Арифметические свойства биномиальных коэффициентов

На конференции Вам будет предложено несколько исследовательских проектов. Цель — как можно дальше продвинуться в каком-то из проектов. Задачи можно решать коллективно, объединившись в любые команды (члены команды могут быть из разных городов). Вы можете решать задачи сразу из нескольких проектов, причем по разным проектам Вы можете участвовать в разных командах. Единственное, чего не следует делать, — это присваивать себе чужие результаты, такое случается, если команда слишком велика и не все из нее активно решают задачи данного проекта.

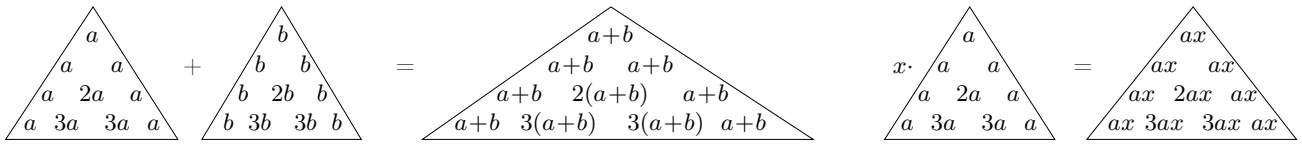
Это ознакомительная подборка задач по теме о биномиальных коэффициентах. Задачи следует решать письменно и сдавать Кохасю К.П. (вагон 15, место 17). В Теберде набор задач будет существенно расширен и все задачи, кроме задачи 1.2, можно будет сдавать и позже. По задаче 1.2 решения принимаются только в поезде, после этого задача снимается с конкурса.

1 Задачи в поезд

1.1. Докажите, что а) $C_{p-1}^k \equiv (-1)^k \pmod p$; б) $C_{2n}^n \equiv (-4)^n C_{\frac{p-1}{2}}^n \pmod p$ при $n \leqslant \frac{p-1}{2}$.

1.2. Докажите, что количество нечетных биномиальных коэффициентов в n-й строке треугольника Паскаля равно 2^r, где r — количество единиц в двоичной записи числа n.

1.3. Зафиксируем натуральное число m. Назовем m-арифметическим треугольником Паскаля треугольник, в котором вместо чисел C_n^k расставлены их остатки по модулю m. Кроме того, мы будем рассматривать похожие треугольники из остатков, у которых вдоль боковых сторон вместо единиц стоят одинаковые остатки a по модулю m. Такие треугольники можно умножать на число, а также складывать (если размеры совпадают), причем будем считать, что операции тоже выполняются по модулю m.



Пусть в s-й строке m-арифметического треугольника Паскаля все элементы, кроме крайних, — нули. Докажите, что тогда этот треугольник имеет вид, показанный на рис. 1. Заштрихованные треугольники состоят из нулей, а треугольники Δ_n^k состоят из s строк и подчинены следующим соотношениям:

1) $\Delta_n^{k-1} + \Delta_n^k = \Delta_{n+1}^k$; 2) $\Delta_n^k = C_n^k \cdot \Delta_0^0 \pmod m$.

Головоломка Ханойская башня представляет собой три стержня, на которые надеваются диски разной величины. Вначале все диски упорядочены по размеру (более крупные — ниже) и находятся на первом стержне. Разрешается снять со стержня один верхний диск и переместить его на другой стержень. При этом запрещается более крупный диск класть на диск меньшего размера. В головоломке требуется переложить все диски с первого стержня на второй.

Пусть количество дисков равно n. Рассмотрим граф TH_n, вершины которого — это всевозможные расположения дисков Ханойской башни, а ребра соединяют те состояния головоломки, которые получаются друг из друга за один ход. Рассмотрим также граф P_n, вершины которого — это единицы, расположенные в первых 2^n строках 2-арифметического треугольника Паскаля, а ребра соединяют соседние единицы (т.е. соседние в строке или в двух смежных строках по диагонали).

1.4. Докажите, что графы TH_n и P_n изоморфны.

1.5. Докажите, что в первых 10^6 строках 2-арифметического треугольника Паскаля единицы составляют меньше 1 %.

1.6. Докажите, что если n делится на $p-1$, то $C_n^{p-1} + C_n^{2(p-1)} + C_n^{3(p-1)} + \ldots + C_n^n \equiv 1 \pmod p$. Или лучше докажите в общем виде: если $1 \leqslant j, k \leqslant p-1$ и $n \equiv k \pmod{p-1}$, то

$$C_n^j + C_n^{(p-1)+j} + C_n^{2(p-1)+j} + C_n^{3(p-1)+j} + \ldots \equiv C_k^j \pmod p.$$

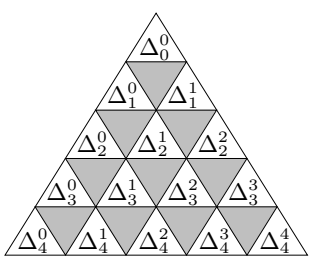


Рис. 1.

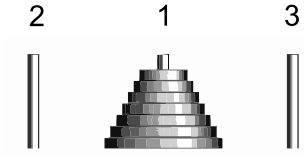


Рис. 2.

Арифметические свойства биномиальных коэффициентов — 2

Официальным "теоретическим материалом" для этого цикла задач служит статья Э. Б. Винберга [1]. В частности, считаются известными следующие теоремы.

1. Теорема Вильсона. Для всех простых p (и только для простых) выполнено сравнение $(p-1)! \equiv -1 \pmod p$.
2. Теорема Люка. Запишем числа n и k в системе счисления по основанию p:
$$n = n_d p^d + n_{d-1}p^{d-1} + \ldots + n_1 p + n_0, \qquad k = k_d p^d + k_{d-1}p^{d-1} + \ldots + k_1 p + k_0. \tag{1}$$
Тогда $C_n^k \equiv C_{n_d}^{k_d} C_{n_{d-1}}^{k_{d-1}} \cdot \ldots \cdot C_{n_1}^{k_1} C_{n_0}^{k_0} \pmod p$.
3. Теорема Куммера. Показатель $\operatorname{ord}_p C_n^k$ равен числу переносов при сложении столбиком чисел k и $\ell = n-k$ в p-ичной записи.
4. Теорема Волстенхолма. При $p \geqslant 5 \quad C_{2p}^p \equiv 2 \pmod{p^3}$ или, что то же самое, $C_{2p-1}^{p-1} \equiv 1 \pmod{p^3}$.

Напомним, что по определению $C_0^0 = 1$, $C_n^k = 0$ при $k > n$ и при $k < 0$.

Всюду буквой p мы обозначаем простое число. Для произвольного натурального числа n обозначим через $(n!)_p$ произведение всех натуральных чисел от 1 до n, не делящихся на p. Если задано число p, то символами n_i, m_i и т. д. обозначаются цифры p-ичной записи чисел n, m и т. д.

* * *

2 Арифметический треугольник и делимость

2.1. a) Докажите, что в первых 3^k строках 3-арифметического треугольника Паскаля содержится $\frac{1}{2}(6^k + 4^k)$ единиц и $\frac{1}{2}(6^k - 4^k)$ двоек.

b) Найдите число нулевых элементов в первых 5^k строках 5-арифметического треугольника Паскаля.

c) Найдите число ненулевых элементов в первых p^k строках p-арифметического треугольника Паскаля.

2.2. Докажите, что количество единиц в первых m строках 2-арифметического треугольника Паскаля равно
$$\sum_{i=0}^{n-1} m_i \cdot 2^{\sum_{k=i+1}^{n-1} m_k} \cdot 3^i.$$
Полагая $m = 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_r}$, где $\alpha_1 > \alpha_2 > \ldots > \alpha_r$, можно то же выражение записать в виде
$$3^{\alpha_1} + 2 \cdot 3^{\alpha_2} + 2^2 \cdot 3^{\alpha_3} + \ldots + 2^{r-1} \cdot 3^{\alpha_r}.$$

2.3. Рассмотрим n-ю строку 2-арифметического треугольника Паскаля как двоичную запись некоторого натурального числа P_n. Докажите, что
$$P_n = F_{i_1} \cdot \ldots \cdot F_{i_s},$$
где $i_1, \ldots, i_s$ — номера разрядов, в которых в двоичной записи числа n стоят единицы, и $F_i = 2^{2^i} + 1$ — i-е число Ферма.

2.4. Докажите, что количество ненулевых элементов в n-й строке p-арифметического треугольника Паскаля равно $\prod_{i=0}^{d} (n_i + 1)$.

2.5. a) Для того чтобы все биномиальные коэффициенты C_n^k, где $0 < k < n$, делились на p, необходимо и достаточно, чтобы n было степенью числа p.

b) Для того чтобы все биномиальные коэффициенты C_n^k, где $0 \leqslant k \leqslant n$, не делились на p, необходимо и достаточно, чтобы $n+1$ делилось на p^d, иными словами, чтобы все цифры p-ичной записи числа n, кроме старшей, были равны $p-1$.

2.6. Пусть $0 < k < n+1$. Докажите, что если $C_n^{k-1} \not\vdots\, p$ и $C_n^k \not\vdots\, p$, то $C_{n+1}^k \not\vdots\, p$, *кроме* случая, когда $n+1$ делится на p.

3 Обобщение теорем Вильсона и Люка

3.1. Докажите, что $\operatorname{ord}_p(n!) = \dfrac{n - (n_d + \ldots + n_1 + n_0)}{p-1}$.

3.2. Докажите следующие обобщения теоремы Вильсона. a) $(-1)^{[n/p]}(n!)_p \equiv n_0! \pmod p$;
b) При $p \geqslant 3$ выполнено сравнение

$$(p^q!)_p \equiv -1 \pmod{p^q},$$

а при $p = 2$, $q \geqslant 3$ выполнено сравнение $(p^q!)_p \equiv 1 \pmod{p^q}$.
c) $\dfrac{n!}{p^\mu} \equiv (-1)^\mu n_0! n_1! \ldots n_d! \pmod p$, где $\mu = \operatorname{ord}_p(n!)$

3.3. Обобщенная теорема Люка. Пусть $r = n - k$, $\ell = \operatorname{ord}_p(C_n^k)$. Тогда

$$\frac{1}{p^\ell} C_n^k \equiv (-1)^\ell \Big(\frac{n_0!}{k_0! r_0!}\Big)\Big(\frac{n_1!}{k_1! r_1!}\Big) \ldots \Big(\frac{n_d!}{k_d! r_d!}\Big) \pmod p$$

3.4. a) Докажите, что $(1+x)^{p^d} \equiv 1 + x^{p^d} \pmod p$ при всех $x = 0, 1, \ldots, p-1$.
b) Докажите теорему Люка алгебраически.

3.5. a) Пусть m, n, k — натуральные числа, причем $(n, k) = 1$. Докажите, что $C_{mn}^k \equiv 0 \pmod n$.
b) Если $n \vdots p^k$, $m \not\vdots p$, то $C_n^m \vdots p^k$.

3.6. Пусть $f_{n,a} = \sum\limits_{k=0}^{n} (C_n^k)^a$. Докажите, что $f_{n,a} \equiv \prod\limits_{i=0}^{d} f_{n_i,a} \pmod p$.

4 Вариации на тему теоремы Волстенхолма

4.1. Докажите, что $\dfrac{1}{1} + \dfrac{1}{2} + \ldots + \dfrac{1}{p-1} \equiv 0 \pmod{p^2}$ при $p \geqslant 5$.

4.2. Пусть $p = 4k + 3$ — простое число. Найдите $\dfrac{1}{0^2+1} + \dfrac{1}{1^2+1} + \ldots + \dfrac{1}{(p-1)^2+1} \pmod p$.

4.3. a) Пусть натуральное число k таково, что для каждого простого делителя p числа m $k \not\vdots (p-1)$. Докажите, что

$$\frac{1}{1^k} + \frac{1}{2^k} + \ldots + \frac{1}{(m-1)^k} \equiv 0 \pmod m.$$

Здесь суммирование распространяется на все слагаемые, знаменатели которых взаимно просты с m.
b) Пусть k нечетно и $(k+1) \not\vdots (p-1)$. Докажите, что $\dfrac{1}{1^k} + \dfrac{1}{2^k} + \ldots + \dfrac{1}{(p-1)^k} \equiv 0 \pmod{p^2}$.

4.4. Докажите, что сравнение (12) из статьи Винберга выполнено по модулю p^4.

4.5. Докажите эквивалентность следующих сравнений. 1) $C_{2p-1}^{p-1} \equiv 1 \pmod{p^4}$;
2) $\dfrac{1}{1} + \dfrac{1}{2} + \ldots + \dfrac{1}{p-1} \equiv 0 \pmod{p^3}$; 3) $\dfrac{1}{1^2} + \dfrac{1}{2^2} + \ldots + \dfrac{1}{(p-1)^2} \equiv 0 \pmod{p^2}$.

4.6. a) Докажите алгебраически, что для всякого простого p и произвольных k и n $(C_{pk}^{pm} - C_k^m) \vdots p^2$. В статье Винберга этот факт доказан комбинаторно.
b) Докажите утверждение (9) из статьи Винберга: для всякого простого $p \geqslant 5$ и произвольных k и n $(C_{pk}^{pm} - C_k^m) \vdots p^3$.

4.7. Пусть $p \geqslant 5$. Докажите, что a) $C_{p^2}^p \equiv C_p^1 \pmod{p^5}$; b) $C_{p^{s+1}}^p \equiv p^s \pmod{p^{2s+3}}$.

4.8. Докажите, что $C_{p^3}^{p^2} \equiv C_{p^2}^p \pmod{p^8}$.

Арифметические свойства биномиальных коэффициентов — 3

Дополнения к предыдущим темам

2.7. Докажите, что $C^k_{p^n-1} \equiv (-1)^{S_k} \pmod p$, где S_k — сумма цифр p-ичной записи числа k.

2.8. Докажите, что если биномиальный коэффициент C^k_n нечетен, (т. е. в обозначениях из (1) $k_i \leqslant n_i$ при всех $i = 0, \ldots, d$), то

$$C^k_n \equiv \prod_{i=1}^{d} (-1)^{k_{i-1}n_i + k_i n_{i-1}} \pmod 4 .$$

2.9. Докажите, что если в двоичной записи числа n нет двух единиц подряд, то все нечетные числа в n-й строке треугольника Паскаля сравнимы с 1 по модулю 4, а в противном случае ровно половина из них сравнима с 1 по модулю 4.

2.10. Докажите, что количество пятерок в каждой строке 8-арифметического треугольника Паскаля равно степени двойки. То же касается единиц, троек и семерок.

2.11. Докажите, что если все элементы двух множеств

$$\{C^1_{2^n-1}, C^3_{2^n-1}, C^5_{2^n-1}, \ldots, C^{2^n-1}_{2^n-1}\} \quad \text{и} \quad \{1, 3, 5, \ldots, 2^n - 1\}$$

рассматривать как остатки по модулю 2^n, то эти множества совпадают.

2.12. Докажите, что элементы одной строки треугольника Паскаля не взаимно просты в следующем довольно сильном смысле. Для каждого числа $\varepsilon > 0$ существует N, такое, что при всех натуральных $n > N$ и $k_1, k_2, \ldots, k_{100} < \varepsilon\sqrt{n}$ верно, что числа

$$C^{n+k_1}_{2n},\ C^{n+k_2}_{2n},\ \ldots,\ C^{n+k_{100}}_{2n}$$

имеют общий делитель.

2.13. а) Даны натуральные числа $m > 1$, n, k. Докажите, что хотя бы одно из чисел C^k_n, C^k_{n+1}, …, C^k_{n+k} не делится на m.

b) Докажите, что для любого k найдется бесконечно много таких n, что все числа C^k_n, C^k_{n+1}, …, C^k_{n+k-1} делятся на m.

4.9. Докажите, что при $n > 1$ $C^{2^n}_{2^{n+1}} - C^{2^{n-1}}_{2^n}$ делится на 2^{2n+2}.

4.10. Докажите, что при $p \geqslant 5$ $(-1)^{\frac{p-1}{2}} C^{\frac{p-1}{2}}_{p-1} \equiv 4^{p-1} \pmod{p^3}$.

Арифметические свойства биномиальных коэффициентов — 4

Дополнения к предыдущим темам

4.11. Пусть m — произвольное натуральное число, $p \geqslant 5$ — простое. Докажите, что

$$\frac{1}{mp+1} + \frac{1}{mp+2} + \cdots + \frac{1}{mp+(p-1)} \equiv 0 \pmod{p^2}.$$

4.12. Пусть p и q — различные простые числа. Докажите, что сравнение $C_{2pq-1}^{pq-1} \equiv 1 \pmod{pq}$ выполнено в том и только в том случае, когда $C_{2p-1}^{p-1} \equiv 1 \pmod{q}$ и $C_{2q-1}^{q-1} \equiv 1 \pmod{p}$.

5 Суммы биномиальных коэффициентов

5.1. a) Докажите, что $\sum\limits_{k=0}^{3^a-1} C_{2k}^k$ делится на 3; b) делится на 3^a.

5.2. Пусть $C_k = \frac{1}{k+1}C_{2k}^k$ — последовательность чисел Каталана. Докажите, что $\sum\limits_{k=1}^{n} C_k \equiv 1 \pmod 3$ тогда и только тогда, когда троичное разложение числа $n+1$ содержит хотя бы одну цифру 2.

5.3. Пусть $p \geqslant 5$, $k = [2p/3]$. Докажите, что сумма $C_p^1 + C_p^2 + \ldots + C_p^k$ делится на p^2.

5.4. Если $n \vdots (p-1)$, где p — нечетное простое, то

$$C_n^{p-1} + C_n^{2(p-1)} + C_n^{3(p-1)} + \ldots \equiv 1 + p(n+1) \pmod{p^2}.$$

5.5. Докажите, что при $0 \leqslant j \leqslant p-1 < n$ и $q = [\frac{n-1}{p-1}]$

$$\sum_{m:m\equiv j \pmod p} (-1)^m C_n^m \equiv 0 \pmod{p^q}.$$

5.6. Докажите, что если p — нечетное простое, то $n \vdots (p+1)$ тогда и только тогда, когда

$$C_n^j - C_n^{j+(p-1)} + C_n^{j+2(p-1)} - C_n^{j+3(p-1)} + \ldots \equiv 0 \pmod p$$

при всех $j = 1, 3, \ldots, p-2$.

Решения

1 Задачи в поезд

1.1. а) Решение 1. $C_{p-1}^k = \dfrac{(p-1)(p-2)\dots(p-k)}{1\cdot 2\cdots k} \equiv \dfrac{(-1)(-2)\dots(-k)}{1\cdot 2\cdots k} \equiv (-1)^k \pmod p$.

Решение 2. По формуле для биномиальных коэффициентов очевидно, что C_p^i при $1 \leqslant i \leqslant p-1$ делится на p. Кроме того, имеет место основное рекуррентное соотношение $C_{p-1}^{k-1} + C_{p-1}^k = C_p^k$. Так как $C_{p-1}^0 = 1 \equiv 1 \pmod p$ и $(C_{p-1}^0 + C_{p-1}^1) \vdots p$, заключаем отсюда, что $C_{p-1}^1 \equiv -1 \pmod p$. Но $C_{p-1}^1 + C_{p-1}^2$ тоже делится на p, значит, $C_{p-1}^2 \equiv 1 \pmod p$ и т.д.

б) Это задача [3, задача 162]. Поскольку дроби C_{2n+2}^{n+1}/C_{2n}^n и $C_{\frac{p-1}{2}}^{n+1}/C_{\frac{p-1}{2}}^n$ сильно сократимы, утверждение легко проверяется по индукции. Но мы предложим прямое вычисление из [3].

Как нетрудно видеть,

$$C_{2n}^n = 2^n \cdot \frac{1\cdot 3\cdots(2n-1)}{n!}$$

При этом

$$1\cdot 3\cdots(2n-1) = (-1)^n(-1)(-3)\cdots(-2n+1) \equiv (-1)^n(p-1)(p-3)\cdots(p-2n+1) =$$

$$= (-1)^n 2^n \Big(\frac{p-1}{2}\Big)\Big(\frac{p-3}{2}\Big)\cdots\Big(\frac{p-2n+1}{2}\Big) = (-1)^n 2^n \Big(\frac{p-1}{2}\Big)\Big(\frac{p-1}{2}-1\Big)\cdots\Big(\frac{p-1}{2}-n+1\Big) =$$

$$= (-1)^n 2^n \frac{(\frac{p-1}{2})!}{(\frac{p-1}{2}-n)!} \pmod p.$$

Таким образом, $C_{2n}^n \equiv (-1)^n 4^n \dfrac{(\frac{p-1}{2})!}{n!(\frac{p-1}{2}-n)!} = (-4)^n C_{\frac{p-1}{2}}^n \pmod p$.

1.2. Это непосредственно следует из самоподобной структуры арифметического треугольника Паскаля, описанной в следующих задачах. Это также сразу следует из теоремы Люка. Доказательство можно прочесть в статье Винберга [1].

1.3. Мы ограничимся небольшим созерцанием, полное решение см. в [3, задача 133].

Поскольку в s-й строке расположен длинный ряд из нулей, в $(s+1)$-й строке под этими нулями также расположен ряд из нулей (на единицу короче), в $(s+2)$-й строке — опять ряд из нулей (снова на 1 короче) и т. д. Этим объясняется наличие серого треугольника снизу от Δ_0^0 (рис. 1).

Далее, ненулевые элементы s-й строки равны 1, тогда ряды чисел, идущих вдоль наклонных границ серого треугольника, состоящего из нулей, — это тоже всё сплошь единицы (по рекуррентному правилу построения треугольника Паскаля). Таким образом, вдоль боковых сторон треугольников Δ_1^0 и Δ_1^1 расположены единицы, и значит, оба этих треугольника идентичны Δ_0^0.

Теперь понятно, как выглядит $2s$-я строка треугольника. Крайние элементы в ней — единицы, остальные элементы — нули, кроме центрального элемента, который равен 2, как сумма двух вышестоящих единиц. Отсюда получаем, что снизу от $2s$-й строки находятся два серых нулевых треугольника, по краям от них — треугольники Δ_2^0 и Δ_2^2, идентичные Δ_0^0, а между ними — треугольник Δ_2^1, у которого вдоль боковых сторон расположены двойки. Как нетрудно понять, это значит, что $\Delta_2^1 = 2\cdot\Delta_0^0$.

Ну и так далее.

1.4. Этот сюжет мы взяли в статье [21], где некоторые факты о биномиальных коэффициентов доказываются с помощью рассмотрения Ханойской башни и графа TH_n.

Пусть на первом стержне самый верхний диск имеет диаметр a, на втором — диаметр b, на третьем — c, $a < b < c$, тогда в этом положении есть три возможных хода: с a на b или на c, либо с b на c; аналогично имеется три хода, если диски занимают лишь два стержня. Если же все диски находятся на одном стержне, возможных ходов только два, обозначим такие конфигурации A_1, A_2, A_3 по номеру стержня, на который нанизаны диски.

Заметим, что 2^s-я строка треугольника Паскаля состоит из одних единиц — это следует из задачи 1.2 или проверяется непосредственно с помощью формулы Лежандра (4). Отсюда следует, что граф P_n имеет поворотную симметрию третьего порядка, поскольку основное соотношение

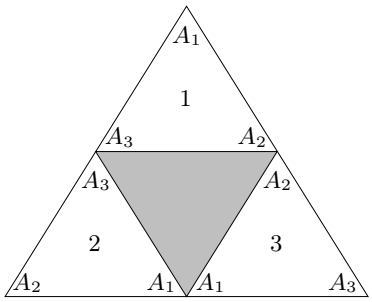


Рис. 3.

$C_n^{k-1} + C_n^k = C_{n+1}^k$, при помощи которого мы строим треугольник Паскаля "сверху вниз", в арифметике по модулю 2 равносильно соотношениям $C_n^{k-1} = C_n^k + C_{n+1}^k$ и $C_n^k = C_n^{k-1} + C_{n+1}^k$, с помощью которых можно аналогично построить треугольник Паскаля "слева снизу — вправо вверх" и "справа снизу — влево вверх". Кроме того, отсюда следует (из предыдущей задачи), что треугольник Паскаля в 2 раза большего размера содержит три копии исходного треугольника.

Докажем по индукции, что существует биекция между TH_n и P_n, при которой вершинам треугольника P_n соответствуют конфигурации A_1, A_2, A_3. База $n = 1$ очевидна.

Докажем переход. Пусть мы уже умеем строить биекцию между TH_n и P_n. Рассмотрим 2-арифметический треугольник треугольник Паскаля со стороной 2^{n+1}, он содержит три копии треугольника со стороной 2^n. Пронумеруем копии и разметим их вершины, как показано на рис. 3. Рассмотрим все положения Ханойской башни, в которых самый крупный диск находится на стержне i. Если мы не двигаем этот диск, то все эти положения и перекладывания остальных дисков задают граф, изоморфный TP_n. С помощью уже имеющейся биекции отождествим этот граф с графом P_n, расположенным в i-й копии треугольника, причем потребуем, чтобы конфигурации A_j были отождествлены в соответствии с разметкой вершин. Перемещение самого крупного диска, скажем, с первого стержня на второй возможно, только если все остальные диски находятся на третьем стержне. Это в точности соответствует ребру, соединяющему соседние вершины A_3 на левой боковой стороне треугольника, аналогично обстоят дела с другими перемещениями самого большого диска. Таким образом, построенное соответствие действительно дает изоморфизм графов TP_{n+1} и P_n.

1.5. Биекция с Ханойской башней дает простую явную формулу (когда число строк – степень двойки): в первых 2^k строках треугольника Паскаля содержится 3^k единиц. Та же формула легко доказывается по индукции из рекуррентности задачи 1.3. Пользуясь этим фактом легко получаем оценку. Так как $10^6 < 2^{20}$, количество элементов в этих строках равно $\frac{1}{2} \cdot 10^6(10^6 + 1)$, а количество единиц не превосходит 3^{20}. Доля единиц не превосходит $\frac{2 \cdot 3^{20}}{10^6(10^6+1)} \ll 0.01$.

1.6. Мы взяли это утверждение в обзоре [18].

Р е ш е н и е 1 ([CSTTVZ]). При $p = 2$ утверждение задачи легко проверяется. Будем далее считать, что p — нечетное простое. Пусть $n = x(p-1) + k$. Будем доказывать утверждение индукцией по x.

База $x = 0$ тривиальна: $C_k^j \equiv C_k^j \pmod p$.

Для доказательства перехода воспользуемся свойством биномиальных коэффициентов

$$C_{a+b}^s = \sum_i C_a^{s-i} C_b^i \qquad \text{(суммирование в естественных границах)},$$

которое выражает два способа подсчета числа вариантов взять s шаров из коробки, в которой лежит a черных и b белых шаров. Пусть $n = m + (p-1)$. Заметим, что

$$C_n^{\ell(p-1)+j} = C_{m+(p-1)}^{\ell(p-1)+j} = \sum_{i=0}^{p-1} C_m^{\ell(p-1)+j-i} C_{p-1}^i \equiv \sum_{i=0}^{p-1} (-1)^i C_m^{\ell(p-1)+j-i} \pmod p$$

(последнее сравнение — по утверждению задачи 1.1 а). Отметим, что в последней сумме первое и

последнее слагаемое присутствуют со знаком плюс. Преобразуем теперь интересующую нас сумму.

$$\sum_\ell C_n^{\ell(p-1)+j} \equiv$$
$$\equiv \left(C_m^j - C_m^{j-1} + \ldots\right) + \left(C_m^{p-1+j} - C_m^{p-1+j-1} + \ldots + C_m^j\right) + \left(C_m^{2(p-1)+j} - C_m^{2(p-1)+j-1} + \ldots + C_m^{(p-1)+j}\right) + \ldots$$
$$= \sum_{i=0}^{m} (-1)^i C_m^i + \sum_\ell C_m^{\ell(p-1)+j} \pmod p.$$

Здесь первая сумма равна нулю, а вторая по предположению индукции сравнима с $C_k^j \pmod p$. ЧТД

Решение 2 (основное рекуррентное тождество, [J], [T]). Утверждение доказывается индукцией по n. База $n \leqslant p-1$ тривиальна: левая часть содержит всего одно слагаемое — то же самое, что и в правой части. Переход:

$$C_n^j + C_n^{(p-1)+j} + \ldots = \left(C_{n-1}^j + C_{n-1}^{j-1}\right) + \left(C_{n-1}^{(p-1)+j} + C_{n-1}^{(p-1)+j-1}\right) + \ldots =$$
$$= \left(C_{n-1}^j + C_{n-1}^{(p-1)+j} + \ldots\right) + \left(C_{n-1}^{j-1} + C_{n-1}^{(p-1)+j-1} + \ldots\right) \equiv C_{k-1}^j + C_{k-1}^{j-1} = C_k^j \pmod p.$$

Но тут следует иметь в виду, что в формулировке утверждения в случае, когда параметры j и k делятся на $p-1$, они приравниваются к $p-1$, а не к 0. Таким образом, выписанное соотношение требует отдельного рассмотрения при $j = 1$ или $k = 1$. Мы ограничимся рассмотрением частного случая, которое проясняет ситуацию. Пусть $p = 5$, $j = 1$ и мы доказываем переход к $n = 13$. Имеем

$$C_1^1 \stackrel{?}{\equiv} C_{13}^1 + C_{13}^6 + C_{13}^{11} = \left(C_{12}^1 + C_{12}^6 + C_{12}^{11}\right) + \left(C_{12}^0 + C_{12}^5 + C_{12}^{10}\right)$$

Здесь первая скобка дает по индукционному предположению остаток C_4^1 (а вовсе не C_0^1, как могло показаться по предыдущему вычислению). Во второй скобке первое слагаемое не участвует в индукционном предположении, а сумма остальных сравнима с C_4^0. Записывая для ясности $p-1$ вместо 4, получаем, что вся сумма сравнима с $C_{n-1}^0 + C_{p-1}^1 + C_{p-1}^0 \equiv C_1^1 \pmod p$, что и требуется.

Решение 3 (алгебраическое рассуждение с теоремой Люка, [18]). Индукция по n. База $n \leqslant p-1$ тривиальна. Пусть теперь $n \geqslant p$, запишем все встречающиеся параметры в системе счисления по основанию p, сумму цифр числа m будем обозначать $\sigma_p(m)$. Очевидно, если $m \equiv j \pmod p$, то $\sigma_p(m) \equiv j \pmod p$. Тогда по теореме Люка интересующая нас сумма равна

$$\sum C_{n_0}^{m_0} C_{n_1}^{m_1} \ldots C_{n_d}^{m_d} \pmod p,$$

где суммирование распространяется на все $m = \overline{m_d \ldots m_1 m_0} \leqslant n$, для которых $\sigma_p(m) \equiv j \pmod p$. Эта сумма в точности равна сумме коэффициентов при x^j, x^{j+p-1}, $x^{j+2(p-1)}$, ... в выражении

$$(1+x)^{n_0}(1+x)^{n_1} \ldots (1+x)^{n_d} = (1+x)^{\sigma_p(n)}.$$

Но очевидно, что указанная сумма коэффициентов равна

$$\sum_{\substack{1 \leqslant r \leqslant \sigma_p(n) \\ r \equiv j \pmod{p-1}}} C_{\sigma_p(n)}^r,$$

которая удовлетворяет индукционному предположению, так как $1 \leqslant \sigma_p(n) \leqslant n-1$, и дает нужное нам сравнение, поскольку $\sigma_p(n) \equiv n \equiv j \pmod p$.

Решение 4 (немного здравого смысла и линейной алгебры, [Д]). Многочлены x, x^2, ..., x^{p-1} линейно независимы над $\mathbb{Z}_p$ и образуют базис в пространстве функций $f\colon \mathbb{Z}_p \to \mathbb{Z}_p, f(0) = 0$. По малой теореме Ферма $(1+x)^n \equiv (1+x)^k \pmod p$. Редуцируя левую часть с помощью соотношений $x^{i+a(p-1)} \equiv x^i$, получаем, что интересующая нас сумма как элемент $\mathbb{Z}_p$ равна коэффициенту при x^j в правой части, т. е. C_k^j.

2 *Арифметический треугольник и делимость*

2.1. a) Это результат Робертса [27]. Обозначим количество единиц в первых 3^k строках через a_k, а количество двоек b_k — через b_k. Пользуясь рекуррентностью из задачи 1.3, получаем соотношения:

$$a_{k+1} = 5a_k + b_k, \qquad b_{k+1} = 5b_k + a_k.$$

Отсюда утверждение задачи легко следует по индукции.

b) Ответ: $\frac{1}{2} \cdot 5^k(5^k + 1) - 15^k$. Обозначая искомую величину a_k, аналогично предыдущей задаче получаем соотношение

$$a_{k+1} = 15a_k + 10 \cdot \frac{5^k(5^k - 1)}{2}.$$

Поскольку в целом треугольник содержит $\frac{5^k(5^k+1)}{2}$ элементов, естественно ввести замену переменных $a_k = \frac{5^k(5^k+1)}{2} - b_k$. Тогда для переменной b_k предыдущее соотношение записывается в виде $b_{k+1} = 15b_k$.

c) Ответ: $\left(\frac{p(p+1)}{2}\right)^k$. Это результат Файна [13]. Он аналогично предыдущим пунктам получается по индукции из рекуррентности задачи 1.3.

2.2. Решение 1. Индукция по α_1. База для $\alpha_1 = 0, 1$ легко проверяется. Пусть для всех $\alpha_1 < a$ утверждение уже доказано. Докажем его для $\alpha_1 = a$. Очевидно, $\tilde{m} - 2^{\alpha_1} < 2^{\alpha_1}$. Пусть в обозначениях задачи 1.3 $s = 2^{\alpha_1}$. Числу $\tilde{m} = 2^{\alpha_2} + 2^{\alpha_3} + \ldots + 2^{\alpha_r}$ соответствует строчка в треугольнике Δ_0^0. В этой строке и в строках над ней по индукционному предположению содержится

$$3^{\alpha_2} + 2 \cdot 3^{\alpha_3} + \ldots + 2^{r-2} \cdot 3^{\alpha_r} \tag{2}$$

единиц. Тогда числу $m = \tilde{m} + 2^{\alpha_1}$ соответствует строчка, пересекающая треугольники Δ_0^1 и Δ_1^1 (идентичные треугольнику Δ_0^0, поскольку у нас 2-арифметика). В этой строке и выше находится целиком треугольник Δ_0^0 (в нем по предположению индукции 3^{α_1} единиц) и два неполных треугольника Δ_0^1 и Δ_1^1, в каждом из которых число единиц задается формулой (2). В сумме получаем

$$3^{\alpha_1} + 2(3^{\alpha_2} + 2 \cdot 3^{\alpha_3} + \ldots + 2^{r-2} \cdot 3^{\alpha_r})$$

единиц, что и требуется.

Решение 2 (комбинаторный смысл коэффициентов — разбиваем на слои, [T]).

Лемма 1. Пусть число единиц в k-й строке равно 2^r (или, что то же самое, бинарная запись числа k содержит r единиц) и пусть $\alpha_1 > \alpha_2 > \cdots > \alpha_m$, $2^{\alpha_m} > k$. Тогда число единиц в строке с номером $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_m} + k$ равно 2^{m+r}.

Доказательство. Очевидно, бинарная запись числа $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_m} + k$ содержит $m + r$ единиц и тогда в строке треугольника Паскаля с этим номером 2^{m+r} единиц. □

Лемма 2. Суммарное количество единиц в строках с номерами

$$2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}}, \quad 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + 1, \quad \ldots, \quad 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + 2^{\alpha_m} - 1,$$

равно $2^k 3^{\alpha_m}$.

Доказательство. По лемме 1 количество единиц в строке с номером $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + i$ равно $2^k x_i$, где x_i — количество единиц в i-й строке. Тогда суммарное число единиц в упомянутых строках равно $2^k \sum x_i$. Но $\sum x_i$ — это число единиц в первых $2^{\alpha_m} - 1$ строках треугольника Паскаля, оно равно 3^{α_m} (это нам известно, например, из задачи 1.4). □

Осталось просуммировать по m количества единиц из леммы 2.

2.3. Мы взяли утверждение задачи из статьи Винберга [1], а решение из статьи Гранвилля [18]. Утверждение выводится из теоремы Люка с помощью следующего наблюдения (тоже упомянутого в [1]): биномиальный коэффициент C_n^k нечетен в том и только том случае, когда единицы в двоичном разложении числа k могут стоять лишь в тех разрядах, где стоят единицы в записи числа n. Отсюда

сразу следует, что $P_n = \sum 2^k$, где суммирование распространяется на все числа k, описанные в предыдущем предложении. В обозначениях формулы (1) при $p = 2$ положим $S_n = \{i : n_i = 1\}$. Тогда

$$P_n = \sum_{I \subseteq S_n} \prod_{i \in I} 2^{2^i} = \prod_{i \in S_n} F_i.$$

2.4. Этот результат Файна [13], 1947 г, — простое следствие теоремы Куммера. Чтобы биномиальный коэффициент C_n^k не делился на p, не должно быть переносов при сложении чисел k и $n-k$, записанных в системе счисления по основанию p. При фиксированном n это означает, что выбор i-й цифры p-ичной записи числа k можно сделать $n_i + 1$ способом.

2.5. a) Это сразу следует из формулы, доказанной в предыдущей задаче, поскольку речь идет о строке, в которой ровно два элемента не делятся на p.

b) [13]. Если $(n+1) \vdots p^d$, то $n = \overline{a(p-1)(p-1)\dots(p-1)}$ в системе счисления по основанию p. Тогда для каждого k, $0 \leqslant k \leqslant n$, каждая цифра числа k не превосходит соответствующей цифры числа n. Тогда все биномиальные коэффициенты $C_{n_i}^{k_i}$ не равны нулю (в том числе, по модулю p) и по теореме Люка C_n^k не делится на p.

В обратную сторону. Пусть все биномиальные коэффициенты C_n^k не делятся на p, но число n является числом вида $\overline{a(p-1)(p-1)\dots(p-1)}$. Это значит, что одна из цифр, скажем n_i, меньше $p-1$. Возьмем $k = (p-1)\cdot p^i$. Тогда $k_i = p-1$, следовательно, $C_{n_i}^{k_i} = 0$ и по теореме Люка C_n^k делится на p. Противоречие.

2.6. Это известное утверждение мы почерпнули в [12].

Решение 1. Допустим, что $C_n^{k-1} \not\vdots p$ и $C_n^k \not\vdots p$, но при этом $C_{n+1}^k = (C_n^{k-1} + C_n^k) \vdots p$. Тогда $C_n^k \equiv -C_n^{k-1} \pmod p$. Так как оба биномиальных коэффициента не делятся на p, мы можем сократить правую и левую части. Получим $\frac{n-k+1}{k} \equiv -1 \pmod p$, откуда $n+1 \equiv 0 \pmod p$.

Решение 2 ([K]). Хотя утверждение выглядит очень естественным, напоминая нам основное тождество для биномиальных коэффициентов, часть "$C_n^{k-1} \not\vdots p$" в нем лишняя. Действительно, если $(n+1) \not\vdots p$, то $0 \leqslant n_0 \leqslant p-2$. Поскольку $C_n^k \not\vdots p$, то по теореме Куммера при всех i верно неравенство $k_i \leqslant n_i$. Но тогда аналогичные неравенства верны и для пары чисел k и $n+1$, поскольку у числа $n+1$ те же цифры, что и у n, кроме цифры в самом младшем разряде, которая у числа $n+1$ на 1 больше. Следовательно, $C_{n+1}^k \not\vdots p$.

2.7. [2]. Сразу следует из теоремы Люка и задачи 1.1.a)

2.8. Задача из статьи Винберга [1]. Индукция по числу цифр. База тривиальна. Для перехода добавляем очередную цифру в конец числа. В силу нечетности биномиального коэффициента $n_i \geqslant k_i$. Пользуясь рекуррентностью $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$, перебирая разные варианты четности n и k с помощью теоремы Куммера и задачи 4.6a) сводим все к индукционному предположению.

Например, при нечетном $n = 2\ell + 1$ и четном $k = 2m$, если $k_1 = 1$, то $k = \dots 10$, $n = \dots 11$ (двоичные записи), Тогда $(n-k) = \dots 01$ (потому что по теореме Куммера не должно было быть переносов), $(k-1)_2 = \dots 01$, значит, по теореме Куммера при сложении $(k-1)_2 + (n-k)_2$ есть ровно 1 перенос, т. е. $C_{n-1}^{k-1} \equiv 2 \pmod 4$, откуда

$$C_n^k = C_{n-1}^{k-1} + C_{n-1}^k \equiv -C_{n-1}^k = -C_{2\ell}^{2m} \equiv -C_\ell^m \pmod 4,$$

последнее — по задаче 4.6a). Этот минус в точности соответствует множителю $(-1)^{k_0 n_1 + k_1 n_0}$.

2.9. Задача из статьи Винберга [1]. Утверждение следует из результата предыдущей задачи. Если в записи n нет двух единиц подряд, то все показатели $k_{i-1}n_i + k_i n_{i-1}$ равны нулю и все биномиальные коэффициенты дают остаток 1 при делении на 4. Если же запись числа n содержит участок из единиц, начинающийся с $n_j = 1$, то у половины нечетных биномиальных коэффициентов $k_j = 0$, а у другой половины $k_j = 1$ и, как нетрудно видеть по формуле из предыдущей задачи, по модулю 4 эти половины отличаются знаком.

2.10. Этому запутанному сюжету посвящены две статьи в Monthly [19, 20].

2.11. Эта задача Д.Джукича была в 2002 г. на олимпиаде 239 школы г. Санкт-Петербурга, а потом засветилась в шорт-листе IMO-2008.

Поскольку все биномиальные коэффициенты из условия задачи нечетны (по теореме Люка), для доказательства утверждения достаточно проверить, что все числа $C_{2^n-1}^1$, $C_{2^n-1}^3$, $\ldots$, $C_{2^n-1}^{2^n-1}$ дают разные остатки при делении на 2^n. Дальше можно действовать по-разному.

Р е ш е н и е 1 ([Д]). Предположим противное, пусть $C_{2^n-1}^k \equiv C_{2^n-1}^m \pmod{2^n}$ при нечетных k и m, $k > m$. Заметим, что

$$C_{2^n-1}^k = C_{2^n}^k - C_{2^n-1}^{k-1} = C_{2^n}^k - C_{2^n}^{k-1} + C_{2^n-1}^{k-2} = \cdots = C_{2^n}^k - C_{2^n}^{k-1} + C_{2^n}^{k-2} - \ldots - C_{2^n}^{m+1} + C_{2^n-1}^m\,.$$

В частности,

$$C_{2^n}^k - C_{2^n}^{k-1} + C_{2^n}^{k-2} - \ldots - C_{2^n}^{m+1} \equiv 0 \pmod{2^n}\,.$$

Теорема Куммера позволяет для каждого r легко вычислить показатель $\operatorname{ord}_2 C_{2^n}^r$, а именно, если $\operatorname{ord}_2 r = a$, то при сложении r и $2^n - r$ произойдет $n - a$ переносов (это очевидно из алгоритма сложения столбиком), и значит, $\operatorname{ord}_2 C_{2^n}^r = n - a$. В частности, $C_{2^n}^r$ делится на 2^n при нечетном r, что позволяет отбросить в последнем сравнении половину слагаемых:

$$C_{2^n}^{k-1} + C_{2^n}^{k-3} + \ldots + C_{2^n}^{m+1} \equiv 0 \pmod{2^n}\,.$$

Другое следствие из приведенных рассуждений состоит в том, что у всех слагаемых $C_{2^n}^i$ в левой части параметр i четный и поэтому $\operatorname{ord}_2 C_{2^n}^x < n$. Докажем теперь, что выполнение этого сравнения невозможно. Выберем x, для которого $\operatorname{ord}_2 C_{2^n}^x$ имеет минимальное значение. Так как $\operatorname{ord}_2 C_{2^n}^x < n$, но при этом вся сумма делится на 2^n, найдется y, для которого $\operatorname{ord}_2 C_{2^n}^x = \operatorname{ord}_2 C_{2^n}^y$. Но тогда бинарные записи чисел x и y оканчиваются на одинаковое число нулей, поэтому между x и y найдется число z, оканчивающееся на большее число нулей. Тогда $\operatorname{ord}_2 C_{2^n}^z < \operatorname{ord}_2 C_{2^n}^x$, что противоречит минимальности.

Р е ш е н и е 2 ([CSTTVZ]). Предположим противное, пусть нашлись числа k и ℓ, $k \neq \ell$, такие что $C_{2^n-1}^{2k+1} \equiv C_{2^n-1}^{2\ell+1} \mod 2^n$, $0 \leqslant k, \ell \leqslant 2^n - 1$. Кроме того, мы будем вести рассуждения по индукции, считая, что для меньших значений n утверждение задачи уже доказано. Заметим, что

$$\begin{aligned} C_{2^n-1}^{2k+1} &= \left(\frac{2^n}{1} - 1\right)\left(\frac{2^n}{2} - 1\right)\ldots\left(\frac{2^n}{2k+1} - 1\right) = \\ &= \left(\frac{2^n}{1} - 1\right)\left(\frac{2^n}{3} - 1\right)\ldots\left(\frac{2^n}{2k+1} - 1\right)\cdot\left(\frac{2^{n-1}}{1} - 1\right)\left(\frac{2^{n-1}}{2} - 1\right)\ldots\left(\frac{2^{n-1}}{k} - 1\right) = \qquad (3) \\ &= \left(\frac{2^n}{1} - 1\right)\left(\frac{2^n}{3} - 1\right)\ldots\left(\frac{2^n}{2k+1} - 1\right)\cdot C_{2^{n-1}-1}^k \equiv \\ &\equiv (-1)^{k+1} C_{2^{n-1}-1}^k \pmod{2^n}\,. \end{aligned}$$

И аналогично $C_{2^n-1}^{2\ell+1} \equiv (-1)^{\ell+1} C_{2^{n-1}-1}^{\ell} \pmod{2^n}$. По индукционному предположению, отсюда следует, что k и ℓ не могут быть оба нечетными. Кроме того, в силу симметрии $C_{2^n-1}^r = C_{2^n-1}^{2^n-1-r}$ утверждение задачи означает также, что все биномиальные коэффициенты с четными показателями — $C_{2^n-1}^{2r}$ — тоже попарно различны и по модулю 2^n образуют то же множество, что и биномиальные коэффициенты с нечетными показателями. Поэтому k и ℓ не могут быть оба четными.

Осталось разобрать случай, когда k и ℓ разной четности, пусть $k = 2a + 1$, $\ell = 2b$. Тогда

$$C_{2^{n-1}-1}^{2a+1} + C_{2^{n-1}-1}^{2b} \equiv 0 \pmod{2^n}\,.$$

При $a = b$ это сравнение невозможно, так как $C_{2^{n-1}-1}^{2a}$ нечетно и

$$C_{2^{n-1}-1}^{2a+1} + C_{2^{n-1}-1}^{2a} = C_{2^{n-1}-1}^{2a}\left(1 + \frac{2^{n-1} - 1 - 2a}{2a+1}\right) = C_{2^{n-1}-1}^{2a}\cdot\frac{2^{n-1}}{2a+1} \equiv 2^{n-1} \pmod{2^n}\,.$$

Если же $b \neq a$, то $C_{2^{n-1}-1}^{2a} \neq C_{2^{n-1}-1}^{2b}$ по индукционному предположению и так как $C_{2^{n-1}-1}^{2a} + C_{2^{n-1}-1}^{2a+1}$ делится на 2^{n-1}, сумма $C_{2^{n-1}-1}^{2b} + C_{2^{n-1}-1}^{2a+1}$ не может делиться на 2^{n-1}.

2.12. Эту задачу нам сообщил А. Белов. Заметим, что

$$C_{2n}^{n+k} = C_{2n}^n \cdot \frac{n(n-1)\ldots(n-k+1)}{(n+1)(n+2)\ldots(n+k)},$$

и таким образом, C_{2n}^{n+k} имеет много общих множителей с C_{2n}^{n}, кроме тех, которые сократились со знаменателем дроби. Заметим, что знаменатель не превосходит $(2n)^k$. Напишем аналогичные равенства для всех биномиальных коэффициентов $C_{2n}^{n+k_1}$, $C_{2n}^{n+k_2}$, …, $C_{2n}^{n+k_{100}}$. Наибольший общий делитель всех знаменателей в правых частях этих равенств не превосходит $(n+1)(n+2)\dots\big(n+[\varepsilon\sqrt{n}\,]\big) < (2n)^{\varepsilon\sqrt{n}}$. Но при больших n биномиальный коэффициент C_{2n}^{n} — существенно более крупное число, поэтому даже если сократить его на наибольший общий делитель всех знаменателей, останется весьма крупное частное, которое и будет общим делителем всех ста биномиальных коэффициентов.

Поясним последнее соображение с помощью оценки. Заметим, что

$$C_{2n}^{n} = \frac{2n}{n}\cdot\frac{2n-1}{n-1}\dots\frac{n+1}{1} > 2^n.$$

При этом $(2n)^{100\varepsilon\sqrt{n}} = 2^{\varepsilon\sqrt{n}\log_2 n+\varepsilon\sqrt{n}}$. Очевидно, для каждого фиксированного ε существует N, такое что при всех $n > N$ будет выполнено неравенство

$$\frac{n}{2} > \varepsilon\sqrt{n}\log_2 n + \varepsilon\sqrt{n}.$$

Если для таких n поделить C_{2n}^{n} на НОД всех знаменателей, частное будет не меньше $2^{n/2}$.

2.13. a) Задача предлагалась в 1977 г. на Ленинградской олимпиаде школьников.

Р е ш е н и е 1 (без теоремы Куммера). Мы приводим решение из замечательной книжки [4].

Допустим, что все эти числа делятся на m. Тогда числа

$$\begin{aligned} C_{n+k-1}^{k-1} &= C_{n+k}^{k} - C_{n+k-1}^{k}\,,\\ C_{n+k-2}^{k-1} &= C_{n+k-1}^{k} - C_{n+k-2}^{k}\,,\\ &\dots\\ C_{n}^{k-1} &= C_{n+1}^{k} - C_{n}^{k} \end{aligned}$$

также делятся на m. Аналогично, на m делятся и все числа C_{n+i}^{j}, где $i \leqslant j$ — произвольные неотрицательные целые числа. Но среди них есть число C_n^0 $(i = j = 0)$, которое равно 1. Противоречие.

Р е ш е н и е 2 (теорема Куммера). Пусть p — простой множитель числа m. Проверим, что одно из чисел C_n^k, C_{n+1}^k, …, C_{n+k}^k не делится на p. Запишем k в системе счисления по основанию p. По теореме Куммера достаточно найти такое число ℓ (где $n-k \leqslant \ell \leqslant n$), чтобы сложение $k+\ell$ в системе счисления по основанию p выполнялось без переносов, тогда биномиальный коэффициент $C_{k+\ell}^{k}$ не будет делиться на p.

Это сделать совсем нетрудно. Мы ограничимся рассуждением на конкретном примере. Пусть $p = 7$, $k = 133$ (здесь и далее числа записаны в семиричной системе счисления). Поскольку диапазон, в котором мы ищем число ℓ, содержит $k+1$ число, нам всегда удастся выбрать ℓ так, чтобы число $k+\ell$ было одним из чисел следующего вида

$$\dots 133,\quad \dots 233,\quad ,\dots,\quad \dots 633.$$

(Напомним, что цифра 6 в нашем примере самая старшая.) Тогда очевидно, что при сложении $k+\ell$ не было ни одного переноса.

b) Утверждение взято из [2]. Такие n нетрудно построить с помощью теоремы Куммера. Пусть $\operatorname{ord}_p m = s$, и запись числа k в системе счисления по основанию p содержит $d+1$ цифр. Пусть $n \vdots p^{d+s+1}$. Тогда числа $n-k$, $n-k+1$, …, $n-1$ содержат в разрядах с $(d+2)$-го по $(d+s+2)$-й цифры $(p-1)$, поэтому при сложении этих чисел с k в указанных разрядах будут возникать переносы. Таким образом, по теореме Куммера получаем, что интересующие нас биномиальные коэффициенты все делятся на p^s.

Поскольку условия, наложенные на n, легко совмещаются для разных p, мы получаем отсюда требуемое.

3 Обобщение теорем Вильсона и Люка

3.1. Как известно, $\operatorname{ord}_p(n!) = \sum_k \left[\frac{n}{p^k}\right]$. Если $n = n_d p^d + n_{d-1}p^{d-1} + \ldots + n_1 p + n_0$ — запись в системе счисления по основанию p, то $\left[\frac{n}{p^k}\right] = n_d p^{d-k} + n_{d-1}p^{d-k-1} + \ldots + n_{k+1}p + n_k$ и формулу для $\operatorname{ord}_p(n!)$ можно записать в виде

$$\operatorname{ord}_p(n!) = \sum_{k=1}^{d}\left(\sum_{i=k}^{d} n_i p^{i-k}\right) = \sum_{i=1}^{d} n_i(p^{i-1}+p^{i-2}+\ldots+p+1) = \sum_{i=1}^{d} n_i\frac{p^i-1}{p-1} = \frac{\sum\limits_{i=0}^{d} n_i p^i - \sum\limits_{i=0}^{d} n_i}{p-1}.$$

Мы получили в точности требуемое выражение.

Утверждение задачи также нетрудно доказать индукцией по n, см. [5].

3.2. а) Разбивая множители, составляющие выражение $n!$, на группы по $(p-1)$ штук, получаем

$$(n!)_p = \prod_{k=0}^{[\frac{n}{p}]-1}\big((kp+1)\cdot(kp+2)\cdots(kp+p-1)\big)\cdot\big([\tfrac{n}{p}]p+1\big)\big([\tfrac{n}{p}]p+2\big)\ldots\big([\tfrac{n}{p}]p+n_0\big) \equiv (-1)^{[\frac{n}{p}]} n_0! \pmod p.$$

б) Это утверждение встречается у Гаусса [15]. В произведение $(p^q!)_p$ вместе с каждым сомножителем входит и его обратный по модулю p^q, и произведение этой пары равно 1 по модулю p^q. Таким образом, нам следует лишь проследить за теми множителями m, которые совпадают со своими обратными, т.е. удовлетворяют сравнению

$$m^2 \equiv 1 \pmod{p^q}.$$

Для нечетного p сравнение имеет 2 решения: ± 1. Для $p=2$, $q \geqslant 3$ сравнение имеет еще пару решений: $2^{q-1} \pm 1$.

с) Так как $n! = (n!)_p \cdot p^{[\frac{n}{p}]}\big([\frac{n}{p}]\big)!$, утверждение легко доказывается по индукции с помощью сравнения из п. а).

3.3. Мы взяли утверждение со странички Гранвилля [17]. Помимо теоремы Куммера, широко известна прямая и не столь симпатичная формула для числа ℓ (формула Лежандра):

$$\ell = \operatorname{ord}_p(C_n^k) = \left(\left[\frac{n}{p}\right] - \left[\frac{k}{p}\right] - \left[\frac{r}{p}\right]\right) + \left(\left[\frac{n}{p^2}\right] - \left[\frac{k}{p^2}\right] - \left[\frac{r}{p^2}\right]\right) + \ldots \tag{4}$$

Обозначим для краткости $\tilde{n} = [n/p]$ и т. п. и напишем формулу для биномиального коэффициента, собрав отдельно все множители, делящиеся на p:

$$C_n^k = \frac{(n!)_p}{(k!)_p (r!)_p}\cdot\frac{p^{[n/p]}}{p^{[k/p]}\cdot p^{[r/p]}}\cdot\frac{\tilde{n}!}{\tilde{k}!\cdot\tilde{r}!}.$$

Здесь первая дробь может быть преобразована по модулю p в соответствии с обобщенной теоремой Вильсона (задача 3.2, б) к выражению $\frac{n_0!}{k_0! r_0!}$, третья дробь позволяет действовать по индукции, а средняя дробь (и знак из обобщенной теоремы Вильсона, который мы не упомянули) по формуле (4) даст все нужные выражения, содержащие ℓ.

3.4. а) Раскрывая скобки в выражении $(1+x)^{p^d}$, мы можем воспользоваться тем, что при $1 \leqslant k \leqslant p^d-1$ биномиальный коэффициент $C_{p^d}^k$ делится на p (аналогично задаче 1.1 или по теореме Куммера).

b) Положим $n = n'p + n_0$, $k = k'p + k_0$. По утверждению п. а) $(1+x)^{pn'} \equiv (1+x^p)^{n'} \pmod p$ Тогда

$$(1+x)^n = (1+x)^{pn'}(1+x)^{n_0} \equiv (1+x^p)^{n'}(1+x)^{n_0} \pmod p.$$

Указанное сравнение надо понимать в том смысле, что мы преобразовываем коэффициенты многочлена с целыми коэффициентами с точки зрения их делимости на p. Коэффициент при x^k в левой части равен C_n^k. При раскрытии скобок в правой части мы видим, что все показатели в первой скобке делятся на p, поэтому единственный способ получить одночлен $x^{pk'+k_0}$ — это перемножить

$x^{pk'}$ из первой скобки и x^{k_0} из второй. Итоговый коэффициент будет равен $C_{n'}^{k'}C_{n_0}^{k_0}$. Таким образом, $C_n^k = C_{n'}^{k'}C_{n_0}^{k_0}$, откуда теорема Люка следует по индукции.

3.5. a, b) Простое следствие теоремы Куммера.

3.6. [9]. В следующем вычислении мы используем то, что $C_{n_i}^{k_i} = 0$ при $k_i > n_i$; это позволяет, применив теорему Люка, отбросить при суммировании большое число слагаемых.

$$f_{n,a} = \sum_{k=0}^{n}(C_n^k)^a \equiv \sum_{k_d=0}^{n_d}\sum_{k_{d-1}=0}^{n_{d-1}} \cdots \sum_{k_0=0}^{n_0}\prod_{i=0}^{d}(C_{n_i}^{k_i})^a \equiv \prod_{i=0}^{d}\sum_{k_i=0}^{n_i}(C_{n_i}^{k_i})^a \equiv \prod_{i=0}^{d} f_{n_i,a} \pmod p.$$

4 *Вариации на тему теоремы Волстенхолма*

4.1. Это упражнение на чтение статьи. Утверждение доказано в статье Винберга, но доказательство не выделено явно. Заметим, что

$$2\sum_{i=1}^{p-1}\frac{1}{i} = \sum_{i=1}^{p-1}\frac{1}{i} + \frac{1}{p-i} = p\sum_{i=1}^{p-1}\frac{1}{i(p-i)}\,.$$

Таким образом, рассматриваемая сумма делится на p. Так как по модулю p выражения $\frac{1}{i}$ и $-\frac{1}{p-i}$ равны, нам остается проверить, что

$$\sum_{i=1}^{p-1}\frac{1}{i^2} \equiv 0 \pmod p.$$

Или, поскольку $\frac{1}{1^2}, \frac{1}{2^2}, \ldots, \frac{1}{(p-1)^2}$ — это тот же набор остатков[1], что и $1^2, 2^2, \ldots, (p-1)^2$, достаточно проверить, что

$$\sum_{i=1}^{p-1} i^2 \equiv 0 \pmod p. \tag{5}$$

Пусть $\sum\limits_{i=1}^{p-1} i^2 \equiv s \pmod p$. При $p > 5$ всегда можно выбрать остаток a, такой что $a^2 \not\equiv 1 \pmod p$. Тогда множества $\{1, 2, \ldots, p-1\}$ и $\{a, 2a, \ldots, (p-1)a\}$ совпадают (доказательство как в сноске) и

$$s \equiv \sum_{i=1}^{p-1} i^2 = \sum_{i=1}^{p-1}(ai)^2 = a^2\sum_{i=1}^{p-1} i^2 \equiv a^2 s \pmod p\,.$$

Поэтому $s \equiv 0 \pmod p$.

Разумеется, этот факт нетрудно доказать непосредственно, пользуясь соображением $\frac{1}{x} \equiv x^{\varphi(m)-1} \pmod m$. Мы используем эту технику в третьем решении следующей задачи.

4.2. Ответ: $2k+2$. Эта задача А. С. Голованова предлагалась на олимпиаде Туймаада в 2012 г. Мы приводим три решения. Отметим, что при $p = 4k+3$ уравнение $x^2 + 1 = 0$ не имеет решений в поле остатков по модулю p, следовательно, знаменатели всех рассматриваемых дробей не равны нулю.

Р е ш е н и е 1. Обозначим $a_i = i^2 + 1$, для $i = 0, \ldots, p-1$. Тогда рассматриваемое выражение равно

$$\frac{\sigma_{p-1}(a_0, a_1, \ldots, a_{p-1})}{\sigma_p(a_0, a_1, \ldots, a_{p-1})},$$

где σ_i — основной симметрический многочлен степени i. Найдем многочлен, корнями которого являются числа a_i, т. е.

$$\prod_{i=0}^{p-1}(x - 1 - i^2).$$

[1] Напомним доказательство: $\frac{1}{1}, \frac{1}{2}, \ldots, \frac{1}{(p-1)}$ и 1, 2, …, $(p-1)$ — это один и тот же набор остатков, потому что и в том, и в другом наборе по $p-1$ элементу, при этом очевидно, что в каждом наборе все остатки различны и не равны нулю, значит, каждый набор содержит все ненулевые остатки по модулю p. Тогда для квадратов утверждение очевидно.

Сделав замену $x-1=t^2$, получим многочлен

$$\prod_{i=0}^{p-1}(t^2-i^2)=\prod_{i=0}^{p-1}(t-i)\prod_{i=0}^{p-1}(t+i)\equiv(t^p-t)(t^p-t)=t^{2p}-2t^{p+1}+t^2.$$

Теперь, сделав обратную замену, получаем для $p=4k+3$

$$\prod_{i=0}^{p-1}(x-1-i^2)\equiv(x-1)^p-2(x-1)^{\frac{p+1}{2}}+(x-1)=x^p+\ldots+(p+2\cdot\tfrac{p+1}{2}+1)x-4.$$

По теореме Виета, $\sigma_p\equiv 4 \pmod p$, $\sigma_{p-1}\equiv 2 \pmod p$, поэтому $\frac{\sigma_{p-1}}{\sigma_p}\equiv\frac{1}{2}\equiv 2k+2 \pmod p$.

Р е ш е н и е 2. Разобьем все ненулевые остатки по модулю p, кроме ±1, на пары взаимно обратных. Тогда получится $2k$ пар и в каждой паре (i,j)

$$ij\equiv 1 \quad\Leftrightarrow\quad i^2j^2\equiv 1 \quad\Leftrightarrow\quad (ij)^2+i^2+j^2+1\equiv i^2+j^2+2 \pmod p.$$

Следовательно,

$$1\equiv\frac{(ij)^2+i^2+j^2+1}{(i^2+1)(j^2+1)}\equiv\frac{i^2+j^2+2}{(i^2+1)(j^2+1)}=\frac{1}{i^2+1}+\frac{1}{j^2+1} \pmod p.$$

Таким образом, наша сумма равна $\frac{1}{0^2+1}+\frac{1}{1^2+1}+\frac{1}{(-1)^2+1}+2k\equiv 2k+2$.

Р е ш е н и е 3. Как мы знаем, благодаря малой теореме Ферма, при вычислении по модулю p операции $x\mapsto x^{-1}$ и $x\mapsto x^{p-2}$ дают одинаковый результат. Таким образом, достаточно вычислить сумму

$$\sum_{x=0}^{p-1}(x^2+1)^{p-2}=\sum_{x=0}^{p-1}\sum_{m=0}^{p-2}C_{p-2}^m x^{2m}=\sum_{m=0}^{p-2}C_{p-2}^m S_{2m}, \tag{6}$$

где $S_{2m}=\sum\limits_{x=0}^{p-1}x^{2m}$. Очевидно, $S_{2m}\equiv -1 \pmod p$ при $m=\frac{p-1}{2}$. Докажем, что $S_{2m}\equiv 0 \pmod p$ при остальных значениях m, не превосходящих $p-1$. Действительно, для каждого такого m можно подобрать ненулевой остаток a, такой что $a^{2m}\not\equiv 1 \pmod p$ и тогда можно провести рассуждение как в (5). Возвращаясь к интересующей нас сумме (6), получаем

$$\sum_{m=0}^{p-2}C_{p-2}^m S_{2m}\equiv -C_{p-2}^{\frac{p-1}{2}}=-C_{4k+1}^{2k+1}=-\frac{(4k+1)\cdot 4k\cdot\ldots\cdot(2k+1)}{1\cdot 2\cdot\ldots\cdot(2k+1)}\equiv-\frac{(-2)\cdot(-3)\ldots(2k+2)}{1\cdot 2\cdot\ldots\cdot(2k+1)}\equiv 2k+2.$$

4.3. Мы нашли оба утверждения в [16].

a) Для каждого простого делителя p числа m подберем число a_p, для которого $(a_p^k-1)\not\vdots\, p$. С помощью китайской теоремы об остатках выберем число a, такое что $a\equiv a_p \pmod p$ при всех p. Теперь результат получается аналогично рассуждениям (5).

b) Заметим, что при нечетных k по формуле бинома $i^k+(p-i^k)\equiv ki^{k-1}p \pmod{p^2}$. Тогда

$$2\sum_{i=1}^{p-1}\frac{1}{i^k}=\sum_{i=1}^{p-1}\left(\frac{1}{i^k}+\frac{1}{(p-i)^k}\right)=\sum_{i=1}^{p-1}\frac{i^k+(p-i)^k}{i^k(p-i)^k}\equiv\sum_{i=1}^{p-1}\frac{ki^{k-1}p}{i^k(-i)^k}\equiv -kp\sum_{i=1}^{p-1}\frac{1}{i^{k+1}} \pmod{p^2}.$$

Сумма в правой части сравнения делится на p в силу утверждения п. a).

4.4. Как доказывается в [24], сравнение выполнено даже по модулю p^7, но мы не будем заходить так далеко. Действуя как в статье Винберга [1], но следя за степенями до p^4, получаем

$$C_{p-1}^{2p-1}=\frac{(2p-1)(2p-2)\cdot\ldots\cdot(p+1)}{p!}=\left(\frac{2p}{1}-1\right)\left(\frac{2p}{2}-1\right)\cdot\ldots\cdot\left(\frac{2p}{p-1}-1\right)\equiv$$
$$\equiv 1-2p\sum_{i=1}^{p-1}\frac{1}{i}+4p^2\sum_{\substack{i,j=1\\ i<j}}^{p-1}\frac{1}{ij}-8p^3\sum_{\substack{i,j,k=1\\ i<j<k}}^{p-1}\frac{1}{ijk} \pmod{p^4}. \tag{7}$$

Выразим последнюю сумму через степенные суммы:

$$\sum_{\substack{i,j,k=1\\ i<j<k}}^{p-1} \frac{1}{ijk} = \frac{S_3}{3} - \frac{S_1S_2}{2} + \frac{S_1^3}{6}\,, \quad \text{где } S_k = \sum_{i=1}^{p-1} \frac{1}{i^k}.$$

Как мы знаем, S_1 и S_3 делятся на p^2 (последнее — из задачи 4.3б). Поэтому последнее слагаемое в формуле (7) можно отбросить.

4.5. Задача из [1], обсуждение вариаций на эту тему можно прочесть в [14].

Поскольку

$$2\sum_{k=1}^{p-1} \frac{1}{k^2} = \sum_{k=1}^{p-1}\Big(\frac{1}{k^2} + \frac{1}{(p-k)^2}\Big) = \sum_{k=1}^{p-1} \frac{k^2+(p-k)^2}{k^2(p-k)^2} \equiv -2\sum_{k=1}^{p-1} \frac{1}{k(p-k)} \pmod{p^2},$$

утверждение 3) эквивалентно соотношению $\displaystyle\sum_{k=1}^{p-1} \frac{1}{k(p-k)} \equiv 0 \pmod{p^2}$. Утверждение 2) тоже эквивалентно этому соотношению, так как $\displaystyle 2\sum_{k=1}^{p-1} \frac{1}{k} = 2\sum_{k=1}^{p-1}\Big(\frac{1}{k} + \frac{1}{p-k}\Big) = 2p\sum_{k=1}^{p-1} \frac{1}{k(p-k)}$. Наконец, как мы знаем из предыдущей задачи,

$$C_{2p-1}^{p-1} \equiv 1 - p^2\sum_{i=1}^{p-1} \frac{1}{i(p-i)} + 4p^2 \sum_{\substack{i,j=1\\ i<j}}^{p-1} \frac{1}{ij} \pmod{p^4}.$$

Таким образом, утверждение 1) эквивалентно сравнению

$$\sum_{i=1}^{p-1} \frac{1}{i(p-i)} \equiv 4 \sum_{\substack{i,j=1\\ i<j}}^{p-1} \frac{1}{ij} \pmod{p^2}. \tag{8}$$

Преобразуем выражение в правой части:

$$4 \sum_{\substack{i,j=1\\ i<j}}^{p-1} \frac{1}{ij} = 2\Big(\sum_{k=1}^{p-1} \frac{1}{i}\Big)^2 - 2\sum_{k=1}^{p-1} \frac{1}{k^2} \equiv 2\Big(\sum_{k=1}^{p-1} \frac{1}{i}\Big)^2 + 2\sum_{k=1}^{p-1} \frac{1}{k(p-k)}.$$

Сумма в скобке делится на p, ее квадрат делится на p^2 и это слагаемое можно отбросить. Подставляя в (8), получаем, что и первое утверждение равносильно сравнению $\displaystyle\sum_{k=1}^{p-1} \frac{1}{k(p-k)} \equiv 0 \pmod{p^2}$.

4.6. а) Р е ш е н и е 1 ([5, предложение 2.12]). Индукция по n. Раскроем скобки в равенстве

$$(a+b)^{pn} = (a+b)^{p(n-1)}(a+b)^p$$

Приравняем коэффициенты при $a^{pm}b^{p(n-m)}$:

$$C_{pn}^{pm} = C_{p(n-1)}^{pm} C_p^0 + C_{p(n-1)}^{pm-1} C_p^1 + \ldots + C_{p(n-1)}^{pm-p+1} C_p^{p-1} + C_{p(n-1)}^{pm-p} C_p^p\,.$$

В правой части все слагаемые, кроме крайних, делятся на p^2, потому что каждый биномиальный коэффициент в них делится на p по теореме Люка. Следовательно,

$$C_{pn}^{pm} \equiv C_{p(n-1)}^{pm} + C_{p(n-1)}^{p(m-1)} \pmod{p^2}.$$

По предположению индукции

$$C_{p(n-1)}^{pm} + C_{p(n-1)}^{p(m-1)} \equiv C_{n-1}^m + C_{n-1}^{m-1} \equiv C_n^m \pmod{p^2}.$$

Р е ш е н и е 2 ([Д]). Докажем, что $C_{kp}^{mp} \equiv C_k^m \pmod{p^2}$ индукцией по m.

База $m = 1$. Требуется проверить, что $C_{pk}^p - C_k^1 \equiv 0 \pmod{p^2}$. Преобразуем эту разность:

$$C_{pk}^p - C_k^1 = \frac{pk(pk-1)\dots(pk-p+1)}{p!} - k = \left(\frac{(pk-1)(pk-1)\dots(pk-p+1)}{(p-1)!} - 1\right). \tag{9}$$

В числителе большой дроби четное число сомножителей. Разобьем их на пары:

$$(pk-i)(pk-p+i) \equiv pi^2 - i^2 \pmod{p^2}.$$

Как видим, по модулю p^2 произведение чисел в парах не зависит от k. Поэтому вычисление разности (9) по модулю p^2 дает одинаковый результат при всех k. Но при $k = 1$ вычисляемое выражение равно 0.

Переход. Пусть $C_{kp}^{(m-1)p} \equiv C_k^{m-1} \pmod{p^2}$.

$$C_{kp}^{mp} = C_{kp}^{(m-1)p} \cdot \frac{(p(k-m)+1)(p(k-m)+1)\dots(p(k-m)+p)}{pm(pm-1)\dots(pm-p+1)} =$$
$$= C_{kp}^{(m-1)p} \cdot \frac{(p(k-m)+1)(p(k-m)+1)\dots(p(k-m)+p-1)}{(pm-1)\dots(pm-p+1)} \cdot \frac{k-m+1}{m} \tag{10}$$

Отметим, что обе дроби корректно определены по модулю p^2. Как и в доказательстве базы, выражение в числителе большой дроби по модулю p^2 не зависит от k. Тогда для вычисления большой дроби можно взять $k = 0$, и мы сразу получим, что по модулю p^2 дробь равна 0. Пользуясь этим соображением и предположением индукции, мы можем заменить правую часть (10) на

$$\equiv C_k^{m-1} \cdot \frac{k-m+1}{m} = C_k^m \pmod{p^2}.$$

b) Р е ш е н и е 1 (комбинаторное). Как и рекомендуется в [1], рассматриваем выборки kp предметов из общего количества pn предметов. Полагаем, что исходное множество предметов разбито на блоки по p штук. Количество блочных выборок равно C_n^k. Таким образом, остается проверить, что количество неблочных выборок делится на p^3. Как объясняется в статье, количество неблочных выборок с тремя и более блоками делится на p^3. Так как при $k > 1$ любая неблочная выборка содержит не менее трех блоков, то в этом случае все доказано. Остается разобрать случай, когда $k = 1$ и мы подсчитываем количество неблочных выборок p предметов из из общего множества в $2p$ предметов. Это количество равно $C_{2p}^p - 2$, что по теореме Волстенхолма делится на p^3.

Р е ш е н и е 2. Напишем формулу для биномиального коэффициента $C_a^b = \frac{a(a-1)\dots(a-b+1)}{b(b-1)\dots1}$, разбив числитель и знаменатель на блоки из p сомножителей, после чего сократим первые множители в каждом блоке, а частные соберем в отдельное выражение:

$$C_{mp}^{kp} = \frac{m\not{p}\cdot(mp-1)\dots\big(mp-(p-1)\big)}{k\not{p}\cdot(kp-1)\dots\big(kp-(p-1)\big)} \cdot \frac{(m-1)\not{p}\cdot\big((m-1)p-1\big)\dots\big((m-1)p-(p-1)\big)}{(k-1)\not{p}\cdot\big((k-1)p-1\big)\dots\big((k-1)p-(p-1)\big)} \cdot \dots \times$$
$$\times \frac{(m-k+1)\not{p}\cdot\big((m-k+1)p-1\big)\dots\big((m-k+1)p-(p-1)\big)}{\not{p}\cdot(p-1)\dots1} =$$
$$= C_m^k \cdot \frac{(mp-1)\dots\big(mp-(p-1)\big)}{(kp-1)\dots\big(kp-(p-1)\big)} \cdot \dots \cdot \frac{\big((m-k+1)p-1\big)\dots\big((m-k+1)p-(p-1)\big)}{(p-1)\dots1}.$$

Осталось проверить, что произведение дробей дает остаток 1 при делении на p^3. Для этого достаточно проверить сравнение

$$\frac{(np-1)\dots\big(np-(p-1)\big)}{(rp-1)\dots\big(rp-(p-1)\big)} \equiv 1 \pmod{p^3}$$

или, лучше, вот такое сравнение

$$\frac{(np-1)\dots\big(np-(p-1)\big)}{(p-1)!} \equiv \frac{(rp-1)\dots\big(rp-(p-1)\big)}{(p-1)!} \pmod{p^3}.$$

Это верно, так как обе части сравнимы с 1 по модулю p^3, что устанавливается аналогично доказательству теоремы Волстенхолма.

4.7. a) [5, теорема 2.14]. Преобразуем разность

$$C_{p^2}^p - C_p^1 = \frac{p^2(p^2-1)\dots(p^2-(p-1))}{1\cdot 2\cdot\ldots\cdot(p-1)p} - p = \frac{p}{(p-1)!}\Big((1-p^2)(2-p^2)\dots((p-1)-p^2) - 1\cdot 2\cdot\ldots\cdot(p-1)\Big).$$

Осталось проверить, что

$$(1-p^2)(2-p^2)\dots((p-1)-p^2) \equiv 1\cdot 2\cdot\ldots\cdot(p-1) \pmod{p^4}.$$

Раскроем скобки в левой части:

$$(1-p^2)(2-p^2)\dots((p-1)-p^2) = 1\cdot 2\cdot\ldots\cdot(p-1) + p^2\Big(1+\frac{1}{2}+\ldots+\frac{1}{p-1}\Big)(p-1)! + \text{члены делящиеся на } p^4.$$

По утверждению задачи 4.1 второе слагаемое делится на p^4.

b) Как нетрудно видеть, $C_{p^{s+1}}^p = p^s \cdot C_{p^{s+1}-1}^{p-1}$, поэтому достаточно проверить, что $C_{p^{s+1}-1}^{p-1} \equiv 1 \pmod{p^{s+3}}$.

$$C_{p^{s+1}-1}^{p-1} = \frac{(p^{s+1}-1)(p^{s+1}-2)\dots(p^{s+1}-(p-1))}{1\cdot 2\cdots(p-1)} = \left(\frac{p^{s+1}}{1}-1\right)\left(\frac{p^{s+1}}{2}-1\right)\dots\left(\frac{p^{s+1}}{p-1}-1\right) \equiv$$
$$\equiv (-1)^{p-1} + p^{s+1}\Big(1+\frac{1}{2}+\ldots+\frac{1}{p-1}\Big) \pmod{p^{s+3}}.$$

Это и есть то, что требуется, поскольку $(-1)^{p-1} = 1$ и $1+\frac{1}{2}+\ldots+\frac{1}{p-1} \equiv 0 \mod p^2$.

В статье [14] доказывается чуть более общий факт.

4.8. Задача из статьи Винберга [1], решение [T].

$$C_{p^3}^{p^2} - C_{p^2}^p = p\left(C_{p^3-1}^{p^2-1} - C_{p^2-1}^{p-1}\right) =$$
$$= p\left(\left(\frac{p^3}{1}-1\right)\left(\frac{p^3}{2}-1\right)\dots\left(\frac{p^3}{p^2-1}-1\right) - \left(\frac{p^2}{1}-1\right)\left(\frac{p^2}{2}-1\right)\dots\left(\frac{p^2}{p-1}-1\right)\right) =$$
$$= p\Big(\frac{p^2}{1}-1\Big)\Big(\frac{p^2}{2}-1\Big)\dots\Big(\frac{p^2}{p-1}-1\Big)\left(\prod_{\substack{k=1\\ p\nmid k}}^{p^2-1}\left(\frac{p^3}{k}-1\right)-1\right).$$

Достаточно проверить, что выражение в последней скобке делится на p^7. Преобразуем произведение

$$\prod_{\substack{k=1\\ p\nmid k}}^{p^2-1}\left(\frac{p^3}{k}-1\right) = \prod_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\left(\frac{p^3}{k}-1\right)\left(\frac{p^3}{p^2-k}-1\right) = \prod_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\left(\frac{p^6-p^5}{k(p^2-k)}+1\right) \equiv 1+p^5(p-1)\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k(p^2-k)} \pmod{p^7}.$$

Осталось проверить, что последняя сумма делится на p^2. Это так, поскольку по задаче 4.3a)

$$\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k(p^2-k)} \equiv -\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k^2} \equiv 0 \pmod{p^2}.$$

4.9. Это [6, теорема 5]. Более общий факт доказан в [7].

Р е ш е н и е 1 ([5, предложение 2.19]). Воспользуемся тем, что разность $C_{2^{k+1}}^{2^k} - C_{2^k}^{2^{k-1}}$ равна коэффициенту при x^{2^k} в многочлене

$$(1+x)^{2^{k+1}} - (1-x^2)^{2^k} = (1+x)^{2^k}\Big((1+x)^{2^k} - (1-x)^{2^k}\Big) =$$
$$= \Big(1 + C_{2^k}^1 x + C_{2^k}^2 x^2 + \ldots + x^{2^k}\Big)\cdot 2\Big(C_{2^k}^1 x + C_{2^k}^3 x^3 + \ldots + C_{2^k}^{2^k-1} x^{2^k-1}\Big).$$

Поскольку второй многочлен содержит только множители нечетной степени, коэффициент при x^{2^k} в произведении равен

$$2\Big(C_{2^k}^1 C_{2^k}^{2^k-1} + C_{2^k}^3 C_{2^k}^{2^k-3} + \ldots + C_{2^k}^{2^k-1} C_{2^k}^1\Big).$$

По утверждению задачи 3.5 б) каждый биномиальный коэффициент в этом выражении делится на 2^k, кроме того, каждое слагаемое в сумме встречается 2 раза, а перед суммой стоит коэффициент 2. В итоге все выражение делится на 2^{2k+2}.

Решение 2 ([CSTTVZ]). Так как $C_{2^{n+1}}^{2^n} = 2C_{2^{n+1}-1}^{2^n-1}$, достаточно доказать соотношение

$$C_{2^{n+1}-1}^{2^n-1} \equiv C_{2^n-1}^{2^{n-1}-1} \pmod{2^{2n+1}}.$$

Аналогично (3) получаем

$$C_{2^{n+1}-1}^{2^n-1} = \left(\frac{2^{n+1}}{1} - 1\right)\left(\frac{2^{n+1}}{3} - 1\right)\ldots\left(\frac{2^{n+1}}{2^n-1} - 1\right)\cdot C_{2^n-1}^{2^{n-1}-1}.$$

Достаточно проверить, что

$$L = \left(\frac{2^{n+1}}{1} - 1\right)\left(\frac{2^{n+1}}{3} - 1\right)\ldots\left(\frac{2^{n+1}}{2^n-1} - 1\right) \equiv 1 \pmod{2^{2n+1}}.$$

Это так, поскольку

$$L \equiv (-1)^{2^{n-1}} - 2^{n+1}\left(\frac{1}{1} + \frac{1}{3} + \frac{1}{5} + \ldots + \frac{1}{2^n-1}\right) \equiv$$
$$\equiv 1 - 2^{n+1}\left(\frac{2^n}{1\cdot(2^n-1)} + \frac{2^n}{3\cdot(2^n-3)} + \ldots + \frac{2^n}{(2^{n-1}-1)(2^{n-1}+1)}\right) \equiv 1 \pmod{2^{2n+1}}.$$

4.10. Это теорема Морли [26].

Решение 1 (авторское решение из статьи 1895 года). Оно лишь чуть-чуть выходит за рамки школьной программы.

Возьмем формулу, с помощью которой $\cos^{2n+1} x$ выражается через косинусы кратных углов,[1] или, как говорили в те времена, запишем $\cos^{2n+1} x$ в виде, удобном для интегрирования:

$$2^{2n}\cos^{2n+1}x = \cos(2n{+}1)x + (2n{+}1)\cos(2n{-}1)x + \frac{(2n{+}1)\cdot 2n}{1\cdot 2}\cos(2n{-}3)x + \ldots + \frac{(2n{+}1)\cdot 2n\ldots(n{+}2)}{n!}\cos x.$$

Ну, а раз этот вид удобен для интегрирования, то и проинтегрируем обе части[2] по промежутку $[0, \frac{\pi}{2}]$:

$$2^{2n}\int \cos^{2n+1} x\,dx = \frac{\sin(2n+1)x}{2n+1} + \frac{2n+1}{2n-1}\sin(2n-1)x + \ldots,$$
$$2^{2n}\int_0^{\pi/2} \cos^{2n+1} x\,dx = (-)^n\left(\frac{1}{2n+1} - \frac{2n+1}{2n-1} + \ldots\right).$$

Но любой первокурсник знает, что куда проще этот интеграл вычисляется с помощью формулы понижения, для получения которой нужно всего лишь проинтегрировать по частям:

$$I_{2n+1} = \int_0^{\pi/2} \cos^{2n+1} x\,dx = \int_0^{\pi/2} \cos^{2n} x\cos x\,dx = \cos^{2n} x\sin x\,\Big|_0^{\pi/2} + 2n\int_0^{\pi/2} \cos^{2n-1} x\sin^2 x\,dx =$$
$$= 0 + 2n\int_0^{\pi/2} \cos^{2n-1} x(1-\cos^2 x)\,dx = 2n\cdot I_{2n-1} - 2n\cdot I_{2n+1},$$

[1] Читатель, интересующийся вопросом "где мы ее возьмем" и не удовлетворенный ответом "в справочнике", может просто воспользоваться формулой Эйлера $\cos\varphi = \frac{1}{2}(e^{i\varphi} + e^{-i\varphi})$ и возвести правую часть в степень $2n+1$ по формуле бинома.

[2] Когда мы учим правила умножения, мы запоминаем формулу "минус на минус будет плюс". В этой формуле мы перемножаем знаки. Значит, если нам нужно перемножить n минусов, кажется вполне уместной запись $(-)^n$. Поэтому мы оставляем старомодное обозначение $(-)^n$, как у автора, вместо современного $(-1)^n$.

откуда находим, что $I_{2n+1} = \dfrac{2n}{2n+1}\cdot I_{2n-1}$. Учитывая что $I_1 = 1$, применяя эту формулу n раз подряд, находим, что

$$\int\limits_0^{\pi/2} \cos^{2n+1} x\,dx = \frac{2n\cdot(2n-2)\dots 2}{(2n+1)(2n-1)\dots 3}\,.$$

Приравнивая эти два способа подсчета интеграла, мы получаем тождество

$$2^{2n}\frac{2n\cdot(2n-2)\dots 2}{(2n+1)(2n-1)\dots 3} = (-)^n\left(\frac{1}{2n+1} - \frac{2n+1}{2n-1} + \dots + \frac{(2n+1)\cdot 2n\dots(n+2)}{n!}\right).$$

Если взять $p = 2n+1$ — простое число, то домножая на p, мы сразу получаем требуемое сравнение

$$2^{2n}\frac{2n\cdot(2n-2)\dots 2}{(2n-1)(2n-3)\dots 3} \equiv (-)^n \pmod{p^2}\,.$$

Р е ш е н и е 2 ([CSTTVZ]). Введем несколько обозначений. Пусть

$$A = \sum_{i=1}^{\frac{p-1}{2}}\frac{1}{i}, \qquad B = \sum_{1\leqslant i<j\leqslant\frac{p-1}{2}}\frac{1}{ij}, \qquad C = \sum_{\substack{1\leqslant i\leqslant p-1\\ i \text{ нечетно}}}\frac{1}{i}.$$

Очевидно, $A^2 = \sum\limits_{i=1}^{\frac{p-1}{2}}\frac{1}{i^2} + 2B \equiv 2B \pmod p$ по задаче 4.3b). Итак, $A^2 \equiv 2B \pmod p$. Далее,

$$2C + A = \sum_{\substack{1\leqslant i\leqslant p-1\\ i \text{ нечетно}}}\frac{2}{i} + \sum_{i=1}^{\frac{p-1}{2}}\frac{2}{2i} = \sum_{i=1}^{p-1}\frac{2}{i} \equiv 0 \pmod{p^2}.$$

Таким образом, $C \equiv -\frac{1}{2}A \pmod{p^2}$.

Теперь преобразуем по модулю p^3 правую и левую части доказываемого сравнения. Левая часть:

$$(-1)^{\frac{p-1}{2}}C_{p-1}^{\frac{p-1}{2}} \equiv \left(1-\frac{p}{1}\right)\left(1-\frac{p}{2}\right)\dots\left(1-\frac{p}{\frac{p-1}{2}}\right) \equiv 1 - pA + p^2B \equiv 1 - pA + \frac{1}{2}p^2A^2 \pmod{p^3}.$$

Для преобразования правой части заметим, что

$$2^{p-1} = \frac{2\cdot 4\cdots(p-1)}{1\cdot 2\cdots\frac{p-1}{2}}\cdot\frac{(p+1)\cdots(2p-2)}{\frac{p+1}{2}\cdots(p-1)} = \frac{(p+1)\cdots(2p-2)}{1\cdot 3\cdot 5\cdots(p-2)} =$$
$$= \left(\frac{p}{1}+1\right)\left(\frac{p}{3}+1\right)\dots\left(\frac{p}{p-1}+1\right) \equiv 1 + pC + \frac{1}{2}p^2C^2 \equiv 1 - \frac{1}{2}pA + \frac{1}{8}p^2A^2 \pmod{p^3}\,.$$

Отсюда получаем

$$4^{p-1} \equiv \left(1 - \frac{1}{2}pA + \frac{1}{8}p^2A^2\right)^2 \equiv 1 - pA + \frac{1}{4}p^2A^2 + 2\cdot\frac{1}{8}P^2A^2 = 1 - pA + \frac{1}{2}p^2A^2 \pmod{p^3}.$$

Таким образом, левая часть эквивалентна правой.

4.11. Мы взяли утверждение в [10].

$$\sum_{k=1}^{p-1}\frac{1}{mp+k} = \frac{1}{2}\sum_{k=1}^{p-1}\left(\frac{1}{mp+k} + \frac{1}{mp+p-k}\right) =$$
$$= p\cdot\frac{2m+1}{2}\cdot\sum_{k=1}^{p-1}\frac{1}{(mp+k)(mp+p-k)} \equiv -p\cdot\frac{2m+1}{2}\cdot\sum_{k=1}^{p-1}\frac{1}{k^2} \equiv 0 \pmod{p^2}.$$

4.12. Мы взяли утверждение в [8]. Так как $2pq-1 = (2q-1)p+p-1$, у числа $2pq-1$ последняя цифра p-ичной записи — это $p-1$, а остальные цифры образуют запись числа $2q-1$. Аналогично в записи числа $pq-1$ последняя цифра — $p-1$, а остальные цифры образуют запись числа $q-1$. По теореме Люка $C_{2pq-1}^{pq-1} \equiv C_{2q-1}^{q-1}C_{p-1}^{p-1} \equiv C_{2q-1}^{q-1} \pmod p$. С другой стороны, очевидно, что так как $C_{2pq-1}^{pq-1} \equiv 1 \pmod{pq}$, то $C_{2pq-1}^{pq-1} \equiv 1 \pmod p$. Таким образом, $C_{2q-1}^{q-1} \equiv 1 \pmod p$. Аналогично $C_{2p-1}^{p-1} \equiv 1 \pmod q$.

В обратную сторону утверждение очевидно.

5 Суммы биномиальных коэффициентов

5.1. a) Это сразу следует из результата задачи 1.3. Если Δ_0^0 — это треугольник из трех первых строк 3-арифметического треугольника Паскаля, то, как нетрудно видеть сумма центральных коэффициентов в нем делится на 3. При произвольном a изучаемая сумма содержит элементы нескольких центральных треугольников, кратных Δ_0^0. Поэтому сумма тоже делится на 3.

Другое решение ([CSTTVZ]) получится, если мы воспользуемся тождеством $C_{2k}^k = \sum\limits_{i=0}^{k} {C_k^i}^2$. Тогда $\sum\limits_{k=0}^{3^a-1} C_{2k}^k = \sum\limits_{k=0}^{3^a-1}\sum\limits_{i=0}^{k} {C_k^i}^2$. Поскольку $1^2 = 2^2 = 1$, $0^2 = 0$ по модулю 3, последняя сумма равна по модулю 3 количеству ненулевых элементов в первых 3^a строках треугольника Паскаля. Это количество подсчитано в задаче 2.1a), оно делится на 3.

b) Приводим решение [Д]. Нужная нам сумма является коэффициентом при x^{3^a-1} многочлена

$$x^{3^a-1}\left(1+\frac{(x+1)^2}{x}+\frac{(x+1)^4}{x^2}+\ldots+\frac{(x+1)^{2(3^a-1)}}{x^{3^a-1}}\right)=\frac{\frac{(x+1)^{2\cdot 3^a}}{x^{3^a}}-1}{\frac{(x+1)^2}{x}-1}\cdot x^{3^a-1}=\frac{(x+1)^{2\cdot 3^a}-x^{3^a}}{x^2+x+1}=$$
$$=\frac{x^{2\cdot 3^a}+C_{2\cdot 3^a}^1\cdot x^{2\cdot 3^a-1}+C_{2\cdot 3^a}^2\cdot x^{2\cdot 3^a-2}+\ldots+1-x^{3^a}}{x^3-1}\cdot(x-1)\,.$$

Чтобы найти нужный коэффициент, достаточно поделить числитель на знаменатель “в столбик”, и потом домножить результат на $(x-1)$. Таким образом, не нужно даже доводить деление до конца, достаточно довести его до нахождения коэффициента при x^{3^a-2}, кроме того, напомним, результат нас интересует лишь по модулю 3^a. Отметим, что при $b \not\vdots 3$ все биномиальные коэффициенты $C_{2\cdot 3^a}^b$ делятся на 3^a по теореме Куммера. Сгруппируем слагаемые с этими коэффициентами и будем делить их сумму на x^3-1 отдельно. Очевидно, все коэффициенты частного будут тоже делиться на 3^a, поэтому все эти слагаемые можно отбросить. Остается выражение

$$\frac{x^{2\cdot 3^a}+C_{2\cdot 3^a}^3\cdot x^{2\cdot 3^a-3}+C_{2\cdot 3^a}^6\cdot x^{2\cdot 3^a-6}+\ldots+1-x^{3^a}}{x^3-1}\cdot(x-1).$$

Здесь все показатели в числителе делятся на 3, после деления на x^3-1 все показатели частного тоже будут делиться на 3, а когда мы домножим частное на $x-1$, у нас не появится ни одного показателя вида $3k+2$. Таким образом, искомый коэффициент по модулю 3^a равен 0.

5.2. Задача была опубликована в Monthly [25]. Так как

$$C_{2n+2}^{n+1}-4C_{2n}^n=2\cdot\frac{2n+1}{n+1}C_{2n}^n-4C_{2n}^n=-2C_n\,,$$

то $C_n \equiv C_{2n+2}^{n+1}-C_{2n}^n \pmod 3$. Поэтому сумма по модулю 3 является телескопической,

$$\sum_{k=1}^{n} C_k \equiv \left(C_{2n+2}^{n+1}-C_{2n}^n\right)+\left(C_{2n}^n-C_{2n-2}^{n-1}+\ldots\right)=C_{2n+2}^{n+1}+1 \pmod 3.$$

Таким образом, по теореме Куммера нам остается выяснить, в каком случае сложение числа $(n+1)$ с самим собой в троичной системе счисления приводит к появлению хотя бы одного переноса. Очевидно, это может быть в том и только том случае, когда в записи $n+1$ есть хотя бы одна двойка.

5.3. Это задача A5 Putnam Mathematical Competition, 1998. Поскольку $\frac{1}{p}C_p^n \equiv \frac{(-1)^{n-1}}{n} \pmod p$, получаем, что

$$\sum_{n=1}^{k}\frac{1}{p}C_p^n \equiv \sum_{n=1}^{k}\frac{(-1)^{n-1}}{n}=\sum_{n=1}^{k}\frac{1}{n}-2\sum_{n=1}^{[k/2]}\frac{1}{2n}\equiv\sum_{n=1}^{k}\frac{1}{n}+\sum_{n=p-[\frac{k}{2}]}^{p-1}\frac{1}{n}\overset{*}{=}\sum_{n=1}^{p-1}\frac{1}{n}\equiv 0 \pmod p.$$

В сумме, расположенной непосредственно слева от равенства, помеченного звездочкой, на самом деле суммирование ведется от $n = k + 1$ (в этом нетрудно убедиться: при $p = 6r + 1$ имеем $k = 4r$ и $p - [\frac{k}{2}] = 4r + 1 = k + 1$, аналогично при $p = 6r + 5$).

5.4. Это утверждение из [11]. Решение [CSTTVZ]. Индукция по n. База тривиальна. Докажем переход от $n' = n - (p - 1)$ к n. Пусть $q = \frac{n}{p-1}$. Так как

$$C_{n'+p-1}^{x(p-1)} = \sum_{i=0}^{p-1} C_{p-1}^{i} C_{n'}^{x(p-1)-i},$$

мы можем записать изучаемую сумму в виде

$$C_n^{p-1} + C_n^{2(p-1)} + C_n^{3(p-1)} + \ldots = \sum_{x=1}^{q} \sum_{i=0}^{p-1} C_{p-1}^{i} C_{n'}^{x(p-1)-i} = \sum_{i=0}^{p-1} \left(C_{p-1}^{i} \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} \right) \tag{11}$$

По утверждению задачи 1.1 а) $C_{p-1}^{i} \equiv (-1)^i \pmod p$, пусть $C_{p-1}^{i} = ap + (-1)^i$. По утверждению задачи 1.6 при $i = 0, 1, \ldots, p-2$ выполнено сравнение $\sum\limits_{x=1}^{q} C_{n'}^{x(p-1)-i} \equiv C_{p-1}^{i} \equiv (-1)^i \pmod p$; пусть $\sum\limits_{x=1}^{q} C_{n'}^{x(p-1)-i} = bp + (-1)^i$. Тогда

$$C_{p-1}^{i} \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} = \big(ap + (-1)^i\big)\big(bp + (-1)^i\big) \equiv 1 + (-1)^i(ap + bp) =$$
$$= 1 + (-1)^i \left(C_{p-1}^{i} + \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} - 2 \cdot (-1)^i \right) = (-1)^i \left(C_{p-1}^{i} + \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} \right) - 1 \pmod{p^2}.$$

Напомним, что это преобразование верно при $0 \leqslant i \leqslant p - 2$. Мы можем продолжить равенство (11), выделив отдельное слагаемое для $i = p - 1$:

$$\sum_{i=0}^{p-1} \left(C_{p-1}^{i} \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} \right) \equiv \sum_{i=0}^{p-2} \left((-1)^i \left(C_{p-1}^{i} + \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} \right) - 1 \right) + \sum_{x=0}^{q-1} C_{n'}^{x(p-1)} =$$
$$= \sum_{i=0}^{p-2} (-1)^i C_{p-1}^{i} + \sum_{i=0}^{p-2} \left((-1)^i \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} \right) - (p-1) + C_{n'}^{0} + \sum_{x=1}^{q-1} C_{n'}^{x(p-1)}.$$

Здесь первая сумма равна -1, так как знакопеременная сумма $C_{p-1}^{0} - C_{p-1}^{1} + C_{p-1}^{2} + \ldots$ равна 0. По той же причине вторая (двойная) сумма вместе со слагаемым $C_{n'}^{0}$ равна 0. Последняя же сумма по предположению индукции равна $1 + p(n' + 1)$. Итого все выражение равно $-1 + 0 - p + 1 + 1 + p(n' + 1) = 1 + pn'$. Это как раз то, что требуется, поскольку $1 + p(n + 1) = 1 + p(n' + p - 1 + 1) \equiv 1 + pn' \pmod{p^2}$.

5.5. Это результат Флека, 1913 г., мы узнали о нем из [18]. Решение [CSTTVZ].

При $p = 2$ сумма не знакопеременная и результат очевиден. Далее считаем, что p нечетно. Индукция по q. База следует из утверждения задачи 2.5 а). Докажем переход от $n' = n - (p-1)$ к n. Ниже выражение $\sum_x$ обозначает суммирование по x в естественных границах (т. е. в границах для которых определены биномиальные коэффициенты под знаком суммирования).

$$\pm \sum_{m:m \equiv j \pmod p} (-1)^m C_n^m = \sum_x (-1)^x C_{n'+p-1}^{xp+j} = \sum_x (-1)^x \sum_{i=0}^{p-1} C_{p-1}^{i} C_{n'}^{xp+j-i} = \sum_{i=0}^{p-1} C_{p-1}^{i} \sum_x (-1)^x C_{n'}^{xp+j-i}$$

По предположению индукции $\sum\limits_x (-1)^x C_{n'}^{xp+j-i}$ делится на p^{q-1}, по утверждению задачи 1.1 а) $C_{p-1}^{i} \equiv (-1)^i \pmod p$, следовательно,

$$\sum_{i=0}^{p-1} C_{p-1}^{i} \sum_x (-1)^x C_{n'}^{xp+j-i} \equiv \sum_{i=0}^{p-1} (-1)^i \sum_x (-1)^x C_{n'}^{xp+j-i} \pmod{p^q}.$$

Внимательно посмотрев на последнюю двойную сумму, можно заметить, что это она равна $C_{n'}^{0} - C_{n'}^{1} + C_{n'}^{2} - C_{n'}^{3} + \ldots = 0$.

5.6. Это результат Баскарана (1965 г.), мы взяли его в [18], решение [CSTTVZ].

Обозначим

$$f(n,j) = C_n^j - C_n^{j+(p-1)} + C_n^{j+2(p-1)} - C_n^{j+3(p-1)} + \dots$$

Индукция по n. База $n = p+1$ тривиальна, отметим лишь, что $C_{p+1}^i \equiv 1 \pmod p$ при $i = 0, 1, p, p+1$, а в остальных случаях этот биномиальный коэффициент делится на p. Докажем индукционный переход от $n' = n - (p+1)$ к n. Благодаря сделанному замечанию,

$$C_{n'+(p+1)}^{j+(p-1)k} = \sum_{i=0}^{p+1} C_{n'}^{j+(p-1)k-i} C_{p+1}^i \equiv \sum_{i\in\{0,1,p,p+1\}} C_{n'}^{j+(p-1)k-i} =$$

$$= C_{n'}^{j+(p-1)k} + \underline{C_{n'}^{j-1+(p-1)k} + C_{n'}^{j-1+(p-1)(k-1)}} + C_{n'}^{j-2+(p-1)(k-1)} \pmod p .$$

Поскольку $f(n,j) = \sum_k (-1)^k C_n^{j+k(p-1)}$ — знакочередующаяся сумма, при суммировании по k подчеркнутые выражения сократятся в типовом слагаемом (а несократившиеся выражения в крайних слагаемых равны 0 по причине некорректности биномиального коэффициента). Таким образом, мы получаем соотношения

$$f(n,j) \equiv f(n',j) - f(n',j-2) \quad \text{при } j > 1 , \qquad f(n,1) \equiv f(n',1) + f(n',p-2) .$$

Теперь часть "только тогда" доказываемого утверждения сразу следует из индукционного предположения, а часть "тогда" в общем-то тоже: если $f(n,j) \equiv 0 \pmod p$ при $j = 1, 3, \dots, p-2$, то

$$f(n',p-2) \equiv f(n',p-4) \equiv \dots \equiv f(n',1) \equiv -f(n',p-2) ,$$

откуда $f(n',j) \equiv 0 \pmod p$ при всех нужных j и тогда $n' \vdots (p+1)$, а тогда и $n \vdots (p+1)$.

Ссылки

Авторы многих приведенных решений — участники конференции, в таких решениях мы ставили ссылки:

[Д] Максим Дидин;
[K] Дмитрий Креков;
[J] Jastin Lim Kai Ze;
[T] Teh Zhao Yang Anzo;
[CSTTVZ] Ćevid Domagoj, Stokić Maksim, Tanasijeviić Ivan, Trifunović Petar, Vukorepa Borna, Žikelić Đorđe

Литература

[1] *Винберг Э. Б.* Удивительные свойства биномиальных коэффициентов. // Мат. просвещение. Третья серия. Вып. 12. 2008

[2] *Гашков С.Б., Чубариков В.Н.* Арифметика. Алгоритмы. Сложность вычислений. М.: Высш. шк., 2000.

[3] *Дынкин Е.Б., Успенский В.А.* Математические беседы. 2-е изд. М.: ФИЗМАТЛИТ, 2004.

[4] Петербургские математические олимпиады, 1961–1993. СПб: Лань, 2007.

[5] *Табачников С.Л., Фукс Д.Б.* Математический дивертисмент. 30 лекций по классической математике. М.: МЦНМО, 2011.

[6] *Фукс Д.Б., Фукс М.Б.* Арифметика биномиальных коэффициентов // Квант. 1970. №6. С. 17–25.

[7] *Ширшов А.И.* Об одном свойстве биномиальных коэффициентов // Квант. 1971. №10. С. 16–20.

[8] *Cai T.X., Granville A.* On the residues of binomial coefficients and their products modulo prime powers /!/ Acta

[9] *Calkin N. J.* Factors of sums of powers of binomial coefficients // Acta Arith. 1998. Vol. 86. P. 17–26.

[10] *Carlitz L.* A note of Wolstenholme's theorem // Amer. Math. Monthly. 1954. Vol. 61. № 3. P. 174–176.

[11] *Dimitrov V., Chapman R.* Binomial coefficient identity: 11118 // Amer. Math. Monthly. 2006. Vol. 113. № 7. P. 657–658.

[12] *Everett W.* Subprime factorization and the numbers of binomial coefficients exactly divided by powers of a prime // Integers. 2011. Vol. 11. # A63. `http://www.integers-ejcnt.org/vol11.html`

[13] *Fine N.* Binomial coefficient modulo a prime // Amer. Math. Monthly. 1947. Vol. 54. № 10. Part 1. P. 589–592.

[14] *Gardiner A.* Four problems on prime power divisibility // Amer. Math. Monthly. 1988. Vol. 95. № 10. P. 926–931.

[15] *Gauss K.* Disquisitiones arithmeticae. 1801. Art. 78.

[16] *Gessel I.* Wolstenholme revisited // Amer. Math. Monthly. 1998. Vol. 105. № 7. P. 657–658.

[17] *Granville A.* Arithmetic properties of binomial coefficients. Доступно по адресу `http://www.dms.umontreal.ca/~andrew/Binomial/`

[18] *Granville A.* Binomial coefficients modulo prime powers.

[19] *Granville A.* Zaphod Beeblebrox's Brian and the Fifty-ninth Row of Pascal's Triangle // Amer. Math. Monthly. 1992. Vol. 99. № 4. P. 318–331.

[20] *Granville A.* Correction to: Zaphod Beeblebrox's Brian and the Fifty-ninth Row of Pascal's Triangle // Amer. Math. Monthly. 1997. Vol. 104. № 9. P. 848–851.

[21] *Hinz A.* Pascal's triangle and tower of Hanoi // Amer. Math. Monthly. 1992. Vol. 99. № 6. P. 538–544.

[22] *Loveless A.* A congruence for products of binomial coefficients modulo a composite // Integers: electronic journal of comb. number theory 7 (2007) # A44

[23] *McIntosh R.* On the converse of Wolstenhome's theorem // Acta Arithmetica. 1995. Vol. 61. №4. P. 381–388.

[24] *Meštrović R.* On the mod p^7 determination of $\binom{2p-1}{p-1}$ // `http://arxiv.org/pdf/1108.1174v1.pdf`

[25] *More Y., Chapman R.* The sum of Catalan numbers, modulo 3: 11165 // Amer. Math. Monthly. 2007. Vol. 114. № 5. P. 454–455.

[26] *Morley F.* Note on the congruences $2^{4n} \equiv (-)^n(2n)!/(n!^2)$, where $2n+1$ is a prime // Annals of Math. 1894-1895. Vol. 9. №1. P. 168–170.

[27] *Roberts J.* On binomial coefficient residues // Canad J. Math. 1957. Vol. 9. P. 363–370.

[28] Sun Z.-W., Wan D. On Fleck quotients // `arXxiv:math.0603462v3`

Amazing properties of binomial coefficients

Several research topics will be set to you at the conference. Your aim is the maximal advance in one of these topics. You can co-operate in the solving of problems, arbitrary teams are allowed (i.e. the team may consist of participants from different cities). If you solve problems in different topics you may take part in different teams. The only thing you should avoid is to sign up the solutions of those problems that you really were not solving (this may happen if the team is too big and not all of its members solve the problems of some topic actively).

The following is the introductory set of problems about binomial coefficients. You may hand in the (written) solutions to Kokahs K. (coach 15, seat 17) In Teberda the set of problems will be enlarged a lot and you may hand in your solutions of this set of problems, except 1.2, in Teberda, too. You can hand in the solutions of the problem 1.2 in train only.

1 Problems for solving in train

1.1. Prove that a) $\binom{p-1}{k} \equiv (-1)^k \pmod p$; b) $\binom{2n}{n} \equiv (-4)^n \binom{\frac{p-1}{2}}{n} \pmod p$ при $n \leqslant \frac{p-1}{2}$.

1.2. Prove that the number of odd binomial coefficients in n-th row of Pascal triangle is equal to 2^r, where r is the number of 1's in the binary expansion of n.

1.3. Fix a positive integer m. By a m-arithmetical Pascal triangle we mean a triangle in which binomial coefficients are replaced by their residues modulo m. We will also consider similar triangles with the arbitrary residues a instead of 1's along the lateral sides of the triangle. The operation of the multiplying by a number and addition of triangles of equal size are correctly defined. We will consider these operations modulo m.

$$\begin{array}{ccccccccccc}
 & & a & & & & b & & & & a+b \\
 & a & & a & + & b & & b & = & a+b & & a+b \\
a & & 2a & & a & & b & 2b & b & a+b & 2(a+b) & & a+b \\
a & 3a & & 3a & a & b & 3b & & 3b & b & a+b & 3(a+b) & 3(a+b) & a+b
\end{array}$$

$$x \cdot \begin{array}{ccccc} & & a & & \\ & a & & a & \\ a & & 2a & & a \\ a & 3a & & 3a & a \end{array} = \begin{array}{ccccc} & & ax & & \\ & ax & & ax & \\ ax & & 2ax & & ax \\ ax & 3ax & & 3ax & ax \end{array}$$

Let all the elements of s-th row of m-arithmetical Pascal triangle except the first and the last one be equal to 0. Prove that the triangle has a form depicted on fig. 1. Shaded triangles consist of zeroes, triangles Δ_n^k consist of s rows and satisfy the following relations

1) $\Delta_n^{k-1} + \Delta_n^k = \Delta_{n+1}^k$; 2) $\Delta_n^k = C_n^k \cdot \Delta_0^0 \pmod m$.

The well known puzzle Tower of Hanoi consists of three rods, and a number of disks of different sizes which can slide onto any rod. The puzzle starts with the disks in a neat stack in ascending order of size on one rod, the smallest at the top, thus making a conical shape. The objective of the puzzle is to move the entire stack to another rod, obeying the following rules: 1) only one disk may be moved at a time; 2) each move consists of taking the upper disk from one of the rods and sliding it onto another rod, on top of the other disks that may already be present on that rod; 3) no disk may be placed on top of a smaller disk.

Let n be the number of disks. Let TH_n be a graph, whose vertices are all possible correct placements of disks onto 3 rods and edges connect placements that can be obtained one from another by 1 move. Consider also graph P_n, whose vertices are 1's located in the first 2^n rows of the 2-arithmetical Pascal triangle and edges connect neighboring 1's (i.e. two adjacent 1's in the same row or neighboring 1's by a diagonal in two adjacent rows)

1.4. prove that graphs TH_n and P_n are isomorphic.

