Combinatorics on Words with Applications

Mark V. Sapir

December 11, 1993

Contents

1 Introduction

1.1 Main Definitions

- A *word* is an arbitrary string of symbols. If we fix an alphabet X then the set of all words in this alphabet including the empty one is a monoid under the concatenation. This means that the concatenation is associative and that the empty word is the identity element. This monoid is denoted by X^* and is called the *free monoid over* X. The semigroup obtained from X^* by removing the empty word, is denoted by X^+ and is called the free semigroup over X.

- If w is a word and $w = puq$ for some words $p,\ u,\ q$ then $p,\ u,\ q$ are called *subwords* of w and the expression $p \cdot u \cdot q$ is called an *occurrence* of u

- Being free means in particular, that every map $X \longrightarrow X^*$ (resp. $X \longrightarrow X^+$) is uniquely extendable to an endomorphism of X^* (resp. X^+). Any map $X \longrightarrow X^+$ is called a *substitution*.

- If an arbitrary universal algebra satisfies this property, it is called a *relatively free* algebra. We shall deal with relatively free algebras later.

1.2 About the Course

To investigate words means to investigate regularities in them. Intuitively it is clear that the word *abababababa* has more regular structure than the word *abbbaabbaaaa*. The goal of our course is to present different types of regularities which one can find in words, and to explore connections between regularities in words and properties of different kind of algebras: semigroups, groups, and associative algebras. We will be mainly concerned with so called Burnside type questions "What makes a semigroup (group, associative algebra) finite.

The course is elementary which does not make it easy. No mathematics background is expected. This course can be taught at high school. This does not mean that every high school student would understand it.

1.3 Where Do We Get Words?

There are (at least) three different sources of words in mathematics: a) Products of elements in algebraic systems, b) Combinatorics of partitions, and c) Topology of manifolds.

Take any semigroup $S =< X >$. Every element a of S is a product of elements of X: $a = x_1 x_2 \dots x_n$. Therefore every element of S is represented by a word. An important question: what are those words? What are the words of minimal length representing a? Given two words, do they represent the same element of the semigroup (the *word problem*)? There are many more questions like these.

Take any partition of first m natural numbers: $1, \dots, m = P_1 \cup P_2 \cup \dots P_n$. Take an $n-$element alphabet $\{p_1, p_2, \dots, p_n\}$. Label each number from 1 to m which belong to P_i by letter p_i. Now read these labels as you scan the numbers from 1 to m, and you get a word. For example, the word $p_1p_2p_1p_2p_1$ corresponds to the partition $\{1,2,3,4,5\} = \{1,3,5\} \cup \{2,4\}$.

Take any differential manifold M and divide it into n pieces M_i. Again, associate a letter m_i with every piece. Suppose that you are traveling on this manifold. Every hour write down a letter which corresponds to the region which you are visiting. After n hours you will get a word of length n. The second example is a particular case of the last one: Take the manifold of real numbers from 1 to m divide it according to the partition P, and travel from 1 to m with a constant speed (1 unit interval in 1 hour). For example you can drive along the US Interstate 80 from Chicago (IL) to Lincoln (NE) with the speed 65 miles per hour and every hour write down the two letter abbreviation of the name of the state you are in. Then you will probably get the following word:

$$ILILIAIAIAIAIANENE$$

(the author has performed this experiment himself).

An amazing thing is that all three ways of getting words are closely connected. This will become clear later.

2 Avoidable Words

2.1 An Old Example

In 1851 M.E.Prouhet studied the following problem.

> Are there arbitrary big numbers m such that the interval $[1, 2^m]$ of natural numbers can be divided into 2 disjoint parts P_i such that the sum of all elements in every P_i is the same, the sum of all squares of elements in every P_i is the same,..., the sum of all $(m-1)-$th powers of elements in every P_i is the same.

This problem belonged to a popular at that time direction of number theory. In particular, Gauss and Euler studied some variations of this problem.

Prouhet came up with a solution. His solution may be interpreted in the following way. Let us consider an example. Take $m = 3$. Consider the word $p_3 = abbabaab$, and produce the following table:

1	2	3	4	5	6	7	8
a	b	b	a	b	a	a	b

Now let P_a be the set of numbers from 1 to 8 which are above a in this table, P_b be the set of numbers which are above b: $P_a = \{1, 4, 6, 7\}$, $P_b = \{2, 3, 5, 8\}$. Let us check the Prouhet conditon:

$$1 + 4 + 6 + 7 = 18 = 2 + 3 + 5 + 8,$$

$$1^2 + 4^2 + 6^2 + 7^2 = 102 = 2^2 + 3^2 + 5^2 + 8^2.$$

If we want to construct the Prouhet decomposition for $m = 4$ we have to take this word $p_3 = abbabaab$, change a by b and b by a (we'll get $baababba$), and concatenate these two words:

$$p_4 = abbabaabbaababba.$$

By induction one can easily define the Prouhet word for every m.

Ex. 1 *Prove that the partition corresponding to the Prouhet word p_m satisfies the Prouhet condition for every m.*

It is amazing that the word p_m was rediscovered several times after Prouhet. A.Thue rediscovered this word in 1906 and he was the first who proved the following result:

Theorem 2.1 *The word p_m does not contain subwords of the form www where w is any nonempty word. Thus words p_m are* cube free.

Arshon, Morse and Hedlund proved the same result in late 30s. Now it is contained in some collections of problems for high school students.

2.2 Proof of Thue's Theorem

Let us consider the following substitution:

$$\phi(a) = ab,\ \phi(b) = ba.$$

Words ab and ba will be called *blocks*.
Let $t_1 = a,\ \ldots,\ t_n = \phi(t_{n-1})$.

Ex. 2 $t_n = p_{n-1}$ *for every* $n \geq 2$.

Lemma 2.1 *If w is cube-free then $\phi(w)$ is also cube-free.*

Proof. Suppose that $\phi(w)$ contains a cube ppp.

Case 1. The length of p is even.

Case 1.1. The first p starts at the beginning of a block. Since $\phi(w)$ is a product of blocks of length 2, and $|p|$ is even, p ends at the end of a block. Then the second and the third p also start at the beginning of a block and end at the end of a block. Thus p is a product of blocks, so $p = \phi(q)$ for some word q. Now let us substitute every block in $\phi(w)$ by the corresponding letter. Then $\phi(w)$ will turn into w and $ppp = \phi(q)\phi(q)\phi(q)$ will turn into qqq. Therefore w contains a cube qqq.

Case 1.2. The first p starts at the end of a block. Then it ends at the beginning of a block, and the same is true for the second and the third p. Without loss of generality assume that p starts with a. This a is the end of the block ba. Since the second p also starts with a and this a is the end of a block, we can conclude that p ends with b. Therefore $p = ap_1b$. Then we have:

$$\phi(w) = \ldots b\ ap_1b\ ap_1b\ ap_1b\ \ldots.$$

Consider the word bap_1. This word has even length, starts with the beginning of a block and repeats 3 times in $\phi(w)$, which is impossible by the previous case.

Remark. The procedure which we did in this case is called "shift". It is used very often in the combinatorics on words, especially when we deal with words divided into "blocks".

Case 2. The word p has odd length. If the first p starts with the first letter of a block then the second p starts with the second letter of a block. If the first p starts with the second letter of a block then the second p starts with the first letter of a block and the third p starts with the second letter of a block.

In any case there are two consecutive occurrences of p such that the first one starts with the first letter of a block and the second one starts with the second letter of a block. Let us denote these copies of p by p_1 and p_2.

It is clear that $|p| \geq 2$.

Suppose that p_1 starts with ab. Then p_2 also starts with ab. This b is a beginning of a block. Therefore the third letter in p_2 is a. Therefore the

third letter in p_1 is also a. This a is the first letter of a block. The second letter of this block is b. Therefore the fourth letter of p_1 is b. Then the fourth letter of p_2 is b, which is the first letter of a block, the fifth letter of p_2 is a, same as the fifth letter in p_1, and so on. Every odd letter in p is a, every even letter in p is b. Since p_1 has odd number of letters, the last letter in p_1 is a. This a is the first letter of a block. The second letter of this block is the first letter of p_2, which is a - a contradiction (we found a block aa).

The proof is complete.

Remark. In the last case of this proof we used the fact that blocks have different first (last) letters.

Ex. 3 *Prove that t_n does not contain subwords of the form $qpqpq$ for any words p and q.*

2.3 Square-Free Words

What is a "regularity"? We can say that a word w is more "regular" than a word w' if w is in some sense simpler. There exists a beautiful definition of the property "being simple", due to Kolmogorov. Let M be an universal Turing machine, for example a PC with an infinite memory. Then the Kolmogorov complexity of a word w is the length of the shortest program (without an input) which prints out the word. This definition does not depend on the hardware of M very much. Roughly speaking, if we change M, the complexities of words will be multiplied by a constant. It is clear that the simplest possible words are periodic words, that is words of the form $uuu\ldots$. The words constructed in the previous section are more complex, they do not contain cubes. Now we are going to consider words which do not contain squares, the *square-free* words.

A word is called *square-free* if it does not contain a subword of the form uu.

Ex. 4 *Every square-free word over an alphabet with 2 letters has length at most 3.*

Theorem 2.2 (Thue, 1906) *There exist arbitrary long square-free words over a 3-letter alphabet.*

Proof. Consider the following substitution:

$$\phi(a) = abcab,\ \phi(b) = acabcb,\ h(c) = acbcacb.$$

Lemma 2.2 *For every square-free word w $\phi(w)$ is square-free.*

This lemma will be a corollary of the following powerful theorem of Thue.

Theorem 2.3 *Let M and N be alphabets and let ϕ be a substitution from M to N^+. If*

(0) $\phi(w)$ is square-free whenever w is a square-free word on M of length no greater than 3,

(1) $a = b$ whenever $a, b \in M$ and $\phi(a)$ is a subword of $\phi(b)$,

Then $\phi(w)$ is square-free whenever w is a square-free word on M.

Proof of theorem 2.3. Let ϕ satisfy (0) and (1). First of all let us prove the following "rigidity" statement:

> If $a, e_1, \ldots, e_n$ are letters from M, $e_1 \ldots e_n$ is a square-free word, and $\phi(e_1 \ldots e_n) \equiv X\phi(a)Y$ then $a = e_j$, $X = \phi(e_1 \ldots e_{j-1})$, $Y = \phi(e_{j+1} \ldots e_n)$.

Suppose this is not true. Since $\phi(e_i)$ cannot be a subword of $\phi(a)$, $\phi(a)$ intersects with at most 2 factors $\phi(e_i)$. Since $\phi(a)$ cannot be a subword of $\phi(e_i)$, $\phi(a)$ intersects with exactly 2 factors, say $\phi(e_j e_{j+1})$. Then $\phi(e_j) \equiv pq$, $\phi(e_{j+1}) \equiv rs$, and $\phi(a) \equiv qr$. Now $\phi(ae_ja) = qrpqqr$ is not square-free. By condition (0) the word ae_ja is not square-free, thus $a = e_j$. On the other hand: $\phi(ae_{j+1}a) = qrrsqr$ also is not square-free. Thus $a = e_{j+1}$. Therefore $e_j = e_{j+1}$ which contradicts the fact that $e_1 \ldots e_n$ is a square-free word.

Ex. 5 *Prove the statement without the condition that $e_1 \ldots e_n$ is a square-free word.*

Now suppose that w is a square-free word on M and $\phi(w) = xyyz$ for some nonempty word y. Let $w = e_0 \ldots e_n$. Let us denote $\phi(e_i)$ by E_i. We have

$$E_0E_1 \ldots E_n \equiv xyyz.$$

If E_0 is contained in x or E_n is contained in z then we can shorten w (delete e_0 or e_n). Therefore we can suppose that

$$E_0 = xE_0', \ E_n = E_n'z, \ yy = E_0'E_1 \dots E_{n-1}E_n'.$$

By condition (0) we have that $n \geq 3$.

The word y is equal to $E_0'E_1 \dots E_j' = E_j''E_{j+1} \dots E_n'$.

If $j = 0$ then E_1E_2 must be a subword of E_0 which is impossible. Similarly, $j \neq n$.

Now by the rigidity statement,

$$E_0' \equiv E_j'', \ E_1 \equiv E_{j+1}, \ \dots, E_j' \equiv E_n',$$

and, in particular, $n = 2j$. Therefore

$$\phi(e_0e_je_n) = E_0E_jE_n = xE_0'E_j'E_j''E_n'z \equiv$$
$$xE_0'E_j'E_0'E_j'z.$$

By condition (0) either $e_0 = e_j$ or $e_j = e_n$. Without loss of generality let $e_0 = e_j$. We also know that $E_1 = E_{j+1}, \dots, E_{j-1} = E_{2j-1}$. Condition (0) implies that ϕ is one-to-one. Therefore $e_0 = e_j, e_1 = e_{j+1}, \dots$. Hence w is not square-free: it is equal to $e_0e_1 \dots e_{j-1}e_0 \dots e_{j-1}e_n$.

The proof of theorem 2.3 is complete. It implies lemma 2.2 and theorem 2.2.

A complete algorithmic characterization of square-free substitutions was found first by Berstel in 1979. The best (in the computational sense) characterization is found by Crochemore "Sharp Characterizations of SquareFree Morphisms", Theoretical Computer Science 18 (1982), 221–226.

Theorem 2.4 *(*Crochemore*) Let ϕ be a substitution, M be the maximal size of a block, m be the minimal size of a block, k be the maximum of* 3 *and the number* $1 + [(M-3)/m]$. *Then ϕ is square-free iff for every square-free word of w length $\leq k$ $\phi(w)$ is square-free.*

Using theorem 2.3 and a computer, one can establish a substitution from an infinite alphabet to $\{a, b, c\}^+$ which is square-free. We will present here a substitution from $\{x_1, \dots\}$ to $\{a, b, c, d, e\}^+$. Let $w_1, w_2, \dots$ be an infinite sequence of distinct square-free words on a three-letter alphabet. Consider the following substitution from $x_1, x_2, \dots$ to $\{a, b, c, d, e\}^+$:

$$x_i \to dw_iew_i.$$

This substitution is square-free by theorem 2.3. Indeed it is clear that a word dw_iew_i cannot be a subword of dw_jew_j because w_i and w_j do not contain d and e. Now if $dw_iew_idw_jew_jdw_kew_k$ contains a square uu then the numbers of d's and e's in uu must be even. Neither of them may be 0, because otherwise one of the words w_i, w_j, w_k would contain a square. So each of them is 2. Therefore each of the copies of u has one d and one e. The first d cannot participate, otherwise u must be equal to dw_iew_i, and $w_i \neq w_j$. Therefore u must start at the middle of the first w_i, and end in the middle of the first w_j. Then u must contain the subword ew_id. But this subword occurs in our product only once, so the second copy of u cannot contain it, a contradiction.

A square-free substitution from $\{a, b, c, d, e\}$ to $\{a, b, c\}^+$ has been found by Bean, Ehrenfeucht and McNulty (BEM), Pacific Journal of Mathematics, v. 85, No. 2, 1979.

2.4 kth power-free substitutions

The following theorem, also proved by BEM, gives a sufficient condition for a substitution to be kth power-free for $k > 2$.

Theorem 2.5 *Let M and N be alphabets and let ϕ be a substitution $M \to N^+$ which satisfies the following three conditions*

(0) $\phi(w)$ is kth power-free whenever w is a $k-$power free word of length no greater than $k+1$.

(1) $a = b$ whenever $\phi(a)$ is a subword of $\phi(b)$.

(2) If $a, b, c \in M$ and $x\phi(a)y \equiv \phi(b)\phi(c)$ then either x is empty and $a = b$ or y is empty and $a = c$.

Then ϕ is kth power-free.

Ex. 6 *Prove theorem 2.5.*

2.5 An Application to the Burnside Problem for Semigroups

Definition. A semigroup is a set with an associative binary operation. Examples: X^+ (operation concatenation), $\mathbf{Z}$ (operation $+$), $\mathbf{Z}$ (operation $\cdot$),

$\{a, b, c\}$ where operation is given by the following *multiplication table*

	a	b	c
a	a	b	c
b	c	c	c
c	c	c	c

.

If x is an element of a semigroup S then we will write x^n instead of $x \cdot x \cdots x$ (n times).

Definition. An element x of a semigroup S is called periodic if there exist natural numbers m_x and n_x such that $x^{m_x} = x^{m_x+n_x}$.

It is clear that finite semigroups are "better" than infinite ones. What makes a semigroup finite? This question is the essence of the so called Burnside type problems.

The following easy theorem shows that every finite semigroup possesses some nice properties.

Theorem 2.6 *If a semigroup is finite then it has finitely many generators and every element x is periodic. Moreover those m_x and n_x may be chosen the same for all x, that is any finite semigroup satisfies an identity $x^m = x^{m+n}$.*

Is the converse statement true? More precisely, we can formulate the following problems.

Problem 1 *(a) Is it true that a semigroup is finite provided it is periodic and has finitely many generators?*

(b) Is it true that a semigroup is finite provided it is finitely generated and satisfies the identity $x^m = x^{n+m}$ for some m and n?

