# Widths of regular and context-free languages

David Mestel
University of Luxembourg
david.mestel@uni.lu

## Abstract

Given a partially-ordered finite alphabet $\Sigma$ and a language $L \subseteq \Sigma^*$, how large can an antichain in $L$ be (where $L$ is given the lexicographic ordering)? More precisely, since $L$ will in general be infinite, we should ask about the rate of growth of maximum antichains consisting of words of length $n$. This fundamental property of partial orders is known as the width, and in a companion work [10] we show that the problem of computing the information leakage permitted by a deterministic interactive system modeled as a finite-state transducer can be reduced to the problem of computing the width of a certain regular language. In this paper, we show that if $L$ is regular then there is a dichotomy between polynomial and exponential antichain growth. We give a polynomial-time algorithm to distinguish the two cases, and to compute the order of polynomial growth, with the language specified as an NFA. For context-free languages we show that there is a similar dichotomy, but now the problem of distinguishing the two cases is undecidable. Finally, we generalise the lexicographic order to tree languages, and show that for regular tree languages there is a trichotomy between polynomial, exponential and doubly exponential antichain growth.

**2012 ACM Subject Classification** Theory of computation → Formal languages and automata theory

**Keywords and phrases** Formal languages, combinatorics on words, information flow

**Digital Object Identifier** 10.4230/LIPIcs.FSTTCS.2019.48

**Related Version** A full version of the paper is available at `https://arxiv.org/abs/1709.08696`.

**Funding** The author is supported by FNR under grant number 11689058 (Q-CoDe).

## 1 Introduction

Computing the size of the largest antichain (set of mutually incomparable elements) is the 'central' problem in the extremal combinatorics of partially ordered sets (posets) [14]. In addition to some general theory [7], it has attracted study for a variety of specific sets, beginning with Sperner's Theorem on subsets of $\{1, \ldots, n\}$ ordered by inclusion [12, 2, 11], and for random posets [1]. The size of the largest antichain in a poset is called its *width*.

In this work we study languages $L$ (regular or context-free) over finite partially ordered alphabets, with the lexicographic partial order. Since such languages will in general contain infinite antichains, we study the sets $L_{=n}$ of words of length $n$, and ask how the width of $L_{=n}$ grows with $n$; we call this the *antichain growth* rate of $L$.

In addition to its theoretical interest, the motivation for this work is the study of quantified information flow in computer security: we wish to know whether a pair of isolated agents interacting with a common central system (for example different programs running on a single computer and communicating with the operating system) can obtain any information about each other's actions, and if so how much. In a companion work [10] we show that if the central system is modeled as a deterministic finite-state transducer then this leakage is equivalent to the width of a certain regular language (roughly speaking, antichains corresponding to consistent sets of observations for one agent). The dichotomy we obtain in this paper thus corresponds to a dichotomy between logarithmic and linear information flow.

In Section 2 we set out basic definitions and results on the lexicographic order, antichains and antichain growth. In Section 3 we show that for regular languages there is a dichotomy

39th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2019).
Editors: Arkadev Chattopadhyay and Paul Gastin; Article No. 48; pp. 48:1–48:14

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

between polynomial and exponential antichain growth, and give a polynomial-time algorithm for distinguishing the two cases. In Section 4 we give a polynomial-time algorithm to compute the order of polynomial antichain growth. In Section 5 we show that for context-free languages there is a similar dichotomy between polynomial and exponential antichain growth, but that the problem of distinguishing the two cases is undecidable. In Section 6 we show that for regular tree languages there is a trichotomy between polynomial, exponential and doubly exponential antichain growth. Finally in Section 7 we discuss open problems.

For reasons of space, many proofs have been omitted or sketched in the conference version of this work; an extended version with full proofs may be found at [9].

## 2 Languages, lexicographic order and antichains

▶ **Definition 1.** *Let $\Sigma$ be a finite alphabet equipped with a partial order $\preceq$. Then the lexicographic partial order induced by $\preceq$ on $\Sigma^*$ is the relation $\preceq$ given by*

**(i)** *$\epsilon \preceq w$ for all $w \in \Sigma^*$ (where $\epsilon$ is the empty word), and*

**(ii)** *For any $x, y \in \Sigma, w, w' \in \Sigma^*$, we have $xw \preceq yw'$ if and only if either $x \prec y$ or $x = y$ and $w \preceq w'$.*

If words $x$ and $y$ are comparable in this partial order we write $x \sim y$. If $x$ is a prefix of $y$ we write $x \leq y$. For a language $L$, we will often write $L_{=n}$ to denote the set $\{w \in L \mid |w| = n\}$ (with corresponding definitions for $L_{<n}$, etc.), and $|L|_{=n}$ for $|L_{=n}|$.

The main subject of this work is *antichains*, that is sets of words which are mutually incomparable. It will sometimes be useful also to consider *quasiantichains*, which are sets of words which are incomparable except that the set may include prefixes (note that this is not a standard term). The opposite of an antichain is a *chain*, in which all elements are comparable.

▶ **Definition 2.** *A language $L$ is an* antichain *if for every $l_1, l_2 \in L$ with $l_1 \neq l_2$ we have $l_1 \not\sim l_2$. $L$ is a* quasiantichain *if for every $l_1, l_2 \in L$ we have either $l_1 \leq l_2$, $l_2 \leq l_1$ or $l_1 \not\sim l_2$. $L$ is a* chain *if for all $l_1, l_2 \in L$ we have $l_1 \sim l_2$.*

It is easy to see that the property of being an antichain is preserved by the operations of prefixing, postfixing and concatenation.

▶ **Lemma 3 (Prefixing).** *Let $w, w_1, w_2$ be any words. Then $w_1 \sim w_2$ if and only if $ww_1 \sim ww_2$. Hence for any language $L$, $wL$ is an antichain (respectively quasiantichain) if and only if $L$ is an antichain (quasiantichain).*

▶ **Lemma 4 (Postfixing).** *Let $w, w_1, w_2$ be any words. Then $w_1 \sim w_2$ if $w_1 w \sim w_2 w$. Hence for any language $L$, $Lw$ is an antichain if $L$ is an antichain.*

▶ **Lemma 5 (Concatenation).** *Let $w_1, w_2, w_1', w_2'$ be any words such that $w_1 \not\leq w_2$ and $w_2 \not\leq w_1$. Then $w_1 w_1' \sim w_2 w_2'$ if and only if $w_1 \sim w_2$. Hence if $L_1$ and $L_2$ are antichains then $L_1 L_2$ is an antichain.*

Clearly the property of being an antichain is not preserved by Kleene star, since $L^*$ will contain prefixes for any non-empty $L$. The best we can hope for is that $L^*$ is a quasiantichain.

▶ **Lemma 6 (Kleene star).** *Let $L$ be an antichain. Then $L^*$ is a quasiantichain.*

Ultimately we are going to care about the size of antichains inside particular languages. Since these will often be unbounded, we choose to ask about the rate of growth; that is,

if $L_1, L_2, L_3, \ldots \subseteq L$ are antichains such that $L_i$ consists of words of length $i$, how quickly can $|L_i|$ grow with $i$? We will call $\bigcup_i L_i$ an *antichain family* and ask whether it grows exponentially, polynomially, etc.

