HAUSDORFF DIMENSION OF OVERLAPPING SELF-SIMILAR SETS AND COMBINATORICS ON WORDS

ELMAR TEUFL

Abstract. **Presented at the conference "Fractals in Graz 2001" on June, 4th, 2001 as contributed talk.**

Let F be an overlapping self-similar set in $\mathbb{R}^2$. We are interested in the case that all similarities $\{\phi_1, \dots, \phi_n\}$ of F have the same contraction factor and all overlappings corresponds to an equation of the form $\phi_{i_1} \circ \cdots \circ \phi_{i_k} = \phi_{j_1} \circ \cdots \circ \phi_{j_k}$ for a fixed value of k in oder to assure a certain degree of regularity in the overlapping structure. Furthermore we require that the set F can also be obtained by cutting out scaled versions of some open set.

Under some additional technical assumptions we present a method for calculating the Hausdorff dimension of F. Combinatorics on words is used in connection with the set of algebraic equations for the similarities in order to apply well known results on cut-out sets.

1. Introduction

Suppose that $F \subseteq \mathbb{R}^2$ is a self-similar set and let $\phi_1, \dots, \phi_m$ be its similarities. We assume that all similarities have the same contraction factor $s > 0$. Furthermore we require that all overlappings of F are produced from a finite set of equations of the form $\phi_{i_1} \circ \cdots \circ \phi_{i_\ell} = \phi_{j_1} \circ \cdots \circ \phi_{j_\ell}$ for a fixed value of ℓ. For $i \neq j$ either $\phi_i(F) \cap \phi_j(F) = \emptyset$ or there is an equation with $i = i_1$ and $j = j_1$, such that

$$\phi_i(F) \cap \phi_j(F) = \phi_{i_1} \circ \cdots \circ \phi_{i_\ell}(F) = \phi_{j_1} \circ \cdots \circ \phi_{j_\ell}(F).$$

Last but not least we require that F can also be obtained by iteratively cutting out similar versions of some fixed set.

The following notation will be used frequently: Let $\mathcal{A} = \{1, \dots, m\}$. For $n \in \mathbb{N}$ let $\mathbb{W}_n = \mathcal{A}^n$ the word space consisting of all words over the alphabet $\mathcal{A}$ of length n and let $\mathbb{W}_0 = \{\epsilon\}$, where ϵ is the empty word. Furthermore we define $\mathbb{W} = \bigcup_{n \geq 0} \mathbb{W}_n$. For a finite word $w \in \mathbb{W}$ of length n we write $w_1, \dots, w_n$ for the letters of w and denote by $|w|$ the length n. We write i^n for the word of length $n \in \mathbb{N}$ consisting only of the letter $i \in \mathcal{A}$. The concatenation vw of $v = v_1 \dots v_n$ and $w = w_1 \dots w_k$ is the word $v_1 \dots v_n w_1 \dots w_k$ of length $n + k$. We call u a factor of w if there are $v, v' \in \mathbb{W}$, such that $w = vuv'$. The composition of the similarites $\phi_{w_1} \circ \cdots \circ \phi_{w_n}$ will be written as ϕ_w. Especially let ϕ_ϵ be the identity map.

We will demonstrate our method on the basis of a modified Sierpiński gasket given by three similarities ϕ_1, ϕ_2 and ϕ_3 with overlap. The similarities shall have the fixed points $(0,0)$, $(1,0)$ and $\frac{1}{2}(1, \sqrt{3})$ and the same scaling factor $\frac{1}{2} \leq s \leq 1$ and shall satisfy the equations

$$\phi_{12^\ell} = \phi_{21^\ell}, \qquad \phi_{13^\ell} = \phi_{31^\ell} \qquad \text{and} \qquad \phi_{23^\ell} = \phi_{32^\ell}$$

Date: 4 June 2001.

Supported by the START project Y96-MAT of the Austrian science fund FWF.