The answers are negative. A counterexample was published by Morse and Hedlund, but the construction which they used is attributed to Dilworth.

Theorem 2.7 *There exist*

(a) a 2-generated infinite semigroup with 0 which satisfies the identity $x^3 = 0$; therefore it satisfies the identity $x^3 = x^4$,

(b) a 3-generated infinite semigroup with 0 which satisfies the identity $x^2 = 0$; therefore it satisfies the identity $x^2 = x^3$.

The Dilworth Construction. Let us take any set of words W closed under taking subwords. Let $S(W)$ be the set W with an additional symbol 0. Define an operation on $S(W)$ by the following rule: $u * v = uv$ if $uv \in W$, and $u * v = 0$ otherwise. The set $S(W)$ with this operation is a semigroup. It satisfies the identity $x^2 = 0$ (resp. $x^3 = 0$) provided W consists of square-free (cube-free) words. Indeed, since xx (resp. xxx) is not a word from W, $x * x$ (resp. $x * x * x$) must be equal to 0.

Remark. Burnside originally formulated his problems for groups in 1902. The original problems turned out to be much more complicated than their semigroup analogs. Golod (1964) used a number theoretic construction of Golod and Shafarevich to obtain the first example of a periodic finitely generated infinite group. We will present his example below. Infinite finitely generated groups of bounded exponents were first constructed by Novikov and Adian in 1968. They proved that there exist such groups for every odd exponent greater than 4380. The paper (300+ pages) contained more than 100 statements which were proved by a simultaneous induction. This is one of the most complicated papers in the history of mathematics. Later Adian lowered the bound (of the exponent, not of the number of pages) to 665. Recently Ivanov and Lysionok proved that there exist finitely generated infinite groups of arbitrary (not only odd) sufficiently large exponent.

Using the Dilworth construction, it is easy to translate statements from the language of words into the language of semigroups. For example we can replace x^2 and x^3 by arbitrary words $u(x_1, \ldots, x_n)$, and ask the following question: When $S(W)$ satisfies identically $u(x_1, \ldots, x_n) = 0$? The answer is almost clear:

Theorem 2.8 *$S(W)$ satisfies the identity $u = 0$ if and only if words from W are $u-$free, that is no one of these words contains $\phi(u)$ for any substitution ϕ.*

If W is $u-$free then we shall say that W *avoids* u. For example square-free words avoid x^2.

We have proved that there are infinitely many square-free words in a 3-letter alphabet. If for some number k there are infinitely many words in a $k-$letter alphabet which avoid a word u then we will say that u is *$(k-)$avoidable*. Therefore x^2 is $3-$avoidable, and x^3 is $2-$avoidable.

Translating the concept of avoidability into the language of semigroups we get

Theorem 2.9 *For every word u the following conditions are equivalent:*

1. *u is $k-$avoidable;*
2. *There exists an infinite semigroup S generated by k elements, which identically satisfies $u = 0$.*

Proof. 1.$\rightarrow$ 2. If u is $k-$avoidable then the set W of all words in a $k-$letter alphabet which avoid u is infinite. It is clear that W is closed under taking subwords. Consider $S(W)$. This semigroup is infinite, $k-$generated, and, by Theorem 2.8 satisfies the identity $u = 0$.

2. $\rightarrow$ 1. Let S be an infinite $k-$generated semigroup which satisfies the identity $u = 0$. Let $\{x_1, \ldots, x_k\}$ be the generators of S. Let $y_1, y_2, \ldots$ be all (infinitely many) non-zero elements of S. Since S is generated by x's, each y_i is represented by a product of x's. This product can be considered as a word (see Section 1.3). Take one such word w_i for each y_i.

Each of these words is written in the $k-$letter alphabet of x's. Suppose that one of these w_i does not avoid u. This means that w_i contains $\phi(u)$ for some substitution ϕ. Therefore $w_i \equiv p\phi(u)q$ where $p,\ q$ are words in our $k-$letter alphabet.

For every word v in this alphabet let $\bar{v}$ be the element of S represented by this word. Then we have that $\bar{w}_i = y_i$. So we have

$$y_i = \bar{p}\bar{\phi}(\bar{u})\bar{q} = \bar{p}u(\bar{\phi}(\bar{x}_1), \ldots, \bar{\phi}(\bar{x}_k))\bar{q}.$$

But the middle term in the last product is 0 since S satisfies $u = 0$ identically. Therefore $y_i = 0$, a contradiction (we took y_i to be non-zero).

Our goal now is to describe avoidable words.

2.6 Examples and Simple Facts

We already know that xx, xxx, ..., x^k are avoidable words. It is clear that x^2yx and $xyxy$ are also avoidable words because, in general,

Lemma 2.3 *If u is a ($k-$avoidable) word and ϕ is any substitution then any word containing $\phi(u)$ is also avoidable.*

The following lemma is also useful.

Lemma 2.4 *For every words u and w, if w avoids u then every subword of w avoids u.*

On the other hand, the word xyx is *unavoidable*. Indeed, let k be any natural number. Suppose that there exists an infinite set of words W in a $k-$letter alphabet which avoid xyx. Then W must contain a word w of length $> 2k$. Since we have only k letters in the alphabet, one of these letters, say a, occurs twice in w, that is w contains a subword apa for some nonempty word p. This word is of the form xyx ($\phi : \ x \to a, \ y \to p$).

Let us consider the following words Z_n which will be called *Zimin words*:

$$Z_1 \equiv x_1, Z_2 \equiv x_1 x_2 x_1, \ldots, Z_n \equiv Z_{n-1} Z_n Z_{n-1}. \tag{1}$$

Ex. 7 *(*BEM+Zimin.*) Z_n is an unavoidable word (for every n).*

Amazingly enough, it will turn out that Z_n are "universal" unavoidable words in a $n-$letter alphabet.

Let us list some properties of these words. For every word u and every substitution ϕ we say that $\phi(u)$ is a *value* of u.

Ex. 8 *1.* $|Z_n| = 2^n - 1$,

2. *Every odd letter in Z_n is x_1, every even letter in Z_n is x_i for $i > 1$,*

3. *For every $k < n$, Z_n is a value of Z_k under the following substitution:*

$$x_1 \to Z_{n-k}, \ x_i \to x_{n-k+i-1}$$

for $i = 2, 3 \ldots$,

4. *If we delete the letter x_1 from Z_n then we get a word which differs from Z_{n-1} only by names of the letters.*

5. *Every word in the $n-$letter alphabet $\{x_1, \ldots, x_n\}$, which properly contains Z_n, contains a square, and so is avoidable.*

2.7 The Zimin Theorem

Theorem 2.10 *(*Zimin*.) A word u is avoidable if and only if Z_n does not contain values of u, where n is the number of distinct letters in u.*

Ex. 9 *Deduce the following two facts from Theorem 2.10*

1. *Every word in a $n-$letter alphabet of length 2^n and greater is avoidable.*

2. *Every unavoidable word u contains a letter which occurs in u only once.*

By Lemma 2.3 and Exercise 7, if Z_n contains a value of u then u is unavoidable. The converse implication will follow from Theorem 2.11 below.

2.8 Fusions, Free Subsets and Free Deletions

Definition. Let u be a word in an alphabet X, let B and C be subsets of X. We call the pair (B, C) a *fusion* in u if for every two-letter subword xy in u

$$x \in B \text{ if and only if } y \in C.$$

We will call B and C *components* of the fusion.

For example the sets $\{x\}$ and $\{y\}$ form a fusion in the word $yxyzxytx$ and in the word $xyxy$. Sets $\{x, z\}$ and $\{z, t\}$ form a fusion in the word $xtxzztxz$.

Definition. If B, C is a fusion in u then any subset $A \subseteq B \backslash C$ is called a *free subset* of u.

For example, $\{x\}$ is a free subset in $xtxzztxz$.

Ex. 10 *Find all fusions and all free sets in Z_n.*

Let u be a word, Y be a subset of letters of u. Consider the word obtained from u by deleting all letters which belong to Y. This word will be denoted by u_Y.

Definition. A deletion of a free set A in a word is called a *free deletion* and is denoted by σ_A.

Definition. A sequence of deletions $\sigma_Y, \sigma_Z, \ldots$ in a word u is called a *sequence of free deletions* if Y is a free set in u, Z is a free set in u_Y, etc.

For example, the sequence $\sigma_{\{x_1\}}, \sigma_{\{x_2\}}, \ldots, \sigma_{\{x_n\}}$ is a sequence of free deletions in Z_n.

Definition. Let $u = x_1 x_2 \dots x_n$ be a word, a be a letter. By $[u,a]_0^0$ we denote the word $x_1 a x_2 a \dots a x_n$. By $[u,a]_0^1$ we denote the word $a[u,a]_0^0$, by $[u,a]_0^1$ we denote the word $[u,a]_0^0 a$. Finally, by $[u,a]_1^1$ we denote the word $a[u,a]_0^0 a$.

Definition. Let u be a word, B, C be a fusion in u, $A \subseteq B \backslash C$. For every substitution ϕ of the word $\sigma_A(u)$ we can define the following substitution ϕ^* of u by the following rules:

$$\phi^*(x) = \begin{cases} a & \text{if } x \in A \\ [\phi(x), a]_0^0 & \text{if } x \in C \backslash B \\ [\phi(x), a]_1^0 & \text{if } x \in B \cap C \\ [\phi(x), a]_1^1 & \text{if } x \in B \backslash (C \cup A) \\ [\phi(x), a]_0^1 & \text{otherwise.} \end{cases} \tag{2}$$

We will call ϕ^* the substitution induced by ϕ relative to the triple (A, B, C).

The following lemma contains the main property of free deletions.

Lemma 2.5 *Then $\phi^*(u) = [\phi(u_A), a]_\beta^\alpha$ for some α and β. If u starts and ends with a letter from $B \backslash C$ then $\alpha = \beta = 1$.*

Ex. 11 *Prove lemma 2.5.*

2.9 The BEM+Zimin Theorem

The following Theorem obviously contains Theorem 2.10.

Theorem 2.11 (BEM, Zimin.) *The following conditions are equivalent for every word u:*

1. *u is unavoidable.*
2. *Z_n contains a value of u, where n is the number of distinct letters in u.*
3. *There exists a sequence of free deletions which reduces u to a 1-letter word.*

Example. Z_n has a sequence of free deletions of length $n-1$ which reduces Z_n to x_1.

Our proof of Theorem 2.11 is simpler than the original proofs of BEM and Zimin. It is also more general: the same ideas will be used later in a more complicated situation.

2.10 The Proof of $3 \to 2$

Ex. 12 *Prove this implication.*

Hint. Use induced substitutions (2).

2.11 The Proof of $2 \to 1$

This implication immediately follows from lemmas 2.3 and 2.4.

2.12 Proof of $1 \to 3$

Suppose that there is no sequence of free deletions which reduce u to a 1-letter word. We shall prove that u is avoidable.

In fact we will construct a substitution γ such that for some letter a words $\gamma(a),\ \gamma^2(a),\ \ldots,..$ are all different and avoid u.

The substitution γ is constructed as follows. Let r be a natural number. Let A denote the alphabet of r^2 letters $a_{ij},\ 1 \leq i,j \leq r$. Consider the $r^2 \times r-$matrix M in which every odd column is equal to

$$(1,1,\ldots,1,2,2,\ldots,2,\ldots,r,r,\ldots,r),$$

and each even column is equal to

$$(1,2,\ldots,r,1,2,\ldots,r,\ldots,1,2,\ldots,r).$$

.

Replace every number i in the $j-$th column by the letter a_{ij}. The resulting matrix is denoted by M'. The rows of M' can be considered as words. Denote these words by $w_1,\ldots,w_{r^2}$ counting from the top to the bottom. Now define the substitution γ by the following rule:

$$\gamma(a_{ij}) = w_{(r-1)j+i}.$$

For every $j = 1,2,\ldots$ let A_j be the set $\{a_{ij} \mid i = 1,2,\ldots r\}$.

Ex. 13 *γ satisfies the following properties:*

1. *The length of each block is r.*

2. *No two different blocks have common 2-letter subwords.*

3. *All letters in each block are different.*

4. *The j-th letter in each block belongs to A_j (that is its second index is j).*

Proposition 2.1 *Let $r > 6n + 1$. Then $\gamma^m(a_{11})$ avoids u for every m.*

Proof. By contradiction, suppose that there exists a word u with n letters such that u cannot be reduced to a 1-letter word by free deletions, and there exists m such that $\gamma^m(a_{11})$ contains $\delta(u)$ for some substitution δ. We may suppose that the pair (n, m) is minimal (that is n is minimal possible number for which such u and m exist, and m is minimal for given u.

The idea of the proof is the following. First of all we will construct a substitution ϵ such that $\epsilon(u)$ has many fusions. Then we prove that if a value of a word has many fusions then so does the word itself.

Let us call images of letters under γ *blocks*, and let us call products of blocks *integral words*.

Let w be an integral word. Let v be a subword of w. Then v is equal to a product of three words $p_1p_2p_3$ where p_1 is an end of a block, p_2 is a product of blocks or lies strictly inside a block, p_3 is a beginning of a block. Some of these words may be empty. The following property of integral words is very important. It easily follows from Exercise 13.

Lemma 2.6 *The decomposition $u = p_1p_2p_3$ does not depend on the particular occurrence of v: if we consider another occurrence of v, the the decomposition $v = p_1p_2p_3$ will be the same).*

Now let $w = \gamma^m(a_{11})$. We have that $\delta(u) < w$. For every letter $x \in c(u)$ let $\delta(x) = p_{x1}p_{x2}p_{x3}$ be the decomposition described above.

Now let us take any block B which intersects with $\delta(u)$. This block can appear in many different places of w. Let us consider all words p_{xi} which are contained in all possible occurrences of B. There are no more than $3n$ such words p_{xi}. Each of them may occur in B at most once because all letters in B are different. Therefore we may consider B as an interval which is covered by at most $3n$ other intervals. The length of B is at least $6n + 2$. Therefore there exists a 2-letter subword of t_B which satisfies the following condition:

(T) *For every word p_{xi} the word t_B either is contained in p_{xi} or does not have common letters with it*

This easily follows from the geometry fact that k intervals on a line divide this line into at most $2k+1$ subintervals.

Now let us replace every t_B in every block B in w by a new letter y_B. We shall get a new word w_1 in the alphabet $A \cup \{y_B | B = \gamma(a_{ij})\}$. This word has the following form:

$$P_0 Q_1 y_1 P_1 Q_2 y_2 P_2 \dots Q_k y_k P_k Q_{k+1},$$

where Q_i is a beginning of a block, P_i is an end of a block, P_i and Q_i do not overlap, and

$$\text{if } y_i = y_j \text{ then } Q_i = Q_j \text{ and } P_i = P_j. \tag{3}$$

Every word with these properties will be called a *quasi-integral* word.

Let us denote a substitution ϵ which is obtained from δ by the following procedure: take each $\delta(x)$ and replace each occurrence of t_B there by y_B. The condition (T) of words t_B implies that words t_B do not intersect, so our definition of ϵ is correct. This condition also implies that $\epsilon(u)$ is a subword of the quasi-integral word w_1.

Now we shall show that w_1 (and any other quasi-integral word) has many fusions and free sets.

Lemma 2.7 *In any quasi-integral word the sequence of deletions σ_{A_1}, $\sigma_{A_2} \dots$, σ_{A_r} is a sequence of free deletions.*

Ex. 14 *Prove this.*

The result of this sequence of deletions is, of course, the deletion of all a-letters. Now we have to understand how free deletions in $\epsilon(u)$ relate to the deletions in u.

First of all we have the following two simple observations.

Let θ be any substitution and let v be any word. Let D be a set of letters of $\theta(v)$. Let D' be a subset of $c(u)$ defined as follows:

$$D' = \{x | c(\theta(x)) \subseteq D\}.$$

Now let us define a substitution θ_D of $v_{D'}$:

$$\theta_D(x) = \theta(x)_D.$$

The following Lemma is obvious:

Lemma 2.8 $\theta_D(v_{D'}) = \theta(v)_D$.

Lemma 2.9 *If D is a free set in $\theta(v)$ then D' is a free set in v.*

Ex. 15 *Prove this.*

This lemma, while simple, is yet another key property of free sets.

Let us apply Lemma 2.7, Lemma 2.8 and Lemma 2.9 to our situation: $\epsilon(u) < w_1$. By these Lemmas, there exists a sequence of free deletions $\sigma_1 \ldots$, σ_r in u. Let $u_{A'}$ be the result of these deletions. Then by Lemma 2.8

$$\epsilon_A(u_{A'}) = (w_1)_A = y_1 y_2 \ldots y_k$$

.

Now, by definition, each y_B determines the block B. Let us consider the substitution α which takes each y_B to B. Then $\alpha((w_1)_A) = w$. Therefore $\alpha\epsilon_A(u_{A'})$ is a subword of $w = \gamma^m(a_{11})$. But an image of each letter from $u_{A'}$ under $\alpha\epsilon$ is a product of blocks. Therefore we can apply γ^{-1} to $\alpha\epsilon_A(u_{A'})$. The result, $\gamma^{-1}\alpha\epsilon_A(u_{A'})$, is a subword of $\gamma^{m-1}(a_{11}$.