▶ **Definition 7.** *A language $L$ is an* antichain family *if for each $n$ the set $L_{=n}$ of words in $L$ of length $n$ is an antichain.*

▶ **Definition 8.** *A language $L$ is* exponential *(or* has exponential growth*) if there exists some $\epsilon > 0$ such that*

$$\limsup_{n\to\infty} \frac{|L|_{=n}}{2^{\epsilon n}} > 0,$$

*and the supremum of the set of $\epsilon$ for which this holds is the* order *of exponential growth.*

*$L$ is* polynomial *(or* has polynomial growth*) if there exists some $k$ such that*

$$\limsup_{n\to\infty} \frac{|L|_{=n}}{n^k} < \infty.$$

*If $0 < \limsup_{n\to\infty} \frac{|L|_{=n}}{n^k} < \infty$ then we say that $L$ has polynomial growth of order $k$.*

*For notational convenience, we will sometimes later adopt the convention that a language $L$ which is finite (and so $\limsup_{n\to\infty} \frac{|L|_{=n}}{n^k} = 0$ for all $k$) has polynomial growth of order $-1$.*

A reasonable alternative choice of notation would have been to define the quantity $w_n$ to be the size of the largest antichain consisting of words of length $n$, and then ask about the growth of the series $w_1, w_2, \ldots$. This is clearly equivalent to the definitions we have given above.

Note that we will sometimes use other characterisations that are clearly equivalent; for instance $L$ has exponential growth if and only if there is some $\epsilon$ such that $|L|_{=n} > 2^{\epsilon n}$ infinitely often. We will sometimes refer to a language which is not polynomial as 'super-polynomial', or as having 'growth beyond all polynomial orders'. Of course there exist languages whose growth rates are neither polynomial nor exponential; for instance $|L|_{=n} = \Theta(2^{\sqrt{n}})$.

▶ **Definition 9.** *A language $L$* has exponential antichain growth *if there is an exponential antichain family $L' \subseteq L$. $L$* has polynomial antichain growth *if for every antichain family $L' \subseteq L$ we have that $L'$ is polynomial.*

Antichain growth generalises the classical notion of language growth, which is just antichain growth with respect to the discrete partial order (in which all elements of $\Sigma$ are incomparable).

Note that we could have chosen to define exponential antichain growth as containing an exponential antichain (rather than an exponential antichain family). We will eventually see (Corollary 17) that for regular languages the two notions are equivalent. However, for general languages they are not; indeed the following proposition shows that the two possible definitions are not equivalent even for context-free languages.

▶ **Proposition 10.** *There exists a context-free language $L$ such that $L$ has exponential antichain growth but all antichains in $L$ are finite.*

**Proof.** Let $\Sigma = \{a, b, 0, 1\}$ with $\prec = \{(a, b)\}$. Let $L = \bigcup_{n=1}^{\infty} L_n = \bigcup_{n=1}^{\infty} a^{n-1}b\{0,1\}^n$. Then each $L_n$ is an antichain of size $2^n$ consisting of words of length $2n$, but we have $L_1 > L_2 > L_3 > \ldots$ so any antichain is a subset of $L_k$ for some $k$ and hence is finite (the notation $L_1 > L_2$ means that for any $w_1 \in L_1$ and $w_2 \in L_2$ we have $w_2 \prec w_1$). Plainly $L$ is a context-free language. ◀

We observed above that Kleene star does not preserve the property of being an antichain. We conclude this section by establishing Lemma 12, which addresses this problem; if our goal is to find a large antichain, it suffices to find a large quasiantichain (where the precise meaning of 'large' is having exponential growth).

As a preliminary, we observe the straightforward fact that taking finite unions does not change the polynomial or exponential growth character of languages.

▶ **Lemma 11.** *Let $L_1, L_2, \ldots, L_k$ be languages, such that $\bigcup_{i=1}^{k} L_i$ has exponential growth of order $\epsilon$ (respectively super-polynomial growth). Then $L_i$ has exponential growth of order $\epsilon$ (respectively super-polynomial growth) for some $i$.*

We are now ready to prove Lemma 12.

▶ **Lemma 12.** *Let $L$ be an exponential quasiantichain. Then there exists an exponential antichain $L' \subseteq L$.*

**Proof sketch.** We construct an exponential subset of $L$ which is prefix-free, and is therefore an exponential antichain. We do this by a Ramsey-style argument: always maintaining the invariant of exponential growth, at each step we pick a fixed word $w$ of length $k$, throw away that word if it is in the set, and also throw away all longer words of which $w$ is *not* a prefix; by Lemma 11 it is always possible to choose $w$ such that this process preserves the invariant. ◀

## 3 Regular languages

The dichotomy between polynomial and exponential language growth for regular languages has been independently discovered at least six times (see citations in [4]), in each case based on the fact that a regular language $L$ has polynomial growth if and only if $L$ is *bounded* (that is, $L \subseteq w_1^* \ldots w_k^*$ for some $w_1, \ldots, w_k$); otherwise $L$ has exponential growth.

In [4], Gawrychowski, Krieger, Rampersad and Shallit describe a polynomial time algorithm for determining whether a language is bounded. The key idea is to consider the sets $L_q$ of words which can be generated beginning and ending at state $q$. $L$ is bounded if and only if for every $q$ we have that $L_q$ is *commutative* (that is, that $L_q \subseteq w^*$ for some $w$), and this can be checked in polynomial time.

In this section, we generalise this idea to the problem of antichain growth by showing that $L$ has polynomial antichain growth if and only if $L_q$ is a chain for every $q$, and otherwise $L$ has exponential antichain growth. This is sufficient to establish the dichotomy theorem (Theorem 16). To give an algorithm for distinguishing the two cases (Theorem 18), we show how to produce an automaton whose language is empty if and only if $L_q$ is a chain (roughly speaking the automaton accepts pairs of incomparable words in $L_q$).

Before proving the main theorems, we first establish (Lemma 13) that if $L_1$ and $L_2$ have polynomial antichain growth then so does $L_1 L_2$. Moreover if the rates of polynomial growth of $L_1$ and $L_2$ are at most $k_1$ and $k_2$ respectively then the rate of polynomial growth of $L_1 L_2$ is at most $k_1 + k_2 + 1$. For the proof of this see the extended version [9].

▶ **Lemma 13.** *Let $L_1, L_2$ be languages with polynomial antichain growth of order at most $k_1$ and $k_2$ respectively. Then $L_1 L_2$ has polynomial antichain growth of order at most $k_1 + k_2 + 1$.*

We are now ready to prove the main theorem, generalising the condition for polynomial language growth (that $L_q$ is commutative for every $q$) to one for polynomial antichain growth: that $L_q$ is a chain for every relevant $q$.

