Codes, Simplifying Words, and Open Set Condition

Ludwig Staiger*
Rhein.–Westf. Techn. Hochschule Aachen,
Lehrstuhl für Informatik II,
Ahornstr. 55,
D–52074 Aachen,
Germany

Abstract

A relationship between formal language theory and fractal geometry describing how codes may be used to generate (infinite) iterated function systems fulfilling certain properties is presented.

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

In the theory of iterated function systems (IFS) the so-called Open Set Condition (OSC) plays a crucial role. Here we show that an analogue of this condition corresponds to certain properties of codes in language theory.

Given an alphabet X, we consider the free monoid of (finite) words X^* generated by X with the concatenation $\cdot$ as operation and with the empty word e as identity element. As is well-known, codes over X are defined as the bases of free submonoids of X^*, that is, $C \subseteq X^*$ is a code if and only if every word w in the submonoid C^* of X^* generated by C has exactly one factorization over C.

An iterated function system is an $(n+1)$-tuple $\mathcal{I} = (\mathcal{M}, \varphi_1, \ldots, \varphi_n)$ where $(\mathcal{M}, \rho)$ is a metric space and $\varphi_1, \ldots, \varphi_n$ are contracting mappings on $\mathcal{M}$, that is mappings $\varphi_i\colon \mathcal{M} \to \mathcal{M}$ satisfying $\rho(\varphi_i(x), \varphi_i(y)) \leq c \cdot \rho(x, y)$ for some $c < 1$ (cf. [Ba88], [Fa90]). Then there is a unique compact set $\emptyset \neq \mathcal{K} \subseteq \mathcal{M}$ which is the solution of the equation

$$\mathcal{K} = \bigcup_{i=1}^{n} \varphi_i(\mathcal{K}) \ .$$

$\mathcal{K}$ is called the attractor of $\mathcal{I}$.
An IFS is referred to as satisfying the OSC provided there is a nonempty open subset $\mathcal{O} \subseteq \mathcal{M}$ such that $\varphi_i(\mathcal{O}) \subseteq \mathcal{O}$ and $\varphi_i(\mathcal{O}) \cap \varphi_j(\mathcal{O}) = \emptyset$ for $i, j \in \{1, \ldots, n\}$ and $i \neq j$, and it is referred to as satisfying the strong OSC (SOSC) if there is such an open set $\mathcal{O}$ with $\mathcal{O} \cap \mathcal{K} \neq \emptyset$.

In order to state the announced relationship between codes iterated function systems and we consider IFS in the space (X^ω, ρ) of all infinite strings over an alphabet X (functions mapping $\omega = \{1, 2, 3, \ldots\}$ to X) with the metric

$$\rho(\xi, \eta) := \inf\{(\operatorname{card} X)^{-i} : i = 0 \vee \forall j (j \leq i \to \xi(j) = \eta(j))\} \ .$$

Then a word $w \in X^*$ in a natural way defines a contracting similitude of the form $\varphi_w(\xi) = w \cdot \xi$, where $w \cdot \xi$ denotes the concatenation of the finite string w with the infinite string ξ.

Thus a set of words (a language) $V \subseteq X^*$ defines a set of contracting (if $w \neq e$) similitudes $\{\varphi_v : v \in V\}$ leading to a possibly infinite IFS (IIFS) with *attractor* $V^\omega := \{v_1 \cdot v_2 \cdot v_3 \cdot \ldots : v_i \in V\}$ in the space (X^ω, ρ) (cf. [St92]).
In the topology on X^ω defined by the metric ρ a set $E \subseteq X^\omega$ is *open* if and only if it is of the form $E = W \cdot X^\omega := \{w \cdot \xi : w \in W \wedge \xi \in X^\omega\}$ for some $W \subseteq X^*$.

Following the above open set condition we say that a language $V \subseteq X^*$ satisfies the OSC iff there is a nonempty $W \subseteq X^*$ such that

$$\forall v (v \in V \to v \cdot W \cdot X^\omega \subseteq W \cdot X^\omega) \ , \quad \text{and} \tag{1}$$

$$\forall v, v' (v, v' \in V \wedge v \neq v' \to v \cdot W \cdot X^\omega \cap v' \cdot W \cdot X^\omega = \emptyset) \tag{2}$$

hold true. We will refer to the language $W \subseteq X^*$ also as an OSC-*witness* for V.
If, moreover, $V^\omega \cap W \cdot X^\omega \neq \emptyset$ we say that V satisfies the Strong Open Set Condition (SOSC).