Now we can complete the proof. The word $u_{A'}$ contains at most the same number of letters as u, and $m-1$ is strictly less than m. By the assumption (we assumed that the pair (n, m) is minimal) there exists a sequence of free deletions which reduces $u_{A'}$ to a 1-letter word. If we combine this sequence of free deletions with the sequence which we used to get $u_{A'}$ from u, we will get a sequence of free deletions which reduces u to a 1-letter word, which is impossible.

The Theorem is proved.

2.13 Simultaneous Avoidability

Definition. A set of words W is said to be (k-)avoidable if there exists an infinite set of words in a finite alphabet (a $k-$letter alphabet) each of which avoids each of words from W.

The following result can be proved similar to Theorem 2.9 above.

Theorem 2.12 *A set of words W is $k-$avoidable if and only if there exists an infinite k generated semigroup satisfying identities $w = 0$ for all $w \in W$.*

Ex. 16 *Prove that a finite system of words W is avoidable if and only if each word in W is avoidable.*

3 Free Semigroups and Varieties

3.1 Free Rees Factor-Semigroups

Let us take a set of words W, and an alphabet A, and consider the set $I(W)$ of all words from A^+ which *do not avoid* W. Then $I(W)$ is an ideal in A^+ which means that $I(W)$ is a subsemigroup closed under products by elements of A^+ from the left and from the right.

Then we can consider the *Rees factor semigroup* $A^+/I(W)$ which consists of the set of words $A^+ \setminus I(W)$ (which is, of course, the set of all words which avoid W) and 0 with an obvious multiplication: $u \cdot v = uv$ if $uv \notin I(W)$ and 0 otherwise. This is a semigroup because it is exactly the semigroup obtained by the Dilworth construction applied to the set of words $A^+ \setminus I(W)$.

Of course, the Rees factor-semigroup may be defined for arbitrary ideal of A^+. But our factor-semigroup satisfies the following remarkable property.

Theorem 3.1 *$A^+/I(W)$ is (relatively) free which means that every map ϕ from the set of generators A into $A^+/I(W)$ may be uniquely extended to an endomorphism of the semigroup to itself.*

Theorem 3.2 *A factor-semigroup A^+/I is free if and only if $I = I(W)$ for some set of words W.*

Proof. Take $W = I$.

Definition. The ideal $I(W)$ is called the *verbal* ideal defined by W.

Theorem 3.3 *An ideal I of A^+ is verbal if and only if it is stable under all endomorphisms of A^+.*

Ex. 17 *Prove this theorem.*

3.2 Free Associative Algebras

Definition An associative algebra over a field K is a K-vector space with associative and distributive binary operation.

The absolutely free associative algebra over an alphabet A is the set KA^+ of all linear combinations of words from A^+ with natural addition and distributive multiplication.

We call linear combinations of words *polynomials*. Accordingly words are sometimes called *monomials*.

We call any map $A \to KA^+$ a *substitution*. Every substitution can be extended to an endomorphism of KA^+ (this is why we call KA^+ free).

We say that polynomial p *does not avoid* polynomial q if $p = r\phi(q)s$ for some substitution ϕ and some polynomials r and s.

If we take any set of polynomials $W \subseteq KA^+$ and consider the set $KI(W)$ of all linear combinations of polynomials which do not avoid W, then $KI(W)$ is an ideal of KA^+ which is called the *verbal* or the $T-$ ideal defined by W. The following Theorem is similar to the Theorem 3.2.

Theorem 3.4 *The factor-algebra* KA^+/I *is free if and only if* $I = KI(W)$ *for some set* W *of polynomials from* KA^+. *An ideal* I *is verbal if and only if it is stable under all endomorphisms of* KA^+.

The difference between theorems 3.2 and 3.4 is that the theorem 3.4 describes all free algebras (because every algebra is a factor-algebra of an absolutely free algebra over some ideal) while theorem 3.2 gives only some free semigroups (because not every semigroup is a Rees factor of an absolutely free semigroup).

3.3 A Description of Free Semigroups

Every semigroup is a factor-semigroup of a free semigroup A^+ over some congruence σ. So we have to describe congruences which give us free semigroups.

Theorem 3.5 *A factor-semigroup* A^+/σ *is free if and only if* σ *is stable under all endomorphisms of* A^+.

Ex. 18 *Prove this theorem.*

We will call congruences with the property from Theorem 3.5 *verbal* congruences.

We will not distinguish between pairs (u, v) and identities. In particular if W is a set of pairs of word then we will say that an identity $u = v$ belongs to W if $(u, v) \in W$.

Notice that if a pair (u, v) is from a verbal congruence σ then $u = v$ is always an identity of the free semigroup A^+/σ. Conversely if $u = v$ is an identity of A^+/σ then $(u, v) \in \sigma$. In particular, if I is a verbal ideal of A^+ and $v \in I$ then A^+/I satisfies the identity $v = 0$.

Now, given a set of identities $W = \{u_i = v_i \mid i \in S\}$ and an alphabet A one can define a congruence $\sigma(W)$ as follows.

We say that a pair of words (u, v) *does not avoid* a pair of words (p, q) if $u = s\phi(p)t, v = s\phi(q)t$ for some substitution ϕ and some words s and t.

Take the set $I(W)$ of all pairs which do not avoid W and take the reflexive, symmetric and transitive closure of this set. This is a congruence which will be denoted by $\sigma(W)$.

Remark. It is interesting to notice the similarity between the semigroup and associative algebra cases. Recall that in the algebra case the verbal ideal consists of all linear combinations of polynomials which do not avoid polynomials from W. If we associate the polynomial $u - v$ with every pair of words (u, v) then the pair (u, v) does not avoid the pair (p, q) if and only if the polynomial $u - v$ does not avoid the polynomial $p - q$ (if we allow only the substitutions $A \to A^+$, and not arbitrary substitutions $A \to KA^+$). Now the pair (v, u) corresponds to the polynomial $v - u = -(u - v) = -1 \cdot (u - v)$. Thus the symmetric closure corresponds to the multiplication by an element from the field K. The transitive closure corresponds to the taking sums. Indeed, if the pair (u, w) is obtained by the transitivity from pairs (u, v) and (v, w) then the polynomial $u - w$ is a sum of polynomials $u - v$ and $v - w$.

Theorem 3.6 *A congruence σ on A^+ is verbal if and only if it is equal to $\sigma(W)$ for some set of identities W.*

Ex. 19 *Prove this theorem.*

We call a semigroup $S =< A >$ free in a class $\mathcal{K}$ of semigroups if every map $A \to T \in \mathcal{K}$ may be extended to a homomorphism from S to T.

It is clear that a semigroup is (relatively) free if and only if it is free in the class $\{S\}$. It turns out that if S is free then it is free in a much bigger class.

Let $S = A^+/\sigma$ be free and $\sigma = \sigma(W)$. Then S satisfies all the identities from W and moreover it satisfies all other identities $u = v$ where $(u, v) \in \sigma$.

Ex. 20 *If S is free in the class $\mathcal{K}$ then every semigroup T from $\mathcal{K}$ satisfies all the identities from W (and from σ). The class of all semigroups which satisfy W is the largest class of semigroups where S is free.*

Definition. Let W be a set of identities. The class of all semigroups which satisfy identities from W is called the *variety* defined by W.

Therefore a semigroup is free if and only if it is free in some variety of semigroups.

Remark. The same is true for associative algebras, and other systems.

Examples of varieties of semigroups and associative algebras.

Ex. 21 *The class of all semigroups with 0 satisfying identically $u = 0$ (u is a fixed word) is a variety. It is defined by two identities $uz = u$, $zu = u$ where z is a letter not in $c(u)$. The verbal congruence on A^+, defined by these identities, coincides with the Rees ideal congruence corresponding to the ideal $I(\{u\})$.*

Ex. 22 *The class of all semigroups satisfying identically $u = 1$ (u is a fixed word) is a variety consisting of groups. This variety is defined by two identities $uz = z$, $zu = z$ where z is a letter not in $c(u)$. All groups of this variety satisfy the identity $x^n = 1$ for some n (n depends on u). What is the minimal n with this property?*

Ex. 23 *The class of all commutative semigroups (algebras) is a variety given by the identity $xy = yx$ (resp. $xy - yx = 0$). The verbal congruence on A^+ (ideal of kA^+) corresponding to this identity consists of all pairs (u, v) (all polynomials $\sum_i \alpha_i u_i$) such that v is a permutation of u (for each u_i the sum of all α_j such that u_j is a permutation of u_i is 0).*

Ex. 24 *Describe the verbal congruence on A^+ (the verbal ideal of kA^+ defined by the identity $xyzt = xzyt$ (resp. $xyzt - xzyt = 0$)*

Ex. 25 *Describe the verbal congruence on A^+, defined by the system of identities:*

a) $xyx = x^2y$.

b) $xyx = yx^2$.

c) $x^2y = xy,\ x^2y^2 = y^2x^2$.

3.4 The Structure Description Of Varieties

The following statement may be found in every universal algebra book.

Theorem 3.7 *(*Birkhoff.*) A class of semigroups (groups, associative algebras, etc.) is a variety if and only if it is closed under taking homomorphisms, direct products and subsemigroups (subgroups, etc.).*

Thus varieties may be defined in a "syntactic" way (by identities) and in a "semantic" way (as classes closed under these three most popular algebraic constructions).

The similar situation may be found in other parts of mathematics. For example, a manifold can be defined ("syntactically") by equations and ("semantically") as a locally Euclidean topological spaces.

The fact that a manifold is locally Euclidean means that if we are on this manifold, and cannot go very far (or if we can not memorize big volumes of data) then we won't be able to distinguish the manifold from the ordinary Euclidean space.

The fact that a variety of semigroups (or other systems) is closed under the three constructions means that we can "live" inside a variety, use these constructions and never need any algebras outside the variety. This is why, by the way, the commutative algebra is considered a separate part of mathematics.

Actually the variety of commutative semigroups (associative algebras) plays in some sense the same role in general algebra as Euclidean spaces in geometry and topology. These are the simplest varieties and all other varieties have some features of the variety of commutative semigroups (associative algebras).

4 The Burnside Problem for Varieties

Recall that we started with the problem of finding a finitely generated infinite periodic semigroup. We found such a semigroup which satisfies even the identity $x^2 = 0$. Then we described all words u such that there exists a finitely generated infinite periodic semigroup satisfying the identity $u = 0$. These are exactly the avoidable words. Then we described finite sets W of words such that there exist a finitely generated infinite periodic semigroup satisfying all identities $u = 0$ where $u \in W$.

Thus we have described all varieties given by finitely many identities of the form $u = 0$, which contain infinite finitely generated periodic semigroups. Actually we found a syntactic characterization of these varieties.

Now that we know that we were dealing with varieties, it is natural to ask

> *What are all the varieties of semigroups which contain infinite periodic finitely generated semigroups?*

This problem is extremely difficult. Indeed, since the class of all groups satisfying the identity $x^n = 1$ is a variety of semigroups (see Ex. 22) this problem "contains" the problem of describing n such that every finitely generated group with the identity $x^n = 1$ is finite. As we saw above, this problem seems to be very difficult, and we still do not know if, say, 5 is such an n.

Nevertheless the following similar problem turned out to be decidable. Notice that all infinite periodic semigroups that we constructed above turned out to be *nil*-, that is they contain 0 and every element in some power is equal to 0. Thus we can ask

> *What are all varieties of semigroups containing infinite finitely generated nil-semigroups?*

It is clear that every nil-semigroup is periodic, but not every periodic semigroup is nil. For example a non-trivial periodic group can not be nil (it does not have 0). Thus if we consider nil-semigroups instead of arbitrary periodic semigroups, we avoid "bad" groups. It turned out that as soon as we do that, the situation becomes much more comfortable, and we have the following result (Sapir, 1987).

Theorem 4.1 *Let V be a variety given by a finite number of identities Σ. Then the following conditions are equivalent:*

1. *V does not contain an infinite finitely generated nil-semigroup.*

2. *V does not contain an infinite finitely generated semigroup satisfying the identity $x^2 = 0$.*

3. *V satisfies an identity $Z_{n+1} = W$ where n is the number of letters in words participating in Σ, W is a word distinct from Z_{n+1}.*

4. *There exists an identity $u = v$ from Σ and a substitution ϕ such that Z_{n+1} contains $\phi(u)$ or $\phi(v)$ and $\phi(u) \not\equiv \phi(v)$.*

The equivalence $3 \equiv 4$ is easy to establish. In fact Z_{n+1} in these conditions may be replaced by an arbitrary word Z. Indeed, if 4 holds then $Z = s\phi(u)t$ and $s\phi(v)t \neq s\phi(u)t$ for some substitution ϕ, words s, t, and identity $u = v \in \Sigma$. Let $W = s\phi(v)t$. Then the pair (Z, W) does not avoid Σ. Therefore $(Z, W) \in \sigma(\Sigma)$. Hence Σ implies $Z = W$.

Conversely, if Σ implies $Z = W$ for some W distinct from Z then the pair (Z, W) belongs to the verbal congruence $\sigma(\Sigma)$ which is a reflexive, transitive, and symmetric closure of the set of pairs $I(\Sigma)$ which do not avoid Σ. Therefore there exists a chain of pairs

$$(Z, W_1), (W_1, W_2), \ldots, (W_k, W)$$

which do not avoid Σ or identities dual to identities from Σ ($u = v$ is dual to $v = u$). Since Z differs from W, we may suppose that W_1 differs from Z. Therefore the pair $Z, W_1)$ satisfies condition 4.

Remark 1. The condition 4 is effective, that is there is an algorithm which verifies this condition.

Remark 2. Let $u = v \in \Sigma$ and u is unavoidable but v is avoidable. Then condition 3 holds. Indeed, since u is unavoidable Z_n contains a value $\phi(u)$ of u for some substitution ϕ, that is $Z_n = s\phi(u)t$ for some words s and t. Since v is avoidable, Z_n cannot contain values of v. In particular, $\phi(u) \neq \phi(v)$. Then the words $Z_n = s\phi(u)t$ and $W = s\phi(v)t$ are distinct. The pair (Z_n, W) does not avoid (u, v). Therefore the identity $Z_n = W$ follows from $u = v$. Hence $Z_{n+1} = W x_{n+1} Z_n$ follows from Σ. Therefore Σ satisfies condition 3.

Remark 3. Suppose that for every $u = v \in \Sigma$ both u and v are avoidable. Then condition 4 does not hold. Indeed, if (Z_{n+1}, W) does not avoid (u, v) or (v, u) then Z_{n+1} does not avoid u or v, and then u or v will be unavoidable.

Therefore the most complicated case is when Σ contains identities $u = v$ with both u and v unavoidable and does not contain identities $u = v$ where one of the words u or v is avoidable and another one isn't. The following example shows that even if both sides of an identity $u = v$ are unavoidable, $u = v$ can imply no identity of the form $Z_{n+1} = W$.

Ex. 26 *Prove that both sides of the Mal'cev identity*

$$axbybxazbxayaxb = bxayaxbzaxbybxa$$

(which in the case of groups defines the class of all nilpotent groups of degree 3) satisfies the following two conditions

1. *Both sides of this identity are unavoidable.*
2. *This identity does not imply any identity $Z_6 = W$ where W differs from Z_6.*

The implication $1 \to 2$ is trivial. Therefore in order to complete the proof of our theorem 4.1 it is enough to prove implications $3 \to 1$ and $2 \to 4$.

We will start with implication $4 \to 1$. It is easy to see that in order to prove this implication it is enough to prove the following lemma.

Lemma 4.1 *Any finitely generated nil-semigroup S satisfying a non-trivial identity $Z_n = W$ is finite.*

We will return to this lemma later, after we get familiar with the so called symbolic dynamics.

4.1 Symbolic Dynamics

A topological dynamical system in general is a compact topological space X with a semigroup S of continuous transformations $X \to X$. The most popular semigroups which appear in applied mathematics are the group of integers $\mathbf{Z}$ (the so called discrete dynamical systems), the group of reals $\mathbf{R}$ or the semigroup of positive reals (continuous dynamical systems).

For example let M be a compact manifold without boundary (like a sphere or a torus) and F be a continuous tangent vector field on it. This vector field determines a flow on M. For every point x on M and for every real number

r we can find the point $\alpha_r(x)$ where x will be in r seconds if we start at the point x. The transformations α_r form a group isomorphic to the additive group of real numbers (if some natural analytical conditions on M and F hold). This is a typical continuous dynamical system. The transformations α_n corresponding to the integers form a discrete dynamical system.

If we take this discrete dynamical system, divide M into parts as we did in Section 1.3, then with a discrete trajectory of a point x on M we can associate an infinite in both direction sequence of labels of regions which are visited by the point x in $\ldots, -3, -2, -1, 0, 1, 2, \ldots$ seconds. If we collect all these sequences, we get a set of sequences which approximates the original dynamical system. This approximation is better if the areas of the regions are smaller (this is just like the process of approximating solutions of differential equations). An important observation made by Hadamard, Poincare and Morse says that under some (not very restrictive) conditions this set of sequences may be considered as a dynamical system itself.

