Rational index of languages with bounded dimension of parse trees*

Ekaterina Shemetova[1,3,4][0000−0002−1577−8347], Alexander Okhotin[1][0000−0002−1615−2725], and Semyon Grigorev[2,4][0000−0002−7966−0698]

[1] Department of Mathematics and Computer Science, St. Petersburg State University, 14th Line V. O., 29, Saint Petersburg 199178, Russia
alexander.okhotin@spbu.ru
[2] Department of Mathematics and Mechanics, St. Petersburg State University, 7/9 Universitetskaya nab., Saint Petersburg 199034, Russia
[3] St. Petersburg Academic University, ul. Khlopina, 8, Saint Petersburg 194021, Russia
[4] JetBrains Research, Primorskiy prospekt 68-70, Building 1, St. Petersburg, 197374, Russia
{ekaterina.shemetova,semyon.grigorev}@jetbrains.com

Abstract. The rational index ρ_L of a language L is an integer function, where $\rho_L(n)$ is the maximum length of the shortest string in $L \cap R$, over all regular languages R recognized by n-state nondeterministic finite automata (NFA). This paper investigates the rational index of languages defined by (context-free) grammars with bounded tree dimension, and shows that it is of polynomial in n. More precisely, it is proved that for a grammar with tree dimension bounded by d, its rational index is $O(n^{2d})$, and that this estimation is asymptotically tight, as there exists a grammar with rational index $\Theta(n^{2d})$.
Keywords. Dimension of a parse tree; Strahler number; rational index; context-free languages; CFL-reachability.

1 Introduction

The notion of a rational index was introduced by Boasson, Courcelle and Nivat [3] as a complexity measure for context-free languages. The rational index ρ_L of a language L is an integer function, where $\rho_L(n)$ is the maximum length of the shortest string in a language of the form $L \cap R$, where R is a regular language recognized by n-state nondeterministic finite automata (NFA), and the maximum is taken over all such languages R with $L \cap R \neq \varnothing$. The rational index plays an important role in determining the parallel complexity of practical problems, such as the CFL-reachability problem and the more general Datalog query evaluation.

The *CFL-reachability problem* is stated as follows: for a context-free grammar G given an NFA A over the same alphabet, determine whether $L(G) \cap L(A)$

* Research supported by the Russian Science Foundation, project 18-11-00100.

is non-empty. With A is regarded as a labelled graph, this is a kind of graph reachability problem with path constraints given by context-free languages. This is an important problem used in static code analysis [16] and graph database query evaluation [19].

The CFL-reachability problem is P-complete already for a fixed context-free grammar [9]. The question on the parallel complexity of this problem was investigated by Ullman and Van Gelder [17] in a much more general case, with a rich logic for database queries instead of grammars, and it was proved that under an assumption called the *polynomial fringe property* the problem is decidable in NC [17]. In the special case of grammars, the result of Ullman and Van Gelder [17] gives an NC^2 algorithm for the CFL-reachability problem, under the assumption that the grammar's rational index is polynomial.

Theoretical properties of the rational index have received some attention in the literature. Pierre and Farinone [15] proved that for every algebraic number $\gamma \geqslant 1$, a language with the rational index in $\Theta(n^\gamma)$ exists. An upper bound on the rational index, shown by Pierre [14], is $2^{\Theta(n^2/\ln n)}$, and this bound is reached on the Dyck language on two pairs of parentheses. For several important subfamilies of grammars, such as the linear and the one-counter languages, there are polynomial upper bounds on the rational index, which imply that the CFL-reachability problem is in NC^2; they can be proved to lie in NL by direct methods not involving the rational index [11, 12].

In this paper we investigate the rational index of a generalization of linear languages: the *languages of bounded tree dimension*, that is, those defined by grammars with a certain limit on branching in the parse trees. The notion of tree dimension is well-known in the literature under different names: Chytil and Monien [6] use the term *k-caterpillar trees*, Esparza et al. [7] call this the *Strahler number* of a tree and mention numerous applications and alternative names for this notion, while Luttenberger and Schlund [13] use the term *tree dimension*, which is adopted in this paper.