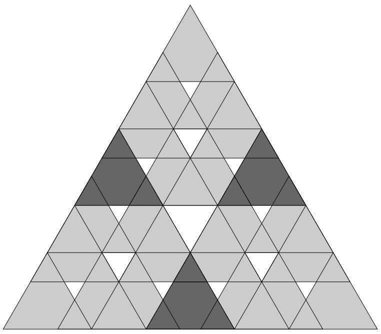
for a fixed $\ell \in \mathbb{N}$. This implies that the scaling factor s is a root of

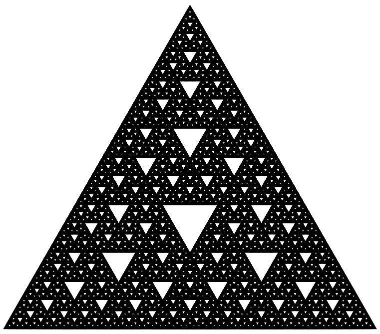
$$s^{\ell} + s^{\ell-1} + \cdots + s = 1.$$

From this equation we can get an asymptotic estimate for s by bootstrapping:

$$s = \tfrac{1}{2}\big(1 + 2^{-\ell-1} + (\ell+1)4^{-\ell-1}\big) + o(4^{-\ell})$$

The following picture shows the asserted structure of the self-similar set:





In this example we see that we have to cut out open triangles in every step. Such an open triangle has at least one address in the word space. In the first ℓ steps every triangle has exactly one address. In the $(\ell+1)$-th step there are three triangles which have two addresses coming from the equations for the similarities.

In order to calculate dimensions we need to know how many triangles we have to cut out in every step. To be precise we only need the exponential asymptotic behaviour.

From the geometry of the set and the equations of the similarities it is clear that we have to count words of length n over the alphabet $\mathcal{A} = \{1, 2, 3\}$, which contain no word of the following set as a factor:

$$U = \{21^{\ell}, 13^{\ell}, 32^{\ell}\}$$

Therefore we need combinatorics on words.

2. Combinatorics on words

We use a method developed by Guibas and Odlyzko [GO81] to solve our combinatorial problem. Because their original work is more general than we need, we will demonstrate their method in a simplified way.

Let $v, w \in \mathbb{W}$ be finite words over the alphabet $\mathcal{A} = \{1, \dots, m\}$. The *correlation* $\mathsf{Cor}(v, w)$ of v and w is defined as follows: The integer $n \in \mathbb{N}$ is a member of $\mathsf{Cor}(v, w)$ if $n \le |v|$ and $v_{|v|+1-k} = w_k$ for $1 \le k \le \min\{n, |w|\}$. In other words, if we place w under v so that the leftmost letter of w is under the n-th letter of v counted from the right and all pairs of letters in the overlapping segment are identical, then $n \in \mathsf{Cor}(v, w)$. Furthermore we introduce the overlapping polynomial

$$p_{v,w}(z) = \sum_{k \in \mathsf{Cor}(v,w)} z^{k-1}.$$

Look at $\mathcal{A} = \{1, 2\}$ and $v = 121221$ and $w = 12212$. Then we get $\mathsf{Cor}(v, w) = \{1, 4\}$ and $p_{v,w}(z) = z^3 + 1$.

121221	k
12212	**1**
12212	2
12212	3
12212	**4**
12212	5
12212	6

12212	k
121221	1
121221	**2**
121221	3
121221	4
121221	5

Because $\mathsf{Cor}(w,v)=\{2\}$ we see that $\mathsf{Cor}(v,w)\neq\mathsf{Cor}(w,v)$ in general!