Let A be a finite alphabet. Consider the set $A^{\mathbf{Z}}$ of all infinite in both directions sequences of letters from A. If $\alpha \in A^{\mathbf{Z}}$, and $m \leq n$ are integers then $\alpha(m, n)$ is the subword of α starting at the position number m and ending at the position number n. One can define a metrics on $A^{\mathbf{Z}}$ by the following rule. Let $\alpha, \beta \in A^{\mathbf{Z}}$. Let n be the largest number such that the word $\alpha(-n, n)$ coincides with the word $\beta(-n, n)$. Then the distance $dist(\alpha, \beta)$ between α and β is equal to $\frac{1}{2^n}$. This metrics makes $A^{\mathbf{Z}}$ a compact topological space (which is homeomorphic to the Cantor set). Let T be a shift on $A^{\mathbf{Z}}$ to the right, that is $T(\alpha)(i, i) = \alpha(i + 1, i + 1)$. It is easy to prove that T and T^{-1} are continuous maps of $A^{\mathbf{Z}}$ onto itself.

Ex. 27 *Prove this.*

Therefore every power T^n of T is continuous, and $(A^{\mathbf{Z}}, < T >)$ is a dynamical system. This system and every its subsystems (that is closed subsets of $A^{\mathbf{Z}}$ which are stable under T) are called *symbolic dynamical systems*. Recall that a closed set of a compact topological space is compact itself. It is easy to see that the set of infinite sequences associated with a partition of the manifold M is stable under the shift T. If the partition is good enough, this set of sequences is closed. Therefore, we indeed have a dynamical system, and, moreover, a symbolic dynamical system.

In 1983 I discovered another source of symbolic dynamical systems: finitely generated semigroups. Let $S =< A >$ be a finitely generated semigroup.

Then as we mentioned in Section 1.3, every element of S is represented by a word over A. For every element s in S take all shortest possible words representing s. These words are called geodesic words: they label geodesics on the Cayley graph of the semigroup. Let W be the set of all geodesic words representing elements of S. Notice that W is closed under taking subwords (a subword of a geodesic word is a geodesic word itself).

Suppose S is infinite. Then W is infinite also. Now, in every word of W, mark a letter which is closest to the center of these word. There must be an infinite subset W_1 of W which have the same marked letters, an infinite subset $W_2 \subseteq W_1$ of words which have the same subwords of length 3 containing the marked letters in the middle, ..., an infinite subset $W_n \subseteq W_{n-1}$ of words which have the same subwords of length $2n-1$ containing the marked words in the middle, and so on. Therefore there is an infinite in both directions sequence α of marked letters from A such that every subword of α is a subword of a word from W. Thus every subword of α is a geodesic word. Infinite sequences with this property will be called *infinite geodesic words*. The set $D(S)$ of all infinite geodesic words of S is a closed subset of $A^{\mathbf{Z}}$ and is stable under the shift T.

Ex. 28 *Prove the last statements.*

Therefore $D(S)$ is a symbolic dynamical system. Notice that if S is finite W is also finite and $D(S)$ is empty. Thus $D(S)$ is not empty if and only if S is infinite.

It is interesting that an arbitrary symbolic dynamical system is $D(S)$ for some S. Indeed, let $D \in A^{\mathbf{Z}}$ be a symbolic dynamical system. Let W be the set of all words over A which are subwords of some sequences from D. Then W is closed under taking subwords. Thus we can apply the Dilworth construction to W and get a semigroup $S(W)$ which we will denote by $S(D)$.

Ex. 29 *Prove that for every symbolic dynamical system D we have*

$$D(S(D)) = D.$$

Thus D is the symbolic dynamical system associated with the semigroup $D(S)$.

Now let us turn to some applications of the theory of dynamical systems to semigroups.

Definition. Let $(D, < T >)$ be a dynamical system. A point $x \in D$ is called *uniformly recurrent* if for every open neighborhood $O(x)$ of x in D there exists a number $N(O(x))$ such that for every $k \in \mathbf{Z}$ one of the points

$$T^k(x), T^{k+1}(x), \ldots, T^{k+N-1}(x)$$

belongs to $O(x)$.

Theorem 4.2 (Poincare.) *Every dynamical system contains a uniformly recurrent point.*

There are some dynamical systems where almost every point is uniformly recurrent. Such are, for example, the systems of molecules of a gas in a closed room. Thus if we put an open bottle with a perfume in a closed room (at almost every point of the room), then there exists a number N depending on the size of the bottle such that after N seconds all molecules of the perfume will gather again in the bottle. The room must be **really** closed, of course.

Lemma 4.2 (Furstenberg.) *If D is a symbolic dynamical system, then a point (sequence) α is uniformly recurrent iff for every subword u of α there exists a number $N(u)$ such that every subword of α of length $N(u)$ contains u.*

Ex. 30 *Prove this lemma.*

Let α be a sequence from $A^{\mathbf{Z}}$. Let $D(\alpha)$ be the closure of all shifts $\ldots, T^{-2}(\alpha), T^{-1}(\alpha), \alpha, T^1(\alpha), T^2(\alpha), \ldots$ of α. This is a closed set and since T and T^{-1} are continuous maps $D(\alpha)$ is stable under the shift. Thus $D(\alpha)$ is a symbolic dynamical system. This is the minimal symbolic dynamical system containing α, thus we will call $D(\alpha)$ the dynamical system *generated by* α.

Lemma 4.3 *If $\beta \in D(\alpha)$ then every subword of β is a subword of α.*

Ex. 31 *Prove this lemma.*

4.2 Application of the Theory of Dynamical Systems to Semigroups

From Theorem 4.2 and Lemma 4.2 one can easily deduce the following statement.

Lemma 4.4 *For every infinite finitely generated semigroup* $S =< A >$ *there exists an infinite uniformly recurrent geodesic word.*

Now let us return to our Theorem 4.1. Recall that we were going to prove Lemma 4.1: if a finitely generated nil-semigroup S satisfies a nontrivial identity of the form $Z_n = W$ then it is finite. The following Lemma gives us a connection between uniformly recurrent sequences and Zimin words Z_n .

Lemma 4.5 *Let* β *be a uniformly recurrent sequence,* U_1aU_2 *be an occurrence of letter* a *in* β *where* U_1 *is a sequence infinite to the left,* U_2 *is a sequence infinite to the right. Then for every natural number* n *there exists a substitution* ϕ *such that* $U_3\phi_n(Z_n\ ) = U_1a$ *for some sequence* U_3 *infinite to the left,* $\phi_n(x_1) = a$*, and* $|\phi_n(Z_n\ )| \leq A(n, U)$ *where number* $A(n, U)$ *depends only on* U *and* n*.*

Proof. Since β is uniformly recurrent , there exists a number $N = N(a)$ such that every subword of β of length N contains a. Therefore one can find another a at most $N + 1$ letters to the left of our occurrence of a. Then we can set $\phi(x_1) = a$, and $\phi(x_2)$ equal to the word between our two occurrences of as. So we get a substitution of Z_2 which satisfies the required condition. Since β is uniformly recurrent , there exists a number $N_1 = N(\phi(Z_2))$ such that every subword of β of length N_1 contains $\phi(Z_2)$. So we can find another occurrence of $\phi(Z_2)$ to the left of the first occurrence, such that the distance between these two occurrences does not exceed $N_1 + 1$. Then we can define $\phi(x_3)$ to be equal to the word between these two occurrences of $\phi(Z_2)$. This gives us a substitution of Z_3 which satisfies the required condition. Now the proof is easy to complete by an induction on n.

Suppose now that S is infinite. Then by Lemma 4.4 there exists a uniformly recurrent geodesic word in $D(S)$. Fix one of these uniformly recurrent geodesic words α.

The following lemma is obvious.

Lemma 4.6 *None of the subwords of α is equal to 0 in S.*

Our goal is to get a contradiction with Lemma 4.6.

Convention If a word differs from Z_n only by the names of its letters, then we will denote it by Z_n'. In particular, words xyx, xzx, and Z_2 are denoted by Z_2'.

Recall that S satisfies the identity $Z_n = W$. First of all let us look at the word W.

Lemma 4.7 *W contains only letters $x_1, \ldots, x_n$.*

Proof. Indeed, suppose W contains an extra letter y. By Lemma 4.5 there exists a substitution ϕ such that $\phi(Z_n)$ is a subword of α . Since every letter of A is an element of S we can consider ϕ as a map into S. Let $\phi(y) = 0$. Then $\phi(W) = 0$ in S. Since S satisfies the identity $Z_n = W$, we have that $\phi(Z_n) = 0$ in S. Hence α has a subword which is equal to 0 in S - a contradiction with Lemma 4.6.

The following lemma is obvious.

Lemma 4.8 *W has one of the following 4 properties.*

1. *$W = [W_1, x_1]_1^1$ where $W_1 = W_{x_1}$.*
2. *W contains x_1^2.*
3. *W contains a subword $x_i x_j$ for some $i, j > 1$.*
4. *W starts or ends not with x_1.*

Notice that since S satisfies the identity $Z_n = W$, it satisfies the following two identities: $Z_{n+1} = Z_n x_{n+1} W$ and $Z_{n+1} = W x_{n+1} Z_n$. Now if W satisfies condition 4 then $Z_n x_{n+1} W$ or $W x_{n+1} Z_n$ satisfies condition 3. Thus we can assume that W satisfies one of the three conditions 1, 2, 3 of Lemma 4.8.

Suppose that W satisfies condition 2 or 3. Then the following statement holds.

Lemma 4.9 *Let β be an arbitrary uniformly recurrent sequence, a be a letter in β. Then β contains a subword u such that $(u, pa^2q) \in I(Z_n, W)$ for some words p and q (recall that $I(Z_n, W)$ is the set of all pairs of the form $(s\phi(Z_n)t, s\phi(W)t)$).*

Proof. If W satisfies condition 2 of Lemma 4.8, the statement is a direct consequence of Lemma 4.5

Suppose W satisfies condition 3. Since β is uniformly recurrent , it can be represented in the form $\dots p_{-2}ap_{-1}ap_1ap_2ap_3\dots$ where lengths of the words p_i are smaller than $N(a)$. Let us introduce a new alphabet $B = \{a, p_1, p_2, \dots\}$ (so we denote words p_i by letters: different words by different letters, equal words by equal letters). Since A is a finite alphabet and there are only finitely many words over A of any given length, B is also a finite alphabet. Let β_1 be the sequence which we get from β by replacing subwords p_i by the corresponding symbols. This sequence, β_1 may not be uniformly recurrent . But let us consider the symbolic dynamical system $D(\beta_1)$ generated by β_1. By the Theorem of Poincare, this system contains a uniformly recurrent sequence β_2. By Lemma 4.3 every subword of β_2 is a subword of β_1. Therefore β_2 has the form $\dots p_{i_1}ap_{i_2}ap_{i_3}a\dots$. Let p_1 be a letter from B occurring in β_2. By Lemma 4.5 there exists a substitution ϕ of the word Z_n such that $\phi(Z_n)$ is a subword of β_2 and $\phi(x_1) = p_1$. Then $\phi(x_i)$, $i = 2, 3, \dots, n$, must start and end with a. Since W contains a subword x_ix_j for $i, j > 1$, we have that $\phi(W)$ contains a^2. The word $\phi(Z_n)$ is a subword of β_1. Let ψ be the substitution which takes a to a and the symbols p_i back to the words denoted by these symbols. Then $\psi(\phi(Z_n))$ is a subword of β and $\psi(\phi(W))$ contains a^2. The lemma is proved.

Convention. If u, v, p, q are words then we will write $u = v\ (Mod\ p = q)$ if (u, v) belongs to $\sigma(p, q)$ (where that $\sigma(p, q)$ is the transitive, reflective, and symmetric closure of the set of pairs which do not avoid (p, q)). Recall that in this case the identity $u = v$ follows from the identity $p = q$, so that if S satisfies $p = q$ then S satisfies $u = v$.

Lemma 4.10 *For every uniformly recurrent sequence β, every natural number n and every letter a occurring in β there exists a subword u of β such that*

$$u = sa^nt\ (Mod\ Z_n = W)$$

for some words s and t.

Ex. 32 *Prove this lemma.*

Now we can finish the proof of Lemma 4.1 in the case when W satisfies one of conditions 2 or 3. Indeed, let us apply Lemma 4.10 to α . Let a be

a letter occurring in α . We can consider a as an element of S. Since S is a nil-semigroup $a^n = 0$ in S for some n. By Lemma 4.10 there exists a subword u in α such that

$$u = sa^n t \ (Mod\ Z_n = W).$$

Hence the identity $u = sa^n t$ follows from $Z_n = W$. Since S satisfies $Z_n = W$ we can conclude that S satisfies the identity $u = sa^n t$. Therefore, in particular, the word u is equal to $sa^n t$ in S. But $sa^n t = 0$ in S. This contradicts Lemma 4.6.

It remains to consider the case when W satisfies condition 1 of Lemma 4.8.

Lemma 4.11 *If* $u = v \ (Mod\ p = q)$ *then* $[u, a]_1^1 = [v, a]_1^1 \ (Mod\ [p, a]_1^1 = [q, a]_1^1)$.

Ex. 33 *Prove this lemma.*

Lemma 4.12 *For every uniformly recurrent sequence* β *there exists a subword* $p < \beta$ *such that for every natural number* n *there exists a subword* u *in* β *such that*

$$u = sp^n t \ (Mod\ Z_n = W)$$

for some words s, t.

Proof. Induction on n. We have proved this lemma in the case when W satisfies condition 2 or 3 of Lemma 4.8 (Lemma 4.10). This is the base of the induction.

Suppose we have proved our lemma for $n - 1$ and that W satisfies the condition 1, that is $W = [W_1, x_1]_1^1$. We have $Z_n = [Z'_{n-1}, x_1]_1^1$. Thus the identity $Z'_{n-1} = W_1$ is nontrivial. So we can suppose that the statement of our lemma holds for this identity. As in the proof of Lemma 4.9, let us represent β in the form $\ldots p_1 a p_2 a p_3 \ldots$, and replace each subword p_i by the corresponding symbol. We get another sequence β_1. Let β_2 be the sequence obtained from β_1 by deleting a. Let β_3 be an uniformly recurrent sequence in $D(\beta_2)$. By the induction hypothesis there exists a word p in β_3 such that for every n there exists a word u in β_3 such that $u = sp^n t \ (Mod\ Z'_{n-1} = W_1)$. Then by Lemma 4.11 we have that $[u, a]_1^1 = [s, a]_1^0([p, a]_1^0)^n [t, a]_1^1 \ (Mod\ Z_n = W)$. The word u is a subword of β_3. By Lemma 4.3 it is a subword of β_2. Then $[u, a]_1^1$

is a subword of β_1. Let ψ be the "return" substitution which takes a to a and every symbol p_i to the word which is denoted by this symbol. Then we have that

$$\psi([u,a]_1^1) = s_1(\psi([p,a]_1^0))^n t_1 \ (Mod\ Z_n = W)$$

for some words s_1 and t_1. The word $\psi([u,a]_1^1)$ is a subword of β, and so $\psi([p,a]_1^0)$ is the desired word p. The lemma is proved.

Ex. 34 *Complete the proof of Lemma 4.1*

Lemma 4.1 gives us the implication $4 \to 1$.
It remains to prove the implication $2 \to 4$.

4.3 Implication $2 \to 4$

Suppose that condition 4 of Theorem 4.1 does not hold. We have to prove that then there exists a finitely generated infinite semigroup S satisfying all identities of Σ and the identity $x^2 = 0$.

Definition. A word w is called an *isoterm* for an identity $u = v$ if for every substitution ϕ $\quad \phi(u) \leq w$ implies $\phi(v) \equiv \phi(u)$.

Remark 1. A word w may be an isoterm for $u = v$ but not for $v = u$. For example, xyx is an isoterm for $x^2 = x$ but not for $x = x^2$.

Remark 2. Condition 4 of Theorem 4.1 may be rewritten in the form:

> Z_{n+1} is not an isoterm for $u = v$ or $v = u$ for some identity $u = v$ in Σ.

Since we assume that this condition does not hold, Z_{n+1} is an isoterm for $u = v$ and $v = u$ for every $u = v \in \Sigma$.

In order to construct an infinite finitely generated semigroup satisfying Σ and $x^2 = 0$, we will employ the Dilworth method again.

The following lemma is an analogue of Theorem 2.8.

Lemma 4.13 *Let W be a set of words closed under taking subwords. The semigroup $S(W)$ satisfies an identity $u = v$ iff every word of W is an isoterm for $u = v$ and $v = u$.*

Ex. 35 *Prove this lemma.*

Thus, *in order to find a finitely generated semigroup which satisfies* $x^2 = 0$ *and* Σ *it is enough to construct an infinite set* W *of square-free words over a finite alphabet which are isoterms for* $u = v$ *and* $v = u$ *for every* $u = v \in \Sigma$.

One can see that, again, we have translated a semantic question about semigroups into a syntactic question about words.

Let γ be the substitution defined in Section 2.11. We will complete the proof of implication $2 \to 4$ if we prove the following result.