Linear languages are languages of tree dimension 1, and their rational index is known to be $O(n^2)$ [3]. It can be derived from the work of Chytil and Monien [6] that languages of tree dimension bounded by d have rational index $O(n^{2d})$: this is explained in Section 3 of this paper. The new result of this paper, presented in Section 4, is that, for every d, there is a language of tree dimension bounded by d with rational index $\Theta(n^{2d})$. Some implications of this result are presented in Section 5: the rational index is asymptotically determined for *superlinear languages* [4], and some bounds are obtained for *languages of bounded oscillation* [8, 18].

2 Definitions

A *(context-free) grammar* is a quadruple $G = (\Sigma, N, R, S)$, where Σ is an alphabet; N is a set of nonterminal symbols; R is a set of rules, each of the form $A \to \alpha$, with $A \in N$ and $\alpha \in (\Sigma \cup N)^*$; and $S \in N$ is the start symbol. A parse tree is a tree, in which every leaf is labelled with a symbol from Σ, while every internal node is labelled with a nonterminal symbol $A \in N$ and has an associated rule $A \to X_1 \dots X_\ell \in R$, so that the node has ℓ ordered children labelled with $X_1, \dots, X_\ell$. The language defined by each nonterminal symbol $A \in N$, denoted by $L_G(A)$, is the set of all strings $w \in \Sigma^*$, for which there exists a parse tree, with A as a root and with the leaves forming the string w, The language defined by the grammar is $L(G) = L_G(S)$.

A grammar G is said to be is in the *Chomsky normal form*, if all rules of R are of the form $A \to BC$, with $B, C \in N$, or of the form $A \to a$, with $a \in \Sigma$.

A *nondeterministic finite automaton* (NFA) is a quintuple $\mathcal{A} = (\Sigma, Q, Q_0, \delta, F)$, where Q is a finite set of states, Σ is a finite set of input symbols, $Q_0 \subseteq Q$ is the set of initial states, $\delta \colon Q \times \Sigma \to 2^Q$ is the transition function, $F \subseteq Q$ is the set of accepting states. It accepts a string $w = a_1 \dots a_n$ if there is a sequence of states $q_0, \dots, q_n \in Q$ with $q_0 \in Q_0$, $q_i \in \delta(q_{i-1}, a_i)$ for all i, and $q_n \in F$. The language of all strings accepted by $\mathcal{A}$ is denoted by $L(\mathcal{A})$.

For a language L over an alphabet Σ, its rational index ρ_L is a function defined as follows:

$$\rho_L(n) = \max_{\substack{\mathcal{A}:\text{NFA with } n \text{ states} \\ L \cap L(\mathcal{A}) \neq \varnothing}} \ \min_{w \in L \cap L(\mathcal{A})} |w|$$

Tree dimension. For each node v in a parse tree t, its *dimension* $\dim v$ is an integer representing the amount of branching in its subtree. It is defined inductively: a leaf v has dimension 0. For an internal node v, if one of its children $v_1, v_2, \dots, v_k$, with $k \geqslant 1$, has a greater dimension than all the others, then v has the same dimension, and if there are multiple children of maximum dimension, then the dimension of v is greater by one.

$$\dim v = \begin{cases} \max_{i \in \{1,\dots,k\}} \dim v_i & \text{if there is a unique maximum} \\ \max_{i \in \{1,\dots,k\}} \dim v_i + 1 & \text{otherwise} \end{cases}$$

The dimension of a parse tree t, denoted by $\dim t$, is the dimension of its root.

Definition 1 (Grammars of bounded tree dimension). *A grammar G is of d-bounded tree dimension if every parse tree t of G has $\dim t \leqslant d$, where d is some constant. This constant is called the dimension of G, denoted by $\dim G = d$.*

Classical transformation to the Chomsky normal form preserves the class of grammars of d-bounded tree dimensions. Languages defined by such grammars are called *languages of d-bounded tree dimension*.

3 Upper bound on the rational index

The first result of this paper is that, if the dimension of trees in a grammar is bounded by a constant d, then the rational index of its language is bounded by $O(n^{2d})$, where the constant factor depends upon the grammar.

Theorem 1. *Let G be a grammar of d-bounded tree dimension, and let $\mathcal{A}$ be an NFA with n states, with non-empty intersection $L(G) \cap L(\mathcal{A})$. Then the length of the shortest string in $L(G) \cap L(\mathcal{A})$ is at most $|G|^d n^{2d}$, where $|G| = \sum_{A \to \alpha}(|\alpha|+2)$ is number of symbols used for the description of the grammar.*

The main component of the proof is the following lemma by Chytil and Monien [6], which they used in their study of unambiguous grammars of finite index.

