Dissecting Power of a Finite Intersection of
Context Free Languages

Josef Rukavicka∗

June 13, 2020
Mathematics Subject Classification: 68Q45

Abstract

Let \( \exp^{k,\alpha} \) denote a tetration function defined as follows: \( \exp^{1,\alpha} = 2^\alpha \) and \( \exp^{k+1,\alpha} = 2^{\exp^{k,\alpha}} \), where \( k, \alpha \) are positive integers. Let \( \Delta_n \) denote an alphabet with \( n \) letters. If \( L \subseteq \Delta_n^* \) is an infinite language such that for each \( u \in L \) there is \( v \in L \) with \( |v| \leq \exp^{k,\alpha} |u| \) then we call \( L \) a language with the growth bounded by \((k, \alpha)\)-tetration.

Given two infinite languages \( L_1, L_2 \in \Delta_n^* \), we say that \( L_1 \) dissect \( L_2 \) if \( |L_1 \cap L_2| = \infty \) and \( |(\Delta_n^* \setminus L_1) \cap L_2| = \infty \).

Let \( L \subseteq \Delta_n^* \) be an infinite language with the growth bounded by \((k, \alpha)\)-tetration, where \( k \geq 2 \). We show that there are \( 3k - 3 \) context free languages \( L_1, L_2, \ldots, L_{3k-3} \) over the alphabet \( \Sigma_{2k-1} \) with \( 2k-1 \) letters, an erasing alphabetical homomorphism \( \upsilon : \Sigma_{2k-1}^* \rightarrow \Delta_1^* \), a nonerasing alphabetical homomorphism \( \pi : \Delta_n^* \rightarrow \Delta_1^* \) such that the homomorphic image \( \upsilon(\bigcap_{i=1}^{3k-3} L_i) \) dissects the homomorphic image \( \pi(L) \).

1 Introduction

In the theory of formal languages, the regular and the context free languages constitute a fundamental concept that attracted a lot of attention in the past several decades. Recall that every regular language is accepted by some

∗Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague (josef.rukavicka@seznam.cz).
deterministic finite automaton and every context free language is accepted by some pushdown automaton.

In contrast to regular languages, the context free languages are closed neither under intersection nor under complement. The intersection of context free languages have been systematically studied in [4, 6, 8, 9, 10]. Let $\text{CFL}_k$ denote the family of all languages such that for each $L \in \text{CFL}_k$ there are $k$ context free languages $L_1, L_2, \ldots, L_k$ with $L = \bigcap_{i=1}^{k} L_i$. For each $k$, it has been shown that there is a language $L \in \text{CFL}_{k+1}$ such that $L \notin \text{CFL}_k$. Thus the $k$-intersections of context free languages form an infinite hierarchy in the family of all formal languages lying between context free and context sensitive languages [6].

One of the topics in the theory of formal languages that has been studied is the dissection of infinite languages. Let $\Delta_n$ be an alphabet with $n$ letters, and let $L_1, L_2 \subseteq \Delta_n^*$ be infinite languages. We say that $L_1$ dissects $L_2$ if $|L_1 \cap L_2| = \infty$ and $|\Delta_n^* \setminus L_1 \cap L_2| = \infty$. Let $\mathcal{C}$ be family of languages. We say that a language $L_1 \in \Delta_n^*$ is $\mathcal{C}$-dissectible if there is a $L_2 \in \mathcal{C}$ such that $L_2$ dissects $L_1$. Let $\text{REG}$ denote the family of regular languages. In [11] the $\text{REG}$-dissectibility has been investigated. Several families of $\text{REG}$-dissectible languages have been presented. Moreover it has been shown that there are infinite languages that cannot be dissected with a regular language. Also some open questions for $\text{REG}$-dissectibility can be found in [11]. For example it is not known if the complement of a context free languages is $\text{REG}$-dissectible.

There is a related longstanding open question in [11]: Given two context free languages $L_1, L_2 \subseteq \Delta_n^*$ such that $L_1 \subseteq L_2$ and $L_2 \setminus L_1$ is an infinite set, is there a context free language $L_3$ such that $L_3 \subseteq L_2$, $L_1 \subseteq L_3$, and both the languages $L_3 \setminus L_1$ and $L_2 \setminus L_3$ are infinite? This question was mentioned also in [11].

Some other results concerning the dissection of infinite languages may be found in [5]. A similar topic is the finding of minimal covers of languages; see [2]. Recall that a language $L_1 \subseteq \Delta_n^*$ is called $\mathcal{C}$-immune if there is no infinite language $L_2 \subseteq L_1$ such that $L_2 \in \mathcal{C}$. The immunity is also related to the dissection of languages; some results on this theme can be found in [3, 7, 10].

Let $\mathbb{N}$ denote the set of all positive integers. An infinite language $L \subseteq \Delta_n^*$ is called constantly growing, if there is a constant $c_0 \in \mathbb{N}$ and a finite set $K \subseteq \mathbb{N}$ such that for each $w \in L$ with $|w| \geq c_0$ there is a word $\bar{w} \in L$ and a constant $c \in K$ such that $|\bar{w}| = |w| + c$. We say also that $L$ is $(c_0, K)$-
constantly growing. In [11], it has been proved that every constantly growing language \( L \) is \( \text{REG} \)-dissectible.

We define a tetration function (a repeated exponentiation) as follows:

\[
\exp^{j,\alpha} = 2^{\alpha} \quad \text{and} \quad \exp^{j+1,\alpha} = 2^{\exp^{j,\alpha}},
\]

where \( j \in \mathbb{N} \). The tetration function is known as a fast growing function. If \( k, \alpha \) are positive integers and \( L \subseteq \Delta_n^* \) is an infinite language such that for each \( u \in L \) there is \( v \in L \) with \( |v| \leq \exp^{k,\alpha} |u| \) then we call \( L \) a language with the growth bounded by \((k, \alpha)\)-tetration.