Lemma 4.14 *Let* $u = v$ *be an identity in* n *variables such that* Z_{n+1} *is an isoterm for* $u = v$ *and* $v = u$. *Then all words* $\gamma^m(a_{11})$, $m \geq 1$ *are isoterms for* $u = v$ *and* $v = u$ *whenever* $r > 6n$.

We shall divide the proof of Lemma 4.14 into a number of steps.

First of all we shall study identities $u = v$ such that Z_m is an isoterm for $u = v$ and $v = u$. Instead of the infinite set of words Z_m, $m \geq 1$ we will consider the infinite sequence

$$Z_\infty = [\ldots[x_1, x_2]_1^1, x_3]_1^1 \ldots$$

which is a "limit" of Z_n'. Any sequence which differs from Z_∞ only by the names of letters, also will be denoted by Z_∞.

The following Lemma is obvious.

Lemma 4.15 *If* Z_m *is not an isoterm for* $u = v$ *then* Z_{m+1} *is not an isoterm for* $u = v$.

Lemma 4.16 *Let* B, C *be a fusion in* u *but not in* v, $|c(u)| = n$. *Let* $A \subseteq B \backslash C$ *and* u_A *is unavoidable. Then* Z_∞ *is not an isoterm for* $u = v$.

Proof By the Theorem of Zimin, there exists a substitution ϕ such that $\phi(u_A) < Z_\infty$. Consider ϕ^*. We have that $\phi^*(u) < Z_\infty$. Since B, C do not form a fusion in v, there exists a subword $xy < v$ such that either $x \notin B$, $y \in C$ or else $x \in B$, $y \notin C$. Suppose $x \notin B$, $y \in C$. Then $\phi^*(x)$ does not end with a, and $\phi^*(y)$ does not start with a (see the definition (2)). Therefore $\phi^*(v)$ contains a subword zt where $z \neq a \neq t$. It follows that $\phi^*(u) \neq \phi^*(v)$ since $\phi^*(u) < Z_\infty$. Suppose $x \in B$, $y \notin C$. Then $\phi^*(x)$ ends and $\phi^*(y)$ starts with a, so that $a^2 < \phi^*(v)$ and again $\phi^*(u) \neq \phi^*(v)$. Thus $\phi^*(u) < Z_\infty$ and $\phi^*(u) \neq \phi^*(v)$, that is Z_∞ is not an isoterm for $u = v$. The Lemma is proved.

Lemma 4.17 *Let A be a free set in u such that Z_m is not an isoterm for $u_A = v_A$. Then Z_{m+1} is not an isoterm for $u = v$.*

Proof. Let ϕ be a substitution such that $\phi(u_A) < Z_m$ and $\phi(u_A) \neq \phi(v_A)$. Then we can define ϕ^* as in (2). By Lemma 2.5, $\phi^*(u)$ is a subword of Z'_{m+1} and $\phi^*(u) \neq \phi^*(v)$. The lemma is proved.

Lemma 4.18 *Let $u = v$ be an identity of n variables such that Z_{n+1} is an isoterm for $u = v$. Then Z_∞ is an isoterm for $u = v$.*

Proof. Let $u = v$ be a counterexample to our statement and let the length $|uv|$ be minimal. Every subword of Z_∞ is contained in some Z'_m. Therefore, since Z_∞ is not an isoterm for $u = v$, there exists a number m such that Z_m is not an isoterm for $u = v$. If $m \leq n + 1$ then we can apply Lemma 4.15. Thus we can assume that $m > n + 1$.

There exists a substitution ϕ such that $\phi(u) \leq Z_m$ and $\phi(u) \neq \phi(v)$. Then u is an unavoidable word and by the Theorem of Zimin (Theorem 2.10) there exists a substitution ψ such that $\psi(u) \leq Z_n$. Since $m > n + 1$, we have that $\psi(u) = \psi(v)$ for every such substitution ψ. Therefore $c(v) \subseteq c(u)$ (otherwise we can define ψ on the extra letters from v in such a way that $\psi(u) \neq \psi(u)$).

Let us prove that $\phi(u)_{x_1} = \phi(v)_{x_1}$. Suppose, by contradiction, that $u' = \phi(u)_{x_1} \neq \phi(v)_{x_1} = v'$. Notice that $u' \leq Z'_{m-1}$. Let $A = \{x \in c(u) \mid \phi(x) = x_1\}$.

If A is empty then u' is a value of u under some substitution $\phi' = \phi_{x_1}$, and $v' = \phi'(v)$ (see Lemma 2.8. Then $\phi'(u) \leq Z_{m-1}$ and $\phi'(u) \neq \phi'(v)$ - a contradiction with the minimality of m.

Thus A is not empty. By Lemma 2.9 A is a free set in u. The word u' is a value of the word u_A under the substitution $\phi' = \phi_{x_1}$, and $v' = \phi'(v_A)$. Therefore Z_{m-1} is not an isoterm for the identity $u_A = v_A$. Since $|u_A v_A| < |uv|$, we can conclude that Z_n is not an isoterm for $u_A = v_A$. By Lemma 4.17 we conclude that Z_{n+1} is not an isoterm for $u = v$ - a contradiction.

Thus, indeed, $\phi(u)_{x_1} = \phi(v)_{x_1}$. Therefore either one of the words $\phi(u)$ and $\phi(v)$ starts or ends with x_1 and another one doesn't, or $\phi(v)$ contains a subword $x_p x_q$ for $p, q \neq 1$, or $\phi(v)$ contains x_1^2. It is clear that these 2-letter subwords cannot occur in $\phi(x)$ for some letter x, otherwise $\phi(u)$ would contain such subwords also, which is impossible since $\phi(u) \leq Z'_m$. Therefore these subwords may occur on the boundaries between $\phi(x)$ and $\phi(y)$.

Let us define a substitution ψ^ϕ by the following rule. For every $x \in c(u)$ we have $\phi(x) = [\phi(x)_{x_1}, x_1]_{\epsilon_x}^{\delta_x}$. Then let $\psi^\phi(x) = [phi(x), a]_{\epsilon_x}^{\delta_x}$.

Now it is easy to see that $\phi(x)$ starts (ends) with x_1 then $\psi^\phi(u)$ starts (ends) with a, and the same holds for v. Therefore if $\phi(u)$ (resp. $\phi(v)$) starts or ends with x_1, but $\phi(v)$ (resp. $\phi(u)$) does not then $\psi^\phi(u)$ (resp. $\psi^\phi(v)$) starts or ends with a, but $\psi^\phi(v)$ (resp. $\psi\phi(u)$) does not, so that $\psi^\phi(u) \neq \psi^\phi(v)$.

Also it is easy to see that $\phi^\psi(u)$ cannot contain a^2 and $x_i x_j$. Therefore

$$\psi^\phi(u) = [\psi^\phi(u)_a, a]_\epsilon^\delta = [\psi(u), a]_\epsilon^\delta$$

for some $\epsilon,\ \delta$, so $\psi^\phi(u) < Z_{n+1}$.

On the other hand, if v contains a subword xy such that $\phi(x)\phi(y)$ contains x_1^2 or $x_i x_j$ for $i, j > 1$, and this word appear on the boundary between $\phi(x)$ and $\phi(y)$, then we can conclude that $\psi^\phi(v) > \psi^\phi(xy)$ and so $\psi^\phi(v)$ contains either a^2 or a word $x_i x_j$.

In all cases $\psi^\phi(u) < Z_{n+1}$ and $\psi^\phi(u) \neq \psi^\phi(v)$ which contradicts the minimality of m. The lemma is proved.

Recall that in the case of avoidable words the crucial role was played by the so called quasi-integral words (see (3)). Now we need more detailed analysis of such words.

Lemma 4.19 *$u = P_0 Q_1 y_1 P_1 Q_2 y_2 P_2 \dots Q_n y_k P_k Q_{k+1}$ be a quasi-integral word, $v = S_1 y_1 S_2 \dots y_k S_{k+1}$ where S_i are words in the alphabet A_r. Let $T = y_1 y_2 \dots y_k$ be an unavoidable word and $u \neq v$. Then Z_∞ is not an isoterm for $u = v$.*

Proof. By contradiction, suppose that Z_∞ is an isoterm for $u = v$.

1. We can assume that u and v start and end with a letter from $Y = \{y_1, \dots, y_k\}$. Otherwise we can multiply this identity by a new letter from the left and by another new letter from the right, and then include these letters in L. All conditions of the lemma will be preserved. Thus we can assume that $P_0 Q_1, S_1, P_k Q_{k+1}, S_{k+1}$ are empty words.

2. Since $u \neq v$, there exists a number ℓ such that $S_\ell \neq P_\ell Q_{\ell+1}$.

3. Let, as before, σ_i be the deletion of letters from A_{ri}. We also denote the deletion of the letter a_{ij} by σ_{ij} and the deletion of all letters from A_r except a_{ij} by σ'_{ij},.

We know that the sequence $\sigma_2, \sigma_3, \ldots, \sigma_r, \sigma_1$ is a sequence of free deletions in any quasi-integral word. As a result of these deletions we will get the word T which is unavoidable. By Theorem of BEM and Zimin we get that u is also unavoidable. Similarly every word which we can get from u by a sequence of deletions $\sigma_{i_1}, \ldots, \sigma_{i_s}$ for $s \leq r$ is unavoidable.

4. Take arbitrary j, $1 \leq j \leq r$. Let $v = \sigma_2\sigma_3 \ldots \sigma_r(v)$. In u_1, the deletion σ'_{j1} is a free deletion because a subset of a free set is free. Sets $B_1 = \{a_{j1}\}$ and $C_1 = \{y_t \mid Q_t \text{ contains } a_{j1}\}$ form a fusion in $\sigma'_{j1}(u_1)$ Therefore σ_1 is a free deletion in $\sigma'(u_1)$. If B_1, C_1 do not form a fusion in $\sigma'_{j1}(v_1)$ then by Lemma 4.16 Z_∞ is not an isoterm for $u_1 = v_1$, and by Lemma 4.17 Z_∞ is not an isoterm for $u = v$, a contradiction.

Thus B_1 and C_1 form a fusion in $\sigma'_{j1}(u_1)$. This implies that S_ℓ contains a_{j1} if and only if $Q_{\ell+1}$ contains this letter, and that this letter appears in S_ℓ only once.

5. Let now $1 \leq i, j \leq r$. Let u_2, v_2 be words obtained from u and v by the deletion of all letters which are not in $A_{ri} \cup A_{r1} \cup Y$. We know that u_2 is obtained from u by a sequence of free deletions.

As above σ'_{ij} is a free deletion in u_2. In the word $\sigma'_{ij}(u_2)$ sets

$$B_2 = \{a_{ij}\} \cup \{y_t \mid P_t \text{ does not contain } a_{ij}\}$$

and

$$C_2 = A_{r1} \cup \{y_t \mid Q_t \text{ either contains } a_{ij} \text{ or empty}\}$$

form a fusion. Therefore they form a fusion in $\sigma'_{ij}(v_2)$. This implies that S_ℓ can be represented in the form PQ where

- P contains the same letters as P_ℓ,
- Q contains the same letters as $Q_{\ell+1}$,
- if $Q_{\ell+1}$ is not empty then Q and $Q_{\ell+1}$ start with the same letter from A_{r1},
- each letter occurs in P (in Q) only once.

Therefore P is a permutation of the word P_ℓ, Q is a permutation of the word $Q_{\ell+1}$.

6. Let $1 \leq j \leq r$. Let j' equal $j+1$ if $j < r$ and 1 if $j = r$. In the word u sets $B_3 = A_{rj} \cup \{y_t \mid P_t \text{ starts with a letter from } A_{rj'}\}$ and $C_3 =$

$A_{rj'} \cup \{y_t \mid Q_t$ ends with a letter from $A_{rj}\}$ form a fusion. Since $\sigma_j(u)$ is unavoidable, these sets form a fusion in v. Therefore, in S_ℓ after every letter of A_{rj} there is a letter from $A_{rj'}$. Since we have proved in 5 that words P and Q are permutations of P_ℓ and $Q_{\ell+1}$, and that Q and $Q_{\ell+1}$ have a common first letter, we can conclude that $P = P_\ell,\ Q = Q_{\ell+1}$, so that $S_\ell = P_\ell Q_{\ell+1}$ which contradicts the assumption in 2.

The lemma is proved.

We will also need the following simple lemma.

Lemma 4.20 *Let u be a word, and let ϕ be a substitution which takes every letter x to a product of distinct letters $x_1 x_2 \ldots x_{k_x}$ (x_i are different for different x). Then u can be obtained from $\phi(u)$ by a sequence of free deletions and by renaming letters.*

Ex. 36 *Prove this lemma.*

Now we are ready to finish the proof of our Lemma 4.14.

Fix some numbers n and $r > 6n+1$. We assume that Lemma 4.14 is false and that m is the smallest number such that $\gamma^m(a_{11})$ is not an isoterm for an identity $u = v$ in n letters, but Z_{n+1} is an isoterm for $u = v$ and $v = u$, and that n is minimal number of letters for which such an m exists.

Since Z_{n+1} is an isoterm for $u = v$ and $v = u$, by Lemma 4.18 Z_∞ is an isoterm for $u = v$ and $v = u$.

By assumption there exists a substitution ϕ such that $\phi(u) \neq \phi(v)$ and $\phi(u) < \gamma^m(a_{11})$. We have met this situation before, in the proof of Theorem 2.11 and we have shown that there exists a substitution ϵ and a substitution ϕ_1 such that

- ϵ takes every letter to a product of at most three different letters. These letters are different for different letters of $c(u)$.
- $\phi_1(x)$ is either a product of $\gamma-$blocks or a subword of a block.
- $\phi = \phi_1\epsilon$.

We also know that in each $\gamma-$block one can find a two-letter subword p such that if p intersects (overlaps) with one of $\phi_1(x)$ then p is a subword of $\phi_1(x)$.

Now we can define the following substitution η for every x in $c(u)$:

$$\eta(x) = \begin{cases} x & \text{if } \phi_1(x) \text{ contains one of these two-letter subwords } p \\ \phi_1(x) & \text{otherwise.} \end{cases}$$

The word $\bar{u} = \eta\epsilon(u)$ is quasi-integral.

The word $\bar{v} = \eta\epsilon(v)$ has the form $S_0x_1S_1 \dots S_fx_fS_{f+1}$ where $x_i \in c(\epsilon(u))$ and $c(S_i) \subseteq A$.

Since $\bar{u}$ is quasi-integral $T_u = \bar{u}_A$ is obtained from $\bar{u}$ by a sequence of free deletions (see an exercise above). The word T_u is obtained from $\epsilon(u)$ by a deletion of some letters. Therefore by Lemma 2.9 T_u can be obtained from $\epsilon(u)$ by a sequence of free deletions. Therefore $T_u = \epsilon_1(u_1)$ for some substitution ϵ_1 and some word u_1 which is obtained from u by a sequence of free deletions. Notice that for every letter x, $\epsilon_1(x)$ is obtained from $\epsilon(x)$ by a deletion of some letters.

Let us denote the word obtained from v by deleting letters from $c(u)\backslash c(u_1)$ by v_1. Then $\epsilon_1(v_1) = T_v$.

Now define a substitution δ of the word T_u by $\delta(x_i) = Q_ix_iP_i$. Then the word $\phi'\delta(x_i)$ is a product of γ-blocks. Therefore we can take $\gamma^{-1}\phi'\delta(x)$. Since $\phi'(T_u)$ is a subword of $\gamma^m(a_{11})$, we have that $\gamma^{-1}\phi'\delta(T_u)$ is a subword of $\gamma^{m-1}(a_{11})$. Since $T_u = \epsilon_1(u_1)$ we have that $\gamma^{m-1}(a_{11})$ contains a value of u_1. In the proof of Theorem 2.11 we have shown that this implies that u_1 is unavoidable. Since u_1 is obtained from u by a series of free deletions and Z_∞ is an isoterm for $u = v$ we can use Lemma 4.17 and conclude that Z_∞ is an isoterm for $u_1 = v_1$.

Suppose that Z_∞ is not an isoterm for $v_1 = u_1$. Then Z_∞ contains a value of v_1, which differs from the corresponding value of u_1. In particular, since Z_∞ contains a value of v_1, this word is unavoidable. By Lemma 4.16 v_1 is obtained from v by a sequence of free deletions (the same sequence was used to get u_1 from u). This contradicts Lemma 4.17. Thus Z_∞ is an isoterm for $u_1 = v_1$ and $v_1 = u_1$.

From the minimality of m we can deduce that $\gamma^{m-1}(a_{11})$ is an isoterm for $u_1 = v_1$ and $v_1 = u_1$. Since $\gamma^{m-1}(a_{11})$ contains a value $\gamma^{-1}\phi'\delta\epsilon_1(u_1)$ of u_1, this value of u_1 must coincide with the corresponding value of v_1. Therefore, in particular, $\gamma^{m-1}(a_{11})$ contains a value of v_1.

As above, this implies that v_1 is unavoidable. As we know v_1 is obtained from v by a series of free deletions. Since Z_∞ is an isoterm for $v = u$, we can

conclude that it is an isoterm for $v_1 = u_1$.