Lemma 1 (Chytil and Monien [6, Lem. 7]). *Let $G = (\Sigma, N, R, S)$ be a grammar, let m be the maximal length of the right-hand side of its rules, and assume that there exists a parse tree of dimension $d \geqslant 1$ in this grammar. Then the grammar defines some string of length at most $(|N|(m-1)+1)^d$.*

The proof proceeds by simplifying the tree of dimension d by removing paths beginning and ending with the same nonterminal symbol. This contraction results in a parse tree of a bounded size, which has the same dimension d [6].

Proof (of Theorem 1). A given grammar G is first transformed to the Chomsky normal form, resulting in a grammar $G' = (\Sigma, N', R', S')$ with the same bound on the dimension of parse trees and with at most $|G|$ nonterminal symbols.

Next, a grammar G'' for the language $L(G) \cap L(\mathcal{A})$ is obtained from G' and $\mathcal{A}$ by the classical construction by Bar-Hillel et al. [2], which produces $|N'| \cdot n^2 + 1$ nonterminal symbols: these are triples of the form (A, p, q), where $A \in N'$ and p, q are two states of the automaton, as well as a new start symbol. The grammar G'' is still in the Chomsky normal form, that is, the maximum length of a right-hand side of any rule is $m = 2$. Since G'' defines at least one string, there exists a parse tree of dimension at most d. Then, by Lemma 1, the length of the shortest string defined by this grammar is at most $(|N'| \cdot n^2 + 1 + 1)^d \leqslant (|G| \cdot n^2)^d$. ◻

4 Lower bound on the rational index

The upper bound $O(n^{2d})$ on the rational index of a language defined by a grammar with tree dimension bounded by d has a matching lower bound $\Omega(n^{2d})$. It is first established for a convenient infinite set of values of n, to be extended to arbitrary n in the following.

Lemma 2. *For every $d \geqslant 1$, there is a grammar G of bounded tree dimension d, such that for every $n \geqslant 2^{d+1}$ divisible by 2^d there is an n-state NFA $\mathcal{B}$, such that the shortest string w in $L(G) \cap L(\mathcal{B})$ is of length at least $\frac{1}{2^{d^2+3d-3}} n^{2d}$.*

Proof. The grammar and the automaton are constructed inductively on d, for every d and only for n divisible by 2^d. Each constructed NFA shall have a unique initial state, which is also the unique accepting state.

Basis. $dim(G) = 1$. The family of languages having dimension $d = 1$ coincides with the family of linear languages. Let G be a linear grammar with the rules $S \to aSb \mid ab$, which defines the language $L(G) = \{\, a^i b^i \mid i \geqslant 1 \,\}$.

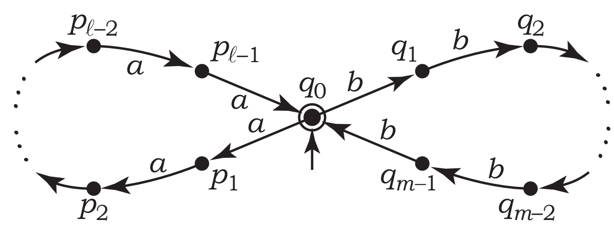

Fig. 1. NFA $\mathcal{B}$ defined in Lemma 2 for $d = 1$.

For every $n \geqslant 4$ divisible by $2^d = 2$, let $\ell = \frac{n}{2}$, $m = \frac{n}{2} + 1$. Then ℓ and m are coprime integers. Define an NFA $\mathcal{B}$ over the alphabet $\{a, b\}$, which consists of two cycles sharing one node, q_0, which is both the initial and the unique accepting state. The cycle of length ℓ has all transitions by a, and the other by b, as shown in Figure 1. The automaton has $\ell + m - 1 = n$ states.

Every string in $L(G) \cap L(\mathcal{B})$ is of the form $a^i b^i$, with $i \geqslant 1$. For the automaton to accept it, i must be divisible both by ℓ and by m. Since the cycle lengths are relatively prime, the shortest string w with this property has $i = \ell m$, and is accordingly of length $2\ell m$. Its growth with n is estimated as follows.

$$|w| = 2\ell m = 2\frac{n}{2} \cdot \left(\frac{n}{2} + 1\right) = \frac{1}{2}n^2 + n$$

This example is well-known to the community [10, 19].