▶ **Definition 14.** *A state $q$ of an automaton $\mathcal{A} = (Q, \Sigma, \Delta, q_0, F)$ is* accessible *if $q$ is reachable from $q_0$ and* co-accessible *if $F$ is reachable from $q$.*

▶ **Definition 15.** *Let $\mathcal{A} = (Q, \Sigma, \Delta, q_0, F)$ be an NFA. Then for each $q_1, q_2 \in Q$, the automaton $\mathcal{A}_{q_1,q_2} \triangleq (Q, \Sigma, \Delta, q_1, \{q_2\})$.*

▶ **Theorem 16.** *Let $\mathcal{A} = (Q, \Sigma, \Delta, q_0, F)$ be an NFA over a partially ordered alphabet. Then*

**(i)** *$\mathcal{L}(\mathcal{A})$ has polynomial antichain growth if and only if $\mathcal{L}(\mathcal{A}_{q,q})$ is a chain for every accessible and co-accessible state $q$, and*

**(ii)** *if $\mathcal{L}(\mathcal{A})$ does not have polynomial antichain growth then it contains an exponential antichain (and hence has exponential antichain growth).*

**Proof.** Suppose that $w_1, w_2 \in \mathcal{L}(\mathcal{A}_{q,q})$ with $w_1 \not\sim w_2$ and $q$ accessible and co-accessible, so $w \in \mathcal{L}(\mathcal{A}_{q_0,q})$ and $w' \in \mathcal{L}(\mathcal{A}_{q,q'})$ for some $w, w'$ and some $q' \in F$. Now by the Kleene star Lemma we have that $(w_1 + w_2)^*$ is an exponential quasiantichain and so by Lemma 12 there is an exponential antichain $L' \subseteq (w_1 + w_2)^*$. Then by the Prefixing and Postfixing Lemmas we have that $wL'w' \subseteq L$ is an exponential antichain.

For the converse, we proceed by induction on $|Q|$. Let $Q' = Q \setminus \{q_0\}, F' = F \setminus \{q_0\}$ and $\Delta'(q, a) = \Delta(q, a) \setminus \{q_0\}$ for all $q \in Q', a \in \Sigma$. For any $q \in Q'$, let $\mathcal{A}'_q = (Q', \Sigma, \Delta', q, F')$. Then by the inductive hypothesis we have that $\mathcal{L}(\mathcal{A}'_q)$ has polynomial antichain growth. Also, since $L_{q_0} = \mathcal{L}(\mathcal{A}_{q_0,q_0})$ is a chain it has polynomial (in particular constant) antichain growth. Now we have

$$\mathcal{L}(\mathcal{A}) \subseteq L_{q_0} \cup \bigcup_{q \in Q'} \bigcup_{a \in \Delta(q_0,q)} L_{q_0} a \mathcal{L}(\mathcal{A}'_q).$$

By Lemma 13, each $L_{q_0} a \mathcal{L}(\mathcal{A}'_q)$ also has polynomial antichain growth, and hence by Lemma 11 so does the finite union. ◀

A trivial restatement of part (ii) of the theorem shows that the two possible definitions of antichain growth are equivalent.

▶ **Corollary 17.** *Let $L$ be a regular language. Then $L$ has exponential (respectively super-polynomial) antichain growth if and only if $L$ contains an exponential (respectively super-polynomial) antichain.*

Using Theorem 16 we can produce an algorithm for distinguishing the two cases.

▶ **Theorem 18.** *There exists a polynomial time algorithm to determine whether the language of a given NFA $\mathcal{A}$ has exponential antichain growth.*

**Proof sketch.** We construct an NFA $\mathcal{B}$ which accepts interleavings of incomparable words over $\Sigma$ and $\Sigma'$ (a fresh copy of the alphabet). We then have that $\mathcal{L}(\mathcal{A}_{q,q})$ is a chain if and only if $\mathcal{L}((\mathcal{A}_{q,q} \mid\mid\mid \mathcal{A}'_{q,q}) \cap \mathcal{B})$ is empty, where $\mathcal{A}'$ is a copy of $\mathcal{A}$ over alphabet $\Sigma'$. This can be checked in polynomial time. ◀

## 4 Precise growth rates

In [4] the authors give an algorithm to compute the order of polynomial language growth for the language of a given NFA; on the other hand efficiently computing the order of exponential growth is an open problem. In this section we give an algorithm to compute the order of polynomial antichain growth for the language of a given NFA. We do this by first giving an algorithm for DFA, and then showing that in fact it also works for NFA. We will assume throughout without loss of generality that all states are accessible and co-accessible.

▶ **Definition 19.** *Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$ be a DFA over a partially ordered alphabet. Let $G_{\mathcal{A}} = (Q, E)$ be the directed graph with vertex-set $Q$ such that $(q, q') \in E$ if and only if $q \xrightarrow{w} q'$ for some $w \in \Sigma^*$.*

*Let $G'_{\mathcal{A}} = (Q, E')$ be the directed graph with $(q, q') \in E'$ if and only if there exist words $w \not\sim w' \in \Sigma^*$ such that $q \xrightarrow{w} q$ and $q \xrightarrow{w'} q'$. We will write $L_{q,q'} \triangleq \mathcal{L}(\mathcal{A}_{q,q'})$.*

We will generally omit the subscript $\mathcal{A}$s from now on, where this will not cause confusion.

Note that by Theorem 16, we have that $G'$ is a directed acyclic graph (DAG) if and only if $\mathcal{L}(\mathcal{A})$ has polynomial antichain growth. By a similar argument to the proof of Theorem 18, the graph $G'$ can be computed in polynomial time. Clearly $G$ can be computed in polynomial time using a flood fill.

▶ **Definition 20.** *Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$ be a DFA with polynomial antichain growth. For a directed path $P = q_0 q_1 \dots q_l$ (not necessarily simple) in $G_{\mathcal{A}}$, let*

$$D(P) = |\{i \in \{0, \dots, l-1\} | (q_i, q_{i+1}) \in E(G'_{\mathcal{A}})\}| + \begin{cases} 1 \text{ if } |L_{q_m,q_l}| = \infty \\ 0 \text{ otherwise.} \end{cases},$$

*where $m = \max\{i+1 | (q_i, q_{i+1}) \in G'_{\mathcal{A}}\}$ if this exists, and 0 otherwise.*

Observe that if $|L_{q_m,q_l}| = \infty$ then we have $ww'^*w'' \subseteq L_{q_m,q_l}$ for some $w, w', w''$.

▶ **Lemma 21.** *Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$ be a DFA with polynomial antichain growth. Let $\mathcal{P}$ be the set of directed paths from $q_0$ to an element of $F$. Then the quantity*

$$D_{\mathcal{A}} = \max_{P \in \mathcal{P}} D(P)$$

*is well-defined and can be computed in polynomial time.*

**Proof.** To show that $D_{\mathcal{A}}$ is well-defined, observe that no directed cycle in $G$ contains an edge in $G'$. Indeed, suppose that $q_1 q_2 \dots q_1$ is a directed cycle in $G$, with $(q_1, q_2) \in E(G')$. Then we have $q_1 \xrightarrow{w} q_1$ and $q_1 \xrightarrow{w'} q_2$ for some $w \not\sim w' \in \Sigma^*$. Also we have $q_2 \xrightarrow{w''} q_1$ for some $w'' \in \Sigma^*$. But then $q_1 \xrightarrow{w'w''} q_1$ and $w'w'' \not\sim w$ by the Concatenation Lemma, contradicting polynomial antichain growth of $\mathcal{L}(\mathcal{A})$. Hence $D(P)$ is bounded.

For a polynomial time algorithm, first expand $G$ and $G'$ by adding a sink vertex $v_f$ for each $f \in F$. For each $q$ such that $|L_{q,f}| = \infty$ put $(q, v_f) \in E(G)$ and $(q, v_f) \in E(G')$. Then add a further vertex $v$ with $(f, v) \in E(G)$ and $(v_f, v) \in E(G)$ for all $f \in F$. Then $D_{\mathcal{A}}$ is precisely the maximum number of edges of $G'$ contained in a directed path from $q_0$ to $v$ in $G$.