In this note we investigate the question which languages satisfy the OSC or SOSC. It turns out that every language satisfying OSC is a code whereas the converse, in general, need not be true. Therefore we ask for properties of codes which are related

to the open set condition.
In [LPRT79] the notion of simplifying word was introduced. We show here that this notion is closely related to the OSC. It allows for a characterization of languages satisfying the OSC or SOSC, thus yielding a close relationship between the Theory of Codes and Fractal Geometry.
First we show that all words contained in an OSC-witness for V are also simplifying for V. Then utilizing the natural partition of the set of simplifying words into simplifying nosubwords and simplifying prefixes we show that the existence of simplifying prefixes is crucial for the SOSC.
Finally, we apply these results in order to show that, in contrast to the finite case, for infinite IFS OSC and SOSC are not equivalent, and we present two classes of codes satisfying the SOSC.
We start with a fundamental property of OSC-languages.

Lemma 1 *If $V \subseteq X^*$ satisfies the OSC then V is a code.*

Proof. First observe that V cannot contain the empty word e. For otherwise $v \cdot W \cdot X^\omega \cap W \cdot X^\omega = \emptyset$ has to hold for all $v \in V$, contradicting Condition (1) $V \cdot W \cdot X^\omega \subseteq W \cdot X^\omega$.
Let W be an OSC-witness for V and suppose V to be no code. Then there are $u_1, \ldots, u_k, v_1, \ldots, v_\ell \in V$ such that $u_1 \neq v_1$ and $u_1 \cdots u_k = v_1 \cdots v_\ell$, that is, in particular $u_1 \cdots u_k \cdot W \cdot X^\omega = v_1 \cdots v_\ell \cdot W \cdot X^\omega$.
Now Condition (1) yields $u_1 \cdots u_{k-1} \cdot (u_k \cdot W \cdot X^\omega) \subseteq u_1 \cdots u_{k-1} \cdot W \cdot X^\omega \subseteq \ldots \subseteq u_1 \cdot W \cdot X^\omega$ and, similarly, $v_1 \cdots v_\ell \cdot W \cdot X^\omega \subseteq v_1 \cdot W \cdot X^\omega$.
Hence $u_1 \cdot W \cdot X^\omega \cap v_1 \cdot W \cdot X^\omega \supseteq u_1 \cdots u_k \cdot W \cdot X^\omega \neq \emptyset$, contradicting Condition (2).
Q.E.D.

The converse of Lemma 1 is not true. To this end we mention the following property of OSC-witnesses for V. Let $\mathbf{A}(V) := \{w : w \in X^* \wedge \exists v(v \in X^* \wedge w \cdot v \in V\}$ be the set of prefixes of words in V.

Claim 2 *If $W \subseteq X^*$ is an OSC-witness for V then $W \cdot V \cap \mathbf{A}(V) = \emptyset$.*

Proof. Assume $W \cdot V \cap \mathbf{A}(V) \neq \emptyset$. Then there are $w \in W$ and $v, v' \in V$ such that $w \cdot v$ is a prefix of v', that is, it holds $v \cdot w \cdot X^\omega \supseteq v' \cdot X^\omega \supseteq v' \cdot w \cdot X^\omega$, contradicting Condition (2).
Q.E.D.

The code C_1 in the following example does not satisfy the OSC.