Inductive step. $dim(G) = d$.

By the induction hypothesis, there is a grammar $\widehat{G} = (\widehat{\Sigma}, \widehat{N}, \widehat{R}, \widehat{S})$ of bounded dimension $dim(\widehat{G}) = d - 1$, which satisfies the statement of the lemma. The new grammar $G = (\Sigma, N, R, S)$ of dimension d is defined over the alphabet $\Sigma = \widehat{\Sigma} \cup \{a, b, c\}$, where $a, b, c \notin \widehat{\Sigma}$ are new symbols. It uses nonterminal symbols $N = \widehat{N} \cup \{S, A\}$, adding two new nonterminals $A, S \notin \widehat{N}$ to those in $\widehat{G}$, where S is the new initial symbol. Its set of rules includes all rules from $\widehat{G}$ and the following new rules.

$$S \to ASc \mid Ac$$
$$A \to aAb \mid a\widehat{S}b$$

To see that the dimension of the new grammar is greater by 1 than the dimension of $\widehat{G}$, first consider the dimension of any parse tree t with the root labeled by the nonterminal A, shown in Figure 2(right). The dimension of the $\widehat{S}$-subtree at the bottom is at most $d-1$ by the properties of $\widehat{G}$. This dimension is inherited by all A-nodes in the tree, because their remaining children are leaves.

Now consider the dimension of a complete parse tree t with the start symbol S in the root, as in Figure 2(left). All A-subtrees in this tree have dimension at most $d - 1$. Then the bottom S-subtree, which uses the rule $S \to Ac$, also has dimension at most $d - 1$. Every S-subtree higher up in the tree uses a rule

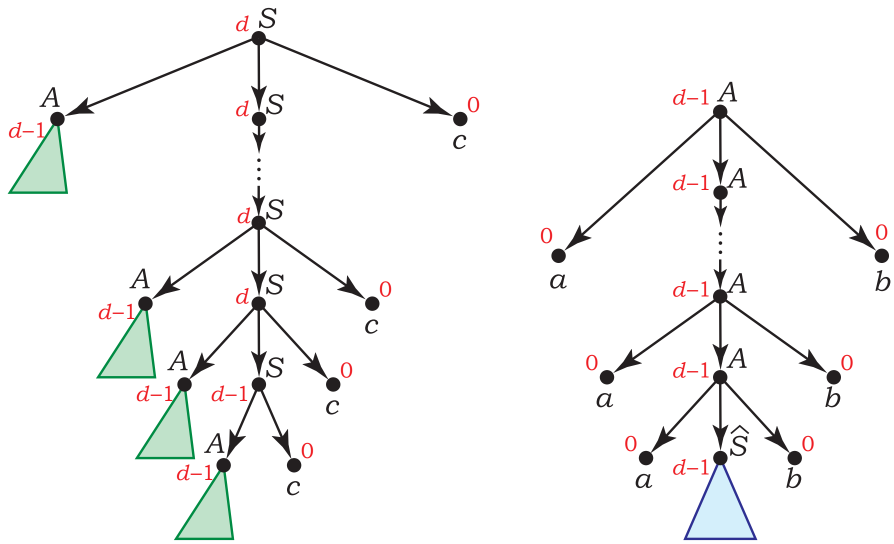

Fig. 2. Parse trees for S and for A, annotated with dimensions of their vertices

$S \to ASc$, and its dimension is at most d, because getting a higher dimension would require two subtrees of dimension d, which is never the case.

Now, for every $n \geqslant 2^{d+1}$ divisible by 2^d, the goal is to construct an n-state NFA over the alphabet Σ, so that the shortest string w in $L(G) \cap L(\mathcal{B})$ is of length at least $\frac{1}{2^{d^2+3d-3}} n^{2d}$. Since the number $\frac{n}{2}$ is at least 2^d and is divisible by 2^{d-1}, the induction hypothesis for the grammar $\widehat{G}$ asserts that there is an NFA $\widehat{\mathcal{B}} = (\widehat{Q}, \widehat{\Sigma}, \widehat{\delta}, \widehat{q}_0, \{\widehat{q}_0\})$, with $\frac{n}{2}$ states, with the shortest string $\widehat{w}$ in $L(\widehat{G}) \cap L(\widehat{\mathcal{B}})$ of length $\frac{1}{2^{(d-1)^2+3(d-1)-3}} (\frac{n}{2})^{2(d-1)}$.