We have already proved that

$$\gamma^{-1}\phi'\delta\epsilon_1(u_1) = \gamma^{-1}\phi'\delta\epsilon_1(v_1).$$

Therefore

$$\phi'\delta\epsilon_1(u_1) = \phi'\delta\epsilon_1(v_1).$$

Let us denote $P_0\delta\epsilon_1(v_1)Q_{f+1}$ by w. We have

$$\phi'(w) = P_0\phi'\delta\epsilon_1(v_1)Q_{f+1} = \phi(u) \neq \phi(v) = \phi'(\bar{v}).$$

Thus, in particular, $w \neq \bar{v}$. By definition, w is a quasi-integral word and T_v is equal to w_A. Recall that T_v is a value of an unavoidable word v_1, and the corresponding substitution ϵ_1 takes every letter to a product of (at most 3) different letters. By Lemma 4.20, T_v is unavoidable. Thus all conditions of Lemma 4.19 hold and we can conclude that Z_∞ is not an isoterm for $w = \bar{v}$.

Therefore there exists a substitution θ such that $\theta(w) < Z_\infty$ and $\theta(w) \neq \theta(\bar{v})$.

We have proved that Z_∞ is an isoterm for $v_1 = u_1$. Since

$$w = P_0\delta\epsilon_1(v_1)Q_{f+1}$$

and

$$\bar{u} = P_0\delta\epsilon_1(u_1)Q_{f+1},$$

we have that Z_∞ is an isoterm for $w = \bar{u}$. Hence

$$\theta(\bar{(u)}) = \theta(w) \neq \theta(\bar{v}).$$

This means that Z_∞ is not an isoterm for $\bar{u} = \bar{v}$. But

$$\bar{u} = \eta\epsilon(u),\ \bar{v} = \eta\epsilon(v).$$

Therefore Z_∞ is not an isoterm for $u = v$, a contradiction.

The theorem is proved.

5 Burnside Problems and the Finite Basis Property

We say that a variety is *finitely based* if is may be defined by a finite set of identities.

All varieties which we met before were finitely based. Now we will use Theorem 4.1 to construct non-finitely based varieties.

We can rewrite this theorem in the following way.

Theorem 5.1 *Let calV be a variety of semigroups which satisfies the following two properties:*

1. *every finitely generated nil-semigroup in calV is finite;*

2. *calV does not satisfy any non-trivial identity of the form $Z_n = W$.*

Then calV cannot be defined by a finite set of identities.

We shall show that it is easy to construct a variety which satisfies both conditions 1 and 2. We know that a variety may be defined syntactically, by identities, and semantically, as a class closed under the three operations: $\mathcal{P}$ - direct products, $\mathcal{S}$ - taking subalgebras, $\mathcal{H}$ - taking homomorphic images. We have considered the syntactic way of defining varieties. Now we turn to the semantic way.

Let C be a class of algebras. Then we can consider the minimal variety containing C. It is clear that such a variety exists because the intersection of varieties is again a variety. We will denote this variety by $varC$. By a Birkhoff theorem (which I do not want to prove here) this variety is precisely the class of algebras which are homomorphic images of subalgebras of direct products of algebras from C. Thus we have the following formula

$$varC = \mathcal{HSP}(C). \tag{4}$$

We will say that the variety $varC$ is *defined by the class C*.

This formula is very useful for proving that a variety defined by a class of algebras satisfies the condition 1 of Theorem 5.1.

Theorem 5.2 *Let C be a class of algebras with finitely many operations. Then the following conditions are equivalent.*

1. *Every finitely generated algebra in $varC$ is finite.*

2. *For every natural number m there exists such number n that the order (i.e. the number of elements) of every m-generated subalgebra of any algebra from C does not exceed n.*

The second condition means that the orders of all m-generated subalgebras of algebras from $varC$ are bounded from the above.

Ex. 37 *a) Prove that the class of all finite semigroups does not satisfy condition 2 of Theorem 5.2, but any class consisting of finite number of finite semigroups satisfies this condition.*

b) Give an example of a class C consisting of one infinite group, which satisfies condition 2.

Proof of Theorem 5.2. $1 \to 2$. Suppose that every finitely generated algebra in $varC$ is finite. Let us take any number m and any m-generated subalgebra $S =< X >$ of an algebra from C. Since every variety is closed under taking subalgebras $S \in varC$. As any variety, $varC$ contains free algebras with any number of generators (we showed this in the case of semigroups, the general case is similar). Let us take the free algebra F_m with m generators. We may suppose that F_m is generated by the same set X. Since every finitely generated algebra in $varC$ is finite, F_m is finite. Let n be the number of elements in F_m. We know that every free algebra is free "outside", that is every map from the set of generators X to any other algebra A in the variety $varC$ is extendable to a homomorphism $F_m \to A$. Therefore there exists a homomorphism $\phi : F_m \to S$ which is identical on X. The image of this homomorphism is a subalgebra of S. It contains the set of generators X, hence it contains all elements of S. Therefore S is an image of F_m, so the number of elements in S does not exceed n (the number of elements in F_m).

$2 \to 1$. Suppose that for every m we have found a number n such that the order of any m-generated subalgebra of any algebra in C does not exceed n. We have to prove that every finitely generated algebra in $varC$ is finite. Let S be an m-generated algebra in $varC$. By formula 4 S is a homomorphic image of a subalgebra T of a direct product $\prod_i A_i$ of algebras from C.

Suppose S is infinite. Then T is also infinite. Notice that we may assume that T is also m-generated. Indeed, we can take a pre-image of each generator or S in T, and generate a subalgebra by these pre-images; S will be a homomorphic image of this subalgebra.

The direct product $\prod_i A_i$ consists of vectors with coordinates from A_i. The projection π_i onto A_i is a homomorphism. $\pi_i(T)$ is generated by m elements (images of generators of T). Therefore the order of $\pi_i(T)$ does not exceed some number $n = n(m)$ (all algebras A_i are from C).

There exists only finitely many algebras of order $\leq n$ (we have only finitely many operations, each of them is defined by the "multiplication table" and there are only finitely many "multiplication tables").

Thus there exists only finitely many images $\pi_i(T)$.

For every finite algebra A there exists only finitely many homomorphisms from T to A. Indeed, each homomorphism is determined by the images of generators of T, we have finitely many generators and A is also finite.

Therefore the number of different kernels of homomorphisms π_i in T is finite. Recall that the kernel of a homomorphism is the partition of T which glues together elements which go to the same place under this homomorphism. Each of these partitions has only finitely many classes. Since T is infinite, there exists two different elements t_1 and t_2 in T which belong to the same class of each of these kernels. Thus $\pi_i(t_1) = \pi_i(t_2)$ for every i. Therefore these two vectors have the same coordinates. This means that they are equal, a contradiction. The theorem is proved.

Now we are ready to find non-finitely based varieties of semigroups.

Let us take all subwords of Z_∞ and construct semigroup $S(Z_\infty)$ using the Dilworth construction.

Ex. 38 *The semigroup $S(Z_\infty)$ satisfies condition 2 of Theorem 5.2. Estimate the function $n(m)$.*

Theorem 5.3 *The variety $varS(Z_\infty)$ is not finitely based.*

Proof. Indeed, by Theorem 5.1 and by the previous exercise it is enough to show that $S(Z_\infty)$ does not satisfy any identity of the form $Z_n = W$. By Lemma 4.13 it is enough to show that for every n some subwords of Z_∞ are not isoterms for $Z_n = W$. But Z_n itself is a subword of Z_∞, and Z_n is, of course, not an isoterm for $Z_n = W$. The theorem is proved.

Now we will give an example of a finite semigroup which generates a non-finitely based variety.

Let B_2^1 be the semigroup of all 2×2-matrix units, 0, and 1:

$$\left\{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 10 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right\}.$$

Then the following theorem holds.

Theorem 5.4 *Every finite semigroup S such that B_2^1 is a homomorphic image of a subsemigroup of S generates a non-finitely based variety.*

Proof. Indeed by Theorem 5.2 every finitely generated semigroup of $varS$ is finite, so the first condition of Theorem 5.1 holds.

Suppose that the second condition does not hold, that is S satisfies an identity $Z_n = W$. Then every subsemigroup of S satisfies this identity, and so does every homomorphic image of every subsemigroup of S. Therefore B_2^1 must satisfy this identity.

Since W differs from Z_n, by Lemma 4.8, W satisfies one of four conditions listed in this lemma. Suppose W contains x_1^2 or $x_i x_j$ where $i, j \neq 1$ or W starts or ends not with x_1. Let $x_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $x_i = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ for all $i \neq 1$. Then it is easy to see that the value of W is not equal to the value of Z_n (which is $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$). This contradicts the fact that $Z_n = W$ is an identity of B_2^1.

If $[W = [W_1, x_1]_1^1$ then $W_1 \neq (Z_n)_{x_1}$. Now let $x_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and all other x_i be arbitrary elements of B_2^1. Since B_2^1 satisfies $Z_n = W$, we will have that $(Z_n)_{x_1} = W_1$ holds for arbitrary choice of $x_i \in B_2^1$. Therefore B_2^1 satisfies the nontrivial identity $(Z_n)_{x_1} = W_1$. But this identity is of the form $Z'_{n-1} = W'$ and we can finish the proof by induction on n. The theorem is proved.

It is easy to see that the multiplicative semigroup of all matrices of order > 1 over arbitrary finite field or ring with unit contains B_2^1. Thus it generates a non-finitely based variety. So does the semigroup of all transformations of a more than 2-element set.

It is interesting that if we consider the set of all matrices over a finite ring as a ring, then it generates a finitely based variety (as any finite associative ring) - by a theorem of L'vov and Kruse.

To complete the picture I'd like to mention some other facts on varieties generated by finite algebras.

- Every finite group generates a finitely based variety (Oates and Powell, 1966).
- Every finite Lie algebra generates a finitely based variety (Bakhturin-Ol'shansky, 1982).
- Every semigroup with less than 6 elements generates a finitely based variety (Trakhtman, 1988).
- Every finite commutative semigroup generates a finitely based variety.
- Every finite nil-semigroup generates a finitely based variety.
- There exist finite semigroups which do not have B_2^1 as a homomorphic image of a subsemigroup, but which generate non-finitely based varieties.
- The description of finite semigroups which generate finitely based varieties is not known. It is not known if the set of these semigroups is recursive.
- Recently McKenzie solved a well known Tarski problem by showing that the class of finite algebras with one binary operation which generate finitely based varieties is not recursive. Therefore there is no algorithm which recognizes if a finite universal algebra generates a finitely based variety.

Ex. 39 *Find finite bases of identities of varieties generated by the following semigroups*

1. $\{a, b\}$ *with the following operation:* $ab = a^2 = a,\ ba = b^2 = b$.
2. $\{a, b\}$ *with the following operation:* $a^2 = a,\ ab = ba = b^2 = b$.
3. $\{a, b, 0\}$ *with the following operation:* $a^2 = a,\ ab = b$, *all other products are equal to* 0.

Hint. In order to find a basis of identities of $varS$ do the following:

1) find some identities Σ of S,

2) describe canonical words in the verbal congruence $\sigma(\Sigma)$ (as we deed in the Exercise 25).

3) Check if two canonical words u and v form an identity of S. If "no" the process stops and Σ is a basis of identities of S. If "yes" then add $u = v$ to Σ and go to step 2.

6 Burnside-type Problems in Associative Algebras with Identities

6.1 Introduction

The Burnside type problem for associative algebras was formulated by Kurosh in 30s:

> Suppose all 1-generated subalgebras of a finitely generated associative algebra A are finite dimensional. Is A finite dimensional?

The answer is negative (Golod) but not as easy as in the case of semigroups. For example in the semigroup case we saw that there exists a 3-generated infinite semigroup satisfying the identity $x^2 = 0$. In the case of associative algebras over a field of characteristic ≥ 3 this is impossible. Indeed, the following theorem holds.

Theorem 6.1 *Every algebra over a field of characteristic ≥ 3 or 0 which satisfies the identity $x^2 = 0$ is nilpotent of class 3, that is every product of 3 elements in this algebra is 0. Every nilpotent finitely generated algebra is finite dimensional.*

Proof. Indeed, let A satisfy the identity $x^2 = 0$. Then for every x and y in A we have $(x + y)^2 = 0$. Let us open parentheses: $x^2 + y^2 + xy + yx = 0$. We know that $x^2 = y^2 = 0$, so we get $xy + yx = 0$. Let us multiply the last identity by x on the right: $xyx + yx^2 = 0$. Since $yx^2 = 0$, we have $xyx = 0$. Now take $x = z + t$ where z and t are arbitrary elements of A. We get $(z+t)y(z+t) = 0$. Again let us open parentheses: $zyz + tyt + zyt + tyz = 0$.

Since we already know that $zyz = tyt = 0$, we have $zyt + tyz = 0$. Now we can use the identity $xy + yx = 0$ which has been established above. Using this identity, we get: $zyt = -yzt = -(yz)t = tyz$, so $tyz + tyz = 2tyz = 0$. Since the characteristic is > 2 or 0 we get $tyz = 0$. Recall that t, y, z were arbitrary elements of A. Thus A is nilpotent of class 3.

Let $A =< X >$ be a finitely generated nilpotent of class k algebra. Then every element of A is a linear combination of products of no more than $k-1$ elements from X. Since X is finite there are finitely many products of length $< k$ of elements of X. Thus A is spanned as a vector space by a finite set of elements, so A is finite dimensional. The theorem is proved.

We will see later that in this theorem one can replace 2 by any other natural number n. This is the so called Dubnov-Ivanov-Nagata-Higman theorem but the proof for arbitrary n is more complicated.

Our next goal is to prove that the answer to Kurosh's question is positive if the algebra satisfies any non-trivial identity. This result was obtained by Kaplansky and Shirshov.

First of all let us clarify what does it mean that every 1-generated subalgebra of an algebra A is finite dimensional.

Lemma 6.1 *Let A be an associative K-algebra, $a \in A$. The subalgebra generated by a is finite dimensional iff a is a root of an equation $f(x) = 0$ where $f(x)$ is a polynomial with coefficients from the field K and with the leading coefficient 1.*

Proof. Indeed, if a is a root of such an equation and f is a polynomial of degree n then every power of a is a linear combination of a^m with $m < n$. Since the subalgebra generated by a is spanned by the powers of a, it is spanned by finitely many powers of a, and so it is finite dimensional.

Conversely, if $< a >$ is finite dimensional then it is spanned by finitely many powers of a, so there exists a natural number m such that a^m is a linear combination of smaller powers of a. Thus a satisfies the equality $a^m + \sum_{i<m} \alpha_i a^i = 0$ for some $\alpha_i \in K$. Thus a is a root of the equation $x^m + \sum_{i<m} \alpha_i x^i = 0$. The lemma is proved.

An algebra where every element is a root of a polynomial with the leading coefficient 1 is called *algebraic*. By Lemma 6.1 every finite dimensional algebra is algebraic (every subspase of a finite dimensional space is finite dimensional).

Now let us turn to identities of associative algebras and prove the following fact.

Lemma 6.2 *Every n-dimensional algebra satisfies the following standard identity $\mathcal{S}_{n+1} = 0$ of degree $n+1$ where $\mathcal{S}_{n+1}$ is the following polynomial:*

$$\sum_{\sigma \in S_{n+1}} (-1)^\sigma x_{\sigma(1)} x_{\sigma(2)} ... x_{\sigma(n+1)} = 0. \tag{5}$$

Here S_{n+1} is the symmetric group on the first $n+1$ numbers, $(-1)^\sigma$ is the oddness of σ: it is 1 if σ is even and -1 if σ is odd.

In particular every 2-dimensional algebra satisfies the identity $xyz - xzy - yxz + yzx + zxy - zyx = 0$.

Proof. First of all notice that $\mathcal{S}_{n+1}$ is a *multilinear* polynomial, that is

$$\mathcal{S}_{n+1}(x_1, \ldots, x_{i-1}, \alpha x + \beta y, x_{i+1}, \ldots, x_{n+1}) =$$

$$\alpha \mathcal{S}_{n+1}(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_{n+1}) +$$
$$\beta \mathcal{S}_{n+1}(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_{n+1})$$

for arbitrary i. Equivalently, every letter occurs in every monomial in $\mathcal{S}_{n+1}$ exactly once.

Now let us take an n-dimensional algebra A spanned by a set X. Every element of A is a linear combination of elements from X. Therefore if we want to prove that $\mathcal{S}_{n+1}$ is identically equal to 0 in A, we have to substitute linear combinations of elements of X for x_i and prove that the result of this substitution is 0. The fact that our identity is multilinear allows us to take elements of X instead of these linear combinations (indeed, the sum of zeroes is zero).

Now let us take $t_1, \ldots, t_{n+1} \in X$. We have to prove that

$$\mathcal{S}_{n+1}(t_1, \ldots, t_{n+1}) = 0.$$

Since X contains n elements at least two of t_i are equal. Let $t_i = t_j$. Let us divide all permutations from S_{n+1} into pairs. Two permutations belong to the same pair if one of them can be obtained from another one by switching i and j. One of the permutations in each pair is even and another one is odd, so the terms of $\mathcal{S}_{n+1}$ corresponding to these permutations have opposite

signs. The "absolute values" of these terms are equal since $t_i = t_j$. Thus these terms cancel. Since every term belongs to one of these pairs, all terms will cancel, and the sum will be equal to 0. The lemma is proved.