Example 1 *Consider the suffix code $C_1 := \{wab^{|w|} : w \in \{a,b\}^*\}$. This code satisfies $\mathbf{A}(C_1) = \{a,b\}^*$ what implies that $C_1 \cdot W \cap \mathbf{A}(C_1) \neq \emptyset$ can never hold.*

As far we have seen that languages satisfying the OSC are codes, but not all codes satisfy the OSC. Next we introduce a property of words (with respect to languages) which is related to the OSC for languages.
According to [LPRT79] (see also [BP85, p.130]) we call a word w *simplifying* for a language V provided

$$\forall u, v(u \in V^* \wedge v \in X^* \wedge u \cdot w \cdot v \in V^* \rightarrow w \cdot v \in V^*) . \tag{3}$$

From the definition it is apparent that we can partition the set of simplifying words for a language V into two classes

nosubwords: all words w for which $V^* \cdot w \cap \mathbf{A}(V^*) = \emptyset$
(For these words the hypothesis of Condition (3) is never satisfied In particular, this concerns all words which are not subwords of a word in V^*.), and

prefixes: the remaining simplifying words with $V^* \cdot w \cap \mathbf{A}(V^*) \neq \emptyset$.
(In virtue of Condition (3) these words belong to $\mathbf{A}(V^*)$.)

We show that OSC-witnesses for a language V contain only simplifying words. In order to prove this proposition we give the following property of OSC-witnesses.

Claim 3 *Let W be an OSC-witness for V and $\emptyset \neq U \subseteq W$. Then $V^* \cdot U$ is also an OSC-witness for V.*

Proof. Condition (1) follows immediately from $V \cdot V^* \cdot U \subseteq V^* \cdot U$, and $v \cdot V^* \cdot U \cdot X^\omega \subseteq v \cdot V^* \cdot W \cdot X^\omega \subseteq v \cdot W \cdot X^\omega$ implies Condition (2).

Q.E.D.

We obtain our result.

Theorem 4 *Let $W \subseteq X^*$ be an OSC-witness for C and let $w \in W$. Then w is a simplifying word for C.*

Proof. If $C^* \cdot w \cap \mathbf{A}(C^*) = \emptyset$ then according to **nosubwords** w is simplifying.
Now assume $C^* \cdot w \cap \mathbf{A}(C^*) \neq \emptyset$. In view of the preceding claim we may consider the OSC-witness $W' := C^* \cdot w$. Assume that w is not a simplifying word for C. Then there are words u, v such that $u, u \cdot w \cdot v \in C^*$ but $w \cdot v \notin C^*$. Let u be a shortest word with this property. Obviously, u is not the empty word.

Thus $u = u_1 \cdot u_2$ where $u_1 \in C, u_2 \in C^*$. Similarly, $u \cdot w \cdot v = u_1' \cdot u_2'$ where $u_1' \in C, u_2' \in C^*$. Since u was assumed to be a shortest word having the given property, $u_1 \neq u_1'$.

From the OSC we know $u_1 \cdot W' \cdot X^\omega \cap u_1' \cdot W' \cdot X^\omega = \emptyset$. On the other hand we have $u \cdot w \cdot v \cdot w \cdot X^\omega \subseteq u \cdot w \cdot X^\omega \subseteq u_1 \cdot W \cdot X^\omega$ and, since $C^i \cdot W' \cdot X^\omega \subseteq W' \cdot X^\omega$, also $u \cdot w \cdot v \cdot w \cdot X^\omega = u_1' \cdot u_2' \cdot w \cdot X^\omega \subseteq u_1' \cdot W' \cdot X^\omega$, contradicting the fact that W' is an OSC-witness for C.

Q.E.D.

The converse of Theorem 4 is not true.

Example 2 *Consider the finite code $C_2 := \{a, ab, bc, bbcc\}$. Then ccc is not a subword of any word in C_2^*, hence a simplifying nosubword. But we have $a \cdot bbcc \cdot ccc \cdot \{a, b, c\}^\omega \subseteq ab \cdot bc \cdot ccc \cdot \{a, b, c\}^\omega$.*
It should be added that this code has also the simplifying words $bbcc \in C_2$ and $bccc \notin \mathbf{A}(C_2^*)$ for which the respective languages $C_2^* \cdot bbcc$ and $C_2^* \cdot bccc$ are OSC-witnesses. Since $a \cdot bbcc \cdot \{a, b, c\}^\omega \supseteq ab \cdot bccc \cdot \{a, b, c\}^\omega$, their union $C_2^* \cdot \{bbcc, bccc\}$, however, is not an OSC-witness for C_2.