Let \( L \subseteq \Delta_n^* \) be an infinite language with the growth bounded by \((k, \alpha)\)-tetration, where \( k \geq 2 \). In the current article we show that there are:

- an alphabet \( \Sigma_{2k-1} \) with \( |\Sigma_{2k-1}| = 2k - 1 \),
- an erasing alphabetical homomorphism \( \nu : \Sigma_{2k-1}^* \to \Delta_1^* \),
- a nonerasing alphabetical homomorphism \( \pi : \Delta_n^* \to \Delta_1^* \), and
- 3\(k-3\) context free languages \( L_1, L_2, \ldots, L_{3k-3} \subseteq \Sigma_{2k-1}^* \) such that the homomorphic image \( \nu(\bigcap_{i=1}^{3k-3} L_i) \) dissects the homomorphic image \( \pi(L) \).

We sketch the basic ideas of our proof. Recall that a non-associative word on the letter \( z \) is a “well parenthesized” word containing a given number of occurrences of \( z \). It is known that the number of non-associative words containing \( n+1 \) occurrences of \( z \) is equal to the \( n \)-th Catalan number. For example for \( n = 3 \) we have 5 distinct non-associative words: \(((zz)z)z\), \(((z)z)z\), \((z(z)z)\), \((zz)(z)z\), and \(((z)zz)z\). Every non-associative word contains the prefix \( z^k \) for some \( k \in \mathbb{N} \), where \( z^k \) denotes the \( k \)-th power of the opening bracket. It is easy to verify that there are non-associative words such that \( k \) equals “approximately” \( \log_2 n \). We construct three context free languages whose intersection accepts such words; we call these words balanced non-associative words. By counting the number of opening brackets of a balanced non-associative word with \( n \) occurrences of \( z \) we can compute a logarithm of \( n \).

Let \( \log_2^{(1)} n = \log_2 n \) and \( \log_2^{(j+1)} n = \log_2^{(j)} (\log_2 n) \). Our construction can be “chained” so that we construct 3\(k-3\) context free languages, whose intersection accepts words with \( n \) occurrences of \( z \) and a prefix \( x^j \bar{z} \), where \( j \) is equal “approximately” to \( \log_2^{(k)} n \) and \( \bar{z} \neq x \). If \( L \) is a language with the growth bounded by a \((k, \alpha)\)-tetration then the language \( \bar{L} = \{z^j \mid j = \ldots \} \)
of dissecting a language with the growth bounded by \([\log_2^k |w|]\) and \(w \in L\) is constantly growing. Less formally said, by means of intersection of \(3k - 3\) context free languages we transform the challenge of dissecting a language with the growth bounded by \((k, \alpha)\)-tetration to the challenge of dissecting a constantly growing language. This approach allows us to prove our result.

2 Preliminaries

Let \(\mathbb{R}^+\) denote the set of all positive real numbers.

Let \(B_k = \{x_1, x_2, \ldots, x_k\}\) be an ordered alphabet (set) of \(k\) distinct opening brackets, and let \(B_k = \{y_1, y_2, \ldots, y_k\}\) be an ordered alphabet (set) of \(k\) distinct closing brackets. We define the alphabet \(\Sigma_{2k-1} = B_k \cup (B_k \setminus \{y_1\})\). The alphabet \(\Sigma_{2k-1}\) contains all opening brackets \(B_k\) and all the closing brackets without the the first one \(B_k \setminus \{y_1\}\). It follows that \(|\Sigma_{2k-1}| = 2k - 1\).

Let \(\epsilon\) denote the empty word. Given a finite alphabet \(S\), let \(S^+\) denote the set of all finite nonempty words over the alphabet \(S\) and let \(S^* = S^+ \cup \{\epsilon\}\).

Given a finite alphabet \(S\), let \(\text{occur}(w, t)\) denote the number of occurrences of the nonempty factor \(t \in S^+\) in the word \(w \in S^*\). Let \(\text{Fac}(w)\) denote the set of all factors a word \(w \in S^*\). We define that \(\epsilon, w \in \text{Fac}(w)\); i.e. the empty word and the word \(w\) are factors of \(w\). Let \(\text{Pref}(w) \subseteq \text{Fac}(w)\) denote the set of all prefixes of \(w \in S^*\). We define that \(\epsilon, w \in \text{Pref}(w)\). Let \(\text{Suf}(w) \subseteq \text{Fac}(w)\) denote the set of all suffixes of \(w \in S^*\). We define that \(\epsilon, w \in \text{Suf}(w)\).

Given two finite alphabets \(S_1, S_2\), a homomorphism from \(S_1^*\) to \(S_2^*\) is a function \(\tau : S_1^* \to S_2^*\) such \(\tau(ab) = \tau(a)\tau(b)\), where \(a, b \in S_1^+\). It follows that in order to define a homomorphism \(\tau\), it suffices to define \(\tau(z)\) for every \(z \in S_1\); such definition “naturally” extends to every word \(a \in S_1^*\). We say that \(\tau\) is an nonerasing alphabetical homomorphism if \(\tau(z) \in S_2\) for every \(z \in S_1\).

We say that \(\tau\) is an erasing alphabetical homomorphism if \(\tau(z) \in S_2 \cup \{\epsilon\}\) for every \(z \in S_1\) and there is at least one \(z \in S_1\) such that \(\tau(z) = \epsilon\).