In particular the algebra of all $m \times m$-matrices satisfies the identity $\mathcal{S}_{n+1} = 0$ where $n = m^2$. Indeed this algebra is spanned by m^2 matrix units.

Thus every finite dimensional algebra is finitely generated, algebraic, and satisfies a non-trivial identity.

The following theorem states that the converse statement also holds.

Theorem 6.2 *(Kaplansky). Every finitely generated algebraic algebra which satisfies a non-trivial identity is finite dimensional.*

This will follow from Shirshov's height theorem stated and proved in the next section.

6.2 Shirshov's Height Theorem

Theorem 6.3 *Let $A =< X >$ be a finitely generated algebra satisfying a non-trivial identity of degree n. Then there exists a finite sequence $v_1, \ldots, v_s$ of not necessarily different words of length $\leq n$ from X^+ such that every element of A is a linear combination of products $v_1^{m_1} v_2^{m_2} \cdots v_s^{m_s}$.*

Ex. 40 *Show that Theorem 6.3 implies Theorem 6.2.*

Let us prove the Shirshov theorem.

First of all let us show that instead of arbitrary identities we can consider only multilinear identities (we have seen that multilinear identities are very convenient).

Lemma 6.3 *Every nontrivial identity implies a multilinear identity of the same or smaller degree.*

Proof. The idea of the proof is similar to one used in the proof of Theorem 6.1. There we proceeded from a non-linear identity $x^2 = 0$ to a multilinear identity $xy + yx = 0$.

Let $f(x_1, \ldots, x_n) = 0$ be a non-multilinear identity. We can represent f as a sum of *normal* components f_i such that all terms (monomials) in f_i have the same content and different f_i have different contents. Then we can take a

component f_i with a minimal (under the inclusion) content. Substitute 0 for variables which are not in the content of f_i. This makes all other components 0. Thus if an algebra satisfies the identity $f = 0$ then it satisfies the identity $f_i = 0$. Therefore we can assume that f is normal.

Suppose that x_j appears in some monomial of f at least twice and the x_j-degree d_j of f is maximal. Consider the polynomial $f' = f(x'_1 + x''_1, x_2, \ldots, x_n) - f(x'_1, x_2, \ldots, x_n) - f(x''_1, x_2, \ldots, x_n)$. This polynomial is identically 0 in every algebra which satisfies the identity $f = 0$, it has the same degree as f and its x'_1-degree and x''_1-degree are smaller than d_j. Thus the number of variables of degree d_j in the polynomial f' is smaller than the same number for f. Continuing this process (which is called the process of *linearization*) we shall finally get a normal polynomial where every variable has degree 1, that is a multilinear polynomial. The lemma is proved.

Thus we can consider only multilinear identities. Let $f = 0$ be a multilinear identity. Then every monomial of f is a product of variables in some order (each variable occurs exactly once in each monomial). Hence all monomials are permutations of the monomial $x_1 x_2 \cdots x_n$. Therefore every multilinear identity has the following form:

$$\sum_{\sigma \in S_n} \alpha_\sigma x_{\sigma(1)} \cdots x_{\sigma(n)} = 0.$$

This identity can be easily transformed to the form

$$\begin{array}{l} x_1 x_2 \cdots x_n = \\ \sum_{\sigma \in S_n \setminus \{1\}} \beta_\sigma x_{\sigma(1)} \cdots x_{\sigma(n)} \end{array} \qquad (6)$$

by renaming variables and dividing this identity by a non-zero coefficient.

Thus if an algebra A satisfies the identity (6) then every product $u_1 u_2 \cdots u_n$ of elements of A is a linear combination of permutations of this product. Theorem 6.3 claims that every element of A is a linear combination of elements of a special kind. Thus we could use an induction if we had an order on the elements of A, and if we proved that every element of A which is not of this kind is a linear combination of smaller elements. This is the main idea of the proof.

Now we are going to make this idea work. Let $A =< X >$ be a finitely generated algebra satisfying an identity (6).

Definition. We fix any order on X and consider the lexicographic order on X^+. It can be defined by the following rule: $u < v$ iff $u = p u_1 q,\ v = p v_1 r$ where p, q, r are (possibly empty) words, $u_1, v_1 \in X$ and $u_1 \leq v_1$.

For example, if $a < b < c \in X$ then $aabccc < aaca$. Indeed, in this case $p = aaa,\ u_1 = b,\ q = ccc,\ v_1 = c,\ r = a$.

Notice that by this definition we cannot compare words u and v if one of these words is an initial segment of the other word. Every two words of the same length are comparable.

One can easily see that every chain of words $w_1 >_l w_2 >_l ...$ of the same content is finite. Thus we can use an induction on the lexicographic order.

Definition. We call a word u *n-divisible* if it can be represented as a product of n words $u_1, \ldots, u_n$ and the word u is greater in the lexicographic order (denoted by $>_l$) than any product

$$u_{\sigma(1)} \cdots u_{\sigma(n)}$$

where $\sigma \in S_n \backslash \{1\}$.

It is clear that if a word has an n-divisible subword and we can apply a multilinear identity of degree n, then this word is a linear combination of smaller words of the same content. So we can proceed by induction. Thus it would be enough to show that if a word over X does not contain n-divisible subwords then it is a short product of powers of short words.

The following lemma is the first step in this direction. It was proved by Shirshov. We present here a different proof of it. This is our variation of the de Luca-Varricchio proof. This is not the shortest proof. Neither it is a very constructive proof. But it employs Zimin words Z_n in a clever way, so it shows an interesting connection between Shirshov's theorem and the things which we discussed above.

Lemma 6.4 *For every natural numbers p, n there exists a number $N(p, n)$ such that every word of length $N(p, n)$ either contains the pth power of some word or contains an n-divisible subword.*

Suppose that for some p, n there is no $N(p, n)$. We will use the fact that Zimin's words Z_h are unavoidable and so every sufficiently long word contains a value of Z_h. We will need some properties of Z_h and their values.

For every h and $i \leq h$ let $w_i = Z_{h-i} x_{h-i+1}$, $w'_i = x_{h-i+1} Z_{h-i}$. Then we have $Z_h = w_1 w_2 \cdots w_h = w'_h w'_{h-1} \cdots w'_1$. For every substitution ϕ we will call the representation of $\phi(Z_h)$ in the form $\phi(w_1)\phi(w_2)\cdots\phi(w_h)$ *the first partition of* $\phi(Z_h)$ and the representation $\phi(w'_h)\cdots\phi(w'_1)$ the second partition of $\phi(Z_h)$.

We will find n-divisions among the first or the second partitions of the values of Z_n.

Lemma 6.5

$$\phi(w_i)\phi(w_{i+1}) >_l \phi(w_{i+1})\phi(w_i) \text{ iff}$$
$$\phi(w'_i)\phi(w'_{i+1}) >_l \phi(w'_{i+1})\phi(w'_i).$$

Ex. 41 *Prove this lemma.*

Lemma 6.6 *If*

$$\phi(w_i)\phi(w_{i+1}) >_l \phi(w_{i+1})\phi(w_i)$$

for arbitrary i, $1 \leq i \leq n-1$ *then*

$$\phi(w_i)\phi(w_j) >_l \phi(w_j)\phi(w_i)$$

for arbitrary $1 \leq i < j \leq n$.

A similar statement holds if we replace $>_l$ *by* $<_l$ *or consider* w'_i *instead of* w_i.

Ex. 42 *Prove this lemma.*

Now suppose that we found a substitution ϕ such that

$$\phi(w_i)\phi(w_{i+1}) >_l \phi(w_{i+1})\phi(w_i)$$

for arbitrary i, $1 \leq i \leq n-1$. Then by Lemma 6.6

$$\phi(w_i)\phi(w_j) >_l \phi(w_j)\phi(w_i)$$

for arbitrary $1 \leq i < j \leq n$. It is easy to see by induction that, starting with the identity permutation, one can obtain an arbitrary permutation from S_n by a series of steps such that at each step one switches two consecutive i and j, $i < j$. If we replace $\phi(w_i)\phi(w_j)$ by $\phi(w_j)\phi(w_i)$ $(i < j)$ then we make the word smaller. Thus in this case the first partition of $\phi(Z_n)$, $\phi(w_1)\phi(w_2)\cdots\phi(w_n)$, is an n-division of $\phi(Z_n)$.

Suppose now that for some substitution ϕ we have that $\phi(w_i)\phi(w_{i+1}) <_l \phi(w_{i+1})\phi(w_i)$ for arbitrary i, $1 \leq i \leq n-1$. Then by Lemma 6.5 we have

that $\phi(w'_i)\phi(w'_{i+1}) <_l \phi(w'_{i+1})\phi(w'_i)$. Then by Lemma 6.6 $\phi(w'_i)\phi(w'_j) <_l \phi(w'_j)\phi(w'_i)$ for arbitrary $1 \leq i < j \leq n$. Therefore the second partition of $\phi(Z_n)$, $w'_n w'_{n-1} \cdots w'_1$, is an n-division.

There is, of course, the third possibility that $\phi(w_i)\phi(w_{i+1}) = \phi(w_{i+1})\phi(w_i)$ for all i. Then we can apply the following lemma which can be easily proved by induction on $|u| + |v|$.

Lemma 6.7 *Let $u, v \in X^+$. If $uv = vu$ then there exists a word p such that $u = p^\ell, v = p^m$ for some ℓ, m.*

Ex. 43 *Using Lemma 6.7 prove that if we have a sequence of words $u_1, \ldots, u_n$ and $u_i u_{i+1} = u_{i+1} u_i$ for all i from 1 to $n-1$ then all u_i are powers of the same word p.*

Thus if the third possibility occurs then all $\phi(w_i)$ are powers of the same word p, and so $\phi(Z_n)$ contains an nth power (at least).

This discussion shows that we have to find a value of Z_n such that one of the three possibilities occurs. In order to do this we need one more property of Z_n.

Lemma 6.8 *Let $1 < j_1 < j_2 < \ldots < j_k < n$ be arbitrary sequence of numbers. Consider $k-1$ words:*

$$u_1 = w_{j_1} \cdots w_{j_2 - 1},$$

$$u_2 = w_{j_2} \cdots w_{j_3 - 1},$$

$$\ldots$$

$$u_{k-1} = w_{j_{k-1}} \cdots w_{j_k}.$$

Then $u_1 u_2 \ldots u_{k-1}$ is the first partition of some value of Z_{k-1}.

Ex. 44 *Prove this statement.*

Thus if we take a very big h then every value of Z_h contains many values of Z_k (if $k << h$). This is a typical "Ramsey situation". We will use the following Ramsey theorem.

Theorem 6.4 *For arbitrary numbers p, k, n there exists a number $R = R(p, k, n)$ such that for every partition of the set of k-element subsets of an R-element set S into n classes there exists a p element subset T of S such that all k-element subsets of T belong to the same class of the partition.*

Let us consider the number $h = R(n + 1, 3, 3)$ and an arbitrary value $\phi(Z_h)$. Let us say that a 3-element subset $\{i, j, l\}$ of $\{1, \ldots, h\}$ $(i < j < l)$ is of type 1 if

$$(w_i \cdots w_{j-1})(w_j \cdots w_{l-1}) >_l (w_j \cdots w_{l-1})(w_i \cdots w_{j-1});$$

of type 2 if

$$(w_i \cdots w_{j-1})(w_j \cdots w_{l-1}) <_l (w_j \cdots w_{l-1})(w_i \cdots w_{j-1});$$

of type 3 if

$$(w_i \cdots w_{j-1})(w_j \cdots w_{l-1}) = (w_j \cdots w_{l-1})(w_i \cdots w_{j-1}).$$

By the Ramsey theorem there exists a subset $T = \{j_1, \ldots, j_{n+1}\}$ of the set $\{1, 2, \ldots, h\}$ such that all triples from this subset are of the same type. Given this subset we can construct words $u_1, \ldots, u_n$ as in Lemma 6.8. By this lemma these words form the first partition of some value of Z_n. By the choice of the subset T, for this value of Z_n one of the three possibilities discussed above occurs.

Now it is easy to choose the number $N(p, n)$ from Lemma 6.4. Let m be the maximum of p and n, let $h = R(m + 1, 3, 3)$. Then $N(p, n)$ may be taken equal to the minimal number t such that every word from X^+ of length t contains a value of Z_h. Such a number exists because the word Z_h is unavoidable. Lemma 6.4 is proved.

From Lemma 6.4, one can deduce by the lexicographic induction that if an algebra $A =< X >$ satisfies a multilinear identity of degree n then every sufficiently long word w over X is equal in this algebra to a linear combination of words containing big powers. This is not exactly what we need. For example the bases of these powers can depend on w, so we do not have finitely many bases as required by Theorem 6.3. This obstacle is taken care of by the following lemma.

Lemma 6.9 *Let u be a word of length $> n$. Then u^{2n} contains an n-divisible subword.*

Proof. Let u_1, u_2,..., u_n be some cyclic permutations of u, that is for every $i = 1, \ldots, n+1$ we have $u = v_i w_i$, $u_i = w_i v_i$. Words u_i have the same length, so they are lexicographically comparable. Let

$$u_{j_1} >_l u_{j_1} >_l \ldots >_l u_{j_n}.$$

Then we can represent u^{2n} in the following form

$$u^{2n} = v_{j_1} w_{j_1} v_{j_1} w_{j_1} v_{j_2} w_{j_2} v_{j_2} w_{j_2} \cdots v_{j_n} w_{j_n} v_{j_n} w_{j_n}.$$

It is clear that the subword

$$(w_{j_1} v_{j_1} w_{j_1} v_{j_2})(w_{j_2} v_{j_2} w_{j_2} v_{j_3}) \cdots (w_{j_{n-1}} v_{j_{n-1}} w_{j_{n-1}} v_{j_n} (w_{j_n} v_{j_n})$$

is n-divisible. The lemma is proved.

From this lemma, it follows that every word over X is a linear combination of words containing big powers of words of length $\leq n$. Thus the number of bases is bounded (there are finitely many words of length $\leq n$). But this is still not quite what we need because we do not have a bound for the number of powers of a given base u in w. This bound is provided by the following lemma. If w is very long, does not contain an n-divisible subwords, and cannot be expressed as a short product of powers of words of length $\leq n$ then w contains many subwords of the form $u^n v$ where v is different from u and has the same length as u, There are finitely many such words, so if a word w is very long it must contain n occurrences of the same word $u^n v$ of this form. The following lemma states that this is impossible.

Lemma 6.10 *Let w be a word containing n different occurrences of a word $u^n v$ where $|v| = |u|$ and $v \neq u$. Then w contains an n-divisible subword.*

Proof. Indeed, by the condition of the lemma $w = p u^n v p_1 u^n v p_2 \cdots u^n v p_n$. Since v has the same length as u, we can compare u and v. Suppose $u > v$. Then consider the following subword of w:

$$(u^n v p_1 u)(u^{n-1} v p_2 u^2)(u^{n-2} v p_3 u^3) \cdots (v p_n)$$

It is easy to see that $u^i v p_{n-i+1} u^{n+1-i} > u^j v p_{n+1-j} u^{n+1-j}$ for every $i > j$. Since arbitrary permutation can be obtained from the trivial permutation by

several switchings of consecutive i, j such that $i < j$ (we have already used this argument), we have that this subword is n-divisible.

Suppose $u < v$. Then we can apply a similar argument to the following subword of w:

$$(vp_1u^{n-1})(uvp_2u^{n-2})(u^2vp_3u^{n-3})\cdots(u^nv).$$

The lemma is proved.

Now we know that every word w which does not contain n-divisible subword may be represented in the following form: $p_0u_1^{k_1}p_1u_2^{k_2}p_2\ldots u_m^{k_m}p_m$ where the number m is bounded by a function in n, the lengths of u_i do not exceed n, and the lengths of p_i are bounded by a function of n. This immediately implies Shirshov's height theorem.

A much more constructive proof of this theorem was given recently by Alexei Belov.

6.3 The Dubnov-Ivanov-Nagata-Higman Theorem

Theorem 6.5 *Every algebra over a field of characteristic $> n$ which satisfies the identity $x^n = 0$ is nilpotent of class $2^n - 1$.*

We have proved this theorem in the case $n = 2$. Notice that the restriction on the characteristic is important. For example one can consider the algebra $\mathbf{Z}_p[X]$ of polynomials in infinitely many variables over $\mathbf{Z}_p$ (p is a prime number), and factorize this algebra by the ideal generated by all polynomials x^p where $x \in X$. The factor algebra satisfies the identity $z^p = 0$ (use the fact that the binomial coefficients $\binom{p}{n}$ are divisible by p), but is not nilpotent: the product $x_1\ldots x_n$ is not equal to 0 for any n $(x_i \in X)$.