From the above Example 2 one might get the impression that the inconsistencies between OSC and simplifying words are caused by those simplifying words which are not in $\mathbf{A}(C^*)$. We are going to show that this is indeed the case. We start with the following.

Lemma 5 *If $w \in \mathbf{A}(C^*)$ is a simplifying word for the code C then $W := C^* \cdot w$ is an OSC-witness for C.*

Proof. By way of contradiction assume $u \cdot C^* \cdot w \cdot X^\omega \cap u' \cdot C^* \cdot w \cdot X^\omega \neq \emptyset$ for some $u, u' \in C, u \neq u'$. Then there are $v, v' \in C^*$ such that $u \cdot v \cdot w \cdot X^\omega \supseteq u' \cdot v' \cdot w \cdot X^\omega$, say. If $u \cdot v = u' \cdot v'$ then C is not a code, so let $|u \cdot v| < |u' \cdot v'|$. Now consider w' with $w \cdot w' \in C^*$. Then $u' \cdot v' \cdot w \cdot w' = u \cdot v \cdot w \cdot w''$ for a suitable $w'' \in X^*$. Since w is simplifying, we obtain $w \cdot w'' \in C^*$ contradicting again the code property of C.

Q.E.D.

Simplifying prefixes (words in $\mathbf{A}(C^*)$) indicate in certain sense a strong fulfillment of the OSC.

Theorem 6 *Let $V \subseteq X^*$. Then V satisfies the SOSC if and only if V is code which has a simplifying word $w \in \mathbf{A}(V^*)$.*

Proof. If V satisfies the SOSC then it has an OSC-witness W such that $V^\omega \cap W \cdot X^\omega \neq \emptyset$. Consequently, $W \cap \mathbf{A}(V^*) \neq \emptyset$, and the assertion follows from Theorem 4.
Conversely, the OSC-witness $W := V^* \cdot w$ constructed in Lemma 5 satisfies the additional condition $V^\omega \cap W \cdot X^\omega \neq \emptyset$.

Q.E.D.

Similarly, it can be shown that the union of OSC-witnesses for a code C is again an OSC-witness provided they contain only words belonging to $\mathbf{A}(C^*)$.

Theorem 7 *Let C be a code and let $\left(W_i\right)_{i=1}^{\infty}$ be a family of OSC-witnesses for C such that $\bigcup_{i=1}^{\infty} W_i \subseteq \mathbf{A}(C^*)$. Then $\bigcup_{i=1}^{\infty} W_i$ is also an OSC-witness for C.*

Proof. We proceed in two steps. First we show that $W_1 \cup W_2$ is an OSC-witness for C, thus proving the theorem for finite unions, and then we prove it for ascending chains $W_1 \subseteq W_2 \subseteq \ldots$.
Suppose that $W_1, W_2 \subseteq \mathbf{A}(C^*)$ are OSC-witnesses for C and suppose $v \cdot (W_1 \cup W_2) \cdot X^\omega \cap v' \cdot (W_1 \cup W_2) \cdot X^\omega \neq \emptyset$ for some $v, v' \in C$. Since W_1 and W_2 both are OSC-witnesses for C, we obtain that $v \cdot w_1 \cdot X^\omega \subseteq v' \cdot w_2 \cdot X^\omega$ for $w_1 \in W_1$ and $w_2 \in W_2$, say. Choose w' such that $w_1 \cdot w' \in C^*$. Then there is w'' satisfying $v \cdot w_1 \cdot w' = v' \cdot w_2 \cdot w'' \in C^*$. According to Theorem 4 w_1 and w_2 both are simplifying words for the code C. This yields $v = v'$.
Now we may assume that the languages in the family $(W_i)_{i=1}^{\infty}$ form a chain $W_1 \subseteq W_2 \subseteq W_3 \subseteq \ldots$. If $v \cdot W \cdot X^\omega \cap v' \cdot W \cdot X^\omega \neq \emptyset$ for $W := \bigcup_{i=1}^{\infty} W_i$ and some $v, v' \in C$ then $v \cdot W_i \cdot X^\omega \cap v' \cdot W_j \cdot X^\omega \neq \emptyset$ for i, j. If $i \geq j$ this implies $v \cdot W_i \cdot X^\omega \cap v' \cdot W_i \cdot X^\omega \neq \emptyset$ and thus $v = v'$.