3 Balanced non-associative words

Suppose \(k, m \in \mathbb{N}\), where \(k, m \geq 2\), and \(k \geq m\). To simplify the notation we define \(x = x_m, y = y_m\), and \(z = x_{m-1}\); it means that \(x\) denotes the \(m\)-th opening bracket, \(y\) denotes the \(m\)-th closing bracket, and \(z\) denotes the
Let $\mu_{k,m} : \Sigma_{2k-1}^* \rightarrow \Sigma_{2k-1}^*$ be an erasing alphabetical homomorphism defined as follows:

- $\mu_{k,m}(z) = z$,
- $\mu_{k,m}(x) = x$,
- $\mu_{k,m}(y) = y$,
- $\mu_{k,m}(a) = \epsilon$, where $a \in \Sigma_{2k-1} \setminus \{x, y, z\}$.

Let $L \subseteq \Sigma_{2k-1}^*$ be a language, we define the language $\mu_{k,m}(L) = \{\mu_{k,m}(w) \mid w \in L\}$.

**Remark 3.1.** For given $k, m$ the erasing alphabetical homomorphism $\mu_{k,m}$ sends all opening and closing brackets from $B_k$ and $B_k$ to the empty string with the exception of $x$, $y$, and $z$.

Let $\text{Naw}_{k,m} \subseteq \Sigma_{2k-1}^*$ be the context free language generated by the following context free grammar, where $S$ is a start non-terminal symbol, $N$ is a non-terminal symbol, and $x, y, z, a$ are terminal symbols (the letters from $\Sigma_{2k-1}^*$).

- $S \rightarrow N x N S S y N \mid N x N z N y N \mid N x N z N z N y N$,
- $N \rightarrow a N \mid \epsilon$, where $a \in \Sigma_{2k-1} \setminus \{x, y, z\}$.

We call the words from $\text{Naw}_{k,m}$ non-associative words over the opening bracket $x$, the closing bracket $y$, and the letter $z$.

**Remark 3.2.** Let $M = \mu_{k,m}(\text{Naw}_{k,m})$. To understand the definition of $\text{Naw}_{k,m}$, note that the language $M$ is generated by the context free grammar defined by: $S \rightarrow x S S y \mid x z y \mid x z z y$. To see this, just remove the non-terminal symbol $N$ in the definition of $\text{Naw}_{k,m}$. The usage of the non-terminal symbol $N$ allows to “insert” between any two letters of a word from $\mu_{k,m}(\text{Naw}_{k,m})$ the words from $K = (\Sigma_{2k-1} \setminus \{x, y, z\})^*$; the set $K$ contains words from $\Sigma_{2k-1}^*$ that have no occurrence of $x, y, z$. It means that if $w = w_1 w_2 \ldots w_n \in \mu_{k,m}(\text{Naw}_{k,m})$, then $t_0 w_1 t_1 w_2 t_2 \ldots t_{n-1} w_n t_n \in \text{Naw}_{k,m}$, where $w_i \in \{x, y, z\}$ and $t_i \in K$.

The reason for the name “non-associative words” is the obvious similarity between the words from $M$ and the “standard non-associative words” mentioned in the introduction section. For our purpose it is necessary that $w_1 x z y w_2 \in M$ if and only if $w_1 x z z y w_2 \in M$ for every $w_1, w_2 \in \{x, z, y\}^*$.
Recall that a pushdown automaton is a 6-tuple \((Q, \Delta, \Gamma, q_0, S, \delta)\), where

- \(Q\) is a set of states,
- \(\Delta\) is an input alphabet,
- \(\Gamma\) is a stack alphabet,
- \(q_0 \in Q\) is an input state,
- \(S \in \Gamma\) is the initial symbol of the stack,
- \(\delta : (Q \times \Delta \times \Gamma) \rightarrow (Q, \Gamma^*)\) is a transition function.

We define that a pushdown automaton accepts a word by the empty stack, hence we do not need to define the set of final states. Given a pushdown automaton \(g\), let \(AL(g) \subseteq \Delta^*\) denotes the language accepted by \(g\).

Let \(\Lambda_{k,m} = AL(g_{k,m}) \subseteq \Sigma_{2k-1}^2\) denote the language accepted by the pushdown automaton \(g_{k,m} = (Q, \Sigma_{2k-1}, \Gamma, q_S, S, F, \delta)\), where:

- \(Q = \{q_S, q_B, q_0, q_x, q_r\}\),
- \(\Gamma = \{S, X\}\),
- \(\delta(q, a, u) \rightarrow (q, u)\), where \(q \in Q, u \in \Gamma\), and \(a \in \Sigma_{2k-1} \setminus \{x, y, z\}\),
- \(\delta(q_S, x, u) \rightarrow (q_B, u)\), where \(u \in \Gamma\),
- \(\delta(q_S, z, u) \rightarrow (q_S, u)\), where \(u \in \Gamma\),
- \(\delta(q_S, y, u) \rightarrow (q_S, u)\), where \(u \in \Gamma\),
- \(\delta(q_B, x, u) \rightarrow (q_x, uXX)\), where \(u \in \Gamma\),
- \(\delta(q_B, z, u) \rightarrow (q_S, u)\), where \(u \in \Gamma\),
- \(\delta(q_B, y, u) \rightarrow (q_S, u)\), where \(u \in \Gamma\),
- \(\delta(q_0, x, u) \rightarrow (q_x, u)\), where \(u \in \Gamma\),
- \(\delta(q_0, z, u) \rightarrow (q_0, u)\), where \(u \in \Gamma\),
- \(\delta(q_0, y, u) \rightarrow (q_0, u)\), where \(u \in \Gamma\),
Remark 3.3. Note that the letters from $\Sigma_{2k-1} \setminus \{x, y, z\}$ change neither the state of $g_{k,m}$ nor the stack. Hence to illuminate the behavior of $g_{k,m}$, we can consider only words over the alphabet $\{x, y, z\}$. Then it is easy to see that the pushdown automaton $g_{k,m}$ pushes $XX$ on the stack on the first occurrence of $xx$. For every other occurrence of $xx$ the pushdown automaton $g_{k,m}$ pushes $X$ on the stack. Once reached the state $q_x$, then for every occurrence of $xz$ one $X$ is removed from the stack. The state $q_r$ works as a refuse state. Note that after reaching the state $q_r$ the stack is not empty, the stack cannot be changed, and no other state can be reached from $q_r$. The states $q_S$ and $q_B$ enable to recognize the first occurrence of $xx$. Once the states $q_x$ are reached, the states $q_S$ and $q_B$ can not be reached any more.