The desired n-state NFA $\mathcal{B} = (\Sigma, Q, q_0, \delta, \{q_0\})$ is constructed as follows. Let $\ell = \frac{n}{4}$ and $m = \frac{n}{4} + 1$, these are two coprime integers. The set of states of $\mathcal{B}$ contains all $\frac{n}{2}$ states from $\widehat{Q}$, in which $\mathcal{B}$ it operates as $\widehat{\mathcal{B}}$, and $m + \ell - 1 = \frac{n}{2}$ new states forming a cycle of length ℓ and a chain of length m, which share a state.

$$Q = \widehat{Q} \cup \{p_1, \ldots, p_{\ell-1}, q_0, \ldots, q_{m-1}\}$$

The new initial state q_0 has a transition by a leading to the initial state of $\widehat{\mathcal{B}}$, from where one can return to q_1 by b.

$$\begin{aligned}
\delta(q_0, a) &= \{\widehat{q}_0\} \\
\delta(\widehat{q}_0, b) &= \{q_1\}
\end{aligned}$$

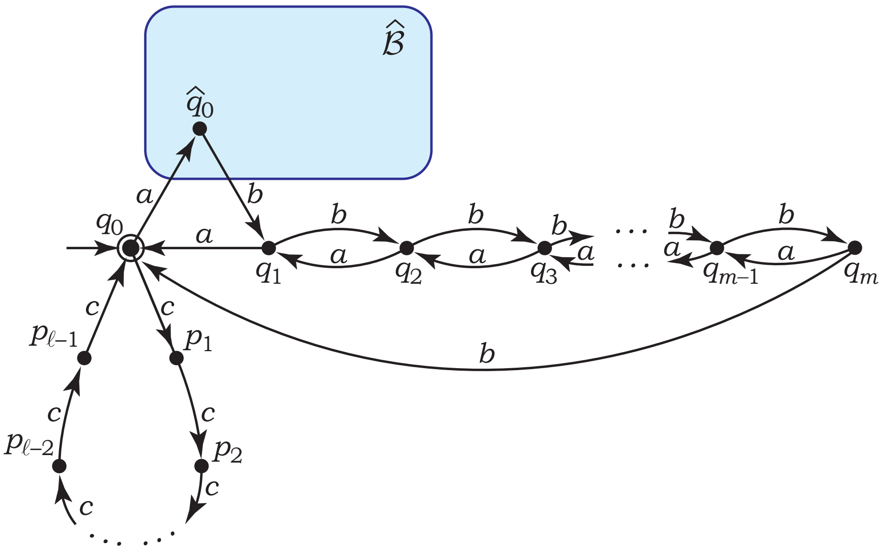

Fig. 3. NFA $\mathcal{B}$ defined in Lemma 2 for d, which incorporates NFA $\widehat{\mathcal{B}}$ for $d-1$.

There is a chain of transitions by a from q_{m-1} to q_0, and another chain b in the opposite direction, from q_1 to q_{m-1} and back to q_0.

$$\begin{aligned}
\delta(q_i, a) &= \{q_{i-1}\}, && \text{with } 1 \leqslant i \leqslant m-1 \\
\delta(q_i, b) &= \{q_{i+1}\}, && \text{with } 1 \leqslant i \leqslant m-2 \\
\delta(q_{m-1}, b) &= \{q_0\}
\end{aligned}$$

There is a cycle by c in the states $q_0, p_1, \dots, p_{\ell-1}$; for uniformity, denote $p_0 = q_0$.

$$\delta(p_i, c) = \{p_{i+1 \bmod \ell}\}, \qquad \text{with } 0 \leqslant i \leqslant \ell - 1$$

The general form of $\mathcal{B}$ is shown in Figure 3.