Form the graph $G''$ on vertex-set $Q \cup \{v\}$ by $(v_1, v_2) \in E(G'')$ if and only if there is a path from $v_1$ to $v_2$ in $G$ containing a single edge of $G'$. Then we have that $G''$ is a DAG (by the first observation), and $D_{\mathcal{A}}$ is the longest path from $q_0$ to $v$ in $G''$, which can be found by a simple dynamic programming algorithm. ◀

We will show that the order of polynomial antichain growth of $\mathcal{L}(\mathcal{A})$ is precisely $D_{\mathcal{A}} - 1$.

▶ **Lemma 22.** *Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$ be a DFA with polynomial antichain growth. Then $\mathcal{L}(\mathcal{A})$ has polynomial antichain growth of order at least $D_{\mathcal{A}} - 1$.*

**Proof.** Let $P = q_0 q_1 \dots q_l$ be a path with $D(P) = D_{\mathcal{A}}$. Let $i_1, \dots, i_k$ be such that $(q_{i_j}, q_{i_j+1}) \in E(G'_{\mathcal{A}})$ for all $j$. Let $w_1, \dots, w_k, w'_1 \dots, w'_k, w \in \Sigma^*$ be such that $w_j \not\sim w'_j$ for all $j$, $q_{i_j} \xrightarrow{w_j} q_{i_j}$ for all $j$, $q_{i_j} \xrightarrow{w'_j} q_{i_{j+1}}$ for all $j < k$, $q_{i_k} \xrightarrow{w'_k} q_l$, and $q_0 \xrightarrow{w} q_{i_1}$.

Suppose that $|L_{q_m,q_l}| = \infty$ (with $m = i_k$ defined as in Definition 20), and let $w', w'', w''' \in \Sigma^*$ be such that $w'w''^*w''' \subseteq L_{q_m,q_l}$. Then $L = ww_1^*w_1'w_2^*w_2' \dots w_k^*w'w''^*w'''$ is an antichain family with polynomial growth of order $k = D_{\mathcal{A}} - 1$. Similarly if $|L_{q_m,q_l}| < \infty$, then $L = ww_1^*w_1'w_2^*w_2' \dots w_k^*w_k'$ is an antichain with polynomial growth of order $k-1 = D_{\mathcal{A}}-1$. ◀

We will now prove the upper bound. Our strategy will be to classify words by the edges of $G'$ they visit. We first show a preliminary lemma, which bounds the antichain growth from regions between edges of $G'$.

▶ **Lemma 23.** *Let $q_1, q_2 \in Q$, and let $L \subseteq L_{q_1,q_2}$ be the set of words such that no edges of $G'$ appear in the runs corresponding to elements of $L$. Then $L$ has antichain growth of order at most 0.*

**Proof.** Without loss of generality we may assume that $\mathcal{A}$ does not have any transitions labelled by more than a single letter (by introducing additional states if necessary; in particular we can set $Q' = Q \times \Sigma$ and ensure that $\delta'(q, x) \in Q \times \{x\}$ for all $x \in \Sigma$).

We will show that $L$ cannot contain two incomparable words that correspond after removal of loops to the same sets of simple paths in $G$.[1] Since $G$ is finite and hence contains only finitely many simple paths, this suffices to establish the result.

Suppose that $w_1 \not\sim w_2$ correspond to the same simple path $P$. Suppose that the first point of divergence of $w_1$ and $w_2$ is at state $q$; that is, that $w_1 = wx_1w_1'$ and $w_2 = wx_2w_2'$ with $x_1 \neq x_2 \in \Sigma$ and $q_1 \xrightarrow{w} q$ (see Figure 1). Without loss of generality we may assume that $q$ and $\delta(q, x_1)$ lie on $P$.

Since the path for $w_2$ corresponds to $P$ after removal of cycles, we must have that $w_2' = w_2''w_2'''$ with $q \xrightarrow{x_2w_2''} q$ and $q \xrightarrow{w_2'''} q_2$. But $w_1 \not\sim w_2$ and $x_1 \neq x_2$ so $x_1 \not\sim x_2$ and so $x_1 \not\sim x_2w_2''$. Hence $(q, \delta(q, x_1)) \in G'$, which is a contradiction. ◀

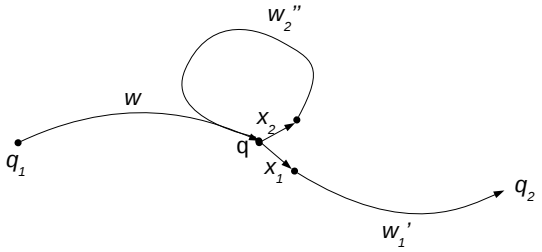


**Figure 1** The proof of Lemma 23

▶ **Lemma 24.** *Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$ be a DFA with polynomial antichain growth. Then $\mathcal{L}(\mathcal{A})$ has polynomial antichain growth of order at most $D_{\mathcal{A}} - 1$.*

**Proof.** We may assume without loss of generality that there is only a single accepting state, say $q_f$ (otherwise consider seperately the automata $\mathcal{A}_1, \dots, \mathcal{A}_{|F|}$ which agree with $\mathcal{A}$ except for having only a single accepting state; then on the one hand we have $D_{\mathcal{A}} = \max D_{\mathcal{A}_i}$, but on the other hand $\mathcal{L}(\mathcal{A}) = \bigcup \mathcal{A}_i$ which is a finite union and hence the order of antichain growth of $\mathcal{L}(\mathcal{A})$ is the maximum of the orders of growth of the $\mathcal{L}(\mathcal{A}_i)$).

[1] Note that since removal of loops may be done in many different ways, a single path may correspond to multiple simple paths. We are asserting that $L$ cannot contain two incomparable words which correspond to precisely the same *sets* of simple paths.

We classify words by the edges of $G'$ that appear in their accepting runs. We shall show that the set of words corresponding to a fixed sequence $P$ of $G'$-edges has antichain growth of order at most $D(P)$ (where $D(P) = |P| - 1$ or $|P|$ depending on whether the set of accepted words beginning at the last vertex of $P$ is finite). Since the number of relevant $G'$-edge sequences is finite (recalling that no edge of $G'$ is contained in a directed cycle in $G$ and so no $G'$-edge can appear more than once), this will suffice to establish the result.

Let $(q_1, q_1'), \ldots, (q_k, q_k')$ be a set of $G'$-edges. Then the set $L$ of words which have this sequence of $G'$-edges in their run is given by

$$L = L'_{q_0,q_1} X_1 L'_{q_1',q_2} X_2 L'_{q_2',q_3} \ldots X_k L'_{q_k',q_f},$$

where $X_i = \{x \in \Sigma \mid \delta(q_i, x) = q_i'\}$ and $L'_{q,q'} \subset L_{q,q'}$ is the set of words whose runs do not include edges of $G'$.