Thus the pushdown automaton $g_{k,m}$ accepts all words, where the number of occurrences of $xz$ after the first occurrence of $xx$ is exactly one more than the number of occurrences of $xx$. Formally, if $w \in \mu_{k,m}(\Sigma_{2k-1})$ then we define $\bar{w}$ as follows:

- If $\text{occur}(w, xx) = 0$ then $\bar{w} = \epsilon$.
- If $\text{occur}(w, xx) \geq 1$ then let $\bar{w} \in \text{Suf}(w)$ be such that $xx \in \text{Pref}(\bar{w})$ and $\text{occur}(\bar{w}, xx) = \text{occur}(w, xx)$.

Clearly $\bar{w}$ is uniquely defined. Then we have that $w \in \mu_{k,m}(\Lambda_{k,m})$ if and only if $\bar{w} = \epsilon$ or $\text{occur}(\bar{w}, xx) + 1 = \text{occur}(\bar{w}, xz)$. It follows that the words without any occurrence of $xx$ are accepted. In the following we will consider the words from the intersection $U = \Lambda_{k,m} \cap \text{Naw}_{k,m}$. Note that there are only two nonempty words $xzy, xzzy \in U$, that have no occurrence of $xx$.

Recall that a “standard” non-associative word can be represented as a full binary rooted tree graph, where every inner node represents a corresponding pair of brackets and every leaf represents the letter $z$. It is know that the number of inner nodes plus one is equal to the number of leaves in a full binary
rooted tree graph. In the case of non-associative words from $Naw_{k,m}$, let the leaves represent the factors $xzy$ and $xzzy$. Then the number of occurrences of $xz$ is equal to the number of leaves and the number of occurrences of $xx$ is equal to the number of inner nodes. Hence the intersection $M \cap Naw_{k,m}$ contains non-associative words that have no “unnecessary” brackets; for example $xzzy, xxzzyy, xxxzzyyy \in Naw_{k,m}$, $xzzy \in M$ and $xxzzyy, xxxzzyyy \notin M$.

Let $Balk,m \subseteq \Sigma^{*}_{2k-1}$ be the context free language generated by the following context free grammar, where $S$ is a start non-terminal symbol, $N, K, V, P$ are non-terminal symbols, and $x, y, z, a$ are terminal symbols (the letters from $\Sigma^{*}_{2k-1}$).

- $S \rightarrow KV P,$
- $V \rightarrow VV | NzN | NzTzN | \epsilon,$
- $T \rightarrow NyNTNxN | \epsilon,$
- $K \rightarrow KK | NxN | \epsilon,$
- $P \rightarrow PP | NyN | \epsilon,$
- $N \rightarrow aN | \epsilon$, where $a \in \Sigma^{*}_{2k-1}\{x, y, z\}$.

We call the words from $Balk,m$ balanced words.

**Remark 3.4.** Let $M = \mu_{k,m}(Balk,m)$. It is easy to see that the words from the language $M$ contains no factor of the form $zy_{i}x_{j}z$, where $i, j$ are distinct positive integers; hence the name “balanced” words. The non-terminal symbols $K, P$ enable that if $w \in M$ then $w$ has a prefix $x^{i}$ and a suffix $y^{j}$ for some $i, j \in \mathbb{N} \cup \{0\}$.

The non-terminal symbol $N$ in the definition of $Balk,m$ has the same purpose like in the definition of $Naw_{k,m}$.

Let

$$\Omega_{k,m} = Naw_{k,m} \cap Balk,m \cap \Lambda_{k,m}.$$ 

We call the words from $\Omega_{k,m}$ balanced non-associative words over the opening bracket $x$, the closing bracket $y$, and a letter $z$.

Let $\Omega_{k,m}(n) = \{w \in \Omega_{k,m} | occur(w, z) = n\}$, where $n \in \mathbb{N}$. The set $\Omega_{k,m}(n)$ contains the balanced non-associative words having exactly $n$ occurrences of the letter $z$. 


Given a word \( w \in \Sigma_{2k-1}^* \) and \( a \in \Sigma_{2k-1} \), let
\[
\text{height}(w, a) = \max\{ j \mid a^j \in \text{Fac}(w) \}.
\]
The height of a word \( w \) is the maximal power of the letter \( a \), that is \( a \) is a factor of \( w \). We show that if \( w \in \mu_{k,m}(\Omega_{k,m}) \) and \( h \) is the height the opening bracket \( x \) in \( w \) then \( x^h \) is a prefix of \( w \) and \( y^h \) is a suffix of \( w \).

**Lemma 3.5.** If \( w \in \mu_{k,m}(\Omega_{k,m}) \) and \( h = \text{height}(w, x) \) then \( x^h \in \text{Pref}(w) \) and \( y^h \in \text{Suf}(w) \).