1.5. Prove that that first 10^6 rows of 2-arithmetical Pascal triangle contain less than 1 % of 1's.

1.6. Prove that if n is divisible by $p-1$, then $\binom{n}{p-1} + \binom{n}{2(p-1)} + \binom{n}{3(p-1)} + \ldots + \binom{n}{n} \equiv 1 \pmod p$. Or, even better prove the general statement: if $1 \leqslant j, k \leqslant p-1$ и $n \equiv k \pmod{p-1}$, then

$$\binom{n}{j} + \binom{n}{(p-1)+j} + \binom{n}{2(p-1)+j} + \binom{n}{3(p-1)+j} + \ldots \equiv \binom{k}{j} \pmod p.$$

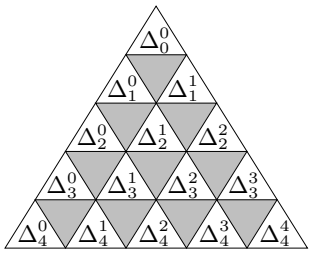


Рис. 1:

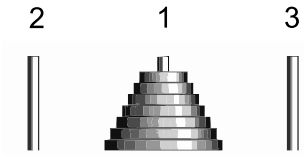


Рис. 2:

Amazing properties of binomial coefficients — 2

"The official theoretical source" for this set of problems is Vinberg's article [1]. Particularly the following theorems are considered to be known.

1. Wilson's theorem. For any prime p (and for primes only) the equivalence holds $(p-1)! \equiv -1 \pmod p$.
2. Lukas' theorem. Write the numbers n and k in base p:
$$n = n_d p^d + n_{d-1} p^{d-1} + \ldots + n_1 p + n_0, \qquad k = k_d p^d + k_{d-1} p^{d-1} + \ldots + k_1 p + k_0. \tag{1}$$
Then $\binom{n}{k} \equiv \binom{n_d}{k_d}\binom{n_{d-1}}{k_{d-1}} \cdot \ldots \cdot \binom{n_1}{k_1}\binom{n_0}{k_0} \pmod p$.
3. Kummer's theorem. The exponent $\operatorname{ord}_p \binom{n}{k}$ is equal to the number of "carries" when we add k and $\ell = n-k$ in base p.
4. Wolstenholme's theorem. If $p \geqslant 5$ then $\binom{2p}{p} \equiv 2 \pmod{p^3}$, or, that is the same, $\binom{2p-1}{p-1} \equiv 1 \pmod{p^3}$.

Remind that $\binom{0}{0} = 1$, $\binom{n}{k} = 0$ for $k > n$ and for $k < 0$ by definition.

We denote by p a prime number. For any natural n denote by $(n!)_p$ the product of all integers from 1 to n not divisible by p. If a number p is given the symbols n_i, m_i etc. denote the digits of numbers n, m etc. in base p.

* * *

2 Arithmetical triangle and divisibility

2.1. a) Prove that the first 3^k rows of 3-arithmetical Pascal triangle contain $\frac{1}{2}(6^k + 4^k)$ residues "1" and $\frac{1}{2}(6^k - 4^k)$ residues "2".

b) Find the number of zero elements in the first 5^k rows of 5-arithmetical Pascal triangle.

c) Find the number of non-zero elements in the first p^k rows of p-arithmetical Pascal triangle.

2.2. Prove that the number of 1's in the first m rows of 2-arithmetical Pascal triangle equals

$$\sum_{i=0}^{n-1} m_i \cdot 2^{\sum_{k=i+1}^{n-1} m_k} \cdot 3^i.$$

If $m = 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_r}$, where $\alpha_1 > \alpha_2 > \ldots > \alpha_r$, then we can rewrite the last expression in the form

$$3^{\alpha_1} + 2 \cdot 3^{\alpha_2} + 2^2 \cdot 3^{\alpha_3} + \ldots + 2^{r-1} \cdot 3^{\alpha_r}.$$

2.3. Consider n-th row of Pascal triangle modulo 2 as binary expansion of some integer P_n. Prove that

$$P_n = F_{i_1} \cdot \ldots \cdot F_{i_s},$$

where $i_1, \ldots, i_s$ are numbers of positions where 1's occur in the binary expansion of n, and $F_i = 2^{2^i} + 1$ is i-th Fermat number.

2.4. Prove that the number of non-zero elements in n-th row of p-arithmetical Pascal triangle equals $\prod_{i=0}^{d}(n_i + 1)$.

2.5. a) All the binomial coefficients $\binom{n}{k}$, where $0 < k < n$, are divisible by p if and only if n is a power of p.

b) All the binomial coefficients $\binom{n}{k}$, where $0 \leqslant k \leqslant n$, are not divisible by p if and only if $n+1$ is divisible by p^d, in other words, all the digits of n, except the leftmost, in base p are equal to $p-1$.

2.6. Let $0 < k < n+1$. Prove that if $\binom{n}{k-1} \not\vdots\, p$ and $\binom{n}{k} \not\vdots\, p$, then $\binom{n+1}{k} \not\vdots\, p$, *except* the case, when $n+1$ is divisible by p.

3 Generalization of Wilson's and Lukas' theorems

3.1. Prove that $\operatorname{ord}_p(n!) = \dfrac{n-(n_d+\ldots+n_1+n_0)}{p-1}$.

3.2. Prove the following generalizations of Wilson's theorem. a) $(-1)^{[n/p]}(n!)_p \equiv n_0! \pmod p$;
b) Prove that for $p \geqslant 3$

$$(p^q!)_p \equiv -1 \pmod{p^q},$$

and for $p=2$, $q \geqslant 3$ $\quad (p^q!)_p \equiv 1 \pmod{p^q}$.
c) $\dfrac{n!}{p^\mu} \equiv (-1)^\mu n_0! n_1! \ldots n_d! \pmod p$, where $\mu = \operatorname{ord}_p(n!)$

3.3. Generalized Lukas' theorem. Let $r = n-k$, $\ell = \operatorname{ord}_p\binom{n}{k}$. Then

$$\frac{1}{p^\ell}\binom{n}{k} \equiv (-1)^\ell \Big(\frac{n_0!}{k_0! r_0!}\Big)\Big(\frac{n_1!}{k_1! r_1!}\Big)\ldots\Big(\frac{n_d!}{k_d! r_d!}\Big) \pmod p$$

3.4. a) Prove that $(1+x)^{p^d} \equiv 1 + x^{p^d} \pmod p$ for all $x = 0, 1, \ldots, p-1$.
b) Prove Laukas' theorem algebraically.

3.5. a) Let m, n, k be nonnegative integers, and $(n,k)=1$. Prove that $C_{mn}^k \equiv 0 \pmod n$.
b) Prove that if $n \vdots p^k$, $m \not\vdots p$, then $\binom{n}{m} \vdots p^k$.

3.6. Let $f_{n,a} = \sum\limits_{k=0}^{n} \big(\binom{n}{k}\big)^a$. Prove that $f_{n,a} \equiv \prod\limits_{i=0}^{d} f_{n_i,a} \pmod p$.

4 Variations on Wolstenholme's theorem

4.1. Prove that $\dfrac{1}{1} + \dfrac{1}{2} + \ldots + \dfrac{1}{p-1} \equiv 0 \pmod{p^2}$.

4.2. Let $p = 4k+3$ be a prime number. Find $\dfrac{1}{0^2+1} + \dfrac{1}{1^2+1} + \ldots + \dfrac{1}{(p-1)^2+1} \pmod p$.

4.3. a) Let k be a nonnegative integer such that for any prime divisor p of the number m $\quad k$ is not divisible by $(p-1)$. Prove that $\dfrac{1}{1^k} + \dfrac{1}{2^k} + \ldots + \dfrac{1}{(p-1)^k} \equiv 0 \pmod m$ (summation over all fractions whose denominators are coprime to m).

b) Let k be odd and $(k+1) \not\vdots (p-1)$. Prove that $\dfrac{1}{1^k} + \dfrac{1}{2^k} + \ldots + \dfrac{1}{(p-1)^k} \equiv 0 \pmod{p^2}$.

4.4. Prove that the equivalence (12) from Vinberg's article holds in fact modulo p^4.

4.5. Prove that the following properties are equivalent 1) $\binom{2p-1}{p-1} \equiv 1 \pmod{p^4}$;
2) $\dfrac{1}{1} + \dfrac{1}{2} + \ldots + \dfrac{1}{p-1} \equiv 0 \pmod{p^3}$; 3) $\dfrac{1}{1^2} + \dfrac{1}{2^2} + \ldots + \dfrac{1}{(p-1)^2} \equiv 0 \pmod{p^2}$.

4.6. a) Prove algebraically that for any prime p and arbitrary k and n $\quad \big(\binom{pk}{pm} - \binom{k}{m}\big) \vdots p^2$. (In Vinberg's article this fact is proven combinatorially.
b) Prove the statement (9) form Vinberg's article: for any prime $p \geqslant 5$ and arbitrary k and n $\quad \big(\binom{pk}{pm} - \binom{k}{m}\big) \vdots p^3$.

4.7. Let $p \geqslant 5$. Prove that a) $\binom{p^2}{p} \equiv \binom{p}{1} \pmod{p^5}$; b) $\binom{p^{s+1}}{p} \equiv p^s \pmod{p^{2s+3}}$.

4.8. Prove that $\binom{p^3}{p^2} \equiv \binom{p^2}{p} \pmod{p^8}$.

Amazing properties of binomial coefficients — 3

Additional problems to previous topics

2.7. Prove that $\binom{p^n-1}{k} \equiv (-1)^{S_k} \pmod p$, where S_k is the sum of digits of k in base p.

2.8. Prove that if the binomial coefficient $\binom{n}{k}$ is odd i.e. $k_i \leqslant n_i$ for all $i = 0, 1, \ldots, d$ in the notations of (1), then

$$\binom{n}{k} \equiv \prod_{i=1}^{d} (-1)^{k_{i-1}n_i + k_i n_{i-1}} \pmod 4.$$

2.9. Prove that if there are no two consecutive 1's in the binary expansion of n then all the odd entries in n-th row $\equiv 1 \pmod 4$, otherwise the number of entries $\equiv 1 \pmod 4$ equals the number of entries $\equiv -1 \pmod 4$.

2.10. Prove that the number of 5's in each row of 8-arithmetical Pascal triangle is a power of 2. Prove the same for 1's, 3's and 7's.

2.11. Prove that if we consider all the elements of the two sets

$$\left\{\binom{2^n-1}{1}, \binom{2^n-1}{3}, \binom{2^n-1}{5}, \ldots, \binom{2^n-1}{2^n-1}\right\} \quad \text{and} \quad \{1, 3, 5, \ldots, 2^n-1\}$$

as a reminders modulo 2^n, then these sets coincide.

2.12. Prove that elements of a row of Pascal triangle are not coprime in the following sence. For any $\varepsilon > 0$ there exists N, such that for all integer $n > N$ and $k_1, k_2, \ldots, k_{100} < \varepsilon\sqrt{n}$ the numbers

$$\binom{2n}{n+k_1}, \binom{2n}{n+k_2}, \ldots, \binom{2n}{n+k_{100}}$$

have a common divisor.

2.13. a) The non negative numbers $m > 1$, n, k are given. Prove that at least one of the numbers $\binom{n}{k}$, $\binom{n+1}{k}$, $\ldots$, $\binom{n+k}{k}$ is not divisible by m.

b) Prove that for each k there exist infinite set of numbers n, such that all the numbers $\binom{n}{k}$, $\binom{n+1}{k}$, $\ldots$, $\binom{n+k-1}{k}$ are divisible by m.

4.9. Prove that for $n > 1$ $\binom{2^{n+1}}{2^n} - \binom{2^n}{2^{n-1}}$ is divisible by 2^{2n+2}.

4.10. Prove that for $p \geqslant 5$ $(-1)^{\frac{p-1}{2}}\binom{p-1}{\frac{p-1}{2}} \equiv 4^{p-1} \pmod{p^3}$.

Amazing properties of binomial coefficients — 4

Additional problems to previous topics

4.11. Let m be a non negative integer, $p \geqslant 5$ be a prime. Prove that

$$\frac{1}{mp+1} + \frac{1}{mp+2} + \cdots + \frac{1}{mp+(p-1)} \equiv 0 \pmod{p^2}.$$

4.12. Let p and q be primes. Prove that $\binom{2pq-1}{pq-1} \equiv 1 \pmod{pq}$ if and only if $\binom{2p-1}{p-1} \equiv 1 \pmod{q}$ and $\binom{2q-1}{q-1} \equiv 1 \pmod{p}$.

5 Sums of binomial coefficients

5.1. a) Prove that the sum $\sum\limits_{k=0}^{3^a-1} \binom{2k}{k}$ is divisible by 3; b) is divisible by 3^a.

5.2. Let $C_k = \frac{1}{k+1}\binom{2k}{k}$ be Catalan numbers. Prove that $\sum\limits_{k=1}^{n} C_k \equiv 1 \pmod 3$ if and only if the number $n+1$ contains at least one digit "2" in base 3.

5.3. Let $p \geqslant 3$, $k = [2p/3]$. Prove that the sum $\binom{p}{1} + \binom{p}{2} + \ldots + \binom{p}{k}$ is divisible by p^2.

5.4. Let $n \vdots (p-1)$, where p is an odd prime. Prove that

$$\binom{n}{p-1} + \binom{n}{2(p-1)} + \binom{n}{3(p-1)} + \ldots \equiv 1 + p(n+1) \pmod{p^2}.$$

5.5. Prove that if $0 \leqslant j \leqslant p-1 < n$ and $q = \frac{n-1}{p-1}]$ then

$$\sum_{m:m\equiv j \pmod p} (-1)^m \binom{n}{m} \equiv 0 \pmod{p^q}.$$

5.6. Let p be an odd roime. Prove that $n \vdots (p+1)$ if and only if

$$\binom{n}{j} - \binom{n}{j+(p-1)} + \binom{n}{j+2(p-1)} - \binom{n}{j+3(p-1)} + \ldots \equiv 0 \pmod p$$

for all $j = 1, 3, \ldots, p-2$.

Solutions

1 Problems for solving in train

1.1. a) Solution 1. $\binom{p-1}{k} = \dfrac{(p-1)(p-2)\dots(p-k)}{1\cdot 2\cdots k} \equiv \dfrac{(-1)(-2)\dots(-k)}{1\cdot 2\cdots k} \equiv (-1)^k \pmod p$.

Solution 2. It is evident by the formula for binomial coefficients that $\binom{p}{i}$ is divisible by p when $1 \leqslant i \leqslant p-1$. Since $\binom{p-1}{k-1} + \binom{p-1}{k} = \binom{p}{k}$ and $\binom{p-1}{0} = 1 \equiv 1 \pmod p$, then $(\binom{p-1}{0} + \binom{p-1}{1}) \vdots p$, and therefore $\binom{p-1}{1} \equiv -1 \pmod p$. But $\binom{p-1}{1} + \binom{p-1}{2}$ is divisible by p also, hence $\binom{p-1}{2} \equiv 1 \pmod p$ etc.

б) This problem is taken from [3, problem 162]. Since the fractions $\binom{2n+2}{n+1}/\binom{2n}{n}$ and $\binom{\frac{p-1}{2}}{n+1}/\binom{\frac{p-1}{2}}{n}$ are highly reducible, the statement can be easily proven by induction. But we suggest a direct calculation from [3].

It easy to see that

$$\binom{2n}{n} = 2^n \cdot \frac{1\cdot 3\cdots(2n-1)}{n!}$$

and

$$1\cdot 3\cdots(2n-1) = (-1)^n(-1)(-3)\cdots(-2n+1) \equiv (-1)^n(p-1)(p-3)\cdots(p-2n+1) =$$

$$= (-1)^n 2^n \Big(\frac{p-1}{2}\Big)\Big(\frac{p-3}{2}\Big)\cdots\Big(\frac{p-2n+1}{2}\Big) = (-1)^n 2^n \Big(\frac{p-1}{2}\Big)\Big(\frac{p-1}{2}-1\Big)\cdots\Big(\frac{p-1}{2}-n+1\Big) =$$

$$= (-1)^n 2^n \frac{(\frac{p-1}{2})!}{(\frac{p-1}{2}-n)!} \pmod p.$$

Therefore $\binom{2n}{n} \equiv (-1)^n 4^n \dfrac{(\frac{p-1}{2})!}{n!(\frac{p-1}{2}-n)!} = (-4)^n \binom{\frac{p-1}{2}}{n} \pmod p$.

1.2. It follows directly from self-similar structure of an arithmetical Pascal triangle, that is described in the next problems. It follows from Lucas' theorem also, you can read the proof in [1].

1.3. We restrict ourselves with small contemplation, the full solution can be found in [3, problem 133].

Since the s-th row contains a long sequence of zeroes, then below these zeroes in $(s+1)$-th row we have the sequence of zeroes, too, (it is one element shorter than the upper sequence); in $(s+2)$-th row there are the sequence of zeroes also (it is one element shorter again) and so on. This explains the presence of the grey triangle below Δ_0^0 (fig. 1).

Further, the non-zero elements of the s-th row are equal to 1, hence the numbers situated along the sloped sides of the grey triangle all are 1's (due to the recurrence for binomial coefficients). So all the numbers along the sloped sides of the triangles Δ_1^0 and Δ_1^1 are 1's, and therefore both triangles are identical to Δ_0^0.

Now it is clear, what is the $(2s)$-th row of the triangle. The left- and the rightmost elements are 1's, all other elements equal 0, except the central element that is equal to 2, because it is a sum of the two upper 1's. Thus we obtain that two grey triangles are situated below $2s$-th row, the triangles Δ_2^0 and Δ_2^2 to the left and to the right of them are identical to Δ_0^0, and the triangle Δ_2^1 with 2's along its sloped sides is equal to $2\cdot\Delta_0^0$.

And so on.

1.4. This statement we found in [21], several facts about binomial coefficients are proven there via Tower of Hanoi and the graph TH_n.

Let a be the diameter of the upper disc on the first rod, b be the diameter of the upper disc on the second rod and c be the diameter of the upper disc on the third rod. W.l.o.g. $a < b < c$, then we have 3 possible moves in this configuration: from a to b or c and from b to c, we analogously have 3 moves if one rod is without discs. If all the discs are placed on one rod then we have 2 possible moves only; let A_1, A_2, A_3 denote the configurations of this type.

Observe that by the problem 1.2 all the elements of 2^s-th row of Pascal triangle are 1's. Therefore graph P_n has the rotational symmetry of the third order, because the recurrence $\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$, that allows us to construct the triangle from top to bottom, is equivalent in arithmetic modulo 2 to the recurrences $\binom{n}{k-1} = \binom{n}{k} + \binom{n+1}{k}$ and $\binom{n}{k} = \binom{n}{k-1} + \binom{n+1}{k}$, that allows us to construct the triangle from the low left

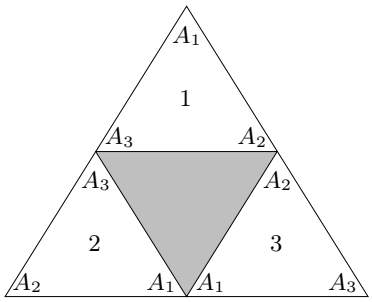


Рис. 3:

corner in the upper right direction and from the low right corner in the upper left direction. It follows also that the triangle of the double size contains 3 copies of the initial triangle.

Now let us prove by induction that there exists a bijection between TH_n and P_n, such that the vertices of the triangle P_n correspond to the configurations A_1, A_2, A_3. The base $n = 1$ is evident.

Proof of the step of induction. Assume that the bijection between TH_n and P_n has been constructed. The 2-arithmetical Pascal triangle with the side length 2^{n+1} contains 3 copies of the triangle with the side length 2^n. Number the copies and mark its vertices as shown on fig.3. Consider all the configurations of the Tower of Hanoi for which the $(n+1)$-th (biggest) disc is placed on rod i. If we fix the placement of this disc then displacements of other discs correspond to the graph that is isomorphic to TP_n. By induction hypothesis we can choose a bijection between this graph and the graph P_n in the i-th copy of the triangle, such that the configurations A_j correspond to the vertices of the triangles with the same marks. When we move the biggest disc, say, from the first rod to the second, all other discs must be on the 3rd rod. This move correspond to the edge connecting two neighboring vertices A_3 on the left sloped side of big triangle. The same reasons concern other moves of the biggest disc. Therefore we obtain an isomorphism between TP_{n+1} and P_n.

1.5. The bijection with Tower of Hanoi gives us a formula (when the number of rows is a power of 2): the first 2^k rows of 2-arithmetical Pascal triangle contain 3^k 1's. The formula can be also proved by induction via recurrence from the problem 1.3. Using this formula we can obtain an estimation. Since $10^6 < 2^{20}$, the total number of elements in these rows equals $\frac{1}{2} \cdot 10^6(10^6 + 1)$, and the number of 1's is at most 3^{20}. The proportion does not exceed $\frac{2 \cdot 3^{20}}{10^6(10^6+1)} \ll 0.01$.

1.6. We found this statement in [18].

Solution 1 ([CSTTVZ]). For $p = 2$ the statement can be easily checked. So we can assume that p is odd prime. Let $n = x(p-1) + k$. We use induction on x.

The base $x = 0$ is trivial: $\binom{k}{j} \equiv \binom{k}{j} \pmod p$.

To prove the step of induction we need the following property of binomial coefficients:

$$\binom{a+b}{s} = \sum_i \binom{a}{s-i}\binom{b}{i} \qquad \text{(summation in natural bounds)},$$

both sides of which calculate in how many ways we can choose s balls in the box that contains a black and b white balls. Let $n = m + (p-1)$. Observe that

$$\binom{n}{\ell(p-1)+j} = \binom{m+(p-1)}{\ell(p-1)+j} = \sum_{i=0}^{p-1} \binom{m}{\ell(p-1)+j-i}\binom{p-1}{i} \equiv \sum_{i=0}^{p-1} (-1)^i \binom{m}{\ell(p-1)+j-i} \pmod p$$

(the last equivalence is due to problem 1.1 a). Remark that the sign of the first and last terms in the last

sum is "plus" . Now transform the sum from the problem statement:

$$\sum_\ell \binom{n}{\ell(p-1)+j} \equiv$$
$$\equiv \left(\binom{m}{j} - \binom{m}{j-1} + \dots\right) + \left(\binom{m}{p-1+j} - \binom{m}{p-1+j-1} + \dots + \binom{m}{j}\right) +$$
$$+ \left(\binom{m}{2(p-1)+j} - \binom{m}{2(p-1)+j-1} + \dots + \binom{m}{2(p-1)+j}\right) + \dots$$
$$= \sum_{i=0}^{m} (-1)^i \binom{m}{i} + \sum_\ell \binom{m}{\ell(p-1)+j} \pmod p.$$

the first sum is equal to 0, the second sum is equivalent $\binom{k}{j} \pmod p$ by the induction hypothesis.

Solution 2 ([J], [T]). Induction by n. The base $n \leqslant p-1$ is trivial: both sides contain the same term. Prove the step of induction.

$$\binom{n}{j} + \binom{n}{(p-1)+j} + \dots = \left(\binom{n-1}{j} + \binom{n-1}{j-1}\right) + \left(\binom{n-1}{(p-1)+j} + \binom{n-1}{(p-1)+j-1}\right) + \dots =$$
$$= \left(\binom{n-1}{j} + \binom{n-1}{(p-1)+j} + \dots\right) + \left(\binom{n-1}{j-1} + \binom{n-1}{(p-1)+j-1} + \dots\right) \equiv$$
$$\equiv \binom{k-1}{j} + \binom{k-1}{j-1} = \binom{k}{j} \pmod p.$$

But it should be accurate in cases when $p-1$ divides j or k, because the induction hypothesis does not hold for $j=0$ or $k=0$ (it uses the value $p-1$ instead of 0). Therefore we must consider more carefully the cases when $j=1$ or $k=1$. We restrict ourselves by consideration of one partial case only. Let $p=5$, $j=1$ and we fulfill step to $n=13$. Then we have

$$\binom{1}{1} \stackrel{?}{\equiv} \binom{13}{1} + \binom{13}{6} + \binom{13}{11} = \left(\binom{12}{1} + \binom{12}{6} + \binom{12}{11}\right) + \left(\binom{12}{0} + \binom{12}{5} + \binom{12}{10}\right).$$

By induction hypothesis the sum in the first parentheses has a residue $\binom{4}{1}$ (and not $\binom{0}{1}$ as the previous calculation shows). In the second parentheses the induction hypothesis covers all the terms except the first one, so the sum has residue $\binom{12}{0} + \binom{4}{0}$. Writing $p-1$ instead of 4 for clarity, we obtain that the whole sum is equivalent to $\binom{n-1}{0} + \binom{p-1}{1} + \binom{p-1}{0} \equiv \binom{1}{1} \pmod p$, as required.

Solution 3 (algebraical reasoning with Luka's theorem, [18]). Induction by n. Base $n \leqslant p-1$ is trivial. Now let $n \geqslant p$, write all parameters in base p, let $\sigma_p(m)$ denotes the sum of digits of m. It is clear that if $m \equiv j \pmod p$, then $\sigma_p(m) \equiv j \pmod p$. The sum under consideration is equal by Luka's theorem to

$$\sum \binom{n_0}{m_0}\binom{n_1}{m_1}\dots\binom{n_d}{m_d} \pmod p,$$

where the summation is over all $m = \overline{m_d \dots m_1 m_0} \leqslant n$, for which $\sigma_p(m) \equiv j \pmod p$. This sum is equal to the sum of coefficients of x^j, x^{j+p-1}, $x^{j+2(p-1)}$, ... in the expression

$$(1+x)^{n_0}(1+x)^{n_1}\dots(1+x)^{n_d} = (1+x)^{\sigma_p(n)}.$$

But it is evident that this sum of coefficients equals

$$\sum_{\substack{1 \leqslant r \leqslant \sigma_p(n) \\ r \equiv j \pmod{p-1}}} \binom{\sigma_p(n)}{r},$$

which satisfy the induction hypothesis because $1 \leqslant \sigma_p(n) \leqslant n-1$, and supply the desired equivalence since $\sigma_p(n) \equiv n \equiv j \pmod p$.

Solution 4 (linear algebra, [D]). The polynomials x, x^2, ..., x^{p-1} are linearly independent over $\mathbb{Z}_p$ and form a basis in the space of functions $f\colon \mathbb{Z}_p \to \mathbb{Z}_p, f(0)=0$. By Fermat's little theorem $(1+x)^n \equiv (1+x)^k \pmod p$. Applying the relations $x^{i+a(p-1)} \equiv x^i$ to the left hand side, we obtain that our sum as an element of $\mathbb{Z}_p$ is equal to the coefficient of x^j in the right hand side, i. e. $\binom{k}{j}$.

2 *Arithmetical triangle and divisibility*

2.1. a) This result is due to Roberts [27]. By a_k denote the number of 1's in the first 3^k rows, and by b_k denote the number of 2's. Due to the recurrence from problem 1.3 we obtain

$$a_{k+1} = 5a_k + b_k, \qquad b_{k+1} = 5b_k + a_k.$$

Now the statement of problem follows by induction.

b) Answer: $\frac{1}{2} \cdot 5^k(5^k + 1) - 15^k$. By a_k denote the number of nonzero elements in the first 5^k rows. As in previous problem we have a recurrence

$$a_{k+1} = 15a_k + 10 \cdot \frac{5^k(5^k - 1)}{2}.$$

Since the whole triangle consists of $\frac{5^k(5^k+1)}{2}$ elements, it is natural to change variables $a_k = \frac{5^k(5^k+1)}{2} - b_k$. Then we can rewrite the previous relation in terms of b_k as $b_{k+1} = 15b_k$.

c) Answer: $\left(\frac{p(p+1)}{2}\right)^k$. This is Fine's result [13]. It can be obtained by induction by means of recurrence of the problem 1.3.

2.2. Solution 1. Induction by α_1. The base $\alpha_1 = 0, 1$ can be easily checked. Let the statement has been proven for all $\alpha_1 < a$. Prove it for $\alpha_1 = a$. Evedently $\tilde{m} - 2^{\alpha_1} < 2^{\alpha_1}$. Let $s = 2^{\alpha_1}$ (in notations of problem 1.3). Consider the $\tilde{m}$-th row in the triangle Δ_0^0, where $\tilde{m} = 2^{\alpha_2} + 2^{\alpha_3} + \ldots + 2^{\alpha_r}$. By the induction hypothesis the number of 1's in this row and above it equals

$$3^{\alpha_2} + 2 \cdot 3^{\alpha_3} + \ldots + 2^{r-2} \cdot 3^{\alpha_r}. \tag{2}$$

Then for the number $m = \tilde{m} + 2^{\alpha_1}$ we have a row that intersects the triangles Δ_0^1 and Δ_1^1 (due to 2-arithmetics they are both identical to triangle Δ_0^0). The part of Pascal triangle from top to this row contains triangle Δ_0^0 (containing 3^{α_1} 1's by induction hypothesis) and partially triangles Δ_0^1 and Δ_1^1 (the number of 1's in them is given by (2)). So the total number of 1's is

$$3^{\alpha_1} + 2(3^{\alpha_2} + 2 \cdot 3^{\alpha_3} + \ldots + 2^{r-2} \cdot 3^{\alpha_r}).$$

Solution 2 (combinatorial sense of coefficients, [T]).

Lemma 1. Let the k-th row contains 2^r 1's (or, equivalently, k contains r 1's in base 2) and let $\alpha_1 > \alpha_2 > \cdots > \alpha_m$, $2^{\alpha_m} > k$. Then the row with number $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_m} + k$ contains 2^{m+r} 1's.

Proof. It is clear that the number $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_m} + k$ in base 2 contains $m + r$ 1's and hence the corresponding row contains 2^{m+r} 1's. □

Lemma 2. The rows with the following numbers

$$2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}}, \quad 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + 1, \quad \ldots, \quad 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + 2^{\alpha_m} - 1,$$

contain $2^k 3^{\alpha_m}$ 1's.

Proof. By lemma 1 the row with number $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + i$ contains $2^k x_i$ 1's, where x_i is the number of 1's in i-th row. Then the total number of 1's in these rows equals $2^k \sum x_i$. But $\sum x_i$ is the number of 1's in the first $2^{\alpha_m} - 1$ rows of Pascal triangle, this number is equal to 3^{α_m} (it is known, for example, by problem 1.4). □

The statement of problem follows from lemma 2.

2.3. The problem is from [1], the solution is from [18]. The problem statement follows from Luka's theorem due to the following observation (it is also mentioned in [1]): a binomial coefficient $\binom{n}{k}$ is odd if and only if the set of 1's in the binary expansion of k is the subset of the set of 1's in the binary expansion of n. Therefore $P_n = \sum 2^k$, where the summation is over all k described in the previous phrase. For $p = 2$ let $S_n = \{i : n_i = 1\}$ in notations of formula (1). Then

$$P_n = \sum_{I \subseteq S_n} \prod_{i \in I} 2^{2^i} = \prod_{i \in S_n} F_i.$$

2.4. This result of Fine [13] (1947) is an easy corollary of Kummer's theorem. If p does not divide $\binom{n}{k}$, then there are no carries when we add k and $n-k$ in base p. For a fixed n it means that we can choose i-th digit of k in base p by n_i+1 ways.

2.5. a) It follows from the formula proven in the previous problem because here we have a row with 2 elements only not divisible by p.

b) [13]. If Если $p^d \mid (n+1)$, then $n = \overline{a(p-1)(p-1)\dots(p-1)}$ in base p. Then for any k, $0 \leqslant k \leqslant n$, each digit of k does not exceed the corresponding digit of n. Therefore all the binomial coefficients $\binom{n_i}{k_i}$ are not equal to 0 and $\not\equiv 0 \pmod p$. By Lukas' theorem $\binom{n}{k}$ is not divisible by p.

The reverse statement. Assume that all the coefficients $\binom{n}{k}$ are not divisible by p, but n is not the number of the form $\overline{a(p-1)(p-1)\dots(p-1)}$. Therefore one of its digits, say, n_i is less than $p-1$. Choose $k = (p-1)\cdot p^i$. Then $k_i = p-1$ and hence $\binom{n_i}{k_i} = 0$, and $p \mid \binom{n}{k}$ by Lukas' theorem. A contradicition.

2.6. This problem we found in [12].

Solution 1. Assume that $\binom{n}{k-1} \not\vdots\, p$ and $\binom{n}{k} \not\vdots\, p$, but $\binom{n+1}{k} = \left(\binom{n}{k-1} + \binom{n}{k}\right) \vdots\, p$. Then $\binom{n}{k} \equiv -\binom{n}{k-1} \pmod p$. Since both binomial coefficients are not divisible by p, we can reduce the equivalence and obtain $\frac{n-k+1}{k} \equiv -1 \pmod p$. Therefore $n+1 \equiv 0 \pmod p$.

Solution 2 ([K]). Though the statement remind us the main recurrence for binomial coefficients, the part "$\binom{n}{k-1} \not\vdots\, p$" is unnecessary. Indeed, if $(n+1) \not\vdots\, p$, then $0 \leqslant n_0 \leqslant p-2$. Since $\binom{n}{k} \not\vdots\, p$, then by Kummer's theorem $k_i \leqslant n_i$ for all i But analogous inequalities hold also for the pair k and $n+1$, because n and $n+1$ have the same digits except the lower ones that differs by 1. Hence $\binom{n+1}{k} \not\vdots\, p$.

2.7. [2]. It follows from Lukas' theorem and problem 1.1.a).

2.8. The problem is from [1]. Induction by number of digits. The base is trivial. For the proof of induction step add one more digit to the rightmost position. Since the binomial coefficient is odd we have the inequalities $n_i \geqslant k_i$. Now we will use the recurrence $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ and consider distinct variants of parity n и k. Applying Kummer's theorem and the problem 4.6a) we will reduce the question to the induction hypothesis.

For example, let $n = 2\ell+1$ be odd and $k = 2m$ be even. Consider a subcase $k_1 = 1$. Then we have binary representations $k = \dots 10$, $n = \dots 11$, $k-1 = \dots 01$ and $n-k = \dots 01$ (the latter because by Kummer's theorem there are no carries when we add k and $n-k$). Now when we add $k-1$ and $n-k$ we have 1 carry, i.e. $\binom{n-1}{k-1} \equiv 2 \pmod 4$, and hence

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \equiv -\binom{n-1}{k} = -\binom{2\ell}{2m} \equiv -\binom{\ell}{m} \pmod 4,$$

the latter equivalence is by problem 4.6a). The minus sign in it corresponds to the multiplier $(-1)^{k_0 n_1 + k_1 n_0}$.

2.9. The problem is from [1]. The statement follows from the previous problem. If the binary representation of n does not contains two consecutive 1's, then for all k all the exponents $k_{i-1}n_i + k_i n_{i-1}$ are equal to 0 and all the binomial coefficients in n-th row have are equivalent 1 modulo 4. But if the binary representation of n contains several consecutive 1's starting from $n_j = 1$ then the one half of all coefficients have $k_j = 0$, and one half of them have $k_j = 1$. By the formula of previous problem these two halves differ by a sign.

2.10. Two articles in Monthly [19, 20] discuss this dark problem.

2.11. This is a problem of D. Dzhukich was presented at the olympiad of 239 school of St.-Petersburg, 2002, and after that appeared at short-list of IMO-2008.

All the binomial coefficients in the problem statement are odd by Lukas' theorem, therefore, it is sufficient to check that all the numbers $\binom{2^n-1}{1}$, $\binom{2^n-1}{3}$, $\dots$, $\binom{2^n-1}{2^n-1}$ have distinct reminders modulo 2^n.

Solution 1 ([D]). Assume by the contrary that $\binom{2^n-1}{k} \equiv \binom{2^n-1}{m} \pmod{2^n}$ for odd k and m, $k > m$. Observe that

$$\binom{2^n-1}{k} = \binom{2^n}{k} - \binom{2^n-1}{k-1} = \binom{2^n}{k} - \binom{2^n}{k-1} + \binom{2^n-1}{k-2} = \dots =$$
$$= \binom{2^n}{k} - \binom{2^n}{k-1} + \binom{2^n}{k-2} - \dots - \binom{2^n}{m+1} + \binom{2^n-1}{m}.$$

In particular

$$\binom{2^n}{k} - \binom{2^n}{k-1} + \binom{2^n}{k-2} - \ldots - \binom{2^n}{m+1} \equiv 0 \pmod{2^n}.$$

Calculate the exponent $\operatorname{ord}_2 \binom{2^n}{r}$ by Kummer's theorem. If $\operatorname{ord}_2 r = a$ then we have $n-a$ carries in addition r and $2^n - r$ (it is clear by the standard algorithm of addition), hence $\operatorname{ord}_2 \binom{2^n}{r} = n - a$. In particular $2^n \mid \binom{2^n}{r}$ for odd r, that allows us to consider only one half of summands:

$$\binom{2^n}{k-1} + \binom{2^n}{k-3} + \ldots + \binom{2^n}{m+1} \equiv 0 \pmod{2^n}.$$

Now all the $\binom{2^n}{i}$ in the left hand side have even parameter i, therefore $\operatorname{ord}_2 \binom{2^n}{x} < n$.

We will prove that this congruence is impossible and obtain a contradiction. Choose x with minimal $\operatorname{ord}_2 \binom{2^n}{x}$. Since $\operatorname{ord}_2 \binom{2^n}{x} < n$ and the whole sum is divisible by 2^n, there exists y, for which $\operatorname{ord}_2 \binom{2^n}{x} = \operatorname{ord}_2 \binom{2^n}{y}$. Then the binary representations of x and y end with equal number of 0's, and hence there exists z between x and y which binary representation ends with bigger number of 0's. Then $\operatorname{ord}_2 \binom{2^n}{z} < \operatorname{ord}_2 \binom{2^n}{x}$, a contradiction.

Solution 2 ([CSTTVZ]). Induction by n. We prove the step of induction. Let the statement be proven for all numbers less than n. Assume by the contrary that there exist k and ℓ, $k \neq \ell$, $0 \leqslant k, \ell \leqslant 2^n - 1$, such that $\binom{2^n-1}{2k+1} \equiv \binom{2^n-1}{2\ell+1} \mod 2^n$. Observe that

$$\begin{aligned}
\binom{2^n-1}{2k+1} &= \left(\frac{2^n}{1}-1\right)\left(\frac{2^n}{2}-1\right)\ldots\left(\frac{2^n}{2k+1}-1\right) = \\
&= \left(\frac{2^n}{1}-1\right)\left(\frac{2^n}{3}-1\right)\ldots\left(\frac{2^n}{2k+1}-1\right)\cdot\left(\frac{2^{n-1}}{1}-1\right)\left(\frac{2^{n-1}}{2}-1\right)\ldots\left(\frac{2^{n-1}}{k}-1\right) = \qquad (3)\\
&= \left(\frac{2^n}{1}-1\right)\left(\frac{2^n}{3}-1\right)\ldots\left(\frac{2^n}{2k+1}-1\right)\cdot\binom{2^{n-1}-1}{k} \equiv \\
&\equiv (-1)^{k+1}\binom{2^{n-1}-1}{k} \pmod{2^n}
\end{aligned}$$

and analogously $\binom{2^n-1}{2\ell+1} \equiv (-1)^{\ell+1}\binom{2^{n-1}-1}{\ell} \pmod{2^n}$. It follows by induction hypothesis that both k and ℓ can not be odd. Besides, due to the symmetry $\binom{2^n-1}{r} = \binom{2^n-1}{2^n-1-r}$ the problem statement means that all the "even" binomial coefficients $\binom{2^n-1}{2r}$ are pairwise distinct modulo 2^n and form the same set of residues as "odd" binomial coefficients $\binom{2^n-1}{2r+1}$. Therefore k and ℓ can not be even simultaneously.

It remains to consider a case when k and ℓ have distinct parity, say $k = 2a+1$, $\ell = 2b$. Then

$$\binom{2^{n-1}-1}{2a+1} + \binom{2^{n-1}-1}{2b} \equiv 0 \pmod{2^n}.$$

If $a = b$ the congruence is impossible because $\binom{2^{n-1}-1}{2a}$ is odd and

$$\binom{2^{n-1}-1}{2a+1} + \binom{2^{n-1}-1}{2a} = \binom{2^{n-1}-1}{2a}\left(1 + \frac{2^{n-1}-1-2a}{2a+1}\right) = \binom{2^{n-1}-1}{2a}\cdot\frac{2^{n-1}}{2a+1} \equiv 2^{n-1} \pmod{2^n}.$$

If $b \neq a$, then $\binom{2^{n-1}-1}{2a} \neq \binom{2^{n-1}-1}{2b}$ by the induction hypothesis, Since $\binom{2^{n-1}-1}{2a} + \binom{2^{n-1}-1}{2a+1}$ is divisible by 2^{n-1}, the sum $\binom{2^{n-1}-1}{2b} + \binom{2^{n-1}-1}{2a+1}$ can not be divisible by 2^{n-1}.

2.12. The author of this problem is A. Belov. Observe that

$$\binom{2n}{n+k} = \binom{2n}{n}\cdot\frac{n(n-1)\ldots(n-k+1)}{(n+1)(n+2)\ldots(n+k)},$$

and therefore $\binom{2n}{n+k}$ have many common divisors with $\binom{2n}{n}$, because the denominator is not very big, more precisely, it does not exceed $(2n)^k$. Write the analogous equalities for all binomial coefficients $\binom{2n}{n+k_1}$,

$\binom{2n}{n+k_2}, \ldots, \binom{2n}{n+k_{100}}$. Then GCD of all denominators in the right hand sides of the equalities does not exceed $(n+1)(n+2)\ldots\left(n+[\varepsilon\sqrt{n}\,]\right) < (2n)^{\varepsilon\sqrt{n}}$. But for big n the binomial coefficient $\binom{2n}{n}$ is much greater, so after reducing by GCD the quotient is very big, and it divides all 100 binomial coefficients.

Explain more accurate the last reasoning. Observe that

$$\binom{2n}{n} = \frac{2n}{n}\cdot\frac{2n-1}{n-1}\ldots\frac{n+1}{1} > 2^n \qquad \text{and} \qquad (2n)^{100\varepsilon\sqrt{n}} = 2^{\varepsilon\sqrt{n}\log_2 n + \varepsilon\sqrt{n}}.$$

For each ε there exists N such that for all $n > N$ we have the equality $\frac{n}{2} > \varepsilon\sqrt{n}\log_2 n + \varepsilon\sqrt{n}$. If we reduce $\binom{2n}{n}$ by GCD for these n, the quotient is at least $2^{n/2}$.

2.13. a) The problem was presented at Leningrad olympiad, 1977.

Solution 1 (without Kummer's theorem). This is solution from the excellent book [4]. Assume that all these numbers are divisible by m. Then the numbers

$$\begin{gathered}\binom{n+k-1}{k-1} = \binom{n+k}{k} - \binom{n+k-1}{k},\\ \binom{n+k-2}{k-1} = \binom{n+k-1}{k} - \binom{n+k-2}{k},\\ \ldots\\ \binom{n}{k-1} = \binom{n+1}{k} - \binom{n}{k}\end{gathered}$$

are also divisible by m. Then analogously m divides all the numbers $\binom{n+i}{j}$, where $i \leqslant j$ are arbitrary nonnegative integers. But $\binom{n}{0}$ $(i = j = 0)$ is not divisible by m. A contradiction.

Solution 2 (Kummer's theorem). Let p be a prime divisor of m. Prove that at least one of the numbers $\binom{n}{k}, \binom{n+1}{k}, \ldots, \binom{n+k}{k}$ is not divisible by p. By Kummer's theorem if we choose ℓ $(n-k \leqslant \ell \leqslant n)$ such that the addition $k+\ell$ fulfills in base p without carries then the binomial coefficient $\binom{k+\ell}{k}$ is not divisible by p.

We will explain how to choose ℓ by giving a concrete example. Let $p = 7$, $k = 133$. We will write all the numbers in base 7. Since we try to choose ℓ in the set of $k+1$ numbers, we can always choose ℓ such that $k+\ell$ to be one of the following numbers

$$\ldots 133, \quad \ldots 233, \quad ,\ldots, \quad \ldots 633.$$

(Remind that 6 is the greatest digit in our example.) It is clear that the addition $k+\ell$ fulfills without carries.

b) We found this problem in [2]. It is not difficult to construct n by Kummer's theorem. Let $\operatorname{ord}_p m = s$, and k have $d+1$ digits in base p. Let $n \vdots p^{d+s+1}$. Then the representations of numbers $n-k$, $n-k+1$, $\ldots$, $n-1$ contain digits $(p-1)$ in positions from $(d+2)$ to $(d+s+2)$. When we add k to these numbers we have carries in these positions. Therefore by Kummer's theorem all the corresponding binomial coefficients are divisible by p^s.

Since it is not difficult to combine our reasoning for distinct p, the statemetn is proven.

3 Generalizations of Wilson's and Lukas' theorems

3.1. It is well known that $\operatorname{ord}_p(n!) = \sum_k \left[\frac{n}{p^k}\right]$. If $n = n_d p^d + n_{d-1}p^{d-1} + \ldots + n_1 p + n_0$ (representation in base p), then $\left[\frac{n}{p^k}\right] = n_d p^{d-k} + n_{d-1}p^{d-k-1} + \ldots + n_{k+1}p + n_k$ and we can rewrite the formula for $\operatorname{ord}_p(n!)$ in the form

$$\operatorname{ord}_p(n!) = \sum_{k=1}^{d}\left(\sum_{i=k}^{d} n_i p^{i-k}\right) = \sum_{i=1}^{d} n_i(p^{i-1}+p^{i-2}+\ldots+p+1) = \sum_{i=1}^{d} n_i\frac{p^i-1}{p-1} = \frac{\sum\limits_{i=0}^{d} n_i p^i - \sum\limits_{i=0}^{d} n_i}{p-1}.$$

This is exactly what we need.

3.2. a) Split the factors of $n!$ on groups of $(p-1)$ factors:

$$(n!)_p = \prod_{k=0}^{[\frac{n}{p}]-1}\big((kp+1)\cdot(kp+2)\cdots(kp+p-1)\big)\ \cdot\ \big([\tfrac{n}{p}]p+1\big)\big([\tfrac{n}{p}]p+2\big)\ldots\big([\tfrac{n}{p}]p+n_0\big)\equiv(-1)^{[\frac{n}{p}]}n_0!\pmod p\,.$$

б) This statement can be found in Gauss works [15]. The product $(p^q!)_p$ contains factors in pairs: a factor and its inverse modulo p^q, the product of each pair is 1 modulo p^q. So we need to watch on those factors m which equals to its inverse, this factors satisfy the congruence

$$m^2\equiv 1\pmod{p^q}.$$

For odd prime p the congruence has 2 solutions: ±1. For $p=2$, $q\geqslant 3$ the congruence has two more solutions: $2^{q-1}\pm 1$.

c) Since $n!=(n!)_p\cdot p^{[\frac{n}{p}]}\big([\frac{n}{p}]\big)!$, the statement can be proven by induction by means of the congruence of statement a) of this problem.

3.3. We found this problem on the web-page of A.Granville [17]. It well known Legendre's formula for the number ℓ is that

$$\ell=\operatorname{ord}_p\binom{n}{k}=\left(\left[\frac{n}{p}\right]-\left[\frac{k}{p}\right]-\left[\frac{r}{p}\right]\right)+\left(\left[\frac{n}{p^2}\right]-\left[\frac{k}{p^2}\right]-\left[\frac{r}{p^2}\right]\right)+\ldots \tag{4}$$

Denote $\tilde n=[n/p]$ for brevity and so forth, and collect all terms divisible by p in the the formula for a binomial coefficient:

$$\binom{n}{k}=\frac{(n!)_p}{(k!)_p(r!)_p}\cdot\frac{p^{[n/p]}}{p^{[k/p]}\cdot p^{[r/p]}}\cdot\frac{\tilde n!}{\tilde k!\cdot\tilde r!}.$$

By generalized Wilson's theorem (problem 3.2, b) the first fraction equals $\pm\frac{n_0!}{k_0!r_0!}\pmod p$, the third fraction allows us to apply induction, and the middle fraction (together with the sign of the first fraction) supply all the expressions containing ℓ by the formula (4).

3.4. a) Expand brackets in $(1+x)^{p^d}$ use the fact that $p\mid\binom{p^d}{k}$ for $1\leqslant k\leqslant p^d-1$ by Kummer's theorem.

b) Let $n=n'p+n_0$, $k=k'p+k_0$. By the previous statement $(1+x)^{pn'}\equiv(1+x^p)^{n'}\pmod p$. Then

$$(1+x)^n=(1+x)^{pn'}(1+x)^{n_0}\equiv(1+x^p)^{n'}(1+x)^{n_0}\pmod p.$$

This congruence means that we transform the coefficients of the polynomial modulo p. The coefficient of x^k at the l.h.s. equals $\binom{n}{k}$. All the exponents in the first brackets at the r.h.s. are divisible by p, hence the only way to obtain the term $x^{pk'+k_0}$ is multiplying the $x^{pk'}$ from the first bracket and x^{k_0} from the second. Thus we obtain $\binom{n'}{k'}\binom{n_0}{k_0}$ and so $\binom{n}{k}=\binom{n'}{k'}\binom{n_0}{k_0}$. Now Lukas' theorem follows by induction.

3.5. a, b) It follows from Kummer's theorem.

3.6. [9]. In the following calculation we use that $\binom{n_i}{k_i}=0$ for $k_i>n_i$; this allows us to apply Lukas' theorem and truncate a lot of summands:

$$f_{n,a}=\sum_{k=0}^{n}\binom{n}{k}^a\equiv\sum_{k_d=0}^{n_d}\sum_{k_{d-1}=0}^{n_{d-1}}\cdots\sum_{k_0=0}^{n_0}\prod_{i=0}^{d}\binom{n_i}{k_i}^a\equiv\prod_{i=0}^{d}\sum_{k_i=0}^{n_i}\binom{n_i}{k_i}^a\equiv\prod_{i=0}^{d}f_{n_i,a}\pmod p.$$

4 Variations on Wolstenholme's theorem

4.1. This is an exercise on reading an article. The statement is proven in article [1]. Observe that

$$2\sum_{i=1}^{p-1}\frac{1}{i}=\sum_{i=1}^{p-1}\frac{1}{i}+\frac{1}{p-i}=p\sum_{i=1}^{p-1}\frac{1}{i(p-i)}\,.$$

Hence the sum under consideration is divisible by p. Since $\frac{1}{i} \equiv -\frac{1}{p-i} \pmod p$, it remains to check that

$$\sum_{i=1}^{p-1} \frac{1}{i^2} \equiv 0 \pmod p.$$

But $\frac{1}{1^2}, \frac{1}{2^2}, \ldots, \frac{1}{(p-1)^2}$ modulo p is the same set as[1], что $1^2, 2^2, \ldots, (p-1)^2$. Therefore it is sufficient to prove that

$$\sum_{i=1}^{p-1} i^2 \equiv 0 \pmod p. \tag{5}$$

Let $\sum\limits_{i=1}^{p-1} i^2 \equiv s \pmod p$. It $p > 5$ we can always choose a, such that $a^2 \not\equiv 1 \pmod p$. Then the sets $\{1, 2, \ldots, p-1\}$ and $\{a, 2a, \ldots, (p-1)a\}$ coincide (the proof is the same as in the footnote) and

$$s \equiv \sum_{i=1}^{p-1} i^2 = \sum_{i=1}^{p-1} (ai)^2 = a^2 \sum_{i=1}^{p-1} i^2 \equiv a^2 s \pmod p.$$

Thus $s \equiv 0 \pmod p$.

4.2. Answer: $2k+2$. This problem of A. Golovanov was presented at Tuimaada-2012 olympiad. Observe that for $p = 4k+3$ the equation $x^2 + 1 = 0$ has no solutions in the set of residues modulo p, and hence the denominators of all fractions are non zero.

Solution 1. Let $a_i = i^2 + 1$, $i = 0, \ldots, p-1$. Then the expression equals

$$\frac{\sigma_{p-1}(a_0, a_1, \ldots, a_{p-1})}{\sigma_p(a_0, a_1, \ldots, a_{p-1})},$$

where σ_i is an elementary symmetrical polynomial of degree i. Find the polynomial for which the numbers a_i are its roots:

$$\prod_{i=0}^{p-1} (x - 1 - i^2).$$

Change the variable $x - 1 = t^2$ and obtain

$$\prod_{i=0}^{p-1} (t^2 - i^2) = \prod_{i=0}^{p-1} (t - i) \prod_{i=0}^{p-1} (t + i) \equiv (t^p - t)(t^p - t) = t^{2p} - 2t^{p+1} + t^2.$$

Now apply the inverse change of variables and obtain for $p = 4k+3$

$$\prod_{i=0}^{p-1} (x - 1 - i^2) \equiv (x-1)^p - 2(x-1)^{\frac{p+1}{2}} + (x-1) = x^p + \ldots + (p + 2 \cdot \tfrac{p+1}{2} + 1)x - 4.$$

By Viete's theorem $\sigma_p \equiv 4 \pmod p$, $\sigma_{p-1} \equiv 2 \pmod p$, therefore $\frac{\sigma_{p-1}}{\sigma_p} \equiv \frac{1}{2} \equiv 2k+2 \pmod p$.

Solution 2. Split all nonzero residues modulo p, except ± 1, on pairs of reciprocal. We obtain $2k$ pairs and in each pair (i, j)

$$ij \equiv 1 \quad \Leftrightarrow \quad i^2 j^2 \equiv 1 \quad \Leftrightarrow \quad (ij)^2 + i^2 + j^2 + 1 \equiv i^2 + j^2 + 2 \pmod p.$$

Therefore,

$$1 \equiv \frac{(ij)^2 + i^2 + j^2 + 1}{(i^2+1)(j^2+1)} \equiv \frac{i^2 + j^2 + 2}{(i^2+1)(j^2+1)} = \frac{1}{i^2+1} + \frac{1}{j^2+1} \pmod p.$$

So, the sum is equal to $\frac{1}{0^2+1} + \frac{1}{1^2+1} + \frac{1}{(-1)^2+1} + 2k \equiv 2k+2$.

[1] These sets coincide because they contain $p-1$ element each, and it is clear that all the reminders in each set are non zero and pairwise distinct.

Solution 3. By Fermat's little theorem the operations $x \mapsto x^{-1}$ and $x \mapsto x^{p-2}$ modulo p coincide. So it is sufficient to calculate the sum

$$\sum_{x=0}^{p-1}(x^2+1)^{p-2} = \sum_{x=0}^{p-1}\sum_{m=0}^{p-2}\binom{p-2}{m}x^{2m} = \sum_{m=0}^{p-2}\binom{p-2}{m}S_{2m}, \tag{6}$$

where $S_{2m} = \sum\limits_{x=0}^{p-1} x^{2m}$. Evidently $S_{2m} \equiv -1 \pmod p$ for $m = \frac{p-1}{2}$. Prove that $S_{2m} \equiv 0 \pmod p$ for all other $m \leqslant p-1$. Indeed, for each m we can choose a non zero residue a such that $a^{2m} \not\equiv 1 \pmod p$ and after that we can reason as in (5). For the sum (6) we have

$$\sum_{m=0}^{p-2}\binom{p-2}{m}S_{2m} \equiv -\binom{p-2}{\frac{p-1}{2}} = -\binom{4k+1}{2k+1} = -\frac{(4k+1)\cdot 4k\cdot\ldots\cdot(2k+1)}{1\cdot 2\cdot\ldots\cdot(2k+1)} \equiv$$
$$\equiv -\frac{(-2)\cdot(-3)\ldots(2k+2)}{1\cdot 2\cdot\ldots\cdot(2k+1)} \equiv 2k+2 \pmod p.$$

4.3. We found these statements in [16].

a) For each prime divisor $p \mid m$ choose a_p such that $p \nmid (a_p^k - 1)$. By the Chinese reminder theorem choose a such that $a \equiv a_p \pmod p$ for all p. Then the result can be proven by reasoning as in (5).

b) Observe that for odd k by the binomial formula we have $i^k + (p-i^k) \equiv ki^{k-1}p \pmod{p^2}$. Then

$$2\sum_{i=1}^{p-1}\frac{1}{i^k} = \sum_{i=1}^{p-1}\left(\frac{1}{i^k}+\frac{1}{(p-i)^k}\right) = \sum_{i=1}^{p-1}\frac{i^k+(p-i)^k}{i^k(p-i)^k} \equiv \sum_{i=1}^{p-1}\frac{ki^{k-1}p}{i^k(-i)^k} \equiv -kp\sum_{i=1}^{p-1}\frac{1}{i^{k+1}} \pmod{p^2}.$$

The sum in the r.h.s is divisible by p by the statement a).

4.4. The congruence holds even modulo p^7 (see [24]), but it goes a bit strong. We can reason as in [1], tracing all powers till p^4, and obtain

$$\binom{p-1}{2p-1} = \frac{(2p-1)(2p-2)\cdot\ldots\cdot(p+1)}{p!} = \left(\frac{2p}{1}-1\right)\left(\frac{2p}{2}-1\right)\cdot\ldots\cdot\left(\frac{2p}{p-1}-1\right) \equiv$$
$$\equiv 1 - 2p\sum_{i=1}^{p-1}\frac{1}{i} + 4p^2\sum_{\substack{i,j=1\\ i<j}}^{p-1}\frac{1}{ij} - 8p^3\sum_{\substack{i,j,k=1\\ i<j<k}}^{p-1}\frac{1}{ijk} \pmod{p^4}. \tag{7}$$

The last sum can be expressed via power sums:

$$\sum_{\substack{i,j,k=1\\ i<j<k}}^{p-1}\frac{1}{ijk} = \frac{S_3}{3} - \frac{S_1S_2}{2} + \frac{S_1^3}{6}, \quad \text{where } S_k = \sum_{i=1}^{p-1}\frac{1}{i^k}.$$

We now that S_1 and S_3 are divisible by p^2 (the latter due to problem 4.3b). Therefore the last term in the formula (7) can be omitted.

4.5. The problem is from [1], variations can be found in [14]. Since

$$2\sum_{k=1}^{p-1}\frac{1}{k^2} = \sum_{k=1}^{p-1}\Big(\frac{1}{k^2}+\frac{1}{(p-k)^2}\Big) = \sum_{k=1}^{p-1}\frac{k^2+(p-k)^2}{k^2(p-k)^2} \equiv -2\sum_{k=1}^{p-1}\frac{1}{k(p-k)} \pmod{p^2},$$

the statement 3) is equivalent to the congruence $\sum\limits_{k=1}^{p-1}\frac{1}{k(p-k)} \equiv 0 \pmod{p^2}$. The statement 2) is equivalent to the same congruence, because $2\sum\limits_{k=1}^{p-1}\frac{1}{k} = 2\sum\limits_{k=1}^{p-1}\Big(\frac{1}{k}+\frac{1}{p-k}\Big) = 2p\sum\limits_{k=1}^{p-1}\frac{1}{k(p-k)}$. Finally we know from the previous problem that

$$\binom{2p-1}{p-1} \equiv 1 - p^2\sum_{i=1}^{p-1}\frac{1}{i(p-i)} + 4p^2\sum_{\substack{i,j=1\\ i<j}}^{p-1}\frac{1}{ij} \pmod{p^4}.$$

So the statement 1) is equivalent to the congruence

$$\sum_{i=1}^{p-1} \frac{1}{i(p-i)} \equiv 4 \sum_{\substack{i,j=1 \\ i<j}}^{p-1} \frac{1}{ij} \pmod{p^2}. \tag{8}$$

Rewrite the expression in the r.h.s.:

$$4 \sum_{\substack{i,j=1 \\ i<j}}^{p-1} \frac{1}{ij} = 2\Big(\sum_{k=1}^{p-1} \frac{1}{i}\Big)^2 - 2\sum_{k=1}^{p-1} \frac{1}{k^2} \equiv 2\Big(\sum_{k=1}^{p-1} \frac{1}{i}\Big)^2 + 2\sum_{k=1}^{p-1} \frac{1}{k(p-k)}.$$

The sum in brackets is divisible by p, its square is divisible by p^2, and we can omit this term. Then from (8) we see that the statement 1) is equivalent to the congruence $\sum_{k=1}^{p-1} \frac{1}{k(p-k)} \equiv 0 \pmod{p^2}$.

4.6. a) Solution 1 ([5, proposition 2.12]). Induction on n. Expand brackets in the equality

$$(a+b)^{pn} = (a+b)^{p(n-1)}(a+b)^p$$

Equate the coefficients of $a^{pm}b^{p(n-m)}$:

$$\binom{pn}{pm} = \binom{p(n-1)}{pm}\binom{p}{0} + \binom{p(n-1)}{pm-1}\binom{p}{1} + \ldots + \binom{p(n-1)}{pm-p+1}\binom{p}{p-1} + \binom{p(n-1)}{pm-p}\binom{p}{p}.$$

All summands except first and last are divisible by p^2, because by Lucas' theorem each binomial coefficient is divisible by p. Hence

$$\binom{pn}{pm} \equiv \binom{p(n-1)}{pm} + \binom{p(n-1)}{p(m-1)} \pmod{p^2}.$$

By the induction hypothesis

$$\binom{p(n-1)}{pm} + \binom{p(n-1)}{p(m-1)} \equiv \binom{n-1}{m} + \binom{n-1}{m-1} \equiv \binom{n}{m} \pmod{p^2}.$$

Solution 2 ([D]). Prove that $\binom{kp}{mp} \equiv \binom{k}{m} \pmod{p^2}$ by induction on m.

To prove the base $m = 1$ we have to check that $\binom{pk}{p} - \binom{k}{1} \equiv 0 \pmod{p^2}$. We have

$$\binom{pk}{p} - \binom{k}{1} = \frac{pk(pk-1)\ldots(pk-p+1)}{p!} - k = \Big(\frac{(pk-1)(pk-1)\ldots(pk-p+1)}{(p-1)!} - 1\Big). \tag{9}$$

Split the multipliers in the numerator onto pairs:

$$(pk-i)(pk-p+i) \equiv pi^2 - i^2 \pmod{p^2}.$$

We see that the product modulo p^2 of each pair does not depend on k. Therefore the difference (9) modulo p^2 does not depend on k, too. Since it is equal to 0 for $k = 1$, it is equal to 0 for all k.

The step of induction. Let $\binom{kp}{(m-1)p} \equiv \binom{k}{m-1} \pmod{p^2}$. We have

$$\begin{aligned}\binom{kp}{mp} &= \binom{kp}{(m-1)p} \cdot \frac{(p(k-m)+1)(p(k-m)+1)\ldots(p(k-m)+p)}{pm(pm-1)\ldots(pm-p+1)} = \\ &= \binom{kp}{(m-1)p} \cdot \frac{(p(k-m)+1)(p(k-m)+1)\ldots(p(k-m)+p-1)}{(pm-1)\ldots(pm-p+1)} \cdot \frac{k-m+1}{m}. \end{aligned} \tag{10}$$

Remark that both fractions are correctly defined modulo p^2. As in the proof of base, the expression in the numerator of big fraction does not depend (modulo p^2) on k. Then we can put $k = 0$ for the calculating

the fraction modulo p^2 and obtain that it is congruent to 0. For the remaining part of the expression we can apply the induction hypothesis and obtain

$$\equiv \binom{k}{m-1} \cdot \frac{k-m+1}{m} = \binom{k}{m} \pmod{p^2}.$$

b) Solution 1 (combinatorial). As it has been suggested in [1], consider samples of kp objects from the set of pn objects. Let the initial set be split on blocks of p objects. The number of block samples equals $\binom{n}{k}$. Hence it remains to check that non block samples is divisible by p^3. But the number of non block samples with 3 or more blocks is divisible by p^3 (see [1]). For $k > 1$ every non block sample consists of at least 3 blocks, so in this case the statement is true. It remains to consider a case when $k = 1$ and we count the number of non block samples of p objects from the set of $2p$ objects. This number equals $\binom{2p}{p} - 2$, by Wolstenholme's theorem it is divisible by p^3.

Solution 2. In the formula $\binom{a}{b} = \frac{a(a-1)\dots(a-b+1)}{b(b-1)\dots1}$ split the numerator and the denominator onto blocks of p terms, reduce the first terms in each block, and collect the quotients in a separate expression:

$$\binom{mp}{kp} = \frac{m\not{p}\cdot(mp-1)\dots\big(mp-(p-1)\big)}{k\not{p}\cdot(kp-1)\dots\big(kp-(p-1)\big)} \cdot \frac{(m-1)\not{p}\cdot\big((m-1)p-1\big)\dots\big((m-1)p-(p-1)\big)}{(k-1)\not{p}\cdot\big((k-1)p-1\big)\dots\big((k-1)p-(p-1)\big)} \cdot \dots \times$$
$$\times \frac{(m-k+1)\not{p}\cdot\big((m-k+1)p-1\big)\dots\big((m-k+1)p-(p-1)\big)}{\not{p}\cdot(p-1)\dots1} =$$
$$= \binom{m}{k} \cdot \frac{(mp-1)\dots\big(mp-(p-1)\big)}{(kp-1)\dots\big(kp-(p-1)\big)} \cdot \dots \cdot \frac{\big((m-k+1)p-1\big)\dots\big((m-k+1)p-(p-1)\big)}{(p-1)\dots1}.$$

It remains to check that the product of fractions is congruent to 1 $\pmod{p^3}$. For this prove the congruence

$$\frac{(np-1)\dots\big(np-(p-1)\big)}{(rp-1)\dots\big(rp-(p-1)\big)} \equiv 1 \pmod{p^3}$$

or, even, it would be better to prove the following congruence

$$\frac{(np-1)\dots\big(np-(p-1)\big)}{(p-1)!} \equiv \frac{(rp-1)\dots\big(rp-(p-1)\big)}{(p-1)!} \pmod{p^3}.$$

This is true because both parts are congruent to 1 $\pmod{p^3}$, that can be shown analogously to the proof of Wolstenholme's theorem.

4.7. a) [5, theorem 2.14]. Transform the difference

$$\binom{p^2}{p} - \binom{p}{1} = \frac{p^2(p^2-1)\dots(p^2-(p-1))}{1\cdot2\cdot\dots\cdot(p-1)p} - p = \frac{p}{(p-1)!}\Big((1-p^2)(2-p^2)\dots((p-1)-p^2) - 1\cdot2\cdot\dots\cdot(p-1)\Big).$$

It remains to check that

$$(1-p^2)(2-p^2)\dots((p-1)-p^2) \equiv 1\cdot2\cdot\dots\cdot(p-1) \pmod{p^4}.$$

Expand brackets in the l.h.s.:

$$(1-p^2)(2-p^2)\dots((p-1)-p^2) = 1\cdot2\cdot\dots\cdot(p-1) + p^2\Big(1+\frac{1}{2}+\dots+\frac{1}{p-1}\Big)(p-1)! + \text{terms divisible by } p^4.$$

By the problem 4.1 the second summand is divisible by p^4.

b) Observe that $\binom{p^{s+1}}{p} = p^s \cdot \binom{p^{s+1}-1}{p-1}$, hence it is sufficient to prove that $\binom{p^{s+1}-1}{p-1} \equiv 1 \pmod{p^{s+3}}$.

$$\binom{p^{s+1}-1}{p-1} = \frac{(p^{s+1}-1)(p^{s+1}-2)\dots(p^{s+1}-(p-1))}{1\cdot2\cdots(p-1)} = \left(\frac{p^{s+1}}{1}-1\right)\left(\frac{p^{s+1}}{2}-1\right)\dots\left(\frac{p^{s+1}}{p-1}-1\right) \equiv$$
$$\equiv (-1)^{p-1} + p^{s+1}\Big(1+\frac{1}{2}+\dots+\frac{1}{p-1}\Big) \pmod{p^{s+3}}.$$

Since $(-1)^{p-1} = 1$ and $1 + \frac{1}{2} + \dots + \frac{1}{p-1} \equiv 0 \pmod{p^2}$ we are done.

4.8. The problem is from [1], we present solution [T].

$$\binom{p^3}{p^2}-\binom{p^2}{p}=p\left(\binom{p^3-1}{p^2-1}-\binom{p^2-1}{p-1}\right)=$$
$$=p\left(\left(\frac{p^3}{1}-1\right)\left(\frac{p^3}{2}-1\right)\dots\left(\frac{p^3}{p^2-1}-1\right)-\left(\frac{p^2}{1}-1\right)\left(\frac{p^2}{2}-1\right)\dots\left(\frac{p^2}{p-1}-1\right)\right)=$$
$$=p\left(\frac{p^2}{1}-1\right)\left(\frac{p^2}{2}-1\right)\dots\left(\frac{p^2}{p-1}-1\right)\left(\prod_{\substack{k=1\\ p\nmid k}}^{p^2-1}\left(\frac{p^3}{k}-1\right)-1\right).$$

It is sufficient to prove that the last bracket is divisible by p^7. Transform the product:

$$\prod_{\substack{k=1\\ p\nmid k}}^{p^2-1}\left(\frac{p^3}{k}-1\right)=\prod_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\left(\frac{p^3}{k}-1\right)\left(\frac{p^3}{p^2-k}-1\right)=\prod_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\left(\frac{p^6-p^5}{k(p^2-k)}+1\right)\equiv 1+p^5(p-1)\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k(p^2-k)}\pmod{p^7}.$$

Now we have to check that the last sum is divisible by p^2. This is true because by problem 4.3a)

$$\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k(p^2-k)}\equiv-\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k^2}\equiv 0\pmod{p^2}.$$

4.9. The statement is taken from [6, theorem 5], its generalization can be found in [7].

Solution 1 ([5, proposition 2.19]). Use the fact that the difference $\binom{2^{k+1}}{2^k}-\binom{2^k}{2^{k-1}}$ is equal to the coefficient of x^{2^k} in the polynomial

$$(1+x)^{2^{k+1}}-(1-x^2)^{2^k}=(1+x)^{2^k}\left((1+x)^{2^k}-(1-x)^{2^k}\right)=$$
$$=\left(1+\binom{2^k}{1}x+\binom{2^k}{2}x^2+\ldots+x^{2^k}\right)\cdot 2\left(\binom{2^k}{1}x+\binom{2^k}{3}x^3+\ldots+\binom{2^k}{2^k-1}x^{2^k-1}\right).$$

Since the second polynomial contains odd exponents only, the coefficient of x^{2^k} in the product equals

$$2\left(\binom{2^k}{1}\binom{2^k}{2^k-1}+\binom{2^k}{3}\binom{2^k}{2^k-3}+\ldots+\binom{2^k}{2^k-1}\binom{2^k}{1}\right).$$

By problem 3.5 b) 2^k divides each binomial coefficient in this expression, moreover each term occurs twice in the sum, and the sum itself is multiplied by 2. Thus all the expression is divisible by 2^{2k+2}.

Solution 2 ([CSTTVZ]). Since $\binom{2^{n+1}}{2^n}=2\binom{2^{n+1}-1}{2^n-1}$, it is sufficient to prove that

$$\binom{2^{n+1}-1}{2^n-1}\equiv\binom{2^n-1}{2^{n-1}-1}\pmod{2^{2n+1}}.$$

Similarly to (3) we obtain

$$\binom{2^{n+1}-1}{2^n-1}=\left(\frac{2^{n+1}}{1}-1\right)\left(\frac{2^{n+1}}{3}-1\right)\dots\left(\frac{2^{n+1}}{2^n-1}-1\right)\cdot\binom{2^n-1}{2^{n-1}-1}.$$

It is sufficient to prove that

$$L=\left(\frac{2^{n+1}}{1}-1\right)\left(\frac{2^{n+1}}{3}-1\right)\dots\left(\frac{2^{n+1}}{2^n-1}-1\right)\equiv 1\pmod{2^{2n+1}}.$$

This is true because

$$L \equiv (-1)^{2^{n-1}} - 2^{n+1}\left(\frac{1}{1}+\frac{1}{3}+\frac{1}{5}+\ldots+\frac{1}{2^n-1}\right) \equiv$$
$$\equiv 1 - 2^{n+1}\left(\frac{2^n}{1\cdot(2^n-1)}+\frac{2^n}{3\cdot(2^n-3)}+\ldots+\frac{2^n}{(2^{n-1}-1)(2^{n-1}+1)}\right) \equiv 1 \pmod{2^{2n+1}}.$$

4.10. This is theorem of Morley [26].

Solution 1 (author's proof, 1895). It goes a bit beyond the school curriculum.

Take the formula which expresses $\cos^{2n+1} x$ via cosines of multiple angles,[1] or, as they were saying in that times, write $\cos^{2n+1} x$ in the form handy for integrating:

$$2^{2n}\cos^{2n+1}x = \cos(2n+1)x + (2n+1)\cos(2n-1)x + \frac{(2n+1)\cdot 2n}{1\cdot 2}\cos(2n-3)x + \ldots + \frac{(2n+1)\cdot 2n\ldots(n+2)}{n!}\cos x.$$

Now integrate it[2] over the interval $[0, \frac{\pi}{2}]$:

$$2^{2n}\int \cos^{2n+1} x\,dx = \frac{\sin(2n+1)x}{2n+1} + \frac{2n+1}{2n-1}\sin(2n-1)x + \ldots,$$
$$2^{2n}\int_0^{\pi/2} \cos^{2n+1} x\,dx = (-)^n\left(\frac{1}{2n+1} - \frac{2n+1}{2n-1} + \ldots\right).$$

Every first grade student of university knows that it is convenient to use integration by parts for calculating this integral:

$$I_{2n+1} = \int_0^{\pi/2} \cos^{2n+1} x\,dx = \int_0^{\pi/2} \cos^{2n} x\cos x\,dx = \cos^{2n} x\sin x\Big|_0^{\pi/2} + 2n\int_0^{\pi/2} \cos^{2n-1} x\sin^2 x\,dx =$$
$$= 0 + 2n\int_0^{\pi/2} \cos^{2n-1} x(1-\cos^2 x)\,dx = 2n\cdot I_{2n-1} - 2n\cdot I_{2n+1},$$

therefore $I_{2n+1} = \frac{2n}{2n+1}\cdot I_{2n-1}$. Since $I_1 = 1$, we can apply the formula n times and obtain

$$\int_0^{\pi/2} \cos^{2n+1} x\,dx = \frac{2n\cdot(2n-2)\ldots 2}{(2n+1)(2n-1)\ldots 3}.$$

Equating of these two results give us the formula

$$2^{2n}\frac{2n\cdot(2n-2)\ldots 2}{(2n+1)(2n-1)\ldots 3} = (-)^n\left(\frac{1}{2n+1} - \frac{2n+1}{2n-1} + \ldots + \frac{(2n+1)\cdot 2n\ldots(n+2)}{n!}\right).$$

Let $p = 2n+1$ be a prime number. We obtain the desired congruence by multiplying the last formula by p:

$$2^{2n}\frac{2n\cdot(2n-2)\ldots 2}{(2n-1)(2n-3)\ldots 3} \equiv (-)^n \pmod{p^2}.$$

Solution 2 ([CSTTVZ]). We will use the following notations:

$$A = \sum_{i=1}^{\frac{p-1}{2}} \frac{1}{i}, \qquad B = \sum_{1\leqslant i<j\leqslant\frac{p-1}{2}} \frac{1}{ij}, \qquad C = \sum_{\substack{1\leqslant i\leqslant p-1\\ i \text{ is odd}}} \frac{1}{i}.$$

[1] The reader who is interested in question "from where do we take it" and not satisfied by the answer "from some text-book" may wish to use the Euler's formula $\cos\varphi = \frac{1}{2}(e^{i\varphi} + e^{-i\varphi})$ and raise its r.h.s in power $2n+1$ by the binomial formula.

[2] When we were learning the rules of multiplication, we just memorized that "minus by minus equals plus". In this formula we multiply signs. If we need to multiply n minuses, the record $(-)^n$ seems to be appropriate. So we leave the old-fashioned notation $(-)^n$, used by the author, instead of the modern one $(-1)^n$.

Then $A^2 = \sum\limits_{i=1}^{\frac{p-1}{2}} \frac{1}{i^2} + 2B \equiv 2B \pmod p$ by the problem 4.3b). So $A^2 \equiv 2B \pmod p$. Further,

$$2C + A = \sum_{\substack{1\leqslant i\leqslant p-1\\ i \text{ нечетно}}} \frac{2}{i} + \sum_{i=1}^{\frac{p-1}{2}} \frac{2}{2i} = \sum_{i=1}^{p-1} \frac{2}{i} \equiv 0 \pmod{p^2}.$$

So $C \equiv -\frac{1}{2}A \pmod{p^2}$.

Now transform modulo p^3 the parts of the given congruence. The l.h.s. is

$$(-1)^{\frac{p-1}{2}} \binom{p-1}{\frac{p-1}{2}} \equiv \Big(1-\frac{p}{1}\Big)\Big(1-\frac{p}{2}\Big)\dots\Big(1-\frac{p}{\frac{p-1}{2}}\Big) \equiv 1 - pA + p^2B \equiv 1 - pA + \frac{1}{2}p^2A^2 \pmod{p^3}.$$

For transforming the r.h.d observe that

$$2^{p-1} = \frac{2\cdot 4\cdots(p-1)}{1\cdot 2\cdots\frac{p-1}{2}} \cdot \frac{(p+1)\cdots(2p-2)}{\frac{p+1}{2}\cdots(p-1)} = \frac{(p+1)\cdots(2p-2)}{1\cdot 3\cdot 5\cdots(p-2)} =$$
$$= \Big(\frac{p}{1}+1\Big)\Big(\frac{p}{3}+1\Big)\dots\Big(\frac{p}{p-1}+1\Big) \equiv 1 + pC + \frac{1}{2}p^2C^2 \equiv 1 - \frac{1}{2}pA + \frac{1}{8}p^2A^2 \pmod{p^3}.$$

Then we have

$$4^{p-1} \equiv \Big(1 - \frac{1}{2}pA + \frac{1}{8}p^2A^2\Big)^2 \equiv 1 - pA + \frac{1}{4}p^2A^2 + 2\cdot\frac{1}{8}P^2A^2 = 1 - pA + \frac{1}{2}p^2A^2 \pmod{p^3}.$$

So the l.h.s. is congruent to the r.h.s.

4.11. We found this statement in [10].

$$\sum_{k=1}^{p-1} \frac{1}{mp+k} = \frac{1}{2}\sum_{k=1}^{p-1}\Big(\frac{1}{mp+k} + \frac{1}{mp+p-k}\Big) =$$
$$= p\cdot\frac{2m+1}{2}\cdot\sum_{k=1}^{p-1}\frac{1}{(mp+k)(mp+p-k)} \equiv -p\cdot\frac{2m+1}{2}\cdot\sum_{k=1}^{p-1}\frac{1}{k^2} \equiv 0 \pmod{p^2}.$$

4.12. We found this statement in [8]. Since $2pq-1 = (2q-1)p + p - 1$, the last digit of the number $2pq-1$ in base p is $p-1$, and the remaining digits form the number $2q-1$. Similarly the last digit of the number $pq-1$ in base p is $p-1$, and the remaining part forms the number $q-1$. By Lukas' theorem $\binom{2pq-1}{pq-1} \equiv \binom{2q-1}{q-1}\binom{p-1}{p-1} \equiv \binom{2q-1}{q-1} \pmod p$. On the other hand since $\binom{2pq-1}{pq-1} \equiv 1 \pmod{pq}$, then $\binom{2pq-1}{pq-1} \equiv 1 \pmod p$. So $\binom{2q-1}{q-1} \equiv 1 \pmod p$. Analogously $\binom{2p-1}{p-1} \equiv 1 \pmod q$.

The inverse statement is trivial.

5 Sums of binomial coefficients

5.1. a) It follows from problem 1.3. If Δ_0^0 is a triangle consisting of the first 3 rows of the 3-arithmetical Pascal triangle, then the sum of its central binomial coefficients is divisible by 3. For arbitrary a the sum under consideration contains elements of several central triangles, which are multiples of Δ_0^0. So the total sum is divisible by 3, too.

Another solution ([CSTTVZ]) we can derive from the identity $\binom{2k}{k} = \sum\limits_{i=0}^{k}\binom{k}{i}^2$. Then $\sum\limits_{k=0}^{3^a-1} C_{2k}^k = \sum\limits_{k=0}^{3^a-1}\sum\limits_{i=0}^{k}\binom{k}{i}^2$. Since $1^2 = 2^2 = 1$, $0^2 = 0 \pmod 3$, the last sum modulo 3 equals the number of nonzero elements in the first 3^a rows of the Pascal triangle. This number is calculated in the problem 2.1a), it is divisible by 3.

b) Solution of [D]. The sum is a coefficient of x^{3^a-1} in the polynomial

$$x^{3^a-1}\left(1+\frac{(x+1)^2}{x}+\frac{(x+1)^4}{x^2}+\ldots+\frac{(x+1)^{2(3^a-1)}}{x^{3^a-1}}\right)=\frac{\frac{(x+1)^{2\cdot 3^a}}{x^{3^a}}-1}{\frac{(x+1)^2}{x}-1}\cdot x^{3^a-1}=\frac{(x+1)^{2\cdot 3^a}-x^{3^a}}{x^2+x+1}=$$
$$=\frac{x^{2\cdot 3^a}+\binom{2\cdot 3^a}{1}\cdot x^{2\cdot 3^a-1}+\binom{2\cdot 3^a}{2}\cdot x^{2\cdot 3^a-2}+\ldots+1-x^{3^a}}{x^3-1}\cdot(x-1)\,.$$

In order to find this coefficient we will perform the long division of the numerator by the denominator and then multiply the result by $(x-1)$. We do not need to find the quotient at whole, it is sufficient to perform the division till the moment when the coefficient of x^{3^a-2} will be found, remind that we are trying to find this coefficient modulo 3^a only. Since for $b \not\vdots 3$ all the binomial coefficient $\binom{2\cdot 3^a}{b}$ are divisible by 3^a (by Kummer's theorem), we can collect all these coefficient in a separate sum. When we divide this sum by x^3-1 all the coefficients of the quotient are divisible by 3^a therefore we can discard this sum. The remaining expression is

$$\frac{x^{2\cdot 3^a}+\binom{2\cdot 3^a}{3}\cdot x^{2\cdot 3^a-3}+\binom{2\cdot 3^a}{6}\cdot x^{2\cdot 3^a-6}+\ldots+1-x^{3^a}}{x^3-1}\cdot(x-1).$$

All the exponents in the numerator are divisible by 3, hence after division by x^3-1 all the exponents of the quotient are divisible by 3, too, and after the multiplying it by $x-1$, there will be no exponents of the form $3k+2$. So the coefficient that we seek equals $0 \pmod{3^a}$.

5.2. This problem was published in Monthly [25]. Since

$$\binom{2n+2}{n+1}-4\binom{2n}{n}=2\cdot\frac{2n+1}{n+1}\binom{2n}{n}-4\binom{2n}{n}=-2C_n\,,$$

then $C_n \equiv \binom{2n+2}{n+1}-\binom{2n}{n} \pmod 3$. Therefore this sum is telescopic modulo 3:

$$\sum_{k=1}^{n}C_k\equiv\left(\binom{2n+2}{n+1}-\binom{2n}{n}\right)+\left(\binom{2n}{n}-\binom{2n-2}{n-1}+\ldots\right)=\binom{2n+2}{n+1}+1\pmod 3.$$

So by Kummer's theorem we have to clarify when we have at least one carry in the addition of the number $(n+1)$ with itself in base 3. It it clear that it happens only if $n+1$ contains at least one 2 in base 2.

5.3. This is problem A5 of Putnam Math. Competition, 1998. Since $\frac{1}{p}\binom{p}{n}\equiv\frac{(-1)^{n-1}}{n} \pmod p$, we have

$$\sum_{n=1}^{k}\frac{1}{p}\binom{p}{n}\equiv\sum_{n=1}^{k}\frac{(-1)^{n-1}}{n}=\sum_{n=1}^{k}\frac{1}{n}-2\sum_{n=1}^{[k/2]}\frac{1}{2n}\equiv\sum_{n=1}^{k}\frac{1}{n}+\sum_{n=p-[\frac{k}{2}]}^{p-1}\frac{1}{n}\overset{*}{=}\sum_{n=1}^{p-1}\frac{1}{n}\equiv 0\pmod p.$$

The summation in the sum to the left of asterisk really starts from $n=k+1$ (it is easy to check: for $p=6r+1$ we have $k=4r$ and $p-[\frac{k}{2}]=4r+1=k+1$, similarly for $p=6r+5$).

5.4. This statement is from [11]. Solution [CSTTVZ]. Induction on n. The base is trivial. Prove the induction step from $n'=n-(p-1)$ to n. Let $q=\frac{n}{p-1}$. Since

$$\binom{n'+p-1}{x(p-1)}=\sum_{i=0}^{p-1}\binom{p-1}{i}\binom{n'}{x(p-1)-i},$$

we can rewrite the sum under consideration in the form

$$\binom{n}{p-1}+\binom{n}{2(p-1)}+\binom{n}{3(p-1)}+\ldots=\sum_{x=1}^{q}\sum_{i=0}^{p-1}\binom{p-1}{i}\binom{n'}{x(p-1)-i}=$$
$$=\sum_{i=0}^{p-1}\left(\binom{p-1}{i}\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right).\quad(11)$$

By the problem 1.1 a) we have $\binom{p-1}{i} \equiv (-1)^i \pmod p$; let $\binom{p-1}{i} = ap + (-1)^i$. By the problem 1.6 we have $\sum\limits_{x=1}^{q} \binom{n'}{x(p-1)-i} \equiv \binom{p-1}{i} \equiv (-1)^i \pmod p$ for $i = 0, 1, \ldots, p-2$; let $\sum\limits_{x=1}^{q} \binom{n'}{x(p-1)-i} = bp + (-1)^i$. Then

$$\binom{p-1}{i}\sum_{x=1}^{q}\binom{n'}{x(p-1)-i} = \big(ap+(-1)^i\big)\big(bp+(-1)^i\big) \equiv 1 + (-1)^i(ap+bp) =$$
$$= 1+(-1)^i\left(\left(\binom{p-1}{i}+\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right) - 2\cdot(-1)^i\right) = (-1)^i\left(\left(\binom{p-1}{i}+\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right)\right)-1 \pmod{p^2}.$$

Remind that these transformations hold for $0 \leqslant i \leqslant p-2$. We can continue equality (11), by separating the summand for $i = p-1$:

$$\sum_{i=0}^{p-1}\left(\binom{p-1}{i}\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right) \equiv \sum_{i=0}^{p-2}\left((-1)^i\left(\binom{p-1}{i}+\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right)-1\right)+\sum_{x=0}^{q-1}\binom{n'}{x(p-1)} =$$
$$= \sum_{i=0}^{p-2}(-1)^i\binom{p-1}{i} + \sum_{i=0}^{p-2}\left((-1)^i\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right) - (p-1) + \binom{n'}{0} + \sum_{x=1}^{q-1}\binom{n'}{x(p-1)}.$$

The first sum here equals -1, because $\binom{p-1}{0} - \binom{p-1}{1} + \binom{p-1}{2} + \ldots = 0$. By the same reasons the second (double) sum together with the summand $\binom{n'}{0}$ equals 0. The last sum equals $1 + p(n'+1)$ by the induction hypothesis. Therefore the whole expression equals $-1 + 0 - p + 1 + 1 + p(n'+1) = 1 + pn'$. This is exactly what we need because $1 + p(n+1) = 1 + p(n'+p-1+1) \equiv 1 + pn' \pmod{p^2}$.

5.5. This is result of Fleck, 1913, it is cited in [18]. Solution [CSTTVZ].

For $p = 2$ the sum is not alternating and the result is trivial. Let p be odd. We use the induction on q. The base follows from the statement 2.5 a). Prove the induction step from $n' = n - (p-1)$ to n. The expression $\sum_x$ below denotes the summation over x in natural bounds (i.e. in bounds for which all the binomial coefficients are correctly defined). We have

$$\pm \sum_{m:m\equiv j \pmod p} (-1)^m\binom{n}{m} = \sum_x (-1)^x\binom{n'+p-1}{xp+j} = \sum_x (-1)^x \sum_{i=0}^{p-1}\binom{p-1}{i}\binom{n'}{xp+j-i} =$$
$$= \sum_{i=0}^{p-1}\binom{p-1}{i}\sum_x (-1)^x\binom{n'}{xp+j-i}.$$

By the induction hypothesis $p^{q-1} \mid \sum_x (-1)^x\binom{n'}{xp+j-i}$; by the problem 1.1 a) $\binom{p-1}{i} \equiv (-1)^i \pmod p$. Therefore

$$\sum_{i=0}^{p-1}\binom{p-1}{i}\sum_x (-1)^x\binom{n'}{xp+j-i} \equiv \sum_{i=0}^{p-1}(-1)^i\sum_x (-1)^x\binom{n'}{xp+j-i} \pmod{p^q}.$$

The last (double sum equals $\binom{n'}{0} - \binom{n'}{1} + \binom{n'}{2} - \binom{n'}{3} + \ldots = 0$.

5.6. The result of Bhaskaran (1965), it is cited in [18], solution [CSTTVZ].

Induction on n. Let

$$f(n,j) = \binom{n}{j} - \binom{n}{j+(p-1)} + \binom{n}{j+2(p-1)} - \binom{n}{j+3(p-1)} + \ldots$$

The base $n = p+1$ is trivial, but observe that $\binom{p+1}{i} \equiv 1 \pmod p$ for $i = 0, 1, p, p+1$, otherwise this binomial coefficient is divisible by p. Prove the step of induction from $n' = n - (p+1)$ to n. By the observation above we have

$$\binom{n'+(p+1)}{j+(p-1)k} = \sum_{i=0}^{p+1}\binom{n'}{j+(p-1)k-i}\binom{p+1}{i} \equiv \sum_{i\in\{0,1,p,p+1\}}\binom{n'}{j+(p-1)k-i} =$$
$$= \binom{n'}{j+(p-1)k} + \underline{\binom{n'}{j-1+(p-1)k} + \binom{n'}{j-1+(p-1)(k-1)}} + \binom{n'}{j-2+(p-1)(k-1)} \pmod p.$$

Since $f(n,j) = \sum_k (-1)^k \binom{n}{j+k(p-1)}$ is an alternating sum, the underlined summands cancel (except the first and the last, but these summands are equal to 0 due to incorrect binomial coefficietns). So we obtain the equalities

$$f(n,j) \equiv f(n',j) - f(n',j-2) \quad \text{при } j > 1\,, \qquad\qquad f(n,1) \equiv f(n',1) + f(n',p-2)\,.$$

Now the part "only if" of the problem statement follows from the induction hypothesis, and the part "if", too: if $f(n,j) \equiv 0 \pmod p$ for $j = 1, 3, \ldots, p-2$, then

$$f(n',p-2) \equiv f(n',p-4) \equiv \ldots \equiv f(n',1) \equiv -f(n',p-2)\,,$$

from where $f(n',j) \equiv 0 \pmod p$ for all required j, and then $n' \vdots (p+1)$, hence $n \vdots (p+1)$.

References

The authors of many solutions are participants of the conference:

[D] Didin Maxim;
[K] Krekov Dmitri;
[J] Jastin Lim Kai Ze;
[T] Teh Zhao Yang Anzo;
[CSTTVZ] Ćevid Domagoj, Stokić Maksim, Tanasijeviić Ivan, Trifunović Petar, Vukorepa Borna, Žikelić Đorđe

Список литературы

[1] *Винберг Э. Б.* Удивительные свойства биномиальных коэффициентов. // Мат. просвещение. Третья серия. Вып. 12. 2008

[2] *Гашков С.Б., Чубариков В.Н.* Арифметика. Алгоритмы. Сложность вычислений. М.: Высш. шк., 2000.

[3] *Дынкин Е.Б., Успенский В.А.* Математические беседы. 2-е изд. М.: ФИЗМАТЛИТ, 2004.

[4] Петербургские математические олимпиады, 1961–1993. СПб: Лань, 2007.

[5] *Табачников С.Л., Фукс Д.Б.* Математический дивертисмент. 30 лекций по классической математике. М.: МЦНМО, 2011.

[6] *Фукс Д.Б., Фукс М.Б.* Арифметика биномиальных коэффициентов // Квант. 1970. №6. С. 17–25.

[7] *Ширшов А.И.* Об одном свойстве биномиальных коэффициентов // Квант. 1971. №10. С. 16–20.

[8] *Cai T.X., Granville A.* On the residues of binomial coefficients and their products modulo prime powers /!/ Acta

[9] *Calkin N. J.* Factors of sums of powers of binomial coefficients // Acta Arith. 1998. Vol. 86. P. 17–26.

[10] *Carlitz L.* A note of Wolstenholme's theorem // Amer. Math. Monthly. 1954. Vol. 61. № 3. P. 174–176.

[11] *Dimitrov V., Chapman R.* Binomial coefficient identity: 11118 // Amer. Math. Monthly. 2006. Vol. 113. № 7. P. 657–658.

[12] *Everett W.* Subprime factorization and the numbers of binomial coefficients exactly divided by powers of a prime // Integers. 2011. Vol. 11. # A63. `http://www.integers-ejcnt.org/vol11.html`

[13] *Fine N.* Binomial coefficient modulo a prime // Amer. Math. Monthly. 1947. Vol. 54. № 10. Part 1. P. 589–592.

[14] *Gardiner A.* Four problems on prime power divisibility // Amer. Math. Monthly. 1988. Vol. 95. № 10. P. 926–931.

[15] *Gauss K.* Disquisitiones arithmeticae. 1801. Art. 78.

[16] *Gessel I.* Wolstenholme revisited // Amer. Math. Monthly. 1998. Vol. 105. № 7. P. 657–658.

[17] *Granville A.* Arithmetic properties of binomial coefficients. `http://www.dms.umontreal.ca/~andrew/Binomial/`

[18] *Granville A.* Binomial coefficients modulo prime powers.

[19] *Granville A.* Zaphod Beeblebrox's Brian and the Fifty-ninth Row of Pascal's Triangle // Amer. Math. Monthly. 1992. Vol. 99. № 4. P. 318–331.

[20] *Granville A.* Correction to: Zaphod Beeblebrox's Brian and the Fifty-ninth Row of Pascal's Triangle // Amer. Math. Monthly. 1997. Vol. 104. № 9. P. 848–851.

[21] *Hinz A.* Pascal's triangle and tower of Hanoi // Amer. Math. Monthly. 1992. Vol. 99. № 6. P. 538–544.

[22] *Loveless A.* A congruence for products of binomial coefficients modulo a composite // Integers: electronic journal of comb. number theory 7 (2007) # A44

[23] *McIntosh R.* On the converse of Wolstenhome's theorem // Acta Arithmetica. 1995. Vol. 61. №4. P. 381–388.

[24] *Meštrović R.* On the mod p^7 determination of $\binom{2p-1}{p-1}$ // `http://arxiv.org/pdf/1108.1174v1.pdf`

[25] *More Y., Chapman R.* The sum of Catalan numbers, modulo 3: 11165 // Amer. Math. Monthly. 2007. Vol. 114. № 5. P. 454–455.

[26] *Morley F.* Note on the congruences $2^{4n} \equiv (-)^n (2n)!/(n!^2)$, where $2n+1$ is a prime // Annals of Math. 1894-1895. Vol. 9. №1. P. 168–170.

[27] *Roberts J.* On binomial coefficient residues // Canad J. Math. 1957. Vol. 9. P. 363–370.

[28] Sun Z.-W., Wan D. On Fleck quotients // `arXxiv:math.0603462v3`

Арифметические свойства биномиальных коэффициентов

На конференции Вам будет предложено несколько исследовательских проектов. Цель — как можно дальше продвинуться в каком-то из проектов. Задачи можно решать коллективно, объединившись в любые команды (члены команды могут быть из разных городов). Вы можете решать задачи сразу из нескольких проектов, причем по разным проектам Вы можете участвовать в разных командах. Единственное, чего не следует делать, — это присваивать себе чужие результаты, такое случается, если команда слишком велика и не все из нее активно решают задачи данного проекта.

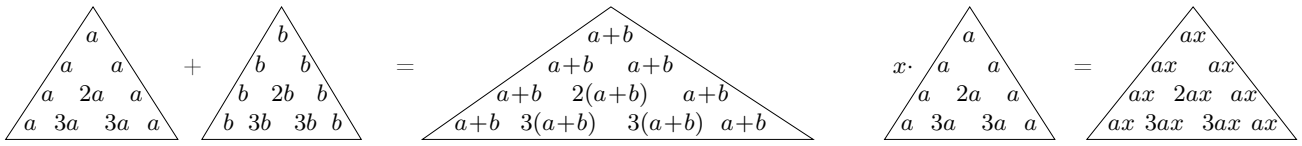
Это ознакомительная подборка задач по теме о биномиальных коэффициентах. Задачи следует решать письменно и сдавать Кохасю К.П. (вагон 15, место 17). В Теберде набор задач будет существенно расширен и все задачи, кроме задачи 1.2, можно будет сдавать и позже. По задаче 1.2 решения принимаются только в поезде, после этого задача снимается с конкурса.

1 Задачи в поезд

1.1. Докажите, что а) $C_{p-1}^k \equiv (-1)^k \pmod p$; б) $C_{2n}^n \equiv (-4)^n C_{\frac{p-1}{2}}^n \pmod p$ при $n \leqslant \frac{p-1}{2}$.

1.2. Докажите, что количество нечетных биномиальных коэффициентов в n-й строке треугольника Паскаля равно 2^r, где r — количество единиц в двоичной записи числа n.

1.3. Зафиксируем натуральное число m. Назовем m-арифметическим треугольником Паскаля треугольник, в котором вместо чисел C_n^k расставлены их остатки по модулю m. Кроме того, мы будем рассматривать похожие треугольники из остатков, у которых вдоль боковых сторон вместо единиц стоят одинаковые остатки a по модулю m. Такие треугольники можно умножать на число, а также складывать (если размеры совпадают), причем будем считать, что операции тоже выполняются по модулю m.



Пусть в s-й строке m-арифметического треугольника Паскаля все элементы, кроме крайних, — нули. Докажите, что тогда этот треугольник имеет вид, показанный на рис. 1. Заштрихованные треугольники состоят из нулей, а треугольники Δ_n^k состоят из s строк и подчинены следующим соотношениям:

1) $\Delta_n^{k-1} + \Delta_n^k = \Delta_{n+1}^k$; 2) $\Delta_n^k = C_n^k \cdot \Delta_0^0 \pmod m$.

Головоломка Ханойская башня представляет собой три стержня, на которые надеваются диски разной величины. Вначале все диски упорядочены по размеру (более крупные — ниже) и находятся на первом стержне. Разрешается снять со стержня один верхний диск и переместить его на другой стержень. При этом запрещается более крупный диск класть на диск меньшего размера. В головоломке требуется переложить все диски с первого стержня на второй.

Пусть количество дисков равно n. Рассмотрим граф TH_n, вершины которого — это всевозможные расположения дисков Ханойской башни, а ребра соединяют те состояния головоломки, которые получаются друг из друга за один ход. Рассмотрим также граф P_n, вершины которого — это единицы, расположенные в первых 2^n строках 2-арифметического треугольника Паскаля, а ребра соединяют соседние единицы (т.е. соседние в строке или в двух смежных строках по диагонали).

1.4. Докажите, что графы TH_n и P_n изоморфны.

1.5. Докажите, что в первых 10^6 строках 2-арифметического треугольника Паскаля единицы составляют меньше 1 %.

1.6. Докажите, что если n делится на $p-1$, то $C_n^{p-1} + C_n^{2(p-1)} + C_n^{3(p-1)} + \ldots + C_n^n \equiv 1 \pmod p$. Или лучше докажите в общем виде: если $1 \leqslant j, k \leqslant p-1$ и $n \equiv k \pmod{p-1}$, то

$$C_n^j + C_n^{(p-1)+j} + C_n^{2(p-1)+j} + C_n^{3(p-1)+j} + \ldots \equiv C_k^j \pmod p.$$

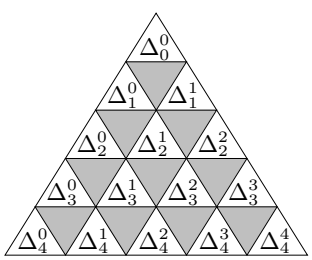


Рис. 1.

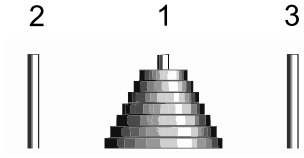


Рис. 2.

Арифметические свойства биномиальных коэффициентов — 2

Официальным "теоретическим материалом" для этого цикла задач служит статья Э. Б. Винберга [1]. В частности, считаются известными следующие теоремы.

1. Теорема Вильсона. Для всех простых p (и только для простых) выполнено сравнение $(p-1)! \equiv -1 \pmod p$.
2. Теорема Люка. Запишем числа n и k в системе счисления по основанию p:
$$n = n_d p^d + n_{d-1}p^{d-1} + \ldots + n_1 p + n_0, \qquad k = k_d p^d + k_{d-1}p^{d-1} + \ldots + k_1 p + k_0. \tag{1}$$
Тогда $C_n^k \equiv C_{n_d}^{k_d} C_{n_{d-1}}^{k_{d-1}} \cdot \ldots \cdot C_{n_1}^{k_1} C_{n_0}^{k_0} \pmod p$.
3. Теорема Куммера. Показатель $\operatorname{ord}_p C_n^k$ равен числу переносов при сложении столбиком чисел k и $\ell = n-k$ в p-ичной записи.
4. Теорема Волстенхолма. При $p \geqslant 5 \quad C_{2p}^p \equiv 2 \pmod{p^3}$ или, что то же самое, $C_{2p-1}^{p-1} \equiv 1 \pmod{p^3}$.

Напомним, что по определению $C_0^0 = 1$, $C_n^k = 0$ при $k > n$ и при $k < 0$.

Всюду буквой p мы обозначаем простое число. Для произвольного натурального числа n обозначим через $(n!)_p$ произведение всех натуральных чисел от 1 до n, не делящихся на p. Если задано число p, то символами n_i, m_i и т. д. обозначаются цифры p-ичной записи чисел n, m и т. д.

* * *

2 Арифметический треугольник и делимость

2.1. a) Докажите, что в первых 3^k строках 3-арифметического треугольника Паскаля содержится $\frac{1}{2}(6^k + 4^k)$ единиц и $\frac{1}{2}(6^k - 4^k)$ двоек.

b) Найдите число нулевых элементов в первых 5^k строках 5-арифметического треугольника Паскаля.

c) Найдите число ненулевых элементов в первых p^k строках p-арифметического треугольника Паскаля.

2.2. Докажите, что количество единиц в первых m строках 2-арифметического треугольника Паскаля равно
$$\sum_{i=0}^{n-1} m_i \cdot 2^{\sum_{k=i+1}^{n-1} m_k} \cdot 3^i.$$
Полагая $m = 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_r}$, где $\alpha_1 > \alpha_2 > \ldots > \alpha_r$, можно то же выражение записать в виде
$$3^{\alpha_1} + 2 \cdot 3^{\alpha_2} + 2^2 \cdot 3^{\alpha_3} + \ldots + 2^{r-1} \cdot 3^{\alpha_r}.$$

2.3. Рассмотрим n-ю строку 2-арифметического треугольника Паскаля как двоичную запись некоторого натурального числа P_n. Докажите, что
$$P_n = F_{i_1} \cdot \ldots \cdot F_{i_s},$$
где $i_1, \ldots, i_s$ — номера разрядов, в которых в двоичной записи числа n стоят единицы, и $F_i = 2^{2^i} + 1$ — i-е число Ферма.

2.4. Докажите, что количество ненулевых элементов в n-й строке p-арифметического треугольника Паскаля равно $\prod_{i=0}^{d} (n_i + 1)$.

2.5. a) Для того чтобы все биномиальные коэффициенты C_n^k, где $0 < k < n$, делились на p, необходимо и достаточно, чтобы n было степенью числа p.

b) Для того чтобы все биномиальные коэффициенты C_n^k, где $0 \leqslant k \leqslant n$, не делились на p, необходимо и достаточно, чтобы $n+1$ делилось на p^d, иными словами, чтобы все цифры p-ичной записи числа n, кроме старшей, были равны $p-1$.

2.6. Пусть $0 < k < n+1$. Докажите, что если $C_n^{k-1} \not\vdots\, p$ и $C_n^k \not\vdots\, p$, то $C_{n+1}^k \not\vdots\, p$, *кроме* случая, когда $n+1$ делится на p.

3 Обобщение теорем Вильсона и Люка

3.1. Докажите, что $\operatorname{ord}_p(n!) = \dfrac{n - (n_d + \ldots + n_1 + n_0)}{p-1}$.

3.2. Докажите следующие обобщения теоремы Вильсона. a) $(-1)^{[n/p]}(n!)_p \equiv n_0! \pmod p$;
b) При $p \geqslant 3$ выполнено сравнение

$$(p^q!)_p \equiv -1 \pmod{p^q},$$

а при $p = 2$, $q \geqslant 3$ выполнено сравнение $(p^q!)_p \equiv 1 \pmod{p^q}$.
c) $\dfrac{n!}{p^\mu} \equiv (-1)^\mu n_0! n_1! \ldots n_d! \pmod p$, где $\mu = \operatorname{ord}_p(n!)$

3.3. Обобщенная теорема Люка. Пусть $r = n - k$, $\ell = \operatorname{ord}_p(C_n^k)$. Тогда

$$\frac{1}{p^\ell} C_n^k \equiv (-1)^\ell \Big(\frac{n_0!}{k_0! r_0!}\Big)\Big(\frac{n_1!}{k_1! r_1!}\Big) \ldots \Big(\frac{n_d!}{k_d! r_d!}\Big) \pmod p$$

3.4. a) Докажите, что $(1+x)^{p^d} \equiv 1 + x^{p^d} \pmod p$ при всех $x = 0, 1, \ldots, p-1$.
b) Докажите теорему Люка алгебраически.

3.5. a) Пусть m, n, k — натуральные числа, причем $(n, k) = 1$. Докажите, что $C_{mn}^k \equiv 0 \pmod n$.
b) Если $n \vdots p^k$, $m \not\vdots p$, то $C_n^m \vdots p^k$.

3.6. Пусть $f_{n,a} = \sum\limits_{k=0}^{n} (C_n^k)^a$. Докажите, что $f_{n,a} \equiv \prod\limits_{i=0}^{d} f_{n_i,a} \pmod p$.

4 Вариации на тему теоремы Волстенхолма

4.1. Докажите, что $\dfrac{1}{1} + \dfrac{1}{2} + \ldots + \dfrac{1}{p-1} \equiv 0 \pmod{p^2}$ при $p \geqslant 5$.

4.2. Пусть $p = 4k + 3$ — простое число. Найдите $\dfrac{1}{0^2+1} + \dfrac{1}{1^2+1} + \ldots + \dfrac{1}{(p-1)^2+1} \pmod p$.

4.3. a) Пусть натуральное число k таково, что для каждого простого делителя p числа m $k \not\vdots (p-1)$. Докажите, что

$$\frac{1}{1^k} + \frac{1}{2^k} + \ldots + \frac{1}{(m-1)^k} \equiv 0 \pmod m.$$

Здесь суммирование распространяется на все слагаемые, знаменатели которых взаимно просты с m.
b) Пусть k нечетно и $(k+1) \not\vdots (p-1)$. Докажите, что $\dfrac{1}{1^k} + \dfrac{1}{2^k} + \ldots + \dfrac{1}{(p-1)^k} \equiv 0 \pmod{p^2}$.

4.4. Докажите, что сравнение (12) из статьи Винберга выполнено по модулю p^4.

4.5. Докажите эквивалентность следующих сравнений. 1) $C_{2p-1}^{p-1} \equiv 1 \pmod{p^4}$;
2) $\dfrac{1}{1} + \dfrac{1}{2} + \ldots + \dfrac{1}{p-1} \equiv 0 \pmod{p^3}$; 3) $\dfrac{1}{1^2} + \dfrac{1}{2^2} + \ldots + \dfrac{1}{(p-1)^2} \equiv 0 \pmod{p^2}$.

4.6. a) Докажите алгебраически, что для всякого простого p и произвольных k и n $(C_{pk}^{pm} - C_k^m) \vdots p^2$. В статье Винберга этот факт доказан комбинаторно.
b) Докажите утверждение (9) из статьи Винберга: для всякого простого $p \geqslant 5$ и произвольных k и n $(C_{pk}^{pm} - C_k^m) \vdots p^3$.

4.7. Пусть $p \geqslant 5$. Докажите, что a) $C_{p^2}^p \equiv C_p^1 \pmod{p^5}$; b) $C_{p^{s+1}}^p \equiv p^s \pmod{p^{2s+3}}$.

4.8. Докажите, что $C_{p^3}^{p^2} \equiv C_{p^2}^p \pmod{p^8}$.

Арифметические свойства биномиальных коэффициентов — 3

Дополнения к предыдущим темам

2.7. Докажите, что $C^k_{p^n-1} \equiv (-1)^{S_k} \pmod p$, где S_k — сумма цифр p-ичной записи числа k.

2.8. Докажите, что если биномиальный коэффициент C^k_n нечетен, (т. е. в обозначениях из (1) $k_i \leqslant n_i$ при всех $i = 0, \ldots, d$), то

$$C^k_n \equiv \prod_{i=1}^{d} (-1)^{k_{i-1}n_i + k_i n_{i-1}} \pmod 4 .$$

2.9. Докажите, что если в двоичной записи числа n нет двух единиц подряд, то все нечетные числа в n-й строке треугольника Паскаля сравнимы с 1 по модулю 4, а в противном случае ровно половина из них сравнима с 1 по модулю 4.

2.10. Докажите, что количество пятерок в каждой строке 8-арифметического треугольника Паскаля равно степени двойки. То же касается единиц, троек и семерок.

2.11. Докажите, что если все элементы двух множеств

$$\{C^1_{2^n-1}, C^3_{2^n-1}, C^5_{2^n-1}, \ldots, C^{2^n-1}_{2^n-1}\} \quad \text{и} \quad \{1, 3, 5, \ldots, 2^n - 1\}$$

рассматривать как остатки по модулю 2^n, то эти множества совпадают.

2.12. Докажите, что элементы одной строки треугольника Паскаля не взаимно просты в следующем довольно сильном смысле. Для каждого числа $\varepsilon > 0$ существует N, такое, что при всех натуральных $n > N$ и $k_1, k_2, \ldots, k_{100} < \varepsilon\sqrt{n}$ верно, что числа

$$C^{n+k_1}_{2n},\ C^{n+k_2}_{2n},\ \ldots,\ C^{n+k_{100}}_{2n}$$

имеют общий делитель.

2.13. а) Даны натуральные числа $m > 1$, n, k. Докажите, что хотя бы одно из чисел C^k_n, C^k_{n+1}, …, C^k_{n+k} не делится на m.

b) Докажите, что для любого k найдется бесконечно много таких n, что все числа C^k_n, C^k_{n+1}, …, C^k_{n+k-1} делятся на m.

4.9. Докажите, что при $n > 1$ $C^{2^n}_{2^{n+1}} - C^{2^{n-1}}_{2^n}$ делится на 2^{2n+2}.