Q.E.D.

As a consequence of Theorem 7 we obtain that provided a code C has a simplifying prefix, or equivalently, satisfies the SOSC, its set of simplifying prefixes, $\hat{S}_C$, is an OSC-witness for C. As we have seen in Example 1 a code may have no simplifying word, and as shall see below the set $\hat{S}_C$ might be empty even if the code C has simplifying words. This latter example shows that there are indeed languages satisfying the OSC but not the SOSC.

Example 3 *Let* $C_3 := \{waab^{|w|} : w \in \{aa, bb\}^*\}$*. Similar to Example 1,* C_3 *is a suffix code. As* $C_3^* \cdot aba \cap \mathbf{A}(C_3^*) = \emptyset$*,* C_3 *has OSC-witness* $C_3^* \cdot aba$*. By construction* $\mathbf{A}(C_3) = \{aa, bb\}^* \cdot \{e, a, b\}$ *and* $C_3^* \subseteq \{aa, bb\}^*$*. Consequently,* $C_3 \cdot W \subseteq \mathbf{A}(C_3)$ *for arbitrary* $W \subseteq \mathbf{A}(C_3^*)$*, which according to Claim 2 implies that no OSC-witness* W *of* C_3 *can contain a word in* $\mathbf{A}(C_3^*)$*.*

Finally, as an illustration, we present two classes of codes satisfying the SOSC.
First we consider regular languages, that is, languages accepted by finite automata.

Theorem 8 *Let* V *be a nonempty regular language. Then the following conditions are equivalent:*

1. *V is a code.*
2. *V satisfies the OSC.*
3. *V satisfies the SOSC.*

Proof. In view of Lemma 1 it remains to verify that a regular code V has a simplifying prefix $w \in \mathbf{A}(V^*)$, which is shown in [LPRT79, Proposition 4.1].

Q.E.D.

Remark. For finite codes the equivalence of 2. and 3. follows also from [Sc94].

As a second class we consider codes of bounded decoding delay (cf. [LPRT79], [St86]). The result follows from [LPRT79] and the above verified fact that if $\hat{S}_C$ is nonempty then it is an OSC-witness for C. We give here a direct proof.

Lemma 9 *Let* C *be a code of bounded decoding delay. Then* C *satisfies the SOSC.*

Proof. Let C have the decoding delay $m \geq 0$. Then according to [St86, Lemma 6] for all $v, v' \in C$ with $v \neq v'$ the sets $v \cdot C^m \cdot X^\omega$ and $v' \cdot C^m \cdot X^\omega$ are disjoint. Hence, $W := C^m$ is an OSC-witness for C.

Q.E.D.

Acknowledgement

I would like to thank Véronique Bruyère for drawing my attention to simplifying words and the paper [LPRT79].

References

[BG92] C. Bandt and S. Graf, *Self-similar sets 7: A characterization of self-similar sets with positive Hausdorff measure.* Proc. Amer. Math. Soc. **114** (1992), 995 – 1001.

[Ba88] M.R. Barnsley, *"Fractals Everywhere"*, Academic Press, Orlando, 1988.

[BP85] J. Berstel and D. Perrin, *Theory of Codes.* Academic Press, Orlando 1985.

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

[LPRT79] A. De Luca, D. Perrin, A. Restivo, and S. Termini, *Synchronization and simplification.* Discrete Mathematics **27** (1979), 297 – 308.

[Sc94] A. Schief, *Separation properties for self-similar sets.* to appear in: Proc. Amer. Math. Soc.

[St86] L. Staiger, *On infinitary finite length codes.* RAIRO Informatique théor. et Appl. **20** (1986), 483 – 494.

[St92] L. Staiger, *Hausdorff dimension of constructively specified sets and applications to image processing.* in: "Topology, Measures, and Fractals" (C. Bandt, J. Flachsmeyer and H. Haase Eds.), Mathematical Research No. **66**, Akademie-Verlag, Berlin 1992, pp. 109 – 120.