**Proof.** Note that \( \mu_{k,m}(\Omega_{k,m}) \subseteq \Omega_{k,m} \). Since \( \Omega_{k,m} \subseteq \text{Naw}_{k,m} \), there is \( \hat{h} \in \mathbb{N} \) such that \( x^h z \in \text{Pref}(w) \). To get a contradiction suppose that \( \hat{h} < h \). Because \( \Omega_{k,m} \subseteq \text{Bal}_{k,m} \) it follows that \( w = x^h w_1 zy^h x^h z w_2 \) for some \( w_1 \in \text{Fac}(w) \) with \( z \in \text{Pref}(w_1 z) \) and \( w_2 \in \text{Suf}(w) \).

Consider the prefix \( r = x^h w_1 z y^h \). Obviously \( w_1 z \in \mu_{k,m}(\text{Bal}_{k,m}) \). It is easy to see that if \( v \in \mu_{k,m}(\text{Bal}_{k,m}) \), \( x \notin \text{Pref}(v) \) and \( y \notin \text{Suf}(v) \) then \( \text{occur}(v, x) = \text{occur}(v, y) \). Thus \( \text{occur}(w_1 z, x) = \text{occur}(w_1 z, y) \). It follows that \( \text{occur}(r, x) < \text{occur}(r, y) \).

This is a contradiction, since for every prefix \( v \in \text{Pref}(w) \) of a non-associative word \( w \in \text{Naw}_{k,m} \) we have that \( \text{occur}(v, x) \geq \text{occur}(v, y) \). We conclude that \( \hat{h} = h \) and \( x^h \in \text{Pref}(w) \). In an analog way we can show that \( y^h \in \text{Suf}(w) \). This completes the proof. \( \square \)

For a word \( w \in \mu_{k,m}(\Omega_{k,m}) \), we show the relation between the height of \( w \) and the number of occurrences of \( z \) in \( w \).

**Proposition 3.6.** If \( w \in \mu_{k,m}(\Omega_{k,m}) \) and \( h = \text{height}(w, x) \) then
\[
2^{h-1} \leq \text{occur}(w, z) \leq 2^h.
\]

**Proof.** We prove the proposition for all \( h \) by induction:

- If \( h = 0 \) then \( w = \epsilon \).
- If \( h = 1 \) then \( w \in \{ xzzy, xzy \} \).
- If \( h = 2 \) then \( w \in \{ xxzyxzzy, xxzzyxzzy, xzzyxzzyy, xzzyxzzy, xzzyxzzyy, xzzyxzzyy \} \).
Thus the proposition holds for $h \leq 2$ and clearly we have that if $h \geq 2$ then $w = xw_1w_2y$, where $w_1, w_2 \in \mu_{k,m}(\Omega_{k,m})$. Suppose the proposition holds for all $h \leq 2$. We prove the proposition holds for $h$.

Let $h_1 = \text{height}(w_1, x)$ and $h_2 = \text{height}(w_2, x)$. Lemma 3.5 implies that $x^{h_1} \in \text{Pref}(w_1)$, $y^{h_1} \in \text{Suf}(w_1)$, $x^{h_2} \in \text{Pref}(w_2)$, and $y^{h_2} \in \text{Suf}(w_2)$. Since $w \in \mu_{k,m}(\text{Bal}_{k,m})$ it follows that $h_1 = h_2$. Because $x^{h_1} \in \text{Pref}(w_1)$ we have that $x^{h_1+1} \in \text{Pref}(w)$. Clearly $\text{occur}(w, x^{h_1+1}) = 1$; note that $\text{occur}(w_1w_2, x^{h_1+1}) = 0$. Thus $h_1 + 1 = h$. For we assumed that the proposition holds for all $h < h$, we can derive that

$$\text{occur}(w, z) = \text{occur}(w_1, z) + \text{occur}(w_2, z) \leq 2^{h_1} = 2^{h_1+1} = 2^h$$

and

$$\text{occur}(w, z) = \text{occur}(w_1, z) + \text{occur}(w_2, z) \leq 2^{h_1} - 1 + 2^{h_1} - 1 = 2^{h_1} = 2^{h-1}.$$ 

This completes the proof. 

Proposition 3.6 have the following obvious corollary.

**Corollary 3.7.** If $n \in \mathbb{N}$, $w \in \mu_{k,m}(\Omega_{k,m}(n))$, and $h = \text{height}(w, x)$ then

$$\log_2 n \leq h \leq 1 + \log_2 n.$$ 

Given $w, u, v \in \Sigma^+_{2k-1}$, let replace$(w, v, u)$ denote the word built from $w$ by replacing the first occurrence of $v$ in $w$ by $u$. Formally, if $\text{occur}(w, v) = 0$ then replace$(w, v, u) = w$. If $\text{occur}(w, v) = j > 0$ and $w = wjuv$, where $\text{occur}(vu, v) = j$ then replace$(w, v, u) = wuuv$.

We prove that the set of balanced non-associative words $\Omega_{k,m}(n)$ having $n$ occurrences of $z$ is nonempty for each $n \in \mathbb{N}$.