Let w be the shortest string in $L(G) \cap L(\mathcal{B})$. Consider how w is formed. Start state is q_0. According to the grammar rule $S \to ASc \mid Ac$, the string w should start with a substring u in $L_G(A)$. There is the only one outgoing edge labeled with a, so the next state is $\widehat{q}_0$. The next part of w should be a symbol a or a string v in $L(\widehat{G})$. As there is no outgoing edge labeled with a, the string v is the shortest string in $L(\widehat{G}) \cap L(\widehat{\mathcal{B}})$, and, hence, $v = \widehat{w}$. Now the first part of w is $a\widehat{w}$. To complete a substring derived by the nonterminal A, there is only one possible transition, which is an edge from $\widehat{q}_0$ to q_1 labeled with b. The next substring should be symbol c (the rule $S \to Ac$) or a string derived by A. The only suitable transition here is from q_1 to q_0 by a, so a substring in $L(A)$ is started. Again, to complete the string generated by A, one goes to the state q_2, and w now starts with $a\widehat{w}baa\widehat{w}bb$. By the construction of NFA $\mathcal{B}$, this process continues until one comes to the state q_0 without starting a substring derived by the nonterminal A (notice that such substrings are the shortest possible). Clearly, it happens after m iterations. Then it is left to read m symbols c by going from q_0 to q_0. But m and ℓ are coprime, so to balance the number of substrings derived by the nonterminal A and the number of symbols c, one needs to repeat the first cycle ℓ times and the second cycle m times.

Accordingly, the shortest string w has the following structure. Let w_i be the shortest string such that there exists computation $q_{i-1} \overset{w_i}{\rightsquigarrow} q_i$ ($q_{m-1} \overset{w_m}{\rightsquigarrow} q_0$ for w_m) for $1 \leqslant i \leqslant m$ in $\mathcal{B}$ and $w_i \in L(A)$. Notice that $w_i = aw_{i-1}b$ and $w_0 = \widehat{w}$, and there exists computation $q_0 \overset{w_1}{\rightsquigarrow} q_1 \overset{w_2}{\rightsquigarrow} q_2 \overset{w_3}{\rightsquigarrow} \ldots \overset{w_{m-1}}{\rightsquigarrow} q_m \overset{w_m}{\rightsquigarrow} q_0$ in $\mathcal{B}$.

Considering the above and the rules $S \to ASc \mid Ac$ of the grammar G, the string w is of the following form:

$$w = \Big(\prod_{i=1}^{m} w_i\Big)^{\ell} c^{m\ell}$$

Then the length of w can be bounded as follows.

$$|w| = \Big(\sum_{i=1}^{m} |w_i|\Big)\ell + \ell m = \Big(\sum_{i=1}^{m} (|\widehat{w}| + 2i)\Big)\ell + \ell m \geqslant \ell m|\widehat{w}|$$

Using the lower bound on the length of $\widehat{w}$, the desired lower bound on the length of w is obtained.

$$\ell m|\widehat{w}| \geqslant \frac{n}{4} \cdot \frac{n}{4} \cdot \frac{1}{2^{(d-1)^2+3(d-1)-3}} \Big(\frac{n}{2}\Big)^{2(d-1)} = \\ = \frac{n^2}{16} \cdot \frac{1}{2^{d^2+d-5}} \cdot \frac{n^{2d-2}}{2^{2d-2}} = \frac{1}{2^{d^2+3d-3}} n^{2d}$$

$\square$

Theorem 2. *For every $d \geqslant 1$, there is a grammar G of bounded tree dimension d, such that for every $n \geqslant 2^{d+1}$ there is an n-state NFA $\mathcal{B}$, such that the shortest string w in $L(G) \cap L(\mathcal{B})$ is of length at least $\frac{1}{2^{d^2+d-3}3^{2d}} n^{2d}$.*

Proof. Let G be the grammar given for d by Lemma 2. Let $2^d r \leqslant n < 2^d(r+1)$, for some integer r. Then $r \geqslant 2$ (for otherwise $n < 2^{d+1}$), and $2^d r \geqslant 2^{d+1}$.

Since $2^d r$ is divisible by 2^d, by Lemma 2, there is an NFA $\mathcal{B}$ with $2^d r \leqslant n$ states, such that the length of the shortest string w in $L(G) \cap L(\mathcal{B})$ is at least $\frac{1}{2^{d^2+3d-3}}(2^d r)^{2d}$. This is the desired n-state NFA.

