Subexponential estimates in Shirshov’s theorem on height

A. Ya. Belov and M. I. Kharitonov

Abstract. Suppose that $F_{2,m}$ is a free 2-generated associative ring with the identity $x^m = 0$. In 1993 Zelmanov put the following question: is it true that the nilpotency degree of $F_{2,m}$ has exponential growth?

We give the definitive answer to Zelmanov’s question by showing that the nilpotency class of an $l$-generated associative algebra with the identity $x^d = 0$ is smaller than $\Psi(d, d, l)$, where

$$\Psi(n, d, l) = 2^{18} l(nd)^{3 \log_3 (nd)+13} d^2.$$ 

This result is a consequence of the following fact based on combinatorics of words. Let $l$, $n$ and $d > n$ be positive integers. Then all words over an alphabet of cardinality $l$ whose length is not less than $\Psi(n, d, l)$ are either $n$-divisible or contain $x^d$; a word $W$ is $n$-divisible if it can be represented in the form $W = W_0 W_1 \cdots W_n$ so that $W_1, \ldots, W_n$ are placed in lexicographically decreasing order. Our proof uses Dilworth’s theorem (according to V. N. Latyshev’s idea). We show that the set of not $n$-divisible words over an alphabet of cardinality $l$ has height $h < \Phi(n, l)$ over the set of words of degree $\leq n - 1$, where

$$\Phi(n, l) = 2^{87} l \cdot n^{12 \log_3 n+48}.$$ 

Bibliography: 40 titles.

Keywords: Shirshov theorem on height, word combinatorics, $n$-divisibility, Dilworth theorem, Burnside-type problems.

§ 1. Introduction

1.1. Shirshov theorem on height. In 1958 Shirshov proved his famous theorem on height [1], [2].

Definition 1.1. A word $W$ is called $n$-divisible if $W$ can be represented in the form $W = v u_1 u_2 \cdots u_n$ so that $u_1 > u_2 > \cdots > u_n$.

In this case any nonidentical permutation $\sigma$ of subwords $u_i$ produces a word $W_{\sigma} = v u_{\sigma(1)} u_{\sigma(2)} \cdots u_{\sigma(n)}$ that is lexicographically smaller than $W$. Some authors take this feature as the definition of $n$-divisibility.

AMS 2010 Mathematics Subject Classification. Primary 16R10, 68R15.
Definition 1.2. A PI-algebra $A$ is called an algebra of bounded height $h = \text{Ht}_Y(A)$ over a set of words $Y = \{u_1, u_2, \ldots\}$ if $h$ is the minimal integer such that any word $x$ from $A$ can be represented in the form

$$x = \sum_i \alpha_i u_{j(i,1)}^{k(i,1)} u_{j(i,2)}^{k(i,2)} \cdots u_{j(i,r_i)}^{k(i,r_i)},$$

where the $\{r_i\}$ do not exceed $h$. The set $Y$ is called a Shirshov basis for $A$.

If no misunderstanding can occur, we use $h$ instead of $\text{Ht}_Y(A)$.

Shirshov Theorem on height ([1], [2]). The set of not $n$-divisible words in a finitely generated algebra with an admissible polynomial identity has bounded height $H$ over the set of words of degree not exceeding $n - 1$.

The Burnside-type problems related to the height theorem are considered in [3]. The authors believe that the Shirshov theorem on height is a fundamental fact in word combinatorics independently of its applications to PI-theory. (All our proofs are elementary and fit in the framework of word combinatorics.) Unfortunately, the experts in combinatorics have not sufficiently appraised this fact yet. As regards the notion of $n$-divisibility itself, it seems to be fundamental as well. Latyshev’s estimates on $\xi_n(k)$, the number of non-$n$-divisible polylinear words in $k$ symbols, have led to fundamental results in PI-theory. At the same time, this number is nothing but the number of arrangements of integers from 1 to $k$ such that no $n$ integers (not necessarily consecutive) are placed in decreasing order. Furthermore it is the number of permutationally ordered sets of diameter $n$ consisting of $k$ elements. (A set is called permutationally ordered if its ordering is the intersection of two linear orderings, the diameter of an ordered set is the length of its maximal antichain.)

The height theorem implies the solution of a number of problems in ring theory. Suppose an associative algebra over a field satisfies a polynomial identity $f(x_1, \ldots, x_n) = 0$. It is possible to prove that then it satisfies an admissible polynomial identity (that is, a polynomial identity with coefficient 1 at some term of higher degree):

$$x_1 x_2 \cdots x_n = \sum_\sigma \alpha_\sigma x_{\sigma(1)} x_{\sigma(2)} \cdots x_{\sigma(n)},$$

where the $\alpha_\sigma$ belong to the ground field. In this case, if $W = vu_1 u_2 \cdots u_n$ is $n$-divisible then for any permutation $\sigma$ the word $W_\sigma = vu_{\sigma(1)} u_{\sigma(2)} \cdots u_{\sigma(n)}$ is lexicographically smaller than $W$, and thus an $n$-divisible word can be represented as a linear combination of lexicographically smaller words. Hence a PI-algebra has a basis consisting of non-$n$-divisible words. By the Shirshov theorem on height, a PI-algebra has bounded height. In particular, if a PI-algebra satisfies $x^n = 0$ then it is nilpotent, that is, all of its words of length exceeding some $N$ are identically zero. Surveys on the height theorem can be found in [4]–[8].

This theorem implies an affirmative solution of the Kurosh problem and of other Burnside-type problems for PI-rings. Indeed, if $Y$ is a Shirshov basis and all its elements are algebraic, then the algebra $A$ is finite-dimensional. Thus the Shirshov theorem explicitly indicates a set of elements whose algebraicity makes the whole algebra finite-dimensional. This theorem implies the following result.
Corollary 1.1 (Berele). Let $A$ be a finitely generated PI-algebra. Then

$$\text{GK}(A) < \infty.$$  

Here $\text{GK}(A)$ is the Gelfand-Kirillov dimension of the algebra $A$, that is,

$$\text{GK}(A) = \lim_{n \to \infty} \frac{\ln V_A(n)}{\ln(n)},$$

where $V_A(n)$ is the growth function of $A$, the dimension of the vector space generated by the words of degree not greater than $n$ in the generators of $A$.

Indeed, it suffices to observe that the number of solutions for the inequality $k_1|v_1| + \cdots + k_h|v_h| \leq n$ with $h \leq H$ exceeds $N^H$, so that

$$\text{GK}(A) \leq \text{Ht}(A).$$

The number $m = \text{deg}(A)$ will mean the degree of the algebra, or the minimal degree of an identity valid in $A$. The number $n = \text{Pid}(A)$ is the complexity of $A$, or the maximal $k$ such that $M_k$, the algebra of matrices of size $k$, belongs to the variety $\text{Var}(A)$ generated by $A$.

Instead of the notion of height, it is more suitable to use the close notion of essential height.

Definition 1.3. An algebra $A$ has essential height $h = H_{\text{Ess}}(A)$ over a finite set $Y$, called an $s$-basis for $A$, if there exists a finite set $D \subset A$ such that $A$ is linearly representable by elements of the form $t_1 \cdots t_l$, where $l \leq 2h+1$, and $\forall i \ t_i \in D \lor t_i = y_i^{k_i}$; $y_i \in Y$ and the set of $i$ such that $t_i \not\in D$ contains at most $h$ elements. The essential height of a set of words is defined similarly.

Informally speaking, any long word is a product of periodic parts and ‘gaskets’ of restricted length. The essential height is the number of periodic parts, and the ordinary height takes account of ‘gaskets’ as well.

The height theorem suggests the following questions.

1. To which classes of rings can the height theorem be extended?
2. Over which $Y$ has the algebra $A$ bounded height? In particular, what sets of words can be taken for $\{v_i\}$?
3. What is the structure of the degree vector $(k_1, \ldots, k_h)$? First of all, what sets of its components are essential, that is, what sets of $k_i$ can be unbounded simultaneously? What is the value of essential height? Is it true that the set of degree vectors has some regularity properties?
4. What estimates for the height are possible?

Let us discuss the above questions.