**Lemma 3.8.** If $n \in \mathbb{N}$ then $\Omega_{k,m}(n) \neq \emptyset$.

**Proof.** If $n = 1$ then $xzy \in \Omega_{k,m}(1)$. Given $n \in \mathbb{N}$ with $n > 1$, let $j \in \mathbb{N}$ be such that $2^{j-1} < n \leq 2^j$. Obviously such $j$ exists and is uniquely determined. Let $w_1 = xzy$. Let $w_{i+1} = xw_iw_iy$. Clearly $\text{occur}(w_j, z) = 2^j$ and $w_j \in \Omega_{k,m}(2^j)$. Note that $\text{occur}(w_j, xzy) = 2^{j-1}$. Let $w_{j,0} = w_j$ and $w_{j,i+1} = \text{replace}(w_{j,i}, xzy, xzy)$. Let $\alpha = 2^j - n$. Then one can easily verify that $\text{occur}(w_{j,\alpha}, z) = n$ and $w_{j,\alpha} \in \Omega_{k,m}(n)$.

Less formally said, we construct a balanced non-associative word $w_j$ having $2^{j-1}$ occurrences of $xzy$ and then we replace a given number of occurrences of $xzy$ with the factor $xyz$ to achieve the required number of occurrences of $z$. This completes the proof. 

10
4 Intersection of balanced non-associative words

Let $\Omega_k = \bigcap_{m=2}^k \Omega_{k,m}$ and let $\Omega_k(n) = \{w \in \Omega_k \mid \text{occur}(w, x_1) = n\}$. We show that for all positive integers $n, k$ with $k \geq 2$ there is a word $w \in \Omega_k$ such that $w$ has $n$ occurrences of the opening bracket $x_1$.

Proposition 4.1. If $k, n \in \mathbb{N}$ and $k \geq 2$ then $\Omega_k(n) \neq \emptyset$.

Proof. Let $h(1) = n$. Let $w_i \in \mu_{k,i}(\Omega_{k,i}(h(i-1)))$ and let $h(i) = \text{height}(w_i, x_i)$, where $i \in \{2, 3, 4, \ldots, k\}$. Lemma 3.8 implies that such $w_i$ exist.

Let $v_2 = w_2$. Let $v_{j+1} = \text{replace}(v_j, x_j^{h(j)}, w_{j+1})$. Note that $\mu_{k,j}(v_j + 1) = \mu_{k,j}(v_j)$. Then it is quite straightforward to see that $v_k \in \Omega_k$ and $\text{occur}(v_k, x_1) = n$. This completes the proof. \qed

To clarify the proof of Proposition 4.1 let us see the following example.

Example 4.2. Let $n = 23$ and $k = 4$. To make the example easy to read, we define $B_4 = \{z, [, <\} and $\bar{B}_4 = \{\bar{z}, ), >\}$. It means that $x_1 = z$, $x_2 = (, x_3 = [, x_4 = <, \bar{x}_1 = \bar{z}, \bar{x}_2 = ), \bar{x}_3 = ]$, and $\bar{x}_4 = >$.

To fit the example into the width of the page, we define auxiliary words $u_1$ and $u_2$:

- $u_1 = z)(z)((z)(z))((z)(z))((z)(z)),$
- $u_2 = (((z)(zz)((zz)(zz))((zz)(zz))((zz)(zz))))$.

Then we have that

- $h(1) = 23; w_2 = ((((u_1u_2; h(2) = 5; w_3 = [[[[]][[]][[][]]])$;
- $w_4 = << [[[> < ]]>; h(4) = 2; v_2 = w_2$;
- $v_3 = [[[[]][[]][[][]]]]u_1u_2$;
- $v_4 = << [>] < [>> ][][]][[][]]u_1u_2$.

This ends the example.

We define two technical functions $\log_2^{(j)} t$ and $\log_2^{[j]} t$ for all $j \in \mathbb{N}$ and $t \in \mathbb{R}^+$ as follows:
• \( \log_2^{(1)} t = \log_2 t \) and \( \log_2^{(j+1)} t = \log_2^{(j)} (\log_2 t) \).

• \( \log_2^{[1]} t = 1 + \log_2 t \) and \( \log_2^{[j+1]} t = \log_2^{[j]} (1 + \log_2 t) \).

It is a simple exercise to prove the following lemma. We omit the proof.

**Lemma 4.3.** If \( j \in \mathbb{N} \) then there is a constants \( c_1 \in \mathbb{R}^+ \) such that for each \( t \in \mathbb{R}^+ \) with \( t \geq 1 \) we have
\[
\log_2^{(j)} t \leq \log_2^{[j]} t \leq c_1 + \log_2^{[j]} t.
\]

**Remark 4.4.** Note in Lemma 4.3 that the constant \( c_1 \) depends on \( j \).

Using the function \( \log_2^{(k)} t \) we present an upper and a lower bound for the height of words from \( \Omega_k \).

**Proposition 4.5.** If \( k \in \mathbb{N}, k \geq 2 \) then there is a constants \( c_1 \in \mathbb{R}^+ \) such for each \( w \in \Omega_k \), \( h = \text{height}(w, x_k) \), and \( n = \text{occur}(w, x_1) \) we have
\[
\log_2^{(k)} n \leq h \leq c_1 + \log_2^{(k)} n.
\]

**Proof.** It follows from Corollary 3.7 that \( \log_2^{(k)} n \leq h \leq \log_2^{[k]} n \). Then the proposition follows from Lemma 4.3. \( \square \)

**Remark 4.6.** Note in Proposition 4.5 that the constant \( c_1 \) depends on \( k \).

## 5 Dissection by a regular language

In [11] it was shown that every constantly growing language can be dissected by some regular language.

**Lemma 5.1.** (see [11, Lemma 3.3]) Every infinite constantly growing language is REG-dissectible.

From the proof of Lemma 3.3 in [11] we can formulate the following Lemma.