The inequality $n < 2^d(r+1)$ implies that $n < 2^d \frac{3r}{2}$, because $r+1$ is at most $\frac{3r}{2}$ for $r \geqslant 2$. Then $2^d r > \frac{2}{3}n$, and the lower bound on the length of w is expressed as a function of n as follows.

$$|w| \geqslant \frac{1}{2^{d^2+3d-3}}(2^d r)^{2d} \geqslant \frac{1}{2^{d^2+3d-3}}\Big(\frac{2}{3}n\Big)^{2d} = \frac{1}{2^{d^2+d-3}3^{2d}} n^{2d}$$

$\square$

For finite automata with fewer than 2^{d+1} states, no lower bounds are given, as the construction in the proof relies on having sufficiently long cycles in the automata.

Overall, the rational index of grammars with tree dimension bounded by d is $\Theta(n^{2d})$ in the worst case.

5 Rational indices for some language families

Superlinear languages. A grammar $G = (\Sigma, N, R, S)$ is *superlinear* (Brzozowski [4]) if its nonterminal symbols split into two classes, $N = N_{lin} \cup N_{nonlin}$, where rules for each nonterminal $A \in N_{lin}$ are of the form $A \to uBv$ or $A \to w$, with $B \in N_{lin}$, $u, v, w \in \Sigma^*$, while rules for a nontermial $A \in N_{nonlin}$ are of the form $A \to \alpha B \beta$, with $B \in N$ and $\alpha, \beta \in (\Sigma \cup N_{lin})^*$. A language is *superlinear* if it is generated by some superlinear grammar.

Corollary 1. *For every superlinear grammar G, the rational index $\rho_{L(G)}$ is at most $|G|^2 \cdot n^4$.*

Proof. Parse trees in a superlinear grammar G have dimension at most 2. Then, by Theorem 1, the rational index $\rho_{L(G)}$ is bounded by $|G|^2 \cdot n^4$.

Turning to a lower bound, note that the grammar constructed in Theorem 2 for $d = 2$ is actually superlinear.

Corollary 2. *There exists a superlinear grammar G with rational index $\rho_{L(G)}(n) \geqslant \frac{1}{648} n^4$.*

Bounded-oscillation languages. The notion of oscillation of runs in pushdown automata, applicable to Turing machines with auxiliary pushdown tape, was introduced by Wechsung [18]. *Languages with oscillation bounded by k* are then a generalization of the linear languages (as one-turn pushdown automata are those with oscillation bounded by $k = 1$).

This family was later studied by Ganty and Valput [8], who introduced the corresponding notion of oscillation in parse trees of grammars. Among other results, they prove that oscillation of a parse tree is closely related to its dimension.

Lemma 3 (Ganty and Valput [8]). *Let $G = (\Sigma, N, R, S)$ be a grammar in the Chomsky normal form, and let t be a parse tree in G. Then,* $\operatorname{osc} t - 1 \leqslant \dim t \leqslant 2 \operatorname{osc} t$.

Thus, k-bounded-oscillation grammars have dimension of parse trees bounded by $2k$, and Theorem 1 gives the following upper bound on the rational index of these languages.

Corollary 3. *Let L be a k-bounded-oscillation language. Then $\rho_L(n) = O(n^{4k})$.*

6 Conclusion and open problems

Languages of bounded tree dimension were proved to have polynomial rational index. This implies, in particular, that the CFL-reachability problem and Datalog query evaluation for these languages is in NC, and the degree of the polynomial becomes a constant factor for the circuit depth.

There is another family of languages which has polynomial rational index, *the one-counter languages*. Their rational index is known to be $O(n^2)$ [5]. Could this class be generalized in the same manner as linear languages, preserving the polynomial order of the rational index? One can consider the Polynomial Stack Lemma by Afrati et al. [1], where some restriction on the PDA stack contents is given, or investigate the properties of the substitution closure of the one-counter languages, which is known to have polynomial rational index [3].

Acknowledgment

The authors are grateful to the anonymous reviewers for numerous helpful remarks and suggestions, and particularly for alerting the authors of the work by Chytil and Monien [6].