Suppose $U\subseteq\mathbb{W}$ be a finite set of forbidden words. We assume that the U is reduced in the sense that no forbidden word is factor of another forbidden word. Let $f(k)=f(U;k)$ denote the number of words in $\mathbb{W}_k$ that do not contain any forbidden word as a factor. Similar we denote by $f_u(k)=f_u(U;k)$ for $u\in U$ the number of words in $\mathbb{W}_k$ that end with u and do not contain any forbidden word as a factor except the single appearance of u at the end. Furthermore we write

$$F(z)=\sum_{k=0}^{\infty}f(k)z^{-k}\qquad\text{resp.}\qquad F_u(z)=\sum_{k=0}^{\infty}f_u(k)z^{-k}$$

for the generating functions of $f(k)$ resp. $f_u(k)$ for $u\in U$. Then the generating functions satisfy

$$(z-m)F(z)+z\sum_{u\in U}F_u(z)=z\qquad\text{and}\qquad F(z)-z\sum_{u\in U}p_{u,v}(z)F_u(z)=0$$

for all $v\in U$. These are $|U|+1$ linear equations for the $|U|+1$ generating functions. The assumption, that U is reduced, guarantees the existence and uniqueness of the solution.

We will give a survey of the proof: If we append a letter to a word counted by $f(k)$, the resulting word will be counted by either $f(k+1)$ or $f_u(k+1)$ for exactly one $u\in U$. Hence

$$mf(k)=f(k+1)+\sum_{u\in U}f_u(k+1).$$

Because of $f(0)=1$ and $f_u(0)=0$ for all $u\in U$ we obtain

$$mF(z)=z(F(z)-1)+z\sum_{u\in U}F_u(z)$$

by multiplying the last equation by z^{-k} and summing over $k\geq 0$.

Let $u\in U$ and suppose $w\in\mathbb{W}_k$ is counted by $f(k)$. Then there is a first occurence of a forbidden word in wu; say $v\in U$ is this first occurence ending at position $t>k$, because w does not contain v as a factor. So the first t letters of wu are counted by $f_v(t)$ and the last $t-k$ letters of v coincide with the first $t-k$ letter of u. Hence we obtain $t-k\in\mathsf{Cor}(v,u)$. Conversely, if $t-k\in\mathsf{Cor}(v,u)$, then any word counted by $f_v(t)$ arises from the concatenation of a word counted by $f(k)$ and the word u. Therefore we get following equation for all $k\geq 0$:

$$f(k)=\sum_{v\in U}\sum_{r\in\mathsf{Cor}(v,u)}f_v(k+r)$$

Since $f_v(k)=0$ for all $k<|v|$ and $r\in\mathsf{Cor}(v,u)$ implies that $r<|v|$, we have

$$\sum_{k=0}^{\infty}\sum_{r\in\mathsf{Cor}(v,u)}f_v(k+r)z^{-k}=z\sum_{r\in\mathsf{Cor}(v,u)}z^{r-1}\sum_{k=0}^{\infty}f_v(k+r)z^{-(k+r)}=z\,p_{v,u}(z)F_v(z)$$

and obtain the remaining equations.

We turn now to our example: Remember that $\mathcal{A} = \{1,2,3\}$ and $U = \{21^\ell, 13^\ell, 32^\ell\}$ for a fixed value of $\ell \in \mathbb{N}$. For the correlations and overlapping polynomials we get:

$$\mathsf{Cor}(21^\ell, 13^\ell) = \mathsf{Cor}(13^\ell, 32^\ell) = \mathsf{Cor}(32^\ell, 21^\ell) = \{1\}$$
$$\mathsf{Cor}(13^\ell, 21^\ell) = \mathsf{Cor}(32^\ell, 13^\ell) = \mathsf{Cor}(21^\ell, 32^\ell) = \emptyset$$
$$\mathsf{Cor}(21^\ell, 21^\ell) = \mathsf{Cor}(13^\ell, 13^\ell) = \mathsf{Cor}(32^\ell, 32^\ell) = \{\ell+1\}$$
$$p_{21^\ell,13^\ell}(z) = p_{13^\ell,32^\ell}(z) = p_{32^\ell,21^\ell}(z) = 1$$
$$p_{13^\ell,21^\ell}(z) = p_{32^\ell,13^\ell}(z) = p_{21^\ell,32^\ell}(z) = 0$$
$$p_{21^\ell,21^\ell}(z) = p_{13^\ell,13^\ell}(z) = p_{32^\ell,32^\ell}(z) = z^\ell$$