**Lemma 5.2.** If \( n, c_0 \in \mathbb{N}, K \subset \mathbb{N}, |K| < \infty, c = \max\{j \in K\}, \) and \( L \subseteq \Delta_n^* \) is a \((c_0, K)\)-constantly growing language then there are \( j_1, j_2 \in \{0, 1, 2, \ldots, c\} \) such that both sets \( H_1, H_2 \) are infinite, where
\[
H_i = \{w \mid w \in L \text{ and } |w| \equiv j_i \pmod{c+1} \} \text{ and } i \in \{1, 2\}.
\]
6 Tetration

Recall that a deterministic finite automaton $g$ is $5$-tuple $(Q, \Delta, q_0, \delta, F)$, where $Q$ is the set of states, $\Delta$ is an input alphabet, $q_0$ is the initial state, $\delta$ is a transition function, and $F$ is the set of accepting states. Let $AL(g)$ denote the language accepted by $g$; $AL(g)$ is a regular language.

We prove that if $L \subseteq \Omega_k$ is an infinite language of balanced non-associative words with the number of occurrences of $x_1$ “bounded” by $(k, \alpha)$-tetration then $L$ can be dissected by a regular language.

**Proposition 6.1.** If $k, \alpha \in \mathbb{N}$, $k \geq 2$, and $L \subseteq \Omega_k$ is an infinite language such that for each $w_1 \in L$ there is $w_2 \in L$ with $\text{occur}(w_2, x_1) \leq \exp^{k,\alpha} \text{occur}(w_1, x_1)$ then there is a regular language $R$ such that $R$ dissects $L$.

**Proof.** Let $w_1, w_2 \in L$ be such that

$$n_2 \leq \exp^k n_1, \quad (1)$$

where $n_1 = \text{occur}(w_1, x_1)$ and $n_2 = \text{occur}(w_2, x_1)$.

Let $h_1 = \text{height}(\mu_{k,k}(w_1), x_k)$ and $h_2 = \text{height}(\mu_{k,k}(w_2), x_k)$. Proposition [4.5] implies that there is a constant $c_1 \in \mathbb{R}^+$ such that

$$h_1 \geq \log^k n_1 \text{ and } h_2 \leq c_1 + \log^k n_2 \quad (2)$$

Note that the constant $c_1$ does not depend on $n_1, n_2, h_1, h_2$. On the other hand, the constant depends on $k$.

From (1) and (2) we have that

$$h_2 \leq c_1 + \log^k n_2 \leq c_1 + \log^k (\exp^{k,\alpha} n_1). \quad (3)$$

Realize that $\log^j (\exp^{j,\alpha} n_1) = \exp^{j-1,\alpha}$ and that if $a, b \in \mathbb{R}^+$ and $a, b \geq 2$ then $a + b \leq ab$. Then we have that

$$\log^j (\exp^{j,\alpha} n_1) = \log^{j-1} (\exp^{j-1,\alpha} + \log^2 n_1) \leq \log^{j-1} (\exp^{j-1,\alpha} \log^2 n_1). \quad (4)$$

From (4) it follows that

$$\log^k (\exp^{k,\alpha} n_1) \leq \log_2 (\exp^{1,\alpha} \log^{k-1} n_1) = \alpha + \log^k n_1. \quad (5)$$
From (2), (3), and (5) we have that

\[ h_2 \leq c_1 + \alpha + \log_2 \left( \frac{k}{n_1} \right) \leq c_1 + \alpha + h_1. \] (6)

The equation (6) says that there is a constant \( c \in \mathbb{N} \) such that for each \( u \in L \) there is \( v \in L \) with \( h_2 \leq c + h_1 \).

Lemma 5.2 implies that there are \( j_1, j_2 \) such that both \( H_1, H_2 \) are infinite sets, where

\[ H_i = \{ v \mid v \in L \text{ and } \text{height}(\mu_{k,k}(v), x_k) \equiv j_i \pmod{c+1} \} \text{ and } i \in \{1, 2\}. \]

Consider the deterministic finite automaton \( g = (Q, \Sigma_{2k-1}, q_0, \delta, F) \), where

- \( Q = \{ q_0, q_1, \ldots, q_c, q_a, q_r \} \),
- \( \delta(q, x) \to (q) \), where \( q \in Q \) and \( x \in \Sigma_{2k-1} \setminus \{x_k, x_k-1\} \),
- \( \delta(q_i, x_k) \to (q_{i+1} \mod c+1) \),
- \( \delta(q_j, x_{k-1}) \to (q_a) \),
- \( \delta(q_i, x_{k-1}) \to (q_r) \), where \( i \neq j \),
- \( \delta(q, x) \to (q) \), where \( q \in \{q_a, q_r\} \) and \( x \in \Sigma_{2k-1} \), and
- \( F = \{q_j, q_a\} \).

The deterministic finite automaton \( g \) implements the modulo operation on the prefix of the form \( x_k^i \). The input letter \( x \in \Sigma_{2k-1} \setminus \{x_k, x_k-1\} \) does not change the state. The input letter \( x_k \) changes the state from \( q_i \) to \( q_{i+1} \mod c+1 \). If the input letter equals \( x_{k-1} \) then the state changes either to accept \( q_a \) or refuse \( q_r \). Realize that if \( w \in \mu_{k,k}(\Omega_k) \), \( a \in \{x_k, y_k, x_k-1\} \), and \( x_k a \in \text{Fac}(w) \) then \( a \in \{x_k, x_{k-1}\} \). Once in the state \( q_a \) or \( q_r \), no other states can be reached. The states \( q_{j_1} \) and \( q_a \) are the accepting states. It is easy to see that \( \text{AL}(g) = H_1 \) and in consequence the regular language \( R = \text{AL}(g) \) dissects \( L \).