1.2. Nonassociative generalizations. The height theorem was extended to some classes of near-associative rings. Pchelintsev [9] proved it for the alternative and the $(-1; 1)$ cases, Mishchenko [10] obtained an analogue of the height theorem for Lie algebras with a sparse identity. Belov [11] proved the height theorem for some class of rings asymptotically close to associative rings. In particular, this class contains alternative and Jordan PI-algebras.
1.3. Shirshov bases. Suppose $A$ is a PI-algebra and a subset $M \subseteq A$ is its $s$-basis. If all elements of $M$ are algebraic over $K$, then $A$ is finite-dimensional (the Kurosh problem). Boundedness of the essential height over $Y$ implies ‘an affirmative solution of the Kurosh problem over $Y’$. The converse is less trivial.

**Theorem 1** (Belov). a) Suppose $A$ is a graded PI-algebra, $Y$ is a finite set of homogeneous elements. If for all $n$ the algebra $A/Y^{(n)}$ is nilpotent, then $Y$ is an $s$-basis for $A$. Moreover, if $Y$ generates $A$ as an algebra, then $Y$ is a Shirshov basis for $A$.

b) Suppose $A$ is a PI-algebra, $M \subseteq A$ is a Kurosh subset in $A$. Then $M$ is an $s$-basis for $A$.

Let $Y^{(n)}$ denote the ideal generated by the $n$th powers of elements from $Y$. A set $M \subseteq A$ is called a Kurosh set if any projection $\pi: A \otimes K[X] \to A'$ such that the image $\pi(M)$ is entire over $\pi(K[X])$ is finite-dimensional over $\pi(K[X])$. The following example motivates this definition. Suppose $A = \mathbb{Q}[x, 1/x]$. Any projection $\pi$ such that $\pi(x)$ is algebraic has a finite-dimensional image. However, the set $\{x\}$ is not an $s$-basis for $\mathbb{Q}[x, 1/x]$. Thus boundedness of the essential height is a noncommutative generalization of the property of entireness.

1.4. Shirshov bases consisting of words. The Shirshov bases consisting of words are described by the following result.

**Theorem 2** ([4], [12]). A set $Y$ of words is a Shirshov basis for an algebra $A$ if and only if for any word $u$ of length not exceeding $m = \text{Pid}(A)$, the complexity of $A$, the set $Y$ contains a word cyclically conjugate to some power of $u$.

A similar result was obtained independently by Ciocanu and Drensky. Problems related to local finiteness of algebras and to algebraic sets of words of degree not exceeding the complexity of the algebra were investigated in [7], [13]–[18]. Questions relating to generalization of the independence theorem were considered in these papers as well.

1.5. Essential height. Clearly the Gelfand-Kirillov dimension is estimated by the essential height. Furthermore an $s$-basis is a Shirshov basis if and only if it generates $A$ as an algebra. In the representable case the converse is also true.

**Theorem 3** (Belov [4]). Suppose $A$ is a finitely generated representable algebra and $H_{\text{Ess}}(A) < \infty$. Then $H_{\text{Ess}}(A) = \text{GK}(A)$.

**Corollary 1.2** (Markov). The Gelfand-Kirillov dimension of a finitely generated representable algebra is an integer.

**Corollary 1.3.** If $H_{\text{Ess}}(A) < \infty$ and $A$ is representable then $H_{\text{Ess}}(A)$ is independent of the choice of the $s$-basis $Y$.

In this case the Gelfand-Kirillov dimension also is equal to the essential height by virtue of the local representability of relatively free algebras.

**Structure of degree vectors.** Although in the representable case the Gelfand-Kirillov dimension and the essential height behave well, even in this case the set of degree vectors may have a bad structure, namely, it can be the complement to the set of solutions of a system of exponential-polynomial Diophantine equations [4].
That is why there exists an instance of a representable algebra with a transcendent Hilbert series. However for a relatively free algebra, the Hilbert series is rational [19].

1.6. \(n\)-divisibility and the Dilworth theorem. The significance of the notion of \(n\)-divisibility goes beyond the limits of Burnside-type problems. This notion also plays a role in the investigation of polylinear words and the estimation of their number; a word is polylinear if each letter occurs in it at most once. Latyshev applied the Dilworth theorem for the estimation of the number of not \(m\)-divisible polylinear words of degree \(n\) over the alphabet \(\{a_1, \ldots, a_n\}\). The estimate is \((m - 1)^{2n}\) and is rather sharp. Let us recall this theorem.

Dilworth’s Theorem. Let \(n\) be the maximal number of elements in an antichain of a given fixed partially ordered set \(M\). Then \(M\) can be divided into \(n\) disjoint chains.

Consider a polylinear word \(W\) consisting of \(n\) letters. Put \(a_i \succ a_j\) if \(i > j\) and the letter \(a_i\) is located in \(W\) to the right of \(a_j\). The condition of not \(k\)-divisibility means the absence of an antichain consisting of \(n\) elements. Then by Dilworth’s theorem all positions (and the letters \(a_i\) as well) split into \(n - 1\) chains. Attach a specific colour to each chain. Then the colouring of positions and of letters uniquely determines the word \(W\). Furthermore, the number of these colourings does not exceed 

\[(n - 1)^k \times (n - 1)^k = (n - 1)^{2k}.
\]

The above estimate implies the validity of polylinear identities corresponding to an irreducible module whose Young diagram includes the square of size \(n^4\). This in turn enables one, firstly, to obtain a transparent proof for Regev’s theorem which asserts that a tensor product of PI-algebras is a PI-algebra as well; secondly, to establish the existence of a sparse identity in the general case and of a Capelli identity in the finitely generated case (and thus to prove the theorem on nilpotency of the radical); and thirdly, to realize Kemer’s ‘supertrick’ that reduces the study of identities in general algebras to that of super-identities in finitely generated superalgebras of zero characteristic. Close questions are considered in [20]–[22].

Problems related to the enumeration of polylinear words which are not \(n\)-divisible are interesting in their own right. (For example, there exists a bijection between not 3-divisible words and Catalan numbers.) On the one hand this is a purely combinatorial problem, but on the other hand, it is related to the set of codimensions for the general matrix algebra. The study of polylinear words seems to be of great importance. Latyshev (see, for instance, [23]) has stated the problem of finite-basedness of the set of leading polylinear words for a \(T\)-ideal with respect to taking overwords and to isotonous substitutions. This problem implies the Specht problem for polylinear polynomials and is closely related to the problem of the weak Noetherian property for the group algebra of an infinite finitary symmetric group over a field of positive characteristic (for zero characteristic this was established by Zalessky). To solve the Latyshev problem it is necessary to translate properties of \(T\)-ideals to the language of polylinear words. In [4], [11] an attempt was made to realize a project of translation of structure properties of algebras to the language of word combinatorics. Translation to the language of polylinear words is simpler and enables one to get some information on words of a general form.
In this paper we transfer Latyshev’s technique to the non-polylinear case, and this enables us to obtain a subexponential estimate in the Shirshov-height theorem. Chelnokov suggested the idea of this transfer in 1996.

1.7. Estimates for the height. The original Shirshov’s proof, being purely combinatorial (it was based on the technique of elimination developed by him for Lie algebras, in particular in the proof of the theorem on freeness), nevertheless implied only primitively recursive estimates. Later Kolotov [24] obtained an estimate $\text{Ht}(A) \leq l^n$ ($n = \text{deg}(A)$, $l$ is the number of generators). Belov in [25] showed that $\text{Ht}(n, l) < 2nl^{n+1}$. The exponential estimate in the Shirshov height theorem was also presented in [12], [26], [27]. The above estimates were sharpened by Klein [28], [29]. In 2001, Chibrikov proved in [30] that $\text{Ht}(4, l) \geq (7k^2 - 2k)$. Kharlamov in [27], [31], [32] obtained estimates for the structure of piecewise periodicity. In 2011, Lopatin [33] obtained the following result.

**Theorem 4.** Let $C_{n,l}$ be the nilpotency degree of a free $l$-generated algebra satisfying $x^n = 0$, and let $p$ be the characteristic of the ground field of the algebra, greater than $\frac{n}{2}$. Then

$$C_{n,l} < 4 \cdot 2^{n/2l}. \quad (1)$$

By definition $C_{n,l} \leq \Psi(n, n, l)$. Observe that for small $n$ the estimate (1) is smaller than the estimate $\Psi(n, n, l)$ established in this paper but for growing $n$ the estimate $\Psi(n, n, l)$ is asymptotically better than (1).