Therefore we have to solve the following linear system of equations:

$$\begin{pmatrix} (z-3) & z & z & z \\ 1 & -z^{\ell+1} & 0 & -z \\ 1 & -z & -z^{\ell+1} & 0 \\ 1 & 0 & -z & -z^{\ell+1} \end{pmatrix} \begin{pmatrix} F(z) \\ F_{21^\ell}(z) \\ F_{13^\ell}(z) \\ F_{32^\ell}(z) \end{pmatrix} = \begin{pmatrix} z \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

From that we can easily compute the solution and by replacing z by $\frac{1}{z}$ we get the usual generating function $F\left(\frac{1}{z}\right)$ for the numbers $f(k)$:

$$F(z) = \frac{1+z^\ell}{1-3z^{\ell-1}+z^\ell} \qquad \text{and} \qquad F\left(\tfrac{1}{z}\right) = \frac{1+z^\ell}{1-3z+z^\ell}$$

By an application of Rouché's theorem we see that there is exactly one root of the denominator $1-3z+z^\ell$ inside the unit circle. By α we denote the modulus of this root. Hence we obtain

$$f(k) = C\alpha^{-k} + O(1)$$

for some constant C. From the equation of $1-3z+z^\ell = 0$ we can get the asymptotics for α by bootstrapping:

$$\alpha = \tfrac{1}{3}\left(1 + 3^{-\ell} + \ell\, 9^{-\ell}\right) + o(9^{-\ell})$$

3. Cut-out sets

We will need some further notation: Suppose that $A \subseteq \mathbb{R}^2$ is a subset of $\mathbb{R}^2$. We denote by A_r the r-neighbourhood of A:

$$A_r = \{x \in \mathbb{R}^2 \ : \ d(x, A) \le r\}$$

Moreover we define $A_r^+ = A_r \setminus A$ and $A_r^- = A_r \cap A$. By λ_1 resp. λ_2 we denote the 1-dimensional resp. 2-dimensional Lebesgue measure and by $p(A) = \lambda_1(\partial A)$ we denote the perimeter of A. Let ϕ be a similarity of $\mathbb{R}^2$ with scaling factor $t \ge 0$. Then we have $\lambda_2(\phi(A)) = t^2\lambda_2(A)$ and $p(\phi(A)) = t\,p(A)$ for all suitable sets $A \in \mathbb{R}^2$.

We follow a method of Falconer [Fa97] in order to calculate the box dimension of cut-out sets in $\mathbb{R}^2$. In our setting the cut-out set F is obtained by removing similar versions E_k ($k \in \mathbb{N}$) of one fixed open and bounded set $E \subseteq \mathbb{R}^2$ from a given compact set $G \subseteq \mathbb{R}^2$. We denote the contraction factor of E_k by t_k and assume, that the sets E_k are sorted in descending order of t_k. We require that the sets E_k are pairwise disjoint. Thus we have

$$F = G \setminus \biguplus_{k\in\mathbb{N}} E_k.$$

We assume the following two conditions:

(1) The perimeter $p(G)$ and the area $\lambda_2(G)$ exist and $\lambda_2(G_r^+) = p(G)r + c(G)r^2$ for sufficiently small $r > 0$ and a constant $c(G)$.

(2) The perimeter $p(E)$ and the area $\lambda_2(E)$ exist and there is a constant $c(E) \in \mathbb{R}$, such that $\lambda_2(\phi(E)_r^-) = t\,p(E)r + c(E)r^2$ holds for every similarity ϕ with scaling factor $t > 0$, if $\phi(E)_r^-$ does not cover $\phi(E)$.

The condition on the set E holds for example for convex polygones and for convex sets whose boundary is twice continuously differentiable and has bounded curvature.

Under these assumptions we can proof the following statement: Write

$$a = -\liminf_{k\to\infty} \frac{\log(t_k)}{\log(k)} \qquad \text{and} \qquad b = -\limsup_{k\to\infty} \frac{\log(t_k)}{\log(k)}.$$

If $\frac{1}{2} < b \le a < 1$, we obtain $\frac{1}{a} \le \dim_B(F) \le \frac{1}{b}$ for the box dimension of F.