We will need one nice observation concerning identities. Suppose an algebra A satisfies an identity $f(x_1,\ldots,x_n) = 0$. We can represent this identity in the form $f_0 + f_1 + \ldots + f_m = 0$ where every monomial from f_i contains exactly i occurrences of x_1. Suppose our field has at least $m+1$ different elements $\alpha_0,\ldots,\alpha_m$. Let us substitute $x_1 \to \alpha_i x_1$ for $i = 1, 2, \ldots, n$. Then we get:

$$\begin{cases} \alpha_0^0 f_0 + \alpha_0^1 f_1 + \ldots + \alpha_0^m f_m = 0 \\ \alpha_1^0 f_0 + \alpha_1^1 f_1 + \ldots + \alpha_1^m f_m = 0 \\ \ldots\ldots \\ \alpha_m^0 f_0 + \alpha_m^1 f_1 + \ldots + \alpha_m^m f_m = 0 \end{cases}$$

This is a system of linear equations with coefficients α_i^j and unknowns f_i. It has the famous Vandermonde determinant which is equal to $\prod_{i<j}(\alpha_i - \alpha_j) \neq 0$. Therefore every f_i is identically equal to 0 on A. Therefore the identity $f = 0$ implies all identities $f_i = 0$, $i = 0, \ldots, m$.

Our identity $x^n = 0$ implies the identity $(x+y)^n = 0$. Let $f(x,y) = (x+y)^n$. Then the identity

$$f_1(x,y) = \sum_{i=0}^{n-1} y^i x y^{n-1-i} = 0$$

also follows from $x^n = 0$. Here we used the fact that the characteristic of the field is $p > n$, and so our field contains at least $n+1$ elements.

Now consider the following sum:

$$\sum_{i,j=1}^{n-1} x^{n-1-i} z y^j x^i y^{n-1-j} =$$

$$\sum_{j=1}^{n-1} (\sum_{i=1}^{n-1} x^{n-1-i} (z y^j) x^i) y^{n-1-j} =$$

$$-\sum_{j=1}^{n-1} x^{n-1} z y^j y^{n-1-j} = -(n-1) x^{n-1} z y^{n-1}.$$

On the other hand

$$\sum_{i,j=1}^{n-1} x^{n-1-i} z y^j x^i y^{n-1-j} =$$

$$\sum_{i=1}^{n-1} x^{n-1-i} z (\sum_{j=1}^{n-1} y^j x^i y^{n-1-j}) =$$

$$-\sum_{i=1}^{n-1} (x^{n-1-i} z x^i) y^{n-1} = x^{n-1} z y^{n-1}.$$

Thus $n x^{n-1} z y^{n-1} = 0$. Since the characteristic is $> n$ we have that $x^{n-1} z y^{n-1} = 0$.

Now let us consider an algebra A satisfying $x^n = 0$. Suppose (by induction) that arbitrary algebra which satisfies the identity $x^{n-1} = 0$ is nilpotent of class $2^{n-1} - 1$. Consider the ideal I of A generated by all powers x^{n-1}.

This ideal is spanned by all products of the form $px^{n-1}q$ for some $p, q \in A$. Since we have the identity $x^{n-1}zy^{n-1} = 0$ we have that $IAI = 0$ (the product of every three elements a, b, c where $a, c \in I,\ b \in A$ is 0). By the induction hypothesis we have that A/I is nilpotent of class $2^{n-1} - 1$. Therefore the product of every $2^{n-1} - 1$ elements of A belongs to I. Hence the product of every $(2^{n-1} - 1) + 1 + (2^{n-1} - 1) = 2^n - 1$ elements of A may be represented as a product of three elements a, b, c where $a, c \in I,\ b \in A$. We have proved that this product is 0, so A is nilpotent of class $2^n - 1$. This completes the proof of the Dubnov-Ivanov-Nagata-Higman theorem.

Notice that the estimate $2^n - 1$ is not optimal. Razmyslov proved that the upper bound is n^2 while Kuzmin proved that the lower bound is $\frac{n(n-1)}{2}$.

6.4 Golod Counterexamples to the Kurosh Problem

We have proved that every finitely generated associative algebra satisfying the identity $x^n = 0$ is finite dimensional. It turns out however that if we allow the exponent n to depend on x then the situation is quite different.

Recall that an algebra A is called a *nil-algebra* if for every $x \in A$ there exists a number n_x such that $x^{n_x} = 0$. The Kurosh problem asks if all finitely generated nil-algebras are nilpotent. In 1964 Golod, using a method of Golod and Shafarevich, constructed a counterexample. The method of Golod and Shafarevich came from a study of some commutative algebra/number theory problems.

Theorem 6.6 *For arbitrary countable field K there exists a finitely generated infinite dimensional nil-algebra over K.*

Let K be a countable field and X a finite set. Consider the free algebra $F = KX^+$.

Every algebra A is a factor-algebra of F over some ideal I. Thus we have to construct an ideal I of F such that

1. F/I is a nil-algebra,

2. F/I is not finite dimensional.

In order to achieve the first goal we have to have the following: for every element $p \in F$ there must be a number n_p such that $p^{n_p} \in I$. In order to

achieve the second goal we want I to be as small as possible (the smaller I we take, the bigger F/I we get). Thus we may suppose that I is generated by all these p^{n_p}. Therefore our problem may be rewritten in the following form.

> List all elements from F (recall that K is countable, X^+ is also countable, so $F = KX^+$ is also countable):
>
> $$p_1, p_2, \ldots .$$
>
> For every $i = 1, 2, \ldots$ choose a natural number n_i. Let I be the ideal generated by $p_i^{n_i}$. The question is: how to make a choice of n_i in order to obtain an infinite dimensional factor-algebra F/I.

To solve this problem we first of all will make it a bit harder.

Every polynomial p in F is a sum of *homogeneous components* $p_1 + \ldots + p_m$ where for every i all monomials in p_i have the same length i which is called the *degree* of p_i. Notice that this decomposition differs from the decomposition considered in the previous section.

Let $|X| = d \geq 3$. Then there exist exactly d^n words over X of length n. Thus for every n the space F_n of homogeneous polynomials of degree n from F has dimension d^n (it is spanned by the words of length n).

It is easy to see that if an ideal I is generated by a set R of homogeneous polynomials then it is spanned by homogeneous polynomials. Indeed by the definition I is spanned by X^*RX^*. Each polynomial in X^*RX^* is homogeneous. Therefore in this case $I = I \cap F_1 \oplus I \cap F_2 \oplus \ldots$. Thus I has a nice decomposition into a sum of finite dimensional subspaces (a subspace of a finite dimensional space is finite dimensional). This shows that ideals generated by homogeneous polynomials are easier to study than arbitrary ideals.

Unfortunately the ideal I generated by $p_i^{n_i}$ won't be generated by homogeneous polynomials. Hence we shall make a sacrifice. Let us generate I not by $p_i^{n_i}$ but by all homogeneous components of $p_i^{n_i}$. For example if $p_i = x + y^2$ and we choose $n_i = 2$ then $p_i^2 = x^2 + y^4 + xy^2 + yx^2$, the homogeneous components will be x^2 (degree 2), $xy^2 + y^2x$ (degree 3), y^4 (degree 4). We will put all these components into I. It is clear that the ideal generated by these components is bigger than the ideal generated by $p_i^{n_i}$, so the factor-algebra will be a nil-algebra also.

Let R be the space spanned by the homogeneous components of $p_i^{n_i}$, I be the ideal generated by I, let X' be the space spanned by X.

Now let us introduce the key concept of the proof of the Golod-Shafarevich theorem. This concept has become one of the key concepts in algebra as a whole.

With every subspace S of F spanned by homogeneous polynomials we associate the following Hilbert series:

$$H_S = \sum_{i=1}^{\infty} h_n t^n$$

where h_n is the dimension of $S \cap \mathcal{F}_n$, t is the unknown. So H_S is a formal series in one unknown. For example

$$H_{X'} = dt,\ H_F = \sum_n d^n t^n = dt/(1-dt). \tag{7}$$

It turns out that the properties of F/I are closely related to the properties of the Hilbert series of R.

Since $I = (I \cap F_1) \oplus (I \cap F_2) \oplus \ldots$, F/I is isomorphic as a vector space to the sum of complements $I^c = (I \cap F_1)^c \oplus (I \cap F_2)^c \oplus \ldots$ where $(I \cap F_j)^c \oplus (I \cap F_j) = F_j$. Thus F/I is finite dimensional if and only if I contains all F_j starting with some F_n. Therefore F/I is finite dimensional if and only if $h_j = 0$ for $j \geq n$ (for some n). In other words F/I is finite dimensional if and only if H_{I^c} is a polynomial. Therefore by a clever choice of n_i we have to make H_{I^c} not a polynomial.

Notice that we can choose n_i as we want. We can, for example, suppose that R does not have polynomials of degree 1 (it is enough to take all $n_i \geq 2$). We also can choose n_i in such a way that all coefficients in H_R are either 0 or 1: just choose n_i big enough so that all homogeneous components of $p_i^{n_i}$ have greater degrees than homogeneous components of $p_j^{n_j}$ for $j < i$.

Now we need some properties of the Hilbert series. We say that $H_S \leq H_T$ if coefficients of H_S do not exceed the corresponding coefficients of H_T.

H1. If $S = T + U$ (a sum of two subspaces) then $H_S \leq H_T + H_U$. If this is a direct sum then $H_S = H_T + H_U$. This follows from the fact that the degree of the sum of two subspaces does not exceed the sum of degrees of these subspaces.

H2. If $S = TU$ (recall that TU consists of linear combinations of products tu where $t \in T,\ u \in U$), then $H_S \leq H_T H_U$. This can be proved by a direct calculation or by using tensor products of subspaces (TU is a homomorphic image of the tensor product $T \otimes U$).

H3. If H, H', H'' are formal power series with nonnegative coefficients and $H \leq H'$ then $HH'' \leq H'H''$. This condition is obvious.

Now let us use these properties. Since I is generated by R, we have that

$$I = R + FR + RF + FRF.$$

Since $I \oplus I^c = F$ we have that

$$I = R + I^c R + IR + FRF.$$

Since $FR \subseteq I$ we have

$$I = R + I^c R + IR + IF.$$

Since $IR \subseteq IF$ we have

$$I = R + I^c R + IF.$$

Now since I is an ideal, for every word $w \in X^+$ we have $Iw \subseteq Ix$ where x is the last letter of w. Therefore $IF = IX'$, so

$$I = R + I^c R + IX. \tag{8}$$

The equality $F = I \oplus I^c$ and property H1 imply $H_I = H_F - H_{I^c}$. Applying H1 and H2 to (8) we obtain:

$$H_F - H_{I^c} \leq H_R + H_{I^c} H_R + (H_F - H_{I^c}) H_{X'}.$$

Open parentheses and move everything to the right:

$$0 \leq H_{I^c} + H_{I^c} H_R - H_{I^c} H_{X'} + H_R - H_F + H_F H_{X'}.$$

Therefore

$$0 \leq H_{I^c}(1 + H_R - H_{X'}) + H_R + H_F H_{X'} - H_F.$$

Recall that $H_{X'} = dt,\ H_F = \sum_n d^n t^n$ (see (7). Therefore $H_F H_{X'} - H_F = -dt$. Hence we have:

$$0 \leq H_{I^c}(1 + H_R - dt) + H_R - dt.$$

Add 1 to both sides of this equality:

$$1 \leq (H_{I^c} + 1)(1 - dt + H_R).$$

Let $P = H_{I^c} + 1$. We can conclude that $1 \leq P(1 - dt + H_R)$. It is obvious that P is a polynomial if and only if H_{I^c} is a polynomial. Notice also that P has non-negative coefficients because H_{I^c} is a Hilbert serie.

Now let us use the fact that every coefficients of H_R is 0 or 1. and the first two coefficients are 0. This implies that

$$1 - dt + H_R \leq 1 - dt + \sum_{n=2}^{\infty} t^n.$$

For the computational reasons it is convenient to notice that since $d \geq 3$, we have $\frac{(d-1)^2}{2} \geq 1$ and so

$$1 - dt + H_R \leq 1 - dt + \frac{(d-1)^2}{2} \sum_{n=2}^{\infty} t^n.$$

It is easy to calculate that

$$1 - dt + \frac{(d-1)^2}{2} \sum_{n=2}^{\infty} t^n = 1 - dt + \frac{(d-1)^2}{2} \frac{t^2}{1-t} = \frac{(\frac{d+1}{2}t - 1)^2}{1-t}.$$

Since coefficients of P are non-negative

$$1 \leq P(\frac{(\frac{d+1}{2}t - 1)^2}{1-t}).$$

Suppose we have proved that the serie $W = \frac{1-t}{(\frac{d+1}{2}t-1)^2}$ has positive coefficients. Then we could multiply the inequality $1 \leq P(\frac{(\frac{d+1}{2}t-1)^2}{1-t})$ by W and get (by virtue of condition H3) $W \leq P$. But a polynomial cannot be greater than or equal to a power serie with positive coefficients. Thus we would prove that P is not a polynomial and coplete the proof of our theorem.

It remains to show that all coefficients of W are positive.

We have already used the fact that $\sum_{n=1}^{\infty} t^n = \frac{1}{1-t}$. Now if we take the derivative of both sides of these equality, we get

$$\frac{1}{(1-t)^2} = \sum_{n=0}^{\infty} (n+1)t^n.$$

Applying this formula we obtain:

$$W = (1-t)\sum_{n=0}^{\infty} (n+1)(\frac{d+1}{2}t)^n.$$

The nth coefficient of the last serie is:

$$(n+1)(\frac{d+1}{2})^n - n(\frac{d+1}{2})^{n-1} > 0$$

for $n \geq 1$ and 1 for $n = 0$. Thus, indeed, all coefficients of W are positive.

This completes the proof of Theorem 6.6.

6.5 A Counterexample to the Unbounded Burnside Problem For Groups

Let K be a countable field of characteristic $p > 0$. Consider the algebra $A = F/I$ constructed in the previous section. Let B be the algebra A with a unit element adjoint. Consider the semigroup G generated by all elements $1+x$ where $x \in X$. Every element of this semigroup has the form $1+a$ where $a \in A$. Since A is a nil-algebra $a^{p^n} = 0$ in A for some n which depends on a. Since the characteristic is equal to p we have that $(1+a)^{p^n} = 1 + a^{p^n} = 1$. Therefore G is a group (every element $1+a$ has an inverse $(1+a)^{p^n-1}$). By definition G is finitely generated. If G were finite then KG would be a finite dimensional algebra. The algebra KG contains 1 and all elements $1 + x, \quad x \in X$. So KG contains all the generators x of A. Since KG is a subalgebra of A we can conclude that $KG = A$. But A is not finite dimensional and KG is finite dimensional, a contradiction. Therefore G is an infinite finitely generated periodic group. This is a counterexample to the unbounded Burnside problem for groups.

Recall that Burnside formulated his problem in 1902, and Golod solved it in 1964. It is amazing how simple may be a solution of a 60 years old problem.

Now there exist many other methods of constructing counterexamples to the unbounded Burnside problem for groups (Aleshin, Grigorchuk, etc.). It is interesting to note that there still exist only one method of constructing counterexamples to the Kurosh problem, the Golod-Shafarevich method.

7 Test Problems

Ex. 45 *Is the following substitution*

$$a \to ab$$
$$b \to a$$

a) cube free;
b) 4th power free;
c) nth power free for some n?

Ex. 46 *Is the following substitution*

$$a \to aba$$
$$b \to a$$

a) cube free;
b) 4th power free;
c) nth power free for some n?

Ex. 47 *Prove or disprove that there exists a finitely generated infinite semigroup satisfying the identities* $x^2 = 0$ *and*

$$xx_1yx_2xx_1xx_3xx_1xx_2yx_1x = xx_1xx_2yx_1xx_3xx_1yx_2xx_1x.$$

Ex. 48 *Prove or disprove that there exists a finitely generated infinite semigroup satisfying the identities* $x^2 = 0$ *and*

$$xx_1yx_2xx_1xx_3xx_1xx_2yx_1x = yx_1xx_2xx_1xx_3xx_1yx_2xx_1x.$$

Ex. 49 *Describe in terms of canonical words the verbal congruence on the free semigroup defined by the identity* $xyx = xy^2$.

Ex. 50 *Describe in terms of canonical words the verbal congruence defined by the identity* $xyx = yxy$.

Ex. 51 *Find a finite basis of identities of the semigroup* $S = \{1, a, 0\}; 1 \cdot 1 = 1, 1 \cdot a = a \cdot 1 = a$ *all other products are equal to* 0.

Ex. 52 *Find a finite basis of identities of the semigroup*

$$S = \{a, b, c, 0\}; a^2 = a, b^2 = b, cb = ac = c$$

all other products are equal to 0.

Ex. 53 *Prove that the algebra of* 2×2*-matrices over any field satisfies the identity* $[[x, y]^2, z] = 0$ *where* $[a, b]$ *stands for* $ab - ba$. *Prove that the algebra of* 2×2*-matrices over a field of characteristic 0 does not satisfy any identity of degree 3.*

Ex. 54 *Prove that the algebra of upper triangular* $n \times n$*-matrices over a field of characteristic 0 satisfies the identity* $[x, y]^n = 0$ *where* $[a, b]$ *stands for* $ab - ba$. *Prove that this algebra does not satisfy any identity of degree* $< n$.