. $$\mathbb{L}_{\dim F}(F) := \sum_{\lambda_{S'}=\lambda_{\max}} \sum_{s\in S'} \mathfrak{s}_{W_{(s,S')}}(\lambda^{-1}_{\max}) \cdot \mathbb{L}_{\dim F}(\mathcal{C}(V^\omega_{(s,S')})) \ .$$

References

[Ar83] A. Arnold, *Rational ω-languages are non-ambiguous*, *Theoret. Comput. Sci.* **26** (1983), 221–223.

[Ba89] C. Bandt, *Self-similar sets 3: Constructions with sofic systems*, *Mh. Math.* **108** (1989), 89 – 102.

[Fa85] K.J. Falconer, *The Geometry of Fractal Sets*, Cambridge University Press, Cambridge 1985.

[KS86] W. Kuich and A. Salomaa, *Semirings, automata, languages*, Springer-Verlag, Berlin 1986.

[MS94] W. Merzenich and L. Staiger, *Fractals, Dimension, and Formal Languages*, *RAIRO Inform. Théor.* **28**(2) (1994), 361 – 386.

[MW88] R.D. Mauldin and S.C. Williams, *Hausdorff dimension in graph directed constructions*, *Trans. Amer. Math. Soc.* **309**(2) (1988), 811 – 829.

[SS78] A. Salomaa and M. Soittola, *Automata-theoretic aspects of formal power series*, Springer-Verlag, New York 1978

[St89] L. Staiger, *Combinatorial properties of the Hausdorff dimension*, *J. Statist. Plann. Inference* **23** (1989), 95 – 100.

[St93] L. Staiger, *Kolmogorov complexity and Hausdorff dimension*, *Inform. and Comput.* **103**(2) (1993), 159 – 194.

[St97] L. Staiger, *ω-languages*, in: Handbook of Formal Languages (G. Rozenberg and A. Salomaa Eds.), Vol. 3, Springer-Verlag, Berlin 1997, 339 – 387.

[SW74] L. Staiger und K. Wagner, *Automatentheoretische und automatenfreie Charakterisierungen topologischer Klassen regulärer Folgenmengen*, *Elektron. Informationsverarb. Kybernetik* **EIK 10**(7) (1974), 379 – 392.

[Th90] W. Thomas, *Automata on Infinite Objects*, in: Handbook of Theoretical Computer Science (J. Van Leeuwen Ed.), Vol. B, Elsevier, Amsterdam 1990, 133 – 191.

[Wa79] K. Wagner, *On ω-regular sets*, *Inform. and Control* **43** (1979), 123 – 177.