Zelmanov put the following question in the Dniester Notebook [34] in 1993:

**Question 1.1.** Let $F_{2,m}$ be the free 2-generated associative ring with identity $x^m = 0$. Is it true that the nilpotency class of $F_{2,m}$ grows exponentially in $m$?

Our paper answers Zelmanov’s question as follows: the nilpotency class in question grows subexponentially.

1.8. The results obtained. The main result of the paper is as follows: the nilpotency class in question grows subexponentially.

**Theorem 5.** The height of the set of not $n$-divisible words over an alphabet of cardinality $l$ relative to the set of words of length less than $n$ does not exceed $\Phi(n, l)$, where

$$\Phi(n, l) = E_1 l \cdot n^{E_2 + 12\log_3 n}, \quad E_1 = 4^{21\log_3 4 + 17}, \quad E_2 = 30\log_3 4 + 10.$$

This theorem after some coarsening and simplification of the estimate implies that for fixed $l$ and $n \to \infty$ we have

$$\Phi(n, l) < 2^{87l} \cdot n^{12\log_3 n + 48} = n^{12(1 + o(1))\log_3 n},$$

and for fixed $n$ and $l \to \infty$ we have

$$\Phi(n, l) < C(n)l.$$

**Corollary 1.4.** The height of an $l$-generated PI-algebra with an admissible polynomial identity of degree $n$ over the set of words of length less than $n$ does not exceed $\Phi(n, l)$. 
Moreover we prove a subexponential estimate which is better for small $n$:

**Theorem 6.** The height of the set of not $n$-divisible words over an alphabet of cardinality $l$ relative to the set of words of length less than $n$ does not exceed $\Phi(n, l)$, where

$$\Phi(n, l) = 2^{40} l \cdot n^{38 + 8 \log_2 n}.$$

In particular we obtain subexponential estimates for the nilpotency index of $l$-generated nil-algebras of degree $n$ for an arbitrary characteristic.

The second main result of our paper is the following theorem.

**Theorem 7.** Let $l, n$ and $d \geq n$ be positive integers. Then all $l$-generated words of length not less than $\Psi(n, d, l)$ either contain $x^d$ or are $n$-divisible. Here

$$\Psi(n, d, l) = 4^{5 + 3 \log_3 4} l (nd)^{3 \log_3 (nd)} + (5 + 6 \log_3 4) d^2.$$

This theorem after some coarsening and simplification of the estimate implies that for fixed $l$ and $nd \to \infty$ we have

$$\Psi(n, d, l) < 2^{18} l (nd)^{3 \log_3 (nd) + 13} d^2 = (nd)^{3(1 + o(1)) \log_3 (nd)},$$

and for fixed $n$ and $l \to \infty$ we have

$$\Psi(n, d, l) < C(n, d).$$

**Corollary 1.5.** Let $l, d$ be positive integers, and let an associative $l$-generated algebra $A$ satisfy $x^d = 0$. Then its nilpotency index is less than $\Psi(d, d, l)$.

Moreover we prove a subexponential estimate which is better for small $n$ and $d$:

**Theorem 8.** Let $l, n$ and $d \geq n$ be positive integers. Then all $l$-generated words of length not less than $\Psi(n, d, l)$ either contain $x^d$ or are $n$-divisible. Here

$$\Psi(n, d, l) = 256 l (nd)^{2 \log_2 (nd) + 10} d^2.$$

For a real number $x$ put $\lceil x \rceil := \lceil -x \rceil$. Thus we replace noninteger numbers by the closest greater integers.

Proving Theorem 5 we also prove the following theorem on estimation of the essential height:

**Theorem 9.** The essential height of an $l$-generated PI-algebra with an admissible polynomial identity of degree $n$ over the set of words of length less than $n$ is less than $\Upsilon(n, l)$, where

$$\Upsilon(n, l) = 2 n^{3 \log_3 n^3 + 4} l.$$

In [35] it is established that the nilpotency index of an $l$-generated nil-semiring of degree $n$ equals the nilpotency index of an $l$-generated nilring of degree $n$, where addition is not necessarily commutative. (The paper also contains examples of non-nilpotent nil-nearrings of index 2.) Thus our results extend to the case of semirings as well.
1.9. On estimates from below. Let us compare the results obtained with the estimate for the height from below. The height of an algebra \( A \) is no less than its Gelfand-Kirillov dimension \( \text{GK}(A) \). For the algebra of \( l \)-generated general matrices of order \( n \) this dimension equals \( (l - 1)n^2 + 1 \) (see [36] as well as [37]). At the same time, the minimal degree of an identity in this algebra is \( 2n \) by the Amitsur-Levitsky theorem. We have the following result.

**Proposition 1.1.** The height of an \( l \)-generated PI-algebra of degree \( n \) and of the set of not \( n \)-divisible words over an alphabet of cardinality \( l \) is no less than \( (l - 1)n^2/4 + 1 \).

Estimates from below for the nilpotency index were established by Kuzmin in [38]. He gave an example of a 2-generated algebra with identity \( x^n = 0 \), such that its nilpotency index exceeds \( (n^2 + n - 2)/2 \). The problem of finding estimates from below is considered in [31].

At the same time, for zero characteristic and a countable set of generators, Razmyslov (see for instance [39]) obtained an upper estimate for the nilpotency index, namely \( n^2 \).

First we will prove Theorem 7, and in the following section we will deal with estimates for the essential height, that is, for the number of distinct periodic pieces in a not \( n \)-divisible word.

The authors are grateful to V. N. Latyshev, A. V. Mikhalev and all participants of the “Ring theory” seminar for their attention to our work, as well as to the participants of the seminar at the Moscow Institute for Physics and Technology under the supervision of A. M. Raigorodskii.

§ 2. Estimates on the occurrence of degrees of subwords

2.1. The outline of the proof for Theorem 7. Lemmas 2.1, 2.2 and 2.3 describe sufficient conditions for the presence of a period of length \( d \) in a not \( n \)-divisible word \( W \). Lemma 2.4 connects \( n \)-divisibility of a word \( W \) with the set of its tails. Further we choose some specific subset in the set of tails of \( W \), such that we can apply Dilworth’s theorem. After that we colour the tails and their first letters according to their location in chains obtained by an application of Dilworth’s theorem.

We have to know the position in any chain where neighbouring tails begin to differ. It is of interest what the ‘frequency’ of this position is in a \( p \)-tail for some \( p \leq n \). Further we somewhat generalize our reasoning dividing tails into segments consisting of several letters each and determining the segment containing the position where neighbouring tails begin to differ. Lemma 3.2 connects the ‘frequencies’ in question for \( p \)-tails and \( kp \)-tails for \( k = 3 \).

To complete the proof, we construct a hierarchical structure based on Lemma 3.2, that is, we consecutively consider segments of \( n \)-tails, subsegments of these segments and so on. Furthermore we consider the greatest possible number of tails in the subset to which Dilworth’s theorem is applied, and then we estimate from above the total number of tails and hence of the letters in the word \( W \).
2.2. Periodicity and $n$-divisibility properties. Let $a_1, \ldots, a_l$ be the alphabet used for constructing words. The ordering $a_1 < a_2 < \cdots < a_l$ induces a lexicographical ordering for words over the alphabet. For convenience, we introduce the following definitions.

**Definition 2.1.** a) If a word $v$ includes a subword of the form $u^t$ then we say that $v$ includes a period of length $t$.

b) If a word $u$ is the beginning of a word $v$ then these words are called incomparable.

c) A word $v$ is a tail of a word $u$ if there exists a word $w$ such that $u = vw$.

d) A word $v$ is a $k$-tail of a word $u$ if $v$ consists of the first $k$ letters of some tail $u$.

d*) A $k$-beginning is the same as a $k$-tail.

e) A word $u$ is to the left of a word $v$ if $u$ begins to the left of the beginning of $v$.

Let $|u|$ denote the length of a word $u$.

The proof uses the following sufficient conditions for the presence of a period.

**Lemma 2.1.** In a word $W$ of length $x$ either the first $[x/d]$ tails are pairwise comparable or $W$ includes a period of length $d$.