The $X_i$ are finite and hence have antichain growth of order $-1$. By Lemma 23 the $L'_{q_i',q_{i+1}}$ and also $L'_{q_0,q_1}$ and $L'_{q_k',q_f}$ have antichain growth of order at most 0. Moreover if $L_{q_k',q_f}$ is finite then so is $L'_{q_k',q_f} \subseteq L_{q_k',q_f}$ and so it has antichain growth of order $-1$. The result follows by Lemma 13. ◀

Combining Lemmas 21, 22 and 24 yields

▶ **Theorem 25.** *Let $\mathcal{A} = (Q, q_0, F, \Sigma, \delta)$ be a DFA with polynomial antichain growth. Then $\mathcal{L}(\mathcal{A})$ has polynomial antichain growth of order exactly $D_{\mathcal{A}} - 1$, which can be computed in polynomial time.*

We now show how to extend this algorithm to the case of NFA. Note that $D_{\mathcal{A}}$ as defined above is well-defined for NFA just as for DFA, and that the algorithm to compute it in polynomial time is equally applicable. It therefore remains to show that for NFA we also have that if $\mathcal{A}$ has polynomial antichain growth then it has antichain growth of order exactly $D_{\mathcal{A}} - 1$.

We do this by showing (Lemma 27) that $D_{\mathcal{A}}$ depends only on the language $\mathcal{L}(\mathcal{A})$, so that if $\mathcal{A}$ and $\mathcal{A}'$ are NFA with $\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}')$ then $D_{\mathcal{A}} = D_{\mathcal{A}'}$. Having shown this we then consider $\mathcal{A}'$ to be the determinisation of $\mathcal{A}$. This is a DFA with $\mathcal{L}(\mathcal{A}') = \mathcal{L}(\mathcal{A})$, and by Theorem 25 we have that $\mathcal{L}(\mathcal{A}')$ has polynomial antichain growth of order $D_{\mathcal{A}'} - 1 = D_{\mathcal{A}} - 1$.

We will first show (Lemma 26) that if $L = v_0 w_1^* v_1 w_2^* v_2 \ldots w_k^* v_k \subseteq \mathcal{L}(\mathcal{A})$ then there exists a single sequence of states $q_1, q_2, \ldots, q_k$ which essentially realises $L$ (that is, up to various offsets we have $v_i \in \mathcal{L}(\mathcal{A}_{q_i,q_{i+1}})$ and $w_i^* \in \mathcal{L}(\mathcal{A}_{q_i,q_i})$).

▶ **Lemma 26.** *Let $\mathcal{A} = (Q, q_0, F, \Sigma, \Delta)$ be an NFA such that $v_0 w_1^* v_1 w_2^* v_2 \ldots w_k^* v_k \subseteq \mathcal{L}(\mathcal{A})$. Then then there exists a sequence of states $q_1, q_2, \ldots, q_{k+1}$ and integers $m_1, m_2, \ldots m_k$, $m_1', m_2', \ldots, m_k'$ and $n_1, n_2, \ldots, n_k$ such that*

- **(i)** *$v_0 w_1^{m_1} \in \mathcal{L}(\mathcal{A}_{q_0,q_1})$ and $w_k^{m_k'} v_k \in \mathcal{L}(\mathcal{A}_{q_k,F})$,*
- **(ii)** *for all $0 < i < k$ we have $w_i^{m_i'} v_i w_{i+1}^{m_{i+1}} \in \mathcal{L}(\mathcal{A}_{q_i,q_{i+1}})$, and*
- **(iii)** *for all $0 < i \leq k$ we have $w_i^{n_i} \in \mathcal{L}(\mathcal{A}_{q_i,q_i})$.*

**Proof.** Consider an accepting run for $v_0 w_1^{|Q|+1} v_1 w_2^{|Q|+1} v_2 \ldots w_k^{|Q|+1} v_k \in \mathcal{L}(\mathcal{A})$, and write $q(s)$ for the state reached in this run after the word $s$. By the pigeon-hole principle, we must have $q(v_0 w_1^{m_1}) = q(v_0 w^{m_1+n_1}) = q_1$ (say) for some $m_1 \geq 0$ and some $n_1 > 0$ with $m_1 + n_1 \leq |Q| + 1$. Let $m_1' = |Q| + 1 - m_1 - n_1$. Similarly for each $i$ we have $q(v_1 w_1^{|Q|+1} v_2 \ldots w_i^{m_i}) = q(v_1 w_1^{|Q|+1} v_2 \ldots w_i^{m_i+n_i}) = q_i$ (say) for some $m_i \geq 0$ and $n_i > 0$ with $m_i + n_i \leq |Q| + 1$. Let $m_i' = |Q| + 1 - m_i - n_i$. Then these $q_i, m_i, m_i'$ and $n_i$ give the result. ◀

▶ **Lemma 27.** *Let $\mathcal{A}$ and $\mathcal{A}'$ be NFA with $\mathcal{L}(\mathcal{A}) = \mathcal{L}(\mathcal{A}')$. Then $D_{\mathcal{A}} = D_{\mathcal{A}'}$.*

**Proof.** Let $\mathcal{A} = (Q, q_0, F, \Sigma, \Delta)$ and $\mathcal{A}' = (Q', q_0', F', \Sigma, \Delta')$.

Suppose that $D_{\mathcal{A}'} = k$. Then by an identical argument to the proof of Lemma 22 we have that $v_0 w_1^* v_1 w_2^* v_2 \dots w_k^* v_k \subseteq \mathcal{L}(\mathcal{A}') = \mathcal{L}(\mathcal{A})$ for some $v_0, \dots, v_k, w_1, \dots, w_k \in \Sigma^*$ with $w_i \not\sim v_i$. Then by Lemma 26 there exists a sequence of states $q_1, q_2, \dots, q_{k+1} \in Q$ and integers $m_1, m_2, \dots, m_k, m_1', m_2', \dots m_k'$ and $n_1, n_2, \dots, n_k$ such that (i)–(iii) in the statement of the lemma hold. Now since $w_i \not\sim v_i$ we have $w_i^{k_i n_i} \not\sim w_i^{m_i'} v_i w_{i+1}^{m_{i+1}}$ for sufficiently large $k_i$ and so $D_{\mathcal{A}} \geq k = D_{\mathcal{A}'}$. Similarly $D_{\mathcal{A}'} \geq D_{\mathcal{A}}$, and hence $D_{\mathcal{A}} = D_{\mathcal{A}'}$. ◀

▶ **Theorem 28.** *Let $\mathcal{A}$ be an NFA with polynomial antichain growth. Then $\mathcal{L}(\mathcal{A})$ has polynomial antichain growth of order exactly $D_{\mathcal{A}} - 1$.*

**Proof.** Let $\mathcal{A}'$ be the powerset determinisation of $\mathcal{A}$, so $\mathcal{A}'$ is a DFA with $\mathcal{L}(\mathcal{A}') = \mathcal{L}(\mathcal{A})$. By Theorem 25, $\mathcal{L}(\mathcal{A}')$ has polynomial antichain growth of order exactly $D_{\mathcal{A}'} - 1$, and by Lemma 27 we have $D_{\mathcal{A}'} = D_{\mathcal{A}}$. ◀

## 5 Context-free languages

In [6], Ginsburg and Spanier show (Theorem 5.1) that a context-free grammar $G$ generates a bounded language if and only if the sets $L_A(G)$ and $R_A(G)$ are commutative for all nonterminals $A$, where $L_A$ and $R_A$ are respectively the sets of possible $w$ and $u$ in productions $A \stackrel{*}{\Rightarrow} wAu$. They also give an algorithm to decide this (which [4] improves to be in polynomial time).

We generalise this to our problem by showing that $G$ generates a language with polynomial antichain growth if and only $L_A(G)$ and also the sets $R_{A,w}(G)$ of possible $u$ for each fixed $w$ are chains, and that otherwise $\mathcal{L}(G)$ has exponential antichain growth. However, we will show that the problem of distinguishing the two cases is undecidable, by reduction from the CFG intersection emptiness problem.