4.10. Докажите, что при $p \geqslant 5$ $(-1)^{\frac{p-1}{2}} C^{\frac{p-1}{2}}_{p-1} \equiv 4^{p-1} \pmod{p^3}$.

Арифметические свойства биномиальных коэффициентов — 4

Дополнения к предыдущим темам

4.11. Пусть m — произвольное натуральное число, $p \geqslant 5$ — простое. Докажите, что

$$\frac{1}{mp+1} + \frac{1}{mp+2} + \cdots + \frac{1}{mp+(p-1)} \equiv 0 \pmod{p^2}.$$

4.12. Пусть p и q — различные простые числа. Докажите, что сравнение $C_{2pq-1}^{pq-1} \equiv 1 \pmod{pq}$ выполнено в том и только в том случае, когда $C_{2p-1}^{p-1} \equiv 1 \pmod{q}$ и $C_{2q-1}^{q-1} \equiv 1 \pmod{p}$.

5 Суммы биномиальных коэффициентов

5.1. a) Докажите, что $\sum\limits_{k=0}^{3^a-1} C_{2k}^k$ делится на 3; b) делится на 3^a.

5.2. Пусть $C_k = \frac{1}{k+1}C_{2k}^k$ — последовательность чисел Каталана. Докажите, что $\sum\limits_{k=1}^{n} C_k \equiv 1 \pmod 3$ тогда и только тогда, когда троичное разложение числа $n+1$ содержит хотя бы одну цифру 2.

5.3. Пусть $p \geqslant 5$, $k = [2p/3]$. Докажите, что сумма $C_p^1 + C_p^2 + \ldots + C_p^k$ делится на p^2.

5.4. Если $n \vdots (p-1)$, где p — нечетное простое, то

$$C_n^{p-1} + C_n^{2(p-1)} + C_n^{3(p-1)} + \ldots \equiv 1 + p(n+1) \pmod{p^2}.$$

5.5. Докажите, что при $0 \leqslant j \leqslant p-1 < n$ и $q = [\frac{n-1}{p-1}]$

$$\sum_{m:m\equiv j \pmod p} (-1)^m C_n^m \equiv 0 \pmod{p^q}.$$

5.6. Докажите, что если p — нечетное простое, то $n \vdots (p+1)$ тогда и только тогда, когда

$$C_n^j - C_n^{j+(p-1)} + C_n^{j+2(p-1)} - C_n^{j+3(p-1)} + \ldots \equiv 0 \pmod p$$

при всех $j = 1, 3, \ldots, p-2$.

Решения

1 Задачи в поезд

1.1. а) Решение 1. $C_{p-1}^k = \dfrac{(p-1)(p-2)\dots(p-k)}{1\cdot 2\cdots k} \equiv \dfrac{(-1)(-2)\dots(-k)}{1\cdot 2\cdots k} \equiv (-1)^k \pmod p$.

Решение 2. По формуле для биномиальных коэффициентов очевидно, что C_p^i при $1 \leqslant i \leqslant p-1$ делится на p. Кроме того, имеет место основное рекуррентное соотношение $C_{p-1}^{k-1} + C_{p-1}^k = C_p^k$. Так как $C_{p-1}^0 = 1 \equiv 1 \pmod p$ и $(C_{p-1}^0 + C_{p-1}^1) \vdots p$, заключаем отсюда, что $C_{p-1}^1 \equiv -1 \pmod p$. Но $C_{p-1}^1 + C_{p-1}^2$ тоже делится на p, значит, $C_{p-1}^2 \equiv 1 \pmod p$ и т.д.

б) Это задача [3, задача 162]. Поскольку дроби C_{2n+2}^{n+1}/C_{2n}^n и $C_{\frac{p-1}{2}}^{n+1}/C_{\frac{p-1}{2}}^n$ сильно сократимы, утверждение легко проверяется по индукции. Но мы предложим прямое вычисление из [3].

Как нетрудно видеть,

$$C_{2n}^n = 2^n \cdot \frac{1\cdot 3\cdots(2n-1)}{n!}$$

При этом

$$1\cdot 3\cdots(2n-1) = (-1)^n(-1)(-3)\cdots(-2n+1) \equiv (-1)^n(p-1)(p-3)\cdots(p-2n+1) =$$

$$= (-1)^n 2^n \Big(\frac{p-1}{2}\Big)\Big(\frac{p-3}{2}\Big)\cdots\Big(\frac{p-2n+1}{2}\Big) = (-1)^n 2^n \Big(\frac{p-1}{2}\Big)\Big(\frac{p-1}{2}-1\Big)\cdots\Big(\frac{p-1}{2}-n+1\Big) =$$

$$= (-1)^n 2^n \frac{(\frac{p-1}{2})!}{(\frac{p-1}{2}-n)!} \pmod p.$$

Таким образом, $C_{2n}^n \equiv (-1)^n 4^n \dfrac{(\frac{p-1}{2})!}{n!(\frac{p-1}{2}-n)!} = (-4)^n C_{\frac{p-1}{2}}^n \pmod p$.

1.2. Это непосредственно следует из самоподобной структуры арифметического треугольника Паскаля, описанной в следующих задачах. Это также сразу следует из теоремы Люка. Доказательство можно прочесть в статье Винберга [1].

1.3. Мы ограничимся небольшим созерцанием, полное решение см. в [3, задача 133].

Поскольку в s-й строке расположен длинный ряд из нулей, в $(s+1)$-й строке под этими нулями также расположен ряд из нулей (на единицу короче), в $(s+2)$-й строке — опять ряд из нулей (снова на 1 короче) и т. д. Этим объясняется наличие серого треугольника снизу от Δ_0^0 (рис. 1).

Далее, ненулевые элементы s-й строки равны 1, тогда ряды чисел, идущих вдоль наклонных границ серого треугольника, состоящего из нулей, — это тоже всё сплошь единицы (по рекуррентному правилу построения треугольника Паскаля). Таким образом, вдоль боковых сторон треугольников Δ_1^0 и Δ_1^1 расположены единицы, и значит, оба этих треугольника идентичны Δ_0^0.

Теперь понятно, как выглядит $2s$-я строка треугольника. Крайние элементы в ней — единицы, остальные элементы — нули, кроме центрального элемента, который равен 2, как сумма двух вышестоящих единиц. Отсюда получаем, что снизу от $2s$-й строки находятся два серых нулевых треугольника, по краям от них — треугольники Δ_2^0 и Δ_2^2, идентичные Δ_0^0, а между ними — треугольник Δ_2^1, у которого вдоль боковых сторон расположены двойки. Как нетрудно понять, это значит, что $\Delta_2^1 = 2\cdot\Delta_0^0$.

Ну и так далее.

1.4. Этот сюжет мы взяли в статье [21], где некоторые факты о биномиальных коэффициентов доказываются с помощью рассмотрения Ханойской башни и графа TH_n.

Пусть на первом стержне самый верхний диск имеет диаметр a, на втором — диаметр b, на третьем — c, $a < b < c$, тогда в этом положении есть три возможных хода: с a на b или на c, либо с b на c; аналогично имеется три хода, если диски занимают лишь два стержня. Если же все диски находятся на одном стержне, возможных ходов только два, обозначим такие конфигурации A_1, A_2, A_3 по номеру стержня, на который нанизаны диски.

Заметим, что 2^s-я строка треугольника Паскаля состоит из одних единиц — это следует из задачи 1.2 или проверяется непосредственно с помощью формулы Лежандра (4). Отсюда следует, что граф P_n имеет поворотную симметрию третьего порядка, поскольку основное соотношение

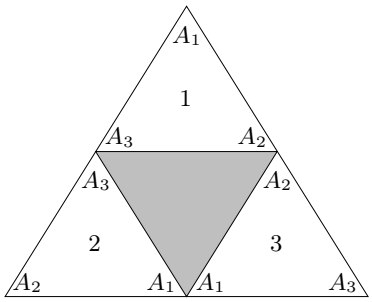


Рис. 3.

$C_n^{k-1} + C_n^k = C_{n+1}^k$, при помощи которого мы строим треугольник Паскаля "сверху вниз", в арифметике по модулю 2 равносильно соотношениям $C_n^{k-1} = C_n^k + C_{n+1}^k$ и $C_n^k = C_n^{k-1} + C_{n+1}^k$, с помощью которых можно аналогично построить треугольник Паскаля "слева снизу — вправо вверх" и "справа снизу — влево вверх". Кроме того, отсюда следует (из предыдущей задачи), что треугольник Паскаля в 2 раза большего размера содержит три копии исходного треугольника.

Докажем по индукции, что существует биекция между TH_n и P_n, при которой вершинам треугольника P_n соответствуют конфигурации A_1, A_2, A_3. База $n = 1$ очевидна.

Докажем переход. Пусть мы уже умеем строить биекцию между TH_n и P_n. Рассмотрим 2-арифметический треугольник треугольник Паскаля со стороной 2^{n+1}, он содержит три копии треугольника со стороной 2^n. Пронумеруем копии и разметим их вершины, как показано на рис. 3. Рассмотрим все положения Ханойской башни, в которых самый крупный диск находится на стержне i. Если мы не двигаем этот диск, то все эти положения и перекладывания остальных дисков задают граф, изоморфный TP_n. С помощью уже имеющейся биекции отождествим этот граф с графом P_n, расположенным в i-й копии треугольника, причем потребуем, чтобы конфигурации A_j были отождествлены в соответствии с разметкой вершин. Перемещение самого крупного диска, скажем, с первого стержня на второй возможно, только если все остальные диски находятся на третьем стержне. Это в точности соответствует ребру, соединяющему соседние вершины A_3 на левой боковой стороне треугольника, аналогично обстоят дела с другими перемещениями самого большого диска. Таким образом, построенное соответствие действительно дает изоморфизм графов TP_{n+1} и P_n.

1.5. Биекция с Ханойской башней дает простую явную формулу (когда число строк – степень двойки): в первых 2^k строках треугольника Паскаля содержится 3^k единиц. Та же формула легко доказывается по индукции из рекуррентности задачи 1.3. Пользуясь этим фактом легко получаем оценку. Так как $10^6 < 2^{20}$, количество элементов в этих строках равно $\frac{1}{2} \cdot 10^6(10^6 + 1)$, а количество единиц не превосходит 3^{20}. Доля единиц не превосходит $\frac{2 \cdot 3^{20}}{10^6(10^6+1)} \ll 0.01$.

1.6. Мы взяли это утверждение в обзоре [18].

Р е ш е н и е 1 ([CSTTVZ]). При $p = 2$ утверждение задачи легко проверяется. Будем далее считать, что p — нечетное простое. Пусть $n = x(p-1) + k$. Будем доказывать утверждение индукцией по x.

База $x = 0$ тривиальна: $C_k^j \equiv C_k^j \pmod p$.

Для доказательства перехода воспользуемся свойством биномиальных коэффициентов

$$C_{a+b}^s = \sum_i C_a^{s-i} C_b^i \qquad \text{(суммирование в естественных границах)},$$

которое выражает два способа подсчета числа вариантов взять s шаров из коробки, в которой лежит a черных и b белых шаров. Пусть $n = m + (p-1)$. Заметим, что

$$C_n^{\ell(p-1)+j} = C_{m+(p-1)}^{\ell(p-1)+j} = \sum_{i=0}^{p-1} C_m^{\ell(p-1)+j-i} C_{p-1}^i \equiv \sum_{i=0}^{p-1} (-1)^i C_m^{\ell(p-1)+j-i} \pmod p$$

(последнее сравнение — по утверждению задачи 1.1 а). Отметим, что в последней сумме первое и

последнее слагаемое присутствуют со знаком плюс. Преобразуем теперь интересующую нас сумму.

$$\sum_\ell C_n^{\ell(p-1)+j} \equiv$$
$$\equiv \left(C_m^j - C_m^{j-1} + \ldots\right) + \left(C_m^{p-1+j} - C_m^{p-1+j-1} + \ldots + C_m^j\right) + \left(C_m^{2(p-1)+j} - C_m^{2(p-1)+j-1} + \ldots + C_m^{(p-1)+j}\right) + \ldots$$
$$= \sum_{i=0}^{m} (-1)^i C_m^i + \sum_\ell C_m^{\ell(p-1)+j} \pmod p.$$

Здесь первая сумма равна нулю, а вторая по предположению индукции сравнима с $C_k^j \pmod p$. ЧТД

Решение 2 (основное рекуррентное тождество, [J], [T]). Утверждение доказывается индукцией по n. База $n \leqslant p-1$ тривиальна: левая часть содержит всего одно слагаемое — то же самое, что и в правой части. Переход:

$$C_n^j + C_n^{(p-1)+j} + \ldots = \left(C_{n-1}^j + C_{n-1}^{j-1}\right) + \left(C_{n-1}^{(p-1)+j} + C_{n-1}^{(p-1)+j-1}\right) + \ldots =$$
$$= \left(C_{n-1}^j + C_{n-1}^{(p-1)+j} + \ldots\right) + \left(C_{n-1}^{j-1} + C_{n-1}^{(p-1)+j-1} + \ldots\right) \equiv C_{k-1}^j + C_{k-1}^{j-1} = C_k^j \pmod p.$$

Но тут следует иметь в виду, что в формулировке утверждения в случае, когда параметры j и k делятся на $p-1$, они приравниваются к $p-1$, а не к 0. Таким образом, выписанное соотношение требует отдельного рассмотрения при $j = 1$ или $k = 1$. Мы ограничимся рассмотрением частного случая, которое проясняет ситуацию. Пусть $p = 5$, $j = 1$ и мы доказываем переход к $n = 13$. Имеем

$$C_1^1 \stackrel{?}{\equiv} C_{13}^1 + C_{13}^6 + C_{13}^{11} = \left(C_{12}^1 + C_{12}^6 + C_{12}^{11}\right) + \left(C_{12}^0 + C_{12}^5 + C_{12}^{10}\right)$$

Здесь первая скобка дает по индукционному предположению остаток C_4^1 (а вовсе не C_0^1, как могло показаться по предыдущему вычислению). Во второй скобке первое слагаемое не участвует в индукционном предположении, а сумма остальных сравнима с C_4^0. Записывая для ясности $p-1$ вместо 4, получаем, что вся сумма сравнима с $C_{n-1}^0 + C_{p-1}^1 + C_{p-1}^0 \equiv C_1^1 \pmod p$, что и требуется.

Решение 3 (алгебраическое рассуждение с теоремой Люка, [18]). Индукция по n. База $n \leqslant p-1$ тривиальна. Пусть теперь $n \geqslant p$, запишем все встречающиеся параметры в системе счисления по основанию p, сумму цифр числа m будем обозначать $\sigma_p(m)$. Очевидно, если $m \equiv j \pmod p$, то $\sigma_p(m) \equiv j \pmod p$. Тогда по теореме Люка интересующая нас сумма равна

$$\sum C_{n_0}^{m_0} C_{n_1}^{m_1} \ldots C_{n_d}^{m_d} \pmod p,$$

где суммирование распространяется на все $m = \overline{m_d \ldots m_1 m_0} \leqslant n$, для которых $\sigma_p(m) \equiv j \pmod p$. Эта сумма в точности равна сумме коэффициентов при x^j, x^{j+p-1}, $x^{j+2(p-1)}$, ... в выражении

$$(1+x)^{n_0}(1+x)^{n_1} \ldots (1+x)^{n_d} = (1+x)^{\sigma_p(n)}.$$

Но очевидно, что указанная сумма коэффициентов равна

$$\sum_{\substack{1 \leqslant r \leqslant \sigma_p(n) \\ r \equiv j \pmod{p-1}}} C_{\sigma_p(n)}^r,$$

которая удовлетворяет индукционному предположению, так как $1 \leqslant \sigma_p(n) \leqslant n-1$, и дает нужное нам сравнение, поскольку $\sigma_p(n) \equiv n \equiv j \pmod p$.

Решение 4 (немного здравого смысла и линейной алгебры, [Д]). Многочлены x, x^2, ..., x^{p-1} линейно независимы над $\mathbb{Z}_p$ и образуют базис в пространстве функций $f\colon \mathbb{Z}_p \to \mathbb{Z}_p, f(0) = 0$. По малой теореме Ферма $(1+x)^n \equiv (1+x)^k \pmod p$. Редуцируя левую часть с помощью соотношений $x^{i+a(p-1)} \equiv x^i$, получаем, что интересующая нас сумма как элемент $\mathbb{Z}_p$ равна коэффициенту при x^j в правой части, т. е. C_k^j.

2 *Арифметический треугольник и делимость*

2.1. a) Это результат Робертса [27]. Обозначим количество единиц в первых 3^k строках через a_k, а количество двоек b_k — через b_k. Пользуясь рекуррентностью из задачи 1.3, получаем соотношения:

$$a_{k+1} = 5a_k + b_k, \qquad b_{k+1} = 5b_k + a_k.$$

Отсюда утверждение задачи легко следует по индукции.

b) Ответ: $\frac{1}{2} \cdot 5^k(5^k + 1) - 15^k$. Обозначая искомую величину a_k, аналогично предыдущей задаче получаем соотношение

$$a_{k+1} = 15a_k + 10 \cdot \frac{5^k(5^k - 1)}{2}.$$

Поскольку в целом треугольник содержит $\frac{5^k(5^k+1)}{2}$ элементов, естественно ввести замену переменных $a_k = \frac{5^k(5^k+1)}{2} - b_k$. Тогда для переменной b_k предыдущее соотношение записывается в виде $b_{k+1} = 15b_k$.

c) Ответ: $\left(\frac{p(p+1)}{2}\right)^k$. Это результат Файна [13]. Он аналогично предыдущим пунктам получается по индукции из рекуррентности задачи 1.3.

2.2. Решение 1. Индукция по α_1. База для $\alpha_1 = 0, 1$ легко проверяется. Пусть для всех $\alpha_1 < a$ утверждение уже доказано. Докажем его для $\alpha_1 = a$. Очевидно, $\tilde{m} - 2^{\alpha_1} < 2^{\alpha_1}$. Пусть в обозначениях задачи 1.3 $s = 2^{\alpha_1}$. Числу $\tilde{m} = 2^{\alpha_2} + 2^{\alpha_3} + \ldots + 2^{\alpha_r}$ соответствует строчка в треугольнике Δ_0^0. В этой строке и в строках над ней по индукционному предположению содержится

$$3^{\alpha_2} + 2 \cdot 3^{\alpha_3} + \ldots + 2^{r-2} \cdot 3^{\alpha_r} \tag{2}$$

единиц. Тогда числу $m = \tilde{m} + 2^{\alpha_1}$ соответствует строчка, пересекающая треугольники Δ_0^1 и Δ_1^1 (идентичные треугольнику Δ_0^0, поскольку у нас 2-арифметика). В этой строке и выше находится целиком треугольник Δ_0^0 (в нем по предположению индукции 3^{α_1} единиц) и два неполных треугольника Δ_0^1 и Δ_1^1, в каждом из которых число единиц задается формулой (2). В сумме получаем

$$3^{\alpha_1} + 2(3^{\alpha_2} + 2 \cdot 3^{\alpha_3} + \ldots + 2^{r-2} \cdot 3^{\alpha_r})$$

единиц, что и требуется.

Решение 2 (комбинаторный смысл коэффициентов — разбиваем на слои, [T]).

Лемма 1. Пусть число единиц в k-й строке равно 2^r (или, что то же самое, бинарная запись числа k содержит r единиц) и пусть $\alpha_1 > \alpha_2 > \cdots > \alpha_m$, $2^{\alpha_m} > k$. Тогда число единиц в строке с номером $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_m} + k$ равно 2^{m+r}.

Доказательство. Очевидно, бинарная запись числа $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_m} + k$ содержит $m + r$ единиц и тогда в строке треугольника Паскаля с этим номером 2^{m+r} единиц. □

Лемма 2. Суммарное количество единиц в строках с номерами

$$2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}}, \quad 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + 1, \quad \ldots, \quad 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + 2^{\alpha_m} - 1,$$

равно $2^k 3^{\alpha_m}$.

Доказательство. По лемме 1 количество единиц в строке с номером $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + i$ равно $2^k x_i$, где x_i — количество единиц в i-й строке. Тогда суммарное число единиц в упомянутых строках равно $2^k \sum x_i$. Но $\sum x_i$ — это число единиц в первых $2^{\alpha_m} - 1$ строках треугольника Паскаля, оно равно 3^{α_m} (это нам известно, например, из задачи 1.4). □

Осталось просуммировать по m количества единиц из леммы 2.

2.3. Мы взяли утверждение задачи из статьи Винберга [1], а решение из статьи Гранвилля [18]. Утверждение выводится из теоремы Люка с помощью следующего наблюдения (тоже упомянутого в [1]): биномиальный коэффициент C_n^k нечетен в том и только том случае, когда единицы в двоичном разложении числа k могут стоять лишь в тех разрядах, где стоят единицы в записи числа n. Отсюда

сразу следует, что $P_n = \sum 2^k$, где суммирование распространяется на все числа k, описанные в предыдущем предложении. В обозначениях формулы (1) при $p = 2$ положим $S_n = \{i : n_i = 1\}$. Тогда

$$P_n = \sum_{I \subseteq S_n} \prod_{i \in I} 2^{2^i} = \prod_{i \in S_n} F_i.$$

2.4. Этот результат Файна [13], 1947 г, — простое следствие теоремы Куммера. Чтобы биномиальный коэффициент C_n^k не делился на p, не должно быть переносов при сложении чисел k и $n-k$, записанных в системе счисления по основанию p. При фиксированном n это означает, что выбор i-й цифры p-ичной записи числа k можно сделать $n_i + 1$ способом.

2.5. a) Это сразу следует из формулы, доказанной в предыдущей задаче, поскольку речь идет о строке, в которой ровно два элемента не делятся на p.

b) [13]. Если $(n+1) \vdots p^d$, то $n = \overline{a(p-1)(p-1)\dots(p-1)}$ в системе счисления по основанию p. Тогда для каждого k, $0 \leqslant k \leqslant n$, каждая цифра числа k не превосходит соответствующей цифры числа n. Тогда все биномиальные коэффициенты $C_{n_i}^{k_i}$ не равны нулю (в том числе, по модулю p) и по теореме Люка C_n^k не делится на p.

В обратную сторону. Пусть все биномиальные коэффициенты C_n^k не делятся на p, но число n является числом вида $\overline{a(p-1)(p-1)\dots(p-1)}$. Это значит, что одна из цифр, скажем n_i, меньше $p-1$. Возьмем $k = (p-1)\cdot p^i$. Тогда $k_i = p-1$, следовательно, $C_{n_i}^{k_i} = 0$ и по теореме Люка C_n^k делится на p. Противоречие.

2.6. Это известное утверждение мы почерпнули в [12].

Решение 1. Допустим, что $C_n^{k-1} \not\vdots p$ и $C_n^k \not\vdots p$, но при этом $C_{n+1}^k = (C_n^{k-1} + C_n^k) \vdots p$. Тогда $C_n^k \equiv -C_n^{k-1} \pmod p$. Так как оба биномиальных коэффициента не делятся на p, мы можем сократить правую и левую части. Получим $\frac{n-k+1}{k} \equiv -1 \pmod p$, откуда $n+1 \equiv 0 \pmod p$.

Решение 2 ([K]). Хотя утверждение выглядит очень естественным, напоминая нам основное тождество для биномиальных коэффициентов, часть "$C_n^{k-1} \not\vdots p$" в нем лишняя. Действительно, если $(n+1) \not\vdots p$, то $0 \leqslant n_0 \leqslant p-2$. Поскольку $C_n^k \not\vdots p$, то по теореме Куммера при всех i верно неравенство $k_i \leqslant n_i$. Но тогда аналогичные неравенства верны и для пары чисел k и $n+1$, поскольку у числа $n+1$ те же цифры, что и у n, кроме цифры в самом младшем разряде, которая у числа $n+1$ на 1 больше. Следовательно, $C_{n+1}^k \not\vdots p$.

2.7. [2]. Сразу следует из теоремы Люка и задачи 1.1.a)

2.8. Задача из статьи Винберга [1]. Индукция по числу цифр. База тривиальна. Для перехода добавляем очередную цифру в конец числа. В силу нечетности биномиального коэффициента $n_i \geqslant k_i$. Пользуясь рекуррентностью $C_n^k = C_{n-1}^{k-1} + C_{n-1}^k$, перебирая разные варианты четности n и k с помощью теоремы Куммера и задачи 4.6a) сводим все к индукционному предположению.

Например, при нечетном $n = 2\ell + 1$ и четном $k = 2m$, если $k_1 = 1$, то $k = \dots 10$, $n = \dots 11$ (двоичные записи), Тогда $(n-k) = \dots 01$ (потому что по теореме Куммера не должно было быть переносов), $(k-1)_2 = \dots 01$, значит, по теореме Куммера при сложении $(k-1)_2 + (n-k)_2$ есть ровно 1 перенос, т. е. $C_{n-1}^{k-1} \equiv 2 \pmod 4$, откуда

$$C_n^k = C_{n-1}^{k-1} + C_{n-1}^k \equiv -C_{n-1}^k = -C_{2\ell}^{2m} \equiv -C_\ell^m \pmod 4,$$

последнее — по задаче 4.6a). Этот минус в точности соответствует множителю $(-1)^{k_0 n_1 + k_1 n_0}$.

2.9. Задача из статьи Винберга [1]. Утверждение следует из результата предыдущей задачи. Если в записи n нет двух единиц подряд, то все показатели $k_{i-1}n_i + k_i n_{i-1}$ равны нулю и все биномиальные коэффициенты дают остаток 1 при делении на 4. Если же запись числа n содержит участок из единиц, начинающийся с $n_j = 1$, то у половины нечетных биномиальных коэффициентов $k_j = 0$, а у другой половины $k_j = 1$ и, как нетрудно видеть по формуле из предыдущей задачи, по модулю 4 эти половины отличаются знаком.

2.10. Этому запутанному сюжету посвящены две статьи в Monthly [19, 20].

2.11. Эта задача Д.Джукича была в 2002 г. на олимпиаде 239 школы г. Санкт-Петербурга, а потом засветилась в шорт-листе IMO-2008.

Поскольку все биномиальные коэффициенты из условия задачи нечетны (по теореме Люка), для доказательства утверждения достаточно проверить, что все числа $C_{2^n-1}^1$, $C_{2^n-1}^3$, $\ldots$, $C_{2^n-1}^{2^n-1}$ дают разные остатки при делении на 2^n. Дальше можно действовать по-разному.

Р е ш е н и е 1 ([Д]). Предположим противное, пусть $C_{2^n-1}^k \equiv C_{2^n-1}^m \pmod{2^n}$ при нечетных k и m, $k > m$. Заметим, что

$$C_{2^n-1}^k = C_{2^n}^k - C_{2^n-1}^{k-1} = C_{2^n}^k - C_{2^n}^{k-1} + C_{2^n-1}^{k-2} = \cdots = C_{2^n}^k - C_{2^n}^{k-1} + C_{2^n}^{k-2} - \ldots - C_{2^n}^{m+1} + C_{2^n-1}^m\,.$$

В частности,

$$C_{2^n}^k - C_{2^n}^{k-1} + C_{2^n}^{k-2} - \ldots - C_{2^n}^{m+1} \equiv 0 \pmod{2^n}\,.$$

Теорема Куммера позволяет для каждого r легко вычислить показатель $\operatorname{ord}_2 C_{2^n}^r$, а именно, если $\operatorname{ord}_2 r = a$, то при сложении r и $2^n - r$ произойдет $n - a$ переносов (это очевидно из алгоритма сложения столбиком), и значит, $\operatorname{ord}_2 C_{2^n}^r = n - a$. В частности, $C_{2^n}^r$ делится на 2^n при нечетном r, что позволяет отбросить в последнем сравнении половину слагаемых:

$$C_{2^n}^{k-1} + C_{2^n}^{k-3} + \ldots + C_{2^n}^{m+1} \equiv 0 \pmod{2^n}\,.$$

Другое следствие из приведенных рассуждений состоит в том, что у всех слагаемых $C_{2^n}^i$ в левой части параметр i четный и поэтому $\operatorname{ord}_2 C_{2^n}^x < n$. Докажем теперь, что выполнение этого сравнения невозможно. Выберем x, для которого $\operatorname{ord}_2 C_{2^n}^x$ имеет минимальное значение. Так как $\operatorname{ord}_2 C_{2^n}^x < n$, но при этом вся сумма делится на 2^n, найдется y, для которого $\operatorname{ord}_2 C_{2^n}^x = \operatorname{ord}_2 C_{2^n}^y$. Но тогда бинарные записи чисел x и y оканчиваются на одинаковое число нулей, поэтому между x и y найдется число z, оканчивающееся на большее число нулей. Тогда $\operatorname{ord}_2 C_{2^n}^z < \operatorname{ord}_2 C_{2^n}^x$, что противоречит минимальности.

Р е ш е н и е 2 ([CSTTVZ]). Предположим противное, пусть нашлись числа k и ℓ, $k \neq \ell$, такие что $C_{2^n-1}^{2k+1} \equiv C_{2^n-1}^{2\ell+1} \mod 2^n$, $0 \leqslant k, \ell \leqslant 2^n - 1$. Кроме того, мы будем вести рассуждения по индукции, считая, что для меньших значений n утверждение задачи уже доказано. Заметим, что

$$\begin{aligned} C_{2^n-1}^{2k+1} &= \left(\frac{2^n}{1} - 1\right)\left(\frac{2^n}{2} - 1\right)\ldots\left(\frac{2^n}{2k+1} - 1\right) = \\ &= \left(\frac{2^n}{1} - 1\right)\left(\frac{2^n}{3} - 1\right)\ldots\left(\frac{2^n}{2k+1} - 1\right)\cdot\left(\frac{2^{n-1}}{1} - 1\right)\left(\frac{2^{n-1}}{2} - 1\right)\ldots\left(\frac{2^{n-1}}{k} - 1\right) = \qquad (3) \\ &= \left(\frac{2^n}{1} - 1\right)\left(\frac{2^n}{3} - 1\right)\ldots\left(\frac{2^n}{2k+1} - 1\right)\cdot C_{2^{n-1}-1}^k \equiv \\ &\equiv (-1)^{k+1} C_{2^{n-1}-1}^k \pmod{2^n}\,. \end{aligned}$$

И аналогично $C_{2^n-1}^{2\ell+1} \equiv (-1)^{\ell+1} C_{2^{n-1}-1}^{\ell} \pmod{2^n}$. По индукционному предположению, отсюда следует, что k и ℓ не могут быть оба нечетными. Кроме того, в силу симметрии $C_{2^n-1}^r = C_{2^n-1}^{2^n-1-r}$ утверждение задачи означает также, что все биномиальные коэффициенты с четными показателями — $C_{2^n-1}^{2r}$ — тоже попарно различны и по модулю 2^n образуют то же множество, что и биномиальные коэффициенты с нечетными показателями. Поэтому k и ℓ не могут быть оба четными.

Осталось разобрать случай, когда k и ℓ разной четности, пусть $k = 2a + 1$, $\ell = 2b$. Тогда

$$C_{2^{n-1}-1}^{2a+1} + C_{2^{n-1}-1}^{2b} \equiv 0 \pmod{2^n}\,.$$

При $a = b$ это сравнение невозможно, так как $C_{2^{n-1}-1}^{2a}$ нечетно и

$$C_{2^{n-1}-1}^{2a+1} + C_{2^{n-1}-1}^{2a} = C_{2^{n-1}-1}^{2a}\left(1 + \frac{2^{n-1} - 1 - 2a}{2a+1}\right) = C_{2^{n-1}-1}^{2a}\cdot\frac{2^{n-1}}{2a+1} \equiv 2^{n-1} \pmod{2^n}\,.$$

Если же $b \neq a$, то $C_{2^{n-1}-1}^{2a} \neq C_{2^{n-1}-1}^{2b}$ по индукционному предположению и так как $C_{2^{n-1}-1}^{2a} + C_{2^{n-1}-1}^{2a+1}$ делится на 2^{n-1}, сумма $C_{2^{n-1}-1}^{2b} + C_{2^{n-1}-1}^{2a+1}$ не может делиться на 2^{n-1}.

2.12. Эту задачу нам сообщил А. Белов. Заметим, что

$$C_{2n}^{n+k} = C_{2n}^n \cdot \frac{n(n-1)\ldots(n-k+1)}{(n+1)(n+2)\ldots(n+k)},$$

и таким образом, C_{2n}^{n+k} имеет много общих множителей с C_{2n}^{n}, кроме тех, которые сократились со знаменателем дроби. Заметим, что знаменатель не превосходит $(2n)^k$. Напишем аналогичные равенства для всех биномиальных коэффициентов $C_{2n}^{n+k_1}$, $C_{2n}^{n+k_2}$, …, $C_{2n}^{n+k_{100}}$. Наибольший общий делитель всех знаменателей в правых частях этих равенств не превосходит $(n+1)(n+2)\dots\big(n+[\varepsilon\sqrt{n}\,]\big) < (2n)^{\varepsilon\sqrt{n}}$. Но при больших n биномиальный коэффициент C_{2n}^{n} — существенно более крупное число, поэтому даже если сократить его на наибольший общий делитель всех знаменателей, останется весьма крупное частное, которое и будет общим делителем всех ста биномиальных коэффициентов.

Поясним последнее соображение с помощью оценки. Заметим, что

$$C_{2n}^{n} = \frac{2n}{n}\cdot\frac{2n-1}{n-1}\dots\frac{n+1}{1} > 2^n.$$

При этом $(2n)^{100\varepsilon\sqrt{n}} = 2^{\varepsilon\sqrt{n}\log_2 n+\varepsilon\sqrt{n}}$. Очевидно, для каждого фиксированного ε существует N, такое что при всех $n > N$ будет выполнено неравенство

$$\frac{n}{2} > \varepsilon\sqrt{n}\log_2 n + \varepsilon\sqrt{n}.$$

Если для таких n поделить C_{2n}^{n} на НОД всех знаменателей, частное будет не меньше $2^{n/2}$.

2.13. a) Задача предлагалась в 1977 г. на Ленинградской олимпиаде школьников.

Р е ш е н и е 1 (без теоремы Куммера). Мы приводим решение из замечательной книжки [4].

Допустим, что все эти числа делятся на m. Тогда числа

$$\begin{aligned} C_{n+k-1}^{k-1} &= C_{n+k}^{k} - C_{n+k-1}^{k}\,,\\ C_{n+k-2}^{k-1} &= C_{n+k-1}^{k} - C_{n+k-2}^{k}\,,\\ &\dots\\ C_{n}^{k-1} &= C_{n+1}^{k} - C_{n}^{k} \end{aligned}$$

также делятся на m. Аналогично, на m делятся и все числа C_{n+i}^{j}, где $i \leqslant j$ — произвольные неотрицательные целые числа. Но среди них есть число C_n^0 $(i = j = 0)$, которое равно 1. Противоречие.

Р е ш е н и е 2 (теорема Куммера). Пусть p — простой множитель числа m. Проверим, что одно из чисел C_n^k, C_{n+1}^k, …, C_{n+k}^k не делится на p. Запишем k в системе счисления по основанию p. По теореме Куммера достаточно найти такое число ℓ (где $n-k \leqslant \ell \leqslant n$), чтобы сложение $k+\ell$ в системе счисления по основанию p выполнялось без переносов, тогда биномиальный коэффициент $C_{k+\ell}^{k}$ не будет делиться на p.

Это сделать совсем нетрудно. Мы ограничимся рассуждением на конкретном примере. Пусть $p = 7$, $k = 133$ (здесь и далее числа записаны в семиричной системе счисления). Поскольку диапазон, в котором мы ищем число ℓ, содержит $k+1$ число, нам всегда удастся выбрать ℓ так, чтобы число $k+\ell$ было одним из чисел следующего вида

$$\dots 133,\quad \dots 233,\quad ,\dots,\quad \dots 633.$$

(Напомним, что цифра 6 в нашем примере самая старшая.) Тогда очевидно, что при сложении $k+\ell$ не было ни одного переноса.

b) Утверждение взято из [2]. Такие n нетрудно построить с помощью теоремы Куммера. Пусть $\operatorname{ord}_p m = s$, и запись числа k в системе счисления по основанию p содержит $d+1$ цифр. Пусть $n \vdots p^{d+s+1}$. Тогда числа $n-k$, $n-k+1$, …, $n-1$ содержат в разрядах с $(d+2)$-го по $(d+s+2)$-й цифры $(p-1)$, поэтому при сложении этих чисел с k в указанных разрядах будут возникать переносы. Таким образом, по теореме Куммера получаем, что интересующие нас биномиальные коэффициенты все делятся на p^s.

Поскольку условия, наложенные на n, легко совмещаются для разных p, мы получаем отсюда требуемое.

3 Обобщение теорем Вильсона и Люка

3.1. Как известно, $\operatorname{ord}_p(n!) = \sum_k \left[\frac{n}{p^k}\right]$. Если $n = n_d p^d + n_{d-1}p^{d-1} + \ldots + n_1 p + n_0$ — запись в системе счисления по основанию p, то $\left[\frac{n}{p^k}\right] = n_d p^{d-k} + n_{d-1}p^{d-k-1} + \ldots + n_{k+1}p + n_k$ и формулу для $\operatorname{ord}_p(n!)$ можно записать в виде

$$\operatorname{ord}_p(n!) = \sum_{k=1}^{d}\left(\sum_{i=k}^{d} n_i p^{i-k}\right) = \sum_{i=1}^{d} n_i(p^{i-1}+p^{i-2}+\ldots+p+1) = \sum_{i=1}^{d} n_i\frac{p^i-1}{p-1} = \frac{\sum\limits_{i=0}^{d} n_i p^i - \sum\limits_{i=0}^{d} n_i}{p-1}.$$

Мы получили в точности требуемое выражение.

Утверждение задачи также нетрудно доказать индукцией по n, см. [5].

3.2. а) Разбивая множители, составляющие выражение $n!$, на группы по $(p-1)$ штук, получаем

$$(n!)_p = \prod_{k=0}^{[\frac{n}{p}]-1}\big((kp+1)\cdot(kp+2)\cdots(kp+p-1)\big)\cdot\big([\tfrac{n}{p}]p+1\big)\big([\tfrac{n}{p}]p+2\big)\ldots\big([\tfrac{n}{p}]p+n_0\big) \equiv (-1)^{[\frac{n}{p}]} n_0! \pmod p.$$

б) Это утверждение встречается у Гаусса [15]. В произведение $(p^q!)_p$ вместе с каждым сомножителем входит и его обратный по модулю p^q, и произведение этой пары равно 1 по модулю p^q. Таким образом, нам следует лишь проследить за теми множителями m, которые совпадают со своими обратными, т.е. удовлетворяют сравнению

$$m^2 \equiv 1 \pmod{p^q}.$$

Для нечетного p сравнение имеет 2 решения: ± 1. Для $p=2$, $q \geqslant 3$ сравнение имеет еще пару решений: $2^{q-1} \pm 1$.

с) Так как $n! = (n!)_p \cdot p^{[\frac{n}{p}]}\big([\frac{n}{p}]\big)!$, утверждение легко доказывается по индукции с помощью сравнения из п. а).

3.3. Мы взяли утверждение со странички Гранвилля [17]. Помимо теоремы Куммера, широко известна прямая и не столь симпатичная формула для числа ℓ (формула Лежандра):

$$\ell = \operatorname{ord}_p(C_n^k) = \left(\left[\frac{n}{p}\right] - \left[\frac{k}{p}\right] - \left[\frac{r}{p}\right]\right) + \left(\left[\frac{n}{p^2}\right] - \left[\frac{k}{p^2}\right] - \left[\frac{r}{p^2}\right]\right) + \ldots \tag{4}$$

Обозначим для краткости $\tilde{n} = [n/p]$ и т. п. и напишем формулу для биномиального коэффициента, собрав отдельно все множители, делящиеся на p:

$$C_n^k = \frac{(n!)_p}{(k!)_p (r!)_p}\cdot\frac{p^{[n/p]}}{p^{[k/p]}\cdot p^{[r/p]}}\cdot\frac{\tilde{n}!}{\tilde{k}!\cdot\tilde{r}!}.$$

Здесь первая дробь может быть преобразована по модулю p в соответствии с обобщенной теоремой Вильсона (задача 3.2, б) к выражению $\frac{n_0!}{k_0! r_0!}$, третья дробь позволяет действовать по индукции, а средняя дробь (и знак из обобщенной теоремы Вильсона, который мы не упомянули) по формуле (4) даст все нужные выражения, содержащие ℓ.

3.4. а) Раскрывая скобки в выражении $(1+x)^{p^d}$, мы можем воспользоваться тем, что при $1 \leqslant k \leqslant p^d-1$ биномиальный коэффициент $C_{p^d}^k$ делится на p (аналогично задаче 1.1 или по теореме Куммера).

b) Положим $n = n'p + n_0$, $k = k'p + k_0$. По утверждению п. а) $(1+x)^{pn'} \equiv (1+x^p)^{n'} \pmod p$ Тогда

$$(1+x)^n = (1+x)^{pn'}(1+x)^{n_0} \equiv (1+x^p)^{n'}(1+x)^{n_0} \pmod p.$$

Указанное сравнение надо понимать в том смысле, что мы преобразовываем коэффициенты многочлена с целыми коэффициентами с точки зрения их делимости на p. Коэффициент при x^k в левой части равен C_n^k. При раскрытии скобок в правой части мы видим, что все показатели в первой скобке делятся на p, поэтому единственный способ получить одночлен $x^{pk'+k_0}$ — это перемножить

$x^{pk'}$ из первой скобки и x^{k_0} из второй. Итоговый коэффициент будет равен $C_{n'}^{k'}C_{n_0}^{k_0}$. Таким образом, $C_n^k = C_{n'}^{k'}C_{n_0}^{k_0}$, откуда теорема Люка следует по индукции.

3.5. a, b) Простое следствие теоремы Куммера.

3.6. [9]. В следующем вычислении мы используем то, что $C_{n_i}^{k_i} = 0$ при $k_i > n_i$; это позволяет, применив теорему Люка, отбросить при суммировании большое число слагаемых.

$$f_{n,a} = \sum_{k=0}^{n}(C_n^k)^a \equiv \sum_{k_d=0}^{n_d}\sum_{k_{d-1}=0}^{n_{d-1}} \cdots \sum_{k_0=0}^{n_0}\prod_{i=0}^{d}(C_{n_i}^{k_i})^a \equiv \prod_{i=0}^{d}\sum_{k_i=0}^{n_i}(C_{n_i}^{k_i})^a \equiv \prod_{i=0}^{d} f_{n_i,a} \pmod p.$$

4 *Вариации на тему теоремы Волстенхолма*

4.1. Это упражнение на чтение статьи. Утверждение доказано в статье Винберга, но доказательство не выделено явно. Заметим, что

$$2\sum_{i=1}^{p-1}\frac{1}{i} = \sum_{i=1}^{p-1}\frac{1}{i} + \frac{1}{p-i} = p\sum_{i=1}^{p-1}\frac{1}{i(p-i)}\,.$$

Таким образом, рассматриваемая сумма делится на p. Так как по модулю p выражения $\frac{1}{i}$ и $-\frac{1}{p-i}$ равны, нам остается проверить, что

$$\sum_{i=1}^{p-1}\frac{1}{i^2} \equiv 0 \pmod p.$$

Или, поскольку $\frac{1}{1^2}, \frac{1}{2^2}, \ldots, \frac{1}{(p-1)^2}$ — это тот же набор остатков[1], что и $1^2, 2^2, \ldots, (p-1)^2$, достаточно проверить, что

$$\sum_{i=1}^{p-1} i^2 \equiv 0 \pmod p. \tag{5}$$

Пусть $\sum\limits_{i=1}^{p-1} i^2 \equiv s \pmod p$. При $p > 5$ всегда можно выбрать остаток a, такой что $a^2 \not\equiv 1 \pmod p$. Тогда множества $\{1, 2, \ldots, p-1\}$ и $\{a, 2a, \ldots, (p-1)a\}$ совпадают (доказательство как в сноске) и

$$s \equiv \sum_{i=1}^{p-1} i^2 = \sum_{i=1}^{p-1}(ai)^2 = a^2\sum_{i=1}^{p-1} i^2 \equiv a^2 s \pmod p\,.$$

Поэтому $s \equiv 0 \pmod p$.

Разумеется, этот факт нетрудно доказать непосредственно, пользуясь соображением $\frac{1}{x} \equiv x^{\varphi(m)-1} \pmod m$. Мы используем эту технику в третьем решении следующей задачи.

4.2. Ответ: $2k+2$. Эта задача А. С. Голованова предлагалась на олимпиаде Туймаада в 2012 г. Мы приводим три решения. Отметим, что при $p = 4k+3$ уравнение $x^2 + 1 = 0$ не имеет решений в поле остатков по модулю p, следовательно, знаменатели всех рассматриваемых дробей не равны нулю.

Р е ш е н и е 1. Обозначим $a_i = i^2 + 1$, для $i = 0, \ldots, p-1$. Тогда рассматриваемое выражение равно

$$\frac{\sigma_{p-1}(a_0, a_1, \ldots, a_{p-1})}{\sigma_p(a_0, a_1, \ldots, a_{p-1})},$$

где σ_i — основной симметрический многочлен степени i. Найдем многочлен, корнями которого являются числа a_i, т. е.

$$\prod_{i=0}^{p-1}(x - 1 - i^2).$$

[1] Напомним доказательство: $\frac{1}{1}, \frac{1}{2}, \ldots, \frac{1}{(p-1)}$ и 1, 2, …, $(p-1)$ — это один и тот же набор остатков, потому что и в том, и в другом наборе по $p-1$ элементу, при этом очевидно, что в каждом наборе все остатки различны и не равны нулю, значит, каждый набор содержит все ненулевые остатки по модулю p. Тогда для квадратов утверждение очевидно.

Сделав замену $x-1=t^2$, получим многочлен

$$\prod_{i=0}^{p-1}(t^2-i^2)=\prod_{i=0}^{p-1}(t-i)\prod_{i=0}^{p-1}(t+i)\equiv(t^p-t)(t^p-t)=t^{2p}-2t^{p+1}+t^2.$$

Теперь, сделав обратную замену, получаем для $p=4k+3$

$$\prod_{i=0}^{p-1}(x-1-i^2)\equiv(x-1)^p-2(x-1)^{\frac{p+1}{2}}+(x-1)=x^p+\ldots+(p+2\cdot\tfrac{p+1}{2}+1)x-4.$$

По теореме Виета, $\sigma_p\equiv 4 \pmod p$, $\sigma_{p-1}\equiv 2 \pmod p$, поэтому $\frac{\sigma_{p-1}}{\sigma_p}\equiv\frac{1}{2}\equiv 2k+2 \pmod p$.

Р е ш е н и е 2. Разобьем все ненулевые остатки по модулю p, кроме ±1, на пары взаимно обратных. Тогда получится $2k$ пар и в каждой паре (i,j)

$$ij\equiv 1 \quad\Leftrightarrow\quad i^2j^2\equiv 1 \quad\Leftrightarrow\quad (ij)^2+i^2+j^2+1\equiv i^2+j^2+2 \pmod p.$$

Следовательно,

$$1\equiv\frac{(ij)^2+i^2+j^2+1}{(i^2+1)(j^2+1)}\equiv\frac{i^2+j^2+2}{(i^2+1)(j^2+1)}=\frac{1}{i^2+1}+\frac{1}{j^2+1} \pmod p.$$

Таким образом, наша сумма равна $\frac{1}{0^2+1}+\frac{1}{1^2+1}+\frac{1}{(-1)^2+1}+2k\equiv 2k+2$.

Р е ш е н и е 3. Как мы знаем, благодаря малой теореме Ферма, при вычислении по модулю p операции $x\mapsto x^{-1}$ и $x\mapsto x^{p-2}$ дают одинаковый результат. Таким образом, достаточно вычислить сумму

$$\sum_{x=0}^{p-1}(x^2+1)^{p-2}=\sum_{x=0}^{p-1}\sum_{m=0}^{p-2}C_{p-2}^m x^{2m}=\sum_{m=0}^{p-2}C_{p-2}^m S_{2m}, \tag{6}$$

где $S_{2m}=\sum\limits_{x=0}^{p-1}x^{2m}$. Очевидно, $S_{2m}\equiv -1 \pmod p$ при $m=\frac{p-1}{2}$. Докажем, что $S_{2m}\equiv 0 \pmod p$ при остальных значениях m, не превосходящих $p-1$. Действительно, для каждого такого m можно подобрать ненулевой остаток a, такой что $a^{2m}\not\equiv 1 \pmod p$ и тогда можно провести рассуждение как в (5). Возвращаясь к интересующей нас сумме (6), получаем

$$\sum_{m=0}^{p-2}C_{p-2}^m S_{2m}\equiv -C_{p-2}^{\frac{p-1}{2}}=-C_{4k+1}^{2k+1}=-\frac{(4k+1)\cdot 4k\cdot\ldots\cdot(2k+1)}{1\cdot 2\cdot\ldots\cdot(2k+1)}\equiv-\frac{(-2)\cdot(-3)\ldots(2k+2)}{1\cdot 2\cdot\ldots\cdot(2k+1)}\equiv 2k+2.$$

4.3. Мы нашли оба утверждения в [16].

a) Для каждого простого делителя p числа m подберем число a_p, для которого $(a_p^k-1)\not\vdots\, p$. С помощью китайской теоремы об остатках выберем число a, такое что $a\equiv a_p \pmod p$ при всех p. Теперь результат получается аналогично рассуждениям (5).

b) Заметим, что при нечетных k по формуле бинома $i^k+(p-i^k)\equiv ki^{k-1}p \pmod{p^2}$. Тогда

$$2\sum_{i=1}^{p-1}\frac{1}{i^k}=\sum_{i=1}^{p-1}\left(\frac{1}{i^k}+\frac{1}{(p-i)^k}\right)=\sum_{i=1}^{p-1}\frac{i^k+(p-i)^k}{i^k(p-i)^k}\equiv\sum_{i=1}^{p-1}\frac{ki^{k-1}p}{i^k(-i)^k}\equiv -kp\sum_{i=1}^{p-1}\frac{1}{i^{k+1}} \pmod{p^2}.$$

Сумма в правой части сравнения делится на p в силу утверждения п. a).

4.4. Как доказывается в [24], сравнение выполнено даже по модулю p^7, но мы не будем заходить так далеко. Действуя как в статье Винберга [1], но следя за степенями до p^4, получаем

$$C_{p-1}^{2p-1}=\frac{(2p-1)(2p-2)\cdot\ldots\cdot(p+1)}{p!}=\left(\frac{2p}{1}-1\right)\left(\frac{2p}{2}-1\right)\cdot\ldots\cdot\left(\frac{2p}{p-1}-1\right)\equiv$$
$$\equiv 1-2p\sum_{i=1}^{p-1}\frac{1}{i}+4p^2\sum_{\substack{i,j=1\\ i<j}}^{p-1}\frac{1}{ij}-8p^3\sum_{\substack{i,j,k=1\\ i<j<k}}^{p-1}\frac{1}{ijk} \pmod{p^4}. \tag{7}$$

Выразим последнюю сумму через степенные суммы:

$$\sum_{\substack{i,j,k=1\\ i<j<k}}^{p-1} \frac{1}{ijk} = \frac{S_3}{3} - \frac{S_1S_2}{2} + \frac{S_1^3}{6}\,, \quad \text{где } S_k = \sum_{i=1}^{p-1} \frac{1}{i^k}.$$

Как мы знаем, S_1 и S_3 делятся на p^2 (последнее — из задачи 4.3б). Поэтому последнее слагаемое в формуле (7) можно отбросить.

4.5. Задача из [1], обсуждение вариаций на эту тему можно прочесть в [14].

Поскольку

$$2\sum_{k=1}^{p-1} \frac{1}{k^2} = \sum_{k=1}^{p-1}\Big(\frac{1}{k^2} + \frac{1}{(p-k)^2}\Big) = \sum_{k=1}^{p-1} \frac{k^2+(p-k)^2}{k^2(p-k)^2} \equiv -2\sum_{k=1}^{p-1} \frac{1}{k(p-k)} \pmod{p^2},$$

утверждение 3) эквивалентно соотношению $\displaystyle\sum_{k=1}^{p-1} \frac{1}{k(p-k)} \equiv 0 \pmod{p^2}$. Утверждение 2) тоже эквивалентно этому соотношению, так как $\displaystyle 2\sum_{k=1}^{p-1} \frac{1}{k} = 2\sum_{k=1}^{p-1}\Big(\frac{1}{k} + \frac{1}{p-k}\Big) = 2p\sum_{k=1}^{p-1} \frac{1}{k(p-k)}$. Наконец, как мы знаем из предыдущей задачи,

$$C_{2p-1}^{p-1} \equiv 1 - p^2\sum_{i=1}^{p-1} \frac{1}{i(p-i)} + 4p^2 \sum_{\substack{i,j=1\\ i<j}}^{p-1} \frac{1}{ij} \pmod{p^4}.$$

Таким образом, утверждение 1) эквивалентно сравнению

$$\sum_{i=1}^{p-1} \frac{1}{i(p-i)} \equiv 4 \sum_{\substack{i,j=1\\ i<j}}^{p-1} \frac{1}{ij} \pmod{p^2}. \tag{8}$$

Преобразуем выражение в правой части:

$$4 \sum_{\substack{i,j=1\\ i<j}}^{p-1} \frac{1}{ij} = 2\Big(\sum_{k=1}^{p-1} \frac{1}{i}\Big)^2 - 2\sum_{k=1}^{p-1} \frac{1}{k^2} \equiv 2\Big(\sum_{k=1}^{p-1} \frac{1}{i}\Big)^2 + 2\sum_{k=1}^{p-1} \frac{1}{k(p-k)}.$$

Сумма в скобке делится на p, ее квадрат делится на p^2 и это слагаемое можно отбросить. Подставляя в (8), получаем, что и первое утверждение равносильно сравнению $\displaystyle\sum_{k=1}^{p-1} \frac{1}{k(p-k)} \equiv 0 \pmod{p^2}$.

4.6. а) Р е ш е н и е 1 ([5, предложение 2.12]). Индукция по n. Раскроем скобки в равенстве

$$(a+b)^{pn} = (a+b)^{p(n-1)}(a+b)^p$$

Приравняем коэффициенты при $a^{pm}b^{p(n-m)}$:

$$C_{pn}^{pm} = C_{p(n-1)}^{pm} C_p^0 + C_{p(n-1)}^{pm-1} C_p^1 + \ldots + C_{p(n-1)}^{pm-p+1} C_p^{p-1} + C_{p(n-1)}^{pm-p} C_p^p\,.$$

В правой части все слагаемые, кроме крайних, делятся на p^2, потому что каждый биномиальный коэффициент в них делится на p по теореме Люка. Следовательно,

$$C_{pn}^{pm} \equiv C_{p(n-1)}^{pm} + C_{p(n-1)}^{p(m-1)} \pmod{p^2}.$$

По предположению индукции

$$C_{p(n-1)}^{pm} + C_{p(n-1)}^{p(m-1)} \equiv C_{n-1}^m + C_{n-1}^{m-1} \equiv C_n^m \pmod{p^2}.$$

Р е ш е н и е 2 ([Д]). Докажем, что $C_{kp}^{mp} \equiv C_k^m \pmod{p^2}$ индукцией по m.

База $m = 1$. Требуется проверить, что $C_{pk}^p - C_k^1 \equiv 0 \pmod{p^2}$. Преобразуем эту разность:

$$C_{pk}^p - C_k^1 = \frac{pk(pk-1)\dots(pk-p+1)}{p!} - k = \left(\frac{(pk-1)(pk-1)\dots(pk-p+1)}{(p-1)!} - 1\right). \tag{9}$$

В числителе большой дроби четное число сомножителей. Разобьем их на пары:

$$(pk-i)(pk-p+i) \equiv pi^2 - i^2 \pmod{p^2}.$$

Как видим, по модулю p^2 произведение чисел в парах не зависит от k. Поэтому вычисление разности (9) по модулю p^2 дает одинаковый результат при всех k. Но при $k = 1$ вычисляемое выражение равно 0.

Переход. Пусть $C_{kp}^{(m-1)p} \equiv C_k^{m-1} \pmod{p^2}$.

$$C_{kp}^{mp} = C_{kp}^{(m-1)p} \cdot \frac{(p(k-m)+1)(p(k-m)+1)\dots(p(k-m)+p)}{pm(pm-1)\dots(pm-p+1)} =$$
$$= C_{kp}^{(m-1)p} \cdot \frac{(p(k-m)+1)(p(k-m)+1)\dots(p(k-m)+p-1)}{(pm-1)\dots(pm-p+1)} \cdot \frac{k-m+1}{m} \tag{10}$$

Отметим, что обе дроби корректно определены по модулю p^2. Как и в доказательстве базы, выражение в числителе большой дроби по модулю p^2 не зависит от k. Тогда для вычисления большой дроби можно взять $k = 0$, и мы сразу получим, что по модулю p^2 дробь равна 0. Пользуясь этим соображением и предположением индукции, мы можем заменить правую часть (10) на

$$\equiv C_k^{m-1} \cdot \frac{k-m+1}{m} = C_k^m \pmod{p^2}.$$

b) Р е ш е н и е 1 (комбинаторное). Как и рекомендуется в [1], рассматриваем выборки kp предметов из общего количества pn предметов. Полагаем, что исходное множество предметов разбито на блоки по p штук. Количество блочных выборок равно C_n^k. Таким образом, остается проверить, что количество неблочных выборок делится на p^3. Как объясняется в статье, количество неблочных выборок с тремя и более блоками делится на p^3. Так как при $k > 1$ любая неблочная выборка содержит не менее трех блоков, то в этом случае все доказано. Остается разобрать случай, когда $k = 1$ и мы подсчитываем количество неблочных выборок p предметов из из общего множества в $2p$ предметов. Это количество равно $C_{2p}^p - 2$, что по теореме Волстенхолма делится на p^3.

Р е ш е н и е 2. Напишем формулу для биномиального коэффициента $C_a^b = \frac{a(a-1)\dots(a-b+1)}{b(b-1)\dots1}$, разбив числитель и знаменатель на блоки из p сомножителей, после чего сократим первые множители в каждом блоке, а частные соберем в отдельное выражение:

$$C_{mp}^{kp} = \frac{m\not{p}\cdot(mp-1)\dots\big(mp-(p-1)\big)}{k\not{p}\cdot(kp-1)\dots\big(kp-(p-1)\big)} \cdot \frac{(m-1)\not{p}\cdot\big((m-1)p-1\big)\dots\big((m-1)p-(p-1)\big)}{(k-1)\not{p}\cdot\big((k-1)p-1\big)\dots\big((k-1)p-(p-1)\big)} \cdot \dots \times$$
$$\times \frac{(m-k+1)\not{p}\cdot\big((m-k+1)p-1\big)\dots\big((m-k+1)p-(p-1)\big)}{\not{p}\cdot(p-1)\dots1} =$$
$$= C_m^k \cdot \frac{(mp-1)\dots\big(mp-(p-1)\big)}{(kp-1)\dots\big(kp-(p-1)\big)} \cdot \dots \cdot \frac{\big((m-k+1)p-1\big)\dots\big((m-k+1)p-(p-1)\big)}{(p-1)\dots1}.$$

Осталось проверить, что произведение дробей дает остаток 1 при делении на p^3. Для этого достаточно проверить сравнение

$$\frac{(np-1)\dots\big(np-(p-1)\big)}{(rp-1)\dots\big(rp-(p-1)\big)} \equiv 1 \pmod{p^3}$$

или, лучше, вот такое сравнение

$$\frac{(np-1)\dots\big(np-(p-1)\big)}{(p-1)!} \equiv \frac{(rp-1)\dots\big(rp-(p-1)\big)}{(p-1)!} \pmod{p^3}.$$

Это верно, так как обе части сравнимы с 1 по модулю p^3, что устанавливается аналогично доказательству теоремы Волстенхолма.

4.7. a) [5, теорема 2.14]. Преобразуем разность

$$C_{p^2}^p - C_p^1 = \frac{p^2(p^2-1)\dots(p^2-(p-1))}{1\cdot 2\cdot\ldots\cdot(p-1)p} - p = \frac{p}{(p-1)!}\Big((1-p^2)(2-p^2)\dots((p-1)-p^2) - 1\cdot 2\cdot\ldots\cdot(p-1)\Big).$$

Осталось проверить, что

$$(1-p^2)(2-p^2)\dots((p-1)-p^2) \equiv 1\cdot 2\cdot\ldots\cdot(p-1) \pmod{p^4}.$$

Раскроем скобки в левой части:

$$(1-p^2)(2-p^2)\dots((p-1)-p^2) = 1\cdot 2\cdot\ldots\cdot(p-1) + p^2\Big(1+\frac{1}{2}+\ldots+\frac{1}{p-1}\Big)(p-1)! + \text{члены делящиеся на } p^4.$$

По утверждению задачи 4.1 второе слагаемое делится на p^4.

b) Как нетрудно видеть, $C_{p^{s+1}}^p = p^s \cdot C_{p^{s+1}-1}^{p-1}$, поэтому достаточно проверить, что $C_{p^{s+1}-1}^{p-1} \equiv 1 \pmod{p^{s+3}}$.

$$C_{p^{s+1}-1}^{p-1} = \frac{(p^{s+1}-1)(p^{s+1}-2)\dots(p^{s+1}-(p-1))}{1\cdot 2\cdots(p-1)} = \left(\frac{p^{s+1}}{1}-1\right)\left(\frac{p^{s+1}}{2}-1\right)\dots\left(\frac{p^{s+1}}{p-1}-1\right) \equiv$$
$$\equiv (-1)^{p-1} + p^{s+1}\Big(1+\frac{1}{2}+\ldots+\frac{1}{p-1}\Big) \pmod{p^{s+3}}.$$

Это и есть то, что требуется, поскольку $(-1)^{p-1} = 1$ и $1+\frac{1}{2}+\ldots+\frac{1}{p-1} \equiv 0 \mod p^2$.

В статье [14] доказывается чуть более общий факт.

4.8. Задача из статьи Винберга [1], решение [T].

$$C_{p^3}^{p^2} - C_{p^2}^p = p\left(C_{p^3-1}^{p^2-1} - C_{p^2-1}^{p-1}\right) =$$
$$= p\left(\left(\frac{p^3}{1}-1\right)\left(\frac{p^3}{2}-1\right)\dots\left(\frac{p^3}{p^2-1}-1\right) - \left(\frac{p^2}{1}-1\right)\left(\frac{p^2}{2}-1\right)\dots\left(\frac{p^2}{p-1}-1\right)\right) =$$
$$= p\Big(\frac{p^2}{1}-1\Big)\Big(\frac{p^2}{2}-1\Big)\dots\Big(\frac{p^2}{p-1}-1\Big)\left(\prod_{\substack{k=1\\ p\nmid k}}^{p^2-1}\left(\frac{p^3}{k}-1\right)-1\right).$$

Достаточно проверить, что выражение в последней скобке делится на p^7. Преобразуем произведение

$$\prod_{\substack{k=1\\ p\nmid k}}^{p^2-1}\left(\frac{p^3}{k}-1\right) = \prod_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\left(\frac{p^3}{k}-1\right)\left(\frac{p^3}{p^2-k}-1\right) = \prod_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\left(\frac{p^6-p^5}{k(p^2-k)}+1\right) \equiv 1+p^5(p-1)\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k(p^2-k)} \pmod{p^7}.$$

Осталось проверить, что последняя сумма делится на p^2. Это так, поскольку по задаче 4.3a)

$$\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k(p^2-k)} \equiv -\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k^2} \equiv 0 \pmod{p^2}.$$

4.9. Это [6, теорема 5]. Более общий факт доказан в [7].

Р е ш е н и е 1 ([5, предложение 2.19]). Воспользуемся тем, что разность $C_{2^{k+1}}^{2^k} - C_{2^k}^{2^{k-1}}$ равна коэффициенту при x^{2^k} в многочлене

$$(1+x)^{2^{k+1}} - (1-x^2)^{2^k} = (1+x)^{2^k}\Big((1+x)^{2^k} - (1-x)^{2^k}\Big) =$$
$$= \Big(1 + C_{2^k}^1 x + C_{2^k}^2 x^2 + \ldots + x^{2^k}\Big)\cdot 2\Big(C_{2^k}^1 x + C_{2^k}^3 x^3 + \ldots + C_{2^k}^{2^k-1} x^{2^k-1}\Big).$$

Поскольку второй многочлен содержит только множители нечетной степени, коэффициент при x^{2^k} в произведении равен

$$2\Big(C_{2^k}^1 C_{2^k}^{2^k-1} + C_{2^k}^3 C_{2^k}^{2^k-3} + \ldots + C_{2^k}^{2^k-1} C_{2^k}^1\Big).$$

По утверждению задачи 3.5 б) каждый биномиальный коэффициент в этом выражении делится на 2^k, кроме того, каждое слагаемое в сумме встречается 2 раза, а перед суммой стоит коэффициент 2. В итоге все выражение делится на 2^{2k+2}.

Решение 2 ([CSTTVZ]). Так как $C_{2^{n+1}}^{2^n} = 2C_{2^{n+1}-1}^{2^n-1}$, достаточно доказать соотношение

$$C_{2^{n+1}-1}^{2^n-1} \equiv C_{2^n-1}^{2^{n-1}-1} \pmod{2^{2n+1}}.$$

Аналогично (3) получаем

$$C_{2^{n+1}-1}^{2^n-1} = \left(\frac{2^{n+1}}{1} - 1\right)\left(\frac{2^{n+1}}{3} - 1\right)\ldots\left(\frac{2^{n+1}}{2^n-1} - 1\right)\cdot C_{2^n-1}^{2^{n-1}-1}.$$

Достаточно проверить, что

$$L = \left(\frac{2^{n+1}}{1} - 1\right)\left(\frac{2^{n+1}}{3} - 1\right)\ldots\left(\frac{2^{n+1}}{2^n-1} - 1\right) \equiv 1 \pmod{2^{2n+1}}.$$

Это так, поскольку

$$L \equiv (-1)^{2^{n-1}} - 2^{n+1}\left(\frac{1}{1} + \frac{1}{3} + \frac{1}{5} + \ldots + \frac{1}{2^n-1}\right) \equiv$$
$$\equiv 1 - 2^{n+1}\left(\frac{2^n}{1\cdot(2^n-1)} + \frac{2^n}{3\cdot(2^n-3)} + \ldots + \frac{2^n}{(2^{n-1}-1)(2^{n-1}+1)}\right) \equiv 1 \pmod{2^{2n+1}}.$$

4.10. Это теорема Морли [26].

Решение 1 (авторское решение из статьи 1895 года). Оно лишь чуть-чуть выходит за рамки школьной программы.

Возьмем формулу, с помощью которой $\cos^{2n+1} x$ выражается через косинусы кратных углов,[1] или, как говорили в те времена, запишем $\cos^{2n+1} x$ в виде, удобном для интегрирования:

$$2^{2n}\cos^{2n+1}x = \cos(2n{+}1)x + (2n{+}1)\cos(2n{-}1)x + \frac{(2n{+}1)\cdot 2n}{1\cdot 2}\cos(2n{-}3)x + \ldots + \frac{(2n{+}1)\cdot 2n\ldots(n{+}2)}{n!}\cos x.$$

Ну, а раз этот вид удобен для интегрирования, то и проинтегрируем обе части[2] по промежутку $[0, \frac{\pi}{2}]$:

$$2^{2n}\int \cos^{2n+1} x\,dx = \frac{\sin(2n+1)x}{2n+1} + \frac{2n+1}{2n-1}\sin(2n-1)x + \ldots,$$
$$2^{2n}\int_0^{\pi/2} \cos^{2n+1} x\,dx = (-)^n\left(\frac{1}{2n+1} - \frac{2n+1}{2n-1} + \ldots\right).$$

Но любой первокурсник знает, что куда проще этот интеграл вычисляется с помощью формулы понижения, для получения которой нужно всего лишь проинтегрировать по частям:

$$I_{2n+1} = \int_0^{\pi/2} \cos^{2n+1} x\,dx = \int_0^{\pi/2} \cos^{2n} x\cos x\,dx = \cos^{2n} x\sin x\,\Big|_0^{\pi/2} + 2n\int_0^{\pi/2} \cos^{2n-1} x\sin^2 x\,dx =$$
$$= 0 + 2n\int_0^{\pi/2} \cos^{2n-1} x(1-\cos^2 x)\,dx = 2n\cdot I_{2n-1} - 2n\cdot I_{2n+1},$$

[1] Читатель, интересующийся вопросом "где мы ее возьмем" и не удовлетворенный ответом "в справочнике", может просто воспользоваться формулой Эйлера $\cos\varphi = \frac{1}{2}(e^{i\varphi} + e^{-i\varphi})$ и возвести правую часть в степень $2n+1$ по формуле бинома.

[2] Когда мы учим правила умножения, мы запоминаем формулу "минус на минус будет плюс". В этой формуле мы перемножаем знаки. Значит, если нам нужно перемножить n минусов, кажется вполне уместной запись $(-)^n$. Поэтому мы оставляем старомодное обозначение $(-)^n$, как у автора, вместо современного $(-1)^n$.

откуда находим, что $I_{2n+1} = \dfrac{2n}{2n+1}\cdot I_{2n-1}$. Учитывая что $I_1 = 1$, применяя эту формулу n раз подряд, находим, что

$$\int\limits_0^{\pi/2} \cos^{2n+1} x\,dx = \frac{2n\cdot(2n-2)\dots 2}{(2n+1)(2n-1)\dots 3}\,.$$

Приравнивая эти два способа подсчета интеграла, мы получаем тождество

$$2^{2n}\frac{2n\cdot(2n-2)\dots 2}{(2n+1)(2n-1)\dots 3} = (-)^n\left(\frac{1}{2n+1} - \frac{2n+1}{2n-1} + \dots + \frac{(2n+1)\cdot 2n\dots(n+2)}{n!}\right).$$

Если взять $p = 2n+1$ — простое число, то домножая на p, мы сразу получаем требуемое сравнение

$$2^{2n}\frac{2n\cdot(2n-2)\dots 2}{(2n-1)(2n-3)\dots 3} \equiv (-)^n \pmod{p^2}\,.$$

Р е ш е н и е 2 ([CSTTVZ]). Введем несколько обозначений. Пусть

$$A = \sum_{i=1}^{\frac{p-1}{2}}\frac{1}{i}, \qquad B = \sum_{1\leqslant i<j\leqslant\frac{p-1}{2}}\frac{1}{ij}, \qquad C = \sum_{\substack{1\leqslant i\leqslant p-1\\ i \text{ нечетно}}}\frac{1}{i}.$$

Очевидно, $A^2 = \sum\limits_{i=1}^{\frac{p-1}{2}}\frac{1}{i^2} + 2B \equiv 2B \pmod p$ по задаче 4.3b). Итак, $A^2 \equiv 2B \pmod p$. Далее,

$$2C + A = \sum_{\substack{1\leqslant i\leqslant p-1\\ i \text{ нечетно}}}\frac{2}{i} + \sum_{i=1}^{\frac{p-1}{2}}\frac{2}{2i} = \sum_{i=1}^{p-1}\frac{2}{i} \equiv 0 \pmod{p^2}.$$

Таким образом, $C \equiv -\frac{1}{2}A \pmod{p^2}$.

Теперь преобразуем по модулю p^3 правую и левую части доказываемого сравнения. Левая часть:

$$(-1)^{\frac{p-1}{2}}C_{p-1}^{\frac{p-1}{2}} \equiv \left(1-\frac{p}{1}\right)\left(1-\frac{p}{2}\right)\dots\left(1-\frac{p}{\frac{p-1}{2}}\right) \equiv 1 - pA + p^2B \equiv 1 - pA + \frac{1}{2}p^2A^2 \pmod{p^3}.$$

Для преобразования правой части заметим, что

$$2^{p-1} = \frac{2\cdot 4\cdots(p-1)}{1\cdot 2\cdots\frac{p-1}{2}}\cdot\frac{(p+1)\cdots(2p-2)}{\frac{p+1}{2}\cdots(p-1)} = \frac{(p+1)\cdots(2p-2)}{1\cdot 3\cdot 5\cdots(p-2)} =$$
$$= \left(\frac{p}{1}+1\right)\left(\frac{p}{3}+1\right)\dots\left(\frac{p}{p-1}+1\right) \equiv 1 + pC + \frac{1}{2}p^2C^2 \equiv 1 - \frac{1}{2}pA + \frac{1}{8}p^2A^2 \pmod{p^3}\,.$$

Отсюда получаем

$$4^{p-1} \equiv \left(1 - \frac{1}{2}pA + \frac{1}{8}p^2A^2\right)^2 \equiv 1 - pA + \frac{1}{4}p^2A^2 + 2\cdot\frac{1}{8}P^2A^2 = 1 - pA + \frac{1}{2}p^2A^2 \pmod{p^3}.$$

Таким образом, левая часть эквивалентна правой.

4.11. Мы взяли утверждение в [10].

$$\sum_{k=1}^{p-1}\frac{1}{mp+k} = \frac{1}{2}\sum_{k=1}^{p-1}\left(\frac{1}{mp+k} + \frac{1}{mp+p-k}\right) =$$
$$= p\cdot\frac{2m+1}{2}\cdot\sum_{k=1}^{p-1}\frac{1}{(mp+k)(mp+p-k)} \equiv -p\cdot\frac{2m+1}{2}\cdot\sum_{k=1}^{p-1}\frac{1}{k^2} \equiv 0 \pmod{p^2}.$$

4.12. Мы взяли утверждение в [8]. Так как $2pq-1 = (2q-1)p+p-1$, у числа $2pq-1$ последняя цифра p-ичной записи — это $p-1$, а остальные цифры образуют запись числа $2q-1$. Аналогично в записи числа $pq-1$ последняя цифра — $p-1$, а остальные цифры образуют запись числа $q-1$. По теореме Люка $C_{2pq-1}^{pq-1} \equiv C_{2q-1}^{q-1}C_{p-1}^{p-1} \equiv C_{2q-1}^{q-1} \pmod p$. С другой стороны, очевидно, что так как $C_{2pq-1}^{pq-1} \equiv 1 \pmod{pq}$, то $C_{2pq-1}^{pq-1} \equiv 1 \pmod p$. Таким образом, $C_{2q-1}^{q-1} \equiv 1 \pmod p$. Аналогично $C_{2p-1}^{p-1} \equiv 1 \pmod q$.

В обратную сторону утверждение очевидно.

5 Суммы биномиальных коэффициентов

5.1. a) Это сразу следует из результата задачи 1.3. Если Δ_0^0 — это треугольник из трех первых строк 3-арифметического треугольника Паскаля, то, как нетрудно видеть сумма центральных коэффициентов в нем делится на 3. При произвольном a изучаемая сумма содержит элементы нескольких центральных треугольников, кратных Δ_0^0. Поэтому сумма тоже делится на 3.

Другое решение ([CSTTVZ]) получится, если мы воспользуемся тождеством $C_{2k}^k = \sum\limits_{i=0}^{k} {C_k^i}^2$. Тогда $\sum\limits_{k=0}^{3^a-1} C_{2k}^k = \sum\limits_{k=0}^{3^a-1}\sum\limits_{i=0}^{k} {C_k^i}^2$. Поскольку $1^2 = 2^2 = 1$, $0^2 = 0$ по модулю 3, последняя сумма равна по модулю 3 количеству ненулевых элементов в первых 3^a строках треугольника Паскаля. Это количество подсчитано в задаче 2.1a), оно делится на 3.

b) Приводим решение [Д]. Нужная нам сумма является коэффициентом при x^{3^a-1} многочлена

$$x^{3^a-1}\left(1+\frac{(x+1)^2}{x}+\frac{(x+1)^4}{x^2}+\ldots+\frac{(x+1)^{2(3^a-1)}}{x^{3^a-1}}\right)=\frac{\frac{(x+1)^{2\cdot 3^a}}{x^{3^a}}-1}{\frac{(x+1)^2}{x}-1}\cdot x^{3^a-1}=\frac{(x+1)^{2\cdot 3^a}-x^{3^a}}{x^2+x+1}=$$
$$=\frac{x^{2\cdot 3^a}+C_{2\cdot 3^a}^1\cdot x^{2\cdot 3^a-1}+C_{2\cdot 3^a}^2\cdot x^{2\cdot 3^a-2}+\ldots+1-x^{3^a}}{x^3-1}\cdot(x-1)\,.$$

Чтобы найти нужный коэффициент, достаточно поделить числитель на знаменатель “в столбик”, и потом домножить результат на $(x-1)$. Таким образом, не нужно даже доводить деление до конца, достаточно довести его до нахождения коэффициента при x^{3^a-2}, кроме того, напомним, результат нас интересует лишь по модулю 3^a. Отметим, что при $b \not\vdots 3$ все биномиальные коэффициенты $C_{2\cdot 3^a}^b$ делятся на 3^a по теореме Куммера. Сгруппируем слагаемые с этими коэффициентами и будем делить их сумму на x^3-1 отдельно. Очевидно, все коэффициенты частного будут тоже делиться на 3^a, поэтому все эти слагаемые можно отбросить. Остается выражение

$$\frac{x^{2\cdot 3^a}+C_{2\cdot 3^a}^3\cdot x^{2\cdot 3^a-3}+C_{2\cdot 3^a}^6\cdot x^{2\cdot 3^a-6}+\ldots+1-x^{3^a}}{x^3-1}\cdot(x-1).$$

Здесь все показатели в числителе делятся на 3, после деления на x^3-1 все показатели частного тоже будут делиться на 3, а когда мы домножим частное на $x-1$, у нас не появится ни одного показателя вида $3k+2$. Таким образом, искомый коэффициент по модулю 3^a равен 0.

5.2. Задача была опубликована в Monthly [25]. Так как

$$C_{2n+2}^{n+1}-4C_{2n}^n=2\cdot\frac{2n+1}{n+1}C_{2n}^n-4C_{2n}^n=-2C_n\,,$$

то $C_n \equiv C_{2n+2}^{n+1}-C_{2n}^n \pmod 3$. Поэтому сумма по модулю 3 является телескопической,

$$\sum_{k=1}^{n} C_k \equiv \left(C_{2n+2}^{n+1}-C_{2n}^n\right)+\left(C_{2n}^n-C_{2n-2}^{n-1}+\ldots\right)=C_{2n+2}^{n+1}+1 \pmod 3.$$

Таким образом, по теореме Куммера нам остается выяснить, в каком случае сложение числа $(n+1)$ с самим собой в троичной системе счисления приводит к появлению хотя бы одного переноса. Очевидно, это может быть в том и только том случае, когда в записи $n+1$ есть хотя бы одна двойка.

5.3. Это задача A5 Putnam Mathematical Competition, 1998. Поскольку $\frac{1}{p}C_p^n \equiv \frac{(-1)^{n-1}}{n} \pmod p$, получаем, что

$$\sum_{n=1}^{k}\frac{1}{p}C_p^n \equiv \sum_{n=1}^{k}\frac{(-1)^{n-1}}{n}=\sum_{n=1}^{k}\frac{1}{n}-2\sum_{n=1}^{[k/2]}\frac{1}{2n}\equiv\sum_{n=1}^{k}\frac{1}{n}+\sum_{n=p-[\frac{k}{2}]}^{p-1}\frac{1}{n}\overset{*}{=}\sum_{n=1}^{p-1}\frac{1}{n}\equiv 0 \pmod p.$$

В сумме, расположенной непосредственно слева от равенства, помеченного звездочкой, на самом деле суммирование ведется от $n = k + 1$ (в этом нетрудно убедиться: при $p = 6r + 1$ имеем $k = 4r$ и $p - [\frac{k}{2}] = 4r + 1 = k + 1$, аналогично при $p = 6r + 5$).

5.4. Это утверждение из [11]. Решение [CSTTVZ]. Индукция по n. База тривиальна. Докажем переход от $n' = n - (p - 1)$ к n. Пусть $q = \frac{n}{p-1}$. Так как

$$C_{n'+p-1}^{x(p-1)} = \sum_{i=0}^{p-1} C_{p-1}^{i} C_{n'}^{x(p-1)-i},$$

мы можем записать изучаемую сумму в виде

$$C_n^{p-1} + C_n^{2(p-1)} + C_n^{3(p-1)} + \ldots = \sum_{x=1}^{q} \sum_{i=0}^{p-1} C_{p-1}^{i} C_{n'}^{x(p-1)-i} = \sum_{i=0}^{p-1} \left(C_{p-1}^{i} \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} \right) \tag{11}$$

По утверждению задачи 1.1 а) $C_{p-1}^{i} \equiv (-1)^i \pmod p$, пусть $C_{p-1}^{i} = ap + (-1)^i$. По утверждению задачи 1.6 при $i = 0, 1, \ldots, p-2$ выполнено сравнение $\sum\limits_{x=1}^{q} C_{n'}^{x(p-1)-i} \equiv C_{p-1}^{i} \equiv (-1)^i \pmod p$; пусть $\sum\limits_{x=1}^{q} C_{n'}^{x(p-1)-i} = bp + (-1)^i$. Тогда

$$C_{p-1}^{i} \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} = \big(ap + (-1)^i\big)\big(bp + (-1)^i\big) \equiv 1 + (-1)^i(ap + bp) =$$
$$= 1 + (-1)^i \left(C_{p-1}^{i} + \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} - 2 \cdot (-1)^i \right) = (-1)^i \left(C_{p-1}^{i} + \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} \right) - 1 \pmod{p^2}.$$

Напомним, что это преобразование верно при $0 \leqslant i \leqslant p - 2$. Мы можем продолжить равенство (11), выделив отдельное слагаемое для $i = p - 1$:

$$\sum_{i=0}^{p-1} \left(C_{p-1}^{i} \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} \right) \equiv \sum_{i=0}^{p-2} \left((-1)^i \left(C_{p-1}^{i} + \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} \right) - 1 \right) + \sum_{x=0}^{q-1} C_{n'}^{x(p-1)} =$$
$$= \sum_{i=0}^{p-2} (-1)^i C_{p-1}^{i} + \sum_{i=0}^{p-2} \left((-1)^i \sum_{x=1}^{q} C_{n'}^{x(p-1)-i} \right) - (p-1) + C_{n'}^{0} + \sum_{x=1}^{q-1} C_{n'}^{x(p-1)}.$$

Здесь первая сумма равна -1, так как знакопеременная сумма $C_{p-1}^{0} - C_{p-1}^{1} + C_{p-1}^{2} + \ldots$ равна 0. По той же причине вторая (двойная) сумма вместе со слагаемым $C_{n'}^{0}$ равна 0. Последняя же сумма по предположению индукции равна $1 + p(n' + 1)$. Итого все выражение равно $-1 + 0 - p + 1 + 1 + p(n' + 1) = 1 + pn'$. Это как раз то, что требуется, поскольку $1 + p(n + 1) = 1 + p(n' + p - 1 + 1) \equiv 1 + pn' \pmod{p^2}$.

5.5. Это результат Флека, 1913 г., мы узнали о нем из [18]. Решение [CSTTVZ].

При $p = 2$ сумма не знакопеременная и результат очевиден. Далее считаем, что p нечетно. Индукция по q. База следует из утверждения задачи 2.5 а). Докажем переход от $n' = n - (p-1)$ к n. Ниже выражение $\sum_x$ обозначает суммирование по x в естественных границах (т. е. в границах для которых определены биномиальные коэффициенты под знаком суммирования).

$$\pm \sum_{m:m \equiv j \pmod p} (-1)^m C_n^m = \sum_x (-1)^x C_{n'+p-1}^{xp+j} = \sum_x (-1)^x \sum_{i=0}^{p-1} C_{p-1}^{i} C_{n'}^{xp+j-i} = \sum_{i=0}^{p-1} C_{p-1}^{i} \sum_x (-1)^x C_{n'}^{xp+j-i}$$

По предположению индукции $\sum\limits_x (-1)^x C_{n'}^{xp+j-i}$ делится на p^{q-1}, по утверждению задачи 1.1 а) $C_{p-1}^{i} \equiv (-1)^i \pmod p$, следовательно,

$$\sum_{i=0}^{p-1} C_{p-1}^{i} \sum_x (-1)^x C_{n'}^{xp+j-i} \equiv \sum_{i=0}^{p-1} (-1)^i \sum_x (-1)^x C_{n'}^{xp+j-i} \pmod{p^q}.$$

Внимательно посмотрев на последнюю двойную сумму, можно заметить, что это она равна $C_{n'}^{0} - C_{n'}^{1} + C_{n'}^{2} - C_{n'}^{3} + \ldots = 0$.

5.6. Это результат Баскарана (1965 г.), мы взяли его в [18], решение [CSTTVZ].

Обозначим

$$f(n,j) = C_n^j - C_n^{j+(p-1)} + C_n^{j+2(p-1)} - C_n^{j+3(p-1)} + \dots$$

Индукция по n. База $n = p+1$ тривиальна, отметим лишь, что $C_{p+1}^i \equiv 1 \pmod p$ при $i = 0, 1, p, p+1$, а в остальных случаях этот биномиальный коэффициент делится на p. Докажем индукционный переход от $n' = n - (p+1)$ к n. Благодаря сделанному замечанию,

$$C_{n'+(p+1)}^{j+(p-1)k} = \sum_{i=0}^{p+1} C_{n'}^{j+(p-1)k-i} C_{p+1}^i \equiv \sum_{i\in\{0,1,p,p+1\}} C_{n'}^{j+(p-1)k-i} =$$

$$= C_{n'}^{j+(p-1)k} + \underline{C_{n'}^{j-1+(p-1)k} + C_{n'}^{j-1+(p-1)(k-1)}} + C_{n'}^{j-2+(p-1)(k-1)} \pmod p .$$

Поскольку $f(n,j) = \sum_k (-1)^k C_n^{j+k(p-1)}$ — знакочередующаяся сумма, при суммировании по k подчеркнутые выражения сократятся в типовом слагаемом (а несократившиеся выражения в крайних слагаемых равны 0 по причине некорректности биномиального коэффициента). Таким образом, мы получаем соотношения

$$f(n,j) \equiv f(n',j) - f(n',j-2) \quad \text{при } j > 1 , \qquad f(n,1) \equiv f(n',1) + f(n',p-2) .$$

Теперь часть "только тогда" доказываемого утверждения сразу следует из индукционного предположения, а часть "тогда" в общем-то тоже: если $f(n,j) \equiv 0 \pmod p$ при $j = 1, 3, \dots, p-2$, то

$$f(n',p-2) \equiv f(n',p-4) \equiv \dots \equiv f(n',1) \equiv -f(n',p-2) ,$$

откуда $f(n',j) \equiv 0 \pmod p$ при всех нужных j и тогда $n' \vdots (p+1)$, а тогда и $n \vdots (p+1)$.

Ссылки

Авторы многих приведенных решений — участники конференции, в таких решениях мы ставили ссылки:

[Д] Максим Дидин;
[K] Дмитрий Креков;
[J] Jastin Lim Kai Ze;
[T] Teh Zhao Yang Anzo;
[CSTTVZ] Ćevid Domagoj, Stokić Maksim, Tanasijeviić Ivan, Trifunović Petar, Vukorepa Borna, Žikelić Đorđe

Литература

[1] *Винберг Э. Б.* Удивительные свойства биномиальных коэффициентов. // Мат. просвещение. Третья серия. Вып. 12. 2008

[2] *Гашков С.Б., Чубариков В.Н.* Арифметика. Алгоритмы. Сложность вычислений. М.: Высш. шк., 2000.

[3] *Дынкин Е.Б., Успенский В.А.* Математические беседы. 2-е изд. М.: ФИЗМАТЛИТ, 2004.

[4] Петербургские математические олимпиады, 1961–1993. СПб: Лань, 2007.

[5] *Табачников С.Л., Фукс Д.Б.* Математический дивертисмент. 30 лекций по классической математике. М.: МЦНМО, 2011.

[6] *Фукс Д.Б., Фукс М.Б.* Арифметика биномиальных коэффициентов // Квант. 1970. №6. С. 17–25.

[7] *Ширшов А.И.* Об одном свойстве биномиальных коэффициентов // Квант. 1971. №10. С. 16–20.

[8] *Cai T.X., Granville A.* On the residues of binomial coefficients and their products modulo prime powers /!/ Acta

[9] *Calkin N. J.* Factors of sums of powers of binomial coefficients // Acta Arith. 1998. Vol. 86. P. 17–26.

[10] *Carlitz L.* A note of Wolstenholme's theorem // Amer. Math. Monthly. 1954. Vol. 61. № 3. P. 174–176.

[11] *Dimitrov V., Chapman R.* Binomial coefficient identity: 11118 // Amer. Math. Monthly. 2006. Vol. 113. № 7. P. 657–658.

[12] *Everett W.* Subprime factorization and the numbers of binomial coefficients exactly divided by powers of a prime // Integers. 2011. Vol. 11. # A63. `http://www.integers-ejcnt.org/vol11.html`

[13] *Fine N.* Binomial coefficient modulo a prime // Amer. Math. Monthly. 1947. Vol. 54. № 10. Part 1. P. 589–592.

[14] *Gardiner A.* Four problems on prime power divisibility // Amer. Math. Monthly. 1988. Vol. 95. № 10. P. 926–931.

[15] *Gauss K.* Disquisitiones arithmeticae. 1801. Art. 78.

[16] *Gessel I.* Wolstenholme revisited // Amer. Math. Monthly. 1998. Vol. 105. № 7. P. 657–658.

[17] *Granville A.* Arithmetic properties of binomial coefficients. Доступно по адресу `http://www.dms.umontreal.ca/~andrew/Binomial/`

[18] *Granville A.* Binomial coefficients modulo prime powers.

[19] *Granville A.* Zaphod Beeblebrox's Brian and the Fifty-ninth Row of Pascal's Triangle // Amer. Math. Monthly. 1992. Vol. 99. № 4. P. 318–331.

[20] *Granville A.* Correction to: Zaphod Beeblebrox's Brian and the Fifty-ninth Row of Pascal's Triangle // Amer. Math. Monthly. 1997. Vol. 104. № 9. P. 848–851.

[21] *Hinz A.* Pascal's triangle and tower of Hanoi // Amer. Math. Monthly. 1992. Vol. 99. № 6. P. 538–544.

[22] *Loveless A.* A congruence for products of binomial coefficients modulo a composite // Integers: electronic journal of comb. number theory 7 (2007) # A44

[23] *McIntosh R.* On the converse of Wolstenhome's theorem // Acta Arithmetica. 1995. Vol. 61. №4. P. 381–388.

[24] *Meštrović R.* On the mod p^7 determination of $\binom{2p-1}{p-1}$ // `http://arxiv.org/pdf/1108.1174v1.pdf`

[25] *More Y., Chapman R.* The sum of Catalan numbers, modulo 3: 11165 // Amer. Math. Monthly. 2007. Vol. 114. № 5. P. 454–455.

[26] *Morley F.* Note on the congruences $2^{4n} \equiv (-)^n(2n)!/(n!^2)$, where $2n+1$ is a prime // Annals of Math. 1894-1895. Vol. 9. №1. P. 168–170.

[27] *Roberts J.* On binomial coefficient residues // Canad J. Math. 1957. Vol. 9. P. 363–370.

[28] Sun Z.-W., Wan D. On Fleck quotients // `arXxiv:math.0603462v3`

Приложения к теории колец, немного истории

А. Я. Белов, М. И. Харитонов

2 августа 2012 г.

Рассмотрим множество, с элементами которого можно проводить операции сложения, причем

1. для любых элементов a, b, c, $a+b=b+a$, $(a+b)+c=a+(b+c)$;
2. есть специальный элемент 0 такой, что $a+0=a$ для всех a;
3. для каждого элемента a существует противоположный $(-a)$ такой, что $(-a)+a=a+(-a)=0$.

   Примеры: различные остатки при делении на n; повороты плоскости на различные углы относительно фиксированной точки и т.д.

   Можно также добавить еще одну естественную операцию, умножение, также удовлетворяющую *ассоциативности*:

4. $a(bc)=(ab)c$ для всех a, b, c.

   Если при всех этих свойствах умножение и сложение удовлетворяют условию *дистрибутивности:*

5. $a(b+c)=ab+ac, (b+c)a=ba+ca$ для всех a, b, c, то множество с указанными операциями называется кольцом.

Примеры: опять же остатки (вычеты по модулю); многочлены от одного или нескольких элементов.

Задача 0.1 *Приведите примеры колец с делителями нуля, то есть колец, где есть элементы a, b такие, что $ab=0$;*

Определение 0.1 Единицей *кольца или* нейтральным элементом *называется элемент E такой, что $EA=AE=A$ для всех $A\in R$.* Обратный элемент A^{-1} *определяется равенством $AA^{-1}=E$.*

Задача 0.2 *Докажите, что в кольце с единицей аксиома коммутативности сложения вытекает из других аксиом.*

Задача 0.3 *Постройте кольцо из 4 элементов, каждый ненулевой элемент которого обратим.*

Задача 0.4 *Постройте некоммутативное кольцо, то есть такое, что существуют a, b такие, что $ab\neq ba$;*

Задача 0.5 *Пусть R кольцо, где любая сумма нескольких единиц не равна нулю. пусть элементы e, f, g таковы, что $ee = e, ff = f, gg = g$ и $e+f+g = 0$. Докажите, что $e = f = g = 0$.*

Определение 0.2 *Многочлен $f = f(x_1, \ldots, x_n)$ называется* тождеством *кольца A, если он тождественно обращается в ноль на A. Тождество f следует из набора тождеств $\{g_i\}$, если везде, где выполняется набор $\{g_i\}$, также выполняется f. Иногда пишут тождество в виде $f = 0$.*

Задача 0.6 *Существуют ли нетривиальные кольца, удовлетворяющие тождествам $x^2 = 0$; $xy = yx$?*

Определение 0.3 *Если каждый ненулевой элемент обратим, то кольцо называется* телом. *Кольцо* коммутативно, *если $ab = ba$ для всех $a, b \in R$. Коммутативное тело есть* поле.

Примеры.

Поле вещественных чисел, поле комплексных чисел, кольцо многочленов, кольцо Z_n остатков по модулю n. Если n простое, то Z_n – поле.

Некоммутативное кольцо матриц: если $A = (a_{ij})$ и $B = (b_{ij})$, то $A + B = A = (a_{ij} + b_{ij})$, $AB = (\sum_j a_{ij}b_{jk})$. Тело кватернионов есть множество выражений вида $ai + bj + ck + d$, где a, b, c, d есть вещественные числа и при этом $ij = -ji = k, ki = -ik = j, jk = -kj = i, i^2 = j^2 = k^2 = -1$.

Задача 0.7 *Проверьте, что приведенные выше объекты действительно являются кольцами.*

Определение 0.4 *Свободная ассоциативная алгебра или кольцо некоммутативных многочленов над кольцом R: это наборы выражений вида $\sum_i a_i v_i$, $a_i \in R$, v_i – слова. Если $v = \sum_i a_i v_i$, $u = \sum_i b_i v_i$, то $u + v = \sum_i (a_i + b_i)v_i$, $uv = \sum_{i,j} a_i b_j v_i v_j$. Свободная ассоциативная алгебра перестаёт быть свободной, если в ней выполняются некоторые тождества. Понятие тождества алгебры приведено ниже.*

Замечание 0.1 *Далее под алгеброй будем подразумевать ассоциативную алгебру.*

Определение 0.5 *Многочлен $f = f(x_1, \ldots, x_n)$ называется* тождеством *алгебры A, если он тождественно обращается в ноль на A. Тождество f следует из набора тождеств $\{g_i\}$, если везде, где выполняется набор $\{g_i\}$, также выполняется f. Иногда пишут тождество в виде $f = 0$.*

Примеры. Тождество *коммутативности* $[a, b] = ab - ba$ выполняется на всех коммутативных кольцах. В поле вычетов по простому p выполняется тождество $x^p - x$ (малая теорема Ферма). $(a + b)^2 = a^2 + ab + ba + b^2$, поэтому из тождества $x^2 = 0$ следует тождество $ab + ba = 0$.

Задача 0.8

1. *Докажите, что в алгебре матриц второго порядка выполняется тождество $[[x, y]^2, z] = 0$ (*тождество Холла*).*
2. *Докажите, что в алгебре матриц второго порядка выполняется тождество $\sum_{\sigma \in S_4} (-1)^\sigma x_{\sigma(1)} \cdots x_{\sigma(4)} = 0$ (*стандартное тождество *степени* 4*).*

Замечание. В алгебре матриц n-го порядка выполняется стандартное тождество степени $2n$ (теорема Амицура-Левицкого). Известно, что все тождества алгебры матриц второго порядка вытекают из стандартного тождества степени 4 и тождества Холла (это довольно трудная теорема, доказанная Ю. П. Размысловым в 1973 году). Однако базис тождеств неизвестен даже для матриц третьего порядка вплоть до настоящего времени.

Определение 0.6 *Алгебра A называется* ниль-алгеброй, *если есть функция $n : A \to \mathbf{N}$ такая, что для любого $x \in A$ выполняется равенство $x^{n(x)} = 0$. Если же в ней выполняется тождество $x^n = 0$, то A называется* ниль-алгеброй индекса n. *A —* нильпотентна индекса k, *если в ней выполняется тождество $x_1 \cdots x_k = 0$, A —* нильпотентна, *если она нильпотентна индекса k при некотором k. τ — элемент алгебры A — называется алгбраичным индекса k, если для некоторых элементов $a_1, a_2, \ldots, a_k$ из алгебры A выполняется равенство $\sum_{i=1}^{k} \tau^i a_i = 0$. Алгебра A* алгебраична индекса k, *если каждый ее элемент алгебраичен индекса k над основным полем, и* алгебраична, *если каждый ее элемент алгебраичен некоторого индекса (зависящего от элемента).*

Задача 0.9 *1. Докажите, что в алгебре, алгебраичной индекса k, выполняется нетривиальное тождество.*

2. Докажите, что алгебра матриц n-го порядка алгебраична индекса n.

3. Докажите равенство (поляризацию)

$$\left(\sum_{i=1}^{n} x_i^n\right)^n - \sum_j (x_1 + \cdots + \widehat{x_j} + \cdots + x_n)^n + \sum_{j<k} (x_1 + \cdots + \widehat{x_j} + \cdots + \widehat{x_k} + \cdots + x_n)^n + \cdots +$$

$$+ (-1)^{n-1} \sum_i x_i^n = \sum_{\sigma \in S_n} x_{\sigma(1)} \cdots x_{\sigma(n)}$$

(Если переменные x_i коммутируют, то в результате получится $n! x_1 \cdots x_n$.)

4. Докажите, что тождество x^n влечет тождество $\sum_{\sigma \in S_n} x_{\sigma(1)} \cdots x_{\sigma(n)}$.

5. Докажите, что каждое тождество имеет полилинейное (т.е. линейное по каждой своей переменной) следствие той же степени.

Задача 0.10 *Пусть в алгебре A выполняется полилинейное тождество степени n. Докажите, что слово, являющееся n-разбиваемым, можно представить в виде линейной комбинации лексикографически меньших слов.*

А. Г. Курош в 1941 году поставил следующий вопрос.

Проблема А.Г.Куроша. *Верно ли, что алгебраическая конечно-порожденная алгебра, в которой выполняется некоторое тождество степени n, конечномерна?*

Первоначальное решение проблемы Куроша, полученное известными математиками Левицким и Капланским десятью годами спустя, было далеко не элементарным, пока А.И.Ширшов не разработал принципиально иной, чисто комбинаторный метод, позволивший решить и проблему Куроша, и вопросы нильпотентности.

Задача 0.11 *Воспользовавшись теоремой Ширшова о высоте, решите проблему Куроша. Докажите также, что l-порожденные ниль-алгебры индекса n нильпотентны индекса $k(n,l)$.*

Наша дальнейшая цель – получение оценок на функцию $k(n,l)$.

Оценки высоты в комбинаторике слов приводят к таким же оценкам в теории колец. Первоначальная оценка А. И. Ширшовабыла очень завышенной, однако его работы содержат глубокие идеи, интересные до сих пор. А.Т.Колотов в 1982 году получил двойную экспоненту (l^{l^n}), где l – число образующих, n – степень тождества. А. Я. Белов в 1990 году получил экспоненциальную оценку порядка $n^3 l^{3n}$, эта оценка улучшалась А.Клейном в 2000.

В 1991 году Е. И. Зельманов поставил следующий вопрос.

"Пусть $F_{2,m}$ – свободное 2-порождённое ассоциативное кольцо с тождеством $x^m = 0$. Верно ли, что класс нильпотентности кольца $F_{2,m}$ растёт экспоненциально по m?"

Задача 0.12 *Докажите, что заключительная задача цикла "Экспоненциальная оценка" проекта "Периодичность и порядочность" дает положительный ответ на вопрос Зельманова.*

В 2010 году А. Я. Белов и М. И. Харитонов получили субэкспоненциальную оценку.

В этой связи возникает следующая нерешенная задача:

Задача 0.13 *Получить полиномиальную оценку на высоту.*

Более того,

Задача 0.14 *Существует ли верхняя оценка на высоту, полиномиальная относительно степени и линейная относительно числа букв в алфавите?*

И, наконец, возникает

Задача 0.15 *Получить как можно более точную нижнюю оценку на высоту.*

Периодичность и порядочность

А. Я. Белов, М. И. Харитонов

August 6, 2012

1. Предисловие

Рассмотрим произвольное слово. Если нам очень повезёт, то оно будет "периодическим", то есть маленьким словом, повторённым много раз подряд. Это маленькое слово называется периодом. Поставим теперь в нашем слове знаки умножения между каждыми соседними буквами (операция умножения букв — некоммутативна, то есть $ab \neq ba$). Теперь о нашем периодическом слове можно говорить как о степени периода.

Однако произвольные слова редко бывают периодическими. Намного чаще слово является произведением нескольких периодических слов. Назовём такое слово кусочно-периодическим. Кроме того, любое слово можно представить как произведение периодических подслов и "прокладок" между ними.

Рассмотрим множество слов над алфавитом $A = \{a_1, \dots, a_s\}$. Порядок на множестве букв $a_1 \prec a_2 \prec \cdots \prec a_s$ индуцирует лексикографический порядок на множестве слов. Пусть $U \prec V$, если первая буква слова U меньше первой буквы V, при их совпадении смотрим на вторые буквы и т.д. Если же одно из слов есть начальная часть другого, то слова U и V считаются *несравнимыми*. Примерно так расположены слова в словаре (только более короткие слова из несравнимых считаются младшими). В словах наблюдаются *беспорядки* – когда старшая буква идет перед младшей. Бывают, однако, беспорядки разной силы. Назовем *k-беспорядком* набор из k неперекрывающихся подслов, идущих в порядке убывания, а слово, имеющее k-беспорядок, — *k-разбиваемым*. Слово называется *k-порядочным*, если оно k-разбиваемо, но не $(k+1)$-разбиваемо. Например, слово $aacbcbb$ — 3-порядочно, так как в нем есть 3 подслова, идущие в порядке убывания, а именно - c, bc, bb.

Оказывается, что степень разбиваемости слова и количество периодических подслов тесно связаны. Данный проект посвящён связи между упорядоченностью и кусочной разбиваемостью в словах.

Теорема Ширшова о высоте утверждает, что существует некоторая функция $H(k, s)$ такая, что все не k-разбиваемые слова можно разбить на $H(k, s)$ кусков, каждый из которых есть подслово периодической последовательности с длиной периода меньше k.

Главной задачей проекта является получение наилучшей оценки на $H(k, s)$.При получении такой оценки мы будем использовать теорему Дилуорса, которая сформулирована в главе 3.

При доказательстве теоремы Ширшова важна лемма, сформулированная как задача 2.10.

Целью проекта является получение конструктивных оценок на функцию $H(k, S)$. Другой нашей целью является перечисление *полилинейных* слов (т.е. каждая буква входит в такое слово только один раз), не имеющих k-беспорядков. Количество полилинейных не 3-разбиваемых слов есть число Каталана.

Мы изучаем также количество кусков фиксированного периода. При этом возникают комбинаторные вопросы теории графов типа теории Рамсея.

Ключевыми задачами до промежуточного финиша являются задачи 2.10, 3.7, 4.11, 4.12.

Отдельно от проекта мы приводим также серию "Приложения к теории колец, немного истории", посвященную приложениям к теории колец. Сама эта серия независима от остального материала, но она дает мотивировку, показывая, как из теоремы о высоте вытекает решение ряда долго стоявших проблем теории колец.

2. Цикл "Комбинаторика"

Задача 2.1. *Карлсон может писать только те слова, которые не содержат подслов из двух различных букв. Сколько слов длины n может написать Карлсон, если в словаре l букв?*

Задача 2.2. *В словаре племени Винни–Пухов 20 слов. В фразах их языка возможны любые сочетания этих слов. Существуют два магических заклинания, "Земля стоит на великом крокодиле" и "Каждый вечер крокодил глотает солнце", которые вызывают ураган, и поэтому вслух можно произносить только такие фразы, в которых эти последовательности слов не встречаются*[1]*. Сколько всего фраз из десяти слов можно произносить вслух?*

Задача 2.3. *В алфавите смешариков l букв. Может ли в их словаре содержаться слово длины l, у которого ровно*

a) $l + 1$

b) $\frac{l(l-1)}{2} - 1$

c) $2l$

различных подслов.

Задача 2.4. *В алфавите индейцев N букв. Из них индейцы составляют слова. Известно, что любое слово, повторенное дважды, означает то же самое, что и само слово, а замена подслова на его квадрат не меняет смысла всего слова. Например,* **гороод** *означает то же, что и* **город**. *Докажите, что в языке индейцев конечное число понятий, если:*

a) $N = 2$*;*

b)* $N = 3$.

*c**) Для произвольного N*

Задача 2.5. *Назовём* запретом *слово, которое мы запрещаем использовать в качестве подслова. Соответственно слово, не содержащее запретов, называем разрешённым. Какое минимальное число запретов нужно задать, чтобы среди стобуквенных слов ровно два — $(ab)^{50}$ и $(ba)^{50}$ — были разрешены?*

[1] Даже если слова в других словарных формах.

Запись u^t означает слово u, написанное t раз подряд.

Задача 2.6. *Пусть k, t – некоторые натуральные числа. Докажите, что если в слове V длины $k \cdot t$ не больше k различных подслов длины k, то для некоторого слова v слово V включает в себя подслово вида v^t.*

Задача 2.7. *Установите биекцию между следующими двумя множествами:*

- *последовательности натуральных чисел $1 \leqslant a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n$, где $a_i \leqslant i$;*
- *перестановки чисел $1, 2, \ldots, n$, у которых длина каждой убывающей последовательности не больше 2.*

Задача 2.8. *Сто людоедов приехали на пир. Обедающий людоед проглатывает целиком себе подобного. Пообедавший людоед, конечно, может и сам сослужить обедом для другого своего собрата. Так и составляются пищевые цепочки. Длиной цепочки назовем количество людоедов, вложенных друг в друга. Вопрос: какой максимальной длины цепочка точно присутствует, если известно, что какие бы десять людоедов мы не взяли, среди них найдутся два экземпляра, один из которых покоится в желудке другого?*

Задачи типа 2.8 Вы можете найти в цикле “Теорема Дилуорса”.

Определение 2.1. *Слово u назовем* нециклическим, *если u нельзя представить в виде v^k, где $k > 1$.*

Задача 2.9. *Пусть u и v – различные нециклические слова длины m и n соответственно. Слово W содержит подслова $u' = u^{m \cdot n}$ и $v' = v^{m \cdot n}$. Докажите, что длина общей части у u' и v' не больше $m + n - 2$.*

Задача 2.10. *На бесконечной ленте в каждой ячейке написаны цифры от 1 до 9. Докажите, что тогда либо из неё можно вырезать 10 непересекающихся тысячезначных чисел в порядке убывания, либо какое-то число длины меньше 10 повторится 50 раз подряд.*

3. Цикл “Теорема Дилуорса”

Задача 3.1. *Из любых ли пяти различных чисел, выписанных в ряд, можно выбрать три, стоящие в этом ряду в порядке убывания или в порядке возрастания?*

Задача 3.2. *Из любых ли девяти различных чисел, выписанных в ряд, можно выбрать четыре, стоящие в этом ряду в порядке убывания или в порядке возрастания?*

Задача 3.3. *Докажите, что из любых десяти различных чисел, выписанных в ряд, можно выбрать четыре, стоящие в этом ряду в порядке убывания или в порядке возрастания.*

Задача 3.4. *Докажите, что среди любых $mn + 1$ различных чисел найдутся либо $m + 1$ в порядке убывания, либо $n + 1$ в порядке возрастания.*

Частично упорядоченное множество (ЧУМ) M — это множество, для любых двух элементов a, b которого известно, находятся они в некотором отношении $\prec$ или нет. При этом должны быть выполнены следующие аксиомы:

X если $a \prec b$ и $b \prec c$, то $a \prec c$;

X если $a \prec b$, то a не равно b.

Задача 3.5. *Возможно ли, что неравенства $a \prec b$ и $b \prec a$ выполнены одновременно?*

Задача 3.6. *Докажите, что слова образуют ЧУМ при отношении, порождённом лексикографическим порядком.*

Определение 3.1. *Множество, любые два элемента которого сравнимы, называют линейно упорядоченным или, коротко, цепью.*

Задача 3.7. *Пусть m, n – некоторые натуральные числа. В частично упорядоченном множестве из $mn+1$ элементов есть либо цепь из идущих в порядке возрастания $m+1$ элементов, либо $n+1$ попарно несравнимых элементов (так называемая антицепь).*

Задача 3.8. *Обозначим через d наибольшее количество элементов цепи данного конечного частично упорядоченного множества M. Тогда M можно разбить на d антицепей.*

Более того, верен следующий факт, который окажет нам серьёзную помощь в дальнейшем:

Теорема Дилуорса. Обозначим через n наибольшее количество элементов антицепи данного конечного частично упорядоченного множества M. Тогда M можно разбить на n цепей.

4. Цикл "Экспоненциальная оценка"

Пусть наш алфавит состоит из букв $a_1, a_2, \dots, a_s$. Будем считать, что $a_1 \prec a_2 \prec \dots \prec a_s$. Таким образом, мы упорядочили буквы алфавита. Рассмотрим теперь два слова u и v. Если одно из них является началом другого, то назовём слова u и v *несравнимыми* (по отношению друг к другу). В противном случае найдутся слова w, u', v' такие, что $u = wu'$, $v = wv'$, причём первые буквы у слов u' и v' – различные (w может быть пустым, u' и v' – нет). Если первая буква u' больше первой буквы v', то считаем слово u больше слова v, в обратном случае считаем u меньше v. Таким образом, мы частично упорядочили слова. Приведённый порядок называется *лексикографическим* (мы уже обсуждали его в предисловии). Не стоит также забывать, что некоторые слова так и остались несравнимыми.