**Proof.** Suppose $W$ includes no word of the form $u^d$. Consider the first $[x/d]$ tails. Suppose some two of them, say $v_1$ and $v_2$, are incomparable and $v_1 = u \cdot v_2$. Then $v_2 = u \cdot v_3$ for some $v_3$. Furthermore $v_1 = u^2 \cdot v_3$. Arguing in this way we obtain that $v_1 = u^d \cdot v_{d+1}$ since $|u| < x/d$, $|v_2| > (d-1)x/d$. A contradiction.

**Lemma 2.2.** If a word $V$ of length $k \cdot t$ includes no more than $k$ different subwords of length $k$ then $V$ includes a period of length $t$.

**Proof.** We use induction in $k$. The base $k = 1$ is obvious. If there are no more than $(k-1)$ different subwords of length $(k-1)$ then we apply the induction assumption. If there exist $k$ different subwords of length $(k-1)$, then every subword of length $k$ is uniquely determined by its first $(k-1)$ letters. Thus $V = v^t$ where $v$ is a $k$-tail of $V$.

**Definition 2.2.** a) A word $W$ is $n$-divisible in the ordinary sense if there exist $u_1, u_2, \ldots, u_n$ such that $W = v \cdot u_1 \cdots u_n$ and $u_1 \succ \cdots \succ u_n$.

b) In our proof we call a word $W$ $n$-divisible in the tail sense if there exist tails $u_1, \ldots, u_n$ such that $u_1 \succ u_2 \succ \cdots \succ u_n$ and for any $i = 1, 2, \ldots, n-1$ the beginning of $u_i$ is to the left of the beginning of $u_{i+1}$. If the contrary is not specified, an $n$-divisible word means $n$-divisible in the tail sense.

c) A word $W$ is $n$-cancellable if either it is $n$-divisible in the ordinary sense or there exists a word of the form $u^d \subseteq W$.

Now we describe a sufficient condition for $n$-cancellability and its connection with $n$-divisibility.

**Lemma 2.3.** If a word $W$ includes $n$ identical disjoint subwords $u$ of length $n \cdot d$ then $W$ is $n$-cancellable.

**Proof.** Suppose the contrary. Consider the tails $u_1, u_2, \ldots, u_n$ of the word $u$ which begin from each of the first $n$ letters of $u$. Renumber the tails to provide the
inequalities $u_1 \succ \ldots \succ u_n$. By Lemma 2.1 the tails are incomparable. Consider the subword $u_1$ in the left-most copy of $u$, the subword $u_2$ in the second copy from the left, $\ldots$, $u_n$ in the $n$th copy from the left. We get an $n$-division of $W$. A contradiction.

**Lemma 2.4.** If a word $W$ is 4nd-divisible then it is $n$-cancellable.

**Proof.** Suppose the contrary. Consider the numbers of positions of letters $a_i$, $a_1 < a_2 < \cdots < a_{4nd}$ that begin the tails $u_i$ dividing $W$. Set $a_{4nd+1} = |W|$. If $W$ is not $n$-cancellable then there exists $i, 1 \leq i \leq 4(n-1)d+1$, such that for any $i \leq b < c \leq d < e \leq i+4d$ the $(a_c-a_b)$-tail $u_b$ is incomparable with the $(a_e-a_d)$-tail $u_d$. Compare $a_{i+2d} - a_i$ and $a_{i+4d} - a_{i+2d}$. We may assume that $a_{i+4d} - a_{i+2d} \geq a_{i+2d} - a_i$. Let $a_{j+1} - a_j = \inf_k (a_{k+1} - a_k)$, $0 \leq j < 2d$. We may assume that $j < d$. By assumption the $(a_{2d} - a_j)$-tail $u_j$ and the $(a_{2d} - a_{j+1})$-tail $u_{j+1}$ are incomparable with the $(a_{4d} - a_{2d})$-tail $u_{2d}$. Since $a_{4d} - a_{2d} \geq a_{2d} - a_j > a_{2d} - a_{j+1}$, the $(a_{2d} - a_{j})$-tail $u_j$ and the $(a_{2d} - a_{j+1})$-tail $u_{j+1}$ are mutually incomparable. Since

$$\frac{a_{2d} - a_j}{a_{2d} - a_{j+1}} \leq \frac{d+1}{d},$$

the $(a_{j+1} - a_j)$-tail $u_j$ in degree $d$ is included into the $(a_{2d} - a_j)$-tail $u_j$. A contradiction.

**Corollary 2.1.** If a word $W$ is not $n$-divisible in the ordinary sense then $W$ is not 4nd-divisible (in the tail sense).

Set $p_{n,d} := 4nd - 1$.

Let $W$ be a not $n$-cancellable word. Then $W$ is not $(p_{n,d}+1)$-divisible. Consider $U$, the $|[W]/d|$-tail of $W$. Let $\Omega$ be the set of tails of $W$ which begin in $U$. Then by Lemma 2.1 any two elements of $\Omega$ are comparable. There is a natural bijection between $\Omega$, the letters of $U$ and positive integers from 1 to $|\Omega| = |U|$.

Let us introduce a word $\theta$ which is lexicographically less than any other word.

**Remark 2.1.** In the current proof of Theorem 7 all tails are assumed to belong to $\Omega$.

§ 3. Estimates on the occurrence of periodic fragments

**An application of Dilworth’s theorem.** For tails $u$ and $v$ put $u \prec v$ if $u \prec v$ and $u$ is to the left of $v$. Then by Dilworth’s theorem, $\Omega$ can be divided into $p_{n,d}$ chains such that in each chain $u \prec v$ if $u$ is to the left of $v$. Paint the initial positions of the tails into $p_{n,d}$ colours according to their occurrence in the chains. Fix a positive integer $p$. To each positive integer $i$ from 1 to $|\Omega|$, assign $B^p(i)$, an ordered set of $p_{n,d}$ words \{f(i,j)\} constructed as follows. For each $j = 1, 2, \ldots, p_{n,d}$ put

$$f(i,j) = \{ \max f \leq i : f \text{ is painted into colour } j \}.$$  

If there is no such $f$ then the word from $B^p(i)$ at position $j$ is assumed to be equal to $\theta$, otherwise equal to the $p$-tail that begins from the $f(i,j)$th letter.

Informally speaking, we observe the speed of ‘evolution’ of tails in their chains when the sequence of positions in $W$ is considered as the time axis.
3.1. The sets $B^p(i)$, and the process at positions.

Lemma 3.1 (on the process). Given a sequence $S$ of length $|S|$ consisting of words of length $(k-1)$. Each word consists of $(k-2)$ symbols ‘0’ and a single symbol ‘1’. Let $S$ satisfy the following condition:

if for some $0 < s \leq k-1$ there exist $p_{n,d}$ words such that ‘1’ occupies the $s$th position, then between the first and the $p_{n,d}$th of these words there exists a word such that ‘1’ occupies a position with number strictly less than $s$.

Let $L(k-1) = \sup |S|$.

Then $L(k-1) \leq p_{n,d}^{k-1} - 1$.

Proof. We have $L(1) \leq p_{n,d} - 1$. Let $L(k-1) \leq p_{n,d}^{k-1} - 1$. We will show that $L(k) \leq p_{n,d}^k - 1$. Consider the words such that ‘1’ occupies the first position. Their number does not exceed $p_{n,d} - 1$. Between any two of them as well as before the first one and after the last one, the number of words does not exceed $L(k-1) \leq p_{n,d}^{k-1} - 1$.

Hence

$L(k) \leq p_{n,d} - 1 + (p_{n,d})(p_{n,d}^{k-1} - 1) = (p_{n,d})^k - 1$,

as required.

We need a quantity which estimates the speed of ‘evolution’ of sets $B^p(i)$. Set

$$\psi(p) := \{\max k : B^p(i) = B^p(i + k - 1)\}.$$ 

In particular, by Lemma 2.2 we have $\psi(p_{n,d}) \leq p_{n,d}d$.

For a given $\alpha$ we divide the sequence of the first $|\Omega|$ positions $i$ of $W$ into equivalence classes $\sim_\alpha$ as follows: $i \sim_\alpha j$ if $B^\alpha(i) = B^\alpha(j)$.

Proposition 3.1. For any positive integers $a < b$ we have $\psi(a) \leq \psi(b)$.