Except where otherwise specified, we will assume all CFGs have starting symbol $S$ and that all nonterminals are accessible and co-accessible: for any nonterminal $A$ we have $S \stackrel{*}{\Rightarrow} uAu'$ for some $u, u' \in \Sigma^*$ and $A \stackrel{*}{\Rightarrow} v$ for some $v \in \Sigma^*$.

▶ **Definition 29.** *Let $G$ be a context-free grammar (CFG) over $\Sigma$. Then for any nonterminal $A$ let*

$$L_A(G) = \{w \in \Sigma^* \mid \exists\, u \in \Sigma^* : A \stackrel{*}{\Rightarrow} wAu\}.$$

▶ **Lemma 30.** *Let $G$ be a CFG over $\Sigma$ and $A$ some nonterminal such that $L_A(G)$ is not a chain. Then $\mathcal{L}(G)$ contains an exponential antichain.*

**Proof.** Since $L_A(G)$ is not a chain, we have $w_1, w_2, u_1, u_2$ with $w_1 \not\sim w_2$ such that $A \stackrel{*}{\Rightarrow} w_1 A u_1$ and $A \stackrel{*}{\Rightarrow} w_2 A u_2$. Now $A$ is accessible and co-accessible so also $S \stackrel{*}{\Rightarrow} uAu'$ and $A \stackrel{*}{\Rightarrow} v$ for some $u, u', v \in \Sigma^*$.

Hence $uw_{i_1} w_{i_2} \dots w_{i_k} v u_{i_k} u_{i_{k-1}} \dots u_{i_1} u' \subseteq \mathcal{L}(G)$, for any $i_1 i_2 \dots i_k \in \{1,2\}^*$. Write $\phi : (w_1 + w_2)^* \to (u_1 + u_2)^*$ for the map $w_{i_1} w_{i_2} \dots w_{i_k} \mapsto u_{i_k} u_{i_{k-1}} \dots u_{i_1}$ (with any ambiguity resolved arbitrarily).

Now $\{w_{i_1} w_{i_2} \dots w_{i_k} | i_1 \dots i_k \in \{1,2\}^*\} = (w_1 + w_2)^*$ is a quasiantichain by Lemma 6, clearly it is exponential and hence by Lemma 12 it contains an exponential antichain $L$. By the Concatenation Lemma we have that $L' = \{lv\phi(l) | l \in L\}$ is an antichain, and it is exponential because there is a bijection between $L$ and $L'$ such that the length of each word in

$L'$ exceeds the length of the corresponding word in $L$ by a factor of at most $\frac{|v|+\max(|u_1|,|u_2|)}{\min(|w_1|,|w_2|)}$. By the Prefixing and Postfixing Lemmas we have that $uL'u' \subseteq \mathcal{L}(G)$ is an exponential antichain. ◀

▶ **Definition 31.** *Let $G$ be a CFG over $\Sigma$. Then for any nonterminal $A$ and any $w \in \Sigma^*$, let*

$$R_{A,w}(G) = \{u \in \Sigma^* | A \stackrel{*}{\Rightarrow} wAu\}.$$

▶ **Lemma 32.** *Let $G$ be a CFG over $\Sigma$, $A$ some nonterminal and $w \in \Sigma^*$ such that $R_{A,w}(G)$ is not a chain. Then $\mathcal{L}(G)$ has exponential antichain growth.*

**Proof.** We have $v, w, u, u' \in \Sigma^*$ and $u_1 \not\sim u_2 \in \Sigma^*$ such that $S \stackrel{*}{\Rightarrow} uAu'$, $A \stackrel{*}{\Rightarrow} v$, $A \stackrel{*}{\Rightarrow} wAu_1$ and $A \stackrel{*}{\Rightarrow} wAu_2$. Let $L_i = uw^{2i}v(u_1u_2 + u_2u_1)^iu'$. Then $L_i$ is an antichain and $\bigcup_{i=1}^{\infty} L_i$ is an exponential antichain family. ◀

▶ **Lemma 33.** *Let $G$ be a CFG over $\Sigma$ such that $L_A(G)$ and $R_{A,w}(G)$ are chains for all nonterminals $A$ and all $w \in \Sigma^*$. Then $\mathcal{L}(G)$ has polynomial antichain growth.*

**Proof sketch.** Induction on the number of nonterminals, similarly to the proof of Theorem 16. ◀

Combining these three lemmas gives:

▶ **Theorem 34.** *Let $L$ be a context-free language. Then either $L$ has exponential antichain growth or $L$ has polynomial antichain growth.*

It is a straightforward exercise to show that the ambiguity of an NFA (the maximum number of accepting paths corresponding to a given word) can be represented as the width of a suitable context-free language, and hence Theorem 34 implies the well-known result that the ambiguity of an NFA has either polynomial or exponential growth (see Theorem 4.1 of [13]).

We now show that the problem of distinguishing the two cases of antichain growth is undecidable for context-free languages, by reduction from the CFG intersection emptiness problem. In fact, it is undecidable even to determine whether a given CFG generates a chain.

▶ **Definition 35.** CFG-Intersection *is the problem of determining whether two given CFGs have non-empty intersection.* CFG-Chain *is the problem of determining whether the language generated by a given CFG is a chain.* CFG-ExpAntichain *is the problem of determining whether the language generated by a given CFG has exponential antichain growth.*

▶ **Lemma 36.** CFG-Intersection *is undecidable.*

**Proof.** [5], Theorem 4.2.1. ◀

▶ **Lemma 37.** *There is a polynomial time reduction from* CFG-Intersection *to* CFG-Chain.

**Proof.** Let $G_1, G_2$ be arbitrary CFGs over alphabet $\Sigma$. Let $\widetilde{\Sigma} = \Sigma \cup \{0, 1\}$, with an arbitrary linear order on $\Sigma$, and $\Sigma < 0, \Sigma < 1$ but 0 and 1 incomparable. Let $\widetilde{G}$ be a CFG such that

$$\mathcal{L}(\widetilde{G}) = (\mathcal{L}(G_1)0) \cup (\mathcal{L}(G_2)1)$$

(which can trivially be constructed with polynomial blowup). Then $\mathcal{L}(\widetilde{G})$ is a chain if and only if $G_1 \cap G_2 = \emptyset$. ◀

▶ **Lemma 38.** *Let $L$ be a prefix-free chain. Then $L^*$ is a chain.*

**Proof.** Let $lw \not\sim l'w'$ be a minimum-length counterexample with $l, l' \in L$ and $w, w' \in L^*$. By minimality and the Prefixing Lemma we have that $l \neq l'$. Then by the Concatenation Lemma since $L$ is prefix-free we have that $l \not\sim l'$, which is a contradiction. ◀

▶ **Lemma 39.** *There is a polynomial time reduction from* CFG-Chain *to* CFG-ExpAntichain.

**Proof.** Let $G$ be a CFG over a partially ordered alphabet $\Sigma$. Let $\widetilde{\Sigma} = \Sigma \cup \{0\}$, with $\Sigma < 0$. Let $\widetilde{G}$ be a CFG such that $\mathcal{L}(\widetilde{G}) = (\mathcal{L}(G)0)^*$.