Задача 4.1. *Пусть алфавит состоит из трёх букв: a, b и c. Введём на них порядок $a < b < c$. Составьте из приведённого ниже списка слов наиболее длинную возрастающую последовательность. Какие пары слов являются несравнимыми?*

$$cb, abc, bac, abb, b, ccc, abc$$

Для дальнейшей работы нам потребуется ввести несколько вспомогательных определений.

Определение 4.1. *Слово W – n-*разбиваемо, *если найдутся слова $u_1, u_2, \dots, u_n$ такие, что $W = v \cdot u_1 \cdots u_n$, при этом $u_1 \succ \dots \succ u_n$.*

Определение 4.2. *Слово называется k-*порядочным, *если оно k-разбиваемо, но не $(k+1)$-разбиваемо.*

Задача 4.2. *Найдите число*

a) 1-порядочных

b) 2-порядочных

слов длины s, все буквы которых различны;

c) 1-порядочных слов длины s, буквы которых необязательно различны.

Задача 4.3. *a) Пусть n – некоторое натуральное число, u – нециклическое слово длины не меньше n. Докажите, что слово u^{2n} является n-разбиваемым.*

b) Пусть u — некоторое слово длины $(n-1)$. Докажите, что слово u^{2n} — не n-разбиваемое.

Определение 4.3. *a) Слово v —* хвост *слова u, если найдется слово w такое, что $u = wv$.*

б) Если в слове v содержится подслово вида u^t, то будем говорить, что в слове v содержится период цикличности *t.*

Задача 4.4. *Пусть x, d — некоторые натуральные числа. Докажите, что в слове W длины x либо первые $[x/d]$ хвостов попарно сравнимы, либо в слове W найдется период длины d.*

Здесь и далее: если в формулировке задачи встречаются числа n и d, то считаем, что $n \leqslant d$.

Определение 4.4. *Слово W — (n,d)-*сократимое*, если оно либо n-разбиваемо, либо найдется u^d — подслово слова W.*

Задача 4.5. *Докажите, что если в слове W найдутся n одинаковых непересекающихся подслов длины n, то W — (n,n)-сократимое.*

Определение 4.5. *Слово W будем называть n-разбиваемым в хвостовом смысле, если найдутся хвосты $u_1,\dots,u_n$ такие, что $u_1 \succ u_2 \succ \dots \succ u_n$ и для любого $i = 1,2,\dots,n-1$ начало u_i слева от начала u_{i+1}.*

Задача 4.6. *Докажите, что если слово W является*
a) n^3d-разбиваемым в хвостовом смысле,
b) $3n^2d$-разбиваемым в хвостовом смысле,
c) $4nd$-разбиваемым в хвостовом смысле,
то оно — либо n-разбиваемо, либо W содержит подслово в степени d.

Задача 4.7. *Для каждой пары натуральных чисел (n,d) (кроме пары $(1,1)$) приведите пример слова W длины $(nd-1)$ такого, что W — не $(n+1,d)$-сократимо и не содержит последовательности возрастающих хвостов длины $(n+1)$.*

Задача 4.8. *Попробуйте улучшить оценку в задаче 4.6.*

Фиксируем алфавит из s букв, слово W длины $r(W)$ над этим алфавитом и натуральные числа $n \leqslant d$. Далее будем считать, что W не содержит подслова в степени d и слово W не $4nd$-разбиваемо в хвостовом смысле. Рассмотрим его первые $[r(W)/d]$ хвостов (далее будем обозначать это множество хвостов за Ω). Тогда по теореме Дилуорса их можно раскрасить в $(4nd-1)$ цветов так, чтобы хвосты одного цвета шли в порядке возрастания. Для решения следующих задач следует пользоваться предыдущими задачами из цикла.

Задача 4.9. *Докажите, что среди любых $4nd^2$ хвостов из Ω найдутся два одноцветных хвоста, у которых отличаются начальные подслова длины $4nd$.*

Задача 4.10. *В бесконечном парламенте у каждого не более 3 врагов. Доказать, что его можно разбить на 2 палаты так, что у каждого будет не более одного врага в своей палате.*

Теперь перестанем фиксировать числа s,n,d и слово W. Будем считать, что s,n,d — некоторые натуральные числа такие, что $n \leqslant d$, а слово W — некоторое слово над алфавитом из s букв.

Задача 4.11 (Лемма Ширшова). *Докажите, что существует функция от натуральных аргументов $f(s,n,d)$ такая, что для любого слова W над алфавитом длины s, не являющегося (n,d)-сократимым, $r(W) < f(s,n,d)$.*

Задача 4.12. *Докажите, что $f(s,n,d) < s(4nd)^{4nd+2}$.*

5. Улучшение экспоненциальной оценки

Для некоторых натуральных чисел s, n, d таких, что $d \geqslant n$, предположим, что слово W над алфавитом длины s — не (n,d)-сократимое (то есть либо n-разбиваемое, либо найдется u^d — подслово слова W). Пусть Ω — множество хвостов слова W, которые начинаются с его первых $[r(W)/d]$ букв. Напомним что хвосты раскрашены в $4nd-1$ цветов (эти понятия были введены в цикле "Экспоненциальная оценка"). Введём обозначение $p_{n,d} = 4nd$. Будем называть Pk-*началом* некоторого слова слово, состоящее из его первых k букв.

Задача 5.1. *Пусть в множестве Ω нашлось $p_{n,d}^2 + p_{n,d} + 1$ хвостов одного цвета с попарно различными $2k$-началами, где k — некоторое натуральное число. Докажите, что среди этих хвостов найдутся два с различными k-началами.*

Задача 5.2. *Докажите, что среди любых $(d+1)p_{n,d}^4$ хвостов из Ω найдутся два одноцветных хвоста с разными $p_{n,d}/2$-началами.*

Задача 5.3. *Докажите, что среди любых $(d+1)p_{n,d}^7$ хвостов из Ω найдутся два одноцветных хвоста с разными $p_{n,d}/4$-началами.*

Будем считать, что $p_{n,d} = 2^t$ для некоторого натурального числа t.

Задача 5.4. *Докажите, что среди любых $(d+1)p_{n,d}^{1+3t}$ хвостов из множества Ω найдутся два одноцветных хвоста с различными 1-началами.*

Задача 5.5. *Докажите, что если длина слова W больше, чем $(d+1)^2 p_{n,d}^{2+3t} s$, то оно либо n-разбиваемое, либо либо найдется u^d — подслово слова W.*

Пользуясь понятиями листка "Приложения к кольцам, немного истории" и задачей 5.5, можно получить отрицательный ответ на вопрос Е. И. Зельманова, поставленный в 1991 году:

"Пусть $F_{2,m}$ — свободное 2-порождённое ассоциативное кольцо с тождеством $x^m = 0$. Верно ли, что класс нильпотентности кольца $F_{2,m}$ растёт экспоненциально по m?"

Будем считать, что $n = 2^q$ для некоторого натурального числа q.

Задача 5.6. *Пусть Γ — (бесконечное) множество, состоящее из слов длины $< n$ над алфавитом длины s и их всевозможных степеней. Пользуясь задачами 4.3а, 5.5, докажите, что если слово W — не n-разбиваемое, то его можно представить в виде произведения менее, чем $n^{100q}s$ слов из множества Gamma.*

Периодичность и порядочность *

А. Я. Белов, М. И. Харитонов, С. Г. Григорьев, А. В. Петухов

11 августа 2012 г.

Рассмотрим множество, с элементами которого можно проводить операции сложения, причем

1. для любых элементов a, b, c, $a+b=b+a$, $(a+b)+c=a+(b+c)$;
2. есть специальный элемент 0 такой, что $a+0=a$ для всех a;
3. для каждого элемента a существует противоположный $(-a)$ такой, что $(-a)+a=a+(-a)=0$.

   Можно также добавить еще одну естественную операцию, умножение, также удовлетворяющую условию *ассоциативности*:

4. $a(bc)=(ab)c$ для всех a, b, c.

   Если при всех этих свойствах умножение и сложение удовлетворяют условию *дистрибутивности:*

5. $a(b+c)=ab+ac, (b+c)a=ba+ca$ для всех a, b, c, то множество с указанными операциями называется кольцом.

Задача 0.1. *Приведите примеры колец с делителями нуля, то есть колец, где есть элементы a, b такие, что $ab=0$;*

Решение. Рассмотрим множество пар (a,b) целых чисел. С поэлементной операцией сложения и поэлементной операцией умножения. Очевидно, это кольцо. Единица в этом кольце соответствует паре $(1,1)$, а ноль — $(0,0)$. Тогда $(1,0)\cdot(0,1)=(0,0)$, т.е. $(1,0)$ и $(0,1)$ — делители нуля. □

Определение 0.1. Единицей *кольца или* нейтральным элементом *называется элемент E такой, что $EA=AE=A$ для всех $A\in R$.* Обратный элемент A^{-1} *определяется равенством $AA^{-1}=E$.*

Задача 0.2. *Докажите, что в кольце с единицей аксиома коммутативности сложения вытекает из других аксиом.*

Решение. Заметим, что

$$ab+a+b+1=(a+1)(b+1)=ab+b+a+1.$$

*Если Вы заметите ошибки в условиях или решениях задач, пишите по адресу krab8nog@yandex.ru.

Следовательно, $a+b=b+a$. □

Задача 0.3. *Постройте кольцо из 4 элементов, каждый ненулевой элемент которого обратим.*

Пример. Обозначим элементы кольца знаками $0, 1, a, b$. Возможные таблицы сложения и умножения кольца приведены ниже

$+$	0	1	a	b
0	0	1	a	b
1	1	0	b	a
a	a	b	0	1
b	b	a	1	0

$\times$	0	1	a	b
0	0	0	0	0
1	0	1	a	b
a	0	a	b	1
b	0	b	1	a

.

Проверка ассоциативности и коммутативности этого сложения и умножения выполняется перебором. Также как и проверка дистрибутивности. □

Задача 0.4. *Постройте некоммутативное кольцо, то есть такое, что существуют a, b такие, что $ab \neq ba$.*

Пример. Кольцо матриц 2×2, вводимое после определения 0.3, является примером некоммутативного кольца. □

Задача 0.5. *Пусть R кольцо, где любая сумма нескольких единиц не равна нулю. Пусть элементы e, f, g таковы, что $ee = e, ff = f, gg = g$ и $e+f+g=0$. Докажите, что $e=f=g=0$.*

Доказательство. Построим кольцо, удовлетворяющее условиям задачи, для которого $e, f, g \neq 0$ в два шага.

Первый шаг. Положим, что e, f, g лежат в кольце 2×2 матриц с коэффициентами в $\mathbb{Z}_2$ и

$$e = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, f = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, f = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Легко видеть, что $e^2 = e, f^2 = f, g^2 = g, e+f+g=0$. К сожалению, удвоенная единица этого кольца равна 0.

Рассмотрим алгебру, состоящую из пар (z, M), где z — целое число, а M — 2×2 матрица с коэффициентами в $\mathbb{Z}_2$. Положим

$$e = (0, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}), f = (0, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}), f = (0, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}).$$

Легко видеть, что $e^2 = e, f^2 = f, g^2 = g, e+f+g=0$. Единица этого кольца $(1, \begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix})$, сложенная с собой любое число раз даёт не ноль. □

Определение 0.2. *Многочлен $f = f(x_1, \ldots, x_n)$ называется* тождеством *кольца A, если он тождественно обращается в ноль на A. Тождество f следует из набора тождеств $\{g_i\}$, если везде, где выполняется набор $\{g_i\}$, также выполняется f. Иногда пишут тождество в виде $f=0$.*

Задача 0.6. *Существуют ли нетривиальные кольца, удовлетворяющие тождествам $x^2=0$; $xy=yx$?*

Пример. Рассмотрим выражения вида (a, b, c), где $a, b \in \mathbb{Z}$, а c — остаток при делении на два (вычет). Определим сложение в этом кольце поточечно, а произведение формулой

$$(a_1, b_1, c_1) \cdot (a_2, b_2, c_2) := (0, 0, a_1 b_2 + a_2 b_1 (\mathrm{mod} 2)).$$

Легко видеть, что указанное кольцо удовлетворяет соотношениям

$$xy = yx \text{ и } x^2 = 0.$$

□

Определение 0.3. *Если каждый ненулевой элемент обратим, то кольцо называется* телом. *Кольцо* коммутативно, *если $ab = ba$ для всех $a, b \in R$. Коммутативное тело есть* поле.

Определение 0.4. *Свободная ассоциативная алгебра или кольцо некоммутативных многочленов над кольцом R: это наборы выражений вида $\sum_i a_i v_i$, $a_i \in R$, v_i – слова. Если $v = \sum_i a_i v_i$, $u = \sum_i b_i v_i$, то $u + v = \sum_i (a_i + b_i) v_i$, $uv = \sum_{i,j} a_i b_j v_i v_j$. Свободная ассоциативная алгебра перестаёт быть свободной, если в ней выполняются некоторые тождества. Понятие тождества алгебры приведено ниже.*

Определение 0.5. *Многочлен $f = f(x_1, \ldots, x_n)$ называется* тождеством *алгебры A, если он тождественно обращается в ноль на A. Тождество f следует из набора тождеств $\{g_i\}$, если везде, где выполняется набор $\{g_i\}$, также выполняется f. Иногда пишут тождество в виде $f = 0$.*

Задача 0.7.

1. *Докажите, что в алгебре матриц второго порядка выполняется тождество $[[x, y]^2, z] = 0$ (*тождество Холла*).*
2. *Докажите, что в алгебре матриц второго порядка выполняется тождество $\sum_{\sigma \in S_4} (-1)^\sigma x_{\sigma(1)} \cdots x_{\sigma(4)} = 0$ (*стандартное тождество *степени 4).*

Доказательство. Доказательство пункта а) распадается в три утверждения, каждое из которых легко доказывается непосредственно.

1. След коммутатора двух матриц равен нулю (след матрицы — это сумма её диагональных элементов);
2. Квадрат матрицы два на два со следом ноль имеет вид $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$;
3. Коммутатор любой матрицы с матрицей указанного выше вида равен 0.

Для матрицы A обозначим её след через $\mathrm{tr}A$. Прямая проверка показывает, что

$$A^2 - A\mathrm{tr}A + \frac{(\mathrm{tr}A)^2 - \mathrm{tr}A^2}{2} = 0$$

для всякой матрицы A (мы рекомендуем начать с проверки этого утверждения для матриц вида $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$). Подставляя вместо A матрицу $[X_1, X_2]$, имеем

$$0 = A^2 - A\mathrm{tr}A + \frac{(\mathrm{tr}A)^2 - \mathrm{tr}A^2}{2} = A^2 - \frac{\mathrm{tr}(A^2)}{2}.$$

То же тождество будет выполнено, если мы подставим вместо A матрицы $[X_3, X_4]$ и $[X_1, X_2] + [X_3, X_4]$. Откуда следует, что

$$[X_1,X_2][X_3,X_4]+[X_3,X_4][X_1,X_2] = \mathrm{tr}\frac{[X_1,X_2][X_3,X_4]+[X_3,X_4][X_1,X_2]}{2} \qquad (1).$$

Назовём *альтернированием* некоммутативного многочлена $f(x_1,...,x_n)$ от n переменных некоммутативный многочлен

$$\mathrm{Alt}(f)(x_1,...,x_n) := \sum_{\sigma\in \mathrm{S}_n} \mathrm{sgn}(\sigma) f(x_{\sigma(1)},...,x_{\sigma(n)}).$$

Тогда $2n$-стандартное тождество запишется как $\mathrm{Alt}(x_1....x_{2n}) = 0$. Применяя операцию Alt к выражению (1) имеем,

$$\mathrm{Alt}(X_1X_2X_3X_4) = \tfrac{1}{2}\mathrm{tr}\ \mathrm{Alt}(X_1X_2X_3X_4).$$

Стандартное тождество степени 4 переписывается, как $\mathrm{Alt}(X_1X_2X_3X_4) = 0$. Таким образом, нам достаточно доказать, что $\mathrm{tr}\ \mathrm{Alt}(X_1X_2X_3X_4) = 0$.

Прямая проверка показывает, что для всяких матриц A и B имеется тождество

$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

Откуда легко видеть, что

$$\begin{array}{l}
\mathrm{tr}(X_1X_2X_3X_4)=\mathrm{tr}(X_4X_1X_2X_3)=\mathrm{tr}(X_3X_4X_1X_2)=\mathrm{tr}(X_2X_3X_4X_1),\\
\mathrm{tr}(X_1X_2X_4X_3)=\mathrm{tr}(X_3X_1X_2X_4)=\mathrm{tr}(X_4X_3X_1X_2)=\mathrm{tr}(X_2X_4X_3X_1),\\
\mathrm{tr}(X_1X_3X_2X_4)=\mathrm{tr}(X_4X_1X_3X_2)=\mathrm{tr}(X_2X_4X_1X_3)=\mathrm{tr}(X_3X_2X_4X_1),\\
\mathrm{tr}(X_1X_3X_4X_2)=\mathrm{tr}(X_2X_1X_3X_4)=\mathrm{tr}(X_4X_2X_1X_3)=\mathrm{tr}(X_3X_4X_2X_1),\\
\mathrm{tr}(X_1X_4X_2X_3)=\mathrm{tr}(X_3X_1X_4X_2)=\mathrm{tr}(X_2X_3X_1X_4)=\mathrm{tr}(X_4X_2X_3X_1),\\
\mathrm{tr}(X_1X_4X_3X_2)=\mathrm{tr}(X_3X_2X_1X_4)=\mathrm{tr}(X_3X_2X_1X_4)=\mathrm{tr}(X_4X_3X_2X_1).
\end{array}$$

И, следовательно, что $\mathrm{Alt}\ \mathrm{tr}(X_1X_2X_3X_4) = 0$. □

Определение 0.6. *Алгебра A называется* ниль-алгеброй, *если есть функция $n : A \to \mathbf{N}$ такая, что для любого $x \in A$ выполняется равенство $x^{n(x)} = 0$. Если же в ней выполняется тождество $x^n = 0$, то A называется* ниль-алгеброй индекса n. *A —* нильпотентна индекса k, *если в ней выполняется тождество $x_1 \cdots x_k = 0$, A —* нильпотентна, *если она нильпотентна индекса k при некотором k. τ — элемент алгебры A — называется алгбраичным индекса k, если для некоторых элементов $a_1, a_2, \ldots, a_k$ из алгебры A выполняется равенство $\sum_{i=1}^{k} \tau^i a_i = 0$. Алгебра A* алгебраична индекса k, *если каждый ее элемент алгебраичен индекса k над основным полем, и* алгебраична, *если каждый ее элемент алгебраичен некоторого индекса (зависящего от элемента).*

Задача 0.8. *1. Докажите, что в алгебре, алгебраичной индекса k, выполняется нетривиальное тождество.*

2. Докажите, что алгебра матриц n-го порядка алгебраична индекса n.

3. Докажите равенство (поляризацию)

$$\left(\sum_{i=1}^{n} x_i^n\right)^n - \sum_j (x_1 + \cdots + \widehat{x_j} + \cdots + x_n)^n + \sum_{j<k} (x_1 + \cdots + \widehat{x_j} + \cdots + \widehat{x_k} + \cdots + x_n)^n + \cdots +$$

$$+(-1)^{n-1} \sum_i x_i^n = \sum_{\sigma\in S_n} x_{\sigma(1)} \cdots x_{\sigma(n)}$$

(Если переменные x_i коммутируют, то в результате получится $n!x_1 \cdots x_n$.)

4. *Докажите, что тождество x^n влечет тождество $\sum_{\sigma \in S_n} x_{\sigma(1)} \cdots x_{\sigma(n)}$.*

5. *Докажите, что каждое тождество имеет полилинейное (т.е. линейное по каждой своей переменной) следствие той же степени.*

Задача 0.9. *Пусть в алгебре A выполняется полилинейное тождество степени n. Докажите, что слово, являющееся n-разбиваемым, можно представить в виде линейной комбинации лексикографически меньших слов.*

А. Г. Курош в 1941 году поставил следующий вопрос.

Проблема А.Г.Куроша. *Верно ли, что алгебраическая конечно-порожденная алгебра, в которой выполняется некоторое тождество степени n, конечномерна?*

Первоначальное решение проблемы Куроша, полученное известными математиками Левицким и Капланским десятью годами спустя, было далеко не элементарным, пока А.И.Ширшов не разработал принципиально иной, чисто комбинаторный метод, позволивший решить и проблему Куроша, и вопросы нильпотентности.

Задача 0.10. *Воспользовавшись теоремой Ширшова о высоте, решите проблему Куроша. Докажите также, что l-порожденные ниль-алгебры индекса n нильпотентны индекса $k(n, l)$.*

Наша дальнейшая цель – получение оценок на функцию $k(n, l)$.

Оценки высоты в комбинаторике слов приводят к таким же оценкам в теории колец. Первоначальная оценка А. И. Ширшовабыла очень завышенной, однако его работы содержат глубокие идеи, интересные до сих пор. А.Т.Колотов в 1982 году получил двойную экспоненту (l^{l^n}), где l – число образующих, n – степень тождества. А. Я. Белов в 1990 году получил экспоненциальную оценку порядка $n^3 l^{3n}$, эта оценка улучшалась А.Клейном в 2000.

В 1991 году Е. И. Зельманов поставил следующий вопрос.

"Пусть $F_{2,m}$ – свободное 2-порождённое ассоциативное кольцо с тождеством $x^m = 0$. Верно ли, что класс нильпотентности кольца $F_{2,m}$ растёт экспоненциально по m?"

Задача 0.11. *Докажите, что заключительная задача цикла "Экспоненциальная оценка" проекта "Периодичность и порядочность" дает положительный ответ на вопрос Зельманова.*

В 2010 году А. Я. Белов и М. И. Харитонов получили субэкспоненциальную оценку.

В этой связи возникает следующая нерешенная задача:

Задача 0.12. *Получить полиномиальную оценку на высоту.*

Более того,

Задача 0.13. *Существует ли верхняя оценка на высоту, полиномиальная относительно степени и линейная относительно числа букв в алфавите?*

И, наконец, возникает

Задача 0.14. *Получить как можно более точную нижнюю оценку на высоту.*

2 Цикл “Комбинаторика”

В некоторых доказательствах присутствует термин “$k-$хвост”, который означает то же самое, что термин “$k-$начало”.

Задача 2.1. *Карлсон может писать только те слова, которые не содержат подслов из двух различных букв. Сколько слов длины n может написать Карлсон, если в словаре l букв?*

Ответ. Карлсон может писать слова, состоящие из всех одинаковых букв, а их ровно l. □

Задача 2.2. *В словаре племени Винни–Пухов 20 слов. В фразах их языка возможны любые сочетания этих слов. Существуют два магических заклинания, “Земля стоит на великом крокодиле” и “Каждый вечер крокодил глотает солнце”, которые вызывают ураган, и поэтому вслух можно произносить только такие фразы, в которых эти последовательности слов не встречаются*[1]*. Сколько всего фраз из десяти слов можно произносить вслух?*

Ответ. $20^{10} - 12 \cdot 20^5 + 4$. □

Задача 2.3. *В алфавите смешариков l букв. Может ли в их словаре содержаться слово длины l, у которого ровно*

a) $l + 1$

b) $\frac{l(l-1)}{2} - 1$

c)* $2l$

различных подслов.

Решение. a) Для $l = 1$ такого слова, очевидно, не существует. Для $l = 2$ такое слово существует: ab. Далее мы считаем, что $l \geqslant 3$. Рассмотрим слова, состоящие из одинаковых букв. В них ровно l различных подслов. Если в слове будет хотя бы 2 различные буквы, то там будет как минимум 2 различных подслова длины 1 и хотя бы 2 длины 2 ($l \geqslant 3$). Тогда всего подслов будет не меньше, чем $l+2$, что уже больше, чем нам нужно.

b) Для $l \leq 4$, все очевидно. Для $l = 5$ пример: ababa. Для l = 6: aaaabb. Для 7: aabbabb.

Примеры для всех $l \geq 8$ строятся по индукции: к слову длины $l-3$ приписывается в конец три буквы, отличающиеся от всех предыдущих.

c) Случай $l \leq 4$ (как более простой) мы оставляем читателю. Рассмотрим случай $l \geq 5$. Для того чтобы слово W имело ровно $2l$ различных подслов, требуется чтобы в нём было хотя бы 2 различные буквы. При этом условии, для любого $1 \leq k < l$ различных подслов длины k хотя бы два (иначе все буквы подслова будут одинаковы). Рассмотрим различные подслова длины два . Если их ровно два, то слово имеет вид:

$$ababab...ab, \text{ или } ababab...aba, \text{ или } abbb...b, \text{ или } aaa...ab.$$

В этих словах меньше, чем $2l$ различных подслов и, следовательно, они не удовлетворяют условию задачи. По аналогичным соображениям подслов длины 3 хотя бы

[1] Даже если слова в других словарных формах.

3. В итоге, различных подслов W по крайней мере $2+3+3+...+2+1$ (всего l слагаемых). Указанная сумма равна $2l+1$ и, следовательно, не существует слова W длины $l \geq 5$, содержащего ровно $2l$ подслов. □

Задача 2.4. *В алфавите индейцев N букв. Из них индейцы составляют слова. Известно, что любое слово, повторенное дважды, означает то же самое, что и само слово, а замена подслова на его квадрат не меняет смысла всего слова. Например,* **гороод** *означает то же, что и* **город**. *Докажите, что в языке индейцев конечное число понятий, если:*

a) $N = 2$;

b) $N = 3$;

c) произвольное N.

Доказательство. Сначала введем одно обозначение. Назовем слово *несократимым*, если нет слова короче с таким же смыслом. Доказывать утверждение задачи будем по индукции по N — числу букв в алфавите. База для N равного единице очевидна. Докажем переход от N к $N+1$. По предположению индукции несократимых слов в алфавите с N буквами конечно. Обозначим за d произвольное число, большее длины каждого из несократимых слов этого алфавита. Теперь рассмотрим алфавит из $N+1$ буквы. Пусть в нём найдется несократимое слово длины хотя бы $(d+1)(N+1)^{d+2}$. Это слово можно разбить на $(N+1)^{d+2}$ блоков длины $d+1$ плюс, возможно, ещё что-то. Среди этих блоков найдутся 2 одинаковых, т.к. всего возможных блоков длины $(d+1)$ ровно $(N+1)^{d+1}$, что меньше $(N+1)^{d+2}$. Обозначим два одинаковых блока длины $(d+1)$ за B (они не пересекаются). Если бы в B было не более, чем N различных букв, то по предположению индукции это слово было бы сократимо. Следовательно, в записи B участвуют все буквы нашего алфавита.

Покажем, что если слово B содержит все буквы алфавита, а C — любое слово, то значение BCB совпадает со значением B (если это так, то слово полученное в предыдущем параграфе сократимо и задача 2.4 решена).

Покажем, что существует такое Y, что слова B и BCY имеют одинаковый смысл (если это так, то слова

$$B \leftrightarrow BCY \leftrightarrow BCBCY \leftrightarrow BCB$$

имеют один и тот же смысл и задача 2.4 решена). Пусть $C = c_1c_2...c_k$. Так как c_1 входит в B,

$$B = Sc_1M.$$

Тогда B совпадает по смыслу с Bc_1M. Так как c_2 входит в B, $B = Ec_2D$ и смысл слов B_1M и $Bc_1c_2Dc_1M$ одинаков. И так далее. Таким образом, и Y имеют одинаковый смысл для некоторого Y.

Значит, в алфавите длины $N+1$ не бывает несократимых слов длины большей, чем $(d+1)(N+1)^{d+2}$, а, следовательно, число слов в этом алфавите конечно. □

Задача 2.5. *Назовём* запретом *слово, которое мы запрещаем использовать в качестве подслова. Соответственно слово, не содержащее запретов, называем разрешённым. Какое минимальное число запретов нужно задать, чтобы среди стобуквенных слов ровно два: $(ab)^{50}$ и $(ba)^{50}$ — были разрешены?*

Решение. Рассмотрим все слова длины 100, состоящие из одинаковых букв. Их ровно l и у них нет общих подслов. Значит, нам требуется как минимум l запретов.

Покажем, что l запретов хватает. Действительно, слова без запретов aa, bb, c, d и всех оставшихся букв — это ровно *abab....ab* и *baba...ba*. □

Запись u^t означает слово u, написанное t раз подряд.

Задача 2.6. *Пусть k, t – некоторые натуральные числа. Докажите, что если в слове V длины $k \cdot t$ не больше k различных подслов длины k, то для некоторого слова v слово V включает в себя подслово вида v^t.*

Доказательство. Докажем лемму индукцией по k. База при $k = 1$ очевидна. Обозначим через V^- слово V, с выброшенной последней буквой. Если в V^- находится не больше, чем $(k-1)$ различное подслово длины $(k-1)$, то применяем индукционное предположение (длина V^- не меньше, чем $(k-1)t$).

Пусть V^- содержит не меньше, чем k подслов длины $k-1$. Так как V^- содержит не больше, чем k различных подслов длины k, то в подслове V длины k последняя буква определяется по $(k-1)$ предыдущей. Таким образом, среди первых $k+1$ подслов длины $(k-1)$ есть не меньше двух одинаковых. Пусть они имеют номера i, j и $i > j$ (отметим, что i-ое подслово и j-ое подслово длины k также совпадают). Тогда i-ая буква совпадает с j-ой, $(i+1)$-ая с $(j+1)$-ой, $(i+k)$-ая с $(j+k)$-ой. Так как $(i-j) \leq k$, V есть подслово $V_{1\to(j-1)}V^{\infty}_{j\to(i-1)}$, где $V_{1\to(j-1)}$ — подслово в V, начинающееся в 1-ой букве, а кончающееся в $j-1$, а $V_{j\to(i-1)}$ — подслово V, начинающееся в j-ой букве, а кончающееся в $(i-1)$-ой. Так как $i - 1 \leqslant k$ и $i - j \leqslant k$, V содержит не меньше, чем t-ую степень подслова $V_{j\to(i-1)}$. Значит, V содержит подслово вида v^t. □

Задача 2.7. *Установите биекцию между следующими двумя множествами:*

- *последовательности натуральных чисел $1 \leqslant a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n$, где $a_i \leqslant i$;*
- *перестановки чисел $1, 2, \ldots, n$, у которых длина каждой убывающей последовательности не больше 2.*

Доказательство. Рассмотрим множество перестановок $S_n^{\vee}$ чисел $1, 2, \ldots, n$, у которых длина каждой убывающей последовательности не больше 2 (мы рассматриваем перестановки, как подмножество множества слов). Для всякой перестановки $\sigma \in S_n^{\vee}$ обозначим через $b(\sigma)$ длину максимальной возрастающей подпоследовательности, конец которой совпадает с концом σ. Для всякого $m < n$ и для каждой перестановки $\sigma \in S_n^{\vee}$ обозначим через $\sigma[m]$ перестановку, получаемую из $\sigma[m]$ выбрасыванием всех чисел, больших m. Каждая перестановка $\sigma \in S_n^{\vee}$ задаёт последовательность чисел

$$b_1 := b(\sigma[1]), b_2 := b(\sigma[2]), ..., b_n := b(\sigma[n]).$$

Заметим, что $b_1 = 1, b_{i+1} \leqslant b_i + 1, 1 \leqslant b_i \leqslant i$. Зададим теперь последовательность $\{a_i\}$ правилом $a_i := i + 1 - b_i$. Заметим, что

$$a_1 = 1,\ 1 \leqslant a_i \leqslant i,\ a_i \leq a_j.$$

Таким образом, каждой перестановке $\sigma \in S_n^{\vee}$, в которой нет убывающей последовательности длины 3, мы сопоставили возрастающую последовательность $1 \leqslant a_1 \leqslant a_2 \leqslant ... \leqslant a_n$, для которой $a_i \leq i$.

Легко видеть, что заданное соответствие биективно. Построение обратной функции оставляется читателю в качестве упражнения. □

Решение этой задачи также находится в проекте Доценко “Числа Каталана и естественные отображения” на одной из предыдущих конференций турнира городов.

Задача 2.8. *Сто людоедов приехали на пир. Обедающий людоед проглатывает целиком себе подобного. Пообедавший людоед, конечно, может и сам сослужить обедом для другого своего собрата. Так и составляются пищевые цепочки. Длиной цепочки назовем количество людоедов, вложенных друг в друга. Вопрос: какой максимальной длины цепочка точно присутствует, если известно, что какие бы десять людоедов мы не взяли, среди них найдутся два экземпляра, один из которых покоится в желудке другого?*

Доказательство. Рассмотрим всех этих людоедов в виде графа: сами людоеды - это вершины графа, а ориентированные ребра идут от пообедавшего людоеда ко всем его жертвам. Заметим, что этот граф является лесом (т.е. объединением деревьев). Подвесим граф за людоедов, которых не съели. По очевидным соображениям на произвольном уровне число вершин не более девяти, а из этого ясно, что есть цепочка длины хотя бы 12. А теперь покажем, что может быть ровно 12: для начала посадим 99 людоедов в 9 комнат по 11 людоедов в каждой, и пусть они пообедают таким образом: второй съедает первого, потом третий второго и так далее. А потом оставшийся людоед съест всех выживших в комнатах. Ура. □

Определение 2.1. *Слово u назовем* нециклическим, *если u нельзя представить в виде v^k, где $k > 1$.*

Задача 2.9. *Пусть u и v – различные нециклические слова длины m и n соответственно. Слово W содержит подслова $u' = u^{m \cdot n}$ и $v' = v^{m \cdot n}$. Докажите, что длина общей части у u' и v' не больше $m + n - 2$.*

Решение. Пусть $m > n$ и пусть пересечение двух периодических подслов u^{mn} и v^{mn} имеет длину хотя бы $m + n - 1$. Обозначим их пересечение за S, а его буквы — $s_1, ..., s_l$ (l — длина S). Покажем, что в этом случае слово u — периодично с периодом $d :=$НОД(m, n).

Достаточно показать, что если $k = l (\text{mod d})$ и $1 \leq k, l < m + n - 1$, то $s_k = s_l$.

Обозначим как r остаток при делении k на d. Пусть $r \neq 0$. Тогда $k - r = an - bm$ для каких-то чисел $a, b \in \mathbb{Z}_{\geq 0}$. Построим последовательности k_i, a_i, b_i по следующим правилам:

1. $k_0 = k, a_0 = a, b_0 = b$;
2. $\begin{cases} k_{i+1} = k_i + n, a_{i+1} = a_i - 1, b_{i+1} = b_i, & \text{если } a_i > 0 \text{ и } k_i < m \\ k_{i+1} = k_i - m, a_{i+1} = a_i, b_{i+1} = b_i - 1, & \text{если } a_i = 0 \text{ и } b_i > 0 \\ i - \text{ый член последовательности — последний}, & \text{если } a_i = b_i = 0 \end{cases}$.

Из определения этой последовательности видно, что

1. $k_i - r = a_i n - b_i m$ для всех $i \geq 0$;
2. $d \not| k_i$ для всех $i \geq 0$;
3. $1 \leq k_i \leq m + n - 1$ для всех $i \geq 0$.
4. Если k_i — последний член последовательности, то $k_i = r$.

Из этих правил следует, что $s_{k_{i+1}} = s_{k_i}$ и, в частности, $s_k = s_l$, если $k = l (\text{mod } d)$.

Пусть $r = 0$. Покажем, что $s_k = s_n = s_m$. Для этого построим последовательности k_i, a_i, b_i, заданные следующими правилами:

1. $k_0 = n$, $a_0 = m/d - 1, b_0 = n/d - 1$;

2. $\begin{cases} k_{i+1} = k_i + n, a_{i+1} = a_i - 1, b_{i+1} = b_i, & \text{если } a_i > 0 \text{ и } k_i < m \\ k_{i+1} = k_i - m, a_{i+1} = a_i, b_{i+1} = b_i - 1, & \text{если } a_i = 0 \text{ и } b_i > 0 \\ i - \text{ый член последовательности — последний}, & \text{если } a_i = b_i = 0 \end{cases}$.

Отметим, что

1. $k_i - n = a_i n - b_i m$ для всех $i \geqslant 0$;

2. $k_i \vdots d$ для всех $i \geqslant 0$;

3. $1 \leq k_i \leq m + n - 1$ для всех $i \geqslant 0$.

Пусть $k_i = n$. Тогда $na_i = mb_i$ и, в частности, или $a_i = 0$, или $a_i \geq m/d$. Так как последнее невозможно, то $a_i = b_i = 0$. Наоборот, если $a_i = b_i = 0$, то $k_i = n$. Таким образом, в последовательностях $\{a_i\}, \{b_i\}, \{k_i\}$ ровно $m/d + n/d - 1$ член. Покажем теперь, что любые два члена последовательности $\{k_i\}$ различны. Действительно, если $k_i = k_j$, то $n(a_i - a_j) = m(b_i - b_j)$. В частности $a_i \geq m/d$. Что невозможно. Таким образом, любые два члена последовательности $\{k_i\}$ попарно различны, принадлежат множеству $\{d, ..., d(m/d + n/d - 1)\}$, а всего элементов в этой последовательности — $m/d + n/d - 1$. Откуда следует, что члены этой последовательности исчерпывают все числа, делящиеся на d, от d до $d(m/d + n/d - 1)$, и, так как $s_{k_i} = s_{k_{i+1}}$, $s_k = s_n = s_m$.

В частности, отсюда следует, что если $k, l \vdots d$ и $k = l \pmod d$, то $s_k = s_l$. Так как пересечение двух периодических множеств имеет длину $m+n-1$ и d-периодично, то каждое из слов u, v (они имеют длину меньшую $m+n-1$!) периодично. Противоречие. Следовательно, пересечение u^{mn} и v^{mn} не может иметь длину, большую $m+n-2$. □

Задача 2.10. *На бесконечной ленте в каждой ячейке написаны цифры от 1 до 9. Докажите, что тогда либо из неё можно вырезать 10 непересекающихся тысячезначных чисел в порядке убывания, либо какое-то число длины меньше 10 повторится 50 раз подряд.*

Доказательство. Рассмотрим момент, с которого все встречаемые 1000-значные числа повторяются бесконечное число раз. Если их хотя бы 10, то найдем наибольшее, затем далее него найдем второй по величине и так далее. А если их менее 10, то тогда существует 2 одинаковых 1000-значных слова, у которых расстояние между началами менее 10. Значит, такое число (частично) периодично с периодом не более 10. А, значит, мы нашли требуемое 50-ти разовое повторение слова длины менее 10. □

3 Цикл "Теорема Дилуорса"

Задача 3.1. *Из любых ли пяти различных чисел, выписанных в ряд, можно выбрать три, стоящие в этом ряду в порядке убывания или в порядке возрастания?*

Доказательство. Является частным случаем задачи 3.4. □

Задача 3.2. *Из любых ли девяти различных чисел, выписанных в ряд, можно выбрать четыре, стоящие в этом ряду в порядке убывания или в порядке возрастания?*

Решение. Нет, не из любых. Например, 3-2-1-6-5-4-9-8-7. □

Задача 3.3. *Докажите, что из любых десяти различных чисел, выписанных в ряд, можно выбрать четыре, стоящие в этом ряду в порядке убывания или в порядке возрастания.*

Доказательство. Является частным случаем задачи 3.4. □

Задача 3.4. *Докажите, что среди любых $mn+1$ различных чисел найдутся либо $m+1$ в порядке убывания, либо $n+1$ в порядке возрастания.*

Доказательство. Положим, что $a \succ b$, если $a > b$ и a стоит после b. А все остальные пары назовем несравнимыми. Из задачи 3.7 мы можем найти либо $m+1$ цепь относительно $\succ$ (что соответствует $m+1$ элементу в порядке возрастания), либо $n+1$ попарно несравнимый элемент относительно $\succ$, что соответствует убывающей последовательности из $n+1$ элемента. □

Частично упорядоченное множество (ЧУМ) M — это множество, для любых двух элементов a, b которого известно, находятся они в некотором отношении $\prec$ или нет. При этом должны быть выполнены следующие аксиомы:

X если $a \prec b$ и $b \prec c$, то $a \prec c$;

X если $a \prec b$, то a не равно b.

Задача 3.5. *Возможно ли, что неравенства $a \prec b$ и $b \prec a$ выполнены одновременно?*

Задача 3.6. *Докажите, что слова образуют ЧУМ при отношении, порождённом лексикографическим порядком.*

Определение 3.1. *Множество, любые два элемента которого сравнимы, называют линейно упорядоченным или, коротко, цепью.*

Задача 3.7. *Пусть m, n – некоторые натуральные числа. В частично упорядоченном множестве из $mn+1$ элементов есть либо цепь из идущих в порядке возрастания $m+1$ элементов, либо $n+1$ попарно несравнимых элементов (так называемая антицепь).*

Доказательство. Будем доказывать по индукции по числу m. База для $m = 0$ очевидна. Докажем переход. Рассмотрим наш ЧУМ. Назовем число максимальным, если нет никакого, большего его. Рассмотрим все максимальные числа. Из определения видно, что никакие 2 максимальных не являются сравнимыми. Если их хотя бы $n+1$, то мы нашли требуемую антицепь. Если из не более, чем n, то забудем временно про них и найдем среди оставшихся (которых не меньше, чем $n(m-1)+1$ либо антицепь длины $n+1$ (уже победа), либо цепь длины m. Во втором случае рассмотрим максимальный элемент этой цепи. Раз он не забыт, то есть забытое число, большее данного, а, соответственно, и всех элементов цепи. Добавим забытое число, и, тем самым, построим требуемую цепь длины $m+1$. □

Задача 3.8. *Обозначим через d наибольшее количество элементов цепи данного конечного частично упорядоченного множества M. Тогда M можно разбить на d антицепей.*

Решение. (Все новые определения есть в решении задачи 3.7) Рассмотрим все максимальные элементы и объединим их в одну антицепь. Забудем про них. Заметим, что не осталось цепей длины n, так как иначе вместе с забытыми мы можем найти цепь длины $n+1$. И будем действовать таким образом и дальше. Ура. □

Более того, верен следующий факт, который окажет нам серьёзную помощь в дальнейшем:

Теорема Дилуорса. Обозначим через n наибольшее количество элементов антицепи данного конечного частично упорядоченного множества M. Тогда M можно разбить на n цепей.

Доказательства теоремы Дилуорса и других задач из посвящённого ей цикла можно прочитать в [17].

4 Цикл "Экспоненциальная оценка"

Пусть наш алфавит состоит из букв $a_1, a_2, \ldots, a_l$. Будем считать, что

$$a_1 \prec a_2 \prec \cdots \prec a_l.$$

Таким образом, мы упорядочили буквы алфавита. Рассмотрим теперь два слова u и v. Если одно из них является началом другого, то назовём слова u и v *несравнимыми* (по отношению друг к другу). В противном случае найдутся слова w, u', v' такие, что $u = wu'$, $v = wv'$, причём первые буквы у слов u' и v' – различные (w может быть пустым, u' и v' – нет). Если первая буква u' больше первой буквы v', то считаем слово u больше слова v, в обратном случае считаем u меньше v. Таким образом, мы частично упорядочили слова. Приведённый порядок называется *лексикографическим* (мы уже обсуждали его в предисловии). Не стоит также забывать, что некоторые слова так и остались несравнимыми.

Решения большинства задач этого параграфа можно найти в статье [15].

Задача 4.1. *Пусть алфавит состоит из трёх букв: a, b и c. Введём на них порядок $a < b < c$. Составьте из приведённого ниже списка слов наиболее длинную возрастающую последовательность. Какие пары слов являются несравнимыми?*

$$cb, abc, bac, abb, b, ccc, abc.$$

Решение. Наибольшая возрастающая подпоследовательность: abb, abc, b, cb, ccc.

Пары несравнимых слов: $b \leftrightarrow bac, abc \leftrightarrow abc$. □

Для дальнейшей работы нам потребуется ввести несколько вспомогательных определений.

Определение 4.1. *Слово W – n-*разбиваемо*, если найдутся слова $u_1, u_2, \ldots, u_n$ такие, что $W = v \cdot u_1 \cdots u_n$, при этом $u_1 \succ \ldots \succ u_n$.*

Определение 4.2. *Слово называется k-*порядочным*, если оно k-разбиваемо, но не $(k+1)$-разбиваемо.*

Задача 4.2. *Найдите число*

a) 1-порядочных слов длины s с попарно различными буквами;

b) 2-порядочных слов, в которых используется ровно l букв.

c) 1-порядочных слов длины s, буквы которых необязательно различны (считать, что в алфавите — l букв).

Решение. a) Ответ: C_s^l.

b) Ответ: $\frac{1}{l+1}\mathrm{C}_{2l}^l$.

Из задачи 2.7 следует, что искомое число слов равно числу последовательностей натуральных чисел $1 \leqslant a_1 \leqslant a_2 \leqslant \ldots \leqslant a_l$, для которых $a_i \leqslant i$. Такие последовательности будем называть *корректными*. Пусть c_n — число корректных последовательностей из n элементов. Для корректной последовательности $\{a_i\}$ положим

$$\mathrm{stup}(a_1, a_2 \ldots, a_{n+1}) = \sup_{1 \leqslant i \leqslant n+1} \{a_i = i\}.$$

Очевидно, что $1 \leqslant \mathrm{stup}(\{a_i\}) \leqslant n$. Тогда корректные последовательности, заданные для некоторого $1 \leqslant j \leqslant n+1$ условием $\mathrm{stup}(a_1, a_2 \ldots, a_{n+1}) = j$, можно задать также следующим набором условий:

$$a_i \leqslant i \text{ для } i < j; \quad a_j = j; \quad a_i \leqslant i-1 \text{ для } i > j.$$

Таким образом, число корректных последовательностей, для которых

$$\mathrm{stup}(a_1, a_2 \ldots, a_{n+1}) = j,$$

равно $c_{j-1}c_{n+1-j}$. Так как j может принимать все натуральные значения в диапазоне $1 \leqslant j \leqslant n+1$, а других значений принимать не может, то получаем, что

$$c_{n+1} = c_0c_n + c_1c_{n-1} + \ldots + c_nc_0. \quad (1)$$

Покажем теперь, что $c_n = \frac{1}{n+1}\mathrm{C}_{2n}^n$. Для этого введём функцию

$$f(x) = c_0 + c_1x + c_2x^2 + \ldots + c_nx^n + \ldots.$$

Тогда из соотношения (1) следует, что $f(x) = c_0 + xf^2(x)$. Откуда

$$f(x) = \frac{1 \pm \sqrt{1-4c_0x}}{2x} = {}^2\frac{1 \pm \sqrt{1-4x}}{2x}.$$

Так как $f(x)$ не имеет полюса в 0,

$$f(x) = \frac{1-(1-4x)^{\frac{1}{2}}}{2x} = \sum_{n \geq 0} \tfrac{1}{2}(-1)^{n+2}4^{n+1}\mathrm{C}_{n+1}^{\frac{1}{2}}x^n = \sum_{n \geq 0} 4^{n+1}\tfrac{(2n-1)!!}{2^{n+2}}{}^3x^n = \sum_{n \geq 0} \tfrac{1}{n+1}\mathrm{C}_{2n}^n x^n.$$

Откуда $c_l = \frac{1}{l+1}\mathrm{C}_{2l}^l$.

c) В 1-порядочном слове буквы идут по возрастанию. Следовательно, 1-порядочные слова находятся во взаимно однозначном соответствии с упорядоченными наборами неотрицательных целых чисел $(k_1, \ldots, k_l)$, для которых $k_1 + \ldots + k_l = s$. Как известно, число таких наборов равно C_{s+l-1}^s. □

Задача 4.3. *a) Пусть n – некоторое натуральное число, u – нециклическое слово длины не меньше n. Докажите, что слово u^{2n} является n-разбиваемым.*

b) Пусть u — некоторое слово длины $(n-1)$. Докажите, что слово u^{2n} — не n-разбиваемое.

[2] То, что c_0=1 проверяется непосредственно.

Доказательство. a) Из слова u длины $m \geqslant n$ с помощью циклических перестановок можно получить m слов: $u_0, u_1, \ldots, u_{m-1}$. Так как слово u — нециклическое, то все слова u_i попарно различны. Предположим, что в лексикографическом смысле $u_{i_0} > u_{i_1} > \ldots > u_{i_{m-1}}$. Представим каждое слово u_i в виде $u_i = v_i w_i$, где $u = w_i v_i$. Рассмотрим теперь слово

$$u^{2n} = w_{i_0} v_{i_0} w_{i_0} v_{i_0} w_{i_1} v_{i_1} w_{i_1} v_{i_1} \ldots w_{i_{m-1}} v_{i_{m-1}} w_{i_{m-1}} v_{i_{m-1}}.$$

Положим $u'_{i_k} = v_{i_k} w_{i_k} v_{i_k} w_{i_{k+1}}$ для $k = 0, 1, \ldots, n-2$;
$u'_{i_{n-1}} = v_{i_{n-1}} w_{i_{n-1}} v_{i_{n-1}}$; $\gamma = w_{i_0}$.
Тогда слово u^{2n} представится в виде $u^{2n} = \gamma u'_{i_0} u'_{i_1} \ldots u'_{i_{n-1}}$. Так как

$$u'_{i_0} > u'_{i_1} > \ldots > u'_{i_{n-1}},$$

получаем, что слово u^{2n} является n-разбиваемым (см. также [18]).

b) Пусть $u = u_1 ... u_s$, где $s \leq n-1$. Пусть u^{2n} — n-разбиваемое слово, то есть содержит непересекающиеся подслова $v_1, ..., v_n$, идущие в порядке убывания. Пусть $r_1, ..., r_n$ — номера, считая с начала слова u^{2n}, первых букв $v_1, ..., v_n$. В силу того что $s < n$, существуют $1 \leqslant i, j \leqslant n$, для которых $r_i = r_j (\mathrm{mod}\ s)$. Тогда либо v_i подслово v_j, либо v_j подслово v_i. В любом случае, v_i и v_j несравнимы либо равны, что противоречит n-разбиваемости u^{2n}. □

Определение 4.3. *a) Слово v —* хвост *слова u, если найдется слово w такое, что $u = wv$.*

б) Если в слове v содержится подслово вида u^t, то будем говорить, что в слове v содержится период цикличности *t.*

Задача 4.4. *Пусть x, d — некоторые натуральные числа. Докажите, что в слове W длины x либо первые $[x/d]$ хвостов попарно сравнимы, либо в слове W найдется период длины d.*

Решение. Пусть в слове W не нашлось слова вида u^d. Рассмотрим первые $[x/d]$ хвостов. Предположим, что среди них нашлись 2 несравнимых хвоста v_1 и v_2. Пусть $v_1 = u \cdot v_2$. Тогда $v_2 = u \cdot v_3$ для некоторого v_3. Тогда $v_1 = u^2 \cdot v_3$. Применяя такие рассуждения, получим, что $v_1 = u^d \cdot v_{d+1}$, так как $|u| < x/d$, $|v_2| \geqslant (d-1)x/d$. Противоречие. Доказательство также написано в работе [15, лемма 2.1]. □

Здесь и далее: если в формулировке задачи встречаются числа n и d, то считаем, что $n \leqslant d$.

Определение 4.4. *Слово W —* (n, d)-сократимое, *если оно либо n-разбиваемо, либо найдется u^d — подслово слова W.*

Задача 4.5. *Докажите, что если в слове W найдутся n одинаковых непересекающихся подслов длины n, то W — n-разбиваемое.*

Доказательство. Предположим противное. Рассмотрим хвосты $u_1, u_2, \ldots, u_n$ слова u, которые начинаются с каждой из его первых n букв. Перенумеруем хвосты так, чтобы выполнялись неравенства: $u_1 \succ \ldots \succ u_n$. Из леммы 1 они несравнимы. Рассмотрим подслово u_1, лежащее в самом левом экземпляре слова u, подслово u_2 – во втором слева,..., u_n – в n-ом слева. Получили n-разбиение слова W. Противоречие (см. также [15, лемма 2.3]). □

Определение 4.5. *Слово W будем называть n-разбиваемым в хвостовом смысле, если найдутся хвосты $u_1, \dots, u_n$ такие, что $u_1 \succ u_2 \succ \dots \succ u_n$ и для любого $i = 1, 2, \dots, n-1$ начало u_i слева от начала u_{i+1}.*

Обозначим через $|W|$ длину слова W.

Задача 4.6. *Докажите, что если слово W является*
а) n^3d-разбиваемым в хвостовом смысле,
b) $3n^2d$-разбиваемым в хвостовом смысле,
*c)** $4nd$-разбиваемым в хвостовом смысле,*
то оно — либо n-разбиваемо, либо W содержит подслово в степени d.

Решение. Мы докажем пункт б) пункт а) из него очевидно следует.

Предположим противное. Рассмотрим порядковые номера позиций букв a_i, где $a_1 < a_2 < \dots < a_{3n^2d}$, с которых начинаются хвосты u_i, разбивающие W. Пусть

$$X_k = \{nd - \text{хвосты } u_i \mid i = 3knd + 1, \dots, 3knd + 2nd\}.$$

Тогда для произвольных различных чисел i и j если $u \in \chi_i, v \in X_j$, то u и v не пересекаются. В этом случае найдется k такое, что любые непересекающиеся слова u, v из X_k несравнимы. Без ограничения общности можно считать, что $k = 1$. Пусть подслово v_i есть $n \cdot d$-хвост u_i. Подслова v_1 и v_{nd+c} не пересекаются для любого $c \in [1, nd]$. Значит $v_{nd+s} = v_{nd+t}$ для любых $1 \leqslant s \leqslant t \leqslant nd$, а так как $a_{nd+t} - a_{nd+s} > n$, то подслова

$$u_1, u_{nd+1}, u_{nd+d+1}, u_{nd+2d+1}, \dots, u_{2nd-d+1}$$

не пересекаются. Следовательно, они несравнимы, а, значит, слово W является n-сократимым. Противоречие. □

Задача 4.7. *Для каждой пары натуральных чисел n, d приведите пример не $(nd-1)$-разбиваемого в хвостовом смысле слова W такого, что W — не $(n+1)$-разбиваемо и не содержит подслова в степени d.*

Решение. Пример, построенный командами “Харьков” и “Девушки”:

Пусть в алфавите 2 буквы $a < b$. Искомое слово W равно

$$W = (a^{n-1}b)^{d-1}a^{n-1}.$$

□

Задача 4.8. *Попробуйте улучшить оценку в задаче 4.6.*

Комментарий. Если такая оценка есть, то она больше $(n-1)(d-1)$, так как для алфавита с буквами $a_1 \prec a_2 \prec \dots \prec a_{n-1} \prec \dots \prec a_s$ слово $a_1^{d-1}a_2^{d-1}\dots a_{n-1}^{d-1}$ является $(n-1)(d-1)$-разбиваемым в хвостовом смысле, но не n-разбиваемо в обычном смысле и не содержит периода в степени d. □

Фиксируем алфавит из l букв, слово W длины $r(W)$ над этим алфавитом и натуральные числа $n \leqslant d$. Далее будем считать, что W не содержит подслова в степени d и слово W не $4nd$-разбиваемо в хвостовом смысле. Рассмотрим его первые $[r(W)/d]$ хвостов (далее будем обозначать это множество хвостов за Ω). Тогда по теореме Дилуорса их можно раскрасить в $(4nd-1)$ цветов так, чтобы хвосты одного цвета шли в порядке возрастания. Для решения следующих задач следует пользоваться предыдущими задачами из цикла.

Задача 4.9. *Докажите, что среди любых $4nd^2$ хвостов из Ω найдутся два одноцветных хвоста, у которых отличаются начальные подслова длины $4nd$.*

Решение. Доказательство написано в работе [15, начало §3]. □

Задача 4.10. [4] *В бесконечном парламенте у каждого не более 3 врагов. Доказать, что его можно разбить на 2 палаты так, что у каждого будет не более одного врага в своей палате.*

Доказательство. Будем говорить, что два парламентёра (далее, для краткости, п.) A и B *связаны далёкой враждой*, если существует цепочка п. $A_0, ..., A_n$, для которых $A_0 = A$, $A_n = B$ и B_{i+1} враждует с B_i для всякого $i < n$. Очевидно, что это отношение симметрично и транзитивно. Также очевидно, что число п., связанных с данным п. далёкой враждой счётно. Таким образом, весь парламент распадается на группы такие, что члены разных групп не враждуют между собой и что число членов в разных группах счётно. Понятно [5], что достаточно доказать, что каждую такую группу можно разбить на две палаты так, чтобы у каждого её члена был не более, чем один враг.

Далее мы считаем, что число п. счётно. Занумеруем их натуральными числами. Будем обозначать через P_n множество из п. с номерами 1,..., n. Разбиение на палаты мы будем понимать как множество функций $f : P_n \to \{1, 2\}$ (для каждого п. мы указываем относится он к первой палате или второй). Множество разбиений на палаты, в котором у каждого п. из P_n неболее одного врага в своей палате обозначим H_n (назовём такие разбиения *допустимыми*). Очевидно, что ограничение допустимого разбиения с множества п. P_n на множество п. P_m $(m < n)$ — допустимо. Мы будем обозначать множество ограничений H_n на H_m как $(H_n)|_m$.

Покажем, что для каждого n множество H_n не пусто. Разместим как-нибудь всех п. в две палаты. Далее, если какой-то п. имеет двух врагов в своей палате, то мы его перемещаем в другую палату. При этом число пар врагов в сумме внутри обеих палатах падает. Так как всего число пар врагов конечно, то после какого-то числа операций у каждого парламентёра будет не более одного врага в своей палате. Т.е. мы показали, что для всякого n множество H_n непусто.

Заметим теперь, что если $m > n > k$, то $((H_m)|_n)|_k = (H_m)_k$, в частности, $(H_m)|_k \subset (H_m)|_n$. Для всякого n положим

$$(H_\infty)_n := \cap_{m \geq n}((H_m)|_n).$$

Покажем, что для всякого $n \geq 1$ множество $(H_\infty)_n$ непусто. Напомним, что если $m_1 > m_2 > n$, то $(H_{m_1})|_n \subset (H_{m_2})_n$. Таким образом, если $(H_\infty)_n = 0$, то $(H_m)|_n = 0$ для какого-то $m > n$. Что невозможно.

Заметим также, что если $m > n$, то

$$(H_\infty)_n = ((H_\infty)_m)|_n. \quad (2)$$

Теперь построим цепочку $\{f_i\}(f_i \in (H_\infty)_i)$ по следующему правилу: $f_{i+1}|_{P_i} = f_i$ (такая функция f_{i+1} обязательно существует для каждой функции f_i, так как $((H_\infty)_{i+1})|_i = (H_\infty)_i$). Эта цепочка задаёт разбиение всех п. на два парламента: i-ый п. находится в $f_i(i)$ палате. □

[4] Эта задача имеет малое отношение к теме проекта. Она достаточно интересна сама по себе и вызвала оживление среди членов жюри.

[5] здесь мы пользуемся аксиомой выбора

Теперь перестанем фиксировать числа l, n, d и слово W. Будем считать, что l, n, d — некоторые натуральные числа такие, что $n \leqslant d$, а слово W — некоторое слово над алфавитом из l букв.

Задача 4.11 (Лемма Ширшова). *Докажите, что существует функция от натуральных аргументов $f(l, n, d)$ такая, что для любого слова W над алфавитом длины l, не являющегося (n, d)-сократимым, $r(W) < f(l, n, d)$.*

Доказательство. Доказательство леммы Ширшова следует из задачи 5.5. Оригинальное доказательство Ширшова содержится в работе [18]. □

Задача 4.12. *Докажите, что $f(l, n, d) < l(4nd)^{4nd+2}$.*

Доказательство. Доказательство леммы Ширшова следует из задачи 5.5.

□

5 Улучшение экспоненциальной оценки

Решения задач 5.1 — 5.5 находится в параграфе 3 работы [15], 5.6 решена в параграфах 4 и 5 той же статьи.

Список литературы

[1] М. И. Харитонов *Оценки на структуру кусочной периодичности в теореме Ширшова о высоте*, Вестник Московского университета, Серия 1, Математика. Механика. №6(2012).

[2] Abraham A. Klein. *Indices of nilpotency in a PI-ring.* Archiv der Mathematik, 1985, vol 44:4.

[3] Abraham A. Klein *Bounds for indices of nilpotency and nility.* Archiv der Mathematik, 2000, vol 74:1 pages 6—10

[4] Е. С. Чибриков. *О высоте Ширшова конечнопорождённой ассоциативной алгебры, удовлетворяющей тождеству степени четыре.* Известия Алтайского государственного университета т. 1(19), 2001, стр. 52—56

[5] М. И. Харитонов *Двусторонние оценки существенной высоты в теореме Ширшова о высоте.* Вестник Московского университета, Серия 1, Математика. Механика., 2(2012), 24—28.

[6] A. A. Lopatin. *On the nilpotency degree of the algebra with identity $x^n = 0$.* arXiv:1106.0950v1.

[7] *Днестровская тетрадь: оперативно-информац. сборник* No 4, Новосибирск, изд. ин–та матем. СО АН СССР, 1993, 73 стр.

[8] И. И. Богданов *Теорема Нагаты-Хигмана для полуколец.* Фундамент. и прикл. матем. т. 7:3, 2001, 651—658.

[9] Колотов А. Г. *О верхней оценке высоты в конечно порожденных алгебрах с тождествами.* Сиб. мат. ж., 1982, т. 23, no 1, стр. 187—189.

[10] Кострикин А. И. *Вокруг Бернсайда. — М.: Наука, 1986*, 232 стр.

[11] Курош А. Г. *Проблемы теории колец, связанные с проблемой Бернсайда о периодических группах.* Изв. АН СССР, сер. мат., 1941, т. 5, стр. 233—240.

[12] Латышев В. Н. *К теореме Регева о тождествах тензорного произведения* PI-*алгебр.* Успехи мат. наук, 1972, т. 27, no 4, стр. 213—214.

[13] Ширшов А. И. *О некоторых неассоциативных нильь-кольцах и алгебраических алгебрах.* Мат. сб., 1957, т. 41, no 3, стр. 381—394.

[14] Ширшов А. И. *О кольцах с тождественными соотношениями.* Мат. сб., 1957, т. 43, no 2, стр. 277—283.

[15] А. Я. Белов, М. И. Харитонов. *Субэкспоненциальные оценки в теореме Ширшова о высоте.* Мат. сб., **203**:4(2012), 81—102.

[16] Belov A. *Some estimations for nilpotence of nil-algebras over field of an arbitrary characteristics and height theorem.* Comm. in Algebra, 1992, vol. 20, N 10, p. 2919—2922.

[17] А. Спивак. *Цепи и антицепи.* Квант, 5(2003), 11—14.

[18] Жевлаков, Слинько, Шестаков, Ширшов. *Кольца близкие к ассоциативным.* М., Наука, 1978.

Amazing properties of binomial coefficients

Several research topics will be set to you at the conference. Your aim is the maximal advance in one of these topics. You can co-operate in the solving of problems, arbitrary teams are allowed (i.e. the team may consist of participants from different cities). If you solve problems in different topics you may take part in different teams. The only thing you should avoid is to sign up the solutions of those problems that you really were not solving (this may happen if the team is too big and not all of its members solve the problems of some topic actively).

The following is the introductory set of problems about binomial coefficients. You may hand in the (written) solutions to Kokahs K. (coach 15, seat 17) In Teberda the set of problems will be enlarged a lot and you may hand in your solutions of this set of problems, except 1.2, in Teberda, too. You can hand in the solutions of the problem 1.2 in train only.

1 Problems for solving in train

1.1. Prove that a) $\binom{p-1}{k} \equiv (-1)^k \pmod p$; b) $\binom{2n}{n} \equiv (-4)^n \binom{\frac{p-1}{2}}{n} \pmod p$ при $n \leqslant \frac{p-1}{2}$.

1.2. Prove that the number of odd binomial coefficients in n-th row of Pascal triangle is equal to 2^r, where r is the number of 1's in the binary expansion of n.

1.3. Fix a positive integer m. By a m-arithmetical Pascal triangle we mean a triangle in which binomial coefficients are replaced by their residues modulo m. We will also consider similar triangles with the arbitrary residues a instead of 1's along the lateral sides of the triangle. The operation of the multiplying by a number and addition of triangles of equal size are correctly defined. We will consider these operations modulo m.

$$\begin{array}{ccccccccccc}
 & & a & & & & b & & & & a+b \\
 & a & & a & + & b & & b & = & a+b & & a+b \\
a & & 2a & & a & & b & 2b & b & a+b & 2(a+b) & & a+b \\
a & 3a & & 3a & a & b & 3b & & 3b & b & a+b & 3(a+b) & 3(a+b) & a+b
\end{array}$$

$$x \cdot \begin{array}{ccccc} & & a & & \\ & a & & a & \\ a & & 2a & & a \\ a & 3a & & 3a & a \end{array} = \begin{array}{ccccc} & & ax & & \\ & ax & & ax & \\ ax & & 2ax & & ax \\ ax & 3ax & & 3ax & ax \end{array}$$

Let all the elements of s-th row of m-arithmetical Pascal triangle except the first and the last one be equal to 0. Prove that the triangle has a form depicted on fig. 1. Shaded triangles consist of zeroes, triangles Δ_n^k consist of s rows and satisfy the following relations

1) $\Delta_n^{k-1} + \Delta_n^k = \Delta_{n+1}^k$; 2) $\Delta_n^k = C_n^k \cdot \Delta_0^0 \pmod m$.

The well known puzzle Tower of Hanoi consists of three rods, and a number of disks of different sizes which can slide onto any rod. The puzzle starts with the disks in a neat stack in ascending order of size on one rod, the smallest at the top, thus making a conical shape. The objective of the puzzle is to move the entire stack to another rod, obeying the following rules: 1) only one disk may be moved at a time; 2) each move consists of taking the upper disk from one of the rods and sliding it onto another rod, on top of the other disks that may already be present on that rod; 3) no disk may be placed on top of a smaller disk.

Let n be the number of disks. Let TH_n be a graph, whose vertices are all possible correct placements of disks onto 3 rods and edges connect placements that can be obtained one from another by 1 move. Consider also graph P_n, whose vertices are 1's located in the first 2^n rows of the 2-arithmetical Pascal triangle and edges connect neighboring 1's (i.e. two adjacent 1's in the same row or neighboring 1's by a diagonal in two adjacent rows)

1.4. prove that graphs TH_n and P_n are isomorphic.

1.5. Prove that that first 10^6 rows of 2-arithmetical Pascal triangle contain less than 1 % of 1's.

1.6. Prove that if n is divisible by $p-1$, then $\binom{n}{p-1} + \binom{n}{2(p-1)} + \binom{n}{3(p-1)} + \ldots + \binom{n}{n} \equiv 1 \pmod p$. Or, even better prove the general statement: if $1 \leqslant j, k \leqslant p-1$ и $n \equiv k \pmod{p-1}$, then

$$\binom{n}{j} + \binom{n}{(p-1)+j} + \binom{n}{2(p-1)+j} + \binom{n}{3(p-1)+j} + \ldots \equiv \binom{k}{j} \pmod p.$$

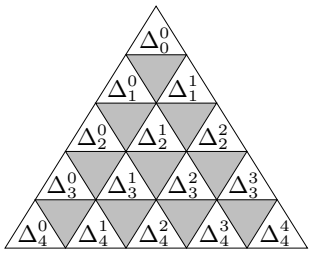


Рис. 1:

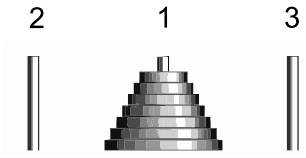


Рис. 2:

Amazing properties of binomial coefficients — 2

"The official theoretical source" for this set of problems is Vinberg's article [1]. Particularly the following theorems are considered to be known.

1. Wilson's theorem. For any prime p (and for primes only) the equivalence holds $(p-1)! \equiv -1 \pmod p$.
2. Lukas' theorem. Write the numbers n and k in base p:
$$n = n_d p^d + n_{d-1} p^{d-1} + \ldots + n_1 p + n_0, \qquad k = k_d p^d + k_{d-1} p^{d-1} + \ldots + k_1 p + k_0. \tag{1}$$
Then $\binom{n}{k} \equiv \binom{n_d}{k_d}\binom{n_{d-1}}{k_{d-1}} \cdot \ldots \cdot \binom{n_1}{k_1}\binom{n_0}{k_0} \pmod p$.
3. Kummer's theorem. The exponent $\operatorname{ord}_p \binom{n}{k}$ is equal to the number of "carries" when we add k and $\ell = n-k$ in base p.
4. Wolstenholme's theorem. If $p \geqslant 5$ then $\binom{2p}{p} \equiv 2 \pmod{p^3}$, or, that is the same, $\binom{2p-1}{p-1} \equiv 1 \pmod{p^3}$.

Remind that $\binom{0}{0} = 1$, $\binom{n}{k} = 0$ for $k > n$ and for $k < 0$ by definition.

We denote by p a prime number. For any natural n denote by $(n!)_p$ the product of all integers from 1 to n not divisible by p. If a number p is given the symbols n_i, m_i etc. denote the digits of numbers n, m etc. in base p.

* * *

2 Arithmetical triangle and divisibility

2.1. a) Prove that the first 3^k rows of 3-arithmetical Pascal triangle contain $\frac{1}{2}(6^k + 4^k)$ residues "1" and $\frac{1}{2}(6^k - 4^k)$ residues "2".

b) Find the number of zero elements in the first 5^k rows of 5-arithmetical Pascal triangle.

c) Find the number of non-zero elements in the first p^k rows of p-arithmetical Pascal triangle.

2.2. Prove that the number of 1's in the first m rows of 2-arithmetical Pascal triangle equals

$$\sum_{i=0}^{n-1} m_i \cdot 2^{\sum_{k=i+1}^{n-1} m_k} \cdot 3^i.$$

If $m = 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_r}$, where $\alpha_1 > \alpha_2 > \ldots > \alpha_r$, then we can rewrite the last expression in the form

$$3^{\alpha_1} + 2 \cdot 3^{\alpha_2} + 2^2 \cdot 3^{\alpha_3} + \ldots + 2^{r-1} \cdot 3^{\alpha_r}.$$

2.3. Consider n-th row of Pascal triangle modulo 2 as binary expansion of some integer P_n. Prove that

$$P_n = F_{i_1} \cdot \ldots \cdot F_{i_s},$$

where $i_1, \ldots, i_s$ are numbers of positions where 1's occur in the binary expansion of n, and $F_i = 2^{2^i} + 1$ is i-th Fermat number.

2.4. Prove that the number of non-zero elements in n-th row of p-arithmetical Pascal triangle equals $\prod_{i=0}^{d}(n_i + 1)$.

2.5. a) All the binomial coefficients $\binom{n}{k}$, where $0 < k < n$, are divisible by p if and only if n is a power of p.

b) All the binomial coefficients $\binom{n}{k}$, where $0 \leqslant k \leqslant n$, are not divisible by p if and only if $n+1$ is divisible by p^d, in other words, all the digits of n, except the leftmost, in base p are equal to $p-1$.

2.6. Let $0 < k < n+1$. Prove that if $\binom{n}{k-1} \not\vdots\, p$ and $\binom{n}{k} \not\vdots\, p$, then $\binom{n+1}{k} \not\vdots\, p$, *except* the case, when $n+1$ is divisible by p.

3 Generalization of Wilson's and Lukas' theorems

3.1. Prove that $\operatorname{ord}_p(n!) = \dfrac{n-(n_d+\ldots+n_1+n_0)}{p-1}$.

3.2. Prove the following generalizations of Wilson's theorem. a) $(-1)^{[n/p]}(n!)_p \equiv n_0! \pmod p$;
b) Prove that for $p \geqslant 3$

$$(p^q!)_p \equiv -1 \pmod{p^q},$$

and for $p=2$, $q \geqslant 3$ $\quad (p^q!)_p \equiv 1 \pmod{p^q}$.
c) $\dfrac{n!}{p^\mu} \equiv (-1)^\mu n_0! n_1! \ldots n_d! \pmod p$, where $\mu = \operatorname{ord}_p(n!)$

3.3. Generalized Lukas' theorem. Let $r = n-k$, $\ell = \operatorname{ord}_p\binom{n}{k}$. Then

$$\frac{1}{p^\ell}\binom{n}{k} \equiv (-1)^\ell \Big(\frac{n_0!}{k_0! r_0!}\Big)\Big(\frac{n_1!}{k_1! r_1!}\Big)\ldots\Big(\frac{n_d!}{k_d! r_d!}\Big) \pmod p$$

3.4. a) Prove that $(1+x)^{p^d} \equiv 1 + x^{p^d} \pmod p$ for all $x = 0, 1, \ldots, p-1$.
b) Prove Laukas' theorem algebraically.

3.5. a) Let m, n, k be nonnegative integers, and $(n,k)=1$. Prove that $C_{mn}^k \equiv 0 \pmod n$.
b) Prove that if $n \vdots p^k$, $m \not\vdots p$, then $\binom{n}{m} \vdots p^k$.

3.6. Let $f_{n,a} = \sum\limits_{k=0}^{n} \big(\binom{n}{k}\big)^a$. Prove that $f_{n,a} \equiv \prod\limits_{i=0}^{d} f_{n_i,a} \pmod p$.

4 Variations on Wolstenholme's theorem

4.1. Prove that $\dfrac{1}{1} + \dfrac{1}{2} + \ldots + \dfrac{1}{p-1} \equiv 0 \pmod{p^2}$.

4.2. Let $p = 4k+3$ be a prime number. Find $\dfrac{1}{0^2+1} + \dfrac{1}{1^2+1} + \ldots + \dfrac{1}{(p-1)^2+1} \pmod p$.

4.3. a) Let k be a nonnegative integer such that for any prime divisor p of the number m $\quad k$ is not divisible by $(p-1)$. Prove that $\dfrac{1}{1^k} + \dfrac{1}{2^k} + \ldots + \dfrac{1}{(p-1)^k} \equiv 0 \pmod m$ (summation over all fractions whose denominators are coprime to m).

b) Let k be odd and $(k+1) \not\vdots (p-1)$. Prove that $\dfrac{1}{1^k} + \dfrac{1}{2^k} + \ldots + \dfrac{1}{(p-1)^k} \equiv 0 \pmod{p^2}$.

4.4. Prove that the equivalence (12) from Vinberg's article holds in fact modulo p^4.

4.5. Prove that the following properties are equivalent 1) $\binom{2p-1}{p-1} \equiv 1 \pmod{p^4}$;
2) $\dfrac{1}{1} + \dfrac{1}{2} + \ldots + \dfrac{1}{p-1} \equiv 0 \pmod{p^3}$; 3) $\dfrac{1}{1^2} + \dfrac{1}{2^2} + \ldots + \dfrac{1}{(p-1)^2} \equiv 0 \pmod{p^2}$.

4.6. a) Prove algebraically that for any prime p and arbitrary k and n $\quad \big(\binom{pk}{pm} - \binom{k}{m}\big) \vdots p^2$. (In Vinberg's article this fact is proven combinatorially.
b) Prove the statement (9) form Vinberg's article: for any prime $p \geqslant 5$ and arbitrary k and n $\quad \big(\binom{pk}{pm} - \binom{k}{m}\big) \vdots p^3$.

4.7. Let $p \geqslant 5$. Prove that a) $\binom{p^2}{p} \equiv \binom{p}{1} \pmod{p^5}$; b) $\binom{p^{s+1}}{p} \equiv p^s \pmod{p^{2s+3}}$.

4.8. Prove that $\binom{p^3}{p^2} \equiv \binom{p^2}{p} \pmod{p^8}$.

Amazing properties of binomial coefficients — 3

Additional problems to previous topics

2.7. Prove that $\binom{p^n-1}{k} \equiv (-1)^{S_k} \pmod p$, where S_k is the sum of digits of k in base p.

2.8. Prove that if the binomial coefficient $\binom{n}{k}$ is odd i.e. $k_i \leqslant n_i$ for all $i = 0, 1, \ldots, d$ in the notations of (1), then

$$\binom{n}{k} \equiv \prod_{i=1}^{d} (-1)^{k_{i-1}n_i + k_i n_{i-1}} \pmod 4.$$

2.9. Prove that if there are no two consecutive 1's in the binary expansion of n then all the odd entries in n-th row $\equiv 1 \pmod 4$, otherwise the number of entries $\equiv 1 \pmod 4$ equals the number of entries $\equiv -1 \pmod 4$.

2.10. Prove that the number of 5's in each row of 8-arithmetical Pascal triangle is a power of 2. Prove the same for 1's, 3's and 7's.

2.11. Prove that if we consider all the elements of the two sets

$$\left\{\binom{2^n-1}{1}, \binom{2^n-1}{3}, \binom{2^n-1}{5}, \ldots, \binom{2^n-1}{2^n-1}\right\} \quad \text{and} \quad \{1, 3, 5, \ldots, 2^n-1\}$$

as a reminders modulo 2^n, then these sets coincide.

2.12. Prove that elements of a row of Pascal triangle are not coprime in the following sence. For any $\varepsilon > 0$ there exists N, such that for all integer $n > N$ and $k_1, k_2, \ldots, k_{100} < \varepsilon\sqrt{n}$ the numbers

$$\binom{2n}{n+k_1}, \binom{2n}{n+k_2}, \ldots, \binom{2n}{n+k_{100}}$$

have a common divisor.

2.13. a) The non negative numbers $m > 1$, n, k are given. Prove that at least one of the numbers $\binom{n}{k}$, $\binom{n+1}{k}$, $\ldots$, $\binom{n+k}{k}$ is not divisible by m.

b) Prove that for each k there exist infinite set of numbers n, such that all the numbers $\binom{n}{k}$, $\binom{n+1}{k}$, $\ldots$, $\binom{n+k-1}{k}$ are divisible by m.

4.9. Prove that for $n > 1$ $\binom{2^{n+1}}{2^n} - \binom{2^n}{2^{n-1}}$ is divisible by 2^{2n+2}.

4.10. Prove that for $p \geqslant 5$ $(-1)^{\frac{p-1}{2}}\binom{p-1}{\frac{p-1}{2}} \equiv 4^{p-1} \pmod{p^3}$.

Amazing properties of binomial coefficients — 4

Additional problems to previous topics

4.11. Let m be a non negative integer, $p \geqslant 5$ be a prime. Prove that

$$\frac{1}{mp+1} + \frac{1}{mp+2} + \cdots + \frac{1}{mp+(p-1)} \equiv 0 \pmod{p^2}.$$

4.12. Let p and q be primes. Prove that $\binom{2pq-1}{pq-1} \equiv 1 \pmod{pq}$ if and only if $\binom{2p-1}{p-1} \equiv 1 \pmod{q}$ and $\binom{2q-1}{q-1} \equiv 1 \pmod{p}$.

5 Sums of binomial coefficients

5.1. a) Prove that the sum $\sum\limits_{k=0}^{3^a-1} \binom{2k}{k}$ is divisible by 3; b) is divisible by 3^a.

5.2. Let $C_k = \frac{1}{k+1}\binom{2k}{k}$ be Catalan numbers. Prove that $\sum\limits_{k=1}^{n} C_k \equiv 1 \pmod 3$ if and only if the number $n+1$ contains at least one digit "2" in base 3.

5.3. Let $p \geqslant 3$, $k = [2p/3]$. Prove that the sum $\binom{p}{1} + \binom{p}{2} + \ldots + \binom{p}{k}$ is divisible by p^2.

5.4. Let $n \vdots (p-1)$, where p is an odd prime. Prove that

$$\binom{n}{p-1} + \binom{n}{2(p-1)} + \binom{n}{3(p-1)} + \ldots \equiv 1 + p(n+1) \pmod{p^2}.$$

5.5. Prove that if $0 \leqslant j \leqslant p-1 < n$ and $q = \frac{n-1}{p-1}]$ then

$$\sum_{m:m\equiv j \pmod p} (-1)^m \binom{n}{m} \equiv 0 \pmod{p^q}.$$

5.6. Let p be an odd roime. Prove that $n \vdots (p+1)$ if and only if

$$\binom{n}{j} - \binom{n}{j+(p-1)} + \binom{n}{j+2(p-1)} - \binom{n}{j+3(p-1)} + \ldots \equiv 0 \pmod p$$

for all $j = 1, 3, \ldots, p-2$.

Solutions

1 Problems for solving in train

1.1. a) Solution 1. $\binom{p-1}{k} = \dfrac{(p-1)(p-2)\dots(p-k)}{1\cdot 2\cdots k} \equiv \dfrac{(-1)(-2)\dots(-k)}{1\cdot 2\cdots k} \equiv (-1)^k \pmod p$.

Solution 2. It is evident by the formula for binomial coefficients that $\binom{p}{i}$ is divisible by p when $1 \leqslant i \leqslant p-1$. Since $\binom{p-1}{k-1} + \binom{p-1}{k} = \binom{p}{k}$ and $\binom{p-1}{0} = 1 \equiv 1 \pmod p$, then $(\binom{p-1}{0} + \binom{p-1}{1}) \vdots p$, and therefore $\binom{p-1}{1} \equiv -1 \pmod p$. But $\binom{p-1}{1} + \binom{p-1}{2}$ is divisible by p also, hence $\binom{p-1}{2} \equiv 1 \pmod p$ etc.

б) This problem is taken from [3, problem 162]. Since the fractions $\binom{2n+2}{n+1}/\binom{2n}{n}$ and $\binom{\frac{p-1}{2}}{n+1}/\binom{\frac{p-1}{2}}{n}$ are highly reducible, the statement can be easily proven by induction. But we suggest a direct calculation from [3].

It easy to see that

$$\binom{2n}{n} = 2^n \cdot \frac{1\cdot 3\cdots(2n-1)}{n!}$$

and

$$1\cdot 3\cdots(2n-1) = (-1)^n(-1)(-3)\cdots(-2n+1) \equiv (-1)^n(p-1)(p-3)\cdots(p-2n+1) =$$

$$= (-1)^n 2^n \Big(\frac{p-1}{2}\Big)\Big(\frac{p-3}{2}\Big)\cdots\Big(\frac{p-2n+1}{2}\Big) = (-1)^n 2^n \Big(\frac{p-1}{2}\Big)\Big(\frac{p-1}{2}-1\Big)\cdots\Big(\frac{p-1}{2}-n+1\Big) =$$

$$= (-1)^n 2^n \frac{(\frac{p-1}{2})!}{(\frac{p-1}{2}-n)!} \pmod p.$$

Therefore $\binom{2n}{n} \equiv (-1)^n 4^n \dfrac{(\frac{p-1}{2})!}{n!(\frac{p-1}{2}-n)!} = (-4)^n \binom{\frac{p-1}{2}}{n} \pmod p$.

1.2. It follows directly from self-similar structure of an arithmetical Pascal triangle, that is described in the next problems. It follows from Lucas' theorem also, you can read the proof in [1].

1.3. We restrict ourselves with small contemplation, the full solution can be found in [3, problem 133].

Since the s-th row contains a long sequence of zeroes, then below these zeroes in $(s+1)$-th row we have the sequence of zeroes, too, (it is one element shorter than the upper sequence); in $(s+2)$-th row there are the sequence of zeroes also (it is one element shorter again) and so on. This explains the presence of the grey triangle below Δ_0^0 (fig. 1).

Further, the non-zero elements of the s-th row are equal to 1, hence the numbers situated along the sloped sides of the grey triangle all are 1's (due to the recurrence for binomial coefficients). So all the numbers along the sloped sides of the triangles Δ_1^0 and Δ_1^1 are 1's, and therefore both triangles are identical to Δ_0^0.

Now it is clear, what is the $(2s)$-th row of the triangle. The left- and the rightmost elements are 1's, all other elements equal 0, except the central element that is equal to 2, because it is a sum of the two upper 1's. Thus we obtain that two grey triangles are situated below $2s$-th row, the triangles Δ_2^0 and Δ_2^2 to the left and to the right of them are identical to Δ_0^0, and the triangle Δ_2^1 with 2's along its sloped sides is equal to $2\cdot\Delta_0^0$.

And so on.

1.4. This statement we found in [21], several facts about binomial coefficients are proven there via Tower of Hanoi and the graph TH_n.

Let a be the diameter of the upper disc on the first rod, b be the diameter of the upper disc on the second rod and c be the diameter of the upper disc on the third rod. W.l.o.g. $a < b < c$, then we have 3 possible moves in this configuration: from a to b or c and from b to c, we analogously have 3 moves if one rod is without discs. If all the discs are placed on one rod then we have 2 possible moves only; let A_1, A_2, A_3 denote the configurations of this type.

Observe that by the problem 1.2 all the elements of 2^s-th row of Pascal triangle are 1's. Therefore graph P_n has the rotational symmetry of the third order, because the recurrence $\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$, that allows us to construct the triangle from top to bottom, is equivalent in arithmetic modulo 2 to the recurrences $\binom{n}{k-1} = \binom{n}{k} + \binom{n+1}{k}$ and $\binom{n}{k} = \binom{n}{k-1} + \binom{n+1}{k}$, that allows us to construct the triangle from the low left

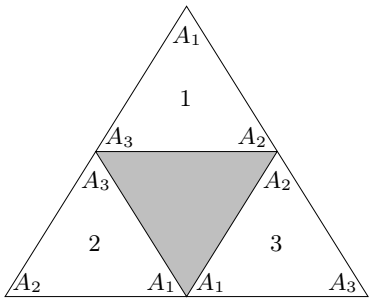


Рис. 3:

corner in the upper right direction and from the low right corner in the upper left direction. It follows also that the triangle of the double size contains 3 copies of the initial triangle.

Now let us prove by induction that there exists a bijection between TH_n and P_n, such that the vertices of the triangle P_n correspond to the configurations A_1, A_2, A_3. The base $n = 1$ is evident.

Proof of the step of induction. Assume that the bijection between TH_n and P_n has been constructed. The 2-arithmetical Pascal triangle with the side length 2^{n+1} contains 3 copies of the triangle with the side length 2^n. Number the copies and mark its vertices as shown on fig.3. Consider all the configurations of the Tower of Hanoi for which the $(n+1)$-th (biggest) disc is placed on rod i. If we fix the placement of this disc then displacements of other discs correspond to the graph that is isomorphic to TP_n. By induction hypothesis we can choose a bijection between this graph and the graph P_n in the i-th copy of the triangle, such that the configurations A_j correspond to the vertices of the triangles with the same marks. When we move the biggest disc, say, from the first rod to the second, all other discs must be on the 3rd rod. This move correspond to the edge connecting two neighboring vertices A_3 on the left sloped side of big triangle. The same reasons concern other moves of the biggest disc. Therefore we obtain an isomorphism between TP_{n+1} and P_n.

1.5. The bijection with Tower of Hanoi gives us a formula (when the number of rows is a power of 2): the first 2^k rows of 2-arithmetical Pascal triangle contain 3^k 1's. The formula can be also proved by induction via recurrence from the problem 1.3. Using this formula we can obtain an estimation. Since $10^6 < 2^{20}$, the total number of elements in these rows equals $\frac{1}{2} \cdot 10^6(10^6 + 1)$, and the number of 1's is at most 3^{20}. The proportion does not exceed $\frac{2 \cdot 3^{20}}{10^6(10^6+1)} \ll 0.01$.

1.6. We found this statement in [18].

Solution 1 ([CSTTVZ]). For $p = 2$ the statement can be easily checked. So we can assume that p is odd prime. Let $n = x(p-1) + k$. We use induction on x.

The base $x = 0$ is trivial: $\binom{k}{j} \equiv \binom{k}{j} \pmod p$.

To prove the step of induction we need the following property of binomial coefficients:

$$\binom{a+b}{s} = \sum_i \binom{a}{s-i}\binom{b}{i} \qquad \text{(summation in natural bounds)},$$

both sides of which calculate in how many ways we can choose s balls in the box that contains a black and b white balls. Let $n = m + (p-1)$. Observe that

$$\binom{n}{\ell(p-1)+j} = \binom{m+(p-1)}{\ell(p-1)+j} = \sum_{i=0}^{p-1} \binom{m}{\ell(p-1)+j-i}\binom{p-1}{i} \equiv \sum_{i=0}^{p-1} (-1)^i \binom{m}{\ell(p-1)+j-i} \pmod p$$

(the last equivalence is due to problem 1.1 a). Remark that the sign of the first and last terms in the last

sum is "plus" . Now transform the sum from the problem statement:

$$\sum_\ell \binom{n}{\ell(p-1)+j} \equiv$$
$$\equiv \left(\binom{m}{j} - \binom{m}{j-1} + \dots\right) + \left(\binom{m}{p-1+j} - \binom{m}{p-1+j-1} + \dots + \binom{m}{j}\right) +$$
$$+ \left(\binom{m}{2(p-1)+j} - \binom{m}{2(p-1)+j-1} + \dots + \binom{m}{2(p-1)+j}\right) + \dots$$
$$= \sum_{i=0}^{m} (-1)^i \binom{m}{i} + \sum_\ell \binom{m}{\ell(p-1)+j} \pmod p.$$

the first sum is equal to 0, the second sum is equivalent $\binom{k}{j} \pmod p$ by the induction hypothesis.

Solution 2 ([J], [T]). Induction by n. The base $n \leqslant p-1$ is trivial: both sides contain the same term. Prove the step of induction.

$$\binom{n}{j} + \binom{n}{(p-1)+j} + \dots = \left(\binom{n-1}{j} + \binom{n-1}{j-1}\right) + \left(\binom{n-1}{(p-1)+j} + \binom{n-1}{(p-1)+j-1}\right) + \dots =$$
$$= \left(\binom{n-1}{j} + \binom{n-1}{(p-1)+j} + \dots\right) + \left(\binom{n-1}{j-1} + \binom{n-1}{(p-1)+j-1} + \dots\right) \equiv$$
$$\equiv \binom{k-1}{j} + \binom{k-1}{j-1} = \binom{k}{j} \pmod p.$$

But it should be accurate in cases when $p-1$ divides j or k, because the induction hypothesis does not hold for $j=0$ or $k=0$ (it uses the value $p-1$ instead of 0). Therefore we must consider more carefully the cases when $j=1$ or $k=1$. We restrict ourselves by consideration of one partial case only. Let $p=5$, $j=1$ and we fulfill step to $n=13$. Then we have

$$\binom{1}{1} \stackrel{?}{\equiv} \binom{13}{1} + \binom{13}{6} + \binom{13}{11} = \left(\binom{12}{1} + \binom{12}{6} + \binom{12}{11}\right) + \left(\binom{12}{0} + \binom{12}{5} + \binom{12}{10}\right).$$

By induction hypothesis the sum in the first parentheses has a residue $\binom{4}{1}$ (and not $\binom{0}{1}$ as the previous calculation shows). In the second parentheses the induction hypothesis covers all the terms except the first one, so the sum has residue $\binom{12}{0} + \binom{4}{0}$. Writing $p-1$ instead of 4 for clarity, we obtain that the whole sum is equivalent to $\binom{n-1}{0} + \binom{p-1}{1} + \binom{p-1}{0} \equiv \binom{1}{1} \pmod p$, as required.

Solution 3 (algebraical reasoning with Luka's theorem, [18]). Induction by n. Base $n \leqslant p-1$ is trivial. Now let $n \geqslant p$, write all parameters in base p, let $\sigma_p(m)$ denotes the sum of digits of m. It is clear that if $m \equiv j \pmod p$, then $\sigma_p(m) \equiv j \pmod p$. The sum under consideration is equal by Luka's theorem to

$$\sum \binom{n_0}{m_0}\binom{n_1}{m_1}\dots\binom{n_d}{m_d} \pmod p,$$

where the summation is over all $m = \overline{m_d \dots m_1 m_0} \leqslant n$, for which $\sigma_p(m) \equiv j \pmod p$. This sum is equal to the sum of coefficients of x^j, x^{j+p-1}, $x^{j+2(p-1)}$, ... in the expression

$$(1+x)^{n_0}(1+x)^{n_1}\dots(1+x)^{n_d} = (1+x)^{\sigma_p(n)}.$$

But it is evident that this sum of coefficients equals

$$\sum_{\substack{1 \leqslant r \leqslant \sigma_p(n) \\ r \equiv j \pmod{p-1}}} \binom{\sigma_p(n)}{r},$$

which satisfy the induction hypothesis because $1 \leqslant \sigma_p(n) \leqslant n-1$, and supply the desired equivalence since $\sigma_p(n) \equiv n \equiv j \pmod p$.

Solution 4 (linear algebra, [D]). The polynomials x, x^2, ..., x^{p-1} are linearly independent over $\mathbb{Z}_p$ and form a basis in the space of functions $f\colon \mathbb{Z}_p \to \mathbb{Z}_p, f(0)=0$. By Fermat's little theorem $(1+x)^n \equiv (1+x)^k \pmod p$. Applying the relations $x^{i+a(p-1)} \equiv x^i$ to the left hand side, we obtain that our sum as an element of $\mathbb{Z}_p$ is equal to the coefficient of x^j in the right hand side, i. e. $\binom{k}{j}$.

2 *Arithmetical triangle and divisibility*

2.1. a) This result is due to Roberts [27]. By a_k denote the number of 1's in the first 3^k rows, and by b_k denote the number of 2's. Due to the recurrence from problem 1.3 we obtain

$$a_{k+1} = 5a_k + b_k, \qquad b_{k+1} = 5b_k + a_k.$$

Now the statement of problem follows by induction.

b) Answer: $\frac{1}{2} \cdot 5^k(5^k + 1) - 15^k$. By a_k denote the number of nonzero elements in the first 5^k rows. As in previous problem we have a recurrence

$$a_{k+1} = 15a_k + 10 \cdot \frac{5^k(5^k - 1)}{2}.$$

Since the whole triangle consists of $\frac{5^k(5^k+1)}{2}$ elements, it is natural to change variables $a_k = \frac{5^k(5^k+1)}{2} - b_k$. Then we can rewrite the previous relation in terms of b_k as $b_{k+1} = 15b_k$.

c) Answer: $\left(\frac{p(p+1)}{2}\right)^k$. This is Fine's result [13]. It can be obtained by induction by means of recurrence of the problem 1.3.

2.2. Solution 1. Induction by α_1. The base $\alpha_1 = 0, 1$ can be easily checked. Let the statement has been proven for all $\alpha_1 < a$. Prove it for $\alpha_1 = a$. Evedently $\tilde{m} - 2^{\alpha_1} < 2^{\alpha_1}$. Let $s = 2^{\alpha_1}$ (in notations of problem 1.3). Consider the $\tilde{m}$-th row in the triangle Δ_0^0, where $\tilde{m} = 2^{\alpha_2} + 2^{\alpha_3} + \ldots + 2^{\alpha_r}$. By the induction hypothesis the number of 1's in this row and above it equals

$$3^{\alpha_2} + 2 \cdot 3^{\alpha_3} + \ldots + 2^{r-2} \cdot 3^{\alpha_r}. \tag{2}$$

Then for the number $m = \tilde{m} + 2^{\alpha_1}$ we have a row that intersects the triangles Δ_0^1 and Δ_1^1 (due to 2-arithmetics they are both identical to triangle Δ_0^0). The part of Pascal triangle from top to this row contains triangle Δ_0^0 (containing 3^{α_1} 1's by induction hypothesis) and partially triangles Δ_0^1 and Δ_1^1 (the number of 1's in them is given by (2)). So the total number of 1's is

$$3^{\alpha_1} + 2(3^{\alpha_2} + 2 \cdot 3^{\alpha_3} + \ldots + 2^{r-2} \cdot 3^{\alpha_r}).$$

Solution 2 (combinatorial sense of coefficients, [T]).

Lemma 1. Let the k-th row contains 2^r 1's (or, equivalently, k contains r 1's in base 2) and let $\alpha_1 > \alpha_2 > \cdots > \alpha_m$, $2^{\alpha_m} > k$. Then the row with number $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_m} + k$ contains 2^{m+r} 1's.

Proof. It is clear that the number $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_m} + k$ in base 2 contains $m + r$ 1's and hence the corresponding row contains 2^{m+r} 1's. □

Lemma 2. The rows with the following numbers

$$2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}}, \quad 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + 1, \quad \ldots, \quad 2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + 2^{\alpha_m} - 1,$$

contain $2^k 3^{\alpha_m}$ 1's.

Proof. By lemma 1 the row with number $2^{\alpha_1} + 2^{\alpha_2} + \ldots + 2^{\alpha_{m-1}} + i$ contains $2^k x_i$ 1's, where x_i is the number of 1's in i-th row. Then the total number of 1's in these rows equals $2^k \sum x_i$. But $\sum x_i$ is the number of 1's in the first $2^{\alpha_m} - 1$ rows of Pascal triangle, this number is equal to 3^{α_m} (it is known, for example, by problem 1.4). □

The statement of problem follows from lemma 2.

2.3. The problem is from [1], the solution is from [18]. The problem statement follows from Luka's theorem due to the following observation (it is also mentioned in [1]): a binomial coefficient $\binom{n}{k}$ is odd if and only if the set of 1's in the binary expansion of k is the subset of the set of 1's in the binary expansion of n. Therefore $P_n = \sum 2^k$, where the summation is over all k described in the previous phrase. For $p = 2$ let $S_n = \{i : n_i = 1\}$ in notations of formula (1). Then

$$P_n = \sum_{I \subseteq S_n} \prod_{i \in I} 2^{2^i} = \prod_{i \in S_n} F_i.$$

2.4. This result of Fine [13] (1947) is an easy corollary of Kummer's theorem. If p does not divide $\binom{n}{k}$, then there are no carries when we add k and $n-k$ in base p. For a fixed n it means that we can choose i-th digit of k in base p by n_i+1 ways.

2.5. a) It follows from the formula proven in the previous problem because here we have a row with 2 elements only not divisible by p.

b) [13]. If Если $p^d \mid (n+1)$, then $n = \overline{a(p-1)(p-1)\dots(p-1)}$ in base p. Then for any k, $0 \leqslant k \leqslant n$, each digit of k does not exceed the corresponding digit of n. Therefore all the binomial coefficients $\binom{n_i}{k_i}$ are not equal to 0 and $\not\equiv 0 \pmod p$. By Lukas' theorem $\binom{n}{k}$ is not divisible by p.

The reverse statement. Assume that all the coefficients $\binom{n}{k}$ are not divisible by p, but n is not the number of the form $\overline{a(p-1)(p-1)\dots(p-1)}$. Therefore one of its digits, say, n_i is less than $p-1$. Choose $k = (p-1)\cdot p^i$. Then $k_i = p-1$ and hence $\binom{n_i}{k_i} = 0$, and $p \mid \binom{n}{k}$ by Lukas' theorem. A contradicition.

2.6. This problem we found in [12].

Solution 1. Assume that $\binom{n}{k-1} \not\vdots\, p$ and $\binom{n}{k} \not\vdots\, p$, but $\binom{n+1}{k} = \left(\binom{n}{k-1} + \binom{n}{k}\right) \vdots\, p$. Then $\binom{n}{k} \equiv -\binom{n}{k-1} \pmod p$. Since both binomial coefficients are not divisible by p, we can reduce the equivalence and obtain $\frac{n-k+1}{k} \equiv -1 \pmod p$. Therefore $n+1 \equiv 0 \pmod p$.

Solution 2 ([K]). Though the statement remind us the main recurrence for binomial coefficients, the part "$\binom{n}{k-1} \not\vdots\, p$" is unnecessary. Indeed, if $(n+1) \not\vdots\, p$, then $0 \leqslant n_0 \leqslant p-2$. Since $\binom{n}{k} \not\vdots\, p$, then by Kummer's theorem $k_i \leqslant n_i$ for all i But analogous inequalities hold also for the pair k and $n+1$, because n and $n+1$ have the same digits except the lower ones that differs by 1. Hence $\binom{n+1}{k} \not\vdots\, p$.

2.7. [2]. It follows from Lukas' theorem and problem 1.1.a).

2.8. The problem is from [1]. Induction by number of digits. The base is trivial. For the proof of induction step add one more digit to the rightmost position. Since the binomial coefficient is odd we have the inequalities $n_i \geqslant k_i$. Now we will use the recurrence $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ and consider distinct variants of parity n и k. Applying Kummer's theorem and the problem 4.6a) we will reduce the question to the induction hypothesis.

For example, let $n = 2\ell+1$ be odd and $k = 2m$ be even. Consider a subcase $k_1 = 1$. Then we have binary representations $k = \dots 10$, $n = \dots 11$, $k-1 = \dots 01$ and $n-k = \dots 01$ (the latter because by Kummer's theorem there are no carries when we add k and $n-k$). Now when we add $k-1$ and $n-k$ we have 1 carry, i.e. $\binom{n-1}{k-1} \equiv 2 \pmod 4$, and hence

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \equiv -\binom{n-1}{k} = -\binom{2\ell}{2m} \equiv -\binom{\ell}{m} \pmod 4,$$

the latter equivalence is by problem 4.6a). The minus sign in it corresponds to the multiplier $(-1)^{k_0 n_1 + k_1 n_0}$.

2.9. The problem is from [1]. The statement follows from the previous problem. If the binary representation of n does not contains two consecutive 1's, then for all k all the exponents $k_{i-1}n_i + k_i n_{i-1}$ are equal to 0 and all the binomial coefficients in n-th row have are equivalent 1 modulo 4. But if the binary representation of n contains several consecutive 1's starting from $n_j = 1$ then the one half of all coefficients have $k_j = 0$, and one half of them have $k_j = 1$. By the formula of previous problem these two halves differ by a sign.

2.10. Two articles in Monthly [19, 20] discuss this dark problem.

2.11. This is a problem of D. Dzhukich was presented at the olympiad of 239 school of St.-Petersburg, 2002, and after that appeared at short-list of IMO-2008.

All the binomial coefficients in the problem statement are odd by Lukas' theorem, therefore, it is sufficient to check that all the numbers $\binom{2^n-1}{1}$, $\binom{2^n-1}{3}$, $\dots$, $\binom{2^n-1}{2^n-1}$ have distinct reminders modulo 2^n.

Solution 1 ([D]). Assume by the contrary that $\binom{2^n-1}{k} \equiv \binom{2^n-1}{m} \pmod{2^n}$ for odd k and m, $k > m$. Observe that

$$\binom{2^n-1}{k} = \binom{2^n}{k} - \binom{2^n-1}{k-1} = \binom{2^n}{k} - \binom{2^n}{k-1} + \binom{2^n-1}{k-2} = \dots =$$
$$= \binom{2^n}{k} - \binom{2^n}{k-1} + \binom{2^n}{k-2} - \dots - \binom{2^n}{m+1} + \binom{2^n-1}{m}.$$

In particular

$$\binom{2^n}{k} - \binom{2^n}{k-1} + \binom{2^n}{k-2} - \ldots - \binom{2^n}{m+1} \equiv 0 \pmod{2^n}.$$

Calculate the exponent $\operatorname{ord}_2 \binom{2^n}{r}$ by Kummer's theorem. If $\operatorname{ord}_2 r = a$ then we have $n-a$ carries in addition r and $2^n - r$ (it is clear by the standard algorithm of addition), hence $\operatorname{ord}_2 \binom{2^n}{r} = n - a$. In particular $2^n \mid \binom{2^n}{r}$ for odd r, that allows us to consider only one half of summands:

$$\binom{2^n}{k-1} + \binom{2^n}{k-3} + \ldots + \binom{2^n}{m+1} \equiv 0 \pmod{2^n}.$$

Now all the $\binom{2^n}{i}$ in the left hand side have even parameter i, therefore $\operatorname{ord}_2 \binom{2^n}{x} < n$.

We will prove that this congruence is impossible and obtain a contradiction. Choose x with minimal $\operatorname{ord}_2 \binom{2^n}{x}$. Since $\operatorname{ord}_2 \binom{2^n}{x} < n$ and the whole sum is divisible by 2^n, there exists y, for which $\operatorname{ord}_2 \binom{2^n}{x} = \operatorname{ord}_2 \binom{2^n}{y}$. Then the binary representations of x and y end with equal number of 0's, and hence there exists z between x and y which binary representation ends with bigger number of 0's. Then $\operatorname{ord}_2 \binom{2^n}{z} < \operatorname{ord}_2 \binom{2^n}{x}$, a contradiction.

Solution 2 ([CSTTVZ]). Induction by n. We prove the step of induction. Let the statement be proven for all numbers less than n. Assume by the contrary that there exist k and ℓ, $k \neq \ell$, $0 \leqslant k, \ell \leqslant 2^n - 1$, such that $\binom{2^n-1}{2k+1} \equiv \binom{2^n-1}{2\ell+1} \mod 2^n$. Observe that

$$\begin{aligned}
\binom{2^n-1}{2k+1} &= \left(\frac{2^n}{1}-1\right)\left(\frac{2^n}{2}-1\right)\ldots\left(\frac{2^n}{2k+1}-1\right) = \\
&= \left(\frac{2^n}{1}-1\right)\left(\frac{2^n}{3}-1\right)\ldots\left(\frac{2^n}{2k+1}-1\right)\cdot\left(\frac{2^{n-1}}{1}-1\right)\left(\frac{2^{n-1}}{2}-1\right)\ldots\left(\frac{2^{n-1}}{k}-1\right) = \qquad (3)\\
&= \left(\frac{2^n}{1}-1\right)\left(\frac{2^n}{3}-1\right)\ldots\left(\frac{2^n}{2k+1}-1\right)\cdot\binom{2^{n-1}-1}{k} \equiv \\
&\equiv (-1)^{k+1}\binom{2^{n-1}-1}{k} \pmod{2^n}
\end{aligned}$$

and analogously $\binom{2^n-1}{2\ell+1} \equiv (-1)^{\ell+1}\binom{2^{n-1}-1}{\ell} \pmod{2^n}$. It follows by induction hypothesis that both k and ℓ can not be odd. Besides, due to the symmetry $\binom{2^n-1}{r} = \binom{2^n-1}{2^n-1-r}$ the problem statement means that all the "even" binomial coefficients $\binom{2^n-1}{2r}$ are pairwise distinct modulo 2^n and form the same set of residues as "odd" binomial coefficients $\binom{2^n-1}{2r+1}$. Therefore k and ℓ can not be even simultaneously.

It remains to consider a case when k and ℓ have distinct parity, say $k = 2a+1$, $\ell = 2b$. Then

$$\binom{2^{n-1}-1}{2a+1} + \binom{2^{n-1}-1}{2b} \equiv 0 \pmod{2^n}.$$

If $a = b$ the congruence is impossible because $\binom{2^{n-1}-1}{2a}$ is odd and

$$\binom{2^{n-1}-1}{2a+1} + \binom{2^{n-1}-1}{2a} = \binom{2^{n-1}-1}{2a}\left(1 + \frac{2^{n-1}-1-2a}{2a+1}\right) = \binom{2^{n-1}-1}{2a}\cdot\frac{2^{n-1}}{2a+1} \equiv 2^{n-1} \pmod{2^n}.$$

If $b \neq a$, then $\binom{2^{n-1}-1}{2a} \neq \binom{2^{n-1}-1}{2b}$ by the induction hypothesis, Since $\binom{2^{n-1}-1}{2a} + \binom{2^{n-1}-1}{2a+1}$ is divisible by 2^{n-1}, the sum $\binom{2^{n-1}-1}{2b} + \binom{2^{n-1}-1}{2a+1}$ can not be divisible by 2^{n-1}.

2.12. The author of this problem is A. Belov. Observe that

$$\binom{2n}{n+k} = \binom{2n}{n}\cdot\frac{n(n-1)\ldots(n-k+1)}{(n+1)(n+2)\ldots(n+k)},$$

and therefore $\binom{2n}{n+k}$ have many common divisors with $\binom{2n}{n}$, because the denominator is not very big, more precisely, it does not exceed $(2n)^k$. Write the analogous equalities for all binomial coefficients $\binom{2n}{n+k_1}$,

$\binom{2n}{n+k_2}, \ldots, \binom{2n}{n+k_{100}}$. Then GCD of all denominators in the right hand sides of the equalities does not exceed $(n+1)(n+2)\ldots\left(n+[\varepsilon\sqrt{n}\,]\right) < (2n)^{\varepsilon\sqrt{n}}$. But for big n the binomial coefficient $\binom{2n}{n}$ is much greater, so after reducing by GCD the quotient is very big, and it divides all 100 binomial coefficients.

Explain more accurate the last reasoning. Observe that

$$\binom{2n}{n} = \frac{2n}{n}\cdot\frac{2n-1}{n-1}\ldots\frac{n+1}{1} > 2^n \qquad \text{and} \qquad (2n)^{100\varepsilon\sqrt{n}} = 2^{\varepsilon\sqrt{n}\log_2 n + \varepsilon\sqrt{n}}.$$

For each ε there exists N such that for all $n > N$ we have the equality $\frac{n}{2} > \varepsilon\sqrt{n}\log_2 n + \varepsilon\sqrt{n}$. If we reduce $\binom{2n}{n}$ by GCD for these n, the quotient is at least $2^{n/2}$.

2.13. a) The problem was presented at Leningrad olympiad, 1977.

Solution 1 (without Kummer's theorem). This is solution from the excellent book [4]. Assume that all these numbers are divisible by m. Then the numbers

$$\begin{gathered}\binom{n+k-1}{k-1} = \binom{n+k}{k} - \binom{n+k-1}{k},\\ \binom{n+k-2}{k-1} = \binom{n+k-1}{k} - \binom{n+k-2}{k},\\ \ldots\\ \binom{n}{k-1} = \binom{n+1}{k} - \binom{n}{k}\end{gathered}$$

are also divisible by m. Then analogously m divides all the numbers $\binom{n+i}{j}$, where $i \leqslant j$ are arbitrary nonnegative integers. But $\binom{n}{0}$ $(i = j = 0)$ is not divisible by m. A contradiction.

Solution 2 (Kummer's theorem). Let p be a prime divisor of m. Prove that at least one of the numbers $\binom{n}{k}, \binom{n+1}{k}, \ldots, \binom{n+k}{k}$ is not divisible by p. By Kummer's theorem if we choose ℓ $(n-k \leqslant \ell \leqslant n)$ such that the addition $k+\ell$ fulfills in base p without carries then the binomial coefficient $\binom{k+\ell}{k}$ is not divisible by p.

We will explain how to choose ℓ by giving a concrete example. Let $p = 7$, $k = 133$. We will write all the numbers in base 7. Since we try to choose ℓ in the set of $k+1$ numbers, we can always choose ℓ such that $k+\ell$ to be one of the following numbers

$$\ldots 133, \quad \ldots 233, \quad ,\ldots, \quad \ldots 633.$$

(Remind that 6 is the greatest digit in our example.) It is clear that the addition $k+\ell$ fulfills without carries.

b) We found this problem in [2]. It is not difficult to construct n by Kummer's theorem. Let $\operatorname{ord}_p m = s$, and k have $d+1$ digits in base p. Let $n \vdots p^{d+s+1}$. Then the representations of numbers $n-k$, $n-k+1$, $\ldots$, $n-1$ contain digits $(p-1)$ in positions from $(d+2)$ to $(d+s+2)$. When we add k to these numbers we have carries in these positions. Therefore by Kummer's theorem all the corresponding binomial coefficients are divisible by p^s.