Lemma 3.2 (basic). For any positive integers $a$ and $k$ we have

$$\psi(a) \leq p_{n,d}^k \psi(k \cdot a) + k \cdot a.$$ 

Proof. Consider the least representative in each class of $\sim_{k,a}$. We get a sequence of positions $\{i_j\}$. Now consider all $i_j$ and $B^{k,a}(i_j)$ from the same equivalence class of $\sim_a$. Suppose it consists of $B^{k,a}(i_j)$ for $i_j \in [b, c)$. Let $\{i_j\}'$ denote the segment of the sequence $\{i_j\}$ such that $i_j \in [b, c - k \cdot a)$.

Fix a positive integer $r$, $1 \leq r \leq p_{n,d}$. All $k \cdot a$-beginnings of colour $r$ that begin from positions of the word $W$ in $\{i_j\}'$ will be called representatives of type $r$. All representatives of type $r$ are pairwise distinct because they begin from the least positions in equivalence classes of $\sim_{k,a}$. Divide each representative of type $r$ into $k$ segments of length $a$. Enumerate segments inside each representative of type $r$ from left to right by integers from zero to $(k-1)$. If there exist $(p_{n,d}+1)$ representatives of type $r$ with the same first $(t - 1)$ segments but with pairwise different $t$th segments, where $1 \leq t \leq k-1$, then there are two $t$th segments such that their first letters are of the same colour. Then the initial positions of these segments belong to different equivalence classes of $\sim_a$.

Now apply Lemma 3.1 as follows: in all representatives of type $r$ except the rightmost one we consider a segment as a unit segment if it contains the least
position where this representative of type \( r \) differs from the preceding one. All other segments are considered as zero segments.

Now we apply the process Lemma 3.1 for the values of parameters as given in the condition of the lemma. We obtain that the sequence \( \{i_j\}' \) contains no more than \( p_{n,d}^{k-1} \) representatives of type \( r \). Then the sequence \( \{i_j\}' \) contains no more than \( p_{n,d}^k \) terms. Thus \( c - b \leq p_{n,d}^k \psi(k \cdot a) + k \cdot a \).

3.2. Completion of the proof for Theorems 7 and 8. Let

\[
    a_0 = 3^{\lfloor \log_3 p_{n,d} \rfloor}, \quad a_1 = 3^{\lfloor \log_3 p_{n,d} \rfloor - 1}, \quad \ldots, \quad a_{\lfloor \log_3 p_{n,d} \rfloor} = 1.
\]

Then \( |W| \leq d|\Omega| + d \) by Lemma 2.1.

Since for the set \( B^1(i) \) no more than \( 1 + p_{n,d}^l \) different values are possible, we have \( |W| \leq d(1 + p_{n,d}^l) \psi(1) + d \). By Lemma 3.2

\[
    \psi(1) < (p_{n,d}^3 + p_{n,d}) \psi(3) < (p_{n,d}^3 + p_{n,d})^2 \psi(9) < \cdots < (p_{n,d}^3 + p_{n,d})^{\lfloor \log_3 p_{n,d} \rfloor} \psi(p_{n,d}) \leq (p_{n,d}^3 + p_{n,d})^{\lfloor \log_3 p_{n,d} \rfloor} p_{n,d}^d.
\]

Take \( p_{n,d} = 4nd - 1 \) to get

\[
    |W| < 4^{5 + 3 \log_3 4} l(nd)^{3 \log_3 (nd) + (5 + 6 \log_3 4)} d^2.
\]

This implies the assertion of Theorem 7.

The proof of Theorem 8 is completed similarly, but instead of the sequence

\[
    a_0 = 3^{\lfloor \log_3 p_{n,d} \rfloor}, \quad a_1 = 3^{\lfloor \log_3 p_{n,d} \rfloor - 1}, \quad \ldots, \quad a_{\lfloor \log_3 p_{n,d} \rfloor} = 1
\]

we have to consider the sequence

\[
    a_0 = 2^{\lfloor \log_2 p_{n,d} \rfloor}, \quad a_1 = 2^{\lfloor \log_2 p_{n,d} \rfloor - 1}, \quad \ldots, \quad a_{\lfloor \log_2 p_{n,d} \rfloor} = 1.
\]

§ 4. An estimate for the essential height

In this section we proceed with the proof of the main Theorem 5. In passing, we prove Theorem 9. We consider positions of letters in the word \( W \) as the time axis. That is, a subword \( u \) occurs before a subword \( v \) if \( u \) is entirely to the left of \( v \) in \( W \).

4.1. Isolation of distinct periodical fragments in the word \( W \). Let \( s \) denote the number of subwords in \( W \) such that each of them includes a period of length less than \( n \) more than \( 2n \) times and each pair of them is separated by subwords of length greater than \( n \), by comparison with the preceding period. Enumerate these from the beginning to the end of the word: \( x_1^{2n}, x_2^{2n}, \ldots, x_s^{2n} \). Thus

\[
    W = y_0 x_1^{2n} y_1 x_2^{2n} \cdots x_s^{2n} y_s.
\]

If there is \( i \) such that the word \( x_i \) has length no less than \( n \), then the word \( x_i^{2n} \) includes \( n \) pairwise comparable tails, hence the word \( x_i^{2n} \) is \( n \)-divisible. Then \( s \) is no less than the essential height of \( W \) over the set of words of length less than \( n \).

Definition 4.1. A word \( u \) will be called noncyclic if \( u \) is not representable in the form \( v^k \) where \( k > 1 \).
Definition 4.2. A word cycle $u$ is the set consisting of the word $u$ and all its cyclic shifts.

Definition 4.3. A word $W$ is strongly $n$-divisible if it is representable in the form $W = W_0 W_1 \cdots W_n$ where the subwords $W_1, \ldots, W_n$ are placed in the lexicographically decreasing order and each of the $W_i, i = 1, 2, \ldots, n$, begins from some word $x_i^k \in \mathbb{Z}$, where all the $z_i$ are distinct.

Lemma 4.1. If there is an integer $m$, $1 \leq m < n$, such that there exist $2n - 1$ pairwise incomparable words of length $m$: $x_{i_1}, \ldots, x_{i_{2n-1}}$, then $W$ is $m$-divisible.

Proof. Put $x := x_{i_1}$. Then $W$ includes disjoint subwords $x^{p_{i_1}} v'_1, \ldots, x^{p_{2n-1}} v'_{2n-1}$, where $p_1, \ldots, p_{2n-1}$ are positive integers greater than $n$, and $v'_1, \ldots, v'_{2n-1}$ are words of length $m$ comparable with $x$, $v'_1 = v_{i_1}$. Hence among the words $v'_1, \ldots, v'_{2n-1}$ either there are $n$ words lexicographically greater than $x$ or there are $n$ words lexicographically smaller than $x$. We may assume that $v'_1, \ldots, v'_n$ are lexicographically greater than $x$. Then $W$ includes subwords $v'_1 x v'_2, \ldots, x^{n-1} v'_n$, which lexicographically decrease from left to right.

Consider an integer $m$, $1 \leq n$. Divide all $x_i$ of length $m$ into equivalence classes relative to strong incomparability and choose a single representative from each class. Let these be $x_{i_1}, \ldots, x_{i'_n}$, where $s'$ is a positive integer. Since the subwords $x_i$ are periods, we consider them as word cycles.

We set $v_k := x_{i_k}$.

Let $v(k, i)$, where $i$ is a positive integer, $1 \leq i \leq m$, be a cyclic shift of a word $v_k$ by $(k - 1)$ positions to the right, that is, $v(k, 1) = v_k$ and the first letter of $v(k, 2)$ is the second letter of $v_k$. Thus $\{v_k(i, i)\}_{i=1}^m$ is a word cycle of $v_k$. Note that for any $1 \leq i_1, i_2 \leq p$, $1 \leq j_1, j_2 \leq m$ the word $v(i_1, j_1)$ is strongly incomparable with $v(i_2, j_2)$.

Remark 4.1. The cases $m = 2, 3, n - 1$ were considered in [31], [27].

4.2. An application of Dilworth’s theorem. Consider a set $\Omega' = \{v(i, j)\}$, where $1 \leq i \leq p$, $1 \leq j \leq m$. Order the words $v(i, j)$ as follows: $v(i_1, j_1) > v(i_2, j_2)$ if $v(i_1, j_1) > v(i_2, j_2)$ and $i_1 > i_2$.

Lemma 4.2. If in the set $\Omega'$ with ordering $>$ there exists an antichain of length $n$ then $W$ is $n$-divisible.