We claim that $\mathcal{L}(\widetilde{G})$ has exponential antichain growth if and only if $\mathcal{L}(G)$ is not a chain. Indeed, suppose that $l_1 \not\sim l_2 \in \mathcal{L}(G)$. Then $l_1 0 \not\sim l_2 0$ and so by Lemmas 6 and 12 we have that $(l_1 0 + l_2 0)^* \subseteq \mathcal{L}(\widetilde{G})$ contains an exponential antichain.

Conversely, suppose that $\mathcal{L}(G)$ is a chain. Then $\mathcal{L}(G)0$ is a prefix-free chain and so by Lemma 38 we have that $\mathcal{L}(\widetilde{G})$ is a chain. ◀

Combining these lemmas gives:

▶ **Theorem 40.** *The problems* CFG-Chain *and* CFG-ExpAntichain *are undecidable.*

## 6 Tree automata

In this section, we generalise the definition of the lexicographic ordering to tree languages, and prove a trichotomy theorem: regular tree languages have antichain growth which is either polynomial, exponential or doubly exponential.

Notation and definitions (other than for the lexicographic ordering) are taken from [3], to which the reader is referred for a more detailed treatment. Results in this section are stated without proof; all proofs may be found in the extended version [9].

▶ **Definition 41.** *Let $\mathcal{F}$ be a finite set of function symbols of arity $\geq 0$, and $\mathcal{X}$ a set of variables. Write $\mathcal{F}_p$ for the set of function symbols of arity $p$. Let $T(\mathcal{F}, \mathcal{X})$ be the set of terms over $\mathcal{F}$ and $\mathcal{X}$. Let $T(\mathcal{F})$ be the set of* ground terms *over $\mathcal{F}$, which is also the set of* ranked ordered trees *labelled by $\mathcal{F}$ (with rank given by arity as function symbols).*

For example, the set of ordered binary trees is $T(\mathcal{F})$, where $\mathcal{F} = \{f, g, c\}$ and $f$ has arity 2, $g$ arity 1 and $c$ arity 0.

Note that this generalises the definition of finite words over an alphabet $\Sigma$, by taking $\mathcal{F} = \Sigma \cup \{\epsilon\}$, giving each $a \in \Sigma$ arity one and $\epsilon$ arity zero.

A term $t$ is *linear* if no free variable appears more than once in $t$. A linear term mentioning $k$ free variables is a *$k$-ary context.*

▶ **Definition 42.** *Let $\mathcal{F}$ be equipped with a partial order $\preceq$. Then the lexicographic partial order induced by $\preceq$ on $T(\mathcal{F})$ is the relation $\preceq$ defined as follows: for any $f \in \mathcal{F}_p, f' \in \mathcal{F}_q$ and any $t_1, \ldots, t_p \in T(\mathcal{F})$ and $t'_1, \ldots, t'_q \in T(\mathcal{F})$ we have $f(t_1, \ldots, t_p) \preceq f'(t'_1, \ldots, t'_q)$ if and only if either $f \prec f'$ or $f = f'$ and $t_i \preceq t'_i$ for all $i$.*

Note that this generalises Definition 1, by taking $\epsilon \preceq a$ for all $a \in \Sigma$. As before we will write $t \sim t'$ if $t, t' \in T(\mathcal{F})$ are related by the lexicographic order; the definitions of *chain* and *antichain* are as before. To quantify antichain growth we need a notion of the size of a tree. The measure we will use will be *height*:

▶ **Definition 43.** *The* height *function $h : T(\mathcal{F}, \mathcal{X}) \to \mathbb{N}$ is defined by $h(x) = 0$ for all $x \in \mathcal{X}$, $h(t) = 1$ for all $t \in \mathcal{F}_0$ and $h(t(t_1, \ldots, t_n)) = 1 + \max(h(t_1, \ldots, t_n))$ for all $t \in \mathcal{F}_n$ $(n \geq 1)$ and $t_1, \ldots, t_n \in T(\mathcal{F}, \mathcal{X})$. For a language $L$, the set $\{t \in L \mid h(t) = k\}$ is denoted $L_{=k}$.*

For example, taking the earlier example of binary trees, ground terms of height 3 include $f(f(c,c),f(c,c))$, $f(c,f(c,c))$ and $g(f(c,c))$.

We say that $L$ has *doubly exponential* antichain growth if there is some $\epsilon$ such that the maximum size antichain in $L_{=n}$ exceeds $2^{2^{\epsilon n}}$ infinitely often.

▶ **Definition 44.** *A* nondeterministic finite tree automaton *(NFTA) over $\mathcal{F}$ is a tuple $\mathcal{A} = (Q, \mathcal{F}, Q_f, \Delta)$ where $Q$ is a set of unary states, $Q_f \subseteq Q$ is a set of final states, and $\Delta$ a set of transition rules of type $f(q_1(x_1), \ldots, q_n(x_n)) \to q(f(x_1, \ldots, x_n))$, for $f \in \mathcal{F}_n$, $q, q_1, \ldots, q_n \in Q$ and $x_1, \ldots, x_n \in \mathcal{X}$. The* move relation $\underset{\mathcal{A}}{\to}$ *is defined by applying a transition rule possibly inside a context and possibly with substitutions for the $x_i$. The reflexive transitive closure of $\underset{\mathcal{A}}{\to}$ is denoted $\overset{*}{\underset{\mathcal{A}}{\to}}$.*

*A tree $t \in T(\mathcal{F})$ is* accepted *by $\mathcal{A}$ if there is some $q \in Q_f$ such that $t \overset{*}{\underset{\mathcal{A}}{\to}} q(t)$. The set of trees accepted by $\mathcal{A}$ is denoted $\mathcal{L}(\mathcal{A})$.*

Again this generalises the definition of an NFA: put in transitions $\epsilon \to q(\epsilon)$ for all accepting states $q$, $a(q(x)) \to q'(a(x))$ whenever $q \in \Delta(q', a)$, and set $Q_f$ as the initial state.

The critical idea for the proof is to find the appropriate analogue of $L_q$. This turns out to be the set $P_q$ of binary contexts such that if the free variables are assigned state $q$ then the root can also be given state $q$. By analogy to the 'trousers decomposition' of differential geometry (also known as the 'pants decomposition'), we refer to such a context as a *pair of trousers*. It turns out that a sufficient condition for $L$ to have doubly exponential antichain growth is for $P_q$ to be non-empty for some $q$ (note that this does not depend on the particular partial order on $\Sigma$). On the other hand, if $P_q$ is empty for all $q$, then there is in a suitable sense no branching and so we have a similar situation to ordinary languages.

▶ **Definition 45.** *Let $\mathcal{A} = (Q, \mathcal{F}, Q_f, \Delta)$ be an NFTA and $q \in Q$. A linear term $t \in T(\mathcal{F}, \{x_1, x_2\})$ is a* pair of trousers *with respect to $q$ if $x_1, x_2$ appear in $t$ and $t[x_1 \leftarrow q(x_1), x_2 \leftarrow q(x_2)] \overset{*}{\underset{\mathcal{A}}{\to}} q(t)$. The set of pairs of trousers with respect to $q$ is denoted $P_q(\mathcal{A})$.*

▶ **Lemma 46.** *Let $\mathcal{A} = (Q, \mathcal{F}, Q_f, \Delta)$ be a reduced NFTA. If there exists some $q \in Q$ such that $P_q(\mathcal{A})$ is non-empty, then $\mathcal{L}(\mathcal{A})$ contains a doubly exponential antichain.*

▶ **Lemma 47.** *Let $\mathcal{A} = (Q, \mathcal{F}, Q_f, \Delta)$ be a reduced NFTA such that $P_q(\mathcal{A}) = \emptyset$ for all $q \in Q$. Then $\mathcal{L}(\mathcal{A})$ has at most exponential growth.*

In the case where there are no pairs of trousers, the situation is essentially equivalent to ordinary NFA, and so we have a further dichotomy between exponential and polynomial antichain growth. To show this, we define a set equivalent to $L_{q,q}$, and show that we have polynomial growth if it is a chain and exponential growth otherwise.