Since it is not difficult to combine our reasoning for distinct p, the statemetn is proven.

3 Generalizations of Wilson's and Lukas' theorems

3.1. It is well known that $\operatorname{ord}_p(n!) = \sum_k \left[\frac{n}{p^k}\right]$. If $n = n_d p^d + n_{d-1}p^{d-1} + \ldots + n_1 p + n_0$ (representation in base p), then $\left[\frac{n}{p^k}\right] = n_d p^{d-k} + n_{d-1}p^{d-k-1} + \ldots + n_{k+1}p + n_k$ and we can rewrite the formula for $\operatorname{ord}_p(n!)$ in the form

$$\operatorname{ord}_p(n!) = \sum_{k=1}^{d}\left(\sum_{i=k}^{d} n_i p^{i-k}\right) = \sum_{i=1}^{d} n_i(p^{i-1}+p^{i-2}+\ldots+p+1) = \sum_{i=1}^{d} n_i\frac{p^i-1}{p-1} = \frac{\sum\limits_{i=0}^{d} n_i p^i - \sum\limits_{i=0}^{d} n_i}{p-1}.$$

This is exactly what we need.

3.2. a) Split the factors of $n!$ on groups of $(p-1)$ factors:

$$(n!)_p = \prod_{k=0}^{[\frac{n}{p}]-1}\big((kp+1)\cdot(kp+2)\cdots(kp+p-1)\big)\ \cdot\ \big([\tfrac{n}{p}]p+1\big)\big([\tfrac{n}{p}]p+2\big)\ldots\big([\tfrac{n}{p}]p+n_0\big)\equiv(-1)^{[\frac{n}{p}]}n_0!\pmod p\,.$$

б) This statement can be found in Gauss works [15]. The product $(p^q!)_p$ contains factors in pairs: a factor and its inverse modulo p^q, the product of each pair is 1 modulo p^q. So we need to watch on those factors m which equals to its inverse, this factors satisfy the congruence

$$m^2\equiv 1\pmod{p^q}.$$

For odd prime p the congruence has 2 solutions: ±1. For $p=2$, $q\geqslant 3$ the congruence has two more solutions: $2^{q-1}\pm 1$.

c) Since $n!=(n!)_p\cdot p^{[\frac{n}{p}]}\big([\frac{n}{p}]\big)!$, the statement can be proven by induction by means of the congruence of statement a) of this problem.

3.3. We found this problem on the web-page of A.Granville [17]. It well known Legendre's formula for the number ℓ is that

$$\ell=\operatorname{ord}_p\binom{n}{k}=\left(\left[\frac{n}{p}\right]-\left[\frac{k}{p}\right]-\left[\frac{r}{p}\right]\right)+\left(\left[\frac{n}{p^2}\right]-\left[\frac{k}{p^2}\right]-\left[\frac{r}{p^2}\right]\right)+\ldots \tag{4}$$

Denote $\tilde n=[n/p]$ for brevity and so forth, and collect all terms divisible by p in the the formula for a binomial coefficient:

$$\binom{n}{k}=\frac{(n!)_p}{(k!)_p(r!)_p}\cdot\frac{p^{[n/p]}}{p^{[k/p]}\cdot p^{[r/p]}}\cdot\frac{\tilde n!}{\tilde k!\cdot\tilde r!}.$$

By generalized Wilson's theorem (problem 3.2, b) the first fraction equals $\pm\frac{n_0!}{k_0!r_0!}\pmod p$, the third fraction allows us to apply induction, and the middle fraction (together with the sign of the first fraction) supply all the expressions containing ℓ by the formula (4).

3.4. a) Expand brackets in $(1+x)^{p^d}$ use the fact that $p\mid\binom{p^d}{k}$ for $1\leqslant k\leqslant p^d-1$ by Kummer's theorem.

b) Let $n=n'p+n_0$, $k=k'p+k_0$. By the previous statement $(1+x)^{pn'}\equiv(1+x^p)^{n'}\pmod p$. Then

$$(1+x)^n=(1+x)^{pn'}(1+x)^{n_0}\equiv(1+x^p)^{n'}(1+x)^{n_0}\pmod p.$$

This congruence means that we transform the coefficients of the polynomial modulo p. The coefficient of x^k at the l.h.s. equals $\binom{n}{k}$. All the exponents in the first brackets at the r.h.s. are divisible by p, hence the only way to obtain the term $x^{pk'+k_0}$ is multiplying the $x^{pk'}$ from the first bracket and x^{k_0} from the second. Thus we obtain $\binom{n'}{k'}\binom{n_0}{k_0}$ and so $\binom{n}{k}=\binom{n'}{k'}\binom{n_0}{k_0}$. Now Lukas' theorem follows by induction.

3.5. a, b) It follows from Kummer's theorem.

3.6. [9]. In the following calculation we use that $\binom{n_i}{k_i}=0$ for $k_i>n_i$; this allows us to apply Lukas' theorem and truncate a lot of summands:

$$f_{n,a}=\sum_{k=0}^{n}\binom{n}{k}^a\equiv\sum_{k_d=0}^{n_d}\sum_{k_{d-1}=0}^{n_{d-1}}\cdots\sum_{k_0=0}^{n_0}\prod_{i=0}^{d}\binom{n_i}{k_i}^a\equiv\prod_{i=0}^{d}\sum_{k_i=0}^{n_i}\binom{n_i}{k_i}^a\equiv\prod_{i=0}^{d}f_{n_i,a}\pmod p.$$

4 Variations on Wolstenholme's theorem

4.1. This is an exercise on reading an article. The statement is proven in article [1]. Observe that

$$2\sum_{i=1}^{p-1}\frac{1}{i}=\sum_{i=1}^{p-1}\frac{1}{i}+\frac{1}{p-i}=p\sum_{i=1}^{p-1}\frac{1}{i(p-i)}\,.$$

Hence the sum under consideration is divisible by p. Since $\frac{1}{i} \equiv -\frac{1}{p-i} \pmod p$, it remains to check that

$$\sum_{i=1}^{p-1} \frac{1}{i^2} \equiv 0 \pmod p.$$

But $\frac{1}{1^2}, \frac{1}{2^2}, \ldots, \frac{1}{(p-1)^2}$ modulo p is the same set as[1], что $1^2, 2^2, \ldots, (p-1)^2$. Therefore it is sufficient to prove that

$$\sum_{i=1}^{p-1} i^2 \equiv 0 \pmod p. \tag{5}$$

Let $\sum\limits_{i=1}^{p-1} i^2 \equiv s \pmod p$. It $p > 5$ we can always choose a, such that $a^2 \not\equiv 1 \pmod p$. Then the sets $\{1, 2, \ldots, p-1\}$ and $\{a, 2a, \ldots, (p-1)a\}$ coincide (the proof is the same as in the footnote) and

$$s \equiv \sum_{i=1}^{p-1} i^2 = \sum_{i=1}^{p-1} (ai)^2 = a^2 \sum_{i=1}^{p-1} i^2 \equiv a^2 s \pmod p.$$

Thus $s \equiv 0 \pmod p$.

4.2. Answer: $2k+2$. This problem of A. Golovanov was presented at Tuimaada-2012 olympiad. Observe that for $p = 4k+3$ the equation $x^2 + 1 = 0$ has no solutions in the set of residues modulo p, and hence the denominators of all fractions are non zero.

Solution 1. Let $a_i = i^2 + 1$, $i = 0, \ldots, p-1$. Then the expression equals

$$\frac{\sigma_{p-1}(a_0, a_1, \ldots, a_{p-1})}{\sigma_p(a_0, a_1, \ldots, a_{p-1})},$$

where σ_i is an elementary symmetrical polynomial of degree i. Find the polynomial for which the numbers a_i are its roots:

$$\prod_{i=0}^{p-1} (x - 1 - i^2).$$

Change the variable $x - 1 = t^2$ and obtain

$$\prod_{i=0}^{p-1} (t^2 - i^2) = \prod_{i=0}^{p-1} (t - i) \prod_{i=0}^{p-1} (t + i) \equiv (t^p - t)(t^p - t) = t^{2p} - 2t^{p+1} + t^2.$$

Now apply the inverse change of variables and obtain for $p = 4k+3$

$$\prod_{i=0}^{p-1} (x - 1 - i^2) \equiv (x-1)^p - 2(x-1)^{\frac{p+1}{2}} + (x-1) = x^p + \ldots + (p + 2 \cdot \tfrac{p+1}{2} + 1)x - 4.$$

By Viete's theorem $\sigma_p \equiv 4 \pmod p$, $\sigma_{p-1} \equiv 2 \pmod p$, therefore $\frac{\sigma_{p-1}}{\sigma_p} \equiv \frac{1}{2} \equiv 2k+2 \pmod p$.

Solution 2. Split all nonzero residues modulo p, except ± 1, on pairs of reciprocal. We obtain $2k$ pairs and in each pair (i, j)

$$ij \equiv 1 \quad \Leftrightarrow \quad i^2 j^2 \equiv 1 \quad \Leftrightarrow \quad (ij)^2 + i^2 + j^2 + 1 \equiv i^2 + j^2 + 2 \pmod p.$$

Therefore,

$$1 \equiv \frac{(ij)^2 + i^2 + j^2 + 1}{(i^2+1)(j^2+1)} \equiv \frac{i^2 + j^2 + 2}{(i^2+1)(j^2+1)} = \frac{1}{i^2+1} + \frac{1}{j^2+1} \pmod p.$$

So, the sum is equal to $\frac{1}{0^2+1} + \frac{1}{1^2+1} + \frac{1}{(-1)^2+1} + 2k \equiv 2k+2$.

[1] These sets coincide because they contain $p-1$ element each, and it is clear that all the reminders in each set are non zero and pairwise distinct.

Solution 3. By Fermat's little theorem the operations $x \mapsto x^{-1}$ and $x \mapsto x^{p-2}$ modulo p coincide. So it is sufficient to calculate the sum

$$\sum_{x=0}^{p-1}(x^2+1)^{p-2} = \sum_{x=0}^{p-1}\sum_{m=0}^{p-2}\binom{p-2}{m}x^{2m} = \sum_{m=0}^{p-2}\binom{p-2}{m}S_{2m}, \tag{6}$$

where $S_{2m} = \sum\limits_{x=0}^{p-1} x^{2m}$. Evidently $S_{2m} \equiv -1 \pmod p$ for $m = \frac{p-1}{2}$. Prove that $S_{2m} \equiv 0 \pmod p$ for all other $m \leqslant p-1$. Indeed, for each m we can choose a non zero residue a such that $a^{2m} \not\equiv 1 \pmod p$ and after that we can reason as in (5). For the sum (6) we have

$$\sum_{m=0}^{p-2}\binom{p-2}{m}S_{2m} \equiv -\binom{p-2}{\frac{p-1}{2}} = -\binom{4k+1}{2k+1} = -\frac{(4k+1)\cdot 4k\cdot\ldots\cdot(2k+1)}{1\cdot 2\cdot\ldots\cdot(2k+1)} \equiv$$
$$\equiv -\frac{(-2)\cdot(-3)\ldots(2k+2)}{1\cdot 2\cdot\ldots\cdot(2k+1)} \equiv 2k+2 \pmod p.$$

4.3. We found these statements in [16].

a) For each prime divisor $p \mid m$ choose a_p such that $p \nmid (a_p^k - 1)$. By the Chinese reminder theorem choose a such that $a \equiv a_p \pmod p$ for all p. Then the result can be proven by reasoning as in (5).

b) Observe that for odd k by the binomial formula we have $i^k + (p-i^k) \equiv ki^{k-1}p \pmod{p^2}$. Then

$$2\sum_{i=1}^{p-1}\frac{1}{i^k} = \sum_{i=1}^{p-1}\left(\frac{1}{i^k}+\frac{1}{(p-i)^k}\right) = \sum_{i=1}^{p-1}\frac{i^k+(p-i)^k}{i^k(p-i)^k} \equiv \sum_{i=1}^{p-1}\frac{ki^{k-1}p}{i^k(-i)^k} \equiv -kp\sum_{i=1}^{p-1}\frac{1}{i^{k+1}} \pmod{p^2}.$$

The sum in the r.h.s is divisible by p by the statement a).

4.4. The congruence holds even modulo p^7 (see [24]), but it goes a bit strong. We can reason as in [1], tracing all powers till p^4, and obtain

$$\binom{p-1}{2p-1} = \frac{(2p-1)(2p-2)\cdot\ldots\cdot(p+1)}{p!} = \left(\frac{2p}{1}-1\right)\left(\frac{2p}{2}-1\right)\cdot\ldots\cdot\left(\frac{2p}{p-1}-1\right) \equiv$$
$$\equiv 1 - 2p\sum_{i=1}^{p-1}\frac{1}{i} + 4p^2\sum_{\substack{i,j=1\\ i<j}}^{p-1}\frac{1}{ij} - 8p^3\sum_{\substack{i,j,k=1\\ i<j<k}}^{p-1}\frac{1}{ijk} \pmod{p^4}. \tag{7}$$

The last sum can be expressed via power sums:

$$\sum_{\substack{i,j,k=1\\ i<j<k}}^{p-1}\frac{1}{ijk} = \frac{S_3}{3} - \frac{S_1S_2}{2} + \frac{S_1^3}{6}, \quad \text{where } S_k = \sum_{i=1}^{p-1}\frac{1}{i^k}.$$

We now that S_1 and S_3 are divisible by p^2 (the latter due to problem 4.3b). Therefore the last term in the formula (7) can be omitted.

4.5. The problem is from [1], variations can be found in [14]. Since

$$2\sum_{k=1}^{p-1}\frac{1}{k^2} = \sum_{k=1}^{p-1}\Big(\frac{1}{k^2}+\frac{1}{(p-k)^2}\Big) = \sum_{k=1}^{p-1}\frac{k^2+(p-k)^2}{k^2(p-k)^2} \equiv -2\sum_{k=1}^{p-1}\frac{1}{k(p-k)} \pmod{p^2},$$

the statement 3) is equivalent to the congruence $\sum\limits_{k=1}^{p-1}\frac{1}{k(p-k)} \equiv 0 \pmod{p^2}$. The statement 2) is equivalent to the same congruence, because $2\sum\limits_{k=1}^{p-1}\frac{1}{k} = 2\sum\limits_{k=1}^{p-1}\Big(\frac{1}{k}+\frac{1}{p-k}\Big) = 2p\sum\limits_{k=1}^{p-1}\frac{1}{k(p-k)}$. Finally we know from the previous problem that

$$\binom{2p-1}{p-1} \equiv 1 - p^2\sum_{i=1}^{p-1}\frac{1}{i(p-i)} + 4p^2\sum_{\substack{i,j=1\\ i<j}}^{p-1}\frac{1}{ij} \pmod{p^4}.$$

So the statement 1) is equivalent to the congruence

$$\sum_{i=1}^{p-1} \frac{1}{i(p-i)} \equiv 4 \sum_{\substack{i,j=1 \\ i<j}}^{p-1} \frac{1}{ij} \pmod{p^2}. \tag{8}$$

Rewrite the expression in the r.h.s.:

$$4 \sum_{\substack{i,j=1 \\ i<j}}^{p-1} \frac{1}{ij} = 2\Big(\sum_{k=1}^{p-1} \frac{1}{i}\Big)^2 - 2\sum_{k=1}^{p-1} \frac{1}{k^2} \equiv 2\Big(\sum_{k=1}^{p-1} \frac{1}{i}\Big)^2 + 2\sum_{k=1}^{p-1} \frac{1}{k(p-k)}.$$

The sum in brackets is divisible by p, its square is divisible by p^2, and we can omit this term. Then from (8) we see that the statement 1) is equivalent to the congruence $\sum_{k=1}^{p-1} \frac{1}{k(p-k)} \equiv 0 \pmod{p^2}$.

4.6. a) Solution 1 ([5, proposition 2.12]). Induction on n. Expand brackets in the equality

$$(a+b)^{pn} = (a+b)^{p(n-1)}(a+b)^p$$

Equate the coefficients of $a^{pm}b^{p(n-m)}$:

$$\binom{pn}{pm} = \binom{p(n-1)}{pm}\binom{p}{0} + \binom{p(n-1)}{pm-1}\binom{p}{1} + \ldots + \binom{p(n-1)}{pm-p+1}\binom{p}{p-1} + \binom{p(n-1)}{pm-p}\binom{p}{p}.$$

All summands except first and last are divisible by p^2, because by Lucas' theorem each binomial coefficient is divisible by p. Hence

$$\binom{pn}{pm} \equiv \binom{p(n-1)}{pm} + \binom{p(n-1)}{p(m-1)} \pmod{p^2}.$$

By the induction hypothesis

$$\binom{p(n-1)}{pm} + \binom{p(n-1)}{p(m-1)} \equiv \binom{n-1}{m} + \binom{n-1}{m-1} \equiv \binom{n}{m} \pmod{p^2}.$$

Solution 2 ([D]). Prove that $\binom{kp}{mp} \equiv \binom{k}{m} \pmod{p^2}$ by induction on m.

To prove the base $m = 1$ we have to check that $\binom{pk}{p} - \binom{k}{1} \equiv 0 \pmod{p^2}$. We have

$$\binom{pk}{p} - \binom{k}{1} = \frac{pk(pk-1)\ldots(pk-p+1)}{p!} - k = \Big(\frac{(pk-1)(pk-1)\ldots(pk-p+1)}{(p-1)!} - 1\Big). \tag{9}$$

Split the multipliers in the numerator onto pairs:

$$(pk-i)(pk-p+i) \equiv pi^2 - i^2 \pmod{p^2}.$$

We see that the product modulo p^2 of each pair does not depend on k. Therefore the difference (9) modulo p^2 does not depend on k, too. Since it is equal to 0 for $k = 1$, it is equal to 0 for all k.

The step of induction. Let $\binom{kp}{(m-1)p} \equiv \binom{k}{m-1} \pmod{p^2}$. We have

$$\begin{aligned}\binom{kp}{mp} &= \binom{kp}{(m-1)p} \cdot \frac{(p(k-m)+1)(p(k-m)+1)\ldots(p(k-m)+p)}{pm(pm-1)\ldots(pm-p+1)} = \\ &= \binom{kp}{(m-1)p} \cdot \frac{(p(k-m)+1)(p(k-m)+1)\ldots(p(k-m)+p-1)}{(pm-1)\ldots(pm-p+1)} \cdot \frac{k-m+1}{m}. \end{aligned} \tag{10}$$

Remark that both fractions are correctly defined modulo p^2. As in the proof of base, the expression in the numerator of big fraction does not depend (modulo p^2) on k. Then we can put $k = 0$ for the calculating

the fraction modulo p^2 and obtain that it is congruent to 0. For the remaining part of the expression we can apply the induction hypothesis and obtain

$$\equiv \binom{k}{m-1} \cdot \frac{k-m+1}{m} = \binom{k}{m} \pmod{p^2}.$$

b) Solution 1 (combinatorial). As it has been suggested in [1], consider samples of kp objects from the set of pn objects. Let the initial set be split on blocks of p objects. The number of block samples equals $\binom{n}{k}$. Hence it remains to check that non block samples is divisible by p^3. But the number of non block samples with 3 or more blocks is divisible by p^3 (see [1]). For $k > 1$ every non block sample consists of at least 3 blocks, so in this case the statement is true. It remains to consider a case when $k = 1$ and we count the number of non block samples of p objects from the set of $2p$ objects. This number equals $\binom{2p}{p} - 2$, by Wolstenholme's theorem it is divisible by p^3.

Solution 2. In the formula $\binom{a}{b} = \frac{a(a-1)\dots(a-b+1)}{b(b-1)\dots1}$ split the numerator and the denominator onto blocks of p terms, reduce the first terms in each block, and collect the quotients in a separate expression:

$$\binom{mp}{kp} = \frac{m\not{p}\cdot(mp-1)\dots\big(mp-(p-1)\big)}{k\not{p}\cdot(kp-1)\dots\big(kp-(p-1)\big)} \cdot \frac{(m-1)\not{p}\cdot\big((m-1)p-1\big)\dots\big((m-1)p-(p-1)\big)}{(k-1)\not{p}\cdot\big((k-1)p-1\big)\dots\big((k-1)p-(p-1)\big)} \cdot \dots \times$$
$$\times \frac{(m-k+1)\not{p}\cdot\big((m-k+1)p-1\big)\dots\big((m-k+1)p-(p-1)\big)}{\not{p}\cdot(p-1)\dots1} =$$
$$= \binom{m}{k} \cdot \frac{(mp-1)\dots\big(mp-(p-1)\big)}{(kp-1)\dots\big(kp-(p-1)\big)} \cdot \dots \cdot \frac{\big((m-k+1)p-1\big)\dots\big((m-k+1)p-(p-1)\big)}{(p-1)\dots1}.$$

It remains to check that the product of fractions is congruent to 1 $\pmod{p^3}$. For this prove the congruence

$$\frac{(np-1)\dots\big(np-(p-1)\big)}{(rp-1)\dots\big(rp-(p-1)\big)} \equiv 1 \pmod{p^3}$$

or, even, it would be better to prove the following congruence

$$\frac{(np-1)\dots\big(np-(p-1)\big)}{(p-1)!} \equiv \frac{(rp-1)\dots\big(rp-(p-1)\big)}{(p-1)!} \pmod{p^3}.$$

This is true because both parts are congruent to 1 $\pmod{p^3}$, that can be shown analogously to the proof of Wolstenholme's theorem.

4.7. a) [5, theorem 2.14]. Transform the difference

$$\binom{p^2}{p} - \binom{p}{1} = \frac{p^2(p^2-1)\dots(p^2-(p-1))}{1\cdot2\cdot\dots\cdot(p-1)p} - p = \frac{p}{(p-1)!}\Big((1-p^2)(2-p^2)\dots((p-1)-p^2) - 1\cdot2\cdot\dots\cdot(p-1)\Big).$$

It remains to check that

$$(1-p^2)(2-p^2)\dots((p-1)-p^2) \equiv 1\cdot2\cdot\dots\cdot(p-1) \pmod{p^4}.$$

Expand brackets in the l.h.s.:

$$(1-p^2)(2-p^2)\dots((p-1)-p^2) = 1\cdot2\cdot\dots\cdot(p-1) + p^2\Big(1+\frac{1}{2}+\dots+\frac{1}{p-1}\Big)(p-1)! + \text{terms divisible by } p^4.$$

By the problem 4.1 the second summand is divisible by p^4.

b) Observe that $\binom{p^{s+1}}{p} = p^s \cdot \binom{p^{s+1}-1}{p-1}$, hence it is sufficient to prove that $\binom{p^{s+1}-1}{p-1} \equiv 1 \pmod{p^{s+3}}$.

$$\binom{p^{s+1}-1}{p-1} = \frac{(p^{s+1}-1)(p^{s+1}-2)\dots(p^{s+1}-(p-1))}{1\cdot2\cdots(p-1)} = \left(\frac{p^{s+1}}{1}-1\right)\left(\frac{p^{s+1}}{2}-1\right)\dots\left(\frac{p^{s+1}}{p-1}-1\right) \equiv$$
$$\equiv (-1)^{p-1} + p^{s+1}\Big(1+\frac{1}{2}+\dots+\frac{1}{p-1}\Big) \pmod{p^{s+3}}.$$

Since $(-1)^{p-1} = 1$ and $1 + \frac{1}{2} + \dots + \frac{1}{p-1} \equiv 0 \pmod{p^2}$ we are done.

4.8. The problem is from [1], we present solution [T].

$$\binom{p^3}{p^2}-\binom{p^2}{p}=p\left(\binom{p^3-1}{p^2-1}-\binom{p^2-1}{p-1}\right)=$$
$$=p\left(\left(\frac{p^3}{1}-1\right)\left(\frac{p^3}{2}-1\right)\dots\left(\frac{p^3}{p^2-1}-1\right)-\left(\frac{p^2}{1}-1\right)\left(\frac{p^2}{2}-1\right)\dots\left(\frac{p^2}{p-1}-1\right)\right)=$$
$$=p\left(\frac{p^2}{1}-1\right)\left(\frac{p^2}{2}-1\right)\dots\left(\frac{p^2}{p-1}-1\right)\left(\prod_{\substack{k=1\\ p\nmid k}}^{p^2-1}\left(\frac{p^3}{k}-1\right)-1\right).$$

It is sufficient to prove that the last bracket is divisible by p^7. Transform the product:

$$\prod_{\substack{k=1\\ p\nmid k}}^{p^2-1}\left(\frac{p^3}{k}-1\right)=\prod_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\left(\frac{p^3}{k}-1\right)\left(\frac{p^3}{p^2-k}-1\right)=\prod_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\left(\frac{p^6-p^5}{k(p^2-k)}+1\right)\equiv 1+p^5(p-1)\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k(p^2-k)}\pmod{p^7}.$$

Now we have to check that the last sum is divisible by p^2. This is true because by problem 4.3a)

$$\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k(p^2-k)}\equiv-\sum_{\substack{k=1\\ p\nmid k}}^{\frac{p^2-1}{2}}\frac{1}{k^2}\equiv 0\pmod{p^2}.$$

4.9. The statement is taken from [6, theorem 5], its generalization can be found in [7].

Solution 1 ([5, proposition 2.19]). Use the fact that the difference $\binom{2^{k+1}}{2^k}-\binom{2^k}{2^{k-1}}$ is equal to the coefficient of x^{2^k} in the polynomial

$$(1+x)^{2^{k+1}}-(1-x^2)^{2^k}=(1+x)^{2^k}\left((1+x)^{2^k}-(1-x)^{2^k}\right)=$$
$$=\left(1+\binom{2^k}{1}x+\binom{2^k}{2}x^2+\ldots+x^{2^k}\right)\cdot 2\left(\binom{2^k}{1}x+\binom{2^k}{3}x^3+\ldots+\binom{2^k}{2^k-1}x^{2^k-1}\right).$$

Since the second polynomial contains odd exponents only, the coefficient of x^{2^k} in the product equals

$$2\left(\binom{2^k}{1}\binom{2^k}{2^k-1}+\binom{2^k}{3}\binom{2^k}{2^k-3}+\ldots+\binom{2^k}{2^k-1}\binom{2^k}{1}\right).$$

By problem 3.5 b) 2^k divides each binomial coefficient in this expression, moreover each term occurs twice in the sum, and the sum itself is multiplied by 2. Thus all the expression is divisible by 2^{2k+2}.

Solution 2 ([CSTTVZ]). Since $\binom{2^{n+1}}{2^n}=2\binom{2^{n+1}-1}{2^n-1}$, it is sufficient to prove that

$$\binom{2^{n+1}-1}{2^n-1}\equiv\binom{2^n-1}{2^{n-1}-1}\pmod{2^{2n+1}}.$$

Similarly to (3) we obtain

$$\binom{2^{n+1}-1}{2^n-1}=\left(\frac{2^{n+1}}{1}-1\right)\left(\frac{2^{n+1}}{3}-1\right)\dots\left(\frac{2^{n+1}}{2^n-1}-1\right)\cdot\binom{2^n-1}{2^{n-1}-1}.$$

It is sufficient to prove that

$$L=\left(\frac{2^{n+1}}{1}-1\right)\left(\frac{2^{n+1}}{3}-1\right)\dots\left(\frac{2^{n+1}}{2^n-1}-1\right)\equiv 1\pmod{2^{2n+1}}.$$

This is true because

$$L \equiv (-1)^{2^{n-1}} - 2^{n+1}\left(\frac{1}{1}+\frac{1}{3}+\frac{1}{5}+\ldots+\frac{1}{2^n-1}\right) \equiv$$
$$\equiv 1 - 2^{n+1}\left(\frac{2^n}{1\cdot(2^n-1)}+\frac{2^n}{3\cdot(2^n-3)}+\ldots+\frac{2^n}{(2^{n-1}-1)(2^{n-1}+1)}\right) \equiv 1 \pmod{2^{2n+1}}.$$

4.10. This is theorem of Morley [26].

Solution 1 (author's proof, 1895). It goes a bit beyond the school curriculum.

Take the formula which expresses $\cos^{2n+1} x$ via cosines of multiple angles,[1] or, as they were saying in that times, write $\cos^{2n+1} x$ in the form handy for integrating:

$$2^{2n}\cos^{2n+1}x = \cos(2n+1)x + (2n+1)\cos(2n-1)x + \frac{(2n+1)\cdot 2n}{1\cdot 2}\cos(2n-3)x + \ldots + \frac{(2n+1)\cdot 2n\ldots(n+2)}{n!}\cos x.$$

Now integrate it[2] over the interval $[0, \frac{\pi}{2}]$:

$$2^{2n}\int \cos^{2n+1} x\,dx = \frac{\sin(2n+1)x}{2n+1} + \frac{2n+1}{2n-1}\sin(2n-1)x + \ldots,$$
$$2^{2n}\int_0^{\pi/2} \cos^{2n+1} x\,dx = (-)^n\left(\frac{1}{2n+1} - \frac{2n+1}{2n-1} + \ldots\right).$$

Every first grade student of university knows that it is convenient to use integration by parts for calculating this integral:

$$I_{2n+1} = \int_0^{\pi/2} \cos^{2n+1} x\,dx = \int_0^{\pi/2} \cos^{2n} x\cos x\,dx = \cos^{2n} x\sin x\Big|_0^{\pi/2} + 2n\int_0^{\pi/2} \cos^{2n-1} x\sin^2 x\,dx =$$
$$= 0 + 2n\int_0^{\pi/2} \cos^{2n-1} x(1-\cos^2 x)\,dx = 2n\cdot I_{2n-1} - 2n\cdot I_{2n+1},$$

therefore $I_{2n+1} = \frac{2n}{2n+1}\cdot I_{2n-1}$. Since $I_1 = 1$, we can apply the formula n times and obtain

$$\int_0^{\pi/2} \cos^{2n+1} x\,dx = \frac{2n\cdot(2n-2)\ldots 2}{(2n+1)(2n-1)\ldots 3}.$$

Equating of these two results give us the formula

$$2^{2n}\frac{2n\cdot(2n-2)\ldots 2}{(2n+1)(2n-1)\ldots 3} = (-)^n\left(\frac{1}{2n+1} - \frac{2n+1}{2n-1} + \ldots + \frac{(2n+1)\cdot 2n\ldots(n+2)}{n!}\right).$$

Let $p = 2n+1$ be a prime number. We obtain the desired congruence by multiplying the last formula by p:

$$2^{2n}\frac{2n\cdot(2n-2)\ldots 2}{(2n-1)(2n-3)\ldots 3} \equiv (-)^n \pmod{p^2}.$$

Solution 2 ([CSTTVZ]). We will use the following notations:

$$A = \sum_{i=1}^{\frac{p-1}{2}} \frac{1}{i}, \qquad B = \sum_{1\leqslant i<j\leqslant\frac{p-1}{2}} \frac{1}{ij}, \qquad C = \sum_{\substack{1\leqslant i\leqslant p-1\\ i \text{ is odd}}} \frac{1}{i}.$$

[1] The reader who is interested in question "from where do we take it" and not satisfied by the answer "from some text-book" may wish to use the Euler's formula $\cos\varphi = \frac{1}{2}(e^{i\varphi} + e^{-i\varphi})$ and raise its r.h.s in power $2n+1$ by the binomial formula.

[2] When we were learning the rules of multiplication, we just memorized that "minus by minus equals plus". In this formula we multiply signs. If we need to multiply n minuses, the record $(-)^n$ seems to be appropriate. So we leave the old-fashioned notation $(-)^n$, used by the author, instead of the modern one $(-1)^n$.

Then $A^2 = \sum\limits_{i=1}^{\frac{p-1}{2}} \frac{1}{i^2} + 2B \equiv 2B \pmod p$ by the problem 4.3b). So $A^2 \equiv 2B \pmod p$. Further,

$$2C + A = \sum_{\substack{1\leqslant i\leqslant p-1\\ i \text{ нечетно}}} \frac{2}{i} + \sum_{i=1}^{\frac{p-1}{2}} \frac{2}{2i} = \sum_{i=1}^{p-1} \frac{2}{i} \equiv 0 \pmod{p^2}.$$

So $C \equiv -\frac{1}{2}A \pmod{p^2}$.

Now transform modulo p^3 the parts of the given congruence. The l.h.s. is

$$(-1)^{\frac{p-1}{2}} \binom{p-1}{\frac{p-1}{2}} \equiv \Big(1-\frac{p}{1}\Big)\Big(1-\frac{p}{2}\Big)\dots\Big(1-\frac{p}{\frac{p-1}{2}}\Big) \equiv 1 - pA + p^2B \equiv 1 - pA + \frac{1}{2}p^2A^2 \pmod{p^3}.$$

For transforming the r.h.d observe that

$$2^{p-1} = \frac{2\cdot 4\cdots(p-1)}{1\cdot 2\cdots\frac{p-1}{2}} \cdot \frac{(p+1)\cdots(2p-2)}{\frac{p+1}{2}\cdots(p-1)} = \frac{(p+1)\cdots(2p-2)}{1\cdot 3\cdot 5\cdots(p-2)} =$$
$$= \Big(\frac{p}{1}+1\Big)\Big(\frac{p}{3}+1\Big)\dots\Big(\frac{p}{p-1}+1\Big) \equiv 1 + pC + \frac{1}{2}p^2C^2 \equiv 1 - \frac{1}{2}pA + \frac{1}{8}p^2A^2 \pmod{p^3}.$$

Then we have

$$4^{p-1} \equiv \Big(1 - \frac{1}{2}pA + \frac{1}{8}p^2A^2\Big)^2 \equiv 1 - pA + \frac{1}{4}p^2A^2 + 2\cdot\frac{1}{8}P^2A^2 = 1 - pA + \frac{1}{2}p^2A^2 \pmod{p^3}.$$

So the l.h.s. is congruent to the r.h.s.

4.11. We found this statement in [10].

$$\sum_{k=1}^{p-1} \frac{1}{mp+k} = \frac{1}{2}\sum_{k=1}^{p-1}\Big(\frac{1}{mp+k} + \frac{1}{mp+p-k}\Big) =$$
$$= p\cdot\frac{2m+1}{2}\cdot\sum_{k=1}^{p-1}\frac{1}{(mp+k)(mp+p-k)} \equiv -p\cdot\frac{2m+1}{2}\cdot\sum_{k=1}^{p-1}\frac{1}{k^2} \equiv 0 \pmod{p^2}.$$

4.12. We found this statement in [8]. Since $2pq-1 = (2q-1)p + p - 1$, the last digit of the number $2pq-1$ in base p is $p-1$, and the remaining digits form the number $2q-1$. Similarly the last digit of the number $pq-1$ in base p is $p-1$, and the remaining part forms the number $q-1$. By Lukas' theorem $\binom{2pq-1}{pq-1} \equiv \binom{2q-1}{q-1}\binom{p-1}{p-1} \equiv \binom{2q-1}{q-1} \pmod p$. On the other hand since $\binom{2pq-1}{pq-1} \equiv 1 \pmod{pq}$, then $\binom{2pq-1}{pq-1} \equiv 1 \pmod p$. So $\binom{2q-1}{q-1} \equiv 1 \pmod p$. Analogously $\binom{2p-1}{p-1} \equiv 1 \pmod q$.

The inverse statement is trivial.

5 Sums of binomial coefficients

5.1. a) It follows from problem 1.3. If Δ_0^0 is a triangle consisting of the first 3 rows of the 3-arithmetical Pascal triangle, then the sum of its central binomial coefficients is divisible by 3. For arbitrary a the sum under consideration contains elements of several central triangles, which are multiples of Δ_0^0. So the total sum is divisible by 3, too.

Another solution ([CSTTVZ]) we can derive from the identity $\binom{2k}{k} = \sum\limits_{i=0}^{k}\binom{k}{i}^2$. Then $\sum\limits_{k=0}^{3^a-1} C_{2k}^k = \sum\limits_{k=0}^{3^a-1}\sum\limits_{i=0}^{k}\binom{k}{i}^2$. Since $1^2 = 2^2 = 1$, $0^2 = 0 \pmod 3$, the last sum modulo 3 equals the number of nonzero elements in the first 3^a rows of the Pascal triangle. This number is calculated in the problem 2.1a), it is divisible by 3.

b) Solution of [D]. The sum is a coefficient of x^{3^a-1} in the polynomial

$$x^{3^a-1}\left(1+\frac{(x+1)^2}{x}+\frac{(x+1)^4}{x^2}+\ldots+\frac{(x+1)^{2(3^a-1)}}{x^{3^a-1}}\right)=\frac{\frac{(x+1)^{2\cdot 3^a}}{x^{3^a}}-1}{\frac{(x+1)^2}{x}-1}\cdot x^{3^a-1}=\frac{(x+1)^{2\cdot 3^a}-x^{3^a}}{x^2+x+1}=$$
$$=\frac{x^{2\cdot 3^a}+\binom{2\cdot 3^a}{1}\cdot x^{2\cdot 3^a-1}+\binom{2\cdot 3^a}{2}\cdot x^{2\cdot 3^a-2}+\ldots+1-x^{3^a}}{x^3-1}\cdot(x-1)\,.$$

In order to find this coefficient we will perform the long division of the numerator by the denominator and then multiply the result by $(x-1)$. We do not need to find the quotient at whole, it is sufficient to perform the division till the moment when the coefficient of x^{3^a-2} will be found, remind that we are trying to find this coefficient modulo 3^a only. Since for $b \not\vdots 3$ all the binomial coefficient $\binom{2\cdot 3^a}{b}$ are divisible by 3^a (by Kummer's theorem), we can collect all these coefficient in a separate sum. When we divide this sum by x^3-1 all the coefficients of the quotient are divisible by 3^a therefore we can discard this sum. The remaining expression is

$$\frac{x^{2\cdot 3^a}+\binom{2\cdot 3^a}{3}\cdot x^{2\cdot 3^a-3}+\binom{2\cdot 3^a}{6}\cdot x^{2\cdot 3^a-6}+\ldots+1-x^{3^a}}{x^3-1}\cdot(x-1).$$

All the exponents in the numerator are divisible by 3, hence after division by x^3-1 all the exponents of the quotient are divisible by 3, too, and after the multiplying it by $x-1$, there will be no exponents of the form $3k+2$. So the coefficient that we seek equals $0 \pmod{3^a}$.

5.2. This problem was published in Monthly [25]. Since

$$\binom{2n+2}{n+1}-4\binom{2n}{n}=2\cdot\frac{2n+1}{n+1}\binom{2n}{n}-4\binom{2n}{n}=-2C_n\,,$$

then $C_n \equiv \binom{2n+2}{n+1}-\binom{2n}{n} \pmod 3$. Therefore this sum is telescopic modulo 3:

$$\sum_{k=1}^{n}C_k\equiv\left(\binom{2n+2}{n+1}-\binom{2n}{n}\right)+\left(\binom{2n}{n}-\binom{2n-2}{n-1}+\ldots\right)=\binom{2n+2}{n+1}+1\pmod 3.$$

So by Kummer's theorem we have to clarify when we have at least one carry in the addition of the number $(n+1)$ with itself in base 3. It it clear that it happens only if $n+1$ contains at least one 2 in base 2.

5.3. This is problem A5 of Putnam Math. Competition, 1998. Since $\frac{1}{p}\binom{p}{n}\equiv\frac{(-1)^{n-1}}{n} \pmod p$, we have

$$\sum_{n=1}^{k}\frac{1}{p}\binom{p}{n}\equiv\sum_{n=1}^{k}\frac{(-1)^{n-1}}{n}=\sum_{n=1}^{k}\frac{1}{n}-2\sum_{n=1}^{[k/2]}\frac{1}{2n}\equiv\sum_{n=1}^{k}\frac{1}{n}+\sum_{n=p-[\frac{k}{2}]}^{p-1}\frac{1}{n}\overset{*}{=}\sum_{n=1}^{p-1}\frac{1}{n}\equiv 0\pmod p.$$

The summation in the sum to the left of asterisk really starts from $n=k+1$ (it is easy to check: for $p=6r+1$ we have $k=4r$ and $p-[\frac{k}{2}]=4r+1=k+1$, similarly for $p=6r+5$).

5.4. This statement is from [11]. Solution [CSTTVZ]. Induction on n. The base is trivial. Prove the induction step from $n'=n-(p-1)$ to n. Let $q=\frac{n}{p-1}$. Since

$$\binom{n'+p-1}{x(p-1)}=\sum_{i=0}^{p-1}\binom{p-1}{i}\binom{n'}{x(p-1)-i},$$

we can rewrite the sum under consideration in the form

$$\binom{n}{p-1}+\binom{n}{2(p-1)}+\binom{n}{3(p-1)}+\ldots=\sum_{x=1}^{q}\sum_{i=0}^{p-1}\binom{p-1}{i}\binom{n'}{x(p-1)-i}=$$
$$=\sum_{i=0}^{p-1}\left(\binom{p-1}{i}\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right).\quad(11)$$

By the problem 1.1 a) we have $\binom{p-1}{i} \equiv (-1)^i \pmod p$; let $\binom{p-1}{i} = ap + (-1)^i$. By the problem 1.6 we have $\sum\limits_{x=1}^{q} \binom{n'}{x(p-1)-i} \equiv \binom{p-1}{i} \equiv (-1)^i \pmod p$ for $i = 0, 1, \ldots, p-2$; let $\sum\limits_{x=1}^{q} \binom{n'}{x(p-1)-i} = bp + (-1)^i$. Then

$$\binom{p-1}{i}\sum_{x=1}^{q}\binom{n'}{x(p-1)-i} = \big(ap+(-1)^i\big)\big(bp+(-1)^i\big) \equiv 1 + (-1)^i(ap+bp) =$$
$$= 1+(-1)^i\left(\left(\binom{p-1}{i}+\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right) - 2\cdot(-1)^i\right) = (-1)^i\left(\left(\binom{p-1}{i}+\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right)\right)-1 \pmod{p^2}.$$

Remind that these transformations hold for $0 \leqslant i \leqslant p-2$. We can continue equality (11), by separating the summand for $i = p-1$:

$$\sum_{i=0}^{p-1}\left(\binom{p-1}{i}\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right) \equiv \sum_{i=0}^{p-2}\left((-1)^i\left(\binom{p-1}{i}+\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right)-1\right)+\sum_{x=0}^{q-1}\binom{n'}{x(p-1)} =$$
$$= \sum_{i=0}^{p-2}(-1)^i\binom{p-1}{i} + \sum_{i=0}^{p-2}\left((-1)^i\sum_{x=1}^{q}\binom{n'}{x(p-1)-i}\right) - (p-1) + \binom{n'}{0} + \sum_{x=1}^{q-1}\binom{n'}{x(p-1)}.$$

The first sum here equals -1, because $\binom{p-1}{0} - \binom{p-1}{1} + \binom{p-1}{2} + \ldots = 0$. By the same reasons the second (double) sum together with the summand $\binom{n'}{0}$ equals 0. The last sum equals $1 + p(n'+1)$ by the induction hypothesis. Therefore the whole expression equals $-1 + 0 - p + 1 + 1 + p(n'+1) = 1 + pn'$. This is exactly what we need because $1 + p(n+1) = 1 + p(n'+p-1+1) \equiv 1 + pn' \pmod{p^2}$.

5.5. This is result of Fleck, 1913, it is cited in [18]. Solution [CSTTVZ].

For $p = 2$ the sum is not alternating and the result is trivial. Let p be odd. We use the induction on q. The base follows from the statement 2.5 a). Prove the induction step from $n' = n - (p-1)$ to n. The expression $\sum_x$ below denotes the summation over x in natural bounds (i.e. in bounds for which all the binomial coefficients are correctly defined). We have

$$\pm \sum_{m:m\equiv j \pmod p} (-1)^m\binom{n}{m} = \sum_x (-1)^x\binom{n'+p-1}{xp+j} = \sum_x (-1)^x \sum_{i=0}^{p-1}\binom{p-1}{i}\binom{n'}{xp+j-i} =$$
$$= \sum_{i=0}^{p-1}\binom{p-1}{i}\sum_x (-1)^x\binom{n'}{xp+j-i}.$$

By the induction hypothesis $p^{q-1} \mid \sum_x (-1)^x\binom{n'}{xp+j-i}$; by the problem 1.1 a) $\binom{p-1}{i} \equiv (-1)^i \pmod p$. Therefore

$$\sum_{i=0}^{p-1}\binom{p-1}{i}\sum_x (-1)^x\binom{n'}{xp+j-i} \equiv \sum_{i=0}^{p-1}(-1)^i\sum_x (-1)^x\binom{n'}{xp+j-i} \pmod{p^q}.$$

The last (double sum equals $\binom{n'}{0} - \binom{n'}{1} + \binom{n'}{2} - \binom{n'}{3} + \ldots = 0$.

5.6. The result of Bhaskaran (1965), it is cited in [18], solution [CSTTVZ].

Induction on n. Let

$$f(n,j) = \binom{n}{j} - \binom{n}{j+(p-1)} + \binom{n}{j+2(p-1)} - \binom{n}{j+3(p-1)} + \ldots$$

The base $n = p+1$ is trivial, but observe that $\binom{p+1}{i} \equiv 1 \pmod p$ for $i = 0, 1, p, p+1$, otherwise this binomial coefficient is divisible by p. Prove the step of induction from $n' = n - (p+1)$ to n. By the observation above we have

$$\binom{n'+(p+1)}{j+(p-1)k} = \sum_{i=0}^{p+1}\binom{n'}{j+(p-1)k-i}\binom{p+1}{i} \equiv \sum_{i\in\{0,1,p,p+1\}}\binom{n'}{j+(p-1)k-i} =$$
$$= \binom{n'}{j+(p-1)k} + \underline{\binom{n'}{j-1+(p-1)k} + \binom{n'}{j-1+(p-1)(k-1)}} + \binom{n'}{j-2+(p-1)(k-1)} \pmod p.$$

Since $f(n,j) = \sum_k (-1)^k \binom{n}{j+k(p-1)}$ is an alternating sum, the underlined summands cancel (except the first and the last, but these summands are equal to 0 due to incorrect binomial coefficietns). So we obtain the equalities

$$f(n,j) \equiv f(n',j) - f(n',j-2) \quad \text{при } j > 1\,, \qquad\qquad f(n,1) \equiv f(n',1) + f(n',p-2)\,.$$

Now the part "only if" of the problem statement follows from the induction hypothesis, and the part "if", too: if $f(n,j) \equiv 0 \pmod p$ for $j = 1, 3, \ldots, p-2$, then

$$f(n',p-2) \equiv f(n',p-4) \equiv \ldots \equiv f(n',1) \equiv -f(n',p-2)\,,$$

from where $f(n',j) \equiv 0 \pmod p$ for all required j, and then $n' \vdots (p+1)$, hence $n \vdots (p+1)$.

References

The authors of many solutions are participants of the conference:

[D] Didin Maxim;
[K] Krekov Dmitri;
[J] Jastin Lim Kai Ze;
[T] Teh Zhao Yang Anzo;
[CSTTVZ] Ćevid Domagoj, Stokić Maksim, Tanasijeviić Ivan, Trifunović Petar, Vukorepa Borna, Žikelić Đorđe

Список литературы

[1] *Винберг Э. Б.* Удивительные свойства биномиальных коэффициентов. // Мат. просвещение. Третья серия. Вып. 12. 2008

[2] *Гашков С.Б., Чубариков В.Н.* Арифметика. Алгоритмы. Сложность вычислений. М.: Высш. шк., 2000.

[3] *Дынкин Е.Б., Успенский В.А.* Математические беседы. 2-е изд. М.: ФИЗМАТЛИТ, 2004.

[4] Петербургские математические олимпиады, 1961–1993. СПб: Лань, 2007.

[5] *Табачников С.Л., Фукс Д.Б.* Математический дивертисмент. 30 лекций по классической математике. М.: МЦНМО, 2011.

[6] *Фукс Д.Б., Фукс М.Б.* Арифметика биномиальных коэффициентов // Квант. 1970. №6. С. 17–25.

[7] *Ширшов А.И.* Об одном свойстве биномиальных коэффициентов // Квант. 1971. №10. С. 16–20.

[8] *Cai T.X., Granville A.* On the residues of binomial coefficients and their products modulo prime powers /!/ Acta

[9] *Calkin N. J.* Factors of sums of powers of binomial coefficients // Acta Arith. 1998. Vol. 86. P. 17–26.

[10] *Carlitz L.* A note of Wolstenholme's theorem // Amer. Math. Monthly. 1954. Vol. 61. № 3. P. 174–176.

[11] *Dimitrov V., Chapman R.* Binomial coefficient identity: 11118 // Amer. Math. Monthly. 2006. Vol. 113. № 7. P. 657–658.

[12] *Everett W.* Subprime factorization and the numbers of binomial coefficients exactly divided by powers of a prime // Integers. 2011. Vol. 11. # A63. `http://www.integers-ejcnt.org/vol11.html`

[13] *Fine N.* Binomial coefficient modulo a prime // Amer. Math. Monthly. 1947. Vol. 54. № 10. Part 1. P. 589–592.

[14] *Gardiner A.* Four problems on prime power divisibility // Amer. Math. Monthly. 1988. Vol. 95. № 10. P. 926–931.

[15] *Gauss K.* Disquisitiones arithmeticae. 1801. Art. 78.

[16] *Gessel I.* Wolstenholme revisited // Amer. Math. Monthly. 1998. Vol. 105. № 7. P. 657–658.

[17] *Granville A.* Arithmetic properties of binomial coefficients. `http://www.dms.umontreal.ca/~andrew/Binomial/`

[18] *Granville A.* Binomial coefficients modulo prime powers.

[19] *Granville A.* Zaphod Beeblebrox's Brian and the Fifty-ninth Row of Pascal's Triangle // Amer. Math. Monthly. 1992. Vol. 99. № 4. P. 318–331.

[20] *Granville A.* Correction to: Zaphod Beeblebrox's Brian and the Fifty-ninth Row of Pascal's Triangle // Amer. Math. Monthly. 1997. Vol. 104. № 9. P. 848–851.

[21] *Hinz A.* Pascal's triangle and tower of Hanoi // Amer. Math. Monthly. 1992. Vol. 99. № 6. P. 538–544.

[22] *Loveless A.* A congruence for products of binomial coefficients modulo a composite // Integers: electronic journal of comb. number theory 7 (2007) # A44

[23] *McIntosh R.* On the converse of Wolstenhome's theorem // Acta Arithmetica. 1995. Vol. 61. №4. P. 381–388.

[24] *Meštrović R.* On the mod p^7 determination of $\binom{2p-1}{p-1}$ // `http://arxiv.org/pdf/1108.1174v1.pdf`

[25] *More Y., Chapman R.* The sum of Catalan numbers, modulo 3: 11165 // Amer. Math. Monthly. 2007. Vol. 114. № 5. P. 454–455.

[26] *Morley F.* Note on the congruences $2^{4n} \equiv (-)^n (2n)!/(n!^2)$, where $2n+1$ is a prime // Annals of Math. 1894-1895. Vol. 9. №1. P. 168–170.

[27] *Roberts J.* On binomial coefficient residues // Canad J. Math. 1957. Vol. 9. P. 363–370.

[28] Sun Z.-W., Wan D. On Fleck quotients // `arXxiv:math.0603462v3`

Applications to ring theory and some history

A. Belov, M. Kharitonov

August 2, 2012

We consider a set S with a binary operation "+" such that

1. for any $a, b, c \in S$, we have $a + b = b + a$, $(a + b) + c = a + (b + c)$;
2. there exists a special element $0 \in S$ such that $a + 0 = a$ for all $a \in S$;
3. for any $a \in S$, there exists an element, denoted by $(-a)$, such that

$$(-a) + a = a + (-a) = 0.$$

Examples: remainders modulo some number; rotations of a plane with a fixed point.

We say that a set with an addition is a *ring*, if it has a multiplication, i.e. a binary operation which satisfies the following conditions

1. $a(bc) = (ab)c$ for all a, b, $c \in S$ (associativity);
2. $a(b + c) = ab + ac, (b + c)a = ba + ca$ for all a, b, $c \in S$ (distributivity).

Examples: remainders modulo some number; polynomials of one or more variables.

Problem 1.1. Provide an example of a ring with a zero-divisor, i.e. such ring that there exists two non-zero elements a, b such that $ab = 0$.

Definition 1.1. *We call an element a* unit of a ring R *or a* neutral element of R *(and denote it 1), if we have $1A = A1 = A$ for all $A \in R$. We say that an element of R is an inverse to A (and denote it A^{-1}), if $AA^{-1} = A^{-1}A = 1$.*

Problem 1.2. Prove that in a ring with a unit commutativity axiom for an operation "+" follows from the other axioms.

Problem 1.3. Construct a ring R with 4 elements such that any non-zero element of R has an inverse.

Problem 1.4. Construct a non-commutative ring R, i.e. such ring R that $ab \neq ba$ for some $a, b \in R$.

Let R be a ring with a unit. Let $n \in \mathbb{Z}_{\geq 1}$ be the smallest number such that

$$1 + 1 + 1 + ... + 1(n \text{ times}) = 0.$$

Then n is called *the order of the unit of* R. If such number n does not exist, we assume that the order of the unit of R equals 0.

Problem 1.5. Let R be a ring such that the order of the unit of R equals 0. Assume that there exist elements $e, f, g \in R$ such that $e^2 = e, f^2 = f, g^2 = g$ and $e + f + g = 0$. Prove that $e = f = g = 0$.

Definition 1.2. *A non-commutative polynomial $f = f(x_1, ..., x_n)$ is called* an identity *of a ring R, if $f(x_1, ..., x_n) = 0$ for any $x_1, ..., x_n \in R$. We say that the identity $f(x_1, ..., x_n)$ follows from a set of identities $\{g_i\}$, if one can deduce $f(x_1, ..., x_n) = 0$ from $g_i(x_1, ..., x_n) = 0$ algebraically.*

Problem 1.6. Provide an example of a ring with a non-zero multiplication in which the identities $x^2 = 0$ and $xy = yx$ are satisfied.

Definition 1.3. *If any non-zero element of a ring R is invertible, then R is called a skew-field. A ring R is called commutative, if R satisfies the identity $xy - yx = 0$. A commutative skew-field is called a field.*

Examples.

1. Field of real numbers, field of complex numbers, ring of polynomials, ring $\mathbb{Z}_n$ of remainders modulo n. If n is prime, then $\mathbb{Z}_n$ is a field.

2. A fundamental example of a non-commutative ring is a ring of matrices. We construct it now. The elements of this ring is $n \times n$ tables filled by numbers. We use index A_{ij} to denote the number puted in i-th line and j-th column. We put

$$(A + B)_{ij} = A_{ij} + B_{ij} \text{ (pointwise addition).}$$

To define multiplication we put

$$(AB)_{ik} = (\sum_j A_{ij} B_{jk}).$$

3. The skew-field of quaternions is a set of elements $ai + bj + ck + d$, where a, b, c, d are real numbers, with a pointwise addition and multiplication defined by the following relations:

$$ij = -ji = k, ki = -ik = j, jk = -kj = i, i^2 = j^2 = k^2 = -1.$$

Problem 1.7. Check that all mentioned sets with addition and multiplication are rings.

Definition 1.4. *By definition, an N-free associative algebra over ring A (or the ring of non-commutative polynomials of the ring R), is a set $\sum_i a_i v_i$, where $a_i \in R$, and v_i are words in alphabet $\{a_1, ..., a_N\}$. If $v = \sum_i a_i v_i$, $u = \sum_i b_i v_i$, then we set*

$$u + v = \sum_i (a_i + b_i) v_i,\ uv = \sum_{i,j} (a_i b_j)(v_i v_j).$$

Remark 1.1. In the rest of the problem set any algebra considered is an associative algebra with a unit.

We put $[a, b] := ab - ba$.

Examples.

1. Commutativity identity $[a, b] = ab - ba$ is satisfied by definition in all commutative rings.

2. Let p be a prime number. Identity $x^p - x = 0$ is satisfied for a ring of remainders modulo p (small Ferma's theorem).

3. Hence $ab+ba=(a+b)^2-a^2-b^2$, the identity $ab+ba=0$ follows from the identity $a^2=0$.

Problem 1.8. 1. Prove that Hall's identity $[[x,y]^2,z]=0$ holds for 2×2 matrices.

2. Prove that standard identity of degree 4

$$\sum_{\sigma\in S_4}(-1)^\sigma x_{\sigma(1)}\cdots x_{\sigma(4)}=0$$

holds for 2×2 matrices.

For any algebra for which some non-trivial identity is satisfied, is satisfied the standard identity of some degree. An algebra of $n\times n$ matrices satisfies the standard identity of degree $2n$.

It is known that for algebra of 2×2 matrices all identities follows from the Hall's identity and the standard identity of degree 4 (this is quite complicated theorem which is proven by Yu. Razmislov in 1973). Even for 3×3-matrices the basis of identities is not known yet.

Definition 1.5. *An algebra is called a nil-algebra, if there exist a function $n: A\to \mathbf{N}$ such that, for all $x\in A$, we have $x^{n(x)}=0$. If for some n the algebra satisfies identity x^n, then it is called a nil-algebra of index n.*

Definition 1.6. *An algebra A is called nilpotent, if the identity $x_1...x_k=0$ is satisfied for some k.*

Definition 1.7. *An element $\tau\in A$ is called algebraic of index k, if $\sum\limits_{i=1}^{k}\tau^i a_i=0$ for some $a_1,...,a_k\in\mathbb{Z}$. An algebra A is called algebraic of index k if any element of A is algebraic of index k. An algebra A is called algebraic if any element of A is algebraic of some index (depending on the element).*

Problem 1.9. 1. Prove that any algebraic algebra of index k satisfies a non-trivial identity.

2. * Prove that algebra of $n\times n$-matrices is algebraic of index n.

3. Fix $n\in\mathbb{Z}_{\geq 0}$. We denote by SS_n the set of all subsets of $\{1,...,n\}$. For a $\tau\in\mathrm{SS}_n$ we put $S_\tau:=\sum_{i\in\tau}x_i$. Prove the following equation (is called polarization):

$$\sum_{\tau\in\mathrm{SS}_n}(-1)^{|\tau|}(S_\tau)^n=\sum_{\sigma\in S_n}x_{\sigma(1)}\cdots x_{\sigma(n)}.$$

(If we assume that x_i commutes on with each other, then the right-hand side will equal $n!x_1\cdots x_n$.)

4. Prove that from the identity x^n follows the identity $\sum_{\sigma\in S_n}x_{\sigma(1)}\cdots x_{\sigma(n)}$.

5. Prove that any identity has a polylinear (i.e. linear over any variable) analogue of the same degree.

Let $a_1, ..., a_l$ be some elements of some algebra A with a unit and $w = a_{i_1}a_{i_2}...a_{i_s}$ be a word of alphabet with letters $\{a_1, ..., a_l\}$. We denote by $w(a)$ the element $a_{i_1} \cdot a_{i_2} \cdot ... \cdot a_{i_s}$ of A.

Problem 1.10. Let A be an algebra which satisfies an identity of degree n and $a_1, ..., a_l$ be some elements of A. Let w be an n-divisible word of alphabet $\{a_1, ..., a_l\}$. Prove that $w(a) = \sum\limits_{w_i \prec w} c_i w_i(a)$. For some finite set of words w_i and some $c_i \in \mathbb{Z}$ [1].

A. Kurosh have posed in 1941 the following question.

Kurosh's problem. *Is it true that any algebraic finitely generated algebra, which satisfies some identity of degree n, is finite-dimensional?*

First solution of Kurosh's problem, obtained by Levitskii and Kaplanskii in 1951, and use highly non-elementary methods. In 1957, A. Shirshov develops purely combinatorial technique, which provides another approach to Kurosh's and other nilpotency-related problems.

Problem 1.11. a) Resolve Kurosh's problem using Shirshov's theorem of height.
b)* Prove that l-generated nil-algebras are n-nilpotent of index $k(n, l)$.

Our next goal is to receive estimates for a function $k(n, l)$. Estimates for height in combinatorics of words straightforwardly leads to estimates of $k(n, l)$. First estimate of A. Shirshov are extremely overwhelming but his works contains deep ideas which still stay in focus of researchers. A. Kolotov receives in 1982 twice exponential estimate for $k(n, l)$ of type (l^{l^n}), where l is a number of generators and n is a degree of identity. A. Belov receives exponential estimate $n^3 l^{3n}$ in 1992, this estimate has been upgraded in works of A. Klein in 2000.

E. Zelmanov have posed the following question in 1991.

Problem 1.12. Let $F_{2,m}$ be a free 2-generated associative ring with the identity $x^m = 0$. Is it true that the nilpotency class of $F_{2,m}$ depends exponentially on m?

Problem 1.13. Prove that the last problem from Section "Exponential estimates" provides a positive answer to Zelmanov's question.

A.Belov and M. Kharitonov receive in 2012 a subexponential estimate for a height. In connection with these results appear the following problem:

Problem 1.14. Receive a polynomial estimate for height.

And another one.

Problem 1.15. Does there exist an estimate for a height which is polynomial with respect to degree and linear with respect to the number of letters in an alphabet?

And the last one.

Problem 1.16. Receive lower height estimate.

[1] Вообще говоря, алгебра определена над полем и над ним же определены соотношения. У нас об этом разговора нет и потому вопрос о том, где живут c_i — сложен.

Periodicity and order

A. Belov, M. Kharitonov

August 10, 2012

1 Introduction

We consider words in some alphabet. We fix some word W. If we are very lucky, W is periodic, i.e. it is some short word w repeated many times. This word w is called *period* of W. We write $c = ba$, for words a, b, c, if c is obtained from b and a by gluing together (we note that any word is a product of several letters an that this product is noncommutative, i.e. $ab \neq ba$, even if a, b are letters). From the point of view of this product, any periodic word is a power of some (short) word.

An arbitrary word does not tend to be periodic. More often a word is a product of several periodic words. We call such words *piecewise-periodic*. Any word can be expressed as a product of several periodic words and some small pieces inbetween them.

Let now fix an alphabet $\mathcal{A} = \{a_1, \ldots, a_s\}$. Order $a_1 \prec a_2 \prec \cdots \prec a_s$ on the set of letters induces a lexicographical order on the set of words. We write $U \prec V$, whenever the first letter of U is smaller than the first letter of V, if they coincide, if the second letter of U is smaller than the second letter of V e.t.c.

If a word U starts from a word V, U and V are *incomparable*, i.e. neither $U \prec V$, nor $V \prec U$. Similarly words are ordered in a dictionary (in this case the shortest from two incomparable words usually goes earlier). A given word A may have a property to be *lexicographically ordered*: this means that a letter with a smaller index always goes earlier than a letter with a bigger index. If a word W is not ordered, it has several *disorders*. There are more powerful and less powerful disorders. We say that a word W has a *k-disorder* if W contains k subwords $W_1, ..., W_k$ such that
1) W_i does not intersect W_j for $i \neq j$;
2) if $i < j$, then W_i goes earlier than W_j;
3) if $i < j$, then $W_i \succ W_j$.
By definition, a word W is *k-divisible*, if W has a k-disorder.

It appear so that the "degree of divisibility" of a word W and the number of periodic subwords of W are closely related. In this project we convert this metamathematical statement to a theorem and try to amuse participants by related facts. We present this theorem right now.

Shirshov's Theorem of height. The set of non k-divisible words is piecewise-periodic with period $(k-1)$ or less, i.e. there exists a function $H(k, s)$ such that any word is either k-divisible, or can be splitted on $H(k, s)$ pieces such that any piece is a subword of a periodic word with a length of period $(k-1)$ or less.

The main goal of this project is to produce estimates for the function $H(k, s)$ of type $H(k, s) \leq sk^{C \ln k}$, where C is some constant, which does not depend on k and s. To

achieve this goal we will need some deep combinatorial results, in particular Dilworth's theorem. Another goal is to count polylinear [1] words, which have no k-disorders [2]. We also estimates the number of subwords with a given period, and here the graphs point of view is useful, in particular we explain a connection with a Ramsey's theory.

The key problems in the first problem set are 2.10, 3.7, 4.11, 4.12.

We also attach to a project a problem set called "Application to ring theory and some history", which is dedicated to the mentioned topics. This problem set is completely independent from all other problem sets, but it provides an area of mathematics, for which the results of our project are also very interesting. It turn's out that the Shirshov's theorem allows to solve some problems of this area, which seems to be completely nonrelated to it and stay unsolved for more than 20 years.

2 De arte kombinatoria

Problem 2.1. Karlsson know how to write only words which does not contain subwords with two or more different letters. How many words of length n Karlsson know how to write, if his alphabet contains l letters?

Problem 2.2. The dictionary of Winnie-the-Pooh tribe has 20 letters. The language of this tribe consider as a phrase any combination of this words. There exists two verbal spells "earth stay on Great Crocodile" and "every evening Crocodile eat Sun", which evoke an earthquake. How many phrases from 10 words does not provide the tribe with an earthquake?

Problem 2.3. The alphabet of a small-wide-tribe "Smeshariki" consists of l letters. May the language of this tribe contains a word of length l which contains precisely

a) $l+1$

b) $\frac{l(l-1)}{2}-1$

c*) 2^l

different subwords?

Problem 2.4. An alphabet of Endeans has N letters, and any word of Endeans consists of letters of their alphabet. It is known that a word W repeated twice means the same with W, and that the meaning of a word $W_1W_2W_3$ is the same with $W_1W_2W_2W_3$. For example "BC" means the same with "BBC". Prove that the number of words with different meanings is finite if

a) $N=2$;

b) $N=3$.

By u^t we denote t copies of a word u putted in one line.

Problem 2.5. Fix an alphabet $\mathcal{A}=\{a,b\}$. What is the minimal number of words $\{W_1,...,W_k\}$ such that the set of all words of length 100, which does not contain $\{W_1,...,W_k\}$ as subwords, consists of $(ab)^{50}$ and $(ba)^{50}$?

Problem 2.6. Let $k,t\in\mathbb{Z}_{\geq 1}$. Prove that, if a word V of length $k\cdot t$ has not more than k different subwords of length k, then for some word v the word V contains a subword v^t.

[1] i.e. such words W that any letter is used in W not more than once

[2] For example, the number of words which are not 3-divisible equals to a Catalan's number

Problem 2.7. Provide a bijection between the following sets:

- sequences of natural numbers $1 \leqslant a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n$, where $a_i \leqslant i$;
- transpositions of numbers $1, 2, \ldots, n$, such that the length of any decreasing sequence is 2 or less.

Problem 2.8. Hundred man-eaters come to a feast. During a feast man-eaters eat themselves. Therefore appear a sequences of man-eaters such that a man-eater eats a man-eater which eats a man-eater which eats a man eater... What is the smallest possible the longest such sequence of man-eaters with additional condition that from any 10 man-eaters any one eats the other one?

Similar problems to Problem 2.8 appear in subsection "Dilworth's theorem".

Definition 2.1. *We call a word u* non-cyclic*, if u is not equal to v^k, for any word v and any $k > 1$.*

Problem 2.9. Let u, v be different non-cyclic words of length m and n respectively. Assume that a word W contains subwords $u' = u^{m \cdot n}$ and $v' = v^{m \cdot n}$. Prove that the length of the common part of u' and v' does no exceed $m + n - 2$.

Problem 2.10. An infinite band is filled by numbers {1,..., 9}. Prove that either one can cut of from it 10 non-intersecting numbers with 1000 digits each, which form an increasing sequence on a band, or there exists a number with 10 or less digits which repeats 50 times in succession.

3 Dilworth's theorem

Problem 3.1. Is it true that, for any sequence of numbers of length 5, there exists a subsequence of length 3 which is ordered (i.e. it is or increasing, or decreasing)?

Problem 3.2. Is it true that, for any sequence of numbers of length 9, there exists a subsequence of length 4 which is ordered (i.e. it is or increasing, or decreasing)?

Problem 3.3. Prove that that for any sequence of numbers of length 10 there exists a subsequence of length 4 which is ordered (i.e. it is or increasing, or decreasing).

Problem 3.4. Prove that that for any sequence of numbers of length $mn+1$ either there exists a decreasing subsequence of length $m + 1$ or there is an increasing subsequence of length $n + 1$.

By definition a *partially ordered set* (POS) is a set M with a relation $\prec$ on it such that, for any two elements a and b of M, or $a \prec b$ is true or is false. This relation should satisfies the following axioms:

1. if $a \prec b$ and $b \prec c$, then $a \prec c$ (transitivity);
2. if $a \prec b$, then a is not b.

Problem 3.5. May $a \prec b$ and $b \prec a$ be true simultaneously?

Problem 3.6. Prove that the set of words with a lexicographical order is a POS.

Definition 3.1. *A POS M, for any elements $a, b \in A$ of which $a = b$, $a \prec b$, $b \prec a$, is called linearly ordered. Such POS are also known as chains.*

Problem 3.7. Let m, n be natural numbers. Prove that in any POS with $mn+1$ elements there exists either a subset with $m+1$ elements which is a chain or there is a subset with $n+1$ elements which is an antichain (i.e. such that any two elements of which are incomparable).

Problem 3.8. Let M be a POS and $c(M)$ be the length of the longest chain of M. Then M can be splitted on $c(M)$ antichains.

The following theorem is in some sense dual to Problem 3.8.

Dilworth's theorem. Let M be a POS and $ad(M)$ be the length of the longest antichain of M. Then M can be splitted on $ad(M)$ chains.

4 Exponential estimates

We fix an alphabet $\mathcal{A} = \{a_1, a_2, \ldots, a_l\}$ and we fix a linear order on $\mathcal{A}$: $a_1 \prec a_2 \prec \cdots \prec a_l$. This order introduces a lexicographical order on the set of words of $\mathcal{A}$. We consider two words u and v. If u begins from v or v begins from u, we call u and v *incomparable* (with respect to each other). Otherwise there exist words w, u', v' such that $u = wu'$, $v = wv'$ and first letters of u' and v' are different (w could be an empty word). If the first letter of u' is greater than the first letter of v', we say that u is greater than v and write $v \prec u$, otherwise we say that v is greater than u and write $v \prec u$. The set of words of $\mathcal{A}$ with respect to $\prec$ is a POS. The order $\prec$ is called *lexicographical* (see also Introduction). It would be significant later that some words are incomparable with respect to the lexicographical order $\prec$.

Problem 4.1. Let alphabet $\mathcal{A}$ consists of letters a, b, c. We introduce an order on them: $a \prec b \prec c$. Find the longest increasing sequence from the following list of words. Which pairs of these words are incomparable?

$$cb, abc, bac, abb, b, ccc, abc$$

The following definitions will be useful later.

Definition 4.1. *A word W is called n-divisible, if there exist words $u_1, ..., u_n$ such that $W = v \cdot u_1 \cdots u_n$ and $u_1 \succ \ldots \succ u_n$.*

Definition 4.2. *A word W is called k-ordered, if W is k-divisible but not $(k+1)$-divisible.*

Problem 4.2. Find a number of a) 1-ordered b) 2-ordered words of length l.

Problem 4.3. Let n be a natural number, u — noncyclic word of length n or less. Prove that the word u^{2n} is not n-divisible.

Definition 4.3. *a) A word v is called a tale of a word u if there exists a word w such that $u = vw$.*

b) If a word v has a subword u^t we say that v has a period of cyclicity t

Problem 4.4. Let l, d be some natural numbers. Prove that, for a word W of length l, or first $[l/d]$ tales are pairwise incomparable, or W has a period of length d.

Further we assume that $n \leqslant d$.

Definition 4.4. *A word W is called* (n, d)-cancellable, *either if W is n-divisible, or if there exists a word u such that u^d is a subword of W.*

Problem 4.5. Prove that if a word W has n pairwise equal non-intersecting subwords of length n, then W is (n, n)-cancellable.

Definition 4.5. *A word W is called n-divisible from tale, is there exists tales $u_1, \ldots, u_n$ such that $u_1 \succ u_2 \succ \ldots \succ u_n$, and, for any $i = 1, 2, \ldots, n-1$, u_i begins earlier than u_{i+1}.*

Problem 4.6. Prove that if a word W is
a) n^3d-divisible from tale,
b) $3n^2d$-divisible from tale,
c) $4nd$-divisible from tale,
then W is either n-divisible, or W has a subword u^d for some nontrivial word u.

Problem 4.7. For any pair of natural numbers (n, d) (except pair (1, 1)), provide an example of a word W of length $(nd-1)$ such that any set of tales of W is not increasing and W is not $(n+1, d)$-cancellable.

Problem 4.8. Try to enhance an estimate from Problem 4.6.

We fix an alphabet $\mathcal{A}$ of length l, a word W of this alphabet of length $r(W)$ and natural numbers $n \leqslant d$.

Further we assume that, for all words u, W does not contain a subword u^d and W is not $4nd$-divisible from tale. We consider first $[r(W)/d]$ tales of W (further we denote this set of words by Ω). Then by Dilworth's theorem we can split Ω on $(4nd-1)$ groups such that tales in one group form a chain.

In solutions of the following problems we expect that you use previous ones.

Problem 4.9. Prove that, from any $4nd^2$ tales of Ω, there exists two tales, for which the first subwords of length $4nd$ are pairwise different.

Problem 4.10. In some infinite parliament any member has not more than 3 enemies among members. Prove that there is a way to divide this parliament into two houses such that a member of any house has not more than 1 enemy among members of his own house.

In the following problems we assume that l, n, d are variables and that W is some word defined over alphabet of l-letters.

Problem 4.11 (Shirshov's lemma)**.** Prove that there exists a function $f(l, n, d)$ such that for any subword W defined over alphabet of l-letters either W is (n, d)-cancellable, or $r(W) < f(l, n, d)$.

Problem 4.12. Prove that $f(l, n, d) < l(4nd)^{4nd+2}$.

5 Advanced estimates

We fix natural numbers s, n, d such that $d \geq n$. We also fix an alphabet $\mathcal{A} := \{a_1, ..., a_s\}$ and a word W of this alphabet of length $r(W)$ such that W is not (n, d)-cancellable (i.e. either W is n-divisible or W has a subword u^d for some non-zero word u). Let $\Omega_d(W)$ be the set of tales of W which begins from the first $[r(W)/d]$-letters of W. By text after Problem 4.8 the tales of Ω are splitted on $4nd - 1$ groups which we call *colors*. We set $p_{n,d} := 4nd$. By definition *a k-start* of a word W is a subword of W which consists from the first k letters of W.

Problem 5.1. Assume that $\Omega_d(W)$ has a subset Ω_c with $p_{n,d}^2 + p_{n,d} + 1$ elements such that all elements of Ω_c have the same color and such that any two elements of Ω_c have different $2k$-starts for some number k. Prove that there exist two subwords of Ω_c which have pairwise different k-starts.

Problem 5.2. Prove that for any subset Ω_c of $\Omega_d(W)$ with $(d+1)p_{n,d}^4$ elements there exist two elements of Ω_c which have the same color and have different $\frac{p_{n,d}}{2}$-starts.

Problem 5.3. Prove that for any subset Ω_c of $\Omega_d(W)$ with $(d+1)p_{n,d}^7$ elements there exist two elements of Ω_c which have the same color and have different $\frac{p_{n,d}}{4}$-starts.

Assume that $p_{n,d} = 2^t$ for some natural number t.

Problem 5.4. Prove that for any subset Ω_c of $\Omega_d(W)$ with $(d+1)p_{n,d}^{1+3t}$ elements there exist two elements of Ω_c which have different 1-starts.

Problem 5.5. Prove that if the length of W is greater than $(d+1)p_{n,d}^{2+3t}s$, then W is either n-divisible or has a subword u^d for some non-zero word d.

Using terms of problem set "Application to ring theory and some history" and problem 5.5 one can solve the problem posed by E. Zelmanov in 1991:

Let $F_{2,m}$ be a free 2-generated associative ring with the identity $x^m = 0$. Is it true that the nilpotency class of $F_{2,m}$ depends exponentially on m?

Assume that $n = 2^q$ for some natural number q.

Problem 5.6. Let $\Gamma_{n,s}^1$ be a finite set which consists from all words of length n or less over alphabet $\mathcal{A}$ with s letters. Let $\Gamma_{n,s}$ be an infinite set which contains $\Gamma_{n,s}^1$ and all powers of all elements of $\Gamma_{n,s}^1$. Prove that if W is not n-divisible, then W is a product of not more than n^{100q} words from $\Gamma_{n,s}$.

Periodicity and order *

A. Belov, M. Kharitonov, S. Grigoriev, A. Petukhov

August 11, 2012

2 De arte kombinatoria

Problem 2.1. Karlsson know how to write only words which does not contain subwords with two or more different letters. How many words of length n Karlsson know how to write, if his alphabet contains l letters?

Solution. Karlsson can write only words with one repeating letter. There are precisely l such words. □

Problem 2.2. The dictionary of Winnie-the-Pooh tribe has 20 letters. The language of this tribe consider as a phrase any combination of this words. There exists two verbal spells "earth stay on Great Crocodile" and "every evening Crocodile eat Sun", which evoke an earthquake. How many phrases from 10 words does not provide the tribe with an earthquake?

Answer. $20^{10} - 12 \cdot 20^5 + 4$. □

Problem 2.3. The alphabet of a small-wide-tribe "Smeshariki" consists of l letters. May the language of this tribe contains a word of length l which contains precisely

a) $l + 1$
b) $\frac{l(l-1)}{2} - 1$
c*) 2^l

different subwords?

Solution. a) There are no such words for $l = 1$. For $l = 2$ such word do exist: ab. Further we assume that $l \geqslant 3$.

We consider words with one repeating letter. They have precisely l subwords. If a word W contains at least two different letters, then W contains at least 2 different subwords of length 1 and at least 2 different subwords of length 2 (note that $l \geqslant 3$). Then the number of subwords of W is greater or equal then $l + 2$.

b) We left this problem for $l \leqslant 4$ as an exercise to a reader. We provide examples of such words for $l = 5, 6, 7$;

$$ababa, aaaabb, aabbabb.$$

*If you have any suggestations or find any bugs please write me on krab8nog@yandex.ru.

If $l \geqslant 8$ we construct such word by induction: we add to the end of a word of this kind of length $l-3$ new letter, which is not used before, 3 times.

c) We left this problem for $l \leqslant 4$ as an exercise to a reader. Further we assume that $l \geqslant 5$.

We will prove that there are know such word W over alphabet of l-letters such that W has precisely $2l$ subwords. Assume that such a word W exists. If the only one letter is used in W, W can't contain $2l$ subwords. Therefore we assume that 2 or more letters are used in W. This implies that for any $1 \leqslant k < l$ there exist at least two different subwords of length k (otherwise the only one letter is used in W). We consider subwords of W of length 2. If only two such words exist, W is of one of the following types:

$$ababab...ab, \text{ or } ababab...aba, \text{ or } abbb...b, \text{ or } aaa...ab.$$

All these words has $2l-1$ subwords or less. Therefore words of these types does not satisfy the conditions of problem. Further we assume that W has at least 3 different subwords of length 2. By the same arguments we assume W has at least 3 different subwords of length 3. Hence the number of different subwords of W is not smaller than $2+3+3+...+2+1$ (l summands). This sum equals $2l+1$ and therefore for $l \geq 5$ there are no words of length l with precisely $2l$ subwords. □

Problem 2.4. An alphabet of Endeans has N letters, and any word of Endeans consists of letters of their alphabet. It is known that a word W repeated twice means the same with W, and that the meaning of a word $W_1W_2W_3$ is the same with $W_1W_2W_2W_3$. For example "BC" means the same with "BBC". Prove that the number of words with different meanings is finite if

a) $N = 2$;

b) $N = 3$;

c) N is arbitrary.

Proof. We start with a definition. We say that a word is *uncancellable*, if any word A', which smaller than A, has different meaning with A.

We use induction by a number N of letters of an alphabet.

The base for $N = 1$ is obviously true.

We prove that from N-th statement follows $(N+1)$-th statement. By the hypothesi of induction the number of uncancellable words in alphabet of N letters is finite. We denote by d a number which is greater then the length of any uncancellable word of this alphabet. Now we consider alphabet $\mathcal{A}_{n+1}$ of $n+1$ letters. Assume that there exists an uncancellable word W of $\mathcal{A}_{n+1}$ of length $(d+1)(N+1)^{d+2}$ or more. This word is a union of $(N+1)^{d+2}$ blocks of length $(d+1)$ with some other piece. Obviously, there are two non-isomorphic non-intersecting blocks B among them. If N or less letters are used in B, then B is cancellable by the induction hypothesi B and thus W is cancellable. Therefore all letters of $\mathcal{A}_{N+1}$ are used in B.

We now show that if all letters of $\mathcal{A}_{N+1}$ are used in a word B and C is any word, then B has the same meaning with BCB (if it is so, then the word W of the previous paragraph is cancellable and we prove the induction hypothesi).

To prove this fact we have to prove that there exists a word Y such that B has the same meaning with BCY (if it is so, then the words

$$B \leftrightarrow BCY \leftrightarrow BCBCY \leftrightarrow BCB$$

have the same meaning and we prove the induction hypothesi). Assume that $C = c_1c_2...c_k$. As c_1 is used in B,

$$B = Sc_1M.$$

Then B has the same meaning with the word Bc_1M. Hence c_2 is used in B, $B = Ec_2D$ and the words B_1M and $Bc_1c_2Dc_1M$ have the same meaning. And so on. Therefore, and Y have the same meaning for some word Y.

Thus there is no uncancellable words of length $(d+1)(N+1)^{d+2}$ or greater over $\mathcal{A}_{N+1}$. We finish with induction.

As a corollary, the number of words over $\mathcal{A}_N$ for any N is finite. □

Problem 2.5. We fix an alphabet $\mathcal{A}_l$ and consider some set of words $W_1, ..., W_s$, which we call {ban-words}. We say that a word W is *permitted*, if $W_1, ..., W_s$ are not subwords of W. What is the minimal number of ban-words $W_1, ..., W_s$ such that the only permitted subwords with 100 letters are $(ab)^{50}$ and $(ba)^{50}$?

Solution. We consider all words over $\mathcal{A}_l$ which are repetition of one letter of length 100. There are precisely l of them and they do not share subwords. Therefore we have to use at least l ban-words.

We show that it is enough to use l ban-words. Namely, the set of ban-words $\{aa, bb, c, d, ...\}$ permit precisely two words: $abab....ab$ and $baba...ba$. □

By u^t we denote t copies of a word u putted in one line.

Problem 2.6. Let $k, t \in \mathbb{Z}_{\geq 1}$. Prove that, if a word V of length $k \cdot t$ has not more than k different subwords of length k, then for some word v the word V contains a subword v^t.

Proof. We prove that problem by induction by k.

The base ($k = 1$) is obvious. We denote by V^- a word V without the last letter. If V^- has $(k-1)$ or less subwords of length $(k-1)$, then by induction hypothesi V^- has a period with t repetitions (note that the length of V^- is bigger or equal than $(k-1)t$).

Assume that V^- contains at least k subwords of length $(k-1)$. Hence V^- has not more than k subwords of length k, in any subword of V of length k, the last letter is determined by the $(k-1)$ previous letters. As we have precisely k different subwords of length k, there are two equal subwords of length k among first $(k+1)$ subwords of length k. Let i, j be the first letters of such subwords. We assume that $i > j$. Then i-the letter coincides with j-the letter, $i+1$-th letter coincides with $j+1$-th letter,..., $i+k$-th letter coincides with $l+k$-th letter. Hence $(i-j) \leq k$, V is a subword of $V_{1\to(j-1)}V^{\infty}_{j\to(i-1)}$, where $V_{1\to(j-1)}$ is a subword of V which starts from the first letter and ends in the $(j-1)$-th letter, and $V_{j\to(i-1)}$ is a subword of V which starts from the j-th letter and ends in the $(i-1)$-th letter. As $i - 1 \leqslant k$ and $i - j \leqslant k$, V contains at least t-th power of $V_{j\to(i-1)}$. Therefore V contains a subword of type v^t. □

Problem 2.7. Provide a bijection between the following sets:

- sequences of natural numbers $1 \leqslant a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n$, where $a_i \leqslant i$;
- transpositions of numbers $1, 2, \ldots, n$, such that the length of any decreasing sequence is 2 or less.

Proof. We consider the set of transpositions $S_n^\vee$ of numbers $1, ..., n$ which have no increasing subsequences of length 3 (we consider transpositions as words in alphabet $\{1, ..., n\}$). We denote by $b(\sigma)$ the length of the longest increasing subsequence of σ which ends in the end of σ for any $\sigma \in S_n^\vee$. We denote by $\sigma[m]$ the transposition of $1, ..., m$ which is obtained from σ by eliminating letters $m+1, ..., n$ for any $\sigma \in S_n^\vee$ and $m \leqslant n$.

We assign to any transposition $\sigma \in S_n^\vee$ the sequence of numbers

$$b_1 := b(\sigma[1]), b_2 := b(\sigma[2]), ..., b_n := b(\sigma[n]).$$

Note that $b_1 = 1, b_{i+1} \leqslant b_i + 1, 1 \leqslant b_i \leqslant i$. Set $a_i := i + 1 - b_i$. Note that

$$a_1 = 1,\ 1 \leqslant a_i \leqslant i,\ a_i \leq a_j.$$

Therefore we assign to a transposition $\sigma \in S_n^\vee$, which have no increasing subsequences of length 3, an increasing sequence of numbers $1 \leqslant a_1 \leqslant a_2 \leqslant ... \leqslant a_n$ such that $a_i \leqslant i$. It is easy to see that this map is bijective. As an exercise, a reader can explicitly construct an inverse map. □

The solution of the previous problem is also presented in the project of V. Dotsenko "Katalan's numbers and natural maps".

Problem 2.8. Hundred man-eaters come to a feast. During a feast man-eaters eat themselves. Therefore appear a sequences of man-eaters such that a man-eater eats a man-eater which eats a man-eater which eats a man eater... What is the smallest possible the longest such sequence of man-eaters with additional condition that from any 10 man-eaters any one eats the other one?

Proof. We produce a graph from the set of man-eaters. We assume that two man-eaters A, B are connected with an oriented edge, if A eats B. Note that this graph is a forest (i.e. is a union of trees). We split man-eaters on the several groups. First group consists of man-eaters which does not eat anyone on the feast. Second group consists of man-eaters which eats only man-eaters from the first group. And so on. Obviously

1. All man-eaters are presented in these groups.
2. The man-eaters from one group does not contain themselves.

We know that from any 10 man-eaters one eat the other one. Therefore the number of man-eaters in one group does not exceed 9. Hence the number of such groups is greater or equal than $[\frac{100}{9}] = 12$. Therefore there is a chain of "man-eaters which eats man-eaters" of length 12 or more.

Now we provide an example of a feast where the longest such chain has length 12. We divide man-eaters onto 9 groups of 11 man-eaters and 1 group of 1 man-eater. Let the second man-eater in a any group eat the first one, let the third one eat the second one and so on. At the end, the man eater from the 10-th group eat all other man-eaters. □

Similar problems to Problem 2.8 appear in subsection "Dilworth's theorem".

Definition 2.1. *We call a word u* non-cyclic, *if u is not equal to v^k, for any word v and any $k > 1$.*

Problem 2.9. Let u, v be different non-cyclic words of length m and n respectively. Assume that a word W contains subwords $u' = u^{m \cdot n}$ and $v' = v^{m \cdot n}$. Prove that the length of the common part of u' and v' does no exceed $m + n - 2$.

Solution. Let $m > n$ and assume that the intersection of two periodic subwords u^{mn} and v^{mn} has length $m+n-1$ or more. We denote this intersection S. We denote the letters of S by $s_1, ..., s_l$, where l is a length of S.

We prove that under these assumptions u is periodic with period $d = \text{GCD}(m,n)$. It is enough to prove that if $k = l (\text{mod } d)$ and $1 \leqslant k, l < m+n-1$, then $s_k = s_l$.

We denote by r the remainder of k modulo d. Assume that $r \neq 0$. Then $k-r = an-bm$ for some $a, b \in \mathbb{Z}_{\geq 0}$. We construct sequences k_i, a_i, b_i by the following inductive rules:

1. $k_0 = k, a_0 = a, b_0 = b$;
2. $\begin{cases} k_{i+1} = k_i + n, a_{i+1} = a_i - 1, b_{i+1} = b_i, & \text{if } a_i > 0 \text{ and } k_i < m \\ k_{i+1} = k_i - m, a_{i+1} = a_i, b_{i+1} = b_i - 1, & \text{if } a_i = 0 \text{ and } b_i > 0 \\ i - \text{th element is the last}, & \text{if } a_i = b_i = 0 \end{cases}$.

It is obvious from definition that

1. $k_i - r = a_i n - b_i m$ for all $i \geq 0$;
2. $d \nmid k_i$ for all $i \geq 0$;
3. $1 \leq k_i \leq m+n-1$ for all $i \geq 0$.
4. If k_i is the last element of the sequence, then $k_i = r$.

From these rules follows that $s_{k_{i+1}} = s_{k_i}$, and in particular $s_k = s_l$, if $k = l (\text{mod } d)$.

Assume that $r = 0$. We show that $s_k = s_n = s_m$. To do this we construct sequences k_i, a_i, b_i by the following inductive rules:

1. $k_0 = n$, $a_0 = m/d - 1, b_0 = n/d - 1$;
2. $\begin{cases} k_{i+1} = k_i + n, a_{i+1} = a_i - 1, b_{i+1} = b_i, & \text{if } a_i > 0 \text{ and } k_i < m \\ k_{i+1} = k_i - m, a_{i+1} = a_i, b_{i+1} = b_i - 1, & \text{if } a_i = 0 \text{ and } b_i > 0 \\ i - \text{th element is the last}, & \text{if } a_i = b_i = 0 \end{cases}$.

Note that

1. $k_i - n = a_i n - b_i m$ for all $i \geqslant 0$;
2. $d \nmid k_i$ for all $i \geqslant 0$;
3. $1 \leq k_i \leq m+n-1$ for all $i \geqslant 0$.

Assume that $k_i = n$. Then $na_i = mb_i$, and in particular either $a_i = 0$, or $a_i \geq m/d$. Hence the last statement is false, $a_i = b_i = 0$. On the other hand if $a_i = b_i = 0$, then $k_i = n$. Therefore the sequences $\{a_i\}, \{b_i\}, \{k_i\}$ has precisely $m/d + n/d - 1$ elements.

Now we prove that these elements that the elements of $\{k_i\}$ are pairwise different. Namely if $k_i = k_j$, then $n(a_i - a_j) = m(b_i - b_j)$. Thus $a_i \geq m/d$. This statement is false. Therefore the elements $\{k_i\}$ are pairwise different and belong to $\{d, ..., d(m/d+n/d-1)\}$. The number of these elements equals $m/d+n/d-1$. Therefore the sequence $\{k_i\}$ consists from numbers $d, 2d, ..., (m+n-1)d$. Hence $s_{k_i} = s_{k_{i+1}}$, $s_k = s_n = s_m$.

As a corollary, if $d \mid k, l$ and $k = l$ (mod d), then $s_k = s_l$. As S is lomger than both u and v and is d-periodic, words u, v are periodic to. This is a contradiction.

Therefore the length of the intersection of two periodic words u^{mn} and v^{mn} does not exceed $m+n-2$. □

Problem 2.10. An infinite band is filled by numbers {1,..., 9}. Prove that either one can cut of from it 10 non-intersecting numbers with 1000 digits each, which form an increasing sequence on a band, or there exists a number with 10 or less digits which repeats 50 times in succession.

Proof. We move from left to right on the band and consider a moment from which all 1000-digit numbers appear infinitely many times. If the number of such 1000-digit numbers

exceed 10, then there exists an increasing sequence of length 10 consisting from non-intersecting 1000-digit numbers.

Further we assume that the number of such 1000-digit numbers is smaller than 10. Then there exist equal 1000-digit numbers which intersect by more than 990 digits. Then these numbers are periodic of period 9 or less. Then the period in these numbers repeats at least $[\frac{1000}{9}]$ times (what is definitely more than 50 times). □

3 Dilworth's theorem

Problem 3.1. Is it true that, for any sequence of numbers of length 5, there exists a subsequence of length 3 which is ordered (i.e. it is or increasing, or decreasing)?

Proof. Is a particular case of Problem 3.4. □

Problem 3.2. Is it true that, for any sequence of numbers of length 9, there exists a subsequence of length 4 which is ordered (i.e. it is or increasing, or decreasing)?

Solution. No, this is not true. The sequence 3-2-1-6-5-4-9-8-7 is a counterexample. □

Problem 3.3. Prove that that for any sequence of numbers of length 10 there exists a subsequence of length 4 which is ordered (i.e. it is or increasing, or decreasing).

Proof. Is a particular case of Problem 3.4. □

Problem 3.4. Prove that that for any sequence of numbers of length $mn+1$ either there exists a decreasing subsequence of length $m+1$ or there is an increasing subsequence of length $n+1$.

Proof. We write $a \succ b$, whenever $a > b$ and a goes earlier than b. All other pairs of numbers are incomparable. There exists either a subsequence from these numbers which is a chain with respect to $\succ$ of length $n+1$ (a chain corresponds to an increasing subsequence) or there exists an antichain, i.e. the set of pairwise incomparable numbers, from these numbers with respect to $\succ$ of length $m+1$ (an antichain corresponds to a decreasing sequence) by Problem 3.7. □

By definition a *partially ordered set* (POS) is a set M with a relation $\prec$ on it such that, for any two elements a and b of M, or $a \prec b$ is true or is false. This relation should satisfies the following axioms:

1. if $a \prec b$ and $b \prec c$, then $a \prec c$ (transitivity);
2. if $a \prec b$, then a is not b.

Problem 3.5. May $a \prec b$ and $b \prec a$ be true simultaneously?

Problem 3.6. Prove that the set of words with a lexicographical order is a POS.

Definition 3.1. *A POS M, for any elements $a, b \in A$ of which $a = b$, $a \prec b$, $b \prec a$, is called linearly ordered. Such POS are also known as chains.*

Problem 3.7. Let m, n be natural numbers. Prove that in any POS with $mn+1$ elements there exists either a subset with $m+1$ elements which is a chain or there is a subset with $n+1$ elements which is an antichain (i.e. such that any two elements of which are incomparable).

Proof. We will prove this statement by induction by m.

The base ($m = 0$) is obviously true.

Now we prove that from the m-th statement follows $(m+1)$-th statement. We fix a POS M with at least $mn+1$ element. We say that an element a of M is *maximal* if any other element is smaller than m or is incomparable with m. We consider the set Max of all maximal elements. By definition, any two maximal elements are incomparable. If the number of elements $|Max|$ in Max is greater or equal than $n+1$, then M is a desired antichain.

Further we assume that $|Max|$ does not exceed n.

We denote M without $|Max|$ as $\tilde{M}$. Obviously, $\tilde{M}$ has at least $n(m-1)+1$ elements and therefore has either a chain of length m or an antichain of length $n+1$ by the induction hypothesi. If the second statement is true, we prove the induction hypothesi. Assume that the first statement is true and $\tilde{M}$ has a chain $\mathcal{C}$ of length $m+1$. Then some element a of Max is greater than the maximal element of $\mathcal{C}$ and therefore $\{a\} \cup \mathcal{C}$ is a chain of length $m+1$. □

Problem 3.8. Let M be a POS and $c(M)$ be the length of the longest chain of M. Then M can be splitted on $c(M)$ antichains.

Solution. (We use definitions of Problem 3.7)We prove the statement by induction by n.

The base ($n = 0$) is obviously true.

We denote by Max the maximal elements of M. Let $\tilde{M}$ be M without Max. As M does not contain chains of $n+1$, $\tilde{M}$ does not contain chains of length n. Therefore $\tilde{M}$ is a union of n antichains. As Max is an antichain, M is a union of $n+1$ antichain. We prove the induction hypothesi. □

The following theorem is in some sense dual to Problem 3.8.

Dilworth's theorem. Let M be a POS and $ad(M)$ be the length of the longest antichain of M. Then M can be splitted on $ad(M)$ chains.

4 Exponential estimates

We fix an alphabet $\mathcal{A} = \{a_1, a_2, \ldots, a_l\}$ and we fix a linear order on $\mathcal{A}$: $a_1 \prec a_2 \prec \cdots \prec a_l$. This order introduces a lexicographical order on the set of words of $\mathcal{A}$. We consider two words u and v. If u begins from v or v begins from u, we call u and v *incomparable* (with respect to each other). Otherwise there exist words w, u', v' such that $u = wu'$, $v = wv'$ and first letters of u' and v' are different (w could be an empty word). If the first letter of u' is greater than the first letter of v', we say that u is greater than v and write $v \prec u$, otherwise we say that v is greater than u and write $v \prec u$. The set of words of $\mathcal{A}$ with respect to $\prec$ is a POS. The order $\prec$ is called *lexicographical* (see also Introduction). It would be significant later that some words are incomparable with respect to the lexicographical order $\prec$.

Problem 4.1. Let alphabet $\mathcal{A}$ consists of letters a, b, c. We introduce an order on them: $a \prec b \prec c$. Find the longest increasing sequence from the following list of words. Which pairs of these words are incomparable?

$$cb, abc, bac, abb, b, ccc, abc$$

Solution. The longest increasing sequence is: abb, abc, b, cb, ccc.

The pairs of incomparable elements: $b \leftrightarrow bac, abc \leftrightarrow abc$. □

The following definitions will be useful later.

Definition 4.1. *A word W is called n-divisible, if there exist words $u_1, ..., u_n$ such that $W = v \cdot u_1 \cdots u_n$ and $u_1 \succ \ldots \succ u_n$.*

Definition 4.2. *A word W is called k-ordered, if W is k-divisible but not $(k+1)$-divisible.*

Problem 4.2. Find a number of

a) 1-ordered;

b) 2-ordered words

of length l with pairwise different letters.

c) 1-ordered letters with not necessarily different letters.

Solution. a) Answer: C_l^s.

b) The number of sequences of natural numbers $1 \leqslant a_1 \leqslant a_2 \leqslant \ldots \leqslant a_l$ such that $a_i \leq i$ by Problem 2.7. We call such a sequence a *correct* sequence. Let c_n be the number of correct sequences of length n. For a correct sequence $\{a_i\}$ we set

$$\mathrm{stup}(a_1, a_2 \ldots, a_{n+1}) = \sup_{1 \leqslant i \leqslant n+1} \{a_i = i\}.$$

Obviously, $1 \leqslant \mathrm{stup}(\{a_i\}) \leqslant n$. Then correct sequences such that, for some $1 \leqslant j \leqslant n+1$, $\mathrm{stup}(a_1, a_2 \ldots, a_{n+1}) = j$, can be described by the following conditions:

$$a_i \leqslant i \text{ для } i < j; \quad a_j = j; \quad a_i \leqslant i - 1 \text{ для } i > j.$$

Therefore the number of the correct sequences such that

$$\mathrm{stup}(a_1, a_2 \ldots, a_{n+1}) = j,$$

equals $c_{j-1}c_{n+1-j}$. Thus we have

$$c_{n+1} = c_0c_n + c_1c_{n-1} + \ldots + c_nc_0. \quad (1)$$

Now we show that $c_n = \frac{1}{n+1}\mathrm{C}_{2n}^n$. To do this we define a function

$$f(x) = c_0 + c_1x + c_2x^2 + ... + c_nx^n + ...$$

Then from the relation (1) follows that $f(x) = c_0 + xf^2(x)$. Therefore

$$f(x) = \frac{1 \pm \sqrt{1 - 4c_0x}}{2x} = {}^1\frac{1 \pm \sqrt{1 - 4x}}{2x}.$$

Hence $f(x)$ has no pole in 0,

$$f(x) = \frac{1-(1-4x)^{\frac{1}{2}}}{2x} = \sum_{n\geq 0} \frac{1}{2}(-1)^{n+2}4^{n+1}\mathrm{C}_{n+1}^{\frac{1}{2}}x^n = \sum_{n\geq 0} 4^{n+1}\frac{(2n-1)!!}{2^{n+2}}{}^2x^n = \sum_{n\geq 0} \frac{1}{n+1}\mathrm{C}_{2n}^n x^n.$$

Hence $c_l = \frac{1}{l+1}\mathrm{C}_{2l}^l$.

c) The letters in a 1-ordered word are ordered. Therefore 1-ordered words are one-to-one correspondence with ordered sequences of nonnegative integers $(k_1, ..., k_l)$ such that $k_1 + ... + k_l = s$. It is well known that the number of such sequences equals C_{s+l-1}^s. □

[1] Obviously, c_0=1.

Problem 4.3. a) Let n be a natural number, u — noncyclic word of length n or more. Prove that the word u^{2n} is n-divisible.
b) Let n be a natural number and u be a word of length $n-1$ or less. Prove that the word u^{2n} is not n-divisible.

Proof. a) Let m be the length of u. Then there is m cyclic rotations of u:

$$u[0], u[1], ..., u[m-1].$$

The word u is non-cyclic, and hence words $u[i]$ are pairwise different. Therefore the set $\{u[i]\}$ with respect to the lexicographical order. We reorder them so that

$$u[i_0] \succ u[i_1] \succ ... \succ u[i_{m-1}].$$

For any i there exist words u_i, w_i such that $u = u_i w_i$ and $u[i] = w_i u_i$. Then

$$u^{2n} = w_{i_0} v_{i_0} w_{i_0} v_{i_0} w_{i_1} v_{i_1} w_{i_1} v_{i_1} \dots w_{i_{m-1}} v_{i_{m-1}} w_{i_{m-1}} v_{i_{m-1}}.$$

Set

$$\begin{cases} u'_{i_k} = v_{i_k} w_{i_k} v_{i_k} w_{i_{k+1}} & k = 0, 1, \dots, n-2 \\ u'_{i_k} = v_{i_k} w_{i_k} v_{i_k} & k = n-1 \end{cases}$$

$\gamma = w_{i_0}$. Then $u^{2n} = \gamma u'_{i_0} u'_{i_1} \dots u'_{i_{n-1}}$. We have $u'_{i_0} > u'_{i_1} > \dots > u'_{i_{n-1}}$, and therefore u^{2n} is n-divisible (see also [18]).

b) Let $u = u_1 ... u_s$, where $s \leq n-1$ and u_i are letters. Assume that u^{2n} is n-divisible, i.e. u contains nonintersecting words $v_1, ..., v_n$ such that v_i goes before v_j and $v_1, ..., v_n$ is a decreasing sequence. Let $r_1, ..., r_n$ be the numbers of the first letters (we count from the left side!) of v_i in u^{2n}. We have $s < n$, and hence there exist i, j such that $1 \leqslant i < j \leqslant n$ and $r_i = r_j (\mathrm{mod}\ s)$. Then either v_i is a subword of v_j, or v_j is a subword of v_i. Anyway v_i and v_j are incomparable or equal one to each other. As $v_j \succ v_i$, this is a contradiction. Therefore u^{2n} is not n-divisible. □

Definition 4.3. *a) A word v is called a tale of a word u if there exists a word w such that $u = vw$.*

b) If a word v has a subword u^t we say that v has a period of cyclicity t

Problem 4.4. Let l, d be some natural numbers. Prove that, for a word W of length l, or first $[l/d]$ tales are pairwise incomparable, or W has a period of length d.

Solution. Assume that W does not contain a subword u^d for any non-zero u. Let Ω_d be the first $[x/d]$ tales. Assume that $v_1, v_2 \in \Omega_d$ are two incomparable tales. Without loss of generality we assume that $v_1 = uv_2$. Then both v_1, v_2 are subwords of u^∞. As $v_2 \in \Omega_d$ and the length of u does not exceed d, W contains u^d (see also [15, lemma 2.1]). □

Further we assume that $n \leqslant d$.

Definition 4.4. *A word W is called* (n, d)-cancellable, *either if W is n-divisible, or if there exists a word u such that u^d is a subword of W.*

Problem 4.5. Prove that if a word W has n pairwise equal non-intersecting subwords of length n, then W is (n, n)-cancellable.

Definition 4.5. *A word W is called n-divisible from tale, is there exists tales $u_1, \dots, u_n$ such that $u_1 \succ u_2 \succ \dots \succ u_n$, and, for any $i = 1, 2, \dots, n-1$, u_i begins earlier than u_{i+1}.*

Problem 4.6. Prove that if a word W is
a) n^3d-divisible from tale,
b) $3n^2d$-divisible from tale,
c)** $4nd$-divisible from tale,
then W is either n-divisible, or W has a subword u^d for some nontrivial word u.

Solution. Parts a), follows from b).

Assume on the contrary that such a word exists. We consider the numbers of the positions of tales a_i, where $a_1 < a_2 < \dots < a_{3n^2d}$, such that tales which starts from u_i divide W. Let

$$X_k = \{nd - \text{tales } u_i \mid i = 3knd + 1, \dots, 3knd + 2nd\}.$$

Then for any numbers i and j, if $u \in \chi_i, v \in X_j$, then u and v does not intersect. Therefore there exists k such that, for any $u, v \in X_k$, if $u \cap v = 0$, then u and v are incomparable. Without loss of generality we assume that $k = 1$. Assume that a subword v_i is $n \cdot d$-tale u_i. The subwords v_1 and v_{nd+c} does not intersect for $c \in [1, nd]$. Thus $v_{nd+s} = v_{nd+t}$ for any $1 \leqslant s \leqslant t \leqslant nd$. As $a_{nd+t} - a_{nd+s} > n$, then the subwords

$$u_1, u_{nd+1}, u_{nd+d+1}, u_{nd+2d+1}, \dots, u_{2nd-d+1}$$

does not intersect. Therefore they are incomparable, and hence a word W is n-divisible. This is a contradiction. □

Problem 4.7. For any pair of natural numbers (n, d) (except pair (1, 1)), provide an example of a word W of length $(nd - 1)$ such that any set of tales of W is not increasing and W is not $(n + 1, d)$-cancellable.

Solution. Two teams from four provide the following example:

Fix an alphabet $\mathcal{A}_2 := \{a < b\}$. The word

$$W = (a^{n-1}b)^{d-1}a^{n-1}$$

has length $(nd - 1)$, W is not $(n + 1, d)$-cancellable and any set of tales of W is not increasing. □

Problem 4.8. Try to enhance an estimate from Problem 4.6.

Comment. If such estimate exists, then it does not exceed $(n-1)(d-1)$, because for an alphabet $\mathcal{A}_s := \{a_1 \prec a_2 \prec \dots \prec a_{n-1} \prec \dots \prec a_l$ the word $a_1^{d-1}a_2^{d-1}\dots a_{n-1}^{d-1}$ is $(n-1)(d-1)$-divisible from tale, but is not n-divisible and has no subword of period d. □

We fix an alphabet $\mathcal{A}$ of length l, a word W of this alphabet of length $r(W)$ and natural numbers $n \leqslant d$.

Further we assume that, for all words u, W does not contain a subword u^d and W is not $4nd$-divisible from tale. We consider first $[r(W)/d]$ tales of W (further we denote this set of words by Ω). Then by Dilworth's theorem we can split Ω on $(4nd - 1)$ groups such that tales in one group form a chain.

In solutions of the following problems we expect that you use previous ones.

Problem 4.9. Prove that, from any $4nd^2$ tales of Ω, there exists two tales, for which the first subwords of length $4nd$ are pairwise different.

Solution. Proof is presented in [15, начало §3]. □

Problem 4.10. In some infinite parliament any member has not more than 3 enemies among members. Prove that there is a way to divide this parliament into two houses such that a member of any house has not more than 1 enemy among members of his own house.

Proof. Without loss of generality [3] we assume that the number of memebers in a parliament is countable.

We enumerate the members of parliament by natural numbers. We denote by P_n the set of members with numbers $1, ..., n$. We understand these houses us a function $f : P_n \to \{1, 2\}$ (any member of a parliament is a member of the first house or of the second house). We denote the set of such functions with an additional condition that any member has not more than 1 enemy in his own house by H_n (we call such maps admissible).

Obviously, the restriction of any admissible function to a subset is admissible. We denote these restriction from P_m to $P_n (m > n)$ by $|_n$.

First we show that H_n is nonempty for any n. We start with some function and if a member has 2 or more enemies in his own house we move him to another house (then the number of the pairs of enemies inside the houses decreases). As the number of the pairs of enemies in P_n is finite, after several such procedures we obtain an admissible function. Therefore H_n is nonempty for any n.

Set $(H_\infty)|_n := \cap_{m \geq n}((H_m)|_n)$. As,

1. for any $m > n$, $(H_m)|_n$ is a non-empty finite subset of P_n, and
2. if $m_1 > m_2 > n$, we have $(H_{m_1})|_n \subset (H_{m_2})_n$,

the set $(H_\infty)_n$ is nonempty. We build a chain of functions $\{f_i\} (f_i \in (H_\infty)_i)$ by the following rule: $f_{i+1}|_{P_i} = f_i$ (for any $f_i \in (H_\infty)|_i$ such a function f_{i+1} always exist, because $((H_\infty)_{i+1})|_i = (H_\infty)_i$). This chain define two houses: i-th memeber is a memeber of $f_i(i)$-th house. □

In the following problems we assume that l, n, d are variables and that W is some word defined over alphabet of l-letters.

Problem 4.11 (Shirshov's lemma)**.** Prove that there exists a function $f(l, n, d)$ such that for any subword W defined over alphabet of l-letters either W is (n, d)-cancellable, or $r(W) < f(l, n, d)$.

Proof. Proof of Shirshov's lemma follows from Problem **??**. The original Shirshov's proof was published in [18]. □

Problem 4.12. Prove that $f(l, n, d) < l(4nd)^{4nd+2}$.

Proof. Proof of this problem follows from Problem 5.5. □

Solutions of Problem's 5.1 — 5.5 are presented in [15, §3], problem 5.6 решена в параграфах 4 и 5 той же статьи.

[3] Here we use an the axiom of choice

References

[1] М. И. Харитонов *Оценки на структуру кусочной периодичности в теореме Ширшова о высоте*, Вестник Московского университета, Серия 1, Математика. Механика. є6(2012).

[2] Abraham A. Klein. *Indices of nilpotency in a PI-ring.* Archiv der Mathematik, 1985, vol 44:4.

[3] Abraham A. Klein *Bounds for indices of nilpotency and nility.* Archiv der Mathematik, 2000, vol 74:1 pages 6—10

[4] Е. С. Чибриков. *О высоте Ширшова конечнопорождчнной ассоциативной алгебры, удовлетворяющей тождеству степени четыре.* Известия Алтайского государственного университета т. 1(19), 2001, стр. 52—56

[5] М. И. Харитонов *Двусторонние оценки существенной выосты в теореме Ширшова о высоте*. Вестник Московского университета, Серия 1, Математика. Механика., 2(2012), 24—28.

[6] A. A. Lopatin. *On the nilpotency degree of the algebra with identity* $x^n = 0$. arXiv:1106.0950v1.

[7] *Днестровская тетрадь: оперативно-информац. сборник* No 4, Новосибирск, изд. ин–та матем. СО АН СССР, 1993, 73 стр.