This completes the proof. \( \Box \)

Given \( n \in \mathbb{N} \), let \( \Delta_n \) be some alphabet with \( n \) letters. Let \( \Delta_1 = B_1 = \{x_1\} \) be the alphabet with the “first” opening bracket. Let \( L \subseteq \Delta_n^* \) be an infinite language with a growing bounded by \( (k, \alpha) \)-tetration. Let \( v : \Sigma_{2k-1}^* \to \Delta_1 \) be an erasing alphabetical homomorphism defined by \( v(x_1) = x_1 \) and \( v(a) = \epsilon \),
where $a \in \Sigma_{2k-1} \setminus \{x_1\}$. Let $\pi : \Delta_n^* \to \Delta_1$ be an alphabetical homomorphism defined by $v(a) = x_1$ for all $a \in \Delta_n$. Note that if $w \in \Delta_n^*$ then $|w| = |\pi(w)|$.

We show that there $3k - 3$ context free languages $L_1, L_2, \ldots, L_{3k-3} \subseteq \Sigma_{2k-1}^*$ such that the homomorphic image $v(\bigcap_{i \leq}^{3k-3} L_i)$ dissects the homomorphic image $\pi(L)$.

**Theorem 6.2.** If $n, \alpha, k \in \mathbb{N}$, $k \geq 2$, $L \subseteq \Delta_n^*$ is an infinite language with the growth bounded by $(k, \alpha)$-tetration then there are $3k - 3$ context free languages $L_1, L_2, \ldots, L_{3k-3}$ such that $v(\bigcap_{i \leq}^{3k-3} L_i)$ dissects $\pi(L)$.

**Proof.** Recall that the language $\Omega_k$ is an intersection of $3k - 3$ context free languages:

$$\Omega_k = \bigcap_{m=2}^{k} (\mathrm{Naw}_{k,m} \cap \mathrm{Bal}_{k,m} \cap \Lambda_{k,m}).$$

Let us denote these languages $L_1, L_2, \ldots, L_{3k-3}$. Let $\pi(L) = \{w \mid w \in L\} \subseteq \Delta_1^*$ and let $\bar{L} = \{w \in \Omega_k \mid v(w) \in \pi(L)\}$. Note that $\bar{L}$ contains $w \in \Omega_k$ if and only if there is $\bar{w} \in L$ such that the number of occurrences of $x_1$ in $w$ is equal to the length of $\bar{w}$; formally $\text{occur}(w, x_1) = |\bar{w}|$.

Since $L$ is a language with the growth bounded by $(k, \alpha)$-tetration, we have that for each $w_1 \in \bar{L}$ there is $w_2 \in \bar{L}$ with

$$\text{occur}(w_2, x_1) \leq \exp^{k,\alpha} \text{occur}(w_1, x_1).$$

Then Proposition 6.1 implies that there is a regular language $R$ that dissects $\bar{L}$. It is well known that intersection of a regular language and a context free language is a context free language. Hence let $\bar{L}_1 = L_1 \cap R$. Then $\bar{L}_1 \cap \left( \bigcap_{i=2}^{3k-3} L_i \right)$ dissects $\bar{L}$. The theorem follows. \[ \square \]

**Acknowledgments**

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS20/183/OHK4/3T/14.

**References**

[1] Bucher, W.: A density problem for context-free languages. Bull. Eur. Assoc. Theor. Comput. Sci. EATCS 10 (1980)
[2] Domaratzki, M., Shallit, J., Yu, S.: Minimal covers of formal languages. In: Developments in Language Theory (2001)

[3] Flajolet, P., Steyaert, J.M.: On sets having only hard subsets. In: Loeckx, J. (ed.) Automata, Languages and Programming. pp. 446–457. Springer Berlin Heidelberg, Berlin, Heidelberg (1974)

[4] Ginsburg, S., Greibach, S.: Deterministic context free languages. Information and Control 9(6), 620 – 648 (1966). https://doi.org/https://doi.org/10.1016/S0019-9958(66)80019-0

[5] Julie, J., Baskar Babujee, J., Masilamani, V.: Dissecting power of certain matrix languages. In: Arumugam, S., Bagga, J., Beineke, L.W., Panda, B. (eds.) Theoretical Computer Science and Discrete Mathematics. pp. 98–105. Springer International Publishing, Cham (2017)

[6] Liu, L., Weiner, P.: An infinite hierarchy of intersections of context-free languages. Math. Systems Theory 7, 185–192. (1973,)

[7] Post, E.L.: Recursively enumerable sets of positive integers and their decision problems. Bull. Amer. Math. Soc. 50(5), 284–316 (05 1944), https://projecteuclid.org:443/euclid.bams/1183505800

[8] Wotschke, D.: The Boolean Closures of the Deterministic and Nondeterministic Context-Free Languages, pp. 113–121. Springer Berlin Heidelberg, Berlin, Heidelberg (1973). https://doi.org/https://doi.org/10.1007/978-3-662-41148-3_11

[9] Wotschke, D.: Nondeterminism and boolean operations in pda's. Journal of Computer and System Sciences 16(3), 456 – 461 (1978). https://doi.org/https://doi.org/10.1016/0022-0000(78)90030-2

[10] Yamakami, T.: Intersection and union hierarchies of deterministic context-free languages and pumping lemmas. In: Leporati, A., Martín-Vide, C., Shapira, D., Zandron, C. (eds.) Language and Automata Theory and Applications. pp. 341–353. Springer International Publishing, Cham (2020)

[11] Yamakami, T., Kato, Y.: The dissecting power of regular languages. Information Processing Letters 113(4), 116 – 122 (2013). https://doi.org/https://doi.org/10.1016/j.ipl.2012.12.006