Proof. Suppose there exists an antichain consisting of $n$ words

$$v(i_1, j_1), v(i_2, j_2), \ldots, v(i_n, j_n), \quad i_1 \leq i_2 \leq \cdots \leq i_n.$$ 

If all inequalities between $i_k$ are strict then $W$ is $n$-divisible by definition.

Suppose that for some $r$ there exist $i_{r+1} = \cdots = i_{r+k}$ such that either $r = 0$ or $i_r < i_{r+1}$. Moreover the positive integer $k$ is such that either $k = n - r$ or $i_r < i_{r+k+1}$.

The word $s_{i_{r+1}}$ is periodic, hence it is representable as a product of $n$ copies of $v_{i_{r+1}}$. The word $v_{i_{r+1}}$ includes a word cycle $v_{i_{r+1}}$. Hence in $s_{i_{r+1}}$ there exist disjoint subwords placed in lexicographically decreasing order and equal to $v(i_{r+1}, j_{r+1}), \ldots, v(i_{r+k}, j_{r+k})$ respectively. Similarly we deal with all sets of equal indices in the sequence $\{i_r\}_{r=1}^n$. The result is $n$-divisibility of $W$. A contradiction.
Thus $\Omega'$ can be divided into $(n - 1)$ chains.

Put $q_n = n - 1$.

### 4.3. The sets $C^\alpha(i)$, the process at positions.

Paint the first letters of the words from $\Omega'$ into $q_n$ colours according to their occurrence in chains. Paint also the integers from 1 to $|\Omega'|$ into the corresponding colours. Fix a positive integer $\alpha \leq m$. To each integer $i$ from 1 to $|\Omega'|$ assign an ordered set $C^\alpha(i)$ of $q_n$ words in the following way:

For each $j = 1, 2, \ldots, q_n$ put $f(i, j) = \{\text{max } f' \leq i : \text{there exists } k \text{ such that } v(f, k) \text{ is painted into colour } j \text{ and the } \alpha\text{-tail beginning from } f \text{ consists only of letters initial in some tails from } \Omega'\}.$

If there is no such $f$ then a word from $C^\alpha(i)$ is assumed to be equal to $\theta$, otherwise we assume it to be equal to the $\alpha$-tail of $v(f, k)$.

Set $\varphi(a) = \{\text{max } k : \text{for some } i \text{ we have } C^\alpha(i) = C^\alpha(i + k - 1)\}.$

For a given $a \leq m$ define a division of the sequence of word cycles $\{i\}$ in $W$ into equivalence classes as follows: $i \sim_a j$ if $C^\alpha(i) = C^\alpha(j)$.

Note that the above construction is rather similar to the construction from the proof of Theorem 7. Observe that $B^\alpha(i)$ and $C^\alpha(i)$ are rather similar as are $\psi(a)$ and $\phi(a)$.

**Lemma 4.3.** $\varphi(m) \leq q_n/m$.

**Proof.** In § 4.1 we have enumerated word cycles. Consider the word cycles with numbers $i, i + 1, \ldots, i + [q_n/m]$. We have shown that each word cycle consists of $m$ distinct words. Now consider words in the word cycles $i, i + 1, \ldots, i + [q_n/m]$ as elements of the set $\Omega'$. Then the first letter in each word cycle gets some position. The total number of the positions in question is no less than $n$. Hence at least two of these positions are of the same colour. Now strong incomparability of word cycles implies the assertion of the lemma.

**Proposition 4.1.** For any positive integers $a < b$ we have $\varphi(a) \leq \varphi(b)$.

**Lemma 4.4** (basic). For positive integers $a, k$ such that $ak \leq m$ we have

$$\varphi(a) \leq q^k_n \varphi(k \cdot a).$$

**Proof.** Consider the minimal representative in each class of $\sim_{k \cdot a}$. We get a sequence of positions $\{i_j\}$. Now consider all $i_j$ and $C^{k \cdot a}(i_j)$ from the same equivalence class of $\sim_{a}$. Suppose it consists of $C^{k \cdot a}(i_j)$ for $i_j \in [b, c)$. Let $\{i_j\}'$ denote the segment of the sequence $\{i_j\}$ such that $i_j \in [b, c)$.

Fix a positive integer $r$, $1 \leq r \leq q_n$. All $k \cdot a$-beginnings of colour $r$ that begin from positions of $W$ in $\{i_j\}'$ will be called representatives of type $r$. All representatives of type $r$ are distinct because they begin at the least positions in equivalence classes of $\sim_{k \cdot a}$. Divide each representative of type $r$ into $k$ segments of length $a$. Enumerate the segments of each representative of type $r$ from left to right by integers from zero to $(k - 1)$. If there exist $(q_n + 1)$ representatives of type $r$ with the same first $(t - 1)$ segments but pairwise different $t$th segments, where $1 \leq t \leq k - 1$, then there are two $t$th segments such that their first letters are of the same colour. Then the initial positions of these segments belong to different equivalence classes of $\sim_{a}$. 
Now apply Lemma 3.1 in the following way: in all representatives of type $r$ except the rightmost one we consider a segment as a unit segment if it contains the least position where this representative of type $r$ differs from the preceding one. All other segments are considered as zero segments.

Now we can apply the process Lemma 3.1 for the values of parameters as given in the condition of the lemma. We obtain that the sequence $\{i_j\}'$ contains no more than $q_n^{k-1}$ representatives of type $r$. Then the sequence $\{i_j\}'$ contains no more than $q_n^k$ terms. Thus $c - b \leq q_n^k \varphi(k \cdot a)$.

4.4. Completion of the proof for Theorem 9. Suppose

$$a_0 = 3^{r \log_3 p_n d \gamma}, \quad a_1 = 3^{r \log_3 p_n d \gamma - 1}, \ldots, \quad a_{r \log_3 p_n d \gamma} = 1.$$  

Substitute these $a_i$ into Lemmas 4.4 and 4.3 to obtain

$$\varphi(1) \leq q_n^3 \varphi(3) \leq q_n^9 \varphi(9) \leq \cdots \leq q_n^{3 r \log_3 m \gamma} \varphi(m) \leq q_n^{3 r \log_3 m \gamma + 1}.$$  

Since $C_1^1$ takes no more than $1 + q_n l$ distinct values, we have

$$|\Omega'| < q_n^{3 r \log_3 m \gamma + 1} (1 + q_n l) < n^{3 r \log_3 n \gamma + 2 l}.$$  

By virtue of Lemma 4.1 the number of subwords $x_i$ of length $m$ is less than $2 n^{3 r \log_3 n \gamma + 3 l}$. Thus the total number of subwords $x_i$ is less than $2 n^{3 r \log_3 n \gamma + 4 l}$, so $s < 2 n^{3 r \log_3 n \gamma + 4 l}$ and Theorem 9 is proved.

§ 5. Proof of the main Theorem 5 and of Theorem 6

5.1. Outline of the proof. Now an $n$-divisible word will mean a word $n$-divisible in the ordinary sense. To start with, we find the necessary number of fragments in $W$ with length of the period no less than $2 n$. For this, it suffices to divide $W$ into subwords of large length and to apply Theorem 7 to them. However the estimate can be improved. For this, we find a periodic fragment $u_1$ in $W$ with the period length no less than $4 n$. Removing $u_1$, we obtain a word $W_1$. In $W_1$ we find a fragment $u_2$ with period length no less than $4 n$ and remove it to get a word $W_2$. Now we again remove a periodic fragment and proceed in this way, as is described in the algorithm below in more detail. Then we restore the original word $W$ using the removed fragments. Further we show that a subword $u_i$ in $W$ usually is not a product of a big number of non-neighbouring subwords. In Lemma 5.1 we prove that an application of the algorithm enables us to find the necessary number of removed subwords of $W$ with period length no less than $2 n$.

5.2. Summing essential heights and nilpotency degrees. Let $Ht(w)$ denote the height of a word $w$ over the set of words of degree not exceeding $n$. Consider a word $W$ of height $Ht(W) > \Phi(n, l)$. Apply the following algorithm to it.