[8] И. И. Богданов *Теорема Нагаты-Хигмана для полуколец*. Фундамент. и прикл. матем. т. 7:3, 2001, 651—658.

[9] Колотов А. Г. *О верхней оценке высоты в конечно порожденных алгебрах с тождествами.* Сиб. мат. ж., 1982, т. 23, no 1, стр. 187—189.

[10] Кострикин А. И. *Вокруг Бернсайда. — М.: Наука, 1986*, 232 стр.

[11] Курош А. Г. *Проблемы теории колец, связанные с проблемой Бернсайда о периодических группах.* Изв. АН СССР, сер. мат., 1941, т. 5, стр. 233—240.

[12] Латышев В. Н. *К теореме Регева о тождествах тензорного произведения* PI-*алгебр.* Успехи мат. наук, 1972, т. 27, no 4, стр. 213—214.

[13] Ширшов А. И. *О некоторых неассоциативных нильь-кольцах и алгебраических алгебрах.* Мат. сб., 1957, т. 41, no 3, стр. 381—394.

[14] Ширшов А. И. *О кольцах с тождественными соотношениями.* Мат. сб., 1957, т. 43, no 2, стр. 277—283.

[15] А. Я. Белов, М. И. Харитонов. *Субэкспоненциальные оценки в теореме Ширшова о высоте.* Мат. сб., **203**:4(2012), 81—102.

[16] Belov A. *Some estimations for nilpotence of nil-algebras over field of an arbitrary characteristics and height theorem.* Comm. in Algebra, 1992, vol. 20, N 10, p. 2919—2922.

[17] А. Спивак. *Цепи и антицепи.* Квант, 5(2003), 11—14.

[18] Жевлаков, Слинько, Шестаков, Ширшов. *Кольца близкие к ассоциативным.* М., Наука, 1978.

PERSPECTIVES ON SHIRSHOV'S HEIGHT THEOREM

ALEXEI KANEL-BELOV AND LOUIS H. ROWEN

In this survey we consider the impact of Shirshov's Height Theorem on algebra. In order to avoid duplication, we often refer to Kemer's survey article [Kem09] in this volume for further details. Proofs of various quoted results are given in the book [BBL97], and in the authors' book [BR05].

1. Historical background to Shirshov's Theorem

Let F denote a field. An F-algebra is called *affine* if it is finitely generated as an algebra. An F-algebra is *algebraic* if each element a satisfies an algebraic equation over F; i.e., if the dimension $[F[a] : F] < \infty$. We say that an algebra A has PI-*degree* n if A satisfies a multilinear polynomial identity (PI) of degree n. One of the early tests of the utility of PI-theory was whether it could provide a framework for a positive solution of the following famous problem of Kurosh:

Are affine algebraic algebras necessarily finite dimensional?

Although now known to be false for associative algebras in general (cf. [Gol64]), Kurosh's problem was solved for associative PI-algebras by Kaplansky [Kap50], building on work of Jacobson and Levitzki, as described in [Kem09]. However, Kaplansky's elegant proof, relying on topology and structure theory, is not constructive.

Digression. In hindsight, Kurosh's problem for PI-algebras has an easy solution using standard results from structure theory. Here is a modification of the argument given in [Pro73]. By [Pro73, Lemma 2.6], if A is not finite dimensional, there is a prime ideal P maximal with respect to A/P not being finite dimensional, so we may assume that A is a prime affine algebraic PI-algebra. But then the center C of A is a field, so A is simple, by [Row88, Corollary 6.1.29], and thus finite dimensional over C, by Kaplansky's Theorem. Then a version of the Artin-Tate Lemma [Row88, Proposition 6.2.5] says the field C is affine and thus finite dimensional over F, implying R is finite dimensional over F. (This argument also works more generally for affine algebras integral over a commutative Noetherian ring.)

A different approach to Kurosh's problem, taken by A. I. Shirshov [Shir57a], [Shir57b], involves the detailed analysis of words and their relations, as given in **Shirshov's Height Theorem**:

Let A be a finitely generated algebra of PI-degree d. Then there exists a finite set $Y \subset A$ and an integer $H \in \mathbb{N}$ such that A is linearly spanned by the set of elements of the form

$$v_1^{k_1} v_2^{k_2} \cdots v_h^{k_h} \quad \text{where } h \leq H,\ v_i \in Y.$$

This research was supported by the Israel Science Foundation, grant #1178/06. The authors would like to thank L. Bokut, A. Kemer, E. Zelmanov, and U. Vishne for helpful comments on drafts of this survey.

For Y we may take the set of words of length $\leq d$. Such Y is called *a Shirshov base* of the algebra A.

The object of this survey is to describe the impact of this pioneering theorem. Shirshov's Theorem immediately yields an independent positive solution of Kurosh's problem and of other related problems for PI-algebras. Specifically, if Y is a Shirshov base consisting of algebraic elements, then the algebra A is finite-dimensional. Thus, Shirshov's theorem explicitly determines the set of elements whose algebraicity implies algebraicity of the whole algebra. (It is worth noting that Procesi [Pro73] later discovered a structural proof of Shirshov's Theorem also, by means of reducing first to prime rings and then utilizing traces). We also have

Corollary 1.1. *If A is a PI-algebra of PI-degree d and all words in its generators of length $\leq d$ are algebraic, then A is locally finite.*

Let us briefly sketch the proof of Shirshov's Theorem. Suppose that $A = F\{a_1, \dots, a_\ell\}$ is an affine algebra. Ordering the letters $a_1 < \cdots < a_\ell$ induces the *lexicographic* order on the set Ω^* of words in the generators $\{a_1, \dots, a_\ell\}$. We consider this as a total order, where a proper subword v of a word w is defined to precede w. But note that this order is not preserved under multiplication; for example $a_2 \prec a_2a_1$ but $a_2^2 \succ (a_2a_1)^2$. A word w is *reducible* if it can be written as a linear combination of smaller words.

Definition 1.2. *A word w is called* d-decomposable *if it contains a subword $w_1 \cdots w_d$ such that $w_1 \cdots w_d \succ w_{\pi(1)} \cdots w_{\pi(d)}$ for any permutation π of $\{1, \dots, d\}$.*

A (multilinear) PI of degree d can be used to rewrite any d-decomposable word as a sum of smaller words; thus, the irreducible words are d-indecomposable. Shirshov proved *Shirshov's Lemma*, which asserts that, for any given $r > 0$, any long enough d-indecomposable word must contain a nonempty word u^r where $|u| \leq d$. Shirshov's height theorem follows from an algorithmic argument given in [BR05, p. 50].

Shirshov's Height Theorem also yields a result about the *Gelfand-Kirillov dimension* $\mathrm{GK}(A)$ of an affine algebra A. Recall that

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

where $V_A(n)$ is the vector space generated by the words of length $\leq n$ in the generators of A. A related concept is the *(Poincaré-)Hilbert Series*

$$H_A = 1 + \sum d_n \lambda^n,$$

where $d_n = \dim(V_A(n)/V_A(n-1))$, the number of irreducible words of length n. (Strictly speaking, H_A depends on the given set of generators of A, whereas $\mathrm{GK}(A)$ is independent of the choice of generators.)

Corollary 1.3 (Berele [Ber93])**.** $\mathrm{GK}(A) < \infty$*, for any affine PI-algebra A.*

To prove the corollary, it suffices to observe that the number of solutions of the inequality $k_1|v_1| + \cdots + k_h|v_h| \leq n$ with $h \leq H$ does not exceed N^H, and therefore $\mathrm{GK}(A) \leq h(A)$.

Shirshov's beautiful theorem, which also is formulated for algebras over arbitrary commutative rings, opened the way to the combinatoric school of PI-theory, which has led to many breakthroughs in recent years. (Ironically, Shirshov's work was unknown

in the West until 1973. Thus, for many years, there was a parallel development of PI-theory on both sides of the former "iron curtain," along mostly combinatoric lines in the former Soviet Union and along structural lines in the West. Although our focus in this survey is on Shirshov's influence, and thus on the Russian school, we also describe parallel results in the West.)

1.1. **The radical of an affine PI-algebra and the Nagata-Higman Theorem.** One of the early applications of Shirshov's Theorem was in a seemingly unrelated direction. Using structure theory, Amitsur [Am57] showed that the Jacobson radical $J(A)$ of an affine PI-algebra is nil. This led to the question of whether $J(A)$ is nilpotent, which was formally raised by Latyshev in his dissertation. Shirshov's Theorem is a key tool in verifying this assertion when R satisfies the PI's of $n \times n$ matrices, as shown by Razmyslov [Raz74a], who also proved that a complete solution is equivalent to the conjecture that every affine PI-algebra satisfies the *standard* PI. Kemer[Kem80] verified this latter conjecture in characteristic 0. Braun [Br84] was the first to prove the nilpotence of $J(A)$ for arbitrary affine A, using the structure of Azumaya algebras. A nice exposition of Braun's theorem can also be found in Lvov[Lv83].

Incidentally, much earlier, Dubnov and Ivanov, and independently, Nagata and Higman [Hig56] showed that in characteristic 0, any nil algebra of bounded index n is nilpotent. The original bounds for the nilpotence index were exponential in n. Better bounds have been obtained as an outgrowth of Shirshov's work. Razmyslov [Raz74b] showed that n^2 is an upper bound, and Kuzmin obtained the lower bound $\frac{n^2+n-2}{2}$, described in [BR05, p. 341].

1.2. **Representable algebras.** An F-algebra is called *representable* if it can be embedded into $M_n(K)$ for some field extension $K \supset F$ and some n. (More generally, we can take K commutative Noetherian, in view of [An92].) Shirshov's Theorem implies that for any representable affine PI-algebra A, one may adjoin the characteristic coefficients of finitely many words of the generators, to obtain a PI-algebra $\hat{A}$, called the *trace ring* or *characteristic closure*, which is finite over its center but also possesses a nonzero ideal contained in A. The use of this "conductor" ideal, discovered by Razmyslov[Raz74a] (and later, independently, by Schelter [Sch76]) is one of the keys to the structure of affine PI-algebras, and is used in Razmyslov's work on the Jacobson radical described above.

Another application of the characteristic closure is to the Hilbert series of an algebra; Answering a question raised by Procesi [Pro73], Belov proved that any relatively free, affine PI-algebra has a rational Hilbert series (with respect to a suitable set of generators); cf. [BR05, Chapter 9] for this and related results. On the other hand, Theorem 3.1 below provides examples of representable algebras with non-rational Hilbert series.

1.3. **Specht's conjecture.** One of the most famous problems in PI-theory was Specht's conjecture, that every set of identities is a consequence of a finite set of identities. (More formally, every T-ideal of the free algebra is finitely generated as a T-ideal.) As described in [Kem09], this question was settled affirmatively by Kemer [Kem87], [Kem90b] whenever the base field F is infinite, and later by Belov for arbitrary affine PI-algebras. The characteristic closure is one component of the proofs, and the nilpotence of the radical is another important aspect, so Shirshov's theorem plays an important role.

The key step of Kemer's theorem is that each affine PI-algebra over an infinite field satisfies the same PI's as a suitable finite dimensional algebra; it follows at once that the corresponding relatively free algebra is representable. (Belov extended this fact to arbitrary fields.)

2. Generalizations to nonassociative algebras

Shirshov's Height Theorem has been extended to various classes of nonassociative algebras. In his original paper, Shirshov applied his theorem to special Jordan algebras. Zelmanov [Zel91] obtained the following analog for ad-identities of Lie algebras:

Say an associative word in X is **special** if it is the leading word appearing in some Lie word (i.e., word with respect to the Lie multiplication). The word w is **Zelmanov d-decomposable** if it can be written as a product of subwords $w = w'w_1w_1'w_2w_2'\cdots w_dw_d'w''$ with each w_i special and $w_1 \succ w_2 \cdots \succ w_d$. Then, for any ℓ, k, d, there is $\beta = \beta(\ell, k, d)$ such that any Zelmanov d-indecomposable word w of length $\geq \beta$ in ℓ letters must contain a nonempty subword of the form u^k, with u special.

Zelmanov's result is a major ingredient in his celebrated solution of the restricted Burnside problem. S. P. Mishchenko [Mis90] obtained an analogue of Shirshov's Height Theorem for Lie algebras with a "sparse" identity. S. V. Pchelintsev [Pch84] proved an analog for alternative and $(-1, 1)$ cases. Belov [Bel88b] proved a version for a certain class of rings asymptotically close to associative rings, including alternative and Jordan PI-algebras.

3. Questions arising in connection with Shirshov's Theorem

Shirshov's Height Theorem also gives rise to various notions, which we examine in turn.

3.1. **d-decomposable words.** We start with d-decomposable words; cf. Definiton 1.2. An equivalent definition: A word w is d-decomposable if it has the form $s_0v_1s_1v_2\ldots s_{-1}v_ds_d$ where $v_1 \succ v_2 \succ \cdots \succ v_d$. The next proposition below demonstrates the importance of the notion of d-decomposability.

Proposition 3.1 (A. I. Shirshov)**.**

a) Suppose that a word w is d-decomposable. Then any word obtained from w by means of a nonidentical permutation is lexicographically less than w.

b) If an algebra A satisfies a PI

$$x_1\cdots x_d = \sum_{\sigma\neq id\in S_d} \alpha_\sigma x_{\sigma(1)}\cdots x_{\sigma(d)}$$

of degree d, then any d-decomposable word w can be written as a linear combination of words of lower order.

Thus in an algebra of PI-degree d, any word not representable as a linear combination of lower-order words is not d-decomposable, and it suffices to check that the set of d-indecomposable words has bounded height.

3.1.1. *d-decomposable words and codimensions.* Regev [Reg72] introduced the *codimension sequence* in order to prove that the tensor product of PI-algebras is a PI-algebra. Namely, let W_n denote the F-space of multilinear polynomials in $x_1, \dots, x_n$, and

$$c_n = \dim_F(W_n/(W_n \cap \mathrm{id}(A));$$

then c_n is exponentially bounded, for any PI-degree n.

A theorem of Dilworth enables one to bound the number of d-indecomposable words of length n by $n^{2(d-1)}$. Latyshev [Lat72] discovered a quicker proof of Regev's tensor product theorem by using Dilworth's Theorem, and showing that $c_n(A)$ is bounded by the number of d-indecomposable multilinear words. This estimate of the codimension series led to the result of Kemer, Regev, and Amitsur that any polynomial identity whose Young tableau contains a rectangle (whose size is a suitably large function of n) is a consequence of any given polynomial identity of degree n. (This is the basis of Kemer's "super-trick" to pass from identities of nonaffine algebras to identities of affine superalgebras.)

On the other hand, there is an interesting refinement of the Hilbert series. The *multivariate Poincaré-Hilbert series* of an affine algebra $A = F\{a_1, \dots, a_\ell\}$ is defined as

$$H(A) = \sum d_{\mathbf{i}} \lambda_1^{i_1} \cdots \lambda_\ell^{i_\ell},$$

where

$$d_{\mathbf{i}} = \dim_F \left(\bar{V}_A(\mathbf{i})\right);$$

here $\mathbf{i} = (i_1, \dots, i_\ell)$, and $\bar{V}_A(\mathbf{i})$ is the vector space spanned by irreducible words of length $\leq i_u$ in the generator a_i of A, for $1 \leq u \leq \ell$.

Kemer [Kem95, §2] proved that the number of d-indecomposable multilinear words of length n equals the codimension of the space of multilinear polynomials of degree n, with traces, of $M_d(F)$. By Formanek [For84], this codimension sequence can be calculated precisely, using the multivariate Hilbert series.

Thus, Shirshov's approach motivates the use of combinatorics to compute codimensions, and to introduce the use of invariants of matrices. In this regard, Razmyslov [Raz74b], Helling[Hel74], and Procesi[Pro76], independently showed in characteristic 0 that every PI is a consequence of the Hamilton-Cayley equation (which can be written as a trace identity). This follows from the two *Fundamental Theorems of Invariant Theory*, which respectively are as follows:

- All invariants can be expressed in terms of traces.
- All relations between invariants are consequences of the Hamilton-Cayley trace identity.

In characteristic $p > 0$ one must study all of the coefficients of the Hamilton-Cayley equation as individual functions, arising from homogeneous forms (not necessarily linear), since they cannot be computed in terms of the trace. Kemer [Kem90b] developed the theory of identities involving these forms. Donkin[Do94] proved the analog of the First Fundamental Theorem of Invariant Theory, and Zubkov [Zubk96] proved the analog of the Second Fundamental Theorem.

In a similar vein, Razmyslov's student Zubrilin developed the technique of incorporating coefficients of the characteristic polynomial into Capelli polynomials, which leads to a combinatoric proof of the Razmyslov-Kemer-Braun theorem, as exposed in [BR05, §2.5].

Kemer [Kem95] showed that, unlike the situation in characteristic 0, any PI-algebra A (not necessarily affine) of characteristic $p > 0$ satisfies all the multilinear identities of a finite dimensional algebra; combining this with the cited work of Donkin, Zubkov, and Zubrilin, yields that A satisfies all PI's of a finite dimensional algebra; cf. [Bel00].

3.2. **Estimates of Shirshov height.** Shirshov's original proof was purely combinatorial (based on an elimination technique he developed for Lie algebras), but did not provide a reasonable estimate for the height. Kolotov [Kol81] obtained an estimate for $ht(A) \leq s^{s^m}$ (m = PI-deg(A), and s is the number of generators). In the Dniester Notebook (most recent version [Dne93]), Zelmanov asked for an exponential bound, which was obtained later by Belov [Bel88a]:

Theorem 3.1. *Suppose A is a PI-algebra of PI-degree d, generated by ℓ elements. Then the height of A over the set of words having length $\leq m$ is bounded by a function $H(m, \ell)$ where $H(m, \ell) < 2m\ell^{m+1}$.*

3.2.1. *Burnside-type problems.* A word $w = u^k$, for $k > 1$, is called *cyclic* or *periodic*. By problems of **Burnside type**, we mean problems related to periodic words. Combinatorics play an important role. The following basic lemma yields computational tools involving subwords which are described in [Bel07] and provide the bounds given in Theorem 3.1. The technique is illustrated in the slightly weaker result given in [BR05, Theorem 2.74].

Lemma 3.2 (on overlapping)**.** *If two periodic words of respective periods m and n contain identical subwords having length $m + n - gcd(m, n)$ then they have identical periods.*

3.3. **The essential height of an algebra.**

Definition 3.3. *An algebra A is said to have* essential height $\leq h$ *over a subset Y, if there is a finite set $S \subset A$ (which may depend on Y) such that A is spanned as a vector space by*

$$Y^{[h],S} = \{s_0 y_1^{m_1} s_1 \cdots s_{t-1} y_t^{m_t} s_t : \ m_i \in \mathbb{N}, \ y_i \in Y, \ s_i \in S, \ t \leq h\}.$$

In this case, Y is called an essential Shirshov base*, and S the* supplementary *set.*

Essential height is an estimate for GK-dimension; also, the converse is true for representable algebras.

Theorem 3.2 (A. Ya. Belov [BBL97])**.** *Suppose A is a finitely generated representable algebra and $H_{EssY}(A) < \infty$. Then $H_{EssY}(A) = \mathrm{GK}(A)$.*

This equality is useful in both directions. First of all, it shows for a representable algebra A that that $H_{EssY}(A)$ independent of the choice of Y. In the other direction, since $H_{EssY}(A)$ must be an integer, one has:

Corollary 3.4 (V. T. Markov)**.** *The Gelfand-Kirillov dimension of a representable affine algebra is an integer.*

Due to the representability of relatively free affine algebras (noted above), the Gelfand-Kirillov dimension of a relatively free algebra also equals the essential height.

Clearly, an s-base is a Shirshov base iff it generates A as an algebra. Boundedness of essential height over Y implies a positive solution of "Kurosh's problem over Y." The converse is much less trivial.

Theorem 3.3 (A. Ya. Belov)**.** *Suppose A is a graded PI-algebra, and Y is a finite set of homogeneous elements. Let $Y^{(n)}$ denote the ideal generated by all nth powers of elements of Y. If the algebra $A/Y^{(n)}$ is nilpotent for each n, then Y is an s-base for A. If in this situation Y generates A as an algebra, then Y is a Shirshov base for A.*

We proceed to formulate a generalization of this theorem for the non-graded case. We must confront the following counterexample to the straightforward converse of Kurosh's problem: Suppose $A = \mathbb{Q}[x, 1/x]$. Each projection π such that $\pi(x)$ is algebraic has finite-dimensional image. Nevertheless the set $\{x\}$ is not an s-base for A.

Thus we need a stronger definition:

Definition 3.5. *A set $M \subset A$ is called a Kurosh set if it satisfies the condition that for any projection $\pi\colon A \otimes K[X] \to A'$, if the image $\pi(M)$ is integral over $\pi(K[X])$, then $\pi(M)$ is finite over $\pi(K[X])$.*

Theorem 3.4 (A. Ya. Belov)**.** *Let A be a PI-algebra, $M \subseteq A$ a Kurosh subset in A. Then M is an s-base for A.*

Thus, boundedness of essential height is a non-commutative generalization of integrality. The following proposition shows that Theorem 3.4 does generalize Theorem 3.3:

Proposition 3.6. *Let A be a graded algebra, Y a set of homogeneous elements. If the algebra $A/Y^{(n)}$ is locally nilpotent for all n, then Y is a Kurosh set.*

3.4. **Normal bases and monomial algebras.** Shirshov's combinatoric approach leads us to the combinatoric study of bases. Let $A = F\{a_1, \dots, a_\ell\}$ be an associative affine algebra. A words is called *reducible* if it can be written as a linear combination of lexicographically smaller words; the *normal base* of the algebra A is the set of all irreducible words in the generators; cf. [BBL97], [BRV06], [Dr00], [Lat88], [Ufn85].

A **monomial algebra** is an algebra that can be described in terms of relations that are monomials in the generators. Any affine algebra A has its *associated monomial algebra* possessing the same Hilbert series; namely one factors the free algebra by the set of reducible words in the generators of A, cf. [BR05, Proposition 9.8]. The associated monomial algebra of an algebra A also has the same Shirshov base, although it may not satisfy the same PI's. Nevertheless, their easier relations make monomial algebras a useful tool in studying Shirshov bases. This discussion follows [BRV06]; the reader should also consult [BBL97].

In case an affine monomial algebra A is PI, it has bounded essential height over a (finite) Shirshov base Y, which we may assume to be a set of words in the generators. Take a supplementary set S as in Definition 3.3 that contains Y. Choose a subset of $Y^{[h],S}$ that spans A. Given

$$w = s_0 y_1^{m_1} s_1 \cdots s_{t-1} y_t^{m_t} s_t \tag{1}$$

(with $y_i \in Y$ and $s_i \in S$, and t bounded by the height), we rewrite it in the same manner with $s_0 \in S$ of maximal possible length, then with $y_1^{m_1}$ of maximal possible length, and so on. $(s_0, y_1, s_1, \dots, s_{t-1}, y_t, s_t)$ is called the *type* of w. The type of a subword of a w of type θ is called a *subtype* of θ.

By an *exponential polynomial* in the variables $m_1, \dots, m_t$ we mean an expression of the form

$$\sum f_j(m_1, \dots, m_t) \alpha_{1j}^{m_1} \cdots \alpha_{tj}^{m_t}$$

where f_j are polynomials over a finite algebraic extension K of F, and $\alpha_{ij} \in K$. For example,

$$P(m_1, \dots, m_t) = (5 - \sqrt{2})^{m_1} - m_2^4 \cdot 3^{m_1}$$

is an exponential polynomial over $\mathbb{Q}$.

Theorem 3.1. *A monomial algebra A over F is representable iff:*

(1) *A has essential height over a finite set Y (with a supplementary set S), such that every word in the generators of A has a unique type, and there are finitely many types.*
(2) *For each type $\theta = (s_0, y_1, s_1, y_2, \dots, y_t, s_t)$, there is a finite system $P_{\theta,j}$ of exponential equations over k, in the variables $m_1, \dots, m_t$, such that*

$$\bigcup_\theta \{s_0 y_1^{m_1} s_1 \cdots y_t^{m_t} s_t : \quad \exists j \ P_{\theta,j}(m_1, \dots, m_t) \neq 0\}$$

is a normal base.

The construction of monomial algebras is thus equivalent to the solution of arbitrary exponential polynomials. But this is algorithmically unsolvable by the celebrated theorem of Davis-Putnam-Robinson [DPR61]. Thus there is no algorithm to determine whether there is an isomorphism (given in terms of the generators) for two monomial subalgebras of the matrix algebra over a polynomial ring of characteristic 0. On the other hand, this isomorphism problem is algorithmically solvable in characteristic $p > 0$. More precisely, Belov and Chilikov [BC00], [BRV06] proved over a field of characteristic p that the set of p-adic representations of exponential equations (with unknowns in $\mathbb{N}$) forms a "regular language." Thus, an inaccessible problem in characteristic 0 becomes algorithmically solvable in positive characteristic.

3.5. **The conjecture of Amitsur and Shestakov.** S. Amitsur and I. P. Shestakov conjectured that if the algebra A satisfies the identities of $M_n(F)$ and all words having length not exceeding n are algebraic, then A is finite-dimensional. I. V. Lvov reduced this assertion to the following:

Let $A = F\{a_1, \dots, a_\ell\}$ be a finite-dimensional subalgebra (without 1) of a matrix algebra of order n. If all words in $a_1, \dots, a_\ell$ of length $\leq n$ are nilpotent, then the algebra A is nilpotent.

Shestakov's conjecture was proved by V. A. Ufnarovsky [Ufn85] and by G. P. Chekanu [Che88]. Their *Independence Theorem* may be formulated as follows [Che88], [Ufn90]:

Theorem 3.5 (Independence Theorem)**.** *Suppose the following is true:*

(1) *a word $w = a_{i_1} \cdots a_{i_n}$ is minimal under the lexicographical order in the set of all nonzero products of length n;*
(2) *all terminal subwords of w are nilpotent.*

Then the initial subwords of w are linearly independent.

Here is a key step. A word is called *extremal* if it does not lexicographically precede any nonzero word.

Lemma 3.7. *Any set of pairwise incomparable subwords of an extremal word is independent.*

To deduce I. P. Shestakov's conjecture (or, equivalently, I. V. L'vov's assertion) from this theorem, we consider the following construction:

Remark 3.8. *Given an algebra A and a right module V, the algebra $\widetilde{A}$ is defined additively as $A \oplus V$, with multiplication defined as follows: $V \cdot V = A \cdot V = 0$, and the product of elements from V and A is given by the module multiplication.*

We take a faithful representation of A acting on an n-dimensional right vector space V. Taking a base $v_1, \dots, v_n$ of this space, then, for some v_i we have $v_i w \neq 0$. Viewing V as a right A-module, we form the algebra $\tilde{A}$ of Remark 3.8, ordering the generators by $v_1 \succ \cdots \succ v_n \succ a_1 \succ \cdots \succ a_s$, and apply the Independence Theorem. Later, Belov and Chekanu showed that we may take the $\{v_i\}$ to be the set of words from Shestakov's conjecture. Another proof of this fact was obtained by V. Drensky.

The original proofs of the Independence Theorem were rather complicated. Application of *hyperwords*, described below, allow a considerable simplification.

Subsequent papers of these authors contained various refinements and generalizations of these theorems. Here is another elegant result of Chekanu [Che96]:

Theorem 3.6. *Suppose a word w is extremal and non-periodic, of length n. If $w^n \neq 0$, then the algebra generated by the letters of w contains a nilpotent element of index exactly n.*

3.6. **Hyperwords in algebras.** Many of the combinatorial results in this survey are most easily proved using infinite words, or *hyperwords*, so we conclude with a discussion of basic auxiliary facts and constructions related to hyperwords in algebras.

Definition 3.9. *A hyperword is a word infinite in both directions; a word infinite only to the left (resp. right) is called a left (*resp. *right) hyperword.*

u^∞ denotes the hyperword having period u, and $u^{\infty/2}$ the left (resp. right) hyperword having period u and terminal (resp. initial) subword u.

The context will always make clear whether we consider a left or right hyperword, so we do not distinguish the notation between them. For example, the expression $u^{\infty/2} w v^{\infty/2}$ indicates that $u^{\infty/2}$ is a left hyperword and $v^{\infty/2}$ is a right hyperword.

Right hyperwords form a linearly ordered set with respect to the lexicographical order. For a right hyperword w, we let $(w)_k$ denote the initial subword of w having length k.

Lemma 3.10 ([BBL97])**.** *Let C be an arbitrary collection of words having unbounded length. Then there exists a hyperword w such that each of its subwords is a subword of a word from C.*

Although evaluating a hyperword in an algebra does not make sense, we can define whether or not it equals 0 (according to whether some subword equals 0), and this leads to the notion of linear independence of hyperwords in A:

Definition 3.11. *a) A hyperword w is called a zero hyperword if it includes a subword of finite length equal to 0, and a nonzero hyperword otherwise.*

b) A finite set of right hyperwords $\{w_i\}$ is called linearly dependent if there exist $\{\alpha_i\}$ such that some of them are not zero and for all sufficiently large k we have

$$\sum \alpha_i (w_i)_k = 0.$$

c) Suppose w is a right hyperword in an algebra A, M is a right A-module, and $m \in M$. We say that $mw \neq 0$ if $m(w)_k \neq 0$ for all k. Otherwise $Mw = 0$.

d) Suppose $\{w_1, \dots, w_n\}$ is a set of right hyperwords in an algebra A, and M is a right A-module. We say that $\sum m_i w_i = 0$ for $m_i \in M$ if $\sum m_i (w_i)_k = 0$ for all sufficiently large k.

Proposition 3.12. *a) A finitely generated non-nilpotent algebra A contains non-zero hyperwords.*

b) Suppose A is a finitely generated algebra, M is a finitely generated right A-module. If $MA^k \neq 0$ for all $k > 0$, then there exist $m \in M$ and a right hyperword w such that $mw \neq 0$.

The existence of a least upper bound and of a greatest lower bound for any set of right hyperwords implies the following

Proposition 3.13. *a) Let w be a hyperword. Then the set of right hyperwords whose subwords are all subwords of w contains maximal and minimal hyperwords.*

b) Suppose $\forall k \quad mA^k \neq 0$. Then the set of right hyperwords w such that $mw \neq 0$ contains a maximal and a minimal hyperword.

c) If A is non-nilpotent, then the set of nonzero right hyperwords in A contains a maximal and a minimal hyperword.

Let u be the maximal word in an algebra A among all nonzero words in A having length $\leq n$. Unfortunately u may have no extension to a word of greater length. Thus, to utilize hyperwords, we need the following construction:

Construction 1. Let A be an algebra having generators $a_s \succ \cdots \succ a_1$. Put $a_1 \succ x$ and consider the free product $A' = A * F\langle x \rangle$.

Each word u in A is an initial subword of some hyperword in A'. If u is the maximal word in A among all words having length at most $|u|$, then the maximal hyperword in A' beginning with u is a hyperword in A. If $\tilde{u}$ is a hyperword in A for which each initial subword has this property, then the maximal hyperword in A' is $\tilde{u}$.

The following construction is useful for treating modules.

Construction 2. Suppose A is an algebra having generators $a_s \succ \cdots \succ a_1$, and V is a finitely generated right A-module having generators $m_k \succ \cdots \succ m_1$. Put $m_1 \succ a_s$, $a_1 \succ x$, and $\widetilde{A}$ as in Remark 3.8. Define $A'' = \widetilde{A} * F\langle x \rangle / I$ where the ideal I is generated by elements of the form xm_i.

In the algebra A'', the maximal right hyperword begins with m_k, and each word in $\widetilde{A}$ may be extended to a hyperword in A''; if $MA^k \neq 0$ for all k, then the maximal hyperword in $\widetilde{A}$ begins with some m_i.

If u is the maximal word in A among all words having length at most $|u|$ that act nontrivially on the generators of the module, then after renumbering the m_i suitably, the maximal hyperword in A'' is a hyperword in $\widetilde{A}$. If u is a hyperword in $\widetilde{A}$ such that each its initial subword has the above property then the maximal hyperword in A'' is u.

Note that if an algebra has no nonzero nilpotent ideals, then any word may be extended to a hyperword. The following observation is useful.

Proposition 3.14. *If an algebra contains no nonzero periodic hyperword, then all of its words are nilpotent.*

The technique of hyperwords seems to lie rather close to the lines of structure theory, as illustrated in the following theorem and its proof, cf. [Bel07].

Theorem 3.7. *The set of irreducible words in a* PI*-algebra A has bounded height over the set of words whose degree does not exceed the PI-degree of A.*

Proof. Suppose m is the minimal degree of identities holding in an algebra A of PI-degree d. Since A has bounded height over the set of words having degree $\leq m$, it suffices to show that if $|u|$ is a nonperiodic word of length $> n$ then the word u^k for sufficiently large k is a linear combination of words of smaller lexicographic order.

Step 1. Consider the right A-module M defined by a generator v and by the relations $vw = 0$ whenever $w \prec u^{\infty/2}$. Our goal is to show that $Mu^k = 0$ for some k. Indeed, some power u^k is spanned by smaller lexicographic words. By virtue of Shirshov's Height Theorem, the set of irreducible words has bounded height over Y_m, the set of words of degree $\leq m$. But if each sufficiently large power of a nonperiodic word having length d may be linearly represented by smaller words, then the words having length $> d$ may be excluded from Y_m.

Step 2. The correspondence $\lambda : vs \to vus$ defines a well-defined endomorphism of the module M, hence M may be considered as an $A[\lambda]$-module. Our goal is to show that $M\lambda^k = 0$ for some k, or equivalently that $\overline{M} = M \otimes \mathbb{F}[\lambda, \lambda^{-1}] = 0$.

Step 3. If $M\lambda^k \in M \cdot J(\operatorname{Ann} M)$ where $J(\operatorname{Ann} M)$ is the Jacobson radical of the annihilator of M, then $M\lambda^{\ell k} \in M \cdot J(\operatorname{Ann} M)^\ell$, and by the nilpotence of the radical, $M\lambda^{\ell k} = 0$ for sufficiently large ℓ. Hence, we may assume that $J(\operatorname{Ann} M) = 0$.

Step 4. Using primary decomposition, we reduce to the case for which M is a faithful module over a primary ring B.

Step 5. Elements of the center $Z(B)$ have trivial annihilator, so we may localize relative to them; replacing $Z(B)$ by an algebraic extension, we reduce to the case for which B is the algebra of some dimension $k \leq n$ over a field, and $\overline{M}$ is a k-dimensional vector space.

Step 6. Since M is a vector space of dimension $< |u|$, the vectors $\vec{v}u_0, \vec{v}u_1, \ldots, \vec{v}u_{n-1}$ are linearly dependent (where u_i is the initial subword of length i in the word u, and $u_0 = 1$). Thus we have the equality

$$\sum_{i \in I} \lambda_i \vec{v}_i u_i = 0 \tag{2}$$

where $I \subseteq \{0, \ldots, n-1\}$, $\lambda_i \in \mathbb{F}\backslash 0$. To each u_i we attach a word $u^{(i)}$ so that $u_i u^{(i)} = u^{|u|}$. Let $u^{(j)}$ be the least of those $u^{(i)}$ which are involved in the formula (2). Write the equality 2 in the form

$$\vec{v}_j u_j = \sum_{i \in I \backslash \{j\}} \beta_i \vec{v}_i u_i \tag{3}$$

where $\beta_i = -\alpha_i/\alpha_j$. But then

$$\vec{v}u^{|u|} = \vec{v}_j u_j u^{(j)} = \sum_{i \in I \backslash \{j\}} \beta_i \vec{v}_i u_i u^{(j)}. \tag{4}$$

If $i \in I\backslash\{j\}$, then $u^{(j)} \prec u^{(i)}$ and $u_i u^{(j)} \prec u_i u^{(i)} = u^{|u|}$; hence $v u_i u^{(j)} = 0$. Thus all terms in the right side of (4) are zero. Hence $\vec{v}u^{|u|} = 0$, as desired. $\square$

Hyperwords facilitate proofs of the Independence Theorem, Shirshov's Height Theorem, nilpotence of the Lie algebra generated by sandwiches [Ufn90], proof of the *Bergman Gap Theorem*, (that any algebra of GK dimension greater than 1, has GK dimension at least 2, together with a description of the base having growth function $V_A(n) = \frac{n(n+3)}{2}$), and also describe various properties of monomial algebras [BBL97] as well as other combinatorial results for semigroups and rings.

References

[Am57] Amitsur, S.A., *A generalization of Hilbert's Nullstellensatz*, Proc. Amer. Math. Soc. **8** (1957), 649–656.

[An92] Anan'in, A.Z., *The representability of finitely generated algebras with chain condition*, Arch. Math. **59** (1992), 1–5.

[Ba87] Bakhturin, Yu. A., Identical relations in Lie algebras. Translated from the Russian by Bakhturin. VNU Science Press, b.v., Utrecht, (1987).

[Bel88a] Belov, A.Ya., *On Shirshov bases in relatively free algebras of complexity n*, Mat. Sb. **135** (1988), no. 3, 373–384.

[Bel88b] Belov, A.Ya., *The height theorem for Jordan and Lie PI-algebras*, in: Tez. Dokl. Sib. Shkoly po Mnogoobr. Algebraicheskih Sistem, Barnaul (1988), pp. 12–13.

[Bel89] Belov, A.Ya., *Estimations of the height and Gelfand-Kirillov dimension of associative PI-algebras*, In: Tez. Dokl. po Teorii Koletz, Algebr i Modulei. Mezhdunar. Konf. po Algebre Pamyati A.I.Mal'tzeva, Novosibirsk (1989), p. 21.

[Bel92] Belov, A.Ya., *Some estimations for nilpotency of nil-algebras over a field of an arbitrary characteristic and height theorem*, Commun. Algebra **20** (1992), no. 10, 2919–2922.

[Bel97] Belov, A.Ya., *Rationality of Hilbert series with respect to free algebras*, Russian Math. Surveys **52** (1997), no. 10, 394–395.

[Bel00] Belov, A.Ya., *Counterexamples to the Specht problem*, Sb. Math. **191** (3-4) (2000), 329-340.

[Bel02] Belov, A.Ya., *Algebras with polynomial idenitities: Representations and combinatorial methods*, Doctor of Science Dissertation, Moscow (2002).

[Bel07] Belov, A.Ya., *Burnside-type problems, and theorems on height and independence (Russian)*, Fundam. Prikl. Mat. 13 (2007), no. 5, 19–79.

[BBL97] Belov, A.Ya., Borisenko, V.V., and Latyshev, V.N., *Monomial algebras. Algebra 4*, J. Math. Sci. (New York) **87** (1997), no. 3, 3463-3575.

[BC00] Belov, A.Ya. and Chilikov, A.A., *Exponential Diophantine equations in rings of positive characteristic* (Russian) Fundam. Prikl. Mat. **6**(3), 649–668, (2000).

[BR05] Belov, A.Ya. and Rowen, L.H. *Computational aspects of polynomial identities*. Research Notes in Mathematics **9**. AK Peters, Ltd., Wellesley, MA, 2005.

[BRV06] Kanel-Belov, A.Ya., Rowen, L.H., and Vishne, U., *Normal bases of PI-algebras*, Adv. in Appl. Math. 37 (2006), no. 3, 378–389.

[Ber93] Berele, A., *Generic verbally prime PI-algebras and their GK-dimensions*, Comm. Algebra **21** (1993), no. 5, 1487–1504.

[Bog01] Bogdanov I., *Nagata-Higman's theorem for hemirings*, Fundam. Prikl. Mat. **7** (2001), no. 3, 651-658 (in Russian).

[BLH88] Bokut', L.A., L'vov, I.V., and Harchenko, V.K., *Noncommutative rings*, In: Sovrem. Probl. Mat. Fundam. Napravl. Vol. 18, Itogi Nauki i Tekhn., All-Union Institute for Scientificand Technical Information (VINITI), Akad. Nauk SSSR, Moscow (1988), 5–116.

[Br82] Braun, A., *The radical in a finitely generated PI-algebra*, Bull. Amer. Math. Soc. **7** (1982), no. 2, 385–386.

[Br84] Braun, A., *The nilpotence of the radical in a finitely generated PI-ring*, J. Algebra **89** (1984), 375–396.

[Che88] Chekanu, G.P., *Local finiteness of algebras. (Russian)* Mat. Issled. **105**, Moduli, Algebry, Topol. (1988), 153–171, 198.

[Che95] Chekanu, G.P., *Independence and quasiregularity in algebras*, Dokl. Akad. Nauk **337** (1994), no. 3, 316-319; translation: Russian Acad. Sci. Dokl. Math. **50** (1995), no. 1, 84–89.

[Che96] Chekanu, G.P., *Independence and quasiregularity in algebras. I. (Moldavian)* Izv. Akad. Nauk Respub. Moldova Mat. 1996, no. 3, 29–39, 120, 122.

[ChUf85] Chekanu, G.P., and Ufnarovski'i, V.A., *Nilpotent matrices*, Mat. Issled. no. 85, Algebry, Kotsa i Topologi (1985), 130–141, 155.

[DPR61] Davis M., Putnam, H., and Robsinson, J. *The decision problem for exponential differential equations*, Annals of Math. **74**, 425-436, (1961).

[Dne93] Dniester Notebook (Dnestrovskaya tetrad), Sobolev Institute of Mathematics, Novosibirsk, 1993.

[Do94] Donkin, S., *Polynomial invariants of representations of quivers*, Comment. Math. Helv. **69** (1994), no. 1, 137–141.

[Dr84a] Drensky, V., *On the Hilbert series of relatively free algebras*, Comm. in Algebra, **12** no. 19 (1984), 2335–2347.

[Dr84b] Drensky, V., *Codimensions of T-ideals and Hilbert series of relatively free algebras*, J. Algebra **91** no. 1 (1984), 1–17.

[Dr00] Drensky, V., *Free Algebras and PI-algebras: Graduate Course in Algebra*, Springer-Verlag, Singapore (2000).

[DrFor04] Drensky, V. and Formanek, E., *Polynomials Identity Rings*, CRM Advanced Courses in Mathematics, Birkhäuser, Basel (2004).

[For84] Formanek, E., *Invariants and the ring of generic matrices*, J. Algebra **89** (1984), no. 1, 178–223.

[Gol64] Golod, E.S., *On nil-algebras and residually finite p-groups*, Izv. Akad. Nauk SSSR **28** (1964), no. 2, 273–276.

[Gri99] Grishin, A.V., *Examples of T-spaces and T-ideals in Characteristic 2 without the Finite Basis Property*, Fundam. Prikl. Mat. **5** (1) (1999), no. 6, 101-118 (in Russian).

[GuKr02] Gupta, C.K., and Krasilnikov, A.N., *A simple example of a non-finitely based system of polynomial identities*, Camm. Algebra **36** (2002), 4851-4866.

[Hel74] Helling, H., *Eine Kennzeichnung von Charakteren auf Gruppen und Assoziativen Algebren*, Comm. in Alg. **1** (1974), 491–501.

[Hig56] Higman, G., *On a conjecture of Nagata*, Proc. Cam. Phil. Soc. **52** (1956), 1-4.

[Ilt91] Iltyakov, A.V., *Finiteness of basis identities of a finitely generated alternative PI-algebra*, Sibir. Mat. Zh. **31** (1991), no. 6, 87–99; English translation: Sib. Math. J. **31** (1991), 948-961.

[Ilt03] Iltyakov, A.V., *Polynomial identities of Finite Dimensional Lie Algebras*, monograph (2003).

[Kap49] Kaplansky, I., *Groups with representations of bounded degree*, Canadian J. Math. **1** (1949), 105–112.

[Kap50] Kaplansky, I., *Topological representation of algebras. II*, Trans. Amer. Math. Soc. **66** (1949), 464–491.

[Kem80] Kemer, A.R., *Capelli identities and the nilpotence of the radical of a finitely generated PI-algebra*, Soviet Math. Dokl. **22** (3) (1980), 750–753.

[Kem87] Kemer, A.R., *Finite basability of identities of associative algebras (Russian)*, Algebra i Logika **26** (1987), 597–641; English translation: Algebra and Logic **26** (1987), 362–397.

[Kem88] Kemer, A.R., *The representability of reduced-free algebras*, Algebra i Logika **27** (1988), no. 3, 274–294.

[Kem90a] Kemer, A.R., *Identities of Associative Algebras*, Transl. Math. Monogr., **87**, Amer. Math. Soc. (1991).

[Kem90b] Kemer, A. R. *Identities of finitely generated algebras over an infinite field* (Russian), Izv. Akad. Nauk SSSR Ser. Mat. **54** (1990), no. 4, 726–753; translation in Math. USSR-Izv. **37** (1991), no. 1, 69–96.

[Kem95] Kemer, A.R., *Multilinear identities of the algebras over a field of characteristic p*, Internat. J. Algebra Comput. **5** (1995), no. 2, 189–197.

[Kem09] Kemer, A.R., *Comments on the Shirshov's Height Theorem*, in this collection.

[Kol81] Kolotov, A.T., *Aperiodic sequences and growth functions in algebras*, Algebra i Logika **20** (1981), no. 2, 138–154.

[KrLe00] Krause, G.R., and Lenagan, T.H., *Growth of Algebras and Gelfand-Kirillov Dimension*, Amer. Math. Soc. Graduate Studies in Mathematics **22** (2000).

[Kuz75] Kuzmin, E.N., *About Nagata-Higman Theorem*, Proceedings dedicated to the 60th birthday of Academician Iliev, Sofia (1975), 101-107 (in Russian).

[Lat72] Latyshev, V.N., *On Regev's theorem on indentities in a tensor product of PI-algebras*, Uspehi Mat. Nauk. **27** (1972), 213–214.

[Lat88] Latyshev, V.N., *Combinatorial Ring Theory. Standard Bases*, Moscow University Press, Moscow (1988), (in Russian).

[Lev46] Levitzki, J., *On a problem of Kurosch*, Bull. Amer. Math. Soc. **52** (1946), 1033–1035.

[Lv83] Lvov, I.V., *Braun's theorem on the radical of PI-algebras*, Institute of Mathematics, Novosibirsk (1983), preprint.

[Mar88] Markov, V.T., *Gelfand-Kirillov dimension: nilpotence, representability, nonmatrix varieties*, In: Tez.Dokl. Sib. Shkola po Mnogoobr. Algebraicheskih Sistem, Barnaul (1988), 43–45.

[Mis90] Mishchenko, S. P. *A variant of a theorem on height for Lie algebras. (Russian)* Mat. Zametki **47** (1990), no. 4, 83–89; translation in Math. Notes **47** (1990), no. 3-4, 368–372.

[Pch84] Pchelintzev, S.V., *The height theorem for alternate algebras*, Mat. Sb. **124** (1984), no. 4, 557–567.

[Pro73] Procesi, C., *Rings with polynomial identities*, Research Notes in Mathematics **917**. Marcel Dekker, New York, 1973.

[Pro76] Procesi, C., *The invariant theory of $n \times n$ matrices*, Advances in Math. **19** (1976), 306-381.

[Raz74a] Razmyslov, Yu.P., Algebra and Logic **13** (1974), no. 3, 192–204.

[Raz74b] Razmyslov, Yu.P., *Trace identities of full matrix algebras over a field of characteristic zero*, Math. USSR Izv. **8** (1974), 724-760.

[Raz89] Razmyslov, Yu.P., *Identities of Algbras and their Representations*, Nauka, Moscow (1989).

[Reg72] Regev, A., *Existence of identities in $A \otimes B$*, Israel J. Math. 11 (1972), 131–152.

[Reg84] Regev, A., *Codimensions and trace codimensions of matrices are asymptotically equal*, Israel J. Math. **47** (1984), 246–250.

[Row88] Rowen, L.H., *Ring Theory II*, Pure and Applied Mathematics **128** Academic Press, New York, 1988.

[Sch76] Schelter, W., *Integral extensions of rings satisfying a polynomial identity*, J. Algebra **40** (1976), 245–257; errata op. cit. **44** (1977), 576.

[Sch78] Schelter, W., *Noncommutative affine PI-algebras are catenary*, J. Algebra **51** (1978), 12–18.

[Shch01] Shchigolev, V.V., *Finite basis property of T-spaces over fields of characteristic zero*, Izv. Ross. Akad. Nauk Ser. Mat. **65** (2001), no. 5, 191–224; translation: Izv. Math. **65** (2001), no. 5, 1041–1071.

[Shir57a] Shirshov, A.I., *On some nonassociative nil-rings and algebraic algebras*, Mat. Sb. **41** (1957), no. 3, 381–394.

[Shir57b] Shirshov, A.I., *On rings with identity relations*, Mat. Sb. **43**, (1957), no. 2, 277–283.

[Sp50] Specht, W., *Gesetze in Ringen I*, Math. Z. **52** (1950), 557–589.

[Ufn80] Ufnarovski'i, V.A., *On Poincaré series of graded algebras*, Mat. Zametki **27** (1980), no. 1, 21–32.

[Ufn85] Ufnarovski'i, V.A., *The independency theorem and its consequences*, Mat. Sb., **128** (1985), no. 1, 124–13.

[Ufn89] Ufnarovski'i, V.A., *On regular words in Shirshov sense*, In: Tez. Dokl. po Teorii Koletz, Algebr i Modulei. Mezhd. Konf. po Algebre Pamyati A.I. Mal'tzeva, Novosibirsk (1989), 140.

[Ufn90] Ufnarovski'i, V.A., *On using graphs for computing bases, growth functions and Hilbert series of associative algebras*, Mat. Sb. **180** (1990), no. 11, 1548–1550.

[VaZel89] Vais, A.Ja., and Zelmanov, E.I., *Kemer's theorem for finitely generated Jordan algebras*, Izv. Vyssh. Uchebn. Zved. Mat. (1989), no. 6, 42–51; translation: Soviet Math. (Iz. VUZ) **33** (1989), no. 6, 38–47.

[Zel91] Zelmanov, E.I., *The solution of the restricted Burnside problem for groups of prime power*, Mimeographed notes, Yale University (1991)

[ZelKos88] Zelmanov, E.I., *On nilpotence of nilalgebras*, Lect. Notes Math. **1352** (1988), 227–240.

[Zubk96] Zubkov, A. N., *On a generalization of the Razmyslov-Procesi theorem.* (Russian) Algebra i Logika **35** (1996), no. 4, 433–457, 498; translation in Algebra and Logic **35** (1996), no. 4, 241–254.

[Zubk00] Zubkov, A. N., *Modules with good filtration and invariant theory.* Algebra—representation theory (Constanta, 2000), 439–460, NATO Sci. Ser. II Math. Phys. Chem. **28** Kluwer Acad. Publ., Dordrecht, 2001.

[Zubr97] Zubrilin, K.A., *On the largest nilpotent ideal in algebras satisfying Capelli identities*, Sb. Math. **188** (1997), 1203–1211.

Deptartment of Mathematics, Bar Ilan University, Ramat Gan 52900, Israel
E-mail address: kanel@mccme.ru

Deptartment of Mathematics, Bar Ilan University, Ramat Gan 52900, Israel
E-mail address: rowen@math.biu.ac.il

Sbornik: *Mathematics* **203**:4 534–553 © 2012 RAS(DoM) and LMS
Matematicheskiĭ Sbornik **203**:4 81–102 DOI 10.1070/SM2012v203n04ABEH004233

Subexponential estimates in Shirshov's theorem on height

A. Ya. Belov and M. I. Kharitonov

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

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

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

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

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

Bibliography: 40 titles.

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

§ 1. Introduction

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

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

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

AMS 2010 *Mathematics Subject Classification.* Primary 16R10, 68R15.

Definition 1.2. A PI-algebra A is called an algebra *of bounded height* $h = \mathrm{Ht}_Y(A)$ *over a set of words* $Y = \{u_1, u_2, \dots\}$ if h is the minimal integer such that any word x from A can be represented in the form

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

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

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

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

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

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

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

where the α_σ belong to the ground field. In this case, if $W = vu_1u_2\cdots u_n$ is n-divisible then for any permutation σ the word $W_\sigma = vu_{\sigma(1)}u_{\sigma(2)}\cdots u_{\sigma(n)}$ is lexicographically smaller than W, and thus an n-divisible word can be represented as a linear combination of lexicographically smaller words. Hence a PI-algebra has a basis consisting of non-n-divisible words. By the Shirshov theorem on height, a PI-algebra has bounded height. In particular, if a PI-algebra satisfies $x^n = 0$ then it is nilpotent, that is, all of its words of length exceeding some N are identically zero. Surveys on the height theorem can be found in [4]–[8].

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

Corollary 1.1 (Berele). *Let A be a finitely generated* PI*-algebra. Then*

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

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

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

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

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

$$\mathrm{GK}(A) \leqslant \mathrm{Ht}(A).$$

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

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

Definition 1.3. An algebra A has *essential height* $h = H_{\mathrm{Ess}}(A)$ over a finite set Y, called an *s-basis* for A, if there exists a finite set $D \subset A$ such that A is linearly representable by elements of the form $t_1 \cdots t_l$, where $l \leqslant 2h+1$, and $\forall\, i\ (t_i \in D \vee t_i = y_i^{k_i}$; $y_i \in Y)$ and the set of i such that $t_i \notin D$ contains at most h elements. The essential height of a set of words is defined similarly.

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

The height theorem suggests the following questions.

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

Let us discuss the above questions.

1.2. Nonassociative generalizations. The height theorem was extended to some classes of near-associative rings. Pchelintsev [9] proved it for the alternative and the $(-1;1)$ cases, Mishchenko [10] obtained an analogue of the height theorem for Lie algebras with a sparse identity. Belov [11] proved the height theorem for some class of rings asymptotically close to associative rings. In particular, this class contains alternative and Jordan PI-algebras.

1.3. Shirshov bases. Suppose A is a PI-algebra and a subset $M \subseteq A$ is its s-basis. If all elements of M are algebraic over K, then A is finite-dimensional (the Kurosh problem). Boundedness of the essential height over Y implies 'an affirmative solution of the Kurosh problem over Y'. The converse is less trivial.

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

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

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

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

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

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

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

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

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

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

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

Structure of degree vectors. Although in the representable case the Gelfand-Kirillov dimension and the essential height behave well, even in this case the set of degree vectors may have a bad structure, namely, it can be the complement to the set of solutions of a system of exponential-polynomial Diophantine equations [4].

That is why there exists an instance of a representable algebra with a transcendent Hilbert series. However for a relatively free algebra, the Hilbert series is rational [19].

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

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

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

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

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

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

In this paper we transfer Latyshev's technique to the non-polylinear case, and this enables us to obtain a subexponential estimate in the Shirshov-height theorem. Chelnokov suggested the idea of this transfer in 1996.

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

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

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

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

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

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

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

1.8. The results obtained. The main result of the paper is as follows.

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

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

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

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

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

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

Corollary 1.4. *The height of an l-generated* PI*-algebra with an admissible polynomial identity of degree n over the set of words of length less than n does not exceed $\Phi(n,l)$.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

For a real number x put $\ulcorner x \urcorner := -[-x]$. Thus we replace noninteger numbers by the closest greater integers.

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

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

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

In [35] it is established that the nilpotency index of an l-generated nil-semiring of degree n equals the nilpotency index of an l-generated nilring of degree n, where addition is not necessarily commutative. (The paper also contains examples of non-nilpotent nil-nearrings of index 2.) Thus our results extend to the case of semirings as well.

1.9. On estimates from below. Let us compare the results obtained with the estimate for the height from below. The height of an algebra A is no less than its Gelfand-Kirillov dimension $\mathrm{GK}(A)$. For the algebra of l-generated general matrices of order n this dimension equals $(l-1)n^2+1$ (see [36] as well as [37]). At the same time, the minimal degree of an identity in this algebra is $2n$ by the Amitsur-Levitsky theorem. We have the following result.

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

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

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

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

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

§ 2. Estimates on the occurrence of degrees of subwords

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

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

To complete the proof, we construct a hierarchical structure based on Lemma 3.2, that is, we consecutively consider segments of n-tails, subsegments of these segments and so on. Furthermore we consider the greatest possible number of tails in the subset to which Dilworth's theorem is applied, and then we estimate from above the total number of tails and hence of the letters in the word W.

2.2. Periodicity and n-divisibility properties. Let $a_1, \dots, a_l$ be the alphabet used for constructing words. The ordering $a_1 \prec a_2 \prec \cdots \prec a_l$ induces a lexicographical ordering for words over the alphabet. For convenience, we introduce the following definitions.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Proof. Suppose the contrary. Consider the tails $u_1, u_2, \dots, u_n$ of the word u which begin from each of the first n letters of u. Renumber the tails to provide the

inequalities $u_1 \succ \ldots \succ u_n$. By Lemma 2.1 the tails are incomparable. Consider the subword u_1 in the left-most copy of u, the subword u_2 in the second copy from the left, $\ldots$, u_n in the nth copy from the left. We get an n-division of W. A contradiction.

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

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

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

the $(a_{j+1} - a_j)$-tail u_j in degree d is included into the $(a_2 d - a_j)$-tail u_j. A contradiction.

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

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

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

Let us introduce a word θ which is lexicographically less than any other word.

Remark 2.1. In the current proof of Theorem 7 all tails are assumed to belong to Ω.

§ 3. Estimates on the occurrence of periodic fragments

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

For each $j = 1, 2, \ldots, p_{n,d}$ put

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

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

Informally speaking, we observe the speed of 'evolution' of tails in their chains when the sequence of positions in W is considered as the time axis.

3.1. The sets $B^p(i)$, and the process at positions.

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

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

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

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

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

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

as required.

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

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

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

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

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

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

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

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

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

Now apply Lemma 3.1 as follows: in all representatives of type r except the rightmost one we consider a segment as a *unit segment* if it contains the least

position where this representative of type r differs from the preceding one. All other segments are considered as *zero segments*.

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

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

$$a_0 = 3^{\lceil \log_3 p_{n,d} \rceil}, \quad a_1 = 3^{\lceil \log_3 p_{n,d} \rceil - 1}, \quad \ldots, \quad a_{\lceil \log_3 p_{n,d} \rceil} = 1.$$

Then $|W| \leqslant d|\Omega| + d$ by Lemma 2.1.

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

$$\begin{aligned}\psi(1) &< (p_{n,d}^3 + p_{n,d})\psi(3) < (p_{n,d}^3 + p_{n,d})^2 \psi(9) \\ &< \cdots < (p_{n,d}^3 + p_{n,d})^{\lceil \log_3 p_{n,d} \rceil} \psi(p_{n,d}) \leqslant (p_{n,d}^3 + p_{n,d})^{\lceil \log_3 p_{n,d} \rceil} p_{n,d} d.\end{aligned}$$

Take $p_{n,d} = 4nd - 1$ to get

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

This implies the assertion of Theorem 7.

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

$$a_0 = 3^{\lceil \log_3 p_{n,d} \rceil}, \quad a_1 = 3^{\lceil \log_3 p_{n,d} \rceil - 1}, \quad \ldots, \quad a_{\lceil \log_3 p_{n,d} \rceil} = 1$$

we have to consider the sequence

$$a_0 = 2^{\lceil \log_2 p_{n,d} \rceil}, \quad a_1 = 2^{\lceil \log_2 p_{n,d} \rceil - 1}, \quad \ldots, \quad a_{\lceil \log_2 p_{n,d} \rceil} = 1.$$

§ 4. An estimate for the essential height

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

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

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

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

Definition 4.1. A word u will be called *noncyclic* if u is not representable in the form v^k where $k > 1$.

Definition 4.2. A *word cycle* u is the set consisting of the word u and all its cyclic shifts.

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

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

Proof. Put $x := x_{i_1}$. Then W includes disjoint subwords $x^{p_1} v_1', \ldots, x^{p_{2n-1}} v_{2n-1}'$, where $p_1, \ldots, p_{2n-1}$ are positive integers greater than n, and $v_1', \ldots, v_{2n-1}'$ are words of length m comparable with x, $v_1' = v_{i_1}$. Hence among the words $v_1', \ldots, v_{2n-1}'$ either there are n words lexicographically greater than x or there are n words lexicographically smaller than x. We may assume that $v_1', \ldots, v_n'$ are lexicographically greater than x. Then W includes subwords $v_1', xv_2', \ldots, x^{n-1} v_n'$, which lexicographically decrease from left to right.

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

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

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

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

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

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

Proof. Suppose there exists an antichain consisting of n words

$$v(i_1, j_1), v(i_2, j_2), \ldots, v(i_n, j_n), \qquad i_1 \leqslant i_2 \leqslant \cdots \leqslant i_n.$$

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

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

The word $s_{i_{r+1}}$ is periodic, hence it is representable as a product of n copies of $v_{i_{r+1}}^2$. The word $v_{i_{r+1}}^2$ includes a word cycle $v_{i_{r+1}}$. Hence in $s_{i_{r+1}}$ there exist disjoint subwords placed in lexicographically decreasing order and equal to $v(i_{r+1}, j_{r+1})$, $\ldots, v(i_{r+k}, j_{r+k})$ respectively. Similarly we deal with all sets of equal indices in the sequence $\{i_r\}_{r=1}^n$. The result is n-divisibility of W. A contradiction.

Thus Ω' can be divided into $(n-1)$ chains.
Put $q_n = n-1$.

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

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

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

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

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

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

Lemma 4.3. $\varphi(m) \leqslant q_n/m$.

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

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

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

$$\varphi(a) \leqslant q_n^k \varphi(k \cdot a).$$

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

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

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

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

4.4. Completion of the proof for Theorem 9. Suppose

$$a_0 = 3^{\lceil \log_3 p_{n,d} \rceil}, \quad a_1 = 3^{\lceil \log_3 p_{n,d} \rceil - 1}, \quad \ldots, \quad a_{\lceil \log_3 p_{n,d} \rceil} = 1.$$

Substitute these a_i into Lemmas 4.4 and 4.3 to obtain

$$\varphi(1) \leqslant q_n^3 \varphi(3) \leqslant q_n^9 \varphi(9) \leqslant \cdots \leqslant q_n^{3^{\lceil \log_3 m \rceil}} \varphi(m) \leqslant q_n^{3^{\lceil \log_3 m \rceil + 1}}.$$

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

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

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

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

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

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

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

$$W_0 = u_1 x_{1'}^{4n + r_1 + r_2} y_1 = u_1 f_1 y_1.$$

In the sequel, the positions contained in f_1 are called *tedious*, the last position of u_1 is called *tedious of type* 1 the second position from the end in u_1 is called *tedious of type* 2, ..., the nth position from the end in u_1 is called *tedious of type* n. Put $W_1 = u_1 y_1$.

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

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

$$W_{k-1} = u'_k x_{k'}^{4n} y'_k.$$

Then put

$$r_1 := \sup\{r \colon u'_k = u_k x_{k'}^r\}, \qquad r_2 := \sup\{r : y'_k = x_{k'}^r y_k\}.$$

(Note that the words involved may be empty.)

Define f_k by the equation

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

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

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

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

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

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

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

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

1) U is a product of words of the form $f_i^{(j)}$;
2) U is not a proper subword of a word which satisfies the above condition 1).

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

If $n_i \geqslant 3$, then $k_i \leqslant k_{i-1} - 1$. If $n_i \leqslant 2$ then $k_i \leqslant k_{i-1} + 1$. Furthermore, $k_1 = 1$, $k_t \geqslant 1 = k_1$. The lemma is proved.

Corollary 5.1.

$$\sum_{k=1}^{\infty} k \cdot s(k) \leqslant 10t \leqslant 5s.$$

Proof. From the proof of Lemma 5.1 we obtain

$$\sum_{n_i \geqslant 3} (n_i - 2) \leqslant 2t.$$

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

Proposition 5.1. *The height of W does not exceed*

$$\Psi(n, 4n, l) + \sum_{k=1}^{\infty} k \cdot s(k) \leqslant \Psi(n, 4n, l) + 5s.$$

In the sequel we consider only f_i with $n_i \leqslant 2$.

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

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

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

Now consider z_j', the subwords in W of the following form:

$$z_j' = x_{(2j-1)''}^{p_{(2j-1)''}+\beth} v_j, \qquad \beth \geqslant 0, \quad |v_j| = |x_{(2j-1)''}|;$$

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

Indeed, if $f''_{2j-1} = f_{m_{2j-1}}$, then put $z_j' = f_{m_{2j-1}} v_j$.

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

Thus we have proved the following lemma.

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

5.3. Completion of the proof for the main Theorem 5 and for Theorem 6. Replace s' in Lemma 5.2 by s from the proof of Theorem 9 to obtain that the height of W does not exceed

$$\Psi(n, 4n, l) + 5s < E_1 l \cdot n^{E_2 + 12 \log_3 n},$$

where $E_1 = 4^{21 \log_3 4 + 17}$, $E_2 = 30 \log_3 4 + 10$.

Thus we have obtained the assertion of the main Theorem 5.

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

$$a_0 = 3^{\lceil \log_3 p_{n,d} \rceil}, \quad a_1 = 3^{\lceil \log_3 p_{n,d} \rceil - 1}, \ldots, \quad a_{\lceil \log_3 p_{n,d} \rceil} = 1$$

by the sequence

$$a_0 = 2^{\lceil \log_2 p_{n,d} \rceil}, \quad a_1 = 2^{\lceil \log_2 p_{n,d} \rceil - 1}, \quad \ldots, \quad a_{\lceil \log_2 p_{n,d} \rceil} = 1,$$

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

§ 6. Comments

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

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

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

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

Bibliography

[1] A. I. Shirshov, "On some nonassociative nilring and algebraic algebras", *Mat. Sb.* **41(83)**:3 (1957), 381–394. (Russian)

[2] A. I. Shirshov, "On rings with identical relations", *Mat. Sb.* **43(85)**:2 (1957), 277–283. (Russian)

[3] E. I. Zel'manov, "On the nilpotency of nil algebras", *Algebra – some current trends* (Varna, 1986), Lecture Notes in Math., vol. 1352, Springer-Verlag, Berlin 1988, pp. 227–240.

[4] A. Ya. Belov, V. V. Borisenko and V. N. Latyshev, "Monomial algebras", *J. Math. Sci. (N. Y.)* **87**:3 (1997), 3463–3575.

[5] A. R. Kemer, "Comments on Shirshov's Height Theorem", *Selected Works of A. I. Shirshov*, Birkhäuser, Basel 2009, pp. 223–230.

[6] A. Belov-Kanel (Kanel-Belov) and L. H. Rowen, "Perspectives on Shirshov's Height Theorem", *Selected Works of A. I. Shirshov*, Birkhäuser, Basel 2009, pp. 185–203.

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

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

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

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

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

[12] A. Kanel-Belov and L. H. Rowen, *Computational aspects of polynomial identities*, Res. Notes Math., vol. 9, Peters, Wellesley, MA 2005.

[13] Gh. Ciocanu, "Independence and quasiregularity in algebras. II", *Izv. Akad. Nauk Respub. Moldova Mat.* **1** (1997), 70–77.

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

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

[16] G. P. Chekanu (Ciocanu), "Independence and quasiregularity in algebras", *Dokl. Ross. Akad. Nauk* **337**:3 (1994), 316–319; English transl. in *Russian Acad. Sci. Dokl. Math.* **50**:1 (1995), 84–89.

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

[18] V. A. Ufnarovskiĭ and Gh. P. Ciocanu, "Nilpotent matrices", *Mat. Issled.* **85** (1985), 130–141. (Russian)

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

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

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

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

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

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

[25] A. Ya. Belov, "Some estimations for nilpotence of nil-algebras over a field of an arbitrary characteristic and height theorem", *Comm. Algebra* **20**:10 (1992), 2919–2922.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

УДК 512.552+512.64+519.1

А. Я. Белов, М. И. Харитонов

Субэкспоненциальные оценки в теореме Ширшова о высоте

Пусть $F_{2,m}$ – свободное 2-порожденное ассоциативное кольцо с тождеством $x^m = 0$. В 1993 г. Е. И. Зельманов поставил вопрос об экспоненциальности роста класса нильпотентности кольца $F_{2,m}$ по m.

Мы отвечаем на вопрос Е. И. Зельманова, установив, что в l-порожденной ассоциативной алгебре с тождеством $x^d = 0$ класс нильпотентности меньше, чем $\Psi(d, d, l)$, где

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

Данный результат является следствием следующего факта, относящегося к комбинаторике слов. Пусть l, n и $d \geqslant n$ – некоторые натуральные числа. Тогда все слова над l-буквенным алфавитом длины не меньше, чем $\Psi(n, d, l)$, либо содержат x^d, либо являются n-разбиваемыми, где слово W называется *n-разбиваемым*, если его можно представить в виде $W = W_0 W_1 \cdots W_n$ так, что подслова $W_1, \ldots, W_n$ идут в порядке лексикографического убывания. В доказательстве используется теорема Дилуорса (идея В. Н. Латышева). Мы показываем, что множество всех не n-разбиваемых слов над l-буквенным алфавитом имеет высоту $h < \Phi(n, l)$ над множеством слов степени не выше $n - 1$, где

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

Библиография: 40 названий.

Ключевые слова: теорема Ширшова о высоте, комбинаторика слов, n-разбиваемость, теоремы Дилуорса, проблемы бернсайдовского типа.

§ 1. Введение

1.1. Теорема Ширшова о высоте. В 1958 г. А. И. Ширшов доказал свою знаменитую теорему о высоте (см. [1], [2]).

Определение 1.1. Назовем слово W *n-разбиваемым*, если W можно представить в виде $W = v u_1 u_2 \cdots u_n$ так, чтобы $u_1 \succ u_2 \succ \cdots \succ u_n$.

В этом случае при любой нетождественной перестановке σ подслов u_i получается слово $W_\sigma = v u_{\sigma(1)} u_{\sigma(2)} \cdots u_{\sigma(n)}$, лексикографически меньшее W. Это свойство некоторые авторы берут за основу определения понятия n-разбиваемости.

Определение 1.2. Назовем PI-алгебру A алгеброй *ограниченной высоты* $h = \mathrm{Ht}_Y(A)$ *над множеством слов* $Y = \{u_1, u_2, \dots\}$, если h – минимальное число такое, что любое слово x из A можно представить в виде

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

причем $\{r_i\}$ не превосходят h. Множество Y называется *базисом Ширшова* для A.

Там, где это не вызывает недоразумений, для обозначения высоты будем использовать h вместо $\mathrm{Ht}_Y(A)$.

Теорема Ширшова о высоте (см. [1], [2]). *Множество всех не n-разбиваемых слов в конечно порожденной алгебре с допустимым полиномиальным тождеством имеет ограниченную высоту H над множеством слов степени не выше $n-1$.*

Проблемы бернсайдовского типа, связанные с теоремой о высоте, рассмотрены в обзоре [3]. Авторы убеждены, что теорема Ширшова о высоте является фундаментальным фактом комбинаторики слов безотносительно приложений к PI-теории. (Все наши доказательства не выходят за рамки работы со словами.) К сожалению, специалисты по комбинаторике данный факт должным образом пока не оценили. Что касается самого понятия n-разбиваемости, то оно также представляется фундаментальным. Оценки, полученные В. Н. Латышевым на $\xi_n(k)$ – количество не n-разбиваемых полилинейных слов от k символов, привели к фундаментальным результатам в PI-теории. Вместе с тем, это количество есть не что иное, как количество расстановок чисел от 1 до k таких, что никакие n из них (не обязательно стоящие подряд) не идут в порядке убывания. Этому же числу равно количество k-элементных перестановочно упорядоченных множеств диаметра n (множество называется *перестановочно упорядоченным*, если его порядок есть пересечение двух линейных порядков, *диаметр* упорядоченного множества – длина его максимальной антицепи).

Из теоремы о высоте вытекает решение ряда проблем теории колец. В самом деле, пусть в ассоциативной алгебре над полем выполняется полиномиальное тождество $f(x_1, \dots, x_n) = 0$. Можно доказать, что тогда в ней выполняется и допустимое полилинейное тождество (т.е. полиномиальное тождество, у которого хотя бы один коэффициент при членах высшей степени равен единице)

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

где α_σ принадлежат основному полю. В этом случае, если $W = vu_1u_2 \cdots u_n$ является n-разбиваемым, то для любой перестановки σ слово $W_\sigma = vu_{\sigma(1)}u_{\sigma(2)} \cdots u_{\sigma(n)}$ лексикографически меньше слова W, т.е. n-разбиваемое слово можно представить в виде линейной комбинации лексикографически меньших слов. Значит, PI-алгебра имеет базис из не n-разбиваемых слов. В силу теоремы Ширшова о высоте PI-алгебра имеет ограниченную высоту. Как следствие имеем, что если в PI-алгебре выполняется тождество $x^n = 0$, то эта алгебра –

нильпотентна, т.е. все ее слова длины больше, чем некоторое N, тождественно равны 0. Обзоры, посвященные теореме о высоте, содержатся в работах [4]–[8].

Из теоремы о высоте вытекает положительное решение проблемы Куроша и других проблем бернсайдовского типа для PI-колец. Ведь если Y – базис Ширшова и все элементы из Y алгебраичны, то алгебра A конечномерна. Тем самым теорема Ширшова дает явное указание множества элементов, алгебраичность которых ведет к конечномерности всей алгебры. Из этой теоремы вытекает

Следствие 1.1 (Берель). *Пусть A – конечно порожденная* PI*-алгебра. Тогда*

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

Здесь $\mathrm{GK}(A)$ – это *размерность Гельфанда–Кириллова* алгебры A:

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

где $V_A(n)$ есть функция роста алгебры A, т.е. размерность векторного пространства, порожденного словами степени не выше n от образующих A.

В самом деле, достаточно заметить, что число решений неравенства $k_1|v_1| + \cdots + k_h|v_h| \leqslant n$, где $h \leqslant H$, превосходит N^H, и потому

$$\mathrm{GK}(A) \leqslant \mathrm{Ht}(A).$$

Число $m = \deg(A)$ будет обозначать *степень алгебры*, или минимальную степень тождества, которое в ней выполняется, число $n = \mathrm{Pid}(A)$ – *сложность* алгебры A, или максимальное k такое, что $\mathbb{M}_k$ – алгебра матриц размера k, принадлежит многообразию $\mathrm{Var}(A)$, порожденному алгеброй A.

Вместо понятия высоты удобнее пользоваться близким понятием существенной высоты.

Определение 1.3. Алгебра A имеет *существенную высоту* $h = H_{\mathrm{Ess}}(A)$ над конечным множеством Y, называемым *s-базисом алгебры* A, если можно выбрать такое конечное множество $D \subset A$, что A линейно представима элементами вида $t_1 \cdots t_l$, где $l \leqslant 2h+1$, и $\forall\, i\ (t_i \in D \vee t_i = y_i^{k_i};\ y_i \in Y)$, причем множество таких i, что $t_i \notin D$, содержит не более h элементов. Аналогично определяется существенная высота множества слов.

Говоря неформально, любое длинное слово есть произведение периодических частей и “прокладок” ограниченной длины. Существенная высота есть число таких периодических кусков, а обычная еще учитывает “прокладки”.

В связи с теоремой о высоте возникли следующие вопросы.

1. На какие классы колец можно распространить теорему о высоте?
2. Над какими Y алгебра A имеет ограниченную высоту? В частности, какие наборы слов можно взять в качестве $\{v_i\}$?
3. Как устроен вектор степеней $(k_1, \ldots, k_h)$? Прежде всего: какие множества компонент этого вектора являются существенными, т.е. какие

наборы k_i могут быть одновременно неограниченными? Какова существенная высота? Верно ли, что множество векторов степеней обладает теми или иными свойствами регулярности?

4. Как оценить высоту?

Перейдем к обсуждению поставленных вопросов.

1.2. Неассоциативные обобщения. Теорема о высоте была распространена на некоторые классы колец, близких к ассоциативным. С. В. Пчелинцев в [9] доказал ее для альтернативного и $(-1,1)$ случаев, С. П. Мищенко в [10] получил аналог теоремы о высоте для алгебр Ли с разреженным тождеством. В работе А. Я. Белова [11] теорема о высоте была доказана для некоторого класса колец, асимптотически близких к ассоциативным, куда входят, в частности, альтернативные и йордановы PI-алгебры.

1.3. Базисы Ширшова. Пусть A – PI-алгебра и подмножество $M \subseteq A$ является ее s-базисом. Тогда если все элементы множества M алгебраичны над K, то алгебра A конечномерна (проблема Куроша). Ограниченность существенной высоты над Y влечет "положительное решение проблемы Куроша над Y". Обратное утверждение менее тривиально.

Теорема 1 (А. Я. Белов). а) *Пусть A – градуированная* PI*-алгебра, Y – конечное множество однородных элементов. Тогда если при всех n алгебра $A/Y^{(n)}$ нильпотентна, то Y есть s-базис A. Если при этом Y порождает A как алгебру, то Y – базис Ширшова алгебры A.*

б) *Пусть A –* PI*-алгебра, $M \subseteq A$ – некоторое курошево подмножество в A. Тогда M – s-базис алгебры A.*

$Y^{(n)}$ обозначает идеал, порожденный n-ми степенями элементов из Y. Множество $M \subset A$ называется *курошевым*, если любая проекция $\pi\colon A \otimes K[X] \to A'$, в которой образ $\pi(M)$ цел над $\pi(K[X])$, конечномерна над $\pi(K[X])$. Мотивировкой этого понятия служит следующий пример. Пусть $A = \mathbb{Q}[x, 1/x]$. Любая проекция π такая, что $\pi(x)$ алгебраичен, имеет конечномерный образ. Однако множество $\{x\}$ не является s-базисом алгебры $\mathbb{Q}[x, 1/x]$. Таким образом, ограниченность существенной высоты есть некоммутативное обобщение свойства целости.

1.4. Базисы Ширшова, состоящие из слов. Описание базисов Ширшова, состоящих из слов, дает следующая теорема.

Теорема 2 (см. [4], [12]). *Множество слов Y является базисом Ширшова алгебры A тогда и только тогда, когда для любого слова и длины не выше $m = \mathrm{Pid}(A)$ – сложности алгебры A – множество Y содержит слово, циклически сопряженное к некоторой степени слова u.*

Аналогичный результат был независимо получен Г. П. Чекану и В. Дренски. Вопросы, связанные с локальной конечностью алгебр, с алгебраическими множествами слов степени не выше сложности алгебры, исследовались в работах [7], [13]–[18]. В этих же работах обсуждались вопросы, связанные с обобщением теоремы о независимости.

1.5. Существенная высота. Ясно, что размерность Гельфанда–Кириллова оценивается существенной высотой и что s-базис является базисом Ширшова тогда и только тогда, когда он порождает A как алгебру. В представимом случае имеет место и обратное утверждение.

Теорема 3 (А. Я. Белов; см. [4]). *Пусть A – конечно порожденная представимая алгебра и пусть $H_{\mathrm{Ess}Y}(A) < \infty$. Тогда $H_{\mathrm{Ess}Y}(A) = \mathrm{GK}(A)$.*

Следствие 1.2 (В. Т. Марков). *Размерность Гельфанда–Кириллова конечно порожденной представимой алгебры есть целое число.*

Следствие 1.3. *Если $H_{\mathrm{Ess}Y}(A) < \infty$ и алгебра A представима, то существенная высота $H_{\mathrm{Ess}Y}(A)$ не зависит от выбора s-базиса Y.*

В этом случае размерность Гельфанда–Кириллова также равна существенной высоте в силу локальной представимости относительно свободных алгебр.

Строение векторов степеней. Хотя в представимом случае размерность Гельфанда–Кириллова и существенная высота ведут себя хорошо, даже тогда множество векторов степеней может быть устроено плохо, а именно может быть дополнением к множеству решений системы экспоненциально-полиномиальных диофантовых уравнений (см. [4]). Вот почему существует пример представимой алгебры с трансцендентным рядом Гильберта. Однако для относительно свободной алгебры ряд Гильберта рационален (см. [19]).

1.6. n-разбиваемость и теорема Дилуорса. Значение понятия n-разбиваемости выходит за рамки проблематики, относящейся к проблемам бернсайдовского типа. Оно играет роль и при изучении полилинейных слов, в оценке их количества, где *полилинейным* называется слово, в которое каждая буква входит не более одного раза. В. Н. Латышев применил теорему Дилуорса для получения оценки числа не m-разбиваемых полилинейных слов степени n над алфавитом $\{a_1, \ldots, a_n\}$. Эта оценка – $(m-1)^{2n}$, и она близка к реальности. Напомним теорему Дилуорса.

Теорема Дилуорса. *Пусть n – наибольшее количество элементов антицепи данного конечного частично упорядоченного множества M. Тогда M можно разбить на n попарно непересекающихся цепей.*

Рассмотрим полилинейное слово W из n букв. Положим $a_i \succ a_j$, если $i > j$ и буква a_i стоит в слове W правее a_j. Условие не k-разбиваемости означает отсутствие антицепи из n элементов. Тогда по теореме Дилуорса все позиции (и, соответственно, буквы a_i) разбиваются на $n-1$ цепь. Сопоставим каждой цепи свой цвет. Тогда раскраска позиций и раскраска букв однозначно определяет слово W, а число таких раскрасок не превосходит

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

Из данной оценки следует, что выполняются полилинейные тождества, отвечающие неприводимому модулю, диаграмма Юнга которого содержит квадрат n^4. Это, в свою очередь, во-первых, позволило получить прозрачное доказательство теоремы Регева о том, что тензорное произведение PI-алгебр снова является PI-алгеброй, во-вторых, установить существование разреженного

тождества в общем случае, а также тождества Капелли в конечно порожденном случае (тем самым, в частности, доказать теорему о нильпотентности радикала), и в-третьих, осуществить "супертрюк" А. Р. Кемера, сводящий изучение тождеств общих алгебр к изучению супер-тождеств конечно порожденных супералгебр в нулевой характеристике. Смежные вопросы рассмотрены в работах [20]–[22].

Вопросы, связанные с перечислением полилинейных слов, не являющихся n-разбиваемыми, имеют самостоятельный интерес. (Например, существует биекция между не 3-разбиваемыми словами и числами Каталана.) С одной стороны, это чисто комбинаторная задача, с другой стороны, она связана с рядом коразмерностей для алгебры общих матриц. Исследование полилинейных слов представляется чрезвычайно важным. В. Н. Латышев (см., например, [23]) поставил проблему конечной базируемости множества старших полилинейных слов для T-идеала относительно взятия надслов и изотонных подстановок. Из этой проблемы вытекает проблема Шпехта для полилинейных многочленов, имеется тесная связь с проблемой слабой нётеровости групповой алгебры бесконечной финитарной симметрической группы над полем положительной характеристики (для нулевой характеристики это было установлено А. Залесским). Для решения проблемы Латышева надо уметь переводить свойства T-идеалов на язык полилинейных слов. В работах [4], [11] была произведена попытка осуществить программу перевода структурных свойств алгебр на язык комбинаторики слов. На язык полилинейных слов такой перевод осуществить проще, в дальнейшем можно получить информацию и о словах общего вида.

В настоящей работе мы переносим технику В. Н. Латышева на не полилинейный случай, что позволяет получить субэкспоненциальную оценку в теореме Ширшова о высоте. Г. Р. Челноков предложил идею этого переноса в 1996 г.

1.7. Оценки высоты. Первоначальное доказательство А. И. Ширшова хотя и было чисто комбинаторным (оно основывалось на технике элиминации, развитой им в алгебрах Ли, в частности, в доказательстве теоремы о свободе), однако оно давало только примитивно рекурсивные оценки. Позднее А. Т. Колотов в [24] получил оценку на $\mathrm{Ht}(A) \leqslant l^{l^n}$ ($n = \deg(A)$, l – число образующих). А. Я. Белов в работе [25] показал, что $\mathrm{Ht}(n, l) < 2nl^{n+1}$. Экспоненциальная оценка теоремы Ширшова о высоте изложена также в работах [12], [26], [27]. Данные оценки улучшались в работах А. Клейна [28], [29]. В 2001 г. Е. С. Чибриков в работе [30] доказал, что $\mathrm{Ht}(4, l) \geqslant (7k^2 - 2k)$. М. И. Харитонов получил в работах [27], [31], [32] оценки на структуру кусочной периодичности. В 2011 г. А. А. Лопатин [33] получил следующий результат.

Теорема 4. *Пусть $C_{n,l}$ – степень нильпотентности свободной l-порожденной алгебры, удовлетворяющей тождеству $x^n = 0$. Пусть p – характеристика базового поля алгебры, $p > \frac{n}{2}$. Тогда*

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

По определению $C_{n,l} \leqslant \Psi(n,n,l)$. Заметим, что для малых n оценка (1) меньше, чем полученная в данной работе оценка $\Psi(n,n,l)$, но при росте n оценка $\Psi(n,n,l)$ асимптотически лучше оценки (1).

Е. И. Зельманов поставил следующий вопрос в Днестровской тетради [34] в 1993 г.

Вопрос 1.1. Пусть $F_{2,m}$ – свободное 2-порожденное ассоциативное кольцо с тождеством $x^m = 0$. Верно ли, что класс нильпотентности кольца $F_{2,m}$ растет экспоненциально по m?

Наша работа отвечает на вопрос Е. И. Зельманова следующим образом: в действительности искомый класс нильпотентности растет субэкспоненциально.

1.8. Полученные результаты. Основной результат работы состоит в следующем.

Теорема 5. *Высота множества не n-разбиваемых слов над l-буквенным алфавитом относительно множества слов длины меньше n не превышает $\Phi(n,l)$, где*

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

Из теоремы 5 путем некоторого огрубления и упрощения оценки получается, что при фиксированном l и $n \to \infty$

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

а при фиксированном n и $l \to \infty$

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

Следствие 1.4. *Высота l-порожденной* PI*-алгебры с допустимым полиномиальным тождеством степени n над множеством слов длины меньше n не превышает $\Phi(n,l)$.*

Кроме того, доказывается субэкспоненциальная оценка, которая лучше при малых n.

Теорема 6. *Высота множества не n-разбиваемых слов над l-буквенным алфавитом относительно множества слов длины меньше n не превышает $\Phi(n,l)$, где*

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

Как следствие получаются субэкспоненциальные оценки на индекс нильпотентности l-порожденных ниль-алгебр степени n для произвольной характеристики.

Другим основным результатом нашей работы является следующая

Теорема 7. *Пусть l, n и $d \geqslant n$ – некоторые натуральные числа. Тогда все l-порожденные слова длины не меньше, чем $\Psi(n,d,l)$, либо содержат x^d, либо являются n-разбиваемыми, где*

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

Из теоремы 7 путем некоторого огрубления и упрощения оценки получается, что при фиксированном l и $nd \to \infty$

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

а при фиксированном n и $l \to \infty$

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

Следствие 1.5. *Пусть l, d – некоторые натуральные числа. Пусть в ассоциативной l-порожденной алгебре A выполнено тождество $x^d = 0$. Тогда ее индекс нильпотентности меньше, чем $\Psi(d,d,l)$.*

Кроме того, доказывается субэкспоненциальная оценка, которая лучше при малых n и d.

Теорема 8. *Пусть l, n и $d \geqslant n$ – некоторые натуральные числа. Тогда все l-порожденные слова длины не меньше, чем $\Psi(n,d,l)$, либо содержат x^d, либо являются n-разбиваемыми, где*

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

Для вещественного числа x положим $\ulcorner x \urcorner := -[-x]$. Таким образом мы округляем нецелые числа в большую сторону.

В процессе доказательства теоремы 5 доказывается следующая теорема, оценивающая существенную высоту.

Теорема 9. *Существенная высота l-порожденной PI-алгебры с допустимым полиномиальным тождеством степени n над множеством слов длины меньше n меньше, чем $\Upsilon(n,l)$, где*

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

В работе [35] установлено, что индекс нильпотентности l-порожденного нильполукольца степени n совпадает с индексом нильпотентности l-порожденного ниль-кольца степени n, причем сложение не обязательно коммутативно. Там же приведены примеры не нильпотентных ниль-почтиколец индекса 2. Таким образом, наши результаты распространяются и на случай полуколец.

1.9. О нижних оценках. Сравним полученные результаты с нижней оценкой для высоты. Высота алгебры A не меньше ее размерности Гельфанда–Кириллова $\mathrm{GK}(A)$. Для алгебры l-порожденных общих матриц порядка n данная размерность равна $(l-1)n^2+1$ (см. [36], а также [37]). В то же время, минимальная степень тождества этой алгебры равна $2n$ в силу теоремы Амицура–Левицкого. Имеет место следующее

Предложение 1.1. *Высота l-порожденной* PI*-алгебры степени n, а также множества не n-разбиваемых слов над l-буквенным алфавитом не меньше, чем $(l-1)n^2/4+1$.*

Нижние оценки на индекс нильпотентности были установлены Е. Н. Кузьминым в работе [38]. Он привел пример 2-порожденной алгебры с тождеством $x^n = 0$, индекс нильпотентности которой строго больше $(n^2 + n - 2)/2$. Вопрос нахождения нижних оценок рассматривается в работе [31].

В то же время для случая нулевой характеристики и счетного числа образующих Ю. П. Размыслов (см., например, [39]) получил верхнюю оценку на индекс нильпотентности равную n^2.

Сначала мы докажем теорему 7, а в § 4 займемся оценками существенной высоты, т.е. количества различных периодических фрагментов в не n-разбиваемом слове.

Авторы признательны В. Н. Латышеву, А. В. Михалеву и всем участникам семинара "Теория колец" за внимание к работе, а также участникам семинара МФТИ под руководством А. М. Райгородского.

§ 2. Оценки на появление степеней подслов

2.1. План доказательства теоремы 7. В леммах 2.1, 2.2 и 2.3 описываются достаточные условия для присутствия периода длины d в не n-разбиваемом слове W. В лемме 2.4 связываются понятия n-разбиваемости слова W и множества его хвостов. После этого определенным образом выбирается подмножество множества хвостов слова W, для которого можно применить теорему Дилуорса. Затем мы раскрашиваем хвосты и их первые буквы в соответствии с принадлежностью к цепям, полученным при применении теоремы Дилуорса.

Необходимо изучить, в какой позиции начинают отличаться соседние хвосты в каждой цепи. Также вызывает интерес то, с какой "частотой" эта позиция попадет в p-хвост для некоторого $p \leqslant n$. Потом мы несколько обобщаем наши рассуждения, деля хвосты на сегменты по несколько букв, а затем рассматривая, в какой сегмент попала позиция, в которой начинают отличаться друг от друга соседние хвосты в цепи. В лемме 3.2 связываются рассматриваемые "частоты" для p-хвостов и kp-хвостов для $k = 3$.

В завершение доказательства строится иерархическая структура на основе применения леммы 3.2, т.е. рассматриваются сначала сегменты n-хвостов, потом подсегменты этих сегментов и т.д. Далее мы рассматриваем наибольшее возможное количество хвостов из подмножества, для которого была применена теорема Дилуорса, после чего оцениваем сверху общее количество хвостов, а, значит, и букв слова W.

2.2. Свойства периодичности и n-разбиваемости. Пусть $a_1, \ldots, a_l$ – алфавит, над которым проводится построение слов. Порядок $a_1 \prec a_2 \prec \cdots \prec a_l$ индуцирует лексикографический порядок на словах над заданным алфавитом. Для удобства введем следующие определения.

Определение 2.1. а) Если в слове v содержится подслово вида u^t, то будем говорить, что в слове v содержится период длины t.

б) Если слово u является началом слова v, то такие слова называют *несравнимыми*.

в) Слово v – *хвост* слова u, если найдется слово w такое, что $u = wv$.

г) Слово v – *k-хвост* слова u, если v состоит из k первых букв некоторого хвоста u.

г*) *k-начало* – то же самое, что и k-хвост.

д) Пусть слово u *левее* слова v, если начало слова u левее начала слова v.

Обозначим через $|u|$ длину слова u.

Для доказательства потребуются следующие достаточные условия наличия периода.

Лемма 2.1. *В слове W длины x либо первые $[x/d]$ хвостов попарно сравнимы, либо в слове W найдется период длины d.*

Доказательство. Пусть в слове W не нашлось слова вида u^d. Рассмотрим первые $[x/d]$ хвостов. Предположим, что среди них нашлись 2 несравнимых хвоста v_1 и v_2. Пусть $v_1 = u \cdot v_2$. Тогда $v_2 = u \cdot v_3$ для некоторого v_3. Тогда $v_1 = u^2 \cdot v_3$. Применяя такие рассуждения, получим, что $v_1 = u^d \cdot v_{d+1}$, так как $|u| < x/d$, $|v_2| \geqslant (d-1)x/d$. Противоречие.

Лемма 2.2. *Если в слове V длины $k \cdot t$ не больше k различных подслов длины k, то V включает в себя период длины t.*

Доказательство. Докажем лемму индукцией по k. База при $k = 1$ очевидна. Если найдется не больше, чем $(k-1)$ различных подслов длины $(k-1)$, то применим индукционное предположение. Если существуют k различных подслов длины $(k-1)$, то каждое подслово длины k однозначно определяется своими первыми $(k-1)$ буквами. Значит, $V = v^t$, где v – k-хвост V.

Определение 2.2. а) Слово W – *n-разбиваемо в обычном смысле*, если найдутся $u_1, u_2, \ldots, u_n$ такие, что $W = v \cdot u_1 \cdots u_n$, при этом $u_1 \succ \cdots \succ u_n$.

б) В текущем доказательстве слово W будем называть *n-разбиваемым в хвостовом смысле*, если найдутся хвосты $u_1, \ldots, u_n$ такие, что $u_1 \succ u_2 \succ \cdots \succ u_n$ и для любого $i = 1, 2, \ldots, n-1$ начало u_i – слева от начала u_{i+1}. Если особо не оговорено противное, то под *n-разбиваемыми* словами мы подразумеваем n-разбиваемые в хвостовом смысле.

в) Слово W – *n-сократимое*, если оно либо n-разбиваемо в обычном смысле, либо найдется слово вида $u^d \subseteq W$.

Теперь опишем достаточное условие n-сократимости и его связь с n-разбиваемостью.

Лемма 2.3. *Если в слове W найдутся n одинаковых непересекающихся подслов u длины $n \cdot d$, то W n-сократимое.*

Доказательство. Предположим противное. Рассмотрим хвосты $u_1, u_2, \ldots, u_n$ слова u, которые начинаются с каждой из его первых n букв. Перенумеруем хвосты так, чтобы выполнялись неравенства $u_1 \succ \cdots \succ u_n$. По лемме 2.1

они несравнимы. Рассмотрим подслово u_1, лежащее в самом левом экземпляре слова u, подслово u_2 – во втором слева, ..., u_n – в n-м слева. Получили n-разбиение слова W. Противоречие.

Лемма 2.4. *Если слово W является $4nd$-разбиваемым, то оно n-сократимое.*

Доказательство. Предположим противное. Рассмотрим порядковые номера позиций букв a_i, где $a_1 < a_2 < \cdots < a_{4nd}$, с которых начинаются хвосты u_i, разбивающие W. Положим $a_{4nd+1} = |W|$. Если W – не n-сократимое, то найдется такое число i, что $1 \leqslant i \leqslant 4(n-1)d+1$ и для любых $i \leqslant b < c \leqslant d < e \leqslant i + 4d$ несравнимы $(a_c - a_b)$-хвост u_b и $(a_e - a_d)$-хвост u_d. Сравним числа $a_{i+2d} - a_i$ и $a_{i+4d} - a_{i+2d}$. Можно считать, что $a_{i+4d} - a_{i+2d} \geqslant a_{i+2d} - a_i$. Пусть $a_{j+1} - a_j = \inf_k (a_{k+1} - a_k)$, $0 \leqslant j < 2d$. Сравним числа j и d. Можно считать, что $j < d$. По предположению $(a_{2d} - a_j)$-хвост u_j и $(a_{2d} - a_{j+1})$-хвост u_{j+1} несравнимы с $(a_{4d} - a_{2d})$-хвостом u_{2d}. Так как $a_{4d} - a_{2d} \geqslant a_{2d} - a_j > a_{2d} - a_{j+1}$, то $(a_2d - a_j)$-хвост u_j и $(a_{2d} - a_{j+1})$-хвост u_{j+1} несравнимы между собой. Так как

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

то $(a_{j+1} - a_j)$-хвост u_j в степени d содержится в $(a_2d - a_j)$-хвосте u_j. Противоречие.

Следствие 2.1. *Если слово W – не n-разбиваемо в обычном смысле, то W не $4nd$-разбиваемо (в хвостовом смысле).*

Положим $p_{n,d} := 4nd - 1$.

Пусть W не n-сократимое слово. Тогда W не $(p_{n,d} + 1)$-разбиваемое. Рассмотрим $U - [|W|/d]$-хвост слова W. Пусть Ω – множество хвостов слова W, которые начинаются в U. Тогда по лемме 2.1 любые два элемента из Ω сравнимы. Естественным образом строится биекция между Ω, буквами U и натуральными числами от 1 до $|\Omega| = |U|$.

Введем слово θ такое, что θ лексикографически меньше любого слова.

Замечание 2.1. В текущем доказательстве теоремы 7 все хвосты мы предполагаем лежащими в Ω.

§ 3. Оценки на появление периодических фрагментов

Применение теоремы Дилуорса. Для хвостов u и v положим $u < v$, если $u \prec v$ и, кроме того, u левее v. Тогда по теореме Дилуорса Ω можно разбить на $p_{n,d}$ цепей, где в каждой цепи $u \prec v$, если u левее v. Покрасим начальные позиции хвостов в $p_{n,d}$ цветов в соответствии с принадлежностью к цепям. Фиксируем натуральное число p. Каждому натуральному числу i от 1 до $|\Omega|$ сопоставим $B^p(i)$ – упорядоченный набор из $p_{n,d}$ слов $\{f(i,j)\}$, построенных по следующему правилу.

Для каждого $j = 1, 2, \ldots, p_{n,d}$ положим

$$f(i,j) = \{\max f \leqslant i : f \textit{ раскрашено в цвет } j\}.$$

Если такого f не найдется, то слово из $B^p(i)$ на позиции j считаем равным θ, в противном случае это слово считаем равным p-хвосту, который начинается с $f(i,j)$-ой буквы.

Неформально говоря, мы наблюдаем, с какой скоростью хвосты "эволюционируют" в своих цепях, если рассматривать последовательность позиций слова W как ось времени.

3.1. Наборы $B^p(i)$, процесс на позициях.

Лемма 3.1 (о процессе). *Дана последовательность S длины $|S|$, составленная из слов длины $k-1$. Каждое из них состоит из $k-2$ символов* "0" *и одной* "1". *Пусть S удовлетворяет следующему условию*:

если для некоторого $0 < s \leqslant k-1$ найдутся $p_{n,d}$ слов, в которых "1" *стоит на s-м месте, то между первым и $p_{n,d}$-м из этих слов найдется слово, в котором* "1" *стоит строго меньше, чем на s-м месте*; $L(k-1) = \sup_S |S|$.

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

Доказательство. $L(1) \leqslant p_{n,d} - 1$. Пусть $L(k-1) \leqslant p_{n,d}^{k-1} - 1$. Покажем, что $L(k) \leqslant p_{n,d}^{k} - 1$. Рассмотрим слова, у которых символ "1" стоит на первом месте; их не больше $p_{n,d} - 1$. Между любыми двумя из них, а также перед первым и после последнего, количество слов не больше $L(k-1) \leqslant p_{n,d}^{k-1} - 1$. Получаем, что

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

Нам требуется ввести некоторую величину, которая бы численно оценивала скорость "эволюции" наборов $B^p(i)$. Положим

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

В частности, по лемме 2.2 $\psi(p_{n,d}) \leqslant p_{n,d}d$.

Для заданного α определим разбиение последовательности первых $|\Omega|$ позиций i слова W на классы эквивалентности $\sim_\alpha$ следующим образом: $i \sim_\alpha j$, если $B^\alpha(i) = B^\alpha(j)$.

Предложение 3.1. *Для любых натуральных $a < b$ имеем $\psi(a) \leqslant \psi(b)$.*

Лемма 3.2 (основная). *Для любых натуральных чисел a, k верно неравенство*

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

Доказательство. Рассмотрим по наименьшему представителю из каждого класса $\sim_{k\cdot a}$. Получена последовательность позиций $\{i_j\}$. Теперь рассмотрим все i_j и $B^{k\cdot a}(i_j)$ из одного класса эквивалентности по $\sim_a$. Пусть он состоит из $B^{k\cdot a}(i_j)$ при $i_j \in [b,c)$. Обозначим за $\{i_j\}'$ отрезок последовательности $\{i_j\}$, для которого $i_j \in [b, c - k\cdot a)$.

Фиксируем некоторое натуральное число r, $1 \leqslant r \leqslant p_{n,d}$. Назовем все $k\cdot a$-начала цвета r, начинающиеся с позиций слова W из $\{i_j\}'$, представителями типа r. Все представители типа r будут попарно различны, так как

они начинаются с наименьших позиций в классах эквивалентности по $\sim_{k\cdot a}$. Разобьем каждый представитель типа r на k сегментов длины a. Пронумеруем сегменты внутри каждого представителя типа r слева направо числами от нуля до $k-1$. Если найдутся $p_{n,d}+1$ представителей типа r, у которых совпадают первые $t-1$ сегментов, но которые попарно различны в t-м, где t – натуральное число, $1 \leqslant t \leqslant k-1$, то найдутся две первых буквы t-го сегмента одного цвета. Тогда позиции, с которых начинаются эти сегменты, входят в разные классы эквивалентности по $\sim_a$.

Применим лемму 3.1 следующим образом: во всех представителях типа r, кроме самого правого, будем считать сегменты *единичными*, если именно в них находится наименьшая позиция, в которой текущий представитель типа r отличается от предыдущего. Остальные сегменты считаем *нулевыми*.

Теперь можно применить лемму 3.1 (о процессе) с параметрами, совпадающими с заданными в условии леммы. Получаем, что в последовательности $\{i_j\}'$ будет не более $p_{n,d}^{k-1}$ представителей типа r. Тогда в последовательности $\{i_j\}'$ будет не более $p_{n,d}^{k}$ членов. Таким образом, $c-b \leqslant p_{n,d}^{k}\psi(k\cdot a)+k\cdot a$.

3.2. Завершение доказательств теорем 7 и 8. Пусть

$$a_0 = 3^{\lceil \log_3 p_{n,d} \rceil}, \quad a_1 = 3^{\lceil \log_3 p_{n,d} \rceil - 1}, \quad \ldots, \quad a_{\lceil \log_3 p_{n,d} \rceil} = 1.$$

При этом $|W| \leqslant d|\Omega| + d$ в силу леммы 2.1.

Так как набор $B^1(i)$ принимает не более $1 + p_{n,d}l$ различных значений, то $|W| \leqslant d(1+p_{n,d}l)\psi(1)+d$. По лемме 3.2

$$\begin{aligned}\psi(1) &< (p_{n,d}^3 + p_{n,d})\psi(3) < (p_{n,d}^3 + p_{n,d})^2\psi(9)\\ &< \cdots < (p_{n,d}^3 + p_{n,d})^{\lceil \log_3 p_{n,d} \rceil}\psi(p_{n,d}) \leqslant (p_{n,d}^3 + p_{n,d})^{\lceil \log_3 p_{n,d} \rceil} p_{n,d} d.\end{aligned}$$

Подставляя $p_{n,d} = 4nd-1$, получаем

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

Отсюда имеем утверждение теоремы 7.

Доказательство теоремы 8 завершается также, только вместо последовательности

$$a_0 = 3^{\lceil \log_3 p_{n,d} \rceil}, \quad a_1 = 3^{\lceil \log_3 p_{n,d} \rceil - 1}, \quad \ldots, \quad a_{\lceil \log_3 p_{n,d} \rceil} = 1$$

рассматривается последовательность

$$a_0 = 2^{\lceil \log_2 p_{n,d} \rceil}, \quad a_1 = 2^{\lceil \log_2 p_{n,d} \rceil - 1}, \quad \ldots, \quad a_{\lceil \log_2 p_{n,d} \rceil} = 1.$$

§ 4. Оценка существенной высоты

В этом параграфе мы продолжаем доказывать основную теорему 5. Попутно доказывается теорема 9. Будем смотреть на позиции букв слова W как на ось времени, т.е. подслово u встретилось раньше подслова v, если u целиком лежит левее v внутри слова W.

4.1. Вычленение различных периодических фрагментов в слове W. Обозначим через s количество подслов слова W с периодом длины меньше n, в которых период повторяется больше $2n$ раз и которые попарно разделены сравнимыми с предыдущим периодом подсловами длины больше n. Пронумеруем их от начала к концу слова: $x_1^{2n}, x_2^{2n}, \dots, x_s^{2n}$. Таким образом,

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

Если найдется i такое, что длина слова x_i не меньше n, то в слове x_i^2 найдутся n попарно сравнимых хвостов, а, значит, слово x_i^{2n} – n-разбиваемое. Получаем, что число s не меньше, чем существенная высота слова W над множеством слов длины меньше n.

Определение 4.1. Слово u назовем *нециклическим*, если u нельзя представить в виде v^k, где $k > 1$.

Определение 4.2. *Слово-цикл* u – слово u со всеми его сдвигами по циклу.

Определение 4.3. Слово W называется *сильно n-разбиваемым*, если его можно представить в виде $W = W_0 W_1 \cdots W_n$, где подслова $W_1, \dots, W_n$ идут в порядке лексикографического убывания, и каждое из слов W_i, $i = 1, 2, \dots, n$, начинается с некоторого слова $z_i^k \in Z$, все z_i различны.

Лемма 4.1. *Если найдется число $m, 1 \leqslant m < n$, такое, что существуют $2n-1$ попарно несравнимых слов длины m $x_{i_1}, \dots, x_{i_{2n-1}}$, то W – n-разбиваемое.*

Доказательство. Положим $x := x_{i_1}$. Тогда в слове W найдутся непересекающиеся подслова $x^{p_1} v_1', \dots, x^{p_{2n-1}} v_{2n-1}'$, где $p_1, \dots, p_{2n-1}$ – некоторые натуральные числа, большие n, а $v_1', \dots, v_{2n-1}'$ – некоторые слова длины m, сравнимые с x, $v_1' = v_{i_1}$. Тогда среди слов $v_1', \dots, v_{2n-1}'$ найдутся либо n лексикографически больших x, либо n лексикографически меньших x. Можно считать, что $v_1', \dots, v_n'$ лексикографически больше x. Тогда в слове W найдутся подслова $v_1', xv_2', \dots, x^{n-1} v_n'$, идущие слева направо в порядке лексикографического убывания.

Рассмотрим некоторое число m, $1 \leqslant m < n$. Разобьем все x_i длины m на эквивалентности по сильной несравнимости и выберем по одному представителю из каждого класса эквивалентности. Пусть это слова $x_{i_1}, \dots, x_{i_{s'}}$, где s' – некоторое натуральное число. Так как подслова x_i являются периодами, будем рассматривать их как слово-циклы.

Обозначим $v_k := x_{i_k}$.

Пусть $v(k, i)$, где i – натуральное число от 1 до m, – циклический сдвиг слова v_k на $(k-1)$ позиций вправо, т.е. $v(k, 1) = v_k$, а первая буква слова $v(k, 2)$ является второй буквой слова v_k. Таким образом, $\{v(k, i)\}_{i=1}^m$ – слово-цикл слова v_k. Заметим, что для любых $1 \leqslant i_1, i_2 \leqslant p$, $1 \leqslant j_1, j_2 \leqslant m$ слово $v(i_1, j_1)$ сильно несравнимо со словом $v(i_2, j_2)$.

Замечание 4.1. Случаи $m = 2, 3, n-1$ рассмотрены в работах [31], [27].

4.2. Применение теоремы Дилуорса. Рассмотрим множество слов $\Omega' = \{v(i,j)\}$, где $1 \leqslant i \leqslant p$, $1 \leqslant j \leqslant m$. Введем следующий порядок на словах $v(i,j)$: $v(i_1,j_1) \succ v(i_2,j_2)$, если $v(i_1,j_1) > v(i_2,j_2)$ и $i_1 > i_2$.

Лемма 4.2. *Если в множестве Ω' для порядка $\succ$ найдется антицепь длины n, то слово W будет n-разбиваемым.*

Доказательство. Пусть нашлась антицепь длины n из слов

$$v(i_1,j_1), v(i_2,j_2), \ldots, v(i_n,j_n), \qquad i_1 \leqslant i_2 \leqslant \cdots \leqslant i_n.$$

Если все неравенства между i_k строгие, то слово W – n-разбиваемое по определению.

Предположим, что для некоторого числа r нашлись $i_{r+1} = \cdots = i_{r+k}$, где либо $r = 0$, либо $i_r < i_{r+1}$. Кроме того, k – такое натуральное число, что либо $k = n - r$, либо $i_{r+k} < i_{r+k+1}$.

Слово $s_{i_{r+1}}$ периодическое, следовательно, оно представляется в виде произведения n экземпляров слова $v^2_{i_{r+1}}$, которое содержит слово-цикл $v_{i_{r+1}}$. Значит, в слове $s_{i_{r+1}}$ можно выбрать непересекающиеся подслова, идущие в порядке лексикографического убывания, равные $v(i_{r+1}, j_{r+1}), \ldots, v(i_{r+k}, j_{r+k})$ соответственно. Таким же образом поступаем со всеми множествами равных индексов в последовательности $\{i_r\}_{r=1}^{n}$. Получаем n-разбиваемость слова W. Противоречие.

Значит, множество Ω' можно разбить на $n-1$ цепь.

Положим $q_n = n - 1$.

4.3. Наборы $C^\alpha(i)$, процесс на позициях. Покрасим первые буквы слов из Ω' в q_n цветов в соответствии с принадлежностью к цепям. Покрасим также числа от 1 до $|\Omega'|$ в соответствующие цвета. Фиксируем натуральное число $\alpha \leqslant m$. Каждому числу i от 1 до $|\Omega'|$ сопоставим упорядоченный набор слов $C^\alpha(i)$, состоящий из q_n слов по следующему правилу.

Для каждого $j = 1, 2, \ldots, q_n$ положим $f(i,j) = \{\max f \leqslant i$: существует k такое, что $v(f,k)$ раскрашено в цвет j и α-хвост, который начинается с f, состоит только из букв, являющихся первыми буквами хвостов из $\Omega'\}$.

Если такого f не найдется, то слово из $C^\alpha(i)$ считаем равным θ, в противном случае это слово считаем равным α-хвосту слова $v(f,k)$.

Положим $\varphi(a) = \{\max k$: для некоторого i верно $C^a(i) = C^a(i+k-1)\}$.

Для заданного $a \leqslant m$ определим разбиение последовательности слово-циклов $\{i\}$ слова W на классы эквивалентности следующим образом: $i \sim_a j$, если $C^a(i) = C^a(j)$.

Заметим, что построенная конструкция во многом аналогична построенной в доказательстве теоремы 7. Обращаем внимание на схожесть $B^a(i)$ и $C^a(i)$, а также $\psi(a)$ и $\varphi(a)$.