We will give a short outline of the proof: There is a constant C_1 such that $t_k \le C_1 k^{-b}$ for sufficiently large k. Let $r > 0$ then there exists an integer $n = n(r)$ and a constant C_2 such that $C_2 t_{n+1} < r \le C_2 t_n$ and the set E_k is covered by $(E_k)_r^-$ if and only if $k > n$. Now we can estimate the area of the r-neighbourhood of F as follows:

$$\begin{aligned}
\lambda_2(F_r) &= \lambda_2(G_r^+) + \sum_{k=1}^{n} \lambda_2((E_k)_r^-) + \sum_{k=n+1}^{\infty} \lambda_2(E_k) \\
&= p(G)r + c(G)r^2 + \sum_{k=1}^{n} (p(E)t_k r + c(E)r^2) + \sum_{k=n+1}^{\infty} t_k^2\, \lambda_2(E) \\
&\le p(G)r + c(G)r^2 + c(E)nr^2 + C_3\, r \sum_{k=1}^{n} k^{-b} + C_4 \sum_{k=n+1}^{\infty} k^{-2b} \\
&\le p(G)r + c(G)r^2 + c(E)\, nr^2 + C_5\, n^{1-b} r + C_6\, n^{1-2b}
\end{aligned}$$

Here we denote by C_3, C_4, …some constants. Because of $n \le C_7\, r^{-1/b}$ for sufficiently small $r > 0$ we obtain

$$\lambda_2(F_r) \le p(G)r + c(G)r^2 + C_8\, r^{2-1/b} \le C_9\, r^{2-1/b}$$

for small values of $r > 0$. From the definition of the box dimension we end up with $\dim_B(F) \le \frac{1}{b}$. A similar argument yields the lower bound.

4. Collecting the pieces

By a theorem of Falconer [Fa97] the Hausdorff dimension is the same as the box dimension for every self-similar set. From this and the last two sections we are ready to calculate the Hausdorff dimension of our example.

We know that the contraction factor s is given by the unique root of $s^\ell + s^{\ell-1} + \cdots + s = 1$ in the intervall $\left(\frac{1}{2}, 1\right)$. Furthermore we know, that we have to remove $f(k) = C\alpha^{-k} + O(1)$ triangles of size s^k in the k-th step, where α is the unique root of the equation $\alpha^\ell - 3\alpha + 1 = 0$ in the intervall $\left(\frac{1}{3}, 1\right)$.

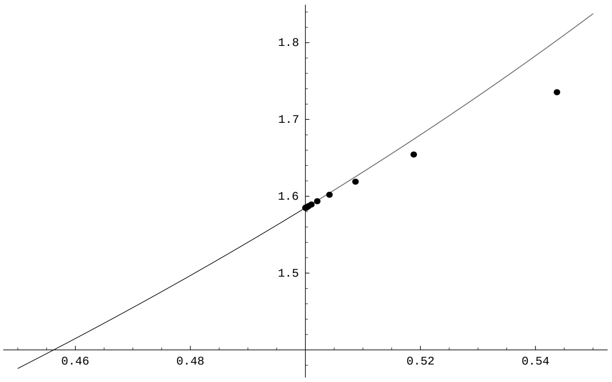
Let $g(n) = f(1) + \cdots + f(n)$ and $g(0) = 0$, then the contraction factor t_k of the k-th removed triangle is given by $t_k = s^n$ if $g(n-1) < k \le g(n)$. Since there are constants C_1 and C_2 sucht that $C_1\alpha^{-n} \le g(n) \le C_2\alpha^{-n}$, we obtain

$$a = -\lim_{k\to\infty} \frac{\log(t_k)}{\log(k)} = -\lim_{n\to\infty} \frac{\log(s^n)}{\log(\alpha^{-n})} = \frac{\log(s)}{\log(\alpha)}$$

and

$$\dim_H(F) = \dim_B(F) = \frac{\log(\alpha)}{\log(s)}.$$

We will now write F_ℓ, s_ℓ and α_ℓ in order to emphasise the dependency of these objects from the value of ℓ. First we remark that s_ℓ tends to $\frac{1}{2}$ and α_ℓ tends to $\frac{1}{3}$ as $\ell \to \infty$, so that $\dim_H(F)$ tends to $\frac{\log(3)}{\log(2)}$. The following picture shows the Hausdorff dimension as a function of the scaling factor s. The points represent the dimension of F_ℓ for some values of ℓ:



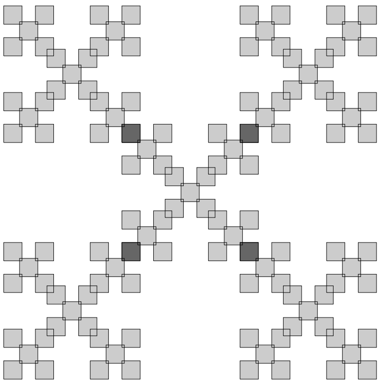
Furthermore we can show using the asymptotics of s_ℓ and α_ℓ that the limit of the difference quotients

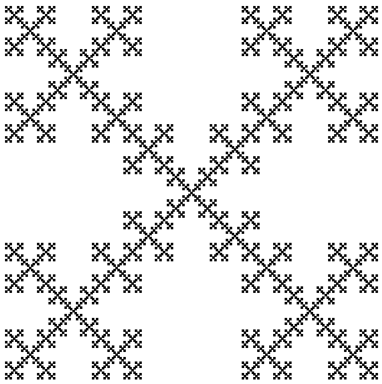
$$\lim_{\ell\to\infty} \frac{\dim_H(F_\ell) - \frac{\log(3)}{\log(2)}}{s_\ell - \frac{1}{2}}$$

is the same as the right derivative at $\frac{1}{2}$ of the Hausdorff dimension of the Sierpiński gasket with contraction factor less than $\frac{1}{2}$.

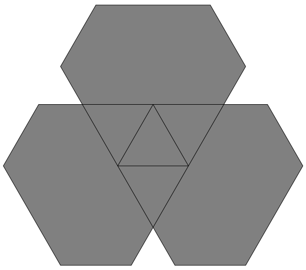
5. Remarks

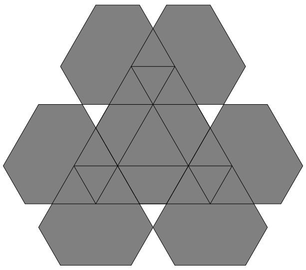
The presented method do not allow a fractal boundary of the cut-out set. But we can make a similar analysis for the box dimension if we cut out pieces from the boundary. As an example we mention a modified Vicsek set with same type of overlap:

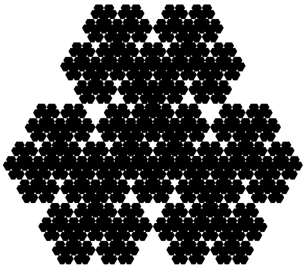




Also the structure of sets which have to cut out may get more involved. In the following example there are three similarities with three equations which are almost the same as for the Sierpiński gasket:







For this example one has to find out a rule, how to cut out the small triangles exactly once. The combinatorial problem is a little bit more complicated as bevor, but our method is still applicable.

If we drop the condition that all similarites have the same contraction factor the size of a set, which have been cutted out, does not only depend on the number of applied similarities but also how often one similarity is applied. Thus both the sorting of these sets and the counting problem are more difficult, and further work is needed to handle this case.

References

[Fa97] K. J. Falconer. *Techniques in Fractal Geometry.* John Wiley & Sons, Chichester, 1997

[GO81] L. J. Guibas, A. M. Odlyzko. *String Overlap, Pattern Matching and Nontransitiv Games.* J. Comb. Theory, Series A **30**, 183-208, 1981

Elmar Teufl, Institut für Mathematik, Technische Universität Graz, Steyrergasse 30, 8010 Graz, Austria

E-mail address: `teufl@weyl.math.tugraz.at`