Algorithm. Step 1. By Theorem 7 the word $W$ includes a subword with period length $4 n$. Suppose $W_0 = W = u'_1 x_1^{4 n} y_1$. The word $x_1$ is not cyclic. Represent $y_1$ in the form $y'_1 = x_1^{r_2} y_1$ where $r_2$ is maximal possible. Represent $u'_1$ as $u'_1 = u_1 x_1^{r_1}$, where $r_1$ is maximal possible. Denote by $f_1$ the word

$$W_0 = u_1 x_1^{4 n + r_1 + r_2} y_1 = u_1 f_1 y_1.$$
In the sequel, the positions contained in $f_1$ are called *tedious*, the last position of $u_1$ is called *tedious of type* 1 the second position from the end in $u_1$ is called *tedious of type* 2, \ldots, the $n$th position from the end in $u_1$ is called *tedious of type* $n$. Put $W_1 = u_1 y_1$.

**Step $k$.** Consider the words $u_{k-1}, y_{k-1}$, $W_{k-1} = u_{k-1} y_{k-1}$ constructed at the preceding step. If $|W_{k-1}| > \Phi(n,l)$, then we apply Theorem 7 to $W$ with the restriction that the process in the main Lemma 3.2 is applied only to nontedious positions and to tedious positions of type greater than $ka$ where $k$ and $a$ are the parameters from Lemma 3.2.

Thus $W_{k-1}$ includes a noncyclic subword with period length $4n$ such that

$$W_{k-1} = u_k x_k y_k.'$$

Then put

$$r_1 := \sup \{ r : u_k' = u_k x_k r \}, \quad r_2 := \sup \{ r : y_k' = x_k r y_k \}.$$

(Note that the words involved may be empty.)

Define $f_k$ by the equation

$$W_{k-1} = u_k x_k^{4n} + r_1 + r_2 \ y_k = u_k f_k y_k.$$

In the sequel, the positions contained in $f_k$ are called *tedious*, the last position of $u_k$ is called *tedious of type* 1 the second position from the end in $u_k$ is called *tedious of type* 2, \ldots, the $n$th position from the end in $u_k$ is called *tedious of type* $n$. If a position occurs to be tedious of two types then the lesser type is chosen for it. Put $W_k = u_k y_k$.

Perform $4t + 1$ steps of the algorithm and consider the original word $W$. For each integer $i$ from the segment $[1, 4t]$ we have

$$W = w_0 f_i^{(1)} w_1 f_i^{(2)} \cdots f_i^{(n_i)} w_{n_i}$$

for some subwords $w_j$. Here $f_i = f_i^{(1)} \cdots f_i^{(n_i)}$. Moreover we assume that for $1 \leq j \leq n_i - 1$ the subword $w_j$ is not empty. Let $s(k)$ be the number of indices $i \in [1, 4t]$ such that $n_i = k$.

To prove Theorem 7 we have to find as many long periodic fragments as possible. For this, we can use the following lemma.

**Lemma 5.1.** $s = s(1) + s(2) \geq 2t$.

**Proof.** A subword $U$ of the word $W$ will be called *monolithic* if

1) $U$ is a product of words of the form $f_i^{(j)}$;

2) $U$ is not a proper subword of a word which satisfies the above condition 1).

Suppose that after the $(i-1)$th step of the algorithm the word $W$ contains $k_{i-1}$ monolithic subwords. Note that $k_i \leq k_{i-1} - n_i + 2$.

If $n_i \geq 3$, then $k_i \leq k_{i-1} - 1$. If $n_i \leq 2$ then $k_i \leq k_{i-1} + 1$. Furthermore, $k_1 = 1$, $k_t \geq 1 = k_1$. The lemma is proved.
Corollary 5.1.
\[
\sum_{k=1}^{\infty} k \cdot s(k) \leq 10t \leq 5s.
\]

Proof. From the proof of Lemma 5.1 we obtain
\[
\sum_{n_i \geq 3} (n_i - 2) \leq 2t.
\]

By definition \(\sum_{k=1}^{\infty} s(k) = 4t\), that is, \(\sum_{k=1}^{\infty} 2s(k) = 8t\). Summing these two inequalities and applying Lemma 5.1 we obtain the required inequality.

Proposition 5.1. The height of \(W\) does not exceed
\[
\Psi(n, 4n, l) + \sum_{k=1}^{\infty} k \cdot s(k) \leq \Psi(n, 4n, l) + 5s.
\]

In the sequel we consider only \(f_i\) with \(n_i \leq 2\).

If \(n_i = 1\) then put \(f_i' := f_i^{(j)}\), where \(f_i^{(j)}\) is the word of maximal length between \(f_i^{(1)}\) and \(f_i^{(2)}\).

Order the words \(f_i'\) according to their distance from the beginning of \(W\). We get a sequence \(f_{m_1}', \ldots, f_{m_s}'\) where \(s' = s(1) + s(2)\). Put \(f_i'' := f_{m_i}'\). Suppose \(f_i'' = w_i'x_i''w_i''\) where at least one of the words \(w_i'\) and \(w_i''\) is empty.

Remark 5.1. We may assume that at starting steps of the algorithm we have chosen all \(f_i\) such that \(n_i = 1\).

Now consider \(z_j'\), the subwords in \(W\) of the following form:
\[
z_j' = x_p^{(2j-1)\nu} v_j, \quad j \geq 0, \quad |v_j| = |x_{(2j-1)\nu}|;
\]

here \(v_j\) is not equal to \(x_{(2j-1)\nu}\), and the beginning of \(z_j'\) coincides with the beginning of a periodic subword in \(f_{2j-1}'\). We will show that the \(z_j'\) are disjoint.

Indeed, if \(f_j''_{2j-1} = f_{m_{2j-1}}\), then put \(z_j = f_{m_{2j-1}} v_j\).

If \(f_j''_{2j-1} = f_{m_{2j-1}}^{(k)}\), \(k = 1, 2\), and \(z_j'\) intersects \(z_{j+1}'\) then \(f_j''_{2j} \subset z_j'\). Since \(x_{(2j)\nu}\) and \(x_{(2j-1)\nu}\) are noncyclic, we have \(|x_{(2j)\nu}| = |x_{(2j-1)\nu}|\). But then the period length in \(z_j'\) is not less than \(4n\), which contradicts Remark 5.1.

Thus we have proved the following lemma.

Lemma 5.2. In a word \(W\) with height not greater than \((\Psi(n, 4n, l) + 5s')\) there exist at least \(s'\) disjoint periodic subwords such that the period occurs in each of them at least \(2n\) times. Furthermore between any two elements of this set of periodic subwords there is a subword with the same period length as the leftmost of these two elements.

5.3. Completion of the proof for the main Theorem 5 and for Theorem 6. Replace \(s'\) in Lemma 5.2 by \(s\) from the proof of Theorem 9 to obtain that the height of \(W\) does not exceed
\[
\Psi(n, 4n, l) + 5s < E_1 l \cdot n^{E_2 + 12 \log_3 n},
\]

where \(E_1 = 4^{21 \log_3 4 + 17}, E_2 = 30 \log_3 4 + 10\).
Thus we have obtained the assertion of the main Theorem 5.

The proof of Theorem 6 is completed similarly but we have to replace in §4.4 the sequence

\[ a_0 = 3^{\log_3 p_{n,d}}, \quad a_1 = 3^{\log_3 p_{n,d} - 1}, \ldots, \quad a_{\log_3 p_{n,d}} = 1 \]

by the sequence

\[ a_0 = 2^{\log_2 p_{n,d}}, \quad a_1 = 2^{\log_2 p_{n,d} - 1}, \ldots, \quad a_{\log_2 p_{n,d}} = 1, \]

and to take the values of \( \Psi(n, 4n, l) \) from Theorem 8.

§ 6. Comments

The technique presented to the reader appears to enable one to improve the estimate obtained in this paper. However this estimate will remain subexponential. A polynomial estimate if it exists, requires new ideas and methods.

At the beginning of the solution presented, subwords of a large word in the application of Shirshov’s theorem are used mainly as a set of independent elements, not as a set of closely related words. Further we use a colouring of letters inside subwords. Account of colouring of first letters only leads to an exponential estimate. Account of colouring of all letters in the subwords results in an exponent as well. This fact is due to the construction of a hierarchical system of subwords. A detailed investigation of the presented connection between subwords together with the solution presented above may improve the presented estimate up to a polynomial one.