Лемма 4.3. *Во введенных ранее обозначениях $\varphi(m) \leqslant q_n/m$.*

Доказательство. В п. 4.1 слово-циклы были пронумерованы. Рассмотрим слово-циклы с номерами $i, i+1, \ldots, i+[q_n/m]$. Ранее было показано, что каждый слово-цикл состоит из m различных слов. Рассмотрим теперь слова в слово-циклах $i, i+1, \ldots, i+[q_n/m]$ как элементы множества Ω'. При таком рассмотрении у первых букв из слово-циклов появляются свои позиции. Всего рассматриваемых позиций не меньше n. Следовательно, среди них найдутся две позиции одного цвета. Тогда в силу сильной несравнимости слово-циклов имеем утверждение леммы.

Предложение 4.1. *Для любых натуральных $a < b$ имеем $\varphi(a) \leqslant \varphi(b)$.*

Лемма 4.4 (основная). *Для натуральных чисел a, k таких, что $ak \leqslant m$, верно неравенство*

$$\varphi(a) \leqslant q_n^k \varphi(k \cdot a).$$

Доказательство. Рассмотрим по наименьшему представителю из каждого класса $\sim_{k\cdot a}$. Получена последовательность позиций $\{i_j\}$. Теперь рассмотрим все i_j и $C^{k\cdot a}(i_j)$ из одного класса эквивалентности по $\sim_a$. Пусть он состоит из $C^{k\cdot a}(i_j)$ при $i_j \in [b, c)$. Обозначим за $\{i_j\}'$ отрезок последовательности $\{i_j\}$, для которого $i_j \in [b, c)$.

Фиксируем некоторое натуральное число r, $1 \leqslant r \leqslant q_n$. Назовем все $k \cdot a$-начала цвета r, начинающиеся с позиций слова W из $\{i_j\}'$, представителями типа r. Все представители типа r будут попарно различны, так как они начинаются с наименьших позиций в классах эквивалентности по $\sim_{k\cdot a}$. Разобьем каждый представитель типа r на k сегментов длины a. Пронумеруем сегменты внутри каждого представителя типа r слева направо числами от нуля до $k-1$. Если найдутся $q_n + 1$ представителей типа r, у которых совпадают первые $t-1$ сегментов, но которые попарно различны в t-м, где t – натуральное число, $1 \leqslant t \leqslant k-1$, то найдутся две первых буквы t-го сегмента одного цвета. Тогда позиции, с которых начинаются эти сегменты, входят в разные классы эквивалентности по $\sim_a$.

Применим лемму 3.1 следующим образом: во всех представителях типа r, кроме самого правого, будем считать сегменты единичными, если именно в них находится наименьшая позиция, в которой текущий представитель типа r отличается от предыдущего. Остальные сегменты считаем нулевыми.

Теперь мы можем применить лемму 3.1 (о процессе) с параметрами, совпадающими с заданными в условии леммы. Получаем, что количество представителей типа r в последовательности $\{i_j\}'$ не превосходит q_n^{k-1}. Тогда в последовательности $\{i_j\}'$ будет не более q_n^k членов. Таким образом, $c - b \leqslant q_n^k \varphi(k \cdot a)$.

4.4. Завершение доказательства теоремы 9. Пусть

$$a_0 = 3^{\lceil \log_3 p_{n,d} \rceil}, \quad a_1 = 3^{\lceil \log_3 p_{n,d} \rceil - 1}, \quad \ldots, \quad a_{\lceil \log_3 p_{n,d} \rceil} = 1.$$

Подставляя эти a_i в леммы 4.4 и 4.3, получаем, что

$$\varphi(1) \leqslant q_n^3 \varphi(3) \leqslant q_n^9 \varphi(9) \leqslant \cdots \leqslant q_n^{3^{\lceil \log_3 m \rceil}} \varphi(m) \leqslant q_n^{3^{\lceil \log_3 m \rceil + 1}}.$$

Так как C_i^1 принимает не более $1+q_nl$ различных значений, то

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

По лемме 4.1 получаем, что количество x_i длины m меньше $2n^{3^{\lceil \log_3 n \rceil+3}}l$.

Имеем, что количество всех x_i меньше $2n^{3^{\lceil \log_3 n \rceil+4}}l$, т.е. $s < 2n^{3^{\lceil \log_3 n \rceil+4}}l$. Таким образом, теорема 9 доказана.

§ 5. Доказательство основной теоремы 5 и теоремы 6

5.1. План доказательства. Будем далее под *n-разбиваемым словом* подразумевать n-разбиваемое в обычном смысле. Сначала мы находим необходимое количество фрагментов с длиной периода не меньше $2n$ в слове W. Это можно сделать, просто разбив слово W на подслова большой длины, к которым применяется теорема 7. Однако мы можем улучшить оценку, если сначала выделим в слове W периодический фрагмент с длиной периода не менее $4n$, затем рассмотрим W_1 – слово W с "вырезанным" периодическим фрагментом u_1. У слова W_1 выделяем фрагмент с длиной периода не менее $4n$, после чего рассматриваем W_2 – слово W_1 с "вырезанным" периодическим фрагментом u_2. У слова W_2 так же вырезаем периодический фрагмент. Далее продолжаем этот процесс, подробнее описанный в алгоритме ниже. Затем мы по вырезанным фрагментам восстанавливаем первоначальное слово W. После этого показывается, что в слове W подслово u_i чаще всего не является произведением большого количества не склеенных подслов. В лемме 5.1 доказывается, что применение алгоритма дает необходимое количество подслов слова W с длиной периода не меньше $2n$ среди вырезанных подслов.

5.2. Суммирование существенной высоты и степени нильпотентности. Пусть $\mathrm{Ht}(w)$ – высота слова w над множеством слов степени не выше n. Рассмотрим слово W с высотой $\mathrm{Ht}(W) > \Phi(n,l)$. Теперь для него проведем следующий алгоритм.

Алгоритм. *Шаг* 1. По теореме 7 в слове W найдется подслово с длиной периода $4n$. Пусть $W_0 = W = u_1'x_{1'}^{4n}y_1'$, причем слово $x_{1'}$ нециклическое. Представим y_1' в виде $y_1' = x_{1'}^{r_2}y_1$, где r_2 – максимально возможное число. Слово u_1' представим как $u_1' = u_1x_{1'}^{r_1}$, где r_1 – наибольшее возможное. Обозначим через f_1 следующее слово:

$$W_0 = u_1x_{1'}^{4n+r_1+r_2}y_1 = u_1f_1y_1.$$

Назовем позиции, входящие в слово f_1, *скучными*, последнюю позицию слова u_1 – *скучной типа* 1, вторую с конца позицию u_1 – *скучной типа* 2, и так далее, n-ю с конца позицию u_1 – *скучной типа* n. Положим $W_1 = u_1y_1$.

Шаг k. Рассмотрим слова u_{k-1}, y_{k-1}, $W_{k-1} = u_{k-1}y_{k-1}$, построенные на предыдущем шаге. Если $|W_{k-1}| \geqslant \Phi(n,l)$, то применим теорему 7 к слову W с тем условием, что процесс в основной лемме 3.2 будет вестись только по не скучным позициям и скучным позициям типа больше ka, где k и a – параметры леммы 3.2.

Таким образом, в слове W_{k-1} найдется нециклическое подслово с длиной периода $4n$, так что

$$W_{k-1} = u'_k x_{k'}^{4n} y'_k.$$

При этом положим

$$r_1 := \sup\{r\colon u'_k = u_k x_{k'}^r\}, \qquad r_2 := \sup\{r : y'_k = x_{k'}^r y_k\}.$$

(Отметим, что слова в наших рассуждениях могут быть пустыми.)

Определим f_k из равенства

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

Назовем позиции, входящие в слово f_k, *скучными*, последнюю позицию слова u_k – *скучной типа* 1, вторую с конца позицию u_k – *скучной типа* 2, и так далее, n-ю с конца позицию u_k – *скучной типа* n. Если позиция в процессе алгоритма определяется как скучная двух типов, то будем считать ее скучной того типа, который меньше. Положим $W_k = u_k y_k$.

Проведем $4t + 1$ шагов алгоритма. Рассмотрим первоначальное слово W. Для каждого натурального i из отрезка $[1, 4t]$ имеет место равенство

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

для некоторых подслов w_j. Здесь $f_i = f_i^{(1)} \cdots f_i^{(n_i)}$. Также мы считаем, что при $1 \leqslant j \leqslant n_i - 1$ подслово w_j непустое. Пусть $s(k)$ – количество индексов $i \in [1, 4t]$ таких, что $n_i = k$.

Для доказательства теоремы 7 требуется найти как можно больше длинных периодических фрагментов. Помочь в этом сможет следующая лемма.

Лемма 5.1. *Во введенных ранее обозначениях* $s = s(1) + s(2) \geqslant 2t$.

Доказательство. Назовем *монолитным* подслово U слова W, если:

1) U является произведением слов вида $f_i^{(j)}$;
2) U не является подсловом слова, для которого выполняется предыдущее свойство 1.

Пусть после $(i-1)$-го шага алгоритма в слове W содержится k_{i-1} монолитных подслов. Заметим, что $k_i \leqslant k_{i-1} - n_i + 2$.

Тогда если $n_i \geqslant 3$, то $k_i \leqslant k_{i-1} - 1$. Если же $n_i \leqslant 2$, то $k_i \leqslant k_{i-1} + 1$. При этом $k_1 = 1$, $k_t \geqslant 1 = k_1$. Лемма доказана.

Следствие 5.1. *Выполнено неравенство*

$$\sum_{k=1}^{\infty} k \cdot s(k) \leqslant 10t \leqslant 5s.$$

Доказательство. Из доказательства леммы 5.1 получаем

$$\sum_{n_i \geqslant 3} (n_i - 2) \leqslant 2t.$$

По определению $\sum_{k=1}^{\infty} s(k) = 4t$, т.е. $\sum_{k=1}^{\infty} 2s(k) = 8t$. Складывая эти два неравенства и применяя лемму 5.1, получаем доказываемое неравенство.

Предложение 5.1. *Высота слова W будет не больше*

$$\Psi(n,4n,l)+\sum_{k=1}^{\infty}k\cdot s(k)\leqslant\Psi(n,4n,l)+5s.$$

Далее будем рассматривать только f_i с $n_i\leqslant 2$. Если $n_i=1$, то положим $f_i':=f_i$. Если же $n_i=2$, то положим $f_i':=f_i^{(j)}$, где $f_i^{(j)}$ – слово с наибольшей длиной между $f_i^{(1)}$ и $f_i^{(2)}$.

Слова f_i' упорядочим в соответствии с их близостью к началу W. Получим последовательность $f'_{m_1},\dots,f'_{m_s}$, где $s'=s(1)+s(2)$; положим $f_i'':=f'_{m_i}$. Пусть $f_i''=w_i'x_{i''}^{p_{i''}}w_i''$, где хотя бы одно из слов w_i', w_i'' пустое.

Замечание 5.1. Можно считать, что мы первыми шагами алгоритма выбрали все те f_i, для которых $n_i=1$.

Теперь рассмотрим z_j' – подслова W следующего вида:

$$z_j'=x_{(2j-1)''}^{p_{(2j-1)''}+\beth}v_j,\qquad \beth\geqslant 0,\quad |v_j|=|x_{(2j-1)''}|,$$

при этом v_j не равно $x_{(2j-1)''}$, начало z_j' совпадает с началом периодического подслова в f''_{2j-1}. Покажем, что z_j' не пересекаются.

В самом деле, если $f''_{2j-1}=f_{m_{2j-1}}$, то $z_j'=f_{m_{2j-1}}v_j$.

Если же $f''_{2j-1}=f^{(k)}_{m_{2j-1}}$, $k=1,2$, а подслово z_j' пересекается с подсловом z'_{j+1}, то $f''_{2j}\subset z_i'$. Так как слова $x_{(2j)''}$ и $x_{(2j-1)''}$ нециклические, то $|x_{(2j)''}|=|x_{(2j-1)''}|$. Но тогда длина периода в z_j' не меньше $4n$, что противоречит замечанию 5.1.

Тем самым доказана следующая лемма.

Лемма 5.2. *В слове W с высотой не более $(\Psi(n,4n,l)+5s')$ найдется не менее s' непересекающихся периодических подслов, в которых период повторится не менее $2n$ раз. Кроме того, между любыми двумя элементами данного множества периодических подслов найдется подслово длины периода более левого из выбранных элементов.*

5.3. Завершение доказательств основной теоремы 5 и теоремы 6. Подставляя в лемму 5.2 вместо числа s' значение s из доказательства теоремы 9 получаем, что высота W не больше, чем

$$\Psi(n,4n,l)+5s<E_1l\cdot n^{E_2+12\log_3 n},$$

где $E_1=4^{21\log_3 4+17}$, $E_2=30\log_3 4+10$.

Тем самым мы получили утверждение основной теоремы 5.

Доказательство теоремы 6 завершается также, только в пункте 4.4 вместо последовательности

$$a_0=3^{\lceil\log_3 p_{n,d}\rceil},\quad a_1=3^{\lceil\log_3 p_{n,d}\rceil-1},\dots,\quad a_{\lceil\log_3 p_{n,d}\rceil}=1$$

рассматривается последовательность

$$a_0=2^{\lceil\log_2 p_{n,d}\rceil},\quad a_1=2^{\lceil\log_2 p_{n,d}\rceil-1},\quad \dots,a_{\lceil\log_2 p_{n,d}\rceil}=1,$$

а значение $\Psi(n,4n,l)$ берется из теоремы 8.

§ 6. Комментарии

Представленная вниманию читателя техника, по всей видимости, позволяет улучшить полученную в настоящей работе оценку, но при этом она останется только субэкспоненциальной. Для получения полиномиальной оценки, если она существует, требуются новые идеи и методы.

В начале представленного решения при использовании теоремы Ширшова подслова большого слова используются прежде всего в качестве множества независимых элементов, а не набора тесно связанных друг с другом слов. Далее используется то, что буквы внутри подслов раскрашены. При учете раскраски только первых букв подслов получается экспоненциальная оценка. При рассмотрении раскраски всех букв подслов опять получается экспонента. Данный факт имеет место из-за построения иерархической системы подслов. Не исключено, что подробное рассмотрение приведенной связи подслов вкупе с изложенным выше решением позволит улучшить полученную оценку вплоть до полиномиальной.

Интересно также получить оценки на высоту алгебры над множеством слов степени не выше сложности алгебры (в англоязычной литературе PI-degree). В работе [4] получены экспоненциальные оценки, а для слов, не являющихся линейной комбинацией лексикографически меньших, в работе [40] получены надэкспоненциальные оценки.

Глубокие идеи оригинальных работ А. И. Ширшова [1], [2], восходящие к технике элиминации в алгебрах Ли, могут оказаться чрезвычайно полезными, в том числе и для улучшения оценок, несмотря на то, что оценки на высоту, полученные в этих работах, являются только примитивно рекурсивными.

Список литературы

[1] А. И. Ширшов, “О некоторых неассоциативных ниль-кольцах и алгебраических алгебрах”, *Матем. сб.*, **41(83)**:3 (1957), 381–394.

[2] А. И. Ширшов, “О кольцах с тождественными соотношениями”, *Матем. сб.*, **43(85)**:2 (1957), 277–283.

[3] E. I. Zel’manov, “On the nilpotency of nil algebras”, *Algebra – some current trends* (Varna, 1986), Lecture Notes in Math., **1352**, Springer-Verlag, Berlin, 1988, 227–240.

[4] A. Ya. Belov, V. V. Borisenko, V. N. Latyshev, “Monomial algebras”, *J. Math. Sci. (New York)*, **87**:3 (1997), 3463–3575.

[5] A. R. Kemer, selected papers of A. I. Shirshov.

[6] A. Kanel-Belov, L. H. Rowen, selected papers of A. I. Shirshov.

[7] V. A. Ufnarovskij, “Combinatorial and asymptotic methods in algebra”, *Algebra VI*, Encyclopaedia Math. Sci., **57**, Springer-Verlag, Berlin, 1995, 1–196.

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

[9] С. В. Пчелинцев, “Теорема о высоте для альтернативных алгебр”, *Матем. сб.*, **124(166)**:4(8) (1984), 557–567; англ. пер.: S. V. Pchelintsev, “A theorem on height for alternative algebras”, *Math. USSR-Sb.*, **52**:2 (1985), 541–551.

[10] С. П. Мищенко, “Вариант теоремы о высоте для алгебр Ли”, *Матем. заметки*, **47**:4 (1990), 83–89; англ. пер.: S. P. Mishchenko, “A variant of the height theorem for Lie algebras”, *Math. Notes*, **47**:4 (1990), 368–372.

[11] А. Я. Белов, "О базисе Ширшова относительно свободных алгебр сложности n", *Матем. сб.*, **135(177)**:3 (1988), 373–384; англ. пер.: A. Ya. Belov, "On a Shirshov basis of relatively free algebras of complexity n", *Math. USSR-Sb.*, **63**:2 (1989), 363–374.

[12] A. Kanel-Belov, L. H. Rowen, *Computational aspects of polynomial identities*, Res. Notes in Math., **9**, Peters, Wellesley, MA, 2005.

[13] Gh. Ciocanu, "Independence and quasiregularity in algebras. II", *Izv. Akad. Nauk Respub. Moldova Mat.*, **1** (1997), 70–77.

[14] Г. П. Чекану, "О локальной конечности алгебр", *Матем. исслед.*, **105** (1988), 153–171.

[15] Г. П. Чекану, Е. П. Козухар, "Независимость и нильпотентность в алгебрах", *Изв. АН Молдовы. Матем.*, **2** (1993), 51–62.

[16] Г. П. Чекану, "Независимость и квазирегулярность в алгебрах", *Докл. РАН*, **337**:3 (1994), 316–319; англ. пер.: G. P. Chekanu, "Independence and quasiregularity in algebras", *Russian Acad. Sci. Dokl. Math.*, **50**:1 (1995), 84–89.

[17] В. А. Уфнаровский, "Теорема о независимости и ее следствия", *Матем. сб.*, **128(170)**:1(9) (1985), 124–132; англ. пер.: V. A. Ufnarovskiĭ, "An independence theorem and its consequences", *Math. USSR-Sb.*, **56**:1 (1987), 121–129.

[18] В. А. Уфнаровский, Г. П. Чекану, "На нильпотентных матрицах", *Матем. исслед.*, **85** (1985), 130–141.

[19] А. Я. Белов, "О рациональности рядов Гильберта относительно свободных алгебр", *УМН*, **52**:2 (1997), 153–154; англ. пер.: A. Ya. Belov, "On the rationality of Hilbert series of relatively free algebras", *Russian Math. Surveys*, **52**:2 (1997), 394–395.

[20] J. Berstel, D. Perrin, "The origins of combinatorics on words", *European J. Combin.*, **28**:3 (2007), 996–1022.

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

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

[23] В. Н. Латышев, "Комбинаторные порождающие полилинейных полиномиальных тождеств", *Фундамент. и прикл. матем.*, **12**:2 (2006), 101–110; англ. пер.: V. N. Latyshev, "Combinatorial generators of the multilinear polynomial identities", *J. Math. Sci.*, **149**:2 (2008), 1107–1112.

[24] А. Г. Колотов, "О верхней оценке высоты в конечно порожденных алгебрах с тождествами", *Сиб. матем. журн.*, **23**:1 (1982), 187–189.

[25] A. Ya. Belov, "Some estimations for nilpotence of nill-algebras over a field of an arbitrary characteristic and height theorem", *Comm. Algebra*, **20**:10 (1992), 2919–2922.

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

[27] М. И. Харитонов, "Оценки на структуру кусочной периодичности в теореме Ширшова о высоте", *Вестн. Моск. ун-та. Сер.* 1. *Матем., мех.*, В печати.

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

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

[30] Е. С. Чибриков, "О высоте Ширшова конечнопорожденной ассоциативной алгебры, удовлетворяющей тождеству степени четыре", *Изв. Алтайского гос. ун-та*, **1** (2001), 52–56.

[31] М. И. Харитонов, "Двусторонние оценки существенной высоты в теореме Ширшова о высоте", *Вестн. Моск. ун-та. Сер.* 1. *Матем., мех.*, 2012, № 2, 24–28.

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

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

[34] *Днестровская тетрадь. Нерешенные проблемы теории колец и модулей*, 4-е изд., Изд. ин-та матем. СО АН СССР, Новосибирск, 1993.

[35] И. И. Богданов, "Теорема Нагаты–Хигмана для полуколец", *Фундамент. и прикл. матем.*, **7**:3 (2001), 651–658.

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

[37] А. Я. Белов, "Размерность Гельфанда–Кириллова относительно свободных ассоциативных алгебр", *Матем. сб.*, **195**:12 (2004), 3–26; англ. пер.: A. Ya. Belov, "The Gel'fand–Kirillov dimension of relatively free associative algebras", *Sb. Math.*, **195**:12 (2004), 1703–1726.

[38] Е. Н. Кузьмин, "О теореме Нагаты–Хигмана", *Сборник трудов, посвященный* 60-*летию акад. Илиева*, София, 1975, 101–107.

[39] Ю. П. Размыслов, *Тождества алгебр и их представлений*, Наука, М., 1989; англ. пер.: Yu. P. Razmyslov, *Identities of algebras and their representations*, Transl. Math. Monogr., **138**, Amer. Math. Soc., Providence, RI, 1992.

[40] А. Я. Белов, "Проблемы бернсайдовского типа, теоремы о высоте и о независимости", *Фундамент. и прикл. матем.*, **13**:5 (2007), 19–79; англ. пер.: A. Ya. Belov, "Burnside-type problems, theorems on height, and independence", *J. Math. Sci.*, **156**:2 (2009), 219–260.

А. Я. Белов (A. Ya. Belov)
Московский институт открытого образования
E-mail: kanel@mccme.ru

М. И. Харитонов (M. I. Kharitonov)
Московский государственный университет
им. М. В. Ломоносова
E-mail: mikhailo.kharitonov@gmail.com

Поступила в редакцию
12.02.2011 и 17.10.2011

Contents

Preface

Anatolii Illarionovich Shirshov (1921–1981) was an outstanding Russian mathematician whose works made a decisive contribution to the theory of associative, Lie, Jordan, and alternative rings. He created a large scientific school whose representatives have worked successfully in many different areas of algebra. For a period of fifteen years (1959–1973), A.I. Shirshov was Deputy Director of the (now Sobolev) Institute of Mathematics of the Siberian Branch of the Russian Academy of Sciences (the Director was S.L. Sobolev) and in this and other positions he made a substantial contribution to the organization and early development of both the Sobolev Institute and the entire Siberian Branch of the Academy.

The present collection contains English translations (by M. Bremner and M. Kochetov) of all the published scientific works of A.I. Shirshov with the exception of his book *Rings that are nearly associative*, M: Nauka, 1978 (With K.A. Zhevlakov, A.M. Slinko and I.P. Shestakov) (translated by H.F. Smith, N.Y.: Academic Press, 1982) and some articles whose content is included in later more extensive publications. The works are ordered chronologically.

February 2009

L.A. Bokut
V. Latyshev
I.P. Shestakov
E.I. Zelmanov

Commentaries

Perspectives on Shirshov's Height Theorem

Alexei Kanel-Belov and Louis H. Rowen

In this survey we consider the impact of Shirshov's Height Theorem on algebra. In order to avoid duplication, we often refer to Kemer's survey article [Kem09] in this volume for further details. Proofs of various quoted results are given in the book [BBL97], and in the authors' book [BR05].

1. Historical background to Shirshov's Theorem

Let F denote a field. An F-algebra is called *affine* if it is finitely generated as an algebra. An F-algebra is *algebraic* if each element a satisfies an algebraic equation over F; i.e., if the dimension $[F[a] : F] < \infty$. We say that an algebra A has PI-*degree* n if A satisfies a multilinear polynomial identity (PI) of degree n. One of the early tests of the utility of PI-theory was whether it could provide a framework for a positive solution of the following famous problem of Kurosh:

Are affine algebraic algebras necessarily finite dimensional?

Although now known to be false for associative algebras in general (cf. [Gol64]), Kurosh's problem was solved for associative PI-algebras by Kaplansky [Kap50], building on work of Jacobson and Levitzki, as described in [Kem09]. However, Kaplansky's elegant proof, relying on topology and structure theory, is not constructive.

Digression. In hindsight, Kurosh's problem for PI-algebras has an easy solution using standard results from structure theory. Here is a modification of the argument given in [Pro73]. By [Pro73, Lemma 2.6], if A is not finite-dimensional, there is a prime ideal P maximal with respect to A/P not being finite-dimensional, so we may assume that A is a prime affine algebraic PI-algebra. But then the center C of A is a field, so A is simple, by [Row88, Corollary 6.1.29], and thus

This research was supported by the Israel Science Foundation, grant #1178/06. The authors would like to thank L. Bokut, A. Kemer, E. Zelmanov, and U. Vishne for helpful comments on drafts of this survey.

finite-dimensional over C, by Kaplansky's Theorem. Then a version of the Artin-Tate Lemma [Row88, Proposition 6.2.5] says the field C is affine and thus finite-dimensional over F, implying R is finite-dimensional over F. (This argument also works more generally for affine algebras integral over a commutative Noetherian ring.)

A different approach to Kurosh's problem, taken by A.I. Shirshov [Shir57a], [Shir57b], involves the detailed analysis of words and their relations, as given in *Shirshov's Height Theorem*:

Let A be a finitely generated algebra of PI-degree d. Then there exists a finite set $Y \subset A$ and an integer $\tilde{h} \in \mathbb{N}$ such that A is linearly spanned by the set of elements of the form

$$v_1^{k_1} v_2^{k_2} \cdots v_h^{k_h} \quad \text{where} \quad h \leq \tilde{h}, \quad v_i \in Y.$$

For Y we may take the set of words of length $\leq d$. Such Y is called a *Shirshov base* of the algebra A, and $\tilde{h}$ is called the *Shirshov height* $h(A)$.

The object of this survey is to describe the impact of this pioneering theorem. Shirshov's theorem immediately yields an independent positive solution of Kurosh's problem and of other related problems for PI-algebras. Specifically, if Y is a Shirshov base consisting of algebraic elements, then the algebra A is finite-dimensional. Thus, Shirshov's theorem explicitly determines the set of elements whose algebraicity implies algebraicity of the whole algebra. (It is worth noting that Procesi [Pro73] later discovered a structural proof of Shirshov's theorem also, by means of reducing first to prime rings and then utilizing traces.) We also have

Corollary 1.1. *If A is a PI-algebra of PI-degree d and all words in its generators of length $\leq d$ are algebraic, then A is locally finite.*

Let us briefly sketch the proof of Shirshov's Theorem. Suppose that $A = F\{a_1, \ldots, a_\ell\}$ is an affine algebra. Ordering the letters $a_1 < \cdots < a_\ell$ induces the *lexicographic* order on the set Ω^* of words in the generators $\{a_1, \ldots, a_\ell\}$. We consider this as a total order, where a proper initial subword v of a word w is defined to precede w. But note that this order is not preserved under multiplication; for example $a_2 \prec a_2 a_1$ but $a_2^2 \succ (a_2 a_1)^2$. A word w is *reducible* if it can be written as a linear combination of smaller words.

Definition 1.2. *A word w is called* d-decomposable *if it contains a subword $w_1 \cdots w_d$ such that $w_1 \cdots w_d \succ w_{\pi(1)} \cdots w_{\pi(d)}$ for any permutation π of $\{1, \ldots, d\}$.*

A (multilinear) PI of degree d can be used to rewrite any d-decomposable word as a sum of smaller words; thus, the irreducible words are d-indecomposable. Shirshov proved *Shirshov's Lemma*, which asserts that, for any given $r > 0$, any long enough d-indecomposable word must contain a nonempty word u^r where $|u| \leq d$. Shirshov's height theorem follows from an algorithmic argument given in [BR05, p. 50].

Shirshov's Height Theorem also yields a result about the *Gelfand-Kirillov dimension* $\mathrm{GK}(A)$ of an affine algebra A. Recall that

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

where $V_A(n)$ is the vector space generated by the words of length $\leq n$ in the generators of A. A related concept is the (*Poincaré-*)*Hilbert Series*

$$H_A = 1 + \sum d_n \lambda^n,$$

where $d_n = \dim(V_A(n)/V_A(n-1))$, the number of irreducible words of length n. (Strictly speaking, H_A depends on the given set of generators of A, whereas $\mathrm{GK}(A)$ is independent of the choice of generators.)

Corollary 1.3 (Berele [Ber93]**).** $\mathrm{GK}(A) < \infty$*, for any affine PI-algebra* A*.*

To prove the corollary, it suffices to observe that the number of solutions of the inequality $k_1|v_1| + \cdots + k_h|v_h| \leq n$ with $h \leq \tilde{h}$ does not exceed $N^{\tilde{h}}$, and therefore $\mathrm{GK}(A) \leq h(A)$.

Shirshov's beautiful theorem, which also is formulated for algebras over arbitrary commutative rings, opened the way to the combinatoric school of PI-theory, which has led to many breakthroughs in recent years. (Ironically, Shirshov's work was unknown in the West until Amitsur brought it to attention in 1973. Thus, for many years, there was a parallel development of PI-theory on both sides of the former "iron curtain," along mostly combinatoric lines in the former Soviet Union and along structural lines in the West. Although our focus in this survey is on Shirshov's influence, and thus on the Russian school, we also describe parallel results in the West.)

1.1. The radical of an affine PI-algebra and the Nagata-Higman Theorem

One of the early applications of Shirshov's Theorem was in a seemingly unrelated direction. Using structure theory, Amitsur [Am57] showed that the Jacobson radical $J(A)$ of an affine PI-algebra is nil. This led to the question of whether $J(A)$ is nilpotent, which was formally raised by Latyshev in his dissertation. Shirshov's Theorem is a key tool in verifying this assertion when R satisfies the PI's of $n \times n$ matrices, as shown by Razmyslov [Raz74a], who also proved that a complete solution is equivalent to the conjecture that every affine PI-algebra satisfies the *standard* PI. Kemer[Kem80] verified this latter conjecture in characteristic 0. Braun [Br84] was the first to prove the nilpotence of $J(A)$ for arbitrary affine A, using the structure of Azumaya algebras. A nice exposition of Braun's theorem can also be found in Lvov [Lv83].

Incidentally, much earlier, Dubnov and Ivanov, and independently, Nagata and Higman [Hig56] showed that in characteristic 0, any nil algebra of bounded index n is nilpotent. The original bounds for the nilpotence index were exponential in n. Better bounds have been obtained as an outgrowth of Shirshov's work.

Razmyslov [Raz74b] showed that n^2 is an upper bound, and Kuzmin obtained the lower bound $\frac{n^2+n-2}{2}$, described in [BR05, p. 341].

1.2. Representable algebras

An F-algebra is called *representable* if it can be embedded into $M_n(K)$ for some field extension $K \supset F$ and some n. (More generally, we can take K commutative Noetherian, in view of [An92].) Shirshov's Theorem implies that for any representable affine PI-algebra A, one may adjoin the characteristic coefficients of finitely many words of the generators, to obtain a PI-algebra $\hat{A}$, called the *trace ring* or *characteristic closure*, which is finite over its center but also possesses a nonzero ideal contained in A. The use of this *conductor ideal*, discovered by Razmyslov [Raz74a] (and later, independently, by Schelter [Sch76]) is one of the keys to the structure of affine PI-algebras, and is used in Razmyslov's work on the Jacobson radical described above.

Another application of the characteristic closure is to the *Hilbert series* of an algebra; Answering a question raised by Procesi [Pro73], Belov proved that any relatively free, affine PI-algebra has a rational Hilbert series (with respect to a suitable set of generators); cf. [BR05, Chapter 9] for this and related results. On the other hand, Theorem 3.5 below provides examples of representable algebras with non-rational Hilbert series.

1.3. Specht's conjecture

One of the most famous problems in PI-theory was Specht's conjecture, that every set of identities is a consequence of a finite set of identities. (More formally, every T-ideal of the free algebra is finitely generated as a T-ideal.) As described in [Kem09], this question was settled affirmatively by Kemer [Kem87], [Kem90b] whenever the base field F is infinite, and later by Belov for arbitrary affine PI-algebras. The characteristic closure is one component of the proofs, and the nilpotence of the radical is another important aspect, so Shirshov's theorem plays an important role. The key step of Kemer's theorem is that each affine PI-algebra over an infinite field satisfies the same PI's as a suitable finite-dimensional algebra; it follows at once that the corresponding relatively free algebra is representable. (Belov extended this fact to affine algebras over arbitrary commutative Noetherian rings.)

2. Generalizations to nonassociative algebras

Shirshov's Height Theorem has been extended to various classes of nonassociative algebras. In his original paper, Shirshov applied his theorem to special Jordan algebras. Zelmanov [Zel91] obtained the following analog for ad-identities of Lie algebras:

Say an associative word in X is *special* if it is the leading word appearing in some Lie word (i.e., word with respect to the Lie multiplication). The word w is *Zelmanov d-decomposable* if it can be written as a product of subwords $w = w'w_1w_1'w_2w_2' \cdots w_dw_d'w''$ with each w_i special and $w_1 \succ w_2 \cdots \succ w_d$. Then, for

any ℓ, k, d, there is $\beta = \beta(\ell, k, d)$ such that any Zelmanov d-indecomposable word w of length $\geq \beta$ in ℓ letters must contain a nonempty subword of the form u^k, with u special.

Zelmanov's result is a major ingredient in his celebrated solution of the restricted Burnside problem. S.P. Mishchenko [Mis90] obtained an analogue of Shirshov's Height Theorem for Lie algebras with a "sparse" identity. S.V. Pchelintsev [Pch84] proved an analog for alternative and $(-1, 1)$ cases. Belov [Bel88b] proved a version for a certain class of rings asymptotically close to associative rings, including alternative and Jordan PI-algebras.

3. Questions arising in connection with Shirshov's Theorem

Shirshov's Height Theorem also gives rise to various notions, which we examine in turn.

3.1. d-decomposable words

We start with d-decomposable words; cf. Definiton 1.2. An equivalent formulation: A word w is d-decomposable if it can be written in the form $s_0 v_1 s_1 v_2 \dots s_{-1} v_d s_d$ where $v_1 \succ v_2 \succ \cdots \succ v_d$. The next proposition below demonstrates the importance of the notion of d-decomposability.

Proposition 3.1 (A.I. Shirshov).

a) *Suppose that a word w is d-decomposable. Then any word obtained from w by means of a nonidentical permutation is lexicographically less than w.*

b) *If an algebra A satisfies a PI*

$$x_1 \cdots x_d = \sum_{\sigma \neq id \in S_d} \alpha_\sigma x_{\sigma(1)} \cdots x_{\sigma(d)}$$

of degree d, then any d-decomposable word w can be written as a linear combination of words of lower order.

Thus in an algebra of PI-degree d, any word not representable as a linear combination of lower-order words is not d-decomposable, and it suffices to check that the set of d-indecomposable words has bounded height.

3.1.1. d-decomposable words and codimensions. Regev [Reg72] introduced the *codimension sequence* in order to prove that the tensor product of PI-algebras is a PI-algebra. Namely, let W_n denote the F-space of multilinear polynomials in $x_1, \dots, x_n$, and

$$c_n = \dim_F(W_n/(W_n \cap \mathrm{id}(A));$$

then c_n is exponentially bounded, for any PI-degree n.

A theorem of Dilworth enables one to bound the number of d-indecomposable words of length n by $n^{2(d-1)}$. Latyshev [Lat72] discovered a quicker proof of Regev's tensor product theorem by using Dilworth's Theorem, and showing that $c_n(A)$ is bounded by the number of d-indecomposable multilinear words. This estimate of

the codimension series led to the result of Kemer, Regev, and Amitsur that any polynomial identity whose Young tableau contains a rectangle (whose size is a suitably large function of n) is a consequence of any given polynomial identity of degree n. (This is the basis of Kemer's "super-trick" to pass from identities of nonaffine algebras to identities of affine superalgebras.)

On the other hand, there is an interesting refinement of the Hilbert series. The *multivariate Poincaré-Hilbert series* of an affine algebra $A = F\{a_1, \dots, a_\ell\}$ is defined as

$$H(A) = \sum d_{\mathbf{i}} \lambda_1^{i_1} \cdots \lambda_\ell^{i_\ell},$$

where

$$d_{\mathbf{i}} = \dim_F \left(\bar{V}_A(\mathbf{i})\right);$$

here $\mathbf{i} = (i_1, \dots, i_\ell)$, and $\bar{V}_A(\mathbf{i})$ is the vector space spanned by irreducible words of length $\leq i_u$ in the generator a_i of A, for $1 \leq u \leq \ell$.

Kemer [Kem95, §2] proved that the number of d-indecomposable multilinear words of length n equals the codimension of the space of multilinear polynomials of degree n, with traces, of $M_d(F)$. By Formanek [For84], this codimension sequence can be calculated precisely, using the multivariate Hilbert series.

Thus, Shirshov's approach motivates the use of combinatorics to compute codimensions, and to introduce the use of invariants of matrices. In this regard, Razmyslov [Raz74b], Helling[Hel74], and Procesi[Pro76], independently showed in characteristic 0 that every PI is a consequence of the Hamilton-Cayley equation (which can be written as a trace identity). This follows from the two *Fundamental Theorems of Invariant Theory*, which respectively are as follows:

- All invariants can be expressed in terms of traces.
- All relations between invariants are consequences of the Hamilton-Cayley trace identity.

In characteristic $p > 0$ one must study all of the coefficients of the Hamilton-Cayley equation as individual functions, arising from homogeneous forms (not necessarily linear), since they cannot be computed in terms of the trace. Kemer [Kem90b] developed the theory of identities involving these forms. Donkin[Do94] proved the analog of the First Fundamental Theorem of Invariant Theory, and Zubkov [Zubk96] proved the analog of the Second Fundamental Theorem.

In a similar vein, Razmyslov's student Zubrilin developed the technique of incorporating coefficients of the characteristic polynomial into Capelli polynomials, which leads to a combinatoric proof of the Razmyslov-Kemer-Braun theorem, as exposed in [BR05, §2.5].

Kemer [Kem95] showed that, unlike the situation in characteristic 0, any PI-algebra A (not necessarily affine) of characteristic $p > 0$ satisfies all the multilinear identities of a finite-dimensional algebra; combining this with the cited work of Donkin, Zubkov, and Zubrilin, yields that A satisfies all PI's of a finite-dimensional algebra; cf. [Bel00].

3.2. Estimates of Shirshov height

Shirshov's original proof was purely combinatorial (based on an elimination technique he developed for Lie algebras), but did not provide a reasonable estimate for the height. Kolotov [Kol81] obtained an estimate for $h(A) \leq s^{s^m}$ ($m = \text{PI-deg}(A)$, and s is the number of generators). In the Dniester Notebook (most recent version [Dne93]), Zelmanov asked for an exponential bound, which was obtained later by Belov [Bel88a]:

Theorem 3.1. *Suppose A is a PI-algebra of PI-degree d, generated by ℓ elements. Then the height of A over the set of words having length $\leq m$ is bounded by a function $h(m, \ell)$ where $h(m, \ell) < 2m\ell^{m+1}$.*

3.2.1. Burnside-type problems. A word $w = u^k$, for $k > 1$, is called *cyclic* or *periodic*. By problems of *Burnside type*, we mean problems related to periodic words. Combinatorics play an important role. The following basic lemma yields computational tools involving subwords which are described in [Bel07] and provide the bounds given in Theorem 3.1. The technique is illustrated in the slightly weaker result given in [BR05, Theorem 2.74].

Lemma 3.2 (on overlapping). *If two periodic words of respective periods m and n contain identical subwords having length $m+n-gcd(m, n)$ then they have identical periods.*

3.3. The essential height of an algebra

Definition 3.3. *An algebra A is said to have* essential height $\leq h$ *over a subset Y, if there is a finite set $S \subset A$ (which may depend on Y) such that A is spanned as a vector space by*

$$Y^{[h],S} = \{s_0 y_1^{m_1} s_1 \cdots s_{t-1} y_t^{m_t} s_t : \ m_i \in \mathbb{N}, \ y_i \in Y, \ s_i \in S, \ t \leq h\}.$$

In this case, Y is called an essential Shirshov base*, and S the* supplementary *set.*

Essential height is an estimate for GK-dimension; also, the converse is true for representable algebras.

Theorem 3.2 (A.Ya. Belov [BBL97]**).** *Suppose A is a finitely generated representable algebra and $H_{Ess_Y}(A) < \infty$. Then $H_{Ess_Y}(A) = \text{GK}(A)$.*

This equality is useful in both directions. First of all, it shows for a representable algebra A that $H_{Ess_Y}(A)$ is independent of the choice of Y. In the other direction, since $H_{Ess_Y}(A)$ must be an integer, one has:

Corollary 3.4 (V.T. Markov). *The Gelfand-Kirillov dimension of a representable affine algebra is an integer.*

Due to the representability of relatively free affine algebras (noted above), the Gelfand-Kirillov dimension of a relatively free algebra also equals the essential height.

Clearly, an essential Shirshov base is a Shirshov base iff it generates A as an algebra. Boundedness of essential height over Y implies a positive solution of "Kurosh's problem over Y." The converse is much less trivial.

Theorem 3.3 (A.Ya. Belov). *Suppose A is a graded PI-algebra, and Y is a finite set of homogeneous elements. Let $Y^{(n)}$ denote the ideal generated by all nth powers of elements of Y. If the algebra $A/Y^{(n)}$ is nilpotent for each n, then Y is an s-base for A. If in this situation Y generates A as an algebra, then Y is a Shirshov base for A.*

We proceed to formulate a generalization of this theorem for the non-graded case. We must confront the following counterexample to the straightforward converse of Kurosh's problem: Suppose $A = \mathbb{Q}[x, 1/x]$. Each projection π such that $\pi(x)$ is algebraic has finite-dimensional image. Nevertheless the set $\{x\}$ is not an s-base for A.

Thus we need a stronger definition:

Definition 3.5. *A set $M \subset A$ is called a Kurosh set if it satisfies the condition that for any projection $\pi\colon A \otimes K[X] \to A'$, if the image $\pi(M)$ is integral over $\pi(K[X])$, then $\pi(M)$ is finite over $\pi(K[X])$.*

Theorem 3.4 (A.Ya. Belov). *Let A be a PI-algebra, $M \subseteq A$ a Kurosh subset in A. Then M is an s-base for A.*

Thus, boundedness of essential height is a non-commutative generalization of integrality. The following proposition shows that Theorem 3.4 does generalize Theorem 3.3:

Proposition 3.6. *Let A be a graded algebra, Y a set of homogeneous elements. If the algebra $A/Y^{(n)}$ is locally nilpotent for all n, then Y is a Kurosh set.*

3.4. Normal bases and monomial algebras

Shirshov's combinatoric approach leads us to the combinatoric study of bases. Let $A = F\{a_1, \dots, a_\ell\}$ be an associative affine algebra. A word is called *reducible* if it can be written as a linear combination of lexicographically smaller words; the *normal base* of the algebra A is the set of all irreducible words in the generators; cf. [BBL97], [BRV06], [Dr00], [Lat88], [Ufn85].

A *monomial algebra* is an algebra that can be described in terms of relations that are monomials in the generators. Any affine algebra A has its *associated monomial algebra* possessing the same Hilbert series; namely one factors the free algebra by the set of reducible words in the generators of A, cf. [BR05, Proposition 9.8]. The associated monomial algebra of an algebra A also has the same Shirshov base, although it may not satisfy the same PI's. Nevertheless, their easier relations make monomial algebras a useful tool in studying Shirshov bases. This discussion follows [BRV06]; the reader should also consult [BBL97].

In case an affine monomial algebra A is PI, it has bounded essential height over a (finite) Shirshov base Y, which we may assume to be a set of words in

the generators. Take a supplementary set S as in Definition 3.3 that contains Y. Choose a subset of $Y^{[h],S}$ that spans A. Given

$$w = s_0 y_1^{m_1} s_1 \cdots s_{t-1} y_t^{m_t} s_t \tag{1}$$

(with $y_i \in Y$ and $s_i \in S$, and t bounded by the height), we rewrite it in the same manner with $s_0 \in S$ of maximal possible length, then with $y_1^{m_1}$ of maximal possible length, and so on. $(s_0, y_1, s_1, \ldots, s_{t-1}, y_t, s_t)$ is called the *type* of w. The type of a subword of a w of type θ is called a *subtype* of θ.

By an *exponential polynomial* in the variables $m_1, \ldots, m_t$ we mean an expression of the form

$$\sum f_j(m_1, \ldots, m_t) \alpha_{1j}^{m_1} \cdots \alpha_{tj}^{m_t}$$

where f_j are polynomials over a finite algebraic extension K of F, and $\alpha_{ij} \in K$. For example,

$$P(m_1, \ldots, m_t) = (5 - \sqrt{2})^{m_1} - m_2^4 \cdot 3^{m_1}$$

is an exponential polynomial over $\mathbb{Q}$.

Theorem 3.5. *A monomial algebra A over F is representable iff:*

1. *A has essential height over a finite set Y (with a supplementary set S), such that every word in the generators of A has a unique type, and there are finitely many types.*
2. *For each type $\theta = (s_0, y_1, s_1, y_2, \ldots, y_t, s_t)$, there is a finite system $P_{\theta,j}$ of exponential equations over k, in the variables $m_1, \ldots, m_t$, such that*

$$\bigcup_\theta \{ s_0 y_1^{m_1} s_1 \cdots y_t^{m_t} s_t : \quad \exists j \ P_{\theta,j}(m_1, \ldots, m_t) \neq 0 \}$$

is a normal base.

The construction of monomial algebras is thus equivalent to the solution of arbitrary exponential polynomials. But this is algorithmically unsolvable by the celebrated theorem of Davis-Putnam-Robinson [DPR61]. Thus there is no algorithm to determine whether there is an isomorphism (given in terms of the generators) for two monomial subalgebras of the matrix algebra over a polynomial ring of characteristic 0. On the other hand, this isomorphism problem is algorithmically solvable in characteristic $p > 0$. More precisely, Belov and Chilikov [BC00], [BRV06] proved over a field of characteristic p that the set of p-adic representations of exponential equations (with unknowns in $\mathbb{N}$) forms a "regular language." Thus, an inaccessible problem in characteristic 0 becomes algorithmically solvable in positive characteristic.

3.5. The conjecture of Amitsur and Shestakov

S. Amitsur and I.P. Shestakov conjectured that if the algebra A satisfies the identities of $M_n(F)$ and all words having length not exceeding n are algebraic, then A is finite-dimensional. I.V. Lvov reduced this assertion to the following:

Let $A = F\{a_1, \dots, a_\ell\}$ be a finite-dimensional subalgebra (without 1) of a matrix algebra of order n. If all words in $a_1, \dots, a_\ell$ of length $\leq n$ are nilpotent, then the algebra A is nilpotent.

Shestakov's conjecture was proved by V.A. Ufnarovsky [Ufn85] and by G.P. Chekanu [Che88]. Their *Independence Theorem* may be formulated as follows [Che88], [Ufn90]:

Theorem 3.6 (Independence Theorem). *Suppose the following is true:*

1. *a word $w = a_{i_1} \cdots a_{i_n}$ is minimal under the lexicographical order in the set of all nonzero products of length n;*
2. *all terminal subwords of w are nilpotent.*

Then the initial subwords of w are linearly independent.

Here is a key step. A word is called *extremal* if it does not lexicographically precede any nonzero word.

Lemma 3.7. *Any set of pairwise incomparable subwords of an extremal word is independent.*

To deduce I.P. Shestakov's conjecture (or, equivalently, I.V. L'vov's assertion) from this theorem, we consider the following construction:

Remark 3.8. *Given an algebra A and a right module V, the algebra $\widetilde{A}$ is defined additively as $A \oplus V$, with multiplication defined as follows: $V \cdot V = A \cdot V = 0$, and the product of elements from V and A is given by the module multiplication.*

We take a faithful representation of A acting on an n-dimensional right vector space V. Taking a base $v_1, \dots, v_n$ of this space, then, for some v_i we have $v_i w \neq 0$. Viewing V as a right A-module, we form the algebra $\tilde{A}$ of Remark 3.8, ordering the generators by $v_1 \succ \cdots \succ v_n \succ a_1 \succ \cdots \succ a_s$, and apply the Independence Theorem. Later, Belov and Chekanu showed that we may take the $\{v_i\}$ to be the set of words from Shestakov's conjecture. Another proof of this fact was obtained by V. Drensky.

The original proofs of the Independence Theorem were rather complicated. Application of *hyperwords*, described below, allow a considerable simplification.

Subsequent papers of these authors contained various refinements and generalizations of these theorems. Here is another elegant result of Chekanu [Che96]:

Theorem 3.7. *Suppose a word w is extremal and non-periodic, of length n. If $w^n \neq 0$, then the algebra generated by the letters of w contains a nilpotent element of index exactly n.*

3.6. Hyperwords in algebras

Many of the combinatorial results in this survey are most easily proved using infinite words, or *hyperwords*, so we conclude with a discussion of basic auxiliary facts and constructions related to hyperwords in algebras.

Definition 3.9. *A hyperword is a word infinite in both directions; a word infinite only to the left (resp. right) is called a left* (resp. *right*) *hyperword.*

u^∞ denotes the hyperword having period u, and $u^{\infty/2}$ the left (resp. right) hyperword having period u and terminal (resp. initial) subword u.

The context will always make clear whether we consider a left or right hyperword, so we do not distinguish the notation between them. For example, the expression $u^{\infty/2}wv^{\infty/2}$ indicates that $u^{\infty/2}$ is a left hyperword and $v^{\infty/2}$ is a right hyperword.

Right hyperwords form a linearly ordered set with respect to the lexicographical order. For a right hyperword w, we let $(w)_k$ denote the initial subword of w having length k.

Lemma 3.10 ([BBL97]**).** *Let C be an arbitrary collection of words having unbounded length. Then there exists a hyperword w such that each of its subwords is a subword of a word from C.*

Although evaluating a hyperword in an algebra does not make sense, we can define whether or not it equals 0 (according to whether some subword equals 0), and this leads to the notion of linear independence of hyperwords in A:

Definition 3.11.

a) *A hyperword w is called a zero hyperword if it includes a subword of finite length equal to 0, and a nonzero hyperword otherwise.*
b) *A finite set of right hyperwords $\{w_i\}$ is called linearly dependent if there exist $\{\alpha_i\}$ such that some of them are not zero and for all sufficiently large k we have*
$$\sum \alpha_i (w_i)_k = 0.$$
c) *Suppose w is a right hyperword in an algebra A, M is a right A-module, and $m \in M$. We say that $mw \neq 0$ if $m(w)_k \neq 0$ for all k. Otherwise $Mw = 0$.*
d) *Suppose $\{w_1, \ldots, w_n\}$ is a set of right hyperwords in an algebra A, and M is a right A-module. We say that $\sum m_i w_i = 0$ for $m_i \in M$ if $\sum m_i (w_i)_k = 0$ for all sufficiently large k.*

Proposition 3.12.

a) *A finitely generated non-nilpotent algebra A contains non-zero hyperwords.*
b) *Suppose A is a finitely generated algebra, M is a finitely generated right A-module. If $MA^k \neq 0$ for all $k > 0$, then there exist $m \in M$ and a right hyperword w such that $mw \neq 0$.*

The existence of a least upper bound and of a greatest lower bound for any set of right hyperwords implies the following

Proposition 3.13.

a) *Let w be a hyperword. Then the set of right hyperwords whose subwords are all subwords of w contains maximal and minimal hyperwords.*
b) *Suppose $\forall k \quad mA^k \neq 0$. Then the set of right hyperwords w such that $mw \neq 0$ contains a maximal and a minimal hyperword.*

c) *If A is non-nilpotent, then the set of nonzero right hyperwords in A contains a maximal and a minimal hyperword.*

Let u be the maximal word in an algebra A among all nonzero words in A having length $\leq n$. Unfortunately u may have no extension to a word of greater length. Thus, to utilize hyperwords, we need the following construction:

Construction 1. Let A be an algebra having generators $a_s \succ \cdots \succ a_1$. Put $a_1 \succ x$ and consider the free product $A' = A * F\langle x\rangle$.

Each word u in A is an initial subword of some hyperword in A'. If u is the maximal word in A among all words having length at most $|u|$, then the maximal hyperword in A' beginning with u is a hyperword in A. If $\tilde{u}$ is a hyperword in A for which each initial subword has this property, then the maximal hyperword in A' is $\tilde{u}$.

The following construction is useful for treating modules.

Construction 2. Suppose A is an algebra having generators $a_s \succ \cdots \succ a_1$, and V is a finitely generated right A-module having generators $m_k \succ \cdots \succ m_1$. Put $m_1 \succ a_s$, $a_1 \succ x$, and $\widetilde{A}$ as in Remark 3.8. Define $A'' = \widetilde{A} * F\langle x\rangle / I$ where the ideal I is generated by elements of the form xm_i.

In the algebra A'', the maximal right hyperword begins with m_k, and each word in $\widetilde{A}$ may be extended to a hyperword in A''; if $MA^k \neq 0$ for all k, then the maximal hyperword in $\widetilde{A}$ begins with some m_i.

If u is the maximal word in A among all words having length at most $|u|$ that act nontrivially on the generators of the module, then after renumbering the m_i suitably, the maximal hyperword in A'' is a hyperword in $\widetilde{A}$. If u is a hyperword in $\widetilde{A}$ such that each of its initial subwords has the above property, then the maximal hyperword in A'' is u.

Note that if an algebra has no nonzero nilpotent ideals, then any word may be extended to a hyperword. The following observation is useful.

Proposition 3.14. *If an algebra contains no nonzero periodic hyperword, then all of its words are nilpotent.*

The technique of hyperwords seems to lie rather close to the lines of structure theory, as illustrated in the following theorem and its proof, cf. [Bel07].

Theorem 3.8. *The set of irreducible words in a* PI*-algebra A has bounded height over the set of words whose degree does not exceed the PI-degree of A.*

Proof. Suppose m is the minimal degree of identities holding in an algebra A of PI-degree d. Since A has bounded height over the set of words having degree $\leq m$, it suffices to show that if $|u|$ is a nonperiodic word of length $> n$ then the word u^k for sufficiently large k is a linear combination of words of smaller lexicographic order.

Step 1. Consider the right A-module M defined by a generator v and by the relations $vw = 0$ whenever $w \prec u^{\infty/2}$. Our goal is to show that $Mu^k = 0$ for some k. Indeed, some power u^k is spanned by smaller lexicographic words. By

virtue of Shirshov's Height Theorem, the set of irreducible words has bounded height over Y_m, the set of words of degree $\leq m$. But if each sufficiently large power of a nonperiodic word having length d may be linearly represented by smaller words, then the words having length $> d$ may be excluded from Y_m.

Step 2. The correspondence $\lambda : vs \to vus$ defines a well-defined endomorphism of the module M, hence M may be considered as an $A[\lambda]$-module. Our goal is to show that $M\lambda^k = 0$ for some k, or equivalently that $\overline{M} = M \otimes \mathbb{F}[\lambda, \lambda^{-1}] = 0$.

Step 3. If $M\lambda^k \in M \cdot J(\operatorname{Ann} M)$ where $J(\operatorname{Ann} M)$ is the Jacobson radical of the annihilator of M, then $M\lambda^{\ell k} \in M \cdot J(\operatorname{Ann} M)^\ell$, and by the nilpotence of the radical, $M\lambda^{\ell k} = 0$ for sufficiently large ℓ. Hence, we may assume that $J(\operatorname{Ann} M) = 0$.

Step 4. Using primary decomposition, we reduce to the case for which M is a faithful module over a primary ring B.

Step 5. Elements of the center $Z(B)$ have trivial annihilator, so we may localize relative to them; replacing $Z(B)$ by an algebraic extension, we reduce to the case for which B is the algebra of some dimension $k \leq n$ over a field, and $\overline{M}$ is a k-dimensional vector space.

Step 6. Since M is a vector space of dimension $< |u|$, the vectors $\vec{v}u_0$, $\vec{v}u_1$, $\ldots$, $\vec{v}u_{n-1}$ are linearly dependent (where u_i is the initial subword of length i in the word u, and $u_0 = 1$). Thus we have the equality

$$\sum_{i \in I} \lambda_i \vec{v}_i u_i = 0 \tag{2}$$

where $I \subseteq \{0, \ldots, n-1\}$, $\lambda_i \in \mathbb{F}\backslash 0$. To each u_i we attach a word $u^{(i)}$ so that $u_i u^{(i)} = u^{|u|}$. Let $u^{(j)}$ be the least of those $u^{(i)}$ which are involved in the formula (2). Write the equality (2) in the form

$$\vec{v}_j u_j = \sum_{i \in I \backslash \{j\}} \beta_i \vec{v}_i u_i \tag{3}$$

where $\beta_i = -\alpha_i/\alpha_j$. But then

$$\vec{v}u^{|u|} = \vec{v}_j u_j u^{(j)} = \sum_{i \in I \backslash \{j\}} \beta_i \vec{v}_i u_i u^{(j)}. \tag{4}$$

If $i \in I\backslash\{j\}$, then $u^{(j)} \prec u^{(i)}$ and $u_i u^{(j)} \prec u_i u^{(i)} = u^{|u|}$; hence $v u_i u^{(j)} = 0$. Thus all terms in the right side of (4) are zero. Hence $\vec{v}u^{|u|} = 0$, as desired. □

Hyperwords facilitate proofs of the Independence Theorem, Shirshov's Height Theorem, nilpotence of the Lie algebra generated by sandwiches [Ufn90], proof of the *Bergman Gap Theorem*, (that any algebra of GK dimension greater than 1, has GK dimension at least 2, together with a description of the base having growth function $V_A(n) = \frac{n(n+3)}{2}$), and also describe various properties of monomial algebras [BBL97] as well as other combinatorial results for semigroups and rings.

References

[Am57] Amitsur, S.A., *A generalization of Hilbert's Nullstellensatz*, Proc. Amer. Math. Soc. **8** (1957), 649–656.

[An92] Anan'in, A.Z., *The representability of finitely generated algebras with chain condition*, Arch. Math. **59** (1992), 1–5.

[Ba87] Bakhturin, Yu.A., Identical relations in Lie algebras. Translated from the Russian by Bakhturin. VNU Science Press, b.v., Utrecht, (1987).

[Bel88a] Belov, A.Ya., *On Shirshov bases in relatively free algebras of complexity n*, Mat. Sb. **135** (1988), no. 3, 373–384.

[Bel88b] Belov, A.Ya., *The height theorem for Jordan and Lie PI-algebras*, in: Tez. Dokl. Sib. Shkoly po Mnogoobr. Algebraicheskih Sistem, Barnaul (1988), pp. 12–13.

[Bel89] Belov, A.Ya., *Estimations of the height and Gelfand-Kirillov dimension of associative PI-algebras*, In: Tez. Dokl. po Teorii Koletz, Algebr i Modulei. Mezhdunar. Konf. po Algebre Pamyati A.I.Mal'tzeva, Novosibirsk (1989), p. 21.

[Bel92] Belov, A.Ya., *Some estimations for nilpotency of nil-algebras over a field of an arbitrary characteristic and height theorem*, Commun. Algebra **20** (1992), no. 10, 2919–2922.

[Bel97] Belov, A.Ya., *Rationality of Hilbert series with respect to free algebras*, Russian Math. Surveys **52** (1997), no. 10, 394–395.

[Bel00] Belov, A.Ya., *Counterexamples to the Specht problem*, Sb. Math. **191** (3–4) (2000), 329–340.

[Bel02] Belov, A.Ya., *Algebras with polynomial idenitities: Representations and combinatorial methods*, Doctor of Science Dissertation, Moscow (2002).

[Bel07] Belov, A.Ya., *Burnside-type problems, and theorems on height and independence* (Russian), Fundam. Prikl. Mat. 13 (2007), no. 5, 19–79.

[BBL97] Belov, A.Ya., Borisenko, V.V., and Latyshev, V.N., *Monomial algebras. Algebra 4*, J. Math. Sci. (New York) **87** (1997), no. 3, 3463–3575.

[BC00] Belov, A.Ya. and Chilikov, A.A., *Exponential Diophantine equations in rings of positive characteristic* (Russian) Fundam. Prikl. Mat. **6**(3), 649–668, (2000).

[BR05] Belov, A.Ya. and Rowen, L.H. *Computational aspects of polynomial identities*. Research Notes in Mathematics **9**. AK Peters, Ltd., Wellesley, MA, 2005.

[BRV06] Kanel-Belov, A.Ya., Rowen, L.H., and Vishne, U., *Normal bases of PI-algebras*, Adv. in Appl. Math. 37 (2006), no. 3, 378–389.

[Ber93] Berele, A., *Generic verbally prime PI-algebras and their GK-dimensions*, Comm. Algebra **21** (1993), no. 5, 1487–1504.

[Bog01] Bogdanov I., *Nagata-Higman's theorem for hemirings*, Fundam. Prikl. Mat. **7** (2001), no. 3, 651–658 (in Russian).

[BLH88] Bokut', L.A., L'vov, I.V., and Harchenko, V.K., *Noncommutative rings*, In: Sovrem. Probl. Mat. Fundam. Napravl. Vol. 18, Itogi Nauki i Tekhn., All-Union Institute for Scientificand Technical Information (VINITI), Akad. Nauk SSSR, Moscow (1988), 5–116.

[Br82] Braun, A., *The radical in a finitely generated PI-algebra*, Bull. Amer. Math. Soc. **7** (1982), no. 2, 385–386.

[Br84] Braun, A., *The nilpotence of the radical in a finitely generated PI-ring*, J. Algebra **89** (1984), 375–396.

[Che88] Chekanu, G.P., *Local finiteness of algebras*. (Russian) Mat. Issled. **105**, Moduli, Algebry, Topol. (1988), 153–171, 198.

[Che95] Chekanu, G.P., *Independence and quasiregularity in algebras*, Dokl. Akad. Nauk **337** (1994), no. 3, 316–319; translation: Russian Acad. Sci. Dokl. Math. **50** (1995), no. 1, 84–89.

[Che96] Chekanu, G.P., *Independence and quasiregularity in algebras. I.* (Moldavian) Izv. Akad. Nauk Respub. Moldova Mat. 1996, no. 3, 29–39, 120, 122.

[ChUf85] Chekanu, G.P., and Ufnarovski'i, V.A., *Nilpotent matrices*, Mat. Issled. no. 85, Algebry, Kotsa i Topologi (1985), 130–141, 155.

[DPR61] Davis M., Putnam, H., and Robsinson, J. *The decision problem for exponential differential equations*, Annals of Math. **74**, 425–436, (1961).

[Dne93] Dniester Notebook (Dnestrovskaya tetrad), Sobolev Institute of Mathematics, Novosibirsk, 1993.

[Do94] Donkin, S., *Polynomial invariants of representations of quivers*, Comment. Math. Helv. **69** (1994), no. 1, 137–141.

[Dr84a] Drensky, V., *On the Hilbert series of relatively free algebras*, Comm. in Algebra, **12** no. 19 (1984), 2335–2347.

[Dr84b] Drensky, V., *Codimensions of T-ideals and Hilbert series of relatively free algebras*, J. Algebra **91** no. 1 (1984), 1–17.

[Dr00] Drensky, V., *Free Algebras and PI-algebras: Graduate Course in Algebra*, Springer-Verlag, Singapore (2000).

[DrFor04] Drensky, V. and Formanek, E., *Polynomials Identity Rings*, CRM Advanced Courses in Mathematics, Birkhäuser, Basel (2004).

[For84] Formanek, E., *Invariants and the ring of generic matrices*, J. Algebra **89** (1984), no. 1, 178–223.

[Gol64] Golod, E.S., *On nil-algebras and residually finite p-groups*, Izv. Akad. Nauk SSSR **28** (1964), no. 2, 273–276.

[Gri99] Grishin, A.V., *Examples of T-spaces and T-ideals in Characteristic 2 without the Finite Basis Property*, Fundam. Prikl. Mat. **5** (1) (1999), no. 6, 101–118 (in Russian).

[GuKr02] Gupta, C.K., and Krasilnikov, A.N., *A simple example of a non-finitely based system of polynomial identities*, Comm. Algebra **36** (2002), 4851–4866.

[Hel74] Helling, H., *Eine Kennzeichnung von Charakteren auf Gruppen und Assoziativen Algebren*, Comm. in Alg. **1** (1974), 491–501.

[Hig56] Higman, G., *On a conjecture of Nagata*, Proc. Cam. Phil. Soc. **52** (1956), 1–4.

[Ilt91] Iltyakov, A.V., *Finiteness of basis identities of a finitely generated alternative PI-algebra*, Sibir. Mat. Zh. **31** (1991), no. 6, 87–99; English translation: Sib. Math. J. **31** (1991), 948–961.

[Ilt03] Iltyakov, A.V., *Polynomial identities of Finite Dimensional Lie Algebras*, monograph (2003).

[Kap49] Kaplansky, I., *Groups with representations of bounded degree*, Canadian J. Math. **1** (1949), 105–112.

[Kap50] Kaplansky, I., *Topological representation of algebras. II*, Trans. Amer. Math. Soc. **66** (1949), 464–491.

[Kem80] Kemer, A.R., *Capelli identities and the nilpotence of the radical of a finitely generated PI-algebra*, Soviet Math. Dokl. **22** (3) (1980), 750–753.

[Kem87] Kemer, A.R., *Finite basability of identities of associative algebras* (Russian), Algebra i Logika **26** (1987), 597–641; English translation: Algebra and Logic **26** (1987), 362–397.

[Kem88] Kemer, A.R., *The representability of reduced-free algebras*, Algebra i Logika **27** (1988), no. 3, 274–294.

[Kem90a] Kemer, A.R., *Identities of Associative Algebras*, Transl. Math. Monogr., **87**, Amer. Math. Soc. (1991).

[Kem90b] Kemer, A.R. *Identities of finitely generated algebras over an infinite field* (Russian), Izv. Akad. Nauk SSSR Ser. Mat. **54** (1990), no. 4, 726–753; translation in Math. USSR-Izv. **37** (1991), no. 1, 69–96.

[Kem95] Kemer, A.R., *Multilinear identities of the algebras over a field of characteristic p*, Internat. J. Algebra Comput. **5** (1995), no. 2, 189–197.

[Kem09] Kemer, A.R., *Comments on the Shirshov's Height Theorem*, in this collection.

[Kol81] Kolotov, A.T., *Aperiodic sequences and growth functions in algebras*, Algebra i Logika **20** (1981), no. 2, 138–154.

[KrLe00] Krause, G.R., and Lenagan, T.H., *Growth of Algebras and Gelfand-Kirillov Dimension*, Amer. Math. Soc. Graduate Studies in Mathematics **22** (2000).

[Kuz75] Kuzmin, E.N., *About Nagata-Higman Theorem*, Proceedings dedicated to the 60th birthday of Academician Iliev, Sofia (1975), 101–107 (in Russian).

[Lat72] Latyshev, V.N., *On Regev's theorem on indentities in a tensor product of PI-algebras*, Uspehi Mat. Nauk. **27** (1972), 213–214.

[Lat88] Latyshev, V.N., *Combinatorial Ring Theory. Standard Bases*, Moscow University Press, Moscow (1988), (in Russian).

[Lev46] Levitzki, J., *On a problem of Kurosch*, Bull. Amer. Math. Soc. **52** (1946), 1033–1035.

[Lv83] Lvov, I.V., *Braun's theorem on the radical of PI-algebras*, Institute of Mathematics, Novosibirsk (1983), preprint.

[Mar88] Markov, V.T., *Gelfand-Kirillov dimension: nilpotence, representability, nonmatrix varieties*, In: Tez.Dokl. Sib. Shkola po Mnogoobr. Algebraicheskih Sistem, Barnaul (1988), 43–45.

[Mis90] Mishchenko, S.P. *A variant of a theorem on height for Lie algebras*. (Russian) Mat. Zametki **47** (1990), no. 4, 83–89; translation in Math. Notes **47** (1990), no. 3–4, 368–372.

[Pch84] Pchelintzev, S.V., *The height theorem for alternate algebras*, Mat. Sb. **124** (1984), no. 4, 557–567.

[Pro73] Procesi, C., *Rings with polynomial identities*, Research Notes in Mathematics **917**. Marcel Dekker, New York, 1973.

[Pro76] Procesi, C., *The invariant theory of $n \times n$ matrices*, Advances in Math. **19** (1976), 306–381.

[Raz74a] Razmyslov, Yu.P., Algebra and Logic **13** (1974), no. 3, 192–204.

[Raz74b] Razmyslov, Yu.P., *Trace identities of full matrix algebras over a field of characteristic zero*, Math. USSR Izv. **8** (1974), 724–760.

[Raz89] Razmyslov, Yu.P., *Identities of Algbras and their Representations*, Nauka, Moscow (1989).

[Reg72] Regev, A., *Existence of identities in $A \otimes B$*, Israel J. Math. 11 (1972), 131–152.

[Reg84] Regev, A., *Codimensions and trace codimensions of matrices are asymptotically equal*, Israel J. Math. **47** (1984), 246–250.

[Row88] Rowen, L.H., *Ring Theory II*, Pure and Applied Mathematics **128** Academic Press, New York, 1988.

[Sch76] Schelter, W., *Integral extensions of rings satisfying a polynomial identity*, J. Algebra **40** (1976), 245–257; errata op. cit. **44** (1977), 576.

[Sch78] Schelter, W., *Noncommutative affine PI-algebras are catenary*, J. Algebra **51** (1978), 12–18.

[Shch01] Shchigolev, V.V., *Finite basis property of T-spaces over fields of characteristic zero*, Izv. Ross. Akad. Nauk Ser. Mat. **65** (2001), no. 5, 191–224; translation: Izv. Math. **65** (2001), no. 5, 1041–1071.

[Shir57a] Shirshov, A.I., *On some nonassociative nil-rings and algebraic algebras*, Mat. Sb. **41** (1957), no. 3, 381–394.

[Shir57b] Shirshov, A.I., *On rings with identity relations*, Mat. Sb. **43**, (1957), no. 2, 277–283.

[Sp50] Specht, W., *Gesetze in Ringen I*, Math. Z. **52** (1950), 557–589.

[Ufn80] Ufnarovski'i, V.A., *On Poincaré series of graded algebras*, Mat. Zametki **27** (1980), no. 1, 21–32.

[Ufn85] Ufnarovski'i, V.A., *The independency theorem and its consequences*, Mat. Sb., **128** (1985), no. 1, 124–13.

[Ufn89] Ufnarovski'i, V.A., *On regular words in Shirshov sense*, In: Tez. Dokl. po Teorii Koletz, Algebr i Modulei. Mezhd. Konf. po Algebre Pamyati A.I. Mal'tzeva, Novosibirsk (1989), 140.

[Ufn90] Ufnarovski'i, V.A., *On using graphs for computing bases, growth functions and Hilbert series of associative algebras*, Mat. Sb. **180** (1990), no. 11, 1548–1550.

[VaZel89] Vais, A.Ja., and Zelmanov, E.I., *Kemer's theorem for finitely generated Jordan algebras*, Izv. Vyssh. Uchebn. Zved. Mat. (1989), no. 6, 42–51; translation: Soviet Math. (Iz. VUZ) **33** (1989), no. 6, 38–47.

[Zel91] Zelmanov, E.I., *The solution of the restricted Burnside problem for groups of prime power*, Mimeographed notes, Yale University (1991)

[ZelKos88] Zelmanov, E.I., *On nilpotence of nilalgebras*, Lect. Notes Math. **1352** (1988), 227–240.

[Zubk96] Zubkov, A.N., *On a generalization of the Razmyslov-Procesi theorem.* (Russian) Algebra i Logika **35** (1996), no. 4, 433–457, 498; translation in Algebra and Logic **35** (1996), no. 4, 241–254.

[Zubk00] Zubkov, A.N., *Modules with good filtration and invariant theory.* Algebra – representation theory (Constanta, 2000), 439–460, NATO Sci. Ser. II Math. Phys. Chem. **28** Kluwer Acad. Publ., Dordrecht, 2001.

[Zubr97] Zubrilin, K.A., *On the largest nilpotent ideal in algebras satisfying Capelli identities*, Sb. Math. **188** (1997), 1203–1211.

On Shirshov's Papers for Lie Algebras

Leonid Bokut

Shirshov published six papers on Lie algebras in which he found the following results (in order of publication, 1953–1962):

- Some years before Witt [84], the "Shirshov-Witt theorem" [1].
- Some years before Lazard [62], the "Lazard-Shirshov elimination process" [1]. This is often called "Lazard elimination"; see for example [79].
- The first example of a Lie ring that is not representable into any associative ring [2]; see also P. Cartier [37] and P.M. Cohn [43].
- In the same year as Chen-Fox-Lyndon [38], the "Lyndon-Shirshov basis" of a free Lie algebra (Lyndon-Shirshov Lie words) [6]. This is often called the "Lyndon basis"; see for example [63], [79], [64].
- Independently of Lyndon [65], the "Lyndon-Shirshov (associative) words" [6]. They are often called "Lyndon words"; see for example [63]. In the literature they are also often called "(Shirshov's) regular words" or "Lyndon-Shirshov words"; see for example [42], [24], [13], [32], [85], [76], [14].
- The algorithmic criterion to recognize Lie polynomials in a free associative algebra over any commutative ring [6]. The algorithm is based on the property that the maximal (in deg-lex ordering) associative word of any Lie polynomial is an associative Lyndon-Shirshov word. The Friedrichs criterion [45] follows from the Shirshov algorithmic criterion (see [6]).
- In the same year as Chen-Fox-Lyndon [38], the "central result on Lyndon-Shirshov words": any word is a unique non-decreasing product of Lyndon-Shirshov words [6]. This is often called the "Lyndon theorem" or the "Chen-Fox-Lyndon theorem".
- The reduction algorithm for Lie polynomials: the elimination of the maximal Lyndon-Shirshov Lie word of a Lie polynomial in a Lyndon-Shirshov Lie word [6]. The algorithm based on the Special Bracketing Lemma [6, Lemma 4], which in turn depends on the "central result on Lyndon-Shirshov words" above.
- The theorem that any Lie algebra of countable dimension is embeddable into two-generated Lie algebra with the same number of defining relations [6].

- Some years before Viennot [82], the "Hall-Shirshov bases" of a free Lie algebra [7]: a series of bases that contains the Hall basis and the Lyndon-Shirshov basis and depends on an ordering of basic Lie words such that $[w] = [[u][v]] > [v]$. They are often called "Hall sets"; see for example [79].
- Some years before Hironaka [53] and Buchberger [35], [36], the "Gröbner-Shirshov basis theory" for Lie polynomials (Lie algebras) explicitly and for noncommutative polynomials (associative algebras) implicitly [9]. This theory includes the definition of composition (s-polynomial), reduction algorithm, algorithm for producing a Gröbner-Shirshov basis (this is an infinite algorithm of Knuth-Bendix types [55]; see also the software implementations in [48], [87], [15]), and "Composition-Diamond Lemma". The Shirshov's "Composition-Diamond Lemma" for associative algebras was formulated explicitly in [25] and rediscovered by G. Bergman [16] under the name "Diamond Lemma for ring theory". The "Gröbner-Shirshov basis theory" for associative algebras was rediscovered by T. Mora [78] under the name "non-commutative Gröbner basis theory". The analogous theory for polynomials (commutative algebras) was found by B. Buchberger [35], [36] under the name "Gröbner basis theory"; similar ideas for (commutative) formal series were found by H. Hironaka [53] under the name "standard basis theory".
- The "Freiheitssatz" and the decidability of the word problem for one-relator Lie algebras [9].
- The first linear basis of the free product of Lie algebras [10].
- The first example showing that an analogue of the Kurosh subgroup theorem is not valid for subalgebras of the free product of Lie algebras [10].

Let us give some comments on these papers and further developments. See also V.K. Kharchenko's comments to some of these papers elsewhere in this volume.

In the paper [1], A.I. Shirshov, an aspirant (Ph.D. student) of A.G. Kurosh, proved that any subalgebra of a free Lie algebra is also free. This result was inspired by Kurosh's theorem [60] that any subalgebra of free non-associative algebra is also free. The former result was independently proved by E. Witt [84] three years later and is now called the Shirshov-Witt theorem. In this paper, Shirshov used the "K_d-lemma" to rewrite, in particular, a basic Lie word on a set $X = \{x_i : i = 1, 2, \ldots\}$ as a basic Lie word on the independent set $[x_i x_1^k] = [\ldots [x_i x_1] \ldots x_1]$ $(i > 1, k \geq 0)$; see Lemma 3 and Corollary 2 in [1]. This is often called the "Lazard elimination process" (Lazard [62]); see Theorem 0.6 of [79], cf. [34].

In the paper [2], Shirshov constructed the first example showing that the PBW theorem is not valid in general for Lie algebras over a commutative ring Σ (Σ-algebras). In this paper, Shirshov was able to construct a Lie Σ-algebra L with an element a in the center of L such that a belongs to the center of any Lie Σ-algebra extension of L. On the other hand, he gives a construction showing that the analogous extension result is not valid in general for associative Σ-algebras. Other counter-examples to the PBW theorem for Lie rings were constructed by P. Cartier [37] and P.M. Cohn [43].

In the paper [4], Shirshov proved that any subalgebra of a free commutative (anti-commutative) non-associative algebra is also free. He established linear bases of free (anti)commutative algebras, and later he used these bases for his "Gröbner-Shirshov basis theory" for (anti)commutative algebras, namely, for "Composition-Diamond Lemmas" for these algebras (see below [8]).

In the paper [5], Shirshov proves that any countably generated special Jordan (non-associative, (anti)commutative) algebra over a commutative ring can be embedded into a two-generated special Jordan (non-associative, (anti)commutative) algebra with the same number of defining relations. For groups, this is the famous Higman-Neumann-Neumann theorem [51]. A.I. Malcev [72] proved an analogous result for associative algebras. The analogous problem for Lie algebras was open until Shirshov's next paper [6].

Speaking about Shirshov's paper [6], I cannot help but cite P.M. Cohn's review (Zbl 0080.25503): "The author varies the usual construction of basic commutators in Lie rings by ordering words lexicographically and not by length [the "Lyndon-Shirshov basis", see also Chen-Fox-Lyndon [38]; in [42], P.M. Cohn credited this basis together with "Lyndon-Shirshov words" to Shirshov alone – L.B.]. This is used to give a very short proof of the theorem (Magnus, this Zbl. 16, 194 [see [69] – L.B.]; Witt, this Zbl 16, 244 [see [83] – L.B.]) that the Lie algebra obtained from a free associative algebra is free, with appropriate modification for the case of restricted Lie algebras. Secondly he derives the Friedrichs criterion (this Zbl. 52, 45 [see [45] – L.B.]) for Lie elements (see also P.M. Cohn [44] and R. Lyndon [66] – L.B.). As the third application he proves that every Lie algebra L can be embedded in a Lie algebra M such that in M any subalgebra of countable dimension is contained in a two-generated subalgebra. This is proved by showing that in the free associative algebra on two generators a, b (over a field), the elements

$$d_k = [[a, [a, b^k]], [a, b]], \quad k = 1, 2, \dots \quad ([x, y] = xy - yx),$$

form a distinguished set in the Lie algebra on two generators a, b (cf. Shirshov, this Zbl 71, 257 [see [5] – L.B.]).

Let us formulate the last statement, Lemma 10 of [6], explicitly. Let $k\langle a, b\rangle$ be the free associative algebra over a field k on two generators a, b, let $\mathrm{Lie}(a, b)$ be the Lie algebra of Lie polynomials of $k\langle a, b\rangle$ (the free Lie algebra on $\{a, b\}$), and let $L_\infty = \mathrm{Lie}(d_k : k = 1, 2, \dots)$ be the Lie subalgebra of $Lie(a, b)$, generated by $\{d_k : k = 1, 2, \dots\}$ above. By the K_d-lemma [1] (the Lazard-Shirshov elimination process), L_∞ is the free Lie algebra on the countable set $\{d_k : k \geq 1\}$. Let S be a subset of L_∞. Then

$$\mathrm{AssoId}_{k\langle a,b\rangle}(S) \cap L_\infty = \mathrm{LieId}_{L_\infty}(S),$$

where the former is the associative ideal (in $k\langle a, b\rangle$) generated by S, and the latter is the Lie ideal (in L_∞) generated by S. Shirshov also noticed that from the last statement the PBW theorem follows. In the proof of Lemma 10, Shirshov used the leading (maximal in the deg-lex order) associative monomials of Lie and associative

polynomials, and Lemma 4 on the "special bracketing" of a Lyndon-Shirshov word with a fixed Lyndon-Shirshov subword. The Special Bracketing Lemma is crucial: it allows him to define the reduction algorithm for Lie polynomials (he used this algorithm in the proof of Lemma 10), and to define later in [9] the notion of composition of two Lie polynomials (an analog of Buchberger's s-polynomial in Gröbner basis theory). By the way, in the proof of the Special Bracketing Lemma, he used the fact that any word c can be uniquely expressed as the product of a non-decreasing series of Lyndon-Shirshov words, $c = c_1 c_2 \dots c_k$ with $c_1 \le c_2 \le \dots \le c_k$ ($k \ge 0$). Actually, this remark is an important theorem often called the Lyndon theorem or the Chen-Fox-Lyndon theorem (see [38]). For example, this result is cited in the following way by Springer Online, Encyclopedia of Mathematics (edited by Michiel Hazewinkel): *Lyndon words – "The central result on Lyndon words is the following Chen-Fox-Lyndon theorem: any word can be expressed as a unique non-decreasing product of Lyndon words".*

All in all, Shirshov's paper [6] can be viewed, in particular, as an important step toward the Gröbner-Shirshov basis theory for associative and Lie algebras [9].

A.I. Shirshov [7] "varies the usual construction of basic commutators" [P. Hall [49] for groups and M. Hall [50] for Lie algebras – L.B.] in a free Lie algebra by ordering basic Lie words $\{[w]\}$ in any way such that $[w] > [v]$ if $[w] = [[u][v]]$. For example, an ordering based on the length (the Hall words), or an ordering based on lexicographical ordering (the Lyndon-Shirshov basis), both enjoy this property. He proves that any ordering of this kind leads to a linear basis of a free Lie algebra. Actually, this paper is a part of Shirshov's Thesis [3]. As mentioned above, Shirshov's series of bases were rediscovered later by Viennot [82] and are now often called "Hall bases" (see [79]). There is another example of "Hall-Shirshov bases", that give bases of free solvable Lie algebras ([18], see also [80] and [79], Ch. 5.3). In the paper [41], a first example of right normed basis of a free Lie algebra is found. Though it is not a Hall-Shirshov basis, it is closely connected to Lyndon–Shirshov words.

In the paper [8], Shirshov invented the "Gröbner-Shirshov basis theory" for (anti)commutative non-associative algebras based on the "Composition-Diamond Lemmas" for those algebras (see Lemma 2 in [8]). In particular, it implies the decidability of the word problem for any finitely presented (anti)commutative non-associative algebra. Also, the reduction algorithm is defined in order to find a "Gröbner-Shirshov basis" of any finitely generated ideal in a free (anti)commutative algebra. Shirshov also mentioned that the same results are valid for non-associative algebras. The decidability of the word problem for non-associative algebras was proved by A.I. Zhukov [86], another student of A.G. Kurosh. Actually, Zhukov invented a kind of "Gröbner-Shirshov basis theory" for non-associative algebras. The difference is that he did not use any linear ordering of non-associative words; for a non-associative polynomial f, he chose any non-associative word of maximal length from f as a "leading monomial" of f.

Shirshov's paper [9] is truly a pioneering paper in the subject. He starts with the definition of the composition of two Lie polynomials f, g (explicitly) and two

associative polynomials (implicitly) via the leading associative words $\bar{f}$, $\bar{g}$ of polynomials in the deg-lex ordering: Let $w = \bar{f}b = a\bar{g}$ for some associative words a, b such that $\bar{f} = ac$, $\bar{g} = cb$ and $c \neq 1$ (where 1 is the empty word). Then the associative composition $(f, g)_c$ (this is Shirshov's original notation; we now use $(f, g)_w$) is defined as follows: $(f, g)_c = fb - ag$. For Lie polynomials f, g, one needs to put extra Lie brackets on fb and ga. This is done according to the above mentioned Special Bracketing Lemma 4 [6]. This is a really important and crucial notion for the Gröbner-Shirshov basis theory for both Lie and associative algebras. Together with the above definition of reduction of one Lie polynomial modulo another (see the same paper [6]), it leads to an infinite algorithm to construct the Gröbner-Shirshov basis S^c starting with any set of Lie (associative) polynomials S. He proves Lemma 3, which is now called the Composition Lemma, or the Composition-Diamond Lemma, for Lie polynomials: if $f \in \text{Ideal}(S)$ then the leading associative word $\bar{f}$ contains as a subword $\bar{s}$ for some $s \in S^c$ (see also [32], [27]). Actually, he assumes the extra condition that S should be stable in some sense (see below), but he did not use the stability condition in the proof of the lemma (this condition is essential in order that S^c should be a recursive set for, say, finite S; he kips the stability condition having in mind the application of his theory to the word problem for Lie algebras). In [24], Shirshov's Composition Lemma for Lie polynomials was formulated in the modern form: Let S be a set of Lie polynomials that is closed under compositions (i.e., a Gröbner-Shirshov basis). If $f \in \text{Ideal}(S)$ then $\bar{f} = a\bar{s}b$ for some $s \in S$ and some associative words a, b. Closure means that any composition (i.e., composition of inclusion and composition of intersection) $(f, g)_w$ of polynomials f, g from S is trivial, i.e., it is zero after the reduction leading words of S. One may use a weaker form of the triviality that $(f, g)_w = \sum \alpha_i(a_i s_i b_i)$ for some $s_i \in S$, $\alpha_i \in k$ (the ground field) and some associative words a_i, b_i (with extra Lie bracketing), such that the leading associative words $a_i\bar{s}_i b_i$ of each expression are strictly less than w. The same Composition Lemma is valid for non-commutative associative polynomials with a much easier proof.

A.I. Shirshov gives three applications of his Composition Lemma for Lie algebras.

Theorem 1. *For any Lie polynomial f, there is no non-zero composition $(f, f)_w$. Then the reduction algorithm gives a solution of the word problem for any one-relator Lie algebra $Lie(X|f = 0)$.*

This is because any one-element set in a free Lie algebra is Gröbner-Shirshov basis. One may apply Shirshov's reduction algorithm for Lie polynomials. For groups, it is the famous result of W. Magnus [67]. S.I. Adjan [12] proved it for any semigroup with one defining relation of the form $u = 1$. V.N. Gerasimov [47] proved the decidability result for an associative one-relator algebra $k\langle X|f(X) = 0\rangle$ over a field k where the maximal homogeneous part $\widetilde{f}$ of $f(X)$ has no a proper two-sided divisor (from $\widetilde{f} = gh = h'g$ it follows $g \in k$).

In the paper [18], there is an application of the Shirshov's theorem: Any Lie algebra L is embeddable into an algebraically closed Lie algebra M (in the sense that any equation $f(x_1, \dots, x_n) = 0$ in the variables $X = \{x_1, x_2, \dots\}$ with coefficients in M has a solution in M; here f belongs to a free Lie product (see [10] below) of a free Lie algebra $\mathrm{Lie}(X)$ and M, $f \notin M$).

Theorem 2. *The word problem is decidable for any Lie algebra with a finite number of homogeneous defining relations.*

This is because any finite homogeneous set of Lie (associative) polynomials is a stable set in the sense of this paper. So, one may find all elements of S^c up to some fixed degree, and then apply Shirshov's reduction algorithm to the polynomial under consideration.

Theorem 3. (Freeness Theorem). *Let L be a Lie algebra with one defining relation $s = 0$. Then any subalgebra of L, generated by all but one letter involved in s, is the free Lie algebra on these free generators.*

For groups, this is the famous "Freiheitssatz" by W. Magnus [68]. The Freeness Theorem is also valid for an associative algebra with one defining relation (L.G. Makar-Limanov [71]). The proof does not use the Gröbner-Shirshov basis theory for associative algebras, but rather the existence of algebraically closed associative algebras (L.G. Makar-Limanov [70]). The Freeness theorem is proved for a pre-Lie (or right-symmetric) one-relator algebra (D. Kozybaev, L. Makar-Limanov, U. Umirbaev [56]).

Shirshov's paper [9] implicitly contains the Gröbner-Shirshov basis theory for associative algebras too, because he constantly used the fact that any Lie polynomial is at the same time a non-commutative polynomial. For example, the maximal term of a Lie polynomial is defined to be its maximal word as a non-commutative polynomial, the definition of the Lie composition (the Lie s-polynomial) of two Lie polynomials begins with their composition as non-commutative polynomials and then puts some special Lie brackets on it, and so on. The main Composition-Diamond Lemma for associative polynomials is actually proved in the paper: we need only to "forget" about the Lie brackets in the proof of this lemma for Lie polynomials (Lemma 3 [9]). The Composition-Diamond Lemma was explicitly formulated much later in papers L.A. Bokut [25] and G. Bergman [16].

We formulate Shirshov's Composition-Diamond Lemma for associative algebras following his paper [9] by "forgetting" the brackets, i.e., with only the change of "Lie polynomials" to "non-commutative polynomials". Let $k\langle X\rangle$ be the free associative algebra over a field k on a set X, such that the free monoid X^* is well-ordered by the deg-lex ordering. For a polynomial f, Shirshov [9] denotes by $\overline{f}$ the maximal word of f. Let f, g be two monic polynomials (possibly equal), let $w \in X^*$ be such that $w = acb$, where $\overline{f} = ac$, $\overline{g} = cb$ and a, b, c are words with c nonempty. Then $(f, g)_c = fb - ag$ is called an (associative) composition of f, g (this is Shirshov's original notation, now we use $(f, g)_w$); for Lie polynomials f, g, Shirshov puts some special brackets into $[fb] - [ag]$ such that $\overline{[fb] - [ag]} < w$.

Let S be a reduced set in $k\langle X\rangle$ and let S^c be a reduced set obtained from S by (transfinite) induction applying the following elementary operations: joining to S a composition of two elements of S and applying the reduction algorithm to the resulting set (until one gets a reduced set with only trivial compositions after the reduction). In current terminology, S^c is a Gröbner-Shirshov basis of the ideal generated by S, and the process of adding compositions is Shirshov's algorithm. He calls S a *stable* set if, at each step, the degree of the composition $(f,g)_w$, after the reduction, is bigger than the degree of f, g (or $(f,g)_w$ is zero after the reduction). Of course, if S is a finite (or recursive) stable set, then S^c is a recursive set and from the next lemma the word problem is solvable in the algebra with defining relations S. Now suppose that S is a Gröbner-Shirshov basis in the sense that S is a reduced set and any composition of intersection of elements of S is zero after the reduction (S is complete or closed under compositions). Hence S is a stable set in the sense of Shirshov. Then Lemma 3 of [9] has the following "forgetting brackets" form (see [25]).

Shirshov's Composition-Diamond Lemma for Associative Algebras. *Let $S \subset k\langle X\rangle$ be a Gröbner-Shirshov basis of the ideal* $\mathrm{Id}(S)$*. If $f \in \mathrm{Id}(S)$, then $\overline{f} = a\overline{s}b$, for some $s \in S$ and $a, b \in X^*$. Hence the set of S-irreducible words* $\mathrm{Irr}(S)$*, that do not contain maximal words of polynomials from S as subwords, is a k-basis of the algebra $k\langle X|S\rangle$.*

It is easy to see that the converse is also true (see [16]).

In the paper [10], Shirshov found a linear basis of a free product of Lie algebras with an amalgamation as an application of his Composition-Diamond Lemma for Lie algebras. Then he found an example proving that an analog of the Kurosh subgroup theorem [61] for a free product of groups, as well as Kurosh's [60] and Gainov's [46] theorems for subalgebras of free products of non-associative or (anti)commutative non-associative algebras, are not valid for subalgebras of free products of Lie algebras. Kukin [59], [58] found a description of subalgebras of free (amalgamated) products of Lie algebras.

In the paper [24], Shirshov's Composition-Diamond Lemma was systematically used in order to prove the following embedding theorem: Let M be any recursively enumerable set of natural numbers. Let

$$L_M = \mathrm{Lie}(a, b, c, a_1, b_1, c_1 \mid [ab^k c] = [a_1 b_1^k c_1], k \in M)$$

be a recursively presented Lie algebra. Then L_M is embeddable into a finitely presented Lie algebra L. If M is not recursive, then the word problem is undecidable in L_M and hence in L.

This gave the negative solution of the word problem for Lie algebras. An explicit example of a finitely presented Lie algebra with undecidable word problem was given by Kukin [57]. The proof in [24] used Matiyasevich's solution of Hilbert's 10th problem [73] and some ideas of the Higman theorem [52] that any recursively presented group is embeddable into a finitely presented group. There remained the problem of whether any recursively presented Lie (associative) algebra can be

embedded into a finitely presented Lie (associative) algebra. V. Belyaev [17] solved positively the problem for associative algebras.

In the paper [25], Shirshov's Composition-Diamond Lemma for associative algebras was used to prove an embedding theorem for associative algebras: For any associative algebras A, A_i $(i = 1, 2, 3, 4)$ with appropriate cardinality conditions, for example, all of them are algebras of countable dimension and A_i $(i = 1, 2, 3, 4)$ is the union of a countable increasing series of subalgebras with factors of countable dimension. Then A can be embedded into a simple associative algebra which is a sum of A_i $(i = 1, 2, 3, 4)$; in particular, A can be embedded into a finitely generated simple associative algebra. By the way, answering a question raised by [25], Shelah [81] constructed an example of an associative algebra of uncountable dimension which is not a union of a countable increasing series of subalgebras with factors of uncountable dimension.

In the paper [26], Shirshov's Composition-Diamond Lemma for Lie algebras was used to prove an embedding theorem for Lie algebras: Any Lie algebra is embeddable into an algebraically closed (in particular simple) Lie algebra which is a sum of four prescribed Lie subalgebras with the same cardinality conditions as in [25] above.

In the papers [20], [21], [23], there were found normal forms of elements of Novikov's and Boone's groups, as well as relative normal forms of some groups of quotients of multiplicative semigroups of some rings. Actually, those normal forms are the (relative) irreducible words for (relative) Gröbner-Shirshov bases of the groups, see [33], [39].

In the papers [74], [75], there were proved Composition-Diamond Lemmas for colore Lie superalgebras, Lie p-algebras and Lie p-superalgebras.

In the papers [54], [40], there were proved Composition-Diamond Lemmas for modules.

In the paper [28], it was proved Composition-Diamond Lemma for associative conformal algebras.

Some other papers on Gröbner–Shirshov bases one may found in surveys [29], [30], [31].

References

[1] Shirshov, A.I., *Subalgebras of free Lie algebras*. (Russian) Mat. Sb., N. Ser. **33(75)**, 441–452 (1953).

[2] Shirshov, A.I. *On representation of Lie rings in associative rings.* (Russian) Usp. Mat. Nauk **8**, No.5 (57), 173–175 (1953).

[3] Shirshov, A.I. *Certain problems of the theory of non-associative algebras.* Thesis, Moscow State University, 1953.

[4] Shirshov, A.I. *Subalgebras of free commutative and free anti-commutative algebras. Mat. Sbornik.*, **34(76)** (1954), 81–88.

[5] Shirshov, A.I., *Some theorems on embedding for rings*. (Russian) Mat. Sb., N. Ser. **40(82)**, 65–72 (1956).

[6] Shirshov, A.I., *On free Lie rings.* (Russian) Mat. Sb., N. Ser. **45(87)**, 113–122 (1958).

[7] Shirshov, A.I., *On bases of a free Lie algebra.* (Russian) Algebra Logika **1**, No.1, 14–19 (1962).

[8] Shirshov, A.I., *Certain algorithmic problems for ϵ-algebras.* (Russian) Sib. Mat. Zh. **3**, 132–137 (1962).

[9] Shirshov, A.I., *Certain algorithmic problems for Lie algebras.* (Russian) Sib. Mat. Zh. **3**, 292–296 (1962). English translation: Shirshov, A.I., *Certain algorithmic problems for Lie algebras.* (English) ACM SIGSAM Bull. **33**, No. 2, 3–6 (1999).

[10] Shirshov, A.I., *On the conjecture of the theory of Lie algebras.* (Russian) Sib. Mat. Zh. **3**, 297–301 (1962).

[11] A.I. Shirshov, *Collected Works. Rings and Algebras*. Nauka, Moscow, 1984.

[12] Adjan, S.I. Defining relations and algorithmic problems for groups and semigroups. (English. Russian original) Proc. Steklov Inst. Math. **85**, 152 p. (1966); translation from Tr. Mat. Inst. Steklov **85**, 123 p. (1966).

[13] Yu.A. Bahturin, A.A. Mikhalev, M.V. Zaicev, and V.M. Petrogradsky, *Infinite Dimensional Lie Superalgebras*. Walter de Gruyter Publ., Berlin, New York, 1992.

[14] Bahturin, Yuri; Mikhalev, Alexander A.; Zaicev, Mikhail Infinite-dimensional Lie superalgebras. (English) Hazewinkel, M. (ed.), Handbook of algebra. Volume 2. Amsterdam: North-Holland. 579–614 (2000).

[15] Backelin, Jörgen; Cojocaru, Svetlana; Ufnarovski, Victor *The computer algebra package Bergman: Current state.* (English) Herzog, Jürgen (ed.) et al., commutative algebra, singularities and computer algebra. Proceedings of the NATO advanced research workshop, Sinaia, Romania, September 17–22, 2002. Dordrecht: Kluwer Academic Publishers. NATO Sci. Ser. II, Math. Phys. Chem. 115, 75–100 (2003).

[16] G.M. Bergman, *The Diamond Lemma for ring theory. Adv. in Math.*, **29**(1978), 178–218.

[17] Belyaev, V. Ya. *Subrings of finitely presented associative rings.* (English) Algebra Logika **17**, 627–638 (1978).

[18] Bokut, L.A., *Embedding of Lie algebras into algebraically closed Lie algebras.* (Russian) Algebra Logika **1**, No.2, 47–53 (1962).

[19] Bokut, L.A., *Bases of free poly-nilpotent Lie algebras.* (Russian) Algebra Logika **2**, No.4, 13–19 (1963).

[20] L.A. Bokut, *On a property of the Boone groups.* Algebra i Logika Sem., **5** (1966), 5, 5–23; **6** (1967), 1, 15–24.

[21] L.A. Bokut, *On Novikov's groups.* Algebra i Logika Sem., **6** (1967), 1, 25–38.

[22] Bokut, L.A., *Degrees of insolvability of the conjugacy problem for finitely presented groups.* (Russian) Algebra Logika **7**, No.5, 4–70; No.6, 4–52 (1968).

[23] L.A. Bokut, *Groups of fractions of multiplication semigroups of certain rings. I–III, Malcev's problem.* Sibir. Math.J., **10**, 2, 246–286; 4, 744–799; 4, 800–819; 5, 965–1005.

[24] L.A. Bokut, *Unsolvability of the word problem, and subalgebras of finitely presented Lie algebras.* Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 1173–1219.

[25] L.A. Bokut, *Imbeddings into simple associative algebras.* Algebra i Logika Sem., **15** (1976), 117–142.

[26] Bokut', L.A., *On algebraically closed and simple Lie algebras.* (Russian, English) Proc. Steklov Inst. Math. **148**, 30–42 (1978).

[27] L.A. Bokut, Yuqun Chen, *Gröbner-Shirshov bases for Lie algebras: after A.I. Shirshov. SEA Bull Math.,* **31** (2007), 811–831.

[28] L.A. Bokut, Y. Fong, W.- F. Ke, *Composition Diamond Lemma for associative conformal algebras. J. Algebra*, **272**(2004), 739–774.

[29] L.A. Bokut, Y. Fong, W.-F. Ke, P.S. Kolesnikov, *Gröbner and Gröbner-Shirshov bases in Algebra and Conformal algebras. Fundamental and Applied Mathematics*, **6**(2000), N3, 669–706 (in Russian).

[30] L.A. Bokut, P.S. Kolesnikov, *Gröbner-Shirshov bases: From Incipient to Nowdays*, *Proceedings of the POMI*, **272**(2000), 26–67.

[31] L.A. Bokut, P.S. Kolesnikov, *Gröbner-Shirshov bases: conformal algebras and pseudoalgebras*, *Journal of Mathematicfal Sciences*, **131**(5)(2005), 5962–6003.

[32] L.A. Bokut, and G.P. Kukin, Algorithmic and combinatorial algebra. Mathematics and its Applications, 255. Kluwer Academic Publishers Group, Dordrecht, 1994.

[33] L.A. Bokut, K.P. Shum, *Relative Gröbner-Shirshov bases for algebras and groups.* Algebra i Analiz **19** (2007), no. 6, 1–21.

[34] Bourbaki, N. Elements de mathematique. Fasc. XXXVII: Groupes et algèbres de Lie. Chap. II: Algèbres de Lie libres. Chap. III: Groupes de Lie. (French) Actualites scientifiques et industrielles 1349. Paris: Hermann. 320 p. (1972)

[35] B. Buchberger, An algorithm for finding a basis for the residue class ring of a zero-dimensional polynomial ideal. (German). Ph.D. thesis, University of Innsbruck, Austria, 1965.

[36] B. Buchberger, *An algorithmical criteria for the solvability of algebraic systems of equations.* (German). Aequationes Math. **4** (1970), 374–383.

[37] Cartier, P. *Remarques sur le theore me de Birkhoff-Witt.* (French) Ann. Sc. Norm. Super. Pisa, Sci. Fis. Mat., III. Ser. **12**, 1–4 (1958).

[38] K.T. Chen, R.H. Fox, and R.C. Lyndon, *Free differential calculus, IV: the quotient groups of the lower central series. Annals of Mathematics* **68** (1958), pp. 81–95.

[39] Yuqun Chen, Wenshu Chen and Runai Luo, *Word problem for Novikov's and Boone's group via Gröbner-Shirshov bases.*, SEA Bull Math., **32**(2008), 5.

[40] E.S. Chibrikov, *On free Lie conformal algebras.* Vestnik Novosib. State Univ., Ser. "Math, Mech, Inform.", **4** (2004), No 1, 65–83 (in Russian).

[41] Chibrikov, E.S. *A right normed basis for free Lie algebras and Lyndon–Shirshov words.* J. Algebra **302**, No. 2, 593–612 (2006).

[42] Cohn, P.M. Universal algebra. (English) Harper's Series in Modern Mathematics. New York-Evanston-London: Harper and Row, Publishers 1965, XV, 333 p. (1965).

[43] Cohn, P.M. *A remark on the Birkhoff-Witt theorem.* (English) J. Lond. Math. Soc. **38**, 197–203 (1963).

[44] Cohn, P.M. *Sur le critère de Friedrichs pour les commutateurs dans une algèbre asociative libre.* Comptes Rendus Acad. Science Paris, **239**, 743–745 (1954).

[45] Friedrichs, K.O. *Mathematical aspects of the quantum theory of fields. V.* (English) Commun. Pure Appl. Math. **6**, 1–72 (1953).

[46] Gainov, A.T. *Free commutative and free anticommutative products of algebras.* (Russian) Sib. Mat. Zh. **3**, 805–833 (1962).

[47] Gerasimov, V.N. *Distributive lattices of subspaces and the equality problem for algebras with a single relation.* Algebra Logic **15** (1976), 238–274 (1977); translation from Algebra Logika **15**, 384–435 (1976).

[48] Gerdt, V.P.; Kornyak, V.V. *Program for constructing a complete system of relations, basis elements, and commutator table for finitely presented Lie algebras and superalgebras.* (English. Russian original) Program. Comput. Softw. **23**, No. 3, 164–172 (1997); translation from Programmirovanie 1997, No.3, 58–71 (1997).

[49] P. Hall, *A contribution to the theory of groups of prime power order. Proc. London Math. Soc. Ser.* 2, **36** (1933), pp. 29–95.

[50] M. Hall, *A basis for free Lie rings and higher commutators in free groups. Proc. Amer. Math. Soc.* **3**(1950), pp. 575–581.

[51] G. Higman, B.H. Neumann, H. Neumann, *Embedding theorems for groups. J. London Math. Soc.* **24** (1949) 247–254.

[52] Higman, G. *Subgroups of finitely presented groups.* (English) Proc. R. Soc. Lond., Ser. A **262**, 455–475 (1961).

[53] H. Hironaka, *Resolution of singularities of an algebraic variety over a field of characteristic zero, I, II. Ann. Math.*, **79**(2) (1964), pp. 109–203, 205–326.

[54] S.-J. Kang, K.-H. Lee, Gröbner–Shirshov bases for irreducible sl_{n+1}-modules, *Journal of Algebra*, **232** (2000), 1–20.

[55] Knuth, D.E.; Bendix, P.B. *Simple word problems in universal algebras.* Comput. Probl. abstract Algebra, Proc. Conf. Oxford 1967, 263–297 (1970).

[56] D. Kozybaev, L. Makar-Limanov, U. Umirbaev, *The Freiheitssatz and the automorphisms of free right-symmetric algebras*, *Asian-European J. Math.* **1** (2008), 2, 243–252.

[57] G.P. Kukin, *On the word problem for Lie algebras.* Sibirsk. Math. Zh. **18** (1977), 1194–1197.

[58] Kukin, G.P. *Subalgebras of a free Lie sum of Lie algebras with an amalgamated subalgebra.* Algebra Logic **11**(1972), 59–86.

[59] Kukin, G.P. *On the Cartesian subalgebras of a free Lie sum of Lie algebras.* Algebra Logika **9**, 701–713 (1970).

[60] Kurosh, A., *Nonassociative free algebras and free products of algebras.* (Russian. English summary) Mat. Sb., N. Ser. **20(62)**, 239–262 (1947).

[61] Kurosch, A. *Die Untergruppen der freien Produkte von beliebigen Gruppen.* (German) Math. Ann. **109**, 647–660 (1934)

[62] Lazard, M. Groupes, anneaux de Lie et problème de Burnside. C.I.M.E., Gruppi, Anelli di Lie e Teoria della Coomologia 60 p. (1960). The same in: Instituto Matematico dell'Universita di Roma (1960).

[63] Lothaire, M. Combinatorics on words. Foreword by Roger Lyndon. Encyclopedia of Mathematics and Its Applications, Vol. 17. Reading, Massachusetts, etc.: Addison-Wesley Publishing Company, Advanced Book Program/World Science Division. XIX, 238 p. (1983).

[64] Lothaire, M. Combinatorics on words. Foreword by Roger Lyndon. 2nd ed. Encyclopedia of Mathematics and Its Applications. 17. Cambridge: Cambridge University Press. xvii, 238 p.(1997).

[65] Lyndon, R.C. *On Burnside's problem.* Trans. Am. Math. Soc. **77**, 202–215 (1954).

[66] Lyndon, R.C. *A theorem of Friedrichs.* Mich. Math. J. **3**, 27–29 (1956).

[67] Magnus, W. *Über diskontinuierliche Gruppen mit einer definierenden Relation* (*Der Freiheitssatz*). *J. Reine Angew. Math* **163***(1930), pp. 141–165.*

[68] Magnus, *W. Das Identitätsproblem für Gruppen mit einer definierenden Relation.* (German) Math. Ann. **106**, 295–307 (1932).,

[69] W. Magnus, *Über Beziehungen zwischen höheren Kommutatoren. J. Reine Angew. Math* **177***(1937), pp. 105–115.*

[70] Makar-Limanov, L. *Algebraically closed skew fields.* J. Algebra 93, 117–135 (1985).

[71] Makar-Limanov, L.G. *On algebras with one relation.* Usp. Mat. Nauk **30**, No.2(182), 217 (1975).

[72] Malcev, A.I.,*On representation of nonassociative rings.* Uspehi Mat. Nauk N.S. **7** (1952), 181–185.

[73] Matiyasevich, Yu.V. *Enumerable sets are diophantine.* Russian original) Sov. Math., Dokl. 11, 354–358 (1970); translation from Dokl. Akad. Nauk SSSR 191, 279–282 (1970).

[74] Mikhalev, A.A., *The junction lemma and the equality problem for color Lie superalgebras.* Vestnik Moskov. Univ. Ser. 1. Mat. Mekh. 1989, no. 5, 88–91. English translation: Moscow Univ. Math. Bull. 44 (1989), 87–90.

[75] A.A. Mikhalev, *Shirshov's composition techniques in Lie superalgebras* (*non-commutative Gröbner bases*). Trudy Sem. Petrovsk. **18** (1995), 277–289.

[76] A.A. Mikhalev and A.A. Zolotykh, *Combinatorial Aspects of Lie Superalgebras*. CRC Press, Boca Raton, New York, 1995.

[77] V.N. Latyshev, *Combinatorial Theory of Rings. Standard Bases*. Moscow State Univ. Publ. House, Moscow, 1988.

[78] T. Mora, *Gröbner bases for non-commutative polynomial rings.* Lecture Notes in Comput. Sci. **229** (1986), 353–362.

[79] C. Reutenauer. Free Lie algebras. London Mathematical Society Monographs. New Series, 7. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1993.

[80] C. Reutenauer. *Dimensions and characters of the derived series of the free Lie algebra.* In M. Lothaire, Mots, Melanges offerts a M.-P. Schützenberger, pp. 171–184. Hermes, Paris.

[81] Shelah, Saharon *On a problem of Kurosh, Jonsson groups, and applications.* (English) Word problems II, Stud. Logic Found. Math. Vol. 95, 373–394 (1980).

[82] Viennot, Gerard. Algèbres de Lie libres et monoides libres. Bases des algèbres de Lie libres et factorisations des monoides libres. (French) Lecture Notes in Mathematics. 691. Berlin-Heidelberg-New York: Springer-Verlag. 124 p. (1978)

[83] E. Witt, *Treue Darstellungen Lieschen Ringe. J. Reine Angew. Math.* **177**(1937), pp. 152–160.

[84] E. Witt, *Subrings of free Lie rings Math. Zeit.*, 64(1956), 195–216.

[85] V.A. Ufnarovski, *Combinatorial and Asymptotic Methods in Algebra*. Encyclopaedia Math. Sci. **57** (1995), 1–196.

[86] A.I. Zhukov, *Reduced systems of defining relations in non-associative algebras Mat. Sb., N. Ser.*, 27(69) (1950), 267–280.

[87] Zolotykh, A.A.; Mikhalev, A.A. *Algorithms for construction of standard Gröbner-Shirshov bases of ideals of free algebras over commutative rings.* (English. Russian original) Program. Comput. Softw. 24, No. 6, 271–272 (1998); translation from Programmirovanie 1998, No.6, 10–11 (1998).