▶ **Definition 48.** *Let $\mathcal{A} = (Q, \mathcal{F}, Q_f, \Delta)$ be an NFTA, and $q \in Q$. Define $\mathcal{L}_q(\mathcal{A}) \subseteq T(\mathcal{F}, \{x_1\})$ to be the set of unary contexts $t$ such that $t[x_1 \leftarrow q(x_1)] \overset{*}{\underset{\mathcal{A}}{\to}} q(t)$.*

Note that unary contexts are linear terms in which *exactly* one free variable appears, so $\mathcal{L}_q(\mathcal{A})$ does not contain ground terms. Note also that $x_1 \in \mathcal{L}_q(\mathcal{A})$ for any $\mathcal{A}$.

To give meaning to the statement '$\mathcal{L}_q(\mathcal{A})$ is a chain', we must extend the definition of the lexicographic order from the set $T(\mathcal{F})$ of ground terms to the set $T(\mathcal{F}, \{x_1\})$ of unary contexts. We do this by extending the relation $\preceq$ on $\mathcal{F}$ to $\mathcal{F} \cup \{x_1\}$ by $x_1 \preceq f$ for all $f \in \mathcal{F}$, and extending this to the lexicographic order as before.

▶ **Lemma 49.** *Let $\mathcal{A} = (Q, \mathcal{F}, Q_f, \Delta)$ be a reduced NFTA such that $P_q(\mathcal{A}) = \emptyset$ for all $q$. Then $\mathcal{L}(\mathcal{A})$ has polynomial antichain growth if $\mathcal{L}_q(\mathcal{A})$ is a chain for all $q$, and otherwise $\mathcal{L}(\mathcal{A})$ has exponential antichain growth.*

Combining these lemmas gives

▶ **Theorem 50.** *Let $L$ be a regular tree language over a partially ordered alphabet. Then $L$ has either doubly exponential antichain growth, singly exponential antichain growth, or polynomial antichain growth.*

The special case of the trivial partial order (in which elements are only comparable to themselves) yields the fact that the language growth of any regular tree language is either polynomial, exponential or doubly exponential, which may not have previously appeared in the literature.

▶ **Corollary 51.** *Let $L$ be a regular tree language. Then $L$ has either doubly exponential language growth, singly exponential language growth or polynomial language growth.*

Finally, we show that there is a polynomial algorithm to detect doubly exponential growth, by determining whether or not the language of a given NFTA contains a pair of trousers.

▶ **Theorem 52.** *There exists a polynomial time algorithm to determine whether the language of a given NFTA has doubly exponential growth.*

## 7 Open problems

It is remarkable that, many decades after the discovery of the dichotomy between polynomial and exponential language growth, and 11 years after the work of Gawrychowski, Krieger, Rampersad and Shallit [4], it remains unknown whether there is an efficient algorithm to compute the order of exponential language growth of a given NFA. Consequently we consider that resolving this question (by providing either a polynomial-time algorithm or an appropriate hardness result) is the most important open problem in this area.

For a DFA, on the other hand, the order of exponential language growth is easily computed as the spectral radius of the transition matrix. However, it is not clear how such 'algebraic' methods can be applied to the case of antichain growth, and so a second open problem is to find a polynomial-time algorithm to compute the order of exponential antichain growth for DFA. Such a result would have immediate application to the field of quantified information flow, since it would allow one to compute the flow rate in the 'dangerous' linear case, at the cost of determinising the automaton representing the system (with overhead roughly corresponding to the amount of hidden state the system contains).

The final problem in this direction is the combination of the preceding two: to find a polynomial-time algorithm to compute the order of exponential antichain growth for a given NFA.

Alternatively we may wish to ask not about growth rates in the asymptotic limit, but instead about the precise width of $L_{=n}$ or $L_{\leq n}$ for given $n$. This is particularly relevant to applications in computer security, where we may want not just an approximation 'for sufficiently large $n$' but a concrete guarantee. For the case of a language given as a DFA and $n$ given in unary there is a straightforward dynamic programming algorithm to compute these quantities (for details see p.89 of [8]), but what about for NFA and for more concise representations of $n$?

Finally we pose a more speculative question: what other phenomena, apart from information flow, can antichains with respect to the lexicographic order usefully represent?

## References


1. Graham Brightwell. Random k-dimensional orders: Width and number of linear extensions. *Order*, 9(4):333–342, 1992.
2. E. Rodney Canfield. On a problem of Rota. *Advances in Mathematics*, 29(1):1–10, 1978.
3. H. Comon, M. Dauchet, R. Gilleron, C. Löding, F. Jacquemard, D. Lugiez, S. Tison, and M. Tommasi. Tree automata techniques and applications. Available on: `http://www.grappa.univ-lille3.fr/tata`, 2007. release October, 12th 2007.
4. Paweł Gawrychowski, Dalia Krieger, Narad Rampersad, and Jeffrey Shallit. Finding the growth rate of a regular of [sic] context-free language in polynomial time. In *Developments in Language Theory*, pages 339–358. Springer, 2008.
5. Seymour Ginsburg. *The Mathematical Theory of Context Free Languages*. McGraw-Hill Book Company, 1966.
6. Seymour Ginsburg and Edwin H. Spanier. Bounded algol-like languages. *Transactions of the American Mathematical Society*, 113(2):333–368, 1964.
7. D. Kleitman, M. Edelberg, and D. Lubell. Maximal sized antichains in partial orders. *Discrete Mathematics*, 1(1):47 – 53, 1971.
8. David Mestel. *Quantifying information flow*. PhD thesis, University of Oxford, 2018.
9. David Mestel. Widths of regular and context-free languages. *CoRR*, abs/1709.08696, 2018. URL: `http://arxiv.org/abs/1709.08696`.
10. David Mestel. Quantifying information flow in interactive systems. In *Proc. 32nd IEEE Computer Security Foundations Symposium (CSF '19)*, pages 414–427, June 2019.
11. G. W. Peck. Maximum antichains of rectangular arrays. *Journal of Combinatorial Theory, Series A*, 27(3):397–400, 1979.
12. Emanuel Sperner. Ein Satz über Untermengen einer endlichen Menge. *Mathematische Zeitschrift*, 27(1):544–548, 1928.
13. Andreas Weber and Helmut Seidl. On the degree of ambiguity of finite automata. *Theor. Comput. Sci.*, 88(2):325–349, October 1991.
14. Douglas B. West. Extremal problems in partially ordered sets. In Ivan Rival, editor, *Ordered Sets: Proceedings of the NATO Advanced Study Institute held at Banff, Canada, August 28 to September 12, 1981*, pages 473–521. Springer Netherlands, Dordrecht, 1982.