References

1. Afrati, F., Papadimitriou, C.: The parallel complexity of simple chain queries. In: Proceedings of the Sixth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems. pp. 210–213. PODS ’87, ACM, New York, NY, USA (1987). https://doi.org/10.1145/28659.28682
2. Bar-Hillel, Y., Perles, M., Shamir, E.: On formal properties of simple phreise structure grammars. STUF - Language Typology and Universals **14**(1-4), 143 – 172 (01 Apr 1961). https://doi.org/10.1524/stuf.1961.14.14.143
3. Boasson, L., Courcelle, B., Nivat, M.: The rational index: a complexity measure for languages. SIAM Journal on Computing **10(2)**, 284–296 (1981)
4. Brzozowski, J.A.: Regular-like expressions for some irregular languages. In: 9th Annual Symposium on Switching and Automata Theory (swat 1968). pp. 278–286 (Oct 1968). https://doi.org/10.1109/SWAT.1968.24
5. Chistikov, D., Czerwinski, W., Hofman, P., Pilipczuk, M., Wehar, M.: Shortest paths in one-counter systems. Log. Methods Comput. Sci. **15**(1) (2019). https://doi.org/10.23638/LMCS-15(1:19)2019
6. Chytil, M., Monien, B.: Caterpillars and context-free languages. In: Choffrut, C., Lengauer, T. (eds.) STACS 90, 7th Annual Symposium on Theoretical Aspects of Computer Science, Rouen, France, February 22-24, 1990, Proceedings. Lecture Notes in Computer Science, vol. 415, pp. 70–81. Springer (1990). https://doi.org/10.1007/3-540-52282-4“˙33
7. Esparza, J., Luttenberger, M., Schlund, M.: A brief history of strahler numbers. In: Dediu, A., Martín-Vide, C., Sierra-Rodríguez, J.L., Truthe, B. (eds.) Language and Automata Theory and Applications - 8th International Conference, LATA 2014, Madrid, Spain, March 10-14, 2014. Proceedings. Lecture Notes in Computer Science, vol. 8370, pp. 1–13. Springer (2014). https://doi.org/10.1007/978-3-319-04921-2“˙1

8. Ganty, P., Valput, D.: Bounded-oscillation pushdown automata. Electronic Proceedings in Theoretical Computer Science **226**, 178–197 (Sep 2016). https://doi.org/10.4204/eptcs.226.13
9. Greenlaw, R., Hoover, H.J., Ruzzo, W.L.: Limits to Parallel Computation: P-completeness Theory. Oxford University Press, Inc., New York, NY, USA (1995)
10. Hellings, J.: Path results for context-free grammar queries on graphs. CoRR **abs/1502.02242** (2015)
11. Holzer, M., Kutrib, M., Leiter, U.: Nodes connected by path languages. In: Mauri, G., Leporati, A. (eds.) Developments in Language Theory. pp. 276–287. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
12. Komarath, B., Sarma, J., Sunil, K.S.: On the complexity of l-reachability. In: Jürgensen, H., Karhumäki, J., Okhotin, A. (eds.) Descriptional Complexity of Formal Systems. pp. 258–269. Springer International Publishing, Cham (2014)
13. Luttenberger, M., Schlund, M.: Convergence of newton's method over commutative semirings. Inf. Comput. **246**, 43–61 (2016). https://doi.org/10.1016/j.ic.2015.11.008
14. Pierre, L.: Rational indexes of generators of the cone of context-free languages. Theoretical Computer Science **95**(2), 279–305 (1992). https://doi.org/10.1016/0304-3975(92)90269-L
15. Pierre, L., Farinone, J.M.: Context-free languages with rational index in $\theta(n^\gamma)$ for algebraic numbers γ. RAIRO - Theoretical Informatics and Applications - Informatique Théorique et Applications **24**(3), 275–322 (1990)
16. Reps, T.W.: Program analysis via graph reachability. Information & Software Technology **40**, 701–726 (1997)
17. Ullman, J.D., Van Gelder, A.: Parallel complexity of logical query programs. In: 27th Annual Symposium on Foundations of Computer Science (sfcs 1986). pp. 438–454 (Oct 1986). https://doi.org/10.1109/SFCS.1986.40
18. Wechsung, G.: The oscillation complexity and a hierarchy of context-free languages. In: Fundamentals of Computation Theory, FCT 1979, Proceedings of the Conference on Algebraic, Arthmetic, and Categorial Methods in Computation Theory, Berlin/Wendisch-Rietz, Germany, September 17-21, 1979. pp. 508–515 (1979)
19. Yannakakis, M.: Graph-theoretic methods in database theory. In: Rosenkrantz, D.J., Sagiv, Y. (eds.) Proceedings of the Ninth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, April 2-4, 1990, Nashville, Tennessee, USA. pp. 230–242. ACM Press (1990). https://doi.org/10.1145/298514.298576