It is also of interest to obtain estimates for the height of an algebra over the set of words whose degrees do not exceed the complexity of the algebra (PI-degree in the English literature). The paper [4] presents exponential estimates, and for words that are not a linear combination of lexicographically smaller words, overexponential estimates were obtained in [40].

The deep ideas of original works by Shirshov [1], [2], which stem from the elimination technique in Lie algebras, may be highly useful, among other issues, for improvement of estimates, despite the fact that the estimates for height in these papers are only primitive recursive.

Bibliography

[1] A. I. Shirshov, “On some nonassociative nilring and algebraic algebras”, Mat. Sb. 41(83):3 (1957), 381–394. (Russian)
[2] A. I. Shirshov, “On rings with identical relations”, Mat. Sb. 43(85):2 (1957), 277–283. (Russian)
[3] E. I. Zel’manov, “On the nilpotency of nil algebras”, Algebra – some current trends (Varna, 1986), Lecture Notes in Math., vol. 1352, Springer-Verlag, Berlin 1988, pp. 227–240.
[4] A. Ya. Belov, V. V. Borisenko and V. N. Latyshev, “Monomial algebras”, J. Math. Sci. (N.Y.) 87:3 (1997), 3463–3575.
[5] A. R. Kemer, “Comments on Shirshov’s Height Theorem”, Selected Works of A. I. Shirshov, Birkhäuser, Basel 2009, pp. 223–230.
[6] A. Belov-Kanel (Kanel-Belov) and L. H. Rowen, “Perspectives on Shirshov’s Height Theorem”, Selected Works of A. I. Shirshov, Birkhäuser, Basel 2009, pp. 185–203.

[7] V. A. Ufnarovskij, “Combinatorial and asymptotic methods in algebra”, Algebra 6, Sovrem. Probl. Mat. Fund. Naprav., vol. 57, VINITI, Moscow 1990, pp. 5–177; English transl. Algebra VI, Encyclopaedia Math. Sci., vol. 57, Springer-Verlag, Berlin 1995, pp. 1–196.

[8] V. Drensky and E. Formanek, Polynomial identity rings, Adv. Courses Math. CRM Barcelona, Birkhäuser, Basel 2004.

[9] S. V. Pchelintsev, “A theorem on height for alternative algebras”, Mat. Sb. 124(166):4(8) (1984), 557–567; English transl. in Math. USSR-Sb. 52:2 (1985), 541–551.

[10] A. Ya. Belov, “On a Shirshov basis of relatively free algebras of complexity n”, Mat. Sb. 135(177):3 (1988), 373–384; English transl. in Math. USSR-Sb. 63:2 (1989), 363–374.

[11] S. P. Mishchenko, “A variant of the height theorem for Lie algebras”, Mat. Zametki 47:4 (1990), 83–89; English transl. in Math. Notes 47:4 (1990), 368–372.

[12] A. Ya. Belov, “On a Shirshov basis of relatively free algebras of complexity n”, Mat. Sb. 135(177):3 (1988), 373–384; English transl. in Math. USSR-Sb. 63:2 (1989), 363–374.

[13] G. P. Chekanu (Ciocanu), “Independence and quasiregularity in algebras”, Izv. Akad. Nauk Respub. Moldova Mat. 1 (1997), 70–77.

[14] Gh. P. Ciocanu, “Local finiteness of algebras”, Mat. Issled. 105 (1988), 153–171. (Russian)

[15] Gh. P. Ciocanu and E. P. Kozukhar, “Independence and nilpotence in algebras”, Izv. Akad. Nauk Moldovy Mat. 2 (1993), 51–62. (Russian)

[16] V. A. Ufnarovskii, “An independence theorem and its consequences”, Mat. Sb. 128(170):1(9) (1985), 124–132; English transl. in Math. USSR-Sb. 56:1 (1987), 121–129.

[17] V. A. Ufnarovskii, “Combinatorial generators of the multilinear polynomial identities”, Fundam. Prikl. Mat. 12:2 (2006), 101–110; English transl. in J. Math. Sci. 149:2 (2008), 1107–1112.

[18] A. Ya. Belov, “On the rationality of Hilbert series of relatively free algebras”, Uspekhi Mat. Nauk 52:2 (1997), 153–154; English transl. in Russian Math. Surveys 52:2 (1997), 394–395.

[19] J. Berstel and D. Perrin, “The origins of combinatorics on words”, European J. Combin. 28:3 (2007), 996–1022.

[20] M. Lothaire, Combinatorics of words (Waterloo, ON, Canada 1982), Encyclopedia Math. Appl., vol. 17, Addison-Wesley, Reading, MA 1983.

[21] M. Lothaire, Algebraic combinatorics on words, Encyclopedia Math. Appl., vol. 90, Cambridge Univ. Press, Cambridge 2002.

[22] V. N. Latyshev, “Combinatorial generators of the multilinear polynomial identities”, Fundam. Prikl. Mat. 12:2 (2006), 101–110; English transl. in J. Math. Sci. 149:2 (2008), 1107–1112.

[23] A. G. Kolotov, “On upper estimate for the height in finitely generated algebras with identities”, Siberian Mat. Zh. 23:1 (1982), 187–189. (Russian)

[24] A. Ya. Belov, “Some estimations for nilpotence of nil-algebras over a field of an arbitrary characteristic and height theorem”, Comm. Algebra 20:10 (1992), 2919–2922.
Subexponential estimates in Shirshov’s theorem on height

[26] V. Drensky, *Free algebras and PI-algebras*. Graduate course in algebra, Springer-Verlag, Singapore 2000.

[27] M. I. Kharitonov, “Estimates for the structure of piecewise periodicity in Shirshov’s height theorem”, *Vestnik Moskov. Univ. Ser. 1 Mat. Mekh.* (to appear).

[28] A. A. Klein, “Indices of nilpotency in a PI-ring”, *Arch. Math. (Basel)* 44:4 (1985), 323–329.

[29] A. A. Klein, “Bounds for indices of nilpotency and nility”, *Arch. Math. (Basel)* 74:1 (2000), 6–10.

[30] E. S. Chibrikov, “Shirshov height of a finitely generated associative algebra satisfying an identity of degree 4”, *Izv. Altai Univ. 1* (2001), 52–56. (Russian)

[31] M. I. Kharitonov, “Two-sided estimates for essential height in Shirshov’s height theorem”, *Vestnik Moskov. Univ. Ser. 1. Mat. Mekh.*, 2012, no. 2, 24–28. (Russian)

[32] M. Kharitonov, *Estimations of the particular periodicity in case of the extremal periods in Shirshov’s height theorem*, arXiv: abs/1108.6295.

[33] A. A. Lopatin, *On the nilpotency degree of the algebra with identity $x^n = 0$*, arXiv: 1106.0950.

[34] *Dniestern notebook: a collection of operative information*, 4th ed., Institute of Mathematics, Siberian Branch of RAS, Novosibirsk 1993. (Russian)

[35] I. I. Bogdanov, “Nagata-Higman theorem for semirings”, *Fundam. Prikl. Mat.* 7:3 (2001), 651–658. (Russian)

[36] C. Procesi, *Rings with polynomial identities*, Marcel Dekker, New York 1973.

[37] A. Ya. Belov, “The Gel’fand-Kirillov dimension of relatively free associative algebras”, *Mat. Sb.* 195:12 (2004), 3–26; English transl. in *Sb. Math.* 195:12 (2004), 1703–1726.

[38] E. N. Kuz’min, “On the Nagata-Higman theorem”, *A collection of papers to the 60th birthday of acad. Iliev*, Sofia 1975, pp. 101–107. (Russian)

[39] Yu. P. Razmyslov, *Identities of algebras and their representations*, Nauka, Moscow 1989; English transl., Transl. Math. Monogr., vol. 138, Amer. Math. Soc., Providence, RI 1992.

[40] A. Ya. Belov, “Burnside-type problems, theorems on height, and independence”, *Fundam. Prikl. Mat.* 13:5 (2007), 19–79; English transl. in *J. Math. Sci.* 156:2 (2009), 219–260.

A. Ya. Belov
Moscow Institute for Open Education
E-mail: kanel@mccme.ru

M. I. Kharitonov
Moscow State University
E-mail: mikhailo.kharitonov@gmail.com

Received 12/DEC/11 and 17/OCT/11
Translated by A. BELOV and M. KHARITONOV