Some of A.I. Shirshov Works

V.K. Kharchenko

In his first published paper "Subalgebras of free Lie algebras" A.I. Shirshov proved for Lie algebras an analog of the famous Nielsen-Schreier theorem: every subalgebra of a free Lie algebra is free. Three years later this theorem was independently proved and extended to restricted Lie algebras by E. Witt [38]. Much later this result was generalized to Lie superalgebras (A.S. Shtern [29]), and to colored Lie superalgebras (A.A. Mikhalev [20, 21, 22]). These results went through further development in the field of quantum algebra as follows. The Shirshov–Witt theorem for Lie algebras over fields of characteristic zero admits an equivalent formulation in terms of a free associative algebra: Every Hopf subalgebra of a free algebra $\mathbf{k}\langle y_i\rangle$ with the coproduct set up by $\Delta(y_i) = y_i \otimes 1 + 1 \otimes y_i$ is free. If we consider the free algebra as a braided Hopf algebra with a very special braiding ($\tau(y_i \otimes y_j) = p_{ij}y_j \otimes y_i,\ p_{ij}p_{ji} = 1$), then we get a reformulation of the Mikhalev-Shtern generalization as well. We may consider the free associative algebra $\mathbf{k}\langle V\rangle$ as a braided Hopf algebra provided that V is a braided space with arbitrary braiding (not necessary invertible). In this setting the braided version of the Shirshov-Witt theorem takes up the following form [12]. If a subalgebra $U \subseteq \mathbf{k}\langle V\rangle$ is a right categorical right coideal, that is $\Delta(U) \subseteq U\underline{\otimes}\mathbf{k}\langle V\rangle,\ \tau(\mathbf{k}\langle V\rangle \otimes U) \subseteq U\otimes\mathbf{k}\langle V\rangle$, then U is a free subalgebra.

A detailed investigation of free generators for subalgebras of a free Lie algebra and their ranks can be found in the papers [15, 23, 25]. An analogue of Schreier's formula was found by V. Petrogradsky for free Lie (super)algebras in terms of formal power series [27, 28]. Description of automorphism groups of free Lie algebras is closely related to the theorem of A.I. Shirshov on subalgebras. In 1964 P. Cohn [4] proved that the automorphisms of free Lie algebras of a finite rank are tame, i.e., the automorphism group is generated by elementary automorpisms. Defining relations of the automorphism group were described in 2007 by U.U. Umirbaev [36]. A detailed investigation of automorphisms of free Lie algebras and their applications can be found in the papers [15, 26, 23, 34, 35, 24, 25, 27]. The Shirshov-Witt theorem on subalgebras gives also the decidability of the occurrence problem for free Lie algebras (see, also [15]). In 1990 U.U. Umirbaev proved

[32] that finitely generated subalgebras of free Lie algebras are finitely separable. The occurrence problem for free Lie algebras and for relatively free algebras was studied in [33, 6, 7].

In a small note "On representation of Lie rings in associative rings" A.I. Shirshov constructed an example of a Lie ring that has no faithful representations in associative rings. This example shows that the Poincare-Birkhoff-Witt theorem may not be extended to Lie algebras over arbitrary commutative rings. Recall that the original proof of the PBW-theorem for Lie algebras over fields remains valid for Lie algebras over commutative rings, provided that the algebra is a free module over the ring of scalars [1, 37]. However it is not evident if a free Lie algebra over any commutative ring indeed is a free module over the ring of scalars. A.I. Shirshov in his fundamental paper "On free Lie rings" showed in particular that this question has an affirmative answer. Independently M. Lazard [17] and P. Cartier [2] proved that every Lie algebra over a Dedekind domain has a representation in an associative ring. If the ring of scalars itself is an algebra over the rationals, the representation exists as well (P. Cohn [3]). Later H.-J. Higgins in [8] found necessary and sufficient conditions for the module structure for a Lie algebra over a commutative ring to have a representation in an associative ring. It should be emphasized that the embedding problems are the most subtle problems located at the interfaces between algebra and logic. Sometimes in this area deep and extensive investigations trace back to publications with serious gaps and errors, see for example the historical notes [30]. Even in our time there appear such publications concerning the representation of generalizations of Lie algebras in associative rings in serious mathematical journals (see a discussion in [12, Section 5]).

In the paper "On free Lie rings" in order to construct a basis of a free Lie algebra (over a commutative ring) A.I. Shirshov introduced a class of words that is fundamental for modern combinatorial theory. This class of words was independently discovered by R. Lyndon several years before [19]. Now these words are called *Lyndon words* or *Lyndon-Shirshov words*, see M. Lothaire [18].

The method of Lyndon-Shirshov words remains a very effective tool for modern investigations in algebra. This allows one to construct a PBW-basis in arbitrary Hopf algebra generated by skew-primitive semi-invariants, [11], or in a braided Hopf algebra with a so-called triangular set of primitive generators [31]. In a more general setting, [5], it is possible to find some kind of factorization of graded Hopf algebras using Lyndon-Shirshov words. An interesting development is due to P. Lalonde and A. Ram. They found an elegant representation of the Lyndon-Shirshov basis for classical finite-dimensional simple Lie algebras, see [16, Figure 1]. More recently the method of Lyndon-Shirshov words has proved to be an extremely important tool for classification of right coideal subalgebras in quantum groups [13, 14].

One more result from the paper "On free Lie rings" that has a reflection to contemporaneity, the Freiderich criterion (Theorem 3 in that paper), shows that the elements of a given free Lie algebra (over a commutative ring Σ) can be distinguished in the enveloping free associative algebra (over Σ) as primitive

elements with respect to the diagonal coproduct. Even though A.I. Shirshov did not introduce the very same coproduct, in Theorem 3 one may replace the commuting variables a_i, a_i' with $a_i \otimes 1$ and $1 \otimes a_i$ respectively. Then the condition

$$f(a_1 + a_1', \ldots, a_n + a_n') = f(a_1, \ldots, a_n) + f(a_1', \ldots, a_n')$$

reduces to $\Delta(f) = f \otimes 1 + 1 \otimes f$, where Δ is the diagonal coproduct defined on the generators via $\Delta(a_i) = a_i \otimes 1 + 1 \otimes a_i$ and extended to the enveloping free algebra as an algebra homomorphism $\Delta : \mathfrak{A}_{\Sigma R} \to \mathfrak{A}_{\Sigma R} \otimes \mathfrak{A}_{\Sigma R}$.

In the paper "On rings with identity relations" A.I. Shirshov in particular proves that every associative PI-ring algebraic over a central subring Z_1 is in some sense finite over Z_1 (Theorem 4 in the paper). This new notion of finiteness is close to but not identical with the notion of finitely generated module. It is interesting that essentially the same notion over a not necessarily central subring appears in the modern noncommutative Galois theory. More precisely a subring $A \subseteq R$ is called (right) *Shirshov finite* over a subring $D \subseteq R$ if there exists a finite number of elements $r_1, r_2, \ldots, r_k$ such that $A \subseteq r_1 D + r_2 D + \cdots + r_k D$. The Shirshov theorem (Theorem 4 in the paper) says that R^n is Shirshov finite over Z_1, where n is the degree of PI-identity of the finitely generated ring R. The same finiteness relation in local form exists between a given semiprime ring R and its Galois subring R^G with respect to a finite group G of automorphisms, [9, 10, Theorem 5.10.1]. In more detail, suppose that semiprime associative ring R has no additive $|G|$-torsion (or more generally G is a Maschke group, [9, 10, Definition 5.4.13]). Then R has an essential two-sided ideal I that is locally finite in the Shirshov sense over the fixed ring $R^G = \{r \in R \mid \forall g \in G,\ g(r) = r\}$. Here the *local finiteness* means that each finitely generated right ideal $A \subseteq I$ is Shirshov finite over R^G as a subring.

References

[1] G. Birkhoff, Representability of Lie algebras and Lie groups by matrices, Annals Math., v.38(1937), 526–532.

[2] P. Cartier, Remarques sur le théorème de Birkhoff–Witt, Ann. Scuola norm sup. Pisa, Sci. fis. mat. v.3, Ser. 12(1958), 1–4.

[3] P.M. Cohn, A remark on the Birkhoff–Witt theorem, J. London Math. Soc., v. 38(1963), 197–203.

[4] P.M. Cohn, Subalgebras of free associative algebras, Proc. London Math. Soc. (3) 14, (1964), 618–632.

[5] M. Graña, I. Heckenberger, On a factorization of graded Hopf algebras using Lyndon words, Journal of Algebra, v. 314, N1(2007), 324–343.

[6] C.K. Gupta and U.U. Umirbaev, Systems of linear equations over associative algebras and the occurence problem for Lie algebras, Commun. Algebra, 27(1999), 411–427.

[7] C.K. Gupta and U.U. Umirbaev, The occurence problem for free metanilpotent Lie algebras, Commun. Algebra, 27(1999), 5857–5876.

[8] H.-J. Higgins, Baer invariants and the Birkhoff–Witt theorem, Journal of Algebra, v. 11(1969), 469–482.

[9] V.K. Kharchenko, Automorphisms and Derivations of Associative Rings, Kluwer Academic Publishers, Dordrecht-Boston-London, 1991.

[10] V.K. Kharchenko, Noncommutative Galois Theory, Nauchnaja Kniga, Novosibirsk, 1996.

[11] V.K. Kharchenko, A quantum analog of the Poincaré-Birkhoff-Witt theorem, Algebra i Logika, 38, N4(1999), 476–507. English translation: Algebra and Logic, 38, N4(1999), 259–276 (QA/0005101).

[12] V.K. Kharchenko, Braided version of Shirshov–Witt theorem, Journal of Algebra, 294, N1(2005), 196–225.

[13] V.K. Kharchenko, PBW-bases of coideal subalgebras and a freeness theorem, Transactions of the American Mmathematical Society, v. 360, N10(2008), 5121–5143.

[14] V.K. Kharchenko, A.V. Lara Sagahon, Right coideal subalgebras in $U_q(\mathfrak{sl}_{n+1})$, Journal of Algebra, v. 319 (2008), 2571–2625.

[15] G.P. Kukin, Primitive elements of free algebras, Algebra i Logika, v. 9, N4(1970), 458–472. English translation: Algebra and Logic, v. 9 (1970), 275–284.

[16] P. Lalonde, A. Ram, Standard Lyndon bases of Lie algebras and enveloping algebras, Transactions of the American Mathematical Society, v. 347, N5(1995), 1821–1830.

[17] M. Lazard, Sur les algèbres enveloppantes universelles de certaines algèbres de Lie, Publ. Sci. Univ. Alger, Sér. A. v. 1(1954), 281–294.

[18] M. Lothaire, Algebraic Combinatorics on Words,Cambridge Univ. Press, 2002.

[19] Lyndon, R.C. On Burnside's problem, Trans. Am. Math. Soc., 77(1954), 202–215.

[20] A.A. Mikhalev, Subalgebras of free color Lie superalgebras, Mat. Zametki 37 N5 (1985) 653–661. English translation: Math. Notes 37 (1985) 356–360.

[21] A.A. Mikhalev, Free color Lie superalgebras, Dokl. Akad. Nauk SSSR, 286, N3 (1986) 551–554. English translation: Soviet Math. Dokl. 33 (1986) 136–139.

[22] A.A. Mikhalev, Subalgebras of free Lie p-superalgebras, Mat. Zametki 43 N2 (1988) 178–191. English translation: Math. Notes 43 (1988) 99–106.

[23] A.A. Mikhalev, Primitive elements and automorphisms of free algebras of Schreier varieties, J. Math. Sci., 102, N6(2000), 4628–4640.

[24] A.A. Mikhalev, U.U. Umirbaev, J.-T. Yu, Automorphic orbits of elements of free nonassociative algebras, Journal of Algebra, 243(2001), 198–223.

[25] A.A. Mikhalev, U. Umirbaev, Jie-Tai Yu, Generic, almost primitive and test elements of free Lie algebras, Proc. Amer. Math. Soc., 130(2002), 1303–1310.

[26] A.A. Mikhalev, A.A. Zolotykh, Rank and primitivity of elements of free color Lie (p-)superalgebras, Intern. J. Algebra and Computation, 4(1994), 617–656.

[27] V.M. Petrogradsky, Schreier's formulae for free Lie algebras, their Applications and Asymptotics, Proceedings of International Algebraic Conference on 90th Birthday of A.G. Kurosh, Moscow, 1998, Ed. by Y. Bahturin, and de Gruyter, Berlin, 2000.

[28] V.M. Petrogradsky, Schreier's formula for free Lie algebras. Arch. Math. (Basel), 75(2000), no. 1, 16–28.

[29] A.S. Shtern, Free Lie superalgebras, Siberian Math. J. 27 (1986) 551–554.

[30] W. Schmid, Poincare and Lie groups, Bull. (N.S.) Amer. Math. Soc. v.6(1982), 175–186.

[31] S. Ufer, PBW bases for a class of braided Hopf algebras, Journal of Algebra, 280, N1(2004), 84–119.

[32] U.U. Umirbaev, On the approximation of free Lie algebras with respect to entry, Monoids, rings and algebras, Tartu: Tartuskij Universitet, Tartu Uelik. Toim., Mat.-Meh.-Alaseid Toeid, 878(1990), 147–152.

[33] U.U. Umirbaev, The occurrence problem for Lie algebras, Algebra Logic, 32 (1993), No. 3, 173–181 ; translation from Algebra Logika, 32(1993), No. 3, 326–340.

[34] U.U. Umirbaev, Partial derivations and endomorphisms of some relatively free Lie algebras, Sib. Math. J., 34(1993), No. 6, 1161–1170; translation from Sib. Mat. Zh. 34(1993), No. 6, 179–188.

[35] U.U. Umirbaev, On Schreier varieties of algebras, Algebra Logic, 33(1994), No. 3, 180–193; translation from Algebra Logika, 33 (1994), No. 3, 317–340 .

[36] U.U. Umirbaev, Definig relations for automorphism groups of free algebras, J. Algebra, 314 (2007), 209–225.

[37] E. Witt, Treue Darstellung Liescher Ringe, J. reine angew. Math. v. 177(1937), 152–160.

[38] E. Witt, Die Unterringe der freien Lieschen Ringe, Math. Zeitschr. Bd. 64 (1956) 195–216.

Comments on Shirshov's Height Theorem

Alexander Kemer

In 1941 A.G. Kurosh [1] posed the problem: Is every finitely-generated algebraic associative algebra finite-dimensional? In 1964 E.S. Golod and I.R. Shafarevich [2, 3] constructed a counterexample: they presented an infinite-dimensional finitely-generated nil-algebra. This counterexample shows that in general finitely-generated algebraic associative algebras are very far from being finite-dimensional.

Every problem can be considered not only as an explicit problem but as a direction of research. In the case of Kurosh's problem such a direction can be formulated in the following way: Find the conditions which imply that a finitely generated algebra is finite-dimensional.

Before the counterexample of Golod-Shafarevich was constructed, many positive results on Kurosh's problem were obtained. In 1945 N. Jacobson [4] solved the problem of Kurosh for algebraic algebras of bounded index. In 1946 J. Levitzky [5] proved that for a finitely generated PI-algebra over a commutative ring, if each element is nilpotent then the algebra is nilpotent. Finally, in 1948 I. Kaplansky [6] solved Kurosh's problem for PI-algebras over a field. All of these results became classical and are included in textbooks on ring theory. The great role of these results in ring theory is well known. In fact, the structure theory of rings developed around the problem of A.G. Kurosh.

In 1957 A.I. Shirshov proved his famous theorem on height:

Theorem (A.I. Shirshov [7]**).** *For any finitely-generated associative PI-algebra A over a commutative ring R with* 1*, there exist a natural number h and elements $a_1, \dots, a_n \in A$ such that any element of A can be represented as an R-linear combination of elements of the form*

$$a_{i_1}^{\alpha_1} \cdots a_{i_k}^{\alpha_k},$$

where $k < h$.

We note that an algebra A over a commutative ring R with 1 is called a PI-algebra if A satisfies some polynomial identity $f = 0$ such that the ideal of the ring R generated by the coefficients of the highest-degree terms of the polynomial f contains 1.

The positive solution of Kurosh's problem for PI-algebras over a ring follows immediately from Shirshov's theorem. Indeed, since the elements $a_1, \ldots, a_n \in A$ are algebraic (the elements $a_1, \ldots, a_n$ are taken from the conclusion of the theorem on height), the degrees α_i are bounded. Hence the algebra A is a finitely-generated R-module.

Comparing the solutions of Kurosh's problem obtained by I. Kaplansky and A.I. Shirshov one notes that the solution of I. Kaplansky is based on the well-developed structure theory of rings, but makes little use of the PI-condition. In fact, the PI-condition is used in two statements: (1) The radical of a finitely-generated algebraic PI-algebra is nilpotent; (2) A matrix algebra of order n does not satisfy a polynomial identity of degree less than $2n$. These statements are quite easy from the contemporary point of view.

The solution of A.I. Shirshov does not use the structure theory at all. Moreover A.I. Shirshov also made little use of algebraicity. It follows from the above that it is sufficient to require algebraicity only for some finite set of elements. But the most important merit of the theorem on height is that it was proved for algebras over a commutative ring. Many of the results in ring theory concerning PI-algebras would not have been obtained if the theorem on height were true only for algebras over fields.

With the first results about PI-algebras it became clear that the PI-condition is a peculiar finiteness condition. In 1957 S. Amitsur [8] proved a remarkable theorem: The radical of a finitely-generated PI-algebra is a nil-ideal. This theorem once again corroborated that the PI-condition is a finiteness condition, and allowed V.N. Latyshev at that time to formulate rather boldly the problem: Is the radical of a finitely-generated PI-algebra nilpotent? (See [9].) A great contribution to the solution of this problem was made by Yu.P. Razmyslov [10] who proved that the radical of finitely-generated PI-algebra over a field is nilpotent if and only if the algebra satisfies some standard identity. To prove this statement, Yu.P. Razmyslov constructed an embedding of certain algebras into algebras which are algebraic over the center and then applied the theorem on height. Yu.P. Razmyslov was the first algebraist to apply the theorem on height very often and deeply. For algebras over a field of characteristic zero, Latyshev's problem was solved by A.R. Kemer [11] who proved that every finitely-generated PI-algebra over a field of characteristic zero satisfies a standard identity of some order. Indeed this result and the theorem of Razmyslov mentioned above imply the positive solution of Latyshev's problem in the case of characteristic zero. In 1982, A. Braun [12] solved Latyshev's problem positively for algebras over a commutative Noetherian ring. At present the theorem on the nilpotency of the radical of a finitely-generated PI-algebra is known as the theorem of Braun-Kemer-Razmyslov.

In 1974, Yu.P. Razmyslov introduced a new concept of trace identity, and proved that each trace identity of the matrix algebra of order n over a field of characteristic 0 follows from the Cayley-Hamilton trace identity of degree n and the identity $\mathrm{Tr}(1) = n$ [13]. Little later C. Procesi [14] proved actually the same result in the terms of invariants.

The Cayley-Hamilton identity of degree n has the form

$$X_n(x) = x^n + b_1(x)x^{n-1} + \cdots + b_n(x) = 0,$$

where the coefficient $b_m(x)$ is a form of degree m. In the case of characteristic zero the coefficients $b_m(x)$ can be represented as linear combinations of trace monomials of the form

$$\mathrm{Tr}(x^{i_1})^{\alpha_1}\mathrm{Tr}(x^{i_2})^{\alpha_2}\cdots\mathrm{Tr}(x^{i_k})^{\alpha_k}.$$

Of course this theorem of Yu.P. Razmyslov does not concern the theorem on height directly, but the idea of trace identities gives a way of embedding (if possible) a finitely-generated PI-algebra over a field into a finite-dimensional algebra (a matrix algebra) over a larger field (such algebras are called representable). Indeed, let a finitely-generated algebra A over a field F be embeddable into the matrix algebra $M_n(K)$, $F \subseteq K$. Consider the F-subalgebra $C = SA$, where S is the F-subalgebra (with unity) of the field K generated by all the elements $b_m(a)$ $(a \in A)$ where the elements $b_m(a)$ are the coefficients of the Cayley-Hamilton identity of degree n. It follows from this that in the case of characteristic zero the algebra A is embeddable into the algebra

$$D = A \otimes T\langle A\rangle / J,$$

where $T\langle A\rangle$ is the commutative algebra generated by the symbols $\mathrm{Tr}(a), a \in A$, the trace on the algebra $A \otimes T\langle A\rangle$ is defined by the formula

$$\mathrm{Tr}\Big(\sum a_k \otimes t_k\Big) = \sum \mathrm{Tr}(a_k)t_k,$$

and the ideal J is generated by the elements $X_n(d)$ $(d \in A \otimes T\langle A\rangle)$. In the case of characteristic p the algebra $A \otimes T\langle A\rangle$ is generated by the symbols $b_m(a)$ $(a \in A)$. The forms $b_m(x)$ are defined in the same manner but with more complicated formulas.

Assume that the algebra A is embeddable into the algebra D. Then the algebra A is embeddable into the algebra

$$D' = A \otimes T'\langle A\rangle / J \cap A \otimes T'\langle A\rangle,$$

where $T'\langle A\rangle$ is the subalgebra of $A \otimes T\langle A\rangle$ generated by the elements $b_m(a_i)$ (the elements a_i are taken from the conclusion of the theorem on height). The algebra D' is finitely-generated and algebraic over the commutative algebra $T'\langle A\rangle$ because it satisfies the Cayley-Hamilton identity. By the theorem on height the algebra D' is a finitely-generated $T'\langle A\rangle$-module. Since the algebra $T'\langle A\rangle$ is noetherian, by a theorem of K. Beidar [15] the algebras D' and A are representable. In 1995 the theorem of Razmyslov in the case of characteristic p was proved by A.R. Kemer at the multilinear level [16] and little later A.N. Zubkov proved this theorem at the homogeneous level [17].

A very important problem in the theory of PI-algebras was posed by W. Specht [18] in 1950: Does every associative algebra over a field of characteristic zero have a finite basis of identities? The finite basis problem makes sense for algebras over any field, and even for rings, groups and arbitrary general algebraic systems.

A positive solution of the finite basis problem for a given class of algebraic systems is a sort of classification of these algebraic systems in the language of identities.

A rather large number of papers have been devoted to Specht's problem for associative algebras over a field of characteristic zero. We note the most important results. In 1977 V.N. Latyshev [19] proved that any associative algebra over a field of characteristic zero satisfying a polynomial identity of the form

$$[x_1, \dots, x_n] \cdots [y_1, \dots, y_n] = 0,$$

has a finite basis of identities. This result was also obtained independently by G. Genov [20] and A. Popov [21].

In 1982 A.R. Kemer reduced the Specht problem to the finite basis problem for graded identities of finitely-generated associative PI-superalgebras [22] and in 1986 he solved the Specht problem positively [23]. The first proof of the theorem on the finite basis of identities was rather complicated. A little later in 1987 A.R. Kemer [24] proved that relatively free finitely-generated associative PI-superalgebras over a field of characteristic zero are representable. This theorem implies the theorem on the finite basis, and explains the reason why the Specht problem has a positive solution. This reason is that finite-generated PI-algebras over a field of characteristic zero cannot be distinguished in the language of identities from finite-dimensional algebras. More precisely, for every finitely-generated PI-algebra A there exists a finitely-dimensional algebra C such that the ideals of identities of these algebras are equal. In 1988 A.R. Kemer proved the same result for algebras over an infinite field of characteristic p [25].

The main idea of the proof of this theorem is to approach step-by-step the given T-ideal Γ by the ideals of identities of finite-dimensional algebras. At the first step there is constructed a finite-dimensional algebra C_0 such that

$$T[C_0] \subseteq \Gamma.$$

The existence of this algebra follows from the theorem on nilpotency of Braun-Kemer-Razmyslov and the theorem of J. Lewin [33]. The most difficult part of the proof is the following statement: If $T[C] \subseteq \Gamma$, $T[C] \neq \Gamma$ (C is finite-dimensional) then there exists a finite-dimensional algebra C' such that

$$T[C] \subseteq T[C'] \subseteq \Gamma, \qquad T[C] \neq T[C'].$$

The proof of this statement uses identities with forms and the standard application of the theorem on height which was described above.

Examples of infinitely-based algebras in the case of characteristic p were constructed in 1999 by V.V. Schigolev [26] and A.Ya. Belov [27].

In 1998 A.Ya. Belov [28] announced a positive solution of the local finite basis problem for algebras over a commutative noetherian ring, and announced a result about the representability of the relatively free algebra over a commutative noetherian ring in some weak sense: The relatively free finitely-generated PI-algebra A over a commutative noetherian ring R is embeddable into some algebra

A' over a commutative noetherian ring R' such that A' is a finitely-generated R'-module ($R \subseteq R'$). In other words the algebra A is embeddable into the algebra of endomorphisms of some finitely generated R'-module.

Regarding the methods of A.Ya. Belov we should note that most of the ideas of A.Ya. Belov are combinatorial, and come from the theorem on height and other results of A.I. Shirshov. A.Ya. Belov developed the combinatorial ideas of A.I. Shirshov which made it possible to consider more complicated combinatorial situations than in the theorem on height. In this sense one can call A.Ya. Belov a successor of A.I. Shirshov.

Another nice idea is applying Zariski closure. This idea was new for PI-theory. The algebras of endomorphisms of finitely generated modules over a ring have a more complicated structure than finite-dimensional algebras, but applying Zariski closure A.Ya. Belov proved that a finitely-generated PI-algebra A over a commutative noetherian ring R has the same identities as some algebra C over a commutative noetherian ring R', $R \subseteq R'$, satisfying the property that the radical of the algebra C splits off and is nilpotent, i.e., $C = P + \operatorname{Rad} C$, where the subalgebra C is semisimple. Applying Zariski closure A.Ya. Belov also obtained a lot of information about the semiprime part P. We note that the main results of A.Ya. Belov are not yet published.

We also mention the results devoted to the estimation of height in the theorem of A.I. Shirshov. The height $h(A)$ of an algebra A depends on the number of generators s and the minimal degree of identities $m = \deg(A)$. The estimate for the height which follows from the proof of the theorem on height is not satisfactory. In 1982 A.G. Kolotov [29] obtained the estimate

$$h(a) \leq s^{s^m}.$$

In [30] E.I. Zelmanov raised a question about the exponential estimate of the height. The positive answer was obtained by A. Ya. Belov in 1988 [31, 32].

References

[1] A.G. Kurosh, Ringtheoretische Probleme, die mit dem Burnsideschen Problem uber periodische Gruppen in Zusammenhang stehen, Izv. Akad. Nauk SSSR. Ser. Mat., v. 5(1941), 233–240. (Russian)

[2] E.S. Golod, On nil-algebras and residually finite p-groups, Izv. Akad. Nauk SSSR. Ser. Mat., v. 28(1964), 273–276; English transl. in Amer. Math. Soc. Transl., v. 48(1965).

[3] E.S. Golod, I.R. Shafarevich, On class field towers, Izv. Akad. Nauk SSSR. Ser. Mat., v. 28(1964), 261–272; English transl. in Amer. Math. Soc. Transl., v. 48(1965).

[4] N. Jacobson, Structure theory for algebraic algebras of bounded degree, Ann. of Math., v. 2(1945), 695–707.

[5] J. Levitzky, On a blem of A. Kurosh, Bull. Amer. Math. Soc., v. 52(1946), 1033–1035.

[6] I. Kaplansky, Rings with a polynomial identity, Bull. Amer. Math. Soc., v. 54(1948), 575–580.

[7] A.I. Shirshov, On rings with polynomial identities, Mat. Sb., v. 43(1957), 277–283; English transl. in Amer. Math. Soc. Transl., v. 119(1983).

[8] S.A. Amitsur, A generalization of Hilbert's Nullstellensatz, Proc. Amer. Math. Soc., v. 8(1957), 649–656.

[9] The Dniester notebook: unsolved problems in the theory of rings and modules, 3rd ed. (V.A. Andrunakievich, editor), Inst. Mat. Sibirsk. Otdel. Akad. Nauk SSSR, Novosibirsk, (1982) (Russian).

[10] Yu.P. Razmyslov, On the Jacobson radical in PI-algebras, Algebra i logika, v. 13(1974), 337–360; English transl. in Algebra and logic, 13(1974).

[11] A.R. Kemer, Capelli identities and the nilpotence of the radical of a finitely-generated PI-algebra, Dokl. Akad. Nauk SSSR, v. 255(1980), 793–797; English transl. in Soviet Math. Dokl., v. 22(1980).

[12] A. Braun, The radical in finitely-generated $P.I.$ algebra, Bull. (New Ser.) Amer. Math. Soc., v. 7(1982), 385–386.

[13] Yu.P. Razmyslov, Trace identities of full matrix algebras over field of characteristic zero, Izv. Akad. Nauk SSSR Ser. Mat., v. 38(1974), 723–756; English transl. in Math. USSR Izv. v. 8.(1974)

[14] C. Procesi, The invariant theory of $n \times n$-matrices, Adv. in Math., v. 19(1076), 306–381.

[15] K.I. Beidar, On theorems of A.I. Mal'tsev concerning matrix representations of algebras, Uspekhi Mat. Nauk, v. 41(1986), 161–162; English transl. in Ruusian Math. Sueveys., v. 41(1986).

[16] A.R. Kemer, Multilinear identities of the algebras over a field of characteristic p, Int. J. of Alg. and Comp., v. 5(1995), 189–197.

[17] A.N. Zubkov, On the generalization of the theorem of Procesi-Razmyslov, Algebra i Logika, v. 35(1996), 433–457. English transl. in Algebra and logic, 35(1996).

[18] W. Specht, Gesetze in Ringen. I, Math. Z., v. 52(1950), 557–589.

[19] V.N. Latyshev, On the finite basis property for the identities of certain rings, Uspekhi Mat. Nauk v. 32(1977), 259–260. (Russian)

[20] G.K. Genov, Some Specht varieties of associative algebras, Pliska Stud. Math. Bulgar v. 2(1981), 30–40. (Russian)

[21] A.P. Popov, On the Specht property for some varieties of assocuative algebras, Pliska Stud. Math. Bulgar v. 2(1981), 41–53. (Russian)

[22] A.R. Kemer, Varieties and Z_2-graded algebras, Izv. Akad. Nauk SSSR. Ser. Mat., v. 48(1982), 1042–1059; English transl. in Math. USSR Izv. v. 25(1985).

[23] A.R. Kemer, Finite bases for identities of associative algebras, Algebra i Logika v. 26(1987), 597–641; English transl. in Algebra and logic, 26(1987).

[24] A.R. Kemer, Representation of relatively free algebras, Algebra i Logika v. 27(1988), 274–294; English transl. in Algebra and logic, 27(1988).

[25] A.R. Kemer, Identities of finitely generated algebras over an infinite field, Izv. Akad. Nauk SSSR. Ser. Mat., v. 54(1990), 726–753; English transl. in Math. USSR Izv. v. 29(1990).

[26] V.V. Schigolev, Examples of infinitely-based T-ideals, Fund. and Appl. Math., v. 5(1999), 307–312.

[27] A.Ya. Belov, On non-spechtian varieties, Fund. and Appl. Math., v. 5(1999), 47–66.

[28] A.Ya. Belov, Local representability of relatively free associative algebras, Kurosh Algebraic conference – 98. Abstracts of talks. Ed. by Yu.A. Bahturin, A.I. Kostrikin, A. Yu. Ol'shansky. Moscow, 1998, 143–144.

[29] A.G. Kolotov, On the upper estimation of the height in finitely-generated PI-algebras, Sibirsk. Mat. Zh., v. 23(1982), 187–189; English transl. in Siberian Math. J. v. 23(1982).

[30] The Dniester notebook: unsolved problems in the theory of rings and modules, 3rd ed. (V.A. Andrunakievich, editor), Inst. Mat. Sibirsk. Otdel. Akad. Nauk SSSR, Novosibirsk, (1993)(Russian).

[31] A.Ya. Belov, Estimations for the height and Gelfand-Kirillov dimension of associative PI-algebras, Abstracts of Int. alg. Maltsev's conf. Novosibirsk, 1989. (Russian)

[32] A.Ya. Belov, Some estimations for nilpotence og nil-algebras over field of an arbitrary characteristic and height theorem, Comm. in algebra. v. 20(1992), 2919–2922.

[33] J. Lewin, A matrix representation for associative algebras. I, II, Trans. Amer. Math. Soc., v. 188(1974), 293–308, 309–317.

Brief Review of the Life and Work of A.I. Shirshov

Evgenii Kuzmin

An outstanding Russian mathematician, Anatoliĭ Illarionovich Shirshov was born on the 8th of August 1921 in the village of Kolyvan in the Novosibirsk Region. Before the war he started his studies in Tomsk University and then went to the front as a volunteer, and after demobilization in 1946 he continued his studies in Voroshilovgrad (now Lugansk) Pedagogical Institute. He combined his studies at the Institute with working as a mathematics teacher at a secondary school.

In 1950, A.I. Shirshov entered the Graduate School of the Faculty of Mechanics and Mathematics at Moscow State University (MSU) where he studied under the supervision of Professor A.G. Kurosh. After successful defence in 1953 of his Candidate of Science thesis, *Some problems in the theory of nonassociative rings and algebras*, he started working at the Department of Higher Algebra at MSU, first as Assistant, and starting in 1955, as Docent. In 1957–1960, A.I. Shirshov worked as the First Deputy Dean of the Faculty (the Dean was A.N. Kolmogorov). These years witnessed the blossoming of his creative scientific activity: in rapid succession he published works in which he laid the foundation for a new direction in modern algebra, the theory of rings that are nearly associative. In 1958, A.I. Shirshov defended his Doctor of Science thesis, *On some classes of rings that are nearly associative*, and in 1961 he was promoted to the rank of Professor.

In 1960, A.I. Shirshov, upon the invitation of Academicians S.L. Sobolev, I.N. Vekua and A.I. Malcev, decided to participate in the realization of an important national program: raising the level of scientific activity in his native region of Siberia. Like many other scientists of the time, who answered the call of the Government, he took an active part in the organization of the Siberian Branch of the Academy of Sciences of the USSR. Together with Academician A.I. Malcev, he became one of the founders of the Siberian school of algebra and logic. By his scientific, administrative and public activities, he made a great contribution to the foundation and development of the Mathematical Institute and the entire Siberian Branch. From 1960 to 1973, he was Deputy Director of the Mathematical

Institute of the Siberian Branch, and from 1967 to his last days, he was the Head of Division of Ring Theory of the Mathematical Institute. Simultaneously he conducted extensive pedagogical work as a Professor of the Department of Algebra and Mathematical Logic of Novosibirsk State University.

In 1964, A.I. Shirshov was elected a Corresponding Member of the Academy of Sciences of the USSR. He became a member of the Bureau of the Mathematical Division of the Academy of Sciences, a member of the National Committee of Soviet Mathematicians, Chairman of the Committee on Algebra of the Academy of Sciences, and also a member of several scientific councils and editorial boards.

The circle of scientific interests of A.I. Shirshov was rather extensive: algebra, mathematical logic, number theory, and projective geometry. However, his creative activity was concentrated mostly on ring theory and problems of algebra on the border with mathematical logic. When A.I. Shirshov started his research in the theory of rings that are nearly associative (1953), this theory simply did not exist: there were merely definitions of various classes of nonassociative rings and some isolated results about these rings. Now, it is a well-developed branch of algebra that includes as its components the theories of infinite-dimensional Lie algebras, the theory of alternative algebras, the theory of Jordan algebras, and also the theories of wider classes of algebras: Malcev algebras, binary-Lie algebras, right-alternative algebras, and others. The theory of rings that are nearly associative owes its modern development largely to the works of A.I. Shirshov and his students.

Already in the first works of A.I. Shirshov on ring theory we find brilliant results that have become classical: the theorem on freeness of subalgebras of free Lie algebras, and the theorem on embedding of an arbitrary Lie algebra with a countable number of generators into a Lie algebra with two generators. The bases of the free Lie algebra constructed by A.I. Shirshov (the Lyndon-Shirshov basis, the Hall-Shirshov bases) have played an important role in the solution of various types of algorithmic problems in the theory of Lie algebras, and also find applications in group theory. The attention of specialists was attracted by A.I. Shirshov's beautiful example of a Lie algebra over a ring which does not have an enveloping associative algebra over the same ring.

In group theory as well as ring theory an important role is played by problems of Burnside type; one of the best-known problems of this kind is the problem posed by A.G. Kurosh: is an associative algebraic algebra necessarily locally finite? As is well known, in the general case this problem of Kurosh was given a negative answer by E.S. Golod: on the other hand, this problem was given a positive answer by Kaplansky in the class of associative algebras which satisfy a polynomial identity. A.I. Shirshov suggested a general combinatorial approach that provides a positive solution to the problem of Kurosh for alternative and special Jordan algebras of bounded degree, and proves local nilpotency in the particular case of nil rings of bounded index. Turning his attention to associative rings with identical relations, A.I. Shirshov proved a theorem on local boundedness of their heights which is an essential strengthening of the theorem of Kaplansky. Introducing natural definitions of algebraicity and local finiteness over a subring of the center, he

obtained another generalization of Kaplansky's theorem: an alternative ring with a non-trivial identity, which is algebraic over a subring of its center, is locally finite over that subring.

Perhaps the most beautiful and difficult theorem of A.I. Shirshov is the statement that any Jordan algebra with two generators is special. This served as the starting point of a long series of works by American authors on Jordan algebras with two and three generators and on identities of Jordan algebras.

An important event for algebra was the publication of the monograph *Rings that are nearly associative* (Moscow, Nauka, 1978) written by A.I. Shirshov in collaboration with his students K.A. Zhevlakov, A.M. Slinko and I.P. Shestakov.

Among the algorithmic problems of algebra, to A.I. Shirshov belongs the solution of the word problem and the proof of the freeness theorem in the classes of commutative and anticommutative algebras and Lie algebras with one defining relation (using what is now called Gröbner-Shirshov basis theory). He also solved the word problem for solvable Lie algebras of index 2.

The works of A.I. Shirshov in the theory of rings that are nearly associative have cleared the way for further investigations in this area. In the works of his students and followers, many problems stated by A.I. Shirshov were solved: decidability of the word problem in the class of all Lie algebras and in the class of solvable Lie algebras; the problem of computing the basis rank of the varieties of alternative and Malcev algebras; the problem of describing the subalgebras of the free product of Lie algebras; the problem of local nilpotency of Jordan nil-algebras of bounded index; and others.

In the last years of his life, A.I. Shirshov was actively engaged in theory of projective planes. He developed a new algebraic approach to the study of projective planes; in particular he constructed a simple explicit "base" of a free projective plane. This approach allowed the formulation of a series of problems and a new viewpoint on the known results and problems in the theory of projective planes. To these problems A.I. Shirshov devoted an extended plenary report at the 14th All-Union Algebra Conference in Novosibirsk in 1977.

A.I. Shirshov devoted much attention and care to the training of the next generation of young scientists; he considered this the duty of a scientist. The school of algebra created by him was an object of personal pride.

A.I. Shirshov died on the 28th of February 1981 after a prolonged serious illness. The profound ideas of his works remain alive.

A Word about the Teacher

Evgenii Kuzmin

Strict and attentive, but at the same time fatherly and warm – a glance above the glasses (in a simple thin frame). He walked through the rows of students looking into notebooks, checking how the problem written on the blackboard was being solved. September 1955: a seminar in higher algebra is in progress for the students of the 104th section of the first year in the Faculty of Mechanics and Mathematics [Mehmat] at MSU. The seminar is run by the Teacher, Anatoly Illarionovich Shirshov, a young assistant in the Department of Higher Algebra at MSU. The Department is headed by Alexander Gennadievich Kurosh, the author of the textbook "A Course in Higher Algebra" and the monograph "Group Theory". A.I. stops next to me, nods with satisfaction and calls me to the blackboard: "Kuzmin, come and tell us how to solve this problem". I go up and explain it. Stopping me before I finish, A.I. asks the audience: "Who knows how to complete the solution? Vinogradov, come to the blackboard."

To be called to the blackboard is an honour; it must be earned. We have some rather strong folks in our class, future doctors of science Sasha Vinogradov, Borya Vainberg, Dima Fuks, Galina Turina (a talented mathematician who, unfortunately, was killed in an untimely accident rafting on northern rivers), Valera Kudryavtsev, Galina Blohina, Vitya Ivnitsky and your humble servant Zhenya (Evgenii) Kuzmin. The studies at Mehmat came easily to me, and I especially enjoyed algebra, with its strict logic of calculations and somewhat dry beauty of algebraic structures and abstract theories. We could not imagine how different higher mathematics is from school mathematics! And it was rare luck to meet your life-long Teacher during the first few days of university studies. A.I. noticed me, started to give me separate, more difficult and interesting homework assignments, and once offered a completely unusual problem:

"There is a theorem of Shirokov which gives a positive answer to the conjecture of Kaplansky on the quasi-nilpotency of the commutator in an associative valuation ring under one extra condition of an algebraic nature. Shirokov proved his theorem using methods of functional analysis. But Kaplansky himself is an

algebraist, and his problem is also formulated algebraically. So I think that there must exist a purely algebraic proof of this theorem. Try to find such a proof!"

After some time I managed to do it! (Later in my diploma thesis I extended Shirokov's theorem to flexible valuation rings, which are a wide generalization of associative rings.) The reaction of Shirshov was unexpected. He brought me to his seminar, where the participants were students one or two years older than me, and said: "Look at this boy. He solved a problem of Kaplansky!" Of course, it was Shirokov who solved the problem of Kaplansky; I merely re-proved his theorem. A.I. simply wanted to praise me, and his words gave me wings. I began to attend his seminar and then his special course in ring theory, where he explained his ideas, amazing in their beauty and complexity, related to alternative, Jordan and Lie rings – the ideas that created a new direction in ring theory and were the basis of Shirshov's doctoral thesis.

The core of Shirshov's seminar consisted of five people: L. A. Bokut, G.V. Dorofeev, E.N. Kuzmin, V.N. Latyshev, and K.A. Zhevlakov, whom somebody called the "magnificent five". These five direct students of A.I. became the basis on which the school of Shirshov emerged; in the framework of this school the well-known doctors of science were formed: V.T. Filippov, A.Ya. Kanel-Belov, A.V. Iltyakov, A.R. Kemer, V.K. Kharchenko, P.S. Kolesnikov, G.P. Kukin, Yu.N. Malcev, Yu.A. Medvedev, A.A. Nikitin, S.V. Pchelintsev, V.V. Shchigolev, I.P. Shestakov, A.M. Slinko, S.R. Sverchkov, U.U. Umirbaev, E.I. Zelmanov, V.N. Zhelyabin – not to mention numerous candidates of science (like members of Shirshov's Ring Theory Department at Sobolev Institute A.Z. Ananin, V.N. Gerasimov, A.T. Kolotov, I.V. Lvov, A.N. Koryukin, V.A. Parfenov, A.P. Pojidaev, V.G. Skosyrskii, O.N. Smirnov, A.I. Valitskas, S.Yu. Vasilovskii).

A distinguishing trait of Shirshov's creative work was its exceptional individuality: he wrote all his main works by himself, without co-authors. This trait was largely inherited by his students. One day, after a regular session of the Academy of Sciences, he recounted that, during a break between meetings, he was approached by I.M. Gelfand, a well-known "co-authorizer", who said, holding Shirshov by a button of his jacket, that he had some ideas about Jordan algebras: "Would you like, Anatoly Illarionovich, to think about them?" A.I. refused; he did not want to join the numerous ranks of co-authors of Izrail Moiseevich.

Something similar also happened to me. After struggling with the problem of existence of an analytic Moufang loop with a given tangent Malcev algebra, I ventured to ask A.I. for help. The answer was like a cold shower: "If you don't want to work on this problem yourself, I will give it to somebody else". In a few years it dawned upon me how to make use of the Campbell-Hausdorff series, and the proof was found! How happy was A.I. for me! He used to say: "This is your second doctoral thesis".

Shirshov's treatment of his students was truly fatherly. He was happy for our successes as a father would be – and what were our successes compared to his truly outstanding achievements?! – and he looked after us even in everyday life. It was impossible not to feel a grateful love for this Man, our great Teacher.

A.I. Shirshov's Works on Alternative and Jordan Algebras

Ivan Shestakov and Efim Zelmanov

This survey is an extended version of Section 3 of the paper [3] by L.A. Bokut and the first author, which was based on the report delivered at the Second International Conference on Algebra in memory of A.I. Shirshov, held in Barnaul, Russia in August 1991.

We consider here the contribution of A.I. Shirshov to the theories of alternative and Jordan algebras. In the middle of the 1950s, when A.I. Shirshov began to investigate these algebras, there was no general structure theory. Only the structure theory of finite-dimensional algebras had been developed in the works of M. Zorn, A.A. Albert, N. Jacobson, R.D. Schafer, and others [9, 20]. As to the infinite-dimensional case, only some isolated results, such as the Bruck-Kleinfeld-Skornyakov theorem on alternative division rings [2, 27], had begun to appear. The results of A.I. Shirshov and, more importantly, the ideas and methods developed in his papers, provided a basis for the creation of structure theories for alternative and Jordan algebras in the general case.

Recall that at that time the structure theory of associative rings was already well-developed. One of its main achievements was I. Kaplansky's solution [11] of the A.G. Kurosh problem for algebraic PI-algebras. Although the Kurosh problem is easily reformulated for alternative or Jordan algebras, the proof of I. Kaplansky could not be translated to these classes of algebras since they lacked any structure theory. As has already been mentioned in other surveys in this volume, A.I. Shirshov looked at the Kurosh problem from the combinatorial point of view, and this approach permitted him not only to obtain more profound results in the associative case, but also to solve the problem for alternative and special Jordan algebras. Furthermore, these works were of fundamental significance for the entire development of the theory of alternative rings. They clearly demonstrated the intrinsic unity of the theories of Jordan and alternative algebras.

Recall that an algebra A is said to be *alternative* if, for all $a, b \in A$, the subalgebra generated by a, b is associative. An algebra J is said to be *Jordan* if it satisfies the identities $xy = yx$ and $(x^2y)x = x^2(yx)$.

The best-known example of an alternative algebra is the algebra of Cayley numbers. The typical example of a Jordan algebra is the algebra $A^{(+)} = \langle A, +, \circ \rangle$, where A is an associative algebra and $a \circ b = \frac{1}{2}(ab + ba)$. If a Jordan algebra J is embeddable into the algebra $A^{(+)}$ for a suitable associative algebra A, then J is called a *special* Jordan algebra; in this case we denote by $A(J) = \mathrm{alg}_A(J)$ the enveloping algebra of J.

A.I. Shirshov proved that if J is an algebraic special Jordan PI-algebra, then the algebra $A(J)$ is locally finite. In particular, the algebra J itself is locally finite in this case. The proof of this striking result should be considered together with the proof of the celebrated Height Theorem (see the other surveys in this volume). Both proofs are based on a Ramsey-type combinatorial statement which has implications far beyond algebra.

A word v is said to be *n-divisible* if it can be represented as $v = v_1 \dots v_n$ where $v > v_{\sigma(1)} \dots v_{\sigma(n)}$ lexicographically for an arbitrary nonidentical permutation σ.

The Shirshov $N(k,s,n)$-lemma. *For arbitrary integers $k, s, n \geq 1$ there exists an integer $N(k,s,n)$ such that an arbitrary word in $x_1, \dots x_k$ of length $N(k,s,n)$ contains a subword u^s or an n-divisible subword.*

The proof of this lemma involves induction on k. Let $k \geq 2$. Modulo the induction assumption it is sufficient to consider only words in the finite set

$$T = \{\, x_k^i x_{i_1} \dots x_{i_r} \mid 1 \leq i < s, 1 \leq i_1, \dots, i_r \leq k-1, r < N(k-1,s,n) \,\}.$$

An $(n-1)$-divisible word in T gives rise to an n-divisible word in the alphabet $x_1, \dots, x_k$. The key observation of Shirshov that allowed him to apply this combinatorics to special Jordan algebras is that an arbitrary T-word is a Jordan word; that is, a lexicographically greatest monomial in a homogeneous Jordan expression in $x_1, \dots, x_k$.

Shirshov's result works for alternative algebras as well. If B is an alternative algebra, then $B^{(+)}$ is a special Jordan algebra, and moreover, an enveloping algebra $A(B^{(+)})$ is isomorphic to the algebra of right multiplications,

$$R(B) = \mathrm{alg}\langle R_b \mid b \in B \rangle, \quad R_b \colon x \mapsto xb.$$

Thus, we have the transitions,

$$\begin{array}{l} B \text{ is an} \\ \text{algebraic} \\ \text{alternative} \\ \text{PI-algebra} \end{array} \Longrightarrow \begin{array}{l} B^{(+)} \text{ is an} \\ \text{algebraic} \\ \text{Jordan} \\ \text{PI-algebra} \end{array} \Longrightarrow \begin{array}{l} A(B^{(+)}) \\ \cong R(B) \text{ is} \\ \text{locally} \\ \text{finite} \end{array} \Longrightarrow \begin{array}{l} B \text{ is} \\ \text{locally} \\ \text{finite,} \end{array}$$

which give a solution to the Kurosh problem for alternative algebras. The idea of the transition from associative algebras to alternative algebras via Jordan algebras,

$$\text{Associative algebras} \overset{\text{Jordan algebras}}{\Longrightarrow} \text{Alternative algebras,}$$

plays a crucial role in many further investigations.

We remark that the reduction of the Restricted Burnside Problem [36, 37, 38] to Engel Lie algebras [35] was based on the Lie analogue of Shirshov's $N(k,s,n)$-lemma. A word in $x_1,\dots,x_k$ is said to be a *Lie word* if it is the lexicographically greatest word in a homogeneous linear combination of commutators in $x_1,\dots,x_k$.

Theorem. *For arbitrary integers $k,s,n\geq 1$, there exists an integer $L(k,s,n)$ such that an arbitrary word in $x_1,\dots,x_k$ of length $L(k,s,n)$ contains a subword u^s where u is a Lie word, or a subword $v_1u_1v_2u_2\dots u_{n-1}v_n$ where $v_1,\dots,v_n$ are Lie words, such that*

$$v_1u_1v_2\dots u_{n-1}v_n > v_{\sigma(1)}u_1v_{\sigma(2)}u_2\dots u_{n-1}v_{\sigma(n)},$$

lexicographically, for any nonidentical permutation σ.

In [34] the Kurosh problem for arbitrary (not necessarily special) Jordan PI-algebras was solved. It is a typical situation for Jordan algebras, when a theorem is first proved for special algebras and then extended to the class of all algebras. In this connection, it is very important to determine conditions sufficient for speciality of an algebra. In this direction, A.I. Shirshov proved the fundamental theorem that *the free Jordan algebra with two generators is special.* Combined with the earlier result by P. Cohn [4], the theorem implies that *every Jordan algebra with two generators is special.* A.I. Shirshov considered this theorem as one of his best results. The claim is quite simple whereas the proof is difficult and sophisticated. The theorem served as a source of diverse research in several directions. The first of them relates to the investigation of the structure of free Jordan algebras $J[x,y,z,\dots]$ with more than two generators.

We ought to say that A.I. Shirshov was always interested in studying problems related to the structure of free algebras. His first works are devoted to free Lie algebras. His last results are concerned with the structure of free projective planes. He also formulated a series of questions on the structure of free Jordan, alternative, Malcev, and other algebras [5].

The first result on the structure of the free Jordan algebra $J[X]$, $|X|\geq 3$, was obtained in 1959 by A.A. Albert and L.J. Paige [1]. They proved that this algebra is neither special nor even a homomorphic image of a special Jordan algebra. (Earlier P. Cohn in [4] showed that the class of special Jordan algebras is not closed under homomorphic images.) This implies that the algebra $J[x,y,z]$ contains nonzero elements vanishing in every special Jordan algebra (such elements are called *s-identities*). In 1966, C.M. Glennie [7] presented a concrete s-identity of degree 8. Until now, no essentially new s-identities have been found, and the question of their description is still open. Moreover, he proved that there are no s-identities of degree ≤ 7 and no homogeneous s-identities in three variables, which are linear in one of them. It is curious that these two facts provided all the identities that are needed for the structure theory [33].

In the case of special algebras, the role of a free algebra is played by the *free special Jordan algebra* $SJ[X]$, which is defined as the minimal subspace of the free associative algebra $\mathrm{Assoc}[X]$ that contains X and is closed with respect to

the Jordan product $a \circ b$. The elements of $SJ[X]$ are called *Jordan elements*. It is easy to see that $SJ[X] \subseteq H(\mathrm{Assoc}[X], *)$, where $H(\mathrm{Assoc}[X], *)$ is the subspace of symmetric elements of $\mathrm{Assoc}[X]$ with respect to the involution $*$ which is the identity on X: $(x_1 x_2 \cdots x_n)^* = x_n \cdots x_2 x_1$. The subspace $H(\mathrm{Assoc}[X], *)$ is closed with respect to the Jordan product and hence may be considered as a Jordan algebra. It is generated as an algebra by the set X and by all *tetrads* $\{x_i x_j x_k x_l\} = x_i x_j x_k x_l + x_l x_k x_j x_i$; when $|X| \leq 3$ then $H(\mathrm{Assoc}[X], *) = SJ[X]$, and when $|X| > 3$ then $H(\mathrm{Assoc}[X], *)$ strictly contains $SJ[X]$ (since tetrads in general are not Jordan elements).

An important tool to "diminish the gap" between $H(\mathrm{Assoc}[X], *)$ and $SJ[X]$ was invented by E. Zelmanov [33]. An element $n \in SJ[X]$ is called a *tetrad-eater* if the tetrad $\{nabc\}$ is a Jordan element for any $a, b, c \in SJ[X]$. E. Zelmanov constructed an ideal I in $SJ[X]$ which consists of tetrad-eaters; it satisfies the condition $I = H(A(I), *)$, that is, I coincides with the subspace of symmetric elements in its enveloping algebra. The tetrad-eater ideal I is essential to the classification of prime Jordan algebras [33]. Among various corollaries of the classification, it was proved that the algebra $J[X]$ is not prime for $|X| > 3$. The generators of I in [33] are of quite large degree. The following example, due to V. Skosyrsky [28], presents a tetrad-eater of minimal known degree:

$$\lambda(x, y, z, t, u) = [[[x, y]^2, x], [[[z, t]^2, z], u]].$$

One can easily check that this is a Jordan element; moreover, the ideal of $SJ[X]$ generated by all homogeneous elements of this type consists of tetrad-eaters.

As of now there are no known criteria to determine when an element of $\mathrm{Assoc}[X]$ is a Jordan element.

A series of interesting results on the structure of the free Jordan algebra $J[X]$ was obtained by Yu.A. Medvedev [16, 17]. He proved in particular that

(1) If $|X| \geq 3$, then the algebra $J[X]$ has nontrivial center and contains Albert subrings (central orders in 27-dimensional exceptional simple Jordan algebras).
(2) If $|X| \geq 32$, then $J[X]$ contains nonzero nilpotent elements and nontrivial nil ideals.

In the joint paper by Yu.A. Medvedev and E. Zelmanov [18], it was proved that

(3) If $|X|$ is infinite, then the nil radical of $J[X]$ is neither nilpotent nor solvable.

The first two results had their analogues in the theory of free alternative algebras [32]. The third is specific for Jordan algebras. As E.I. Zelmanov and I.P. Shestakov showed [39], the nil radical of a free alternative algebra over a field of characteristic zero is nilpotent. It is interesting that nilpotency of the radical in the alternative case as well as nonnilpotency of the radical in the Jordan case was proved by analyzing the structure of simple superalgebras and their identities.

Another direction stemming from the Shirshov theorem on two-generated Jordan algebras relates to investigating the problems of speciality, finding certain criteria of speciality, and studying the influence of identities of an algebra on its

speciality. Together with A.I. Shirshov, P. M. Cohn was a pioneer in this direction [4]. The direction was further developed in the works by A.M. Slin'ko [29] and S.R. Sverchkov [30, 31]. In the papers [14, 22, 19, 13, 8, 23, 24] this approach was extended to Jordan superalgebras and to other classes of algebras. We present one of the results of S.R. Sverchkov [30]: The class of special Jordan algebras regarded as a quasivariety cannot be determined by a set of quasi-identities (that is, expressions of the type "$f(x) = 0 \Rightarrow g(x) = 0$") in finitely many variables.

One of the important and difficult problems in the theory of nonassociative algebras is the construction of effective bases of free algebras. A.I. Shirshov constructed bases for free Lie algebras, and free commutative and anticommutative algebras, and formulated this problem for free alternative, free Jordan, free Malcev, and other free algebras [5, problem 1.160]. In the case of free Jordan algebras, no effective bases are known for $J[X]$, $|X| > 2$ and $SJ[X]$, $|X| > 3$. In the case of alternative algebras, a base for the free algebra $Alt[x, y, z]$ was constructed by A. Iltiakov [10] who also proved that this algebra has no nilpotent elements, contrary to $Alt[X]$ for $|X| > 3$. In [25, 26] bases of free Malcev and alternative superalgebras on one odd generator are constructed.

The structure of the free alternative algebras $Alt[X]$ for $|X| > 3$ was studied by I.P. Shestakov (see [32]). In particular, in [21] he solved the following problem of A.I. Shirshov [5, problem 1.159]: *Let Alt_n denote the variety generated by a free alternative algebra with n generators. Does the chain of varieties*

$$Alt_1 \subseteq Alt_2 \subseteq \cdots \subseteq Alt_n \subseteq Alt_{n+1} \subseteq \cdots ,$$

stabilize after a finite number of steps? The answer turned out to be negative. It was proved in [21] that $Alt_n \subset Alt_{2^n+1}$ strictly for any n. Later, V. T. Filippov [6] showed that if the base field has characteristic different from 2 and 3 then $Alt_n \subset Alt_{n+1}$ strictly for any n.

A.I. Shirshov posed the analogous problem for free Jordan, free Malcev, and other free algebras. A negative answer for the variety Mal of Malcev algebras was obtained in [21] by I.P. Shestakov; later V.T. Filippov refined this result in [6] by proving that $Mal_n \subset Mal_{n+1}$ strictly for any $n \neq 3$. For $n = 3$ the question is still open. The corresponding problem for the variety Jor of Jordan algebras remains open; it is known only that $Jor_1 \subset Jor_2 \subset Jor_3$ strictly. In the light of the above results, it seems very interesting to construct bases of the free Jordan and free Malcev algebras on three generators. In particular, are these algebras semiprime like $Alt[x, y, z]$?

It seems natural to reformulate the problem above on the chain of varieties in the framework of superalgebras. Recall that a variety of algebras $\mathcal{M}$ is said to have finite *basic rank* if it can be generated by a finitely generated algebra; the minimal number of generators in this case is called the basic rank of $\mathcal{M}$. For example, the varieties of all associative and all Lie algebras have basic rank 2, the varieties Alt and Mal, or the variety generated by a Grassmann algebra on an infinite number of generators, have infinite basic rank. Similarly, we will say that a variety $\mathcal{M}$ has a finite *basic superrank* if the corresponding variety of $\mathcal{M}$-superalgebras is

generated by a finitely generated superalgebra; a pair (m, n) of m even and n odd generators of such a superalgebra we call a basic superrank of $\mathcal{M}$ if it is minimal right lexicographically.

The notion of basic superrank is a more refined characteristic of a variety; this fact is evidenced by the following theorem by A.R. Kemer [12] which played a crucial role in his solution of the Specht problem: *Every variety of associative algebras over a field of characteristic 0 has a finite basic superrank*. The variety of alternative algebras which are solvable of index 2 provides a nonassociative example: it has infinite basic rank but its basic superrank is $(0, 1)$. In this connection, the following question arises: *What is the value of basic superrank for the variety Alt of alternative alebras? Is it finite?*

Finally, we want to mention one work by A.I. Shirshov which greatly influenced the development of the theory of nonassociative algebras. This is the survey *Some questions of the theory of rings that are nearly associative*. Many students of A.I. Shirshov, and the students of his students, began their acquaintance with ring theory while perusing this article. On the one hand, it is accessible for beginners, on the other hand, it contains a whole program of further study, and a series of attractive and still open problems.

In recent years, the theory of nonassociative algebras has gained wide recognition; its methods penetrate deeply into other domains of mathematics, not only into algebra but also into geometry, analysis, and theoretical physics. A great part of the merit for this belongs to A.I. Shirshov, who was a harbinger of the theory and whose marvelous theorems will adorn it forever.

References

[1] A.A. Albert and L.J. Paige, On a homomorphism property of certain Jordan algebras, Trans. Amer. Math. Soc. 92 (1959), 20–29.

[2] R.H. Bruck and E. Kleinfeld, The structure of alternative division rings, Proc. Amer. Math. Soc. 2, no. 6 (1951), 878–890.

[3] L.A. Bokut' and I.P. Shestakov, Some results by A.I. Shirshov and his school, Contemporary Mathematics, 184, 1995, 1–12.

[4] P. Cohn, On homomorphic images of special Jordan algebras, Canad. J. Math. 6 (1954), 253–264.

[5] Dnestrovskaya Tetrad', Open problems in the theory of rings and modules, Institute of Mathematics, Novosibirsk, 1993 (in Russian). English translation: Lect. Notes Pure Appl. Math., 246, Non-associative algebra and its applications, 461–516, Chapman & Hall / CRC, Boca Raton, FL, 2006.

[6] V.T. Filippov, On varieties of Malcev and alternative algebras generated by algebras of finite rank, Trudy Inst. Mat. SOAN SSSR, Novosibirsk, v. 4, 1984, 139–156.

[7] C.M. Glennie, Some identities valid in special Jordan algebras but not valid in all Jordan algebras, Pacific J. Math. 16, no. 1 (1966), 47–59.

[8] A.N. Grishkov, I.P. Shestakov, Speciality of Lie-Jordan Algebras, J. Algebra, 237 (2001), 621–636.

[9] N. Jacobson, Structure and Representations of Jordan Algebras, AMS colloquium Publ., vol. 39, Providence, R.I., 1968.

[10] A.V. Iltiakov, Free alternative algebras of rank 3. Algebra i Logika 23, No. 2 (1984), 136–158.

[11] I. Kaplansky, Topological representations of algebras. II, Trans. Amer. Math. Soc. 68, no. 1 (1950), 62–75.

[12] A.R. Kemer, Varieties and Z_2-graded algebras, Izv. Akad. Nauk SSSR, Ser. Mat. 48(1984), 1042–1059.

[13] M.C. López Díaz, I.P. Shestakov and S.R. Sverchkov, On speciality of Bernstein Jordan algebras, Communications in Algebra, 28 (2000), no. 9, 4375–4387.

[14] K. McCrimmon, Speciality and nonspeciality of two Jordan superalgebras, J. Algebra 149 (1992), no. 2, 326–351.

[15] K. McCrimmon, Taste of Jordan Algebras, Springer Berlin Heidelberg, 2004.

[16] Yu.A. Medvedev, Free Jordan algebras, Algebra i Logika, 27, no. 2 (1988), 172–200.

[17] Yu.A. Medvedev, On nilpotent elements of a free Jordan algebra, Sibirsk. Mat. Zh. 26, no.2 (1985), 1402–1408.

[18] Yu.A. Medvedev, E.I. Zelmanov, Some counterexamples in the theory of Jordan algebras. Nonassociative algebraic models (Zaragoza, 1989), 1–16, Nova Sci. Publ., Commack, NY, 1992.

[19] C. Martínez, I. Shestakov and E. Zelmanov, Jordan superalgebras defined by brackets, J. London Math. Soc. (2) 64 (2001), no. 2, 357–368.

[20] R.D. Schafer, An Introduction to Nonassociative Algebras, Acad. Press., New York, 1966.

[21] I.P. Shestakov, On a problem by Shirshov, Algebra i Logika 16, no. 2 (1977), 227–246.

[22] I.P. Shestakov, A quantization of Poisson superalgebras and a speciality of Jordan Poisson superalgebras, Algebra i Logika, 32, N 5 (1993), 572–585.

[23] I.P. Shestakov, The speciality problem for Malcev algebras and deformations of Malcev Poisson algebras, in "Non-Associative Algebra and Its Applications", Proceedings of the IV International Conference on Non-Associative Algebra and Its Applications, July 1998, São Paulo, 365–371, Marcel Dekker, NY; Series Name: Lecture Notes in Pure and Applied Mathematics, v. 211, 2000.

[24] I.P. Shestakov, Every Akivis algebra is linear, Geometriae dedicata, 77 (1999), no. 2, 215–223.

[25] I.P. Shestakov, Free Malcev superalgebra on one odd generator, Algebra and Applications, 2 (2003), no. 4, 451–461.

[26] I. Shestakov, N. Zhukavets, The free alternative superalgebra on one odd generator, International Journal of Algebra and Computation (IJAC) 17, no. 5/6 (2007), 1215–1247.

[27] L.A. Skornyakov, Alternative skew fields, Ukrain. Mat. Zh. 2, no. 1 (1950), 70–85.

[28] V.G. Skosyrsky, Strongly prime noncommutative Jordan algebras, Trudy Inst. Mat. SOAN SSSR, Novosibirsk, v. 16, 1989, 131–164.

[29] A.M. Slin'ko, On special varieties of Jordan algebras, Mat. Zametki 26, no. 3 (1979), 337–344.

[30] S.R. Sverchkov, On the quasivariety of special Jordan algebras, Algebra i Logika 24, no. 5 (1983), 563–573.

[31] S.R. Sverchkov, Varieties of special algebras, Comm. in Algebra 16, no. 9 (1988), 1877–1920.

[32] K.A. Zhevlakov, A.M. Slin'ko, I.P. Shestakov, A.I. Shirshov, Rings that are nearly associative, Nauka, Moscow, 1978.

[33] E.I. Zelmanov, On prime Jordan algebras. II, Sibirsk. Mat. Zh. 24 (1983), 83–104.

[34] E.I. Zelmanov, Absolute zero divisors and algebraic Jordan algebras, Sibirsk. Mat. Zh. 23, no. 6 (1982), 100–116.

[35] E.I. Zelmanov, Some problems in the theory of groups and Lie algebras, Math. Ussr-Sb. 66 (1990), no. 1, 159–168

[36] E.I. Zelmanov, Solution of the restricted Burnside problem for groups of odd exponent, Izv. Akad. Nauk SSSR, Ser. Mat. 54, no. 1 (1991), 41–60.

[37] E.I. Zelmanov, Solution of the restricted Burnside problem for 2-groups, Mat. Sb. 182, no. 4 (1991), 568–592.

[38] E.I. Zelmanov, On the restricted Burnside problem. Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990), 395–402, Math. Soc. Japan, Tokyo, 1991.

[39] E.I. Zelmanov and I.P. Shestakov, Prime alternative superalgebras and nilpotency of the radical of a free alternative algebra, Izv. Akad. Nauk SSSR, Ser. Mat. 53, no. 1 (1990), 42–59.

Publications of A.I. Shirshov

Subalgebras of Free Lie Algebras

A.I. Shirshov

1. Introduction

In the work of A.G. Kurosh [2] it is proved that every subalgebra of a free nonassociative algebra is free. It would be natural to investigate the possibility of transferring this theorem to the most important classes of relatively free algebras whose general definition was given in the work of A.I. Malcev [3].

The widest class of such algebras that includes all classes of algebras that have been studied sufficiently deeply is the class of power associative algebras, i.e., the algebras in which each element generates an associative subalgebra. However, the corresponding theorem for this class of algebras is false, because the free associative algebra with one generator already contains subalgebras that are not free (see A.G. Kurosh [2]). For the same reason, this theorem does not hold for Jordan algebras, for alternative algebras, and also for right or left alternative algebras. It is not difficult to convince oneself that this theorem does not hold for power-commutative or flexible algebras either, for reasons similar to those stated above.

These considerations, however, are not valid for free Lie algebras, since in them a single element generates a one-dimensional subspace with zero multiplication, for which the theorem on subalgebras holds trivially. In the present work, it is proved that every subalgebra of every free Lie algebra is free.

This work was carried out under the supervision of A.G. Kurosh, to whom I find it my pleasant duty to express deep gratitude.

2. Preliminary concepts

Let $R = \{a_\alpha\}$ be a set of symbols where α ranges over some nonempty set of indices. From elements of R one can form nonassociative words of various lengths as is done in the work of A.G. Kurosh [2].

Mat. Sbornik N.S. 33 (75), (1953), no. 2, 441–452.

Definition 1. We will call words of length 1, i.e., elements of R, *regular words*, and we will order them arbitrarily. Assuming that regular words of length less than n, $n > 1$, are already defined and ordered by the relation $\leq$ in such a way that shorter words precede longer words, we call a word w of length n *regular* if the following conditions are satisfied:

1) $w = uv$ where u and v are regular words and $u > v$;
2) if $u = u_1u_2$ then $u_2 \leq v$.

We will order arbitrarily the regular words of length n defined in this way, and declare that they are greater than shorter words.

Definition 2. Suppose we have a regular word d. We will call a regular word w, $w > d$, *d-reducible* if $w = uv$, $v > d$, and *d-irreducible* otherwise.

Obviously, for each regular word w, $w > d$, one can determine if it is d-reducible or d-irreducible. If it turns out that w is d-reducible, then $w = uv$ where each word u, v is regular and greater than d, and thus one can determine if each is d-reducible or d-irreducible. Continuing this process, we will clearly arrive at a unique representation of the word w as a product (with some arrangement of brackets) of d-irreducible words. We will call this representation a *d-factorization* of w.

Definition 3. We will say that two nonassociative words u and v *have the same content relative to* R if each element $a_\alpha \in R$ occurs in u and v the same number of times.

Clearly, the words that have the same content relative to R also have the same length.

Let $\mathcal{A}$ be a free Lie algebra over a field P with the same set R of free generators. The elements of $\mathcal{A}$ are linear combinations of nonassociative words formed from elements of R with coefficients from the field P; in this case, two elements are considered equal if one can be obtained from the other by a finite number of applications of the distributive laws and the identical relations:

$$x^2 = 0, \tag{1}$$

$$(xy)z + (yz)x + (zx)y = 0, \tag{2}$$

or identical transformations in the additive group.

Hall [1] proved:

Theorem 1. *Regular words, for any fixed choice of ordering in the definition, form a basis of the algebra $\mathcal{A}$.*

The proof of this theorem can found in the cited work of Hall. (It is easy to see that Hall's assumption of finiteness of the number of generators of the algebra $\mathcal{A}$ is not essential.) In the following, it is important that the process used in that proof allows one to express each word in the algebra $\mathcal{A}$ as a linear combination of regular words of the same content relative to R.

Theorem 1 and the above remark imply the following result of a combinatorial nature:

Corollary 1. *The number of regular words of the same given content relative to R does not depend on the choice of ordering in the definition of regular words.*

Indeed, let regular words be defined in two different ways, and let M_i $(i = 1, 2)$ be the sets of all words which are regular according to the first (respectively second) sense and have the same given content relative to R. By Theorem 1 the elements of each set M_i in $\mathcal{A}$ are linearly independent over P, and any element of each of these sets is a linear combination of the elements of the other set, which proves the corollary.

Given an arbitrary Lie algebra $\mathcal{L}$, one can speak of a *regular form* of its elements. For this, one must fix some set $M = \{v_\gamma\}$ of generators and consider the homomorphism of the free Lie algebra $\overline{\mathcal{L}}$, with the set $\overline{M} = \{\overline{v_\gamma}\}$ of free generators which are in one-to-one correspondence with the elements of M, onto $\mathcal{L}$.

An M-word, i.e., an element of $\mathcal{L}$ of the form $w = v_{\gamma_1} v_{\gamma_2} \cdots v_{\gamma_k}$ where $v_{\gamma_j} \in M$ with some arrangement of brackets, will be called *M_τ-regular* if, for the set $\overline{M}$ the regular words have been defined in some way τ and the word $\overline{w} = \overline{v_{\gamma_1}}\,\overline{v_{\gamma_2}} \cdots \overline{v_{\gamma_k}}$ in elements of $\overline{M}$ is regular. Generally speaking, for an element of $\mathcal{L}$, an *M_τ-regular form*, i.e., a representation as a linear combination of M_τ-regular words, is not uniquely defined, but for any M-word w there exists an expression as a linear combination of M_τ-regular words with the same content relative to M as w. To find such an expression, one must find an analogous expression for the word $\overline{w}$ and then pass to the homomorphic image.

For consistency of notation in what follows, we will denote by $\overline{\mathcal{D}}$ the free Lie algebra on the set of free generators that are in one-to-one correspondence with the generators of the given Lie algebra $\mathcal{D}$.

Definition 4. We will say that a set $\mathcal{R}$ of elements of the free Lie algebra $\mathcal{A}$ is *independent* if $\mathcal{R}$ generates a free subalgebra of $\mathcal{A}$ and is a system of free generators of that subalgebra.

For example, the set R itself is independent. In what follows we will assume that for the set R the regular words are defined in some fixed way and we will call those words *R-regular*.

Let d be a fixed R-regular word, and K_d the set of d-irreducible words. The set K_d generates some subalgebra $\mathcal{A}_d$ of $\mathcal{A}$. The set K_d consists of R-regular words, and thus it is already ordered by the fixed order of R-regular words. We will transfer this order to the set $\overline{K_d}$ of free generators of the free Lie algebra $\overline{\mathcal{A}_d}$, and starting with this order we will define in some fixed way $\overline{K_d}$-regular $\overline{K_d}$-words. After that, it also makes sense to speak of K_d-regular K_d-words. As was shown above, there exists a representation of each K_d-word as a linear combination of regular K_d words of the same content relative to K_d.

Lemma 1. *Every K_d-word can be represented as a linear combination of K_d-words of the same content relative to K_d which are in fact R-regular.*

This lemma is obvious for K_d-words whose K_d-length (i.e., length relative to K_d) is 1, since the elements of K_d are in fact R-regular.

Suppose the lemma has been proved for K_d-words whose K_d-length is less than n, $n > 1$. A word w whose K_d-length is equal to n can be represented as a product of two K_d-words of smaller K_d-length which can, by the inductive hypothesis, be rewritten in R-regular form with the same content relative to K_d. Therefore, we can assume that $w = uv$ where u and v are R-regular K_d-words; we can also assume that $u > v$ in the sense of the ordering of R-regular words because in the contrary case we would have written $w = -vu$. If u is a K_d-word of K_d-length 1, then w is already R-regular because u and v are R-regular, $u > v$, and if $u = u_1u_2$ then $u_2 \leq d < v$ by definition of d-irreducibility. If the K_d-length of u is greater than 1, then it suffices to consider the case when $u = u_1u_2$ and $u_2 > v$, since in the contrary case w would already be R-regular.

So let $w = (u_1u_2)v$ where u_1, u_2, v are R-regular K_d-words, $u_1 > u_2 > v$. By relation (2),

$$w = (u_1u_2)v = (u_1v)u_2 + u_1(u_2v). \tag{3}$$

Since the lengths of the words u_1v and u_2v are greater than the length of v, rewriting u_1v and u_2v in R-regular form we obtain K_d-words that are greater than v relative to the ordering of words in R. Applying distributivity and removing words of the form uu if they appear, and using anticommutativity to make the right factor less than the left factor, we obtain an expression of w as a linear combination of words, each of which, as w itself, consists of two R-regular factors with the right factor less than the left factor but now greater than v. We do the same with each of these words as with w. Because of the finiteness of the number of words with a given content, this process will terminate after a finite number of steps; this means that we have obtained the required expression for w.

Lemma 2. *K_d-regular K_d-words are linearly independent in $\mathcal{A}$.*

For the proof of Lemma 2 it suffices to prove linear independence of K_d-regular K_d-words with the same content relative to K_d, since by Lemma 1 each K_d-regular K_d-word is a linear combination of R-regular K_d-words of the same content which are linearly independent by Theorem 1.

For K_d-words of K_d-length 1, the statement of Lemma 2 is obvious. Assume by induction that, in any free Lie algebra $\mathcal{A}_0$, for any R_0-regular word d_0, K_{d_0}-regular K_{d_0}-words of K_{d_0}-length less than n are linearly independent.

Suppose there exists a linear dependence between K_d-regular K_d-words of K_d-length n, $n > 1$, that have given content relative to K_d. Now let w be the smallest element of K_d that appears in these linearly dependent words. Subject the K_d-regular K_d-words under consideration to w-factorization, which makes sense in $\overline{\mathcal{A}_d}$ and also in $\mathcal{A}_d$ by the homomorphism $\overline{\mathcal{A}_d} \to \mathcal{A}_d$. All w-irreducible words that can appear here will have the form u or $[\cdots(uw)\cdots]w$ where $u \in K_d$, $u \neq w$. Therefore they will be R-regular, i.e., belong to the set K_w of R-regular w-irreducible words.

The elements K_w will be ordered in a different way depending on whether we consider them as R-words or as K_d-words. Thus we introduce two definitions of regular words in $\overline{\mathcal{A}_w}$ and we will distinguish $\overline{K_{wR}}$-regular $\overline{K_w}$-words and $\overline{K_{wd}}$-regular $\overline{K_w}$-words, depending on whether the ordering in K_w is induced by the ordering of the regular words of $\mathcal{A}$ or the ordering of the regular words of $\overline{\mathcal{A}_d}$. In this sense we will speak of K_{wR}-regular and K_{wd}-regular K_w-words in the subalgebra $\mathcal{A}_w$ generated by the set K_w.

In view of the fact that w by assumption occurs in each of our linearly dependent K_d-regular K_d-words, and since for w itself w-reducibility or w-irreducibility does not make sense, it follows that the K_w-length of the K_d-regular K_d-words under consideration will be less than n, and thus the assumed linear dependence is at the same time a linear dependence between K_d-regular K_w-words of K_w-length less than n. By the inductive hypothesis, K_{wR}-regular K_w-words of length less than n are linearly independent. By Corollary 1 the number of K_{wR}-regular K_w-words of a fixed content is equal to the number of K_{wd}-regular K_w-words of the same content. From the possibility of representing a K_{wR}-regular K_w-word as a linear combination of K_{wd}-regular K_w-words of the same content, and vice versa, it follows that the K_{wd}-regular K_w-words of K_w-length less than n are linearly independent.

Applying Lemma 1 to the algebra $\overline{\mathcal{A}_d}$ it is possible to express any $\overline{K_{wd}}$-regular word as a linear combination of $\overline{K_d}$-regular words of the same content relative to $\overline{K_w}$. On the other hand, it is obvious that every $\overline{K_w}$-word is a linear combination of $\overline{K_{wd}}$-regular $\overline{K_w}$-words of the same content. Passing to the homomorphic images we obtain the corresponding statement for the subalgebra $\mathcal{A}_d$.

By the inductive hypothesis, $\overline{K_{wd}}$-regular $\overline{K_{wd}}$-words of $\overline{K_w}$-length less than n are linearly independent in the algebra $\overline{\mathcal{A}_d}$; therefore the numbers of $\overline{K_{wd}}$-regular and $\overline{K_d}$-regular $\overline{K_w}$-words of $\overline{K_w}$-length less than n and the same content are equal.

An analogous statement holds also for K_w-words. Therefore the K_w-words of K_w-length less than n that are K_d-regular are linearly independent, which however contradicts the above-mentioned linear dependence of these words. This proves Lemma 2.

Lemma 3. *The set K_d is independent.*

The homomorphism $\overline{\mathcal{A}_d} \to \mathcal{A}_d$ is, by Lemma 2, an isomorphism, since only the zero element of $\overline{\mathcal{A}_d}$ is mapped to the zero element of $\mathcal{A}_d$. The existence of an isomorphism between $\mathcal{A}_d$ and the free Lie algebra $\overline{\mathcal{A}_d}$ proves Lemma 3.

Corollary 2. *In the free Lie algebra with two generators there exists a subalgebra that is a free Lie algebra with a countably infinite set of generators.*

Let a and b be the generators of the free Lie algebra. Then the countable set of words of the form ab, $(ab)b$, $[(ab)b)b]$, ... is independent since each of these words belongs to the independent set K_b of b-irreducible words. From this the desired conclusion follows.

In the free Lie algebra $\mathcal{A}$ with the set of free generators R, to each element w there corresponds uniquely a natural number $n(w)$, the *degree of the element* w. The degree of w can be defined as the greatest length of regular words in the representation of w in terms of the basis of regular words. Obviously, this does not depend on the definition of regular words. The sum of the terms in this representation of w whose length is equal to $n(w)$ will be called the *highest part* of w. The element w will be called *homogeneous* if it coincides with its highest part. In an analogous sense, we can define degree, highest part, and homogeneity relative to one of the free generators of the algebra $\mathcal{A}$.

3. Main theorem

Let $\mathcal{B}$ be an arbitrary subalgebra of the free Lie algebra $\mathcal{A}$. We will construct a finite or countably infinite increasing sequence of integers k_n $(n = 0, 1, 2, \ldots)$ and a sequence of subalgebras $\mathcal{B}_n \subset \mathcal{B}$ similarly to the way it is done in the work of A.G. Kurosh [2]: define $k_0 = 0$ and $\mathcal{B}_0 = 0$; if k_m and $\mathcal{B}_m$ are already defined for all $m = 0, 1, \ldots, n-1$, let k_n be the least degree of elements in $\mathcal{B}$ that do not belong to $\mathcal{B}_{n-1}$, and let $\mathcal{B}_n$ be the subalgebra of $\mathcal{B}$ generated by all elements whose degree does not exceed k_n.

Lemma 4. *In $\mathcal{B}$ it is possible to choose a subset $\mathcal{M}$ such that*

(1) *no element $a \in \mathcal{M}$ has its highest part in the subalgebra generated by the highest parts of the elements of $\mathcal{M} \setminus \{a\}$, and*
(2) *the subalgebra $\mathcal{B}$ is generated by the set $\mathcal{M}$.*

The set $\mathcal{K}_n$ of elements of the subalgebra $\mathcal{B}_n$ whose degree does not exceed k_n is a linear subspace and the set $\mathcal{K}'_n$ of elements of the subalgebra $\mathcal{B}_{n-1}$ whose degree does not exceed k_n is a linear subspace of $\mathcal{K}_n$.

Choose arbitrarily one representative for each coset in a basis of the linear space $\mathcal{K}_n/\mathcal{K}'_n$ and let $\mathcal{M}_n$ be this set. Now let $\mathcal{M} = \bigcup_{n\geq 1} \mathcal{M}_n$. We will prove that the set $\mathcal{M}$ satisfies the requirements of Lemma 4.

We will denote the elements of $\mathcal{M}$ by b_β and their highest parts by b'_β. Suppose that for $b_\beta \in \mathcal{M}_n$ the following equality holds:

$$b'_\beta = \sum_{\gamma\neq\beta} \alpha_\gamma b'_\gamma + \sum_{\gamma,\delta\neq\beta} \alpha_{\gamma\delta} b'_\gamma b'_\delta + \cdots + \sum_{\gamma,\delta,\ldots,\nu\neq\beta} \alpha_{\gamma\delta\ldots\nu} b'_\gamma b'_\delta \cdots b'_\nu, \tag{4}$$

where some bracket arrangement is assumed for each summand with more than two factors, and the α's with subscripts are elements of the field P.

The second and following summations on the right-hand side of equation (4) may contain factors of degree greater than the degree of b'_β. Then, when we rewrite these products in the regular form, they will either become zero or will keep the same degree. In view of the linear independence of regular words, all such terms must cancel each other, and hence we may assume that the first summation contains only the elements of the same degree as b'_β, and that the remaining elements

b' appearing on the right-hand side of equation (4) have degree strictly less than the degree of b'_β, but their products have the same degree as b'_β.

The highest part of the element

$$b_\beta - \sum_{\gamma\neq\beta} \alpha_\gamma b_\gamma - \sum_{\gamma,\delta\neq\beta} \alpha_{\gamma\delta} b_\gamma b_\delta - \cdots - \sum_{\gamma,\delta,\ldots,\nu\neq\beta} \alpha_{\gamma\delta\ldots\nu} b_\gamma b_\delta \cdots b_\nu$$

of the subalgebra $\mathcal{B}_n$, has degree less than k_n, and thus this element already belongs to the subalgebra $\mathcal{B}_{n-1}$, which leads to a contradiction with the linear independence of the cosets from which we chose the elements of $\mathcal{M}_n$. Requirement (1) for the set $\mathcal{M}$ has been proved.

To prove that requirement (2) holds, we observe that the subalgebra $\mathcal{B}_n$ is generated by the subalgebra $\mathcal{B}_{n-1}$ and the set $\mathcal{M}_n$, from which it follows by induction that the subalgebra $\mathcal{B}_n$ is generated by the set $\bigcup_{k=1}^{n} \mathcal{M}_k$ for all n. Since for each $c \in \mathcal{B}$ there exists a natural number q such that $c \in \mathcal{B}_q$, requirement (2) has been proved.

By a *nonassociative polynomial* we mean an element of the free nonassociative algebra S over the field P with a countably infinite set of free generators x_1, x_2, ... Let $\mathcal{S}$ be the free Lie algebra over the same field with free generators a_1, a_2, ..., where regular words in $\mathcal{S}$ have been defined in some way. There exists a natural homomorphism of S onto $\mathcal{S}$ that sends the polynomial $f(x_{i_1}, x_{i_2}, \ldots)$ to the element $f(a_{i_1}, a_{i_2}, \ldots)$. We will call two polynomials in S *equivalent* if their images in $\mathcal{S}$ are equal. We will call a polynomial $f(x_{i_1}, x_{i_2}, \ldots)$ *non-trivial* if its image $f(a_{i_1}, a_{i_2}, \ldots)$ is nonzero. Let $\varphi(a_{i_1}, a_{i_2}, \ldots)$ be the regular form of this image. Then the polynomial $\varphi(x_{i_1}, x_{i_2}, \ldots)$ equivalent to the polynomial $f(x_{i_1}, x_{i_2}, \ldots)$ will be called *regular*. Clearly, any part of a regular polynomial is non-trivial.

Theorem 2. *Any subalgebra $\mathcal{B}$ of a free Lie algebra $\mathcal{A}$ is free.*

Suppose we are given a free Lie algebra $\mathcal{A}$ over the field P with the set R of free generators, and a subalgebra $\mathcal{B}$. According to Lemma 4, we choose a set $\mathcal{M}$ and we will prove that it is independent.

Assume that for some finite system of elements b_1, b_2, ..., b_q in $\mathcal{M}$, there exists a non-trivial relation $F(b_1, b_2, \ldots, b_q) = 0$, i.e., $F(x_1, x_2, \ldots, x_q)$ is a non-trivial polynomial which we may take to be regular; from this we will derive a contradiction. We may assume that $n(b_i) \leq n(b_j)$ for $i < j$.

Lemma 5. *Under the above assumption, there exists a finite set $\mathcal{M}_1$ of homogeneous elements of the algebra $\mathcal{A}$ that satisfies requirement* (1) *of Lemma* 4*, and some non-trivial relation $F_1 = 0$ among the elements of $\mathcal{M}_1$.*

The regular polynomial $F(x_1, x_2, \ldots, x_q)$ can be represented as the following sum of two polynomials:

$$F(x_1, x_2, \ldots, x_q) = F_1(x_1, x_2, \ldots, x_q) + F'_2(x_1, x_2, \ldots, x_q).$$

To each term of the polynomial F under the substitution of b_i for x_i $(i = 1, 2, \ldots, q)$ there corresponds a natural number, namely the sum of the degrees relative to R

of all factors of the form b_i that occur in the given term. Then, we denote by F_1 the sum of all terms for which this sum of degrees is maximal.

Let $b_i = b'_i + b''_i$ where b'_i is the leading term of b_i $(i = 1, 2, \ldots, q)$. Then, from the relation

$$\begin{aligned} F(b_1, \ldots, b_q) &= F(b'_1 + b''_1, \ldots, b'_q + b''_q) \\ &= F(b'_1, \ldots, b'_q) + F'(b'_1, \ldots, b'_q, b''_1, \ldots, b''_q) \\ &= F_1(b'_1, \ldots, b'_q) + F'_2(b'_1, \ldots, b'_q) + F'(b'_1, \ldots, b'_q, b''_1, \ldots, b''_q) \\ &= 0, \end{aligned}$$

it follows that $F_1(b'_1, \ldots, b'_q) = 0$ by the definition of the polynomial F_1. The non-triviality of the polynomial F_1 follows from the fact that it is regular as part of the regular polynomial F. The required set $\mathcal{M}_1$ is b'_1, b'_2, $\ldots$, b'_q.

Lemma 6. *Suppose there exists a set $\mathcal{M}_1 = \{b'_i\}$ $(i = 1, 2, \ldots, q)$ and a non-trivial relation*

$$F_1(b'_1, \ldots, b'_q) = 0,$$

that satisfy the conditions of Lemma 5. Suppose that the elements of the set $\mathcal{M}'_2 = \{c_i\}$ $(i = 1, 2, \ldots, q)$ are in one-to-one correspondence with the elements of the set $\mathcal{M}_1$ and have the form $c_i = b'_i + v_i$ $(i = 1, 2, \ldots, q)$ where v_i is an element of the subalgebra generated by the elements b'_k with $k < i$, and v_i either is zero or has the same degree relative to R as b'_i. Then there exists a non-trivial relation $F_2(c_1, \ldots, c_q) = 0$ and the set $\mathcal{M}'_2$ satisfies the same conditions as the set $\mathcal{M}_1$.

First of all, let us prove that there exists a representation $b_i = c_i + v'_i$ $(i = 1, 2, \ldots, q)$ where v'_i is zero or an element of the subalgebra generated by the elements c_j $(j < i)$ whose degree is equal to the degree of v_i. We set $b'_1 = c_1$. Suppose we have found the required representation for all b'_k with $k < m$. Then, from the equality $b'_m = c_m - v_m$, after replacing all b'_j $(j < m)$ in v_m by the already found expressions, it follows that there exists the required expression for b'_m.

We separate, from the non-trivial polynomial $F_1(b'_1, \ldots, b'_q)$ which we may suppose regular, the part F_{11} that has the highest degree relative to b'_q, and then from F_{11} we separate the part F_{12} that has the highest degree relative to b'_{q-1}, and so on; finally, from $F_{1,q-1}$ we separate the part F_{1q} that has the highest degree relative to b'_1. Let us substitute the expressions we have found for b'_k into the relation $f_1 = 0$:

$$\begin{aligned} F(b'_1, \ldots, b'_q) &= F_{1q}(b'_1, \ldots, b'_q) + \overline{F}_1(b'_1, \ldots, b'_q) \\ &= F_{1q}(c_1 + v'_1, \ldots, c_q + v'_q) + \overline{F}_1(c_1 + v'_1, \ldots, c_q + v'_q) \\ &= F_{1q}(c_1, \ldots, c_q) + \varphi(c_1, \ldots, c_q) \\ &= F_2(c_1, \ldots, c_q) \\ &= 0. \end{aligned}$$

The polynomial F_{1q} is non-trivial since it is regular; and obviously it does not have terms of the same content relative to $\mathcal{M}'_2$ as any term of the polynomial φ. It follows that the polynomial F_2 is non-trivial.

Now we prove that the element $c_j \in \mathcal{M}'_2$ does not belong to the subalgebra generated by the set $\mathcal{M}'_2 \setminus c_j$. Assuming the contrary, we obtain the equation

$$c_j = \sum_{k_1 \neq j} \alpha_{k_1} c_{k_1} + \sum_{k_1, k_2 \neq j} \alpha_{k_1 k_2} c_{k_1} c_{k_2} + \cdots + \sum_{k_1, \ldots, k_n \neq j} \alpha_{k_1 \ldots k_n} c_{k_1} c_{k_2} \cdots c_{k_n},$$

where we assume for each product with $n > 2$ there is some arrangement of brackets.

Repeating verbatim what was said above about equation (4), we will assume that the element c_j and all elements c_{k_1} that occur in the first summation have the same degree, and all factors in the second and following summations on the right-hand side have strictly smaller degrees. Let c_ℓ have the greatest index among the elements c_j, c_{k_1}. Then, replacing all c_i by their expressions in terms of b'_i, we obtain that b'_i belongs to the subalgebra generated by the other elements of the set $\mathcal{M}_1$, which contradicts Lemma 5. This completes the proof.

Lemma 7. *Under the conditions of Lemma* 6*, there exists a set* $\mathcal{M}_2$ *of elements which satisfy requirement* (1) *of Lemma* 4*, are homogeneous in each element of* R*, and satisfy some non-trivial relation.*

We choose arbitrarily some generator $a_\alpha \in R$ from among the elements of the set $\mathcal{M}_1$. Each element $b'_i \in \mathcal{M}_1$ can be written in the form

$$b'_i = b_{i1} + b_{i2} + \cdots + b_{in_i},$$

where b_{ik} is the part of the element b'_i that has degree k relative to a_α ($i = 1, 2, \ldots, q$; $k = 0, 1, \ldots, n_i$). If b_{2n_2} belongs to the subalgebra generated by the element b_{1n_1}, i.e., $b_{2n_2} = \gamma b_{1n_1}$, $\gamma \in P$, then we replace the element b'_2 in $\mathcal{M}_1$ by the element $b'_2 - \gamma b'_1$ and denote the resulting set $\mathcal{M}_{12}$, using for symmetry the notation $\mathcal{M}_{11} = \mathcal{M}_1$; otherwise, we set $\mathcal{M}_{12} = \mathcal{M}_{11}$. Suppose the sets $\mathcal{M}_{1r}$ ($r = 1, 2, \ldots, \ell$; $\ell < q$) have already been constructed. If, in the set $\mathcal{M}_{1\ell}$ the element $b_{\ell+1, n_{\ell+1}}$, that is a part of the element $b'_{\ell+1}$, does not belong to the subalgebra generated by the highest parts, relative to a_α, of the preceding elements of $\mathcal{M}_{1\ell}$, then we will set $\mathcal{M}_{1,\ell+1} = \mathcal{M}_{1\ell}$. If, on the other hand, $b_{\ell+1, n_{\ell+1}}$ belongs to that subalgebra, then we replace the element $b'_{\ell+1}$ by the element $b'_{\ell+1} - v_{\ell+1}$ where $v_{\ell+1}$ is an element of the subalgebra generated by the elements of $\mathcal{M}_{1\ell}$ preceding the element $b'_{\ell+1}$, whose highest part relative to a_α is the same as for $b'_{\ell+1}$. We denote the resulting set by $\mathcal{M}_{1,\ell+1}$. We may assume that the highest part relative to a_α of the element $b'_{\ell+1} - v_{\ell+1}$ does not belong to the subalgebra generated by the highest parts relative to a_α of the elements of $\mathcal{M}_{1\ell}$ that precede $b'_{\ell+1}$, because this can be easily achieved by an appropriate choice of $v_{\ell+1}$. Finally, we will obtain a set $\mathcal{M}_{1q} = \mathcal{M}'$ such that the highest part of each element relative to a_α does not belong to the subalgebra generated by the highest parts (relative to a_α) of the preceding elements. In fact, the highest part relative to a_α of each element of

$\mathcal{M}'$ does not belong to the subalgebra generated by the similar parts of the other elements, since assuming the contrary immediately leads to a contradiction as in the proof of Lemma 6.

Applying Lemma 6 at each step of the above construction we obtain that no element of the set $\mathcal{M}'$ belongs to the subalgebra generated by the other elements, and we also obtain a certain non-trivial relation $F'' = 0$ for the elements of this set. We write each element $c'_k \in \mathcal{M}$ in the form $c'_k = c'_{k1} + c'_{k2}$ where c'_{k1} is the highest part of the element c'_k relative to a_α, and separate in each polynomial F'' the highest part F''_1 relative to a_α. Then we will have

$$\begin{aligned} F''(c'_1, \dots, c'_q) &= F''_1(c'_1, \dots, c'_q) + F''_2(c'_1, \dots, c'_q) \\ &= F''_1(c'_{11}, \dots, c'_{q1}) + \varphi''(c'_{11}, \dots, c'_{q1}, c'_{12}, \dots, c'_{q2}) \\ &= 0. \end{aligned}$$

In view of the fact that each term of $F''_1(c'_{11}, \dots, c'_{q1})$ has the highest degree in a_α, these terms cannot cancel with the terms of the polynomial φ''; moreover, F''_1 is non-trivial as a part of a regular polynomial.

Thus we have obtained the set $\mathcal{M}'' = \{c'_{i1}\}$ of elements which are homogeneous in a_α, and a non-trivial relation $F''_1 = 0$ satisfied by these elements. Enumerating one by one all the generators that occur in the elements of the set $\mathcal{M}_1$ we find obtain the desired set $\mathcal{M}_2$ and some non-trivial relation for its elements.

Lemmas 5, 6 and 7 allow us to assume that the set $\mathcal{M}_1 = \{b'_i\}$ $(i = 1, 2, \dots, q)$ consists of elements that are homogeneous in each generator and satisfy requirement (1) of Lemma 4.

If $\mathcal{M}_1$ contains elements of degree 1, then by homogeneity they must have the form γa_μ where $\gamma \in P$, $a_\mu \in R$. Therefore we can assume that such elements have the form $a_\mu \in R$, i.e., they are simply free generators.

The ordered q-tuple $(\nu_1; \nu_2; \dots; \nu_q)$ of natural numbers, where ν_k is the degree of b'_k, will be called the *height* of the set $\mathcal{M}_1$. We order the set of all possible heights lexicographically, and assume that for the sets with smaller height there are no non-trivial relations if those sets satisfy requirement (1) of Lemma 4. This assumption is justified by considering the sets of height $\varepsilon = (1; 1; \dots; 1)$ that consist only of free generators.

Assume $(\nu_1; \nu_2; \dots; \nu_q) > (1; 1; \dots; 1)$; this means that some $\nu_k > 1$. Then, in the element b'_k, we can find a generator a_λ that is not one of the b'_m, since otherwise requirement (1) of Lemma 4 would be violated.

Let us reorder the generators to make a_λ the smallest if this is not already the case, and rewrite all b'_i in regular form relative to some new definition of regular words that depends on this order. After this, we subject the words in the elements of the set $\mathcal{M}_1$ to a_λ-factorization. By Lemma 3, a_λ-irreducible words form an independent set; thus all our considerations can be transferred to the free Lie algebra $\mathcal{A}_{a_\lambda}$ generated by the set K_{a_λ} of a_λ-irreducible words. Since a_λ is the smallest of the generators, all other generators will be a_λ-irreducible; therefore, the degree of each word relative to the new system of free generators of $\mathcal{A}_{a_\lambda}$

will be equal to the difference between its degree relative to the old system of free generators of the algebra $\mathcal{A}$ and its degree relative to a_λ. It follows that the elements of $\mathcal{M}_1$ which are homogeneous in each of the old generators will also be homogeneous relative to the new systems of generators, but the set $\mathcal{M}_1$ itself will have a smaller height. Obviously, the height will not become zero and also the set will retain a non-trivial relation. This contradicts the inductive hypothesis and consequently proves the theorem.

The theorem on subalgebras of free Lie algebras proved above cannot be transferred to rings, since for example the subring, of the free Lie ring with generators a and b, generated by the elements $2a$, b, ab is not free because the generators $2a$, b, ab satisfy the relation

$$(2a)b - 2(ab) = 0,$$

and as can be easily seen there is no other system of generators for this subring that would not satisfy a non-trivial relation.

References

[1] M. Hall, *A basis for free Lie rings and higher commutators in free groups*, Proc. Amer. Math. Soc. 1 (1950) 575–581.

[2] A.G. Kurosh, *Nonassociative free algebras and free products of algebras*, Mat. Sbornik N.S. 20 (1947) 239–262.

[3] A.I. Malcev, *On algebras defined by identical relations*, Mat. Sbornik N.S. 26 (1950) 19–33.

On the Representation of Lie Rings in Associative Rings

A.I. Shirshov

V.M. Kurochkin [1] has formulated the following theorem: *Every* Σ*-operator Lie ring* L *has a faithful representation in an associative* Σ*-operator ring* $\mathcal{A}$*, where* Σ *is an arbitrary domain of operators for the ring* L*.* In a subsequent note [2], V.M. Kurochkin pointed out the insufficient rigor of the proof he proposed for this theorem.

In the present paper, an example is constructed which demonstrates that, for the above formulation, the theorem is not valid.

Consider a linear space A with basis elements a_i $(i = 1, 2, \ldots, 13)$ over the field $GF(2)$. We make the space A into a ring by defining multiplication according to the following formulas:

$$a_1a_2 = a_2a_1 = a_{11}; \qquad a_1a_3 = a_3a_1 = a_{13}; \qquad a_2a_3 = a_3a_2 = a_{12};$$
$$a_1a_8 = a_8a_1 = a_2a_6 = a_6a_2 = a_3a_5 = a_5a_3 = a_{10};$$

and in all remaining cases $a_ia_j = 0$. Since the following equations hold identically as a consequence of the multiplication table,

$$x^2 = 0; \qquad (xy)z = 0;$$

the ring A is a Lie ring.

Now let Σ be the linear space over the same field with basis elements e_i $(i = 0, 1, 2, 3)$. Define a multiplication in Σ as follows:

$$e_ie_0 = e_0e_i = e_i,\ i = 0, 1, 2, 3; \qquad e_ie_j = 0,\ i, j \neq 0.$$

Define an action of the elements of Σ on the elements of A in the following way:

$e_0a_i = a_i$, $i = 1, 2, \ldots, 13$;

$e_1a_1 = a_4$; $e_1a_2 = a_5$; $e_1a_3 = a_6$; $e_1a_{12} = a_{10}$; $e_1a_k = 0$, $3 < k < 12$, $k = 13$;

$e_2a_1 = a_5$; $e_2a_2 = a_7$; $e_2a_3 = a_8$; $e_2a_{13} = a_{10}$; $e_2a_t = 0$, $3 < t < 13$;

Uspekhi Mat. Nauk N.S. 8, (1953), no. 5 (57), 173–175.

$$e_3a_1 = a_6;\ e_3a_2 = a_8;\ e_3a_3 = a_9;\ e_3a_{11} = a_{10};\ e_3a_\ell = 0,\ 3 < \ell < 11,\ \ell = 12, 13.$$

By distributivity we define the action of any element of Σ on any element of A.

In this way, the ring A becomes a Σ-operator Lie ring. Indeed, from the displayed table it follows that $(e_ie_j)a_k = e_i(e_ja_k)$ for $i \neq 0$, $j \neq 0$. Obviously, for $i = 0$ and $j = 0$ this equation also holds. Therefore, $(\sigma_1\sigma_2)b = \sigma_1(\sigma_2 b)$ where $\sigma_1, \sigma_2 \in \Sigma$ and $b \in A$. Further, if $a_ia_j = a_k$ then $k = 10, 11, 12, 13$. Suppose $k = 10$; then the equation $(e_ra_i)a_j = a_i(e_ra_j) = e_ra_{10}$ can be easily verified directly. For $k = 11$, it is sufficient to consider the equation $a_1a_2 = a_{11}$. In this case also, $(e_ra_1)a_2 = a_1(e_ra_2) = e_ra_{11}$, where a nonzero result is possible only for $r = 0$ and $r = 3$. The situation is similar for $k = 12$ and $k = 13$. Now, if $a_ia_j = 0$ then, for example, for $i = 1$ we will have $j = 1, 4, \ldots, 7, 9, 13$ and $(e_ra_1)a_j = a_1(e_ra_j) = 0$. Similarly for $i = 2$ and $i = 3$. From this it easily follows that $(\sigma b_1)b_2 = b_1(\sigma b_2) = \sigma(b_1b_2)$ where $\sigma \in \Sigma$, $b_1, b_2 \in A$, which completes the proof of the fact that A is a Σ-operator ring.

We now show that, in no matter which Σ-operator Lie ring $\mathcal{A}$ we embed the ring A, the element a_{10} will always be an absolute zero-divisor[1] of $\mathcal{A}$.

Indeed, let x be an arbitrary element of $\mathcal{A}$. Then,

$$\begin{aligned}
0 &= [x(e_1a_2 + a_5)]a_3 + (xa_1)(e_2a_3 + a_8) + [x(a_6 + e_1a_3)]a_2 \\
&= [x(e_1a_2)]a_3 + (xa_5)a_3 + (xa_1)(e_2a_3) + (xa_1)a_8 + (xa_6)a_2 + [x(e_1a_3)]a_2 \\
&= (xa_3)(e_1a_2) + x[(e_1a_2)a_3] + [x(e_2a_1)]a_3 + [x(e_2a_1)]a_3 + (xa_1)(e_3a_2) \\
&\quad + [x(e_3a_1)]a_2 + (xa_3)(e_1a_2) \\
&= x[(e_1a_2)a_3] \\
&= xa_{10},
\end{aligned}$$

where we have used the Jacobi identity, the fact that $\mathcal{A}$ is a Σ-operator ring, and the fact that all elements of the additive group of A have order 2.

Suppose there exists an associative Σ-operator ring $\mathcal{B}$ whose commutator Lie ring $\mathcal{B}^-$ contains A as a Σ-admissible subring. Then it is obvious that to the element a_{10} there corresponds some element of the center of $\mathcal{B}$, such that for any embedding of the ring $\mathcal{B}$ into any other associative Σ-operator ring $\overline{\mathcal{B}}$, this element is mapped to the center of $\overline{\mathcal{B}}$.

We now obtain a contradiction from the following result:

Lemma 1. *Any associative Σ-operator ring $\mathcal{B}$, such that $e_0\ell = \ell$ for any $\ell \in \mathcal{B}$, can be embedded into some associative Σ-operator ring $\overline{\mathcal{B}}$ such that the intersection of the center Z of $\overline{\mathcal{B}}$ with $\mathcal{B}$ equals zero*[2].

For the proof it suffices to consider the case in which Σ is a commutative associative ring with identity element e_0 acting on $\mathcal{B}$ as the identity automorphism.

[1] That is, a central element. [Translators]

[2] To make the condition of the lemma hold in our case, it suffices to consider, instead of the ring $\mathcal{B}$, the Σ-admissible subring generated by all elements of A.

Consider the collection $\overline{\mathcal{B}}$ of symbols of the form $(\sigma_i, b_{i1}, b_{i2}, b_{i3})$ where $\sigma_i \in \Sigma$, and $b_{ik} \in \mathcal{B}$, $k = 1, 2, 3$. We will regard two symbols $(\sigma_i, b_{i1}, b_{i2}, b_{i3})$ and $(\sigma_j, b_{j1}, b_{j2}, b_{j3})$ as equal if and only if $\sigma_i = \sigma_j$, $b_{ik} = b_{jk}$, $k = 1, 2, 3$.

We make the collection $\overline{\mathcal{B}}$ into a Σ-operator ring by defining addition, multiplication, and the action of $\sigma \in \Sigma$ on an element $\overline{b} \in \overline{\mathcal{B}}$ by the following formulas:

$$(\sigma_i, b_{i1}, b_{i2}, b_{i3}) + (\sigma_j, b_{j1}, b_{j2}, b_{j3}) = (\sigma_i + \sigma_j, b_{i1} + b_{j1}, b_{i2} + b_{j2}, b_{i3} + b_{j3});$$

$$(\sigma_i, b_{i1}, b_{i2}, b_{i3}) \cdot (\sigma_j, b_{j1}, b_{j2}, b_{j3}) = (0, \sigma_j b_{i3} + b_{i3} b_{j1}, \sigma_i b_{j3} + b_{i2} b_{j3}, b_{i3} b_{j3});$$

$$\sigma(\sigma_i, b_{i1}, b_{i2}, b_{i3}) = (\sigma\sigma_i, \sigma b_{i1}, \sigma b_{i2}, \sigma b_{i3}).$$

It is easy to verify that all the axioms of a Σ-operator ring are satisfied.

The ring $\overline{\mathcal{B}}$ is an associative ring, since

$$\begin{aligned}
&[(\sigma_i, b_{i1}, b_{i2}, b_{i3}) \cdot (\sigma_j, b_{j1}, b_{j2}, b_{j3})] \cdot (\sigma_k, b_{k1}, b_{k2}, b_{k3}) \\
&= (\sigma_i, b_{i1}, b_{i2}, b_{i3}) \cdot [(\sigma_j, b_{j1}, b_{j2}, b_{j3}) \cdot (\sigma_k, b_{k1}, b_{k2}, b_{k3})] \\
&= (0, \sigma_k b_{i3} b_{j3} + b_{i3} b_{j3} b_{k1}, \sigma_i b_{j3} b_{k3} + b_{i2} b_{j3} b_{k3}, b_{i3} b_{j3} b_{k3}),
\end{aligned}$$

and it contains a subring of symbols $(0, 0, 0, b_{i3})$ that is isomorphic to the ring $\mathcal{B}$.

On the other hand, for $b_{i3} \neq 0$, from the equations

$$(e_0, 0, 0, 0) \cdot (0, 0, 0, b_{i3}) = (0, 0, b_{i3}, 0), \text{ and}$$

$$(0, 0, 0, b_{i3}) \cdot (e_0, 0, 0, 0) = (0, b_{i3}, 0, 0),$$

it follows that

$$(e_0, 0, 0, 0) \cdot (0, 0, 0, b_{i3}) \neq (0, 0, 0, b_{i3}) \cdot (e_0, 0, 0, 0),$$

which completes the proof of the lemma.

From the contradiction just obtained, it follows that the Σ-operator ring A cannot be faithfully represented in any associative Σ-operator ring. This example also shows that Ado's theorem cannot be generalized to rings over an arbitrary ring of operators.

It would be interesting to find necessary and sufficient conditions for the existence of a faithful representation of a given Σ-operator Lie ring R. Lazard [3] proved that if Σ is a principal ideal ring, then such a representation exists for any R. One can also prove the following theorem: *If no element $\sigma \in \Sigma$, $\sigma \neq 0$, annihilates an absolute zero-divisor of R, then a faithful representation always exists.*

References

[1] V.M. Kurochkin, *The representation of Lie rings by associative rings*, Mat. Sbornik N.S. 28 (1951) 467–472.

[2] V.M. Kurochkin, *Correction to the paper "The representation of Lie rings by associative rings"*, Mat. Sbornik N.S. 30 (1952) 463.

[3] M. Lazard, *Sur les algèbres enveloppantes universelles de certaines algèbres de Lie*, C. R. Acad. Sci. Paris 234 (1952) 788–791.

Subalgebras of Free Commutative and Free Anticommutative Algebras

A.I. Shirshov

1. It is known (see A.G. Kurosh [2]) that any subalgebra of the free nonassociative algebra is free. It is natural to ask the corresponding question for relatively free algebras (see A.I. Malcev [3]), of course restricting oneself to the most important classes of algebras.

In the work of the present author [4], it is proved that every subalgebra of a free Lie algebra is also free. In the same paper it is pointed out that the analogous theorem is not valid for free associative, alternative, right- or left-alternative, or Jordan algebras, and also for flexible algebras, and power-associative or power-commutative algebras. It is easy to see that this theorem is valid for free nilpotent algebras of class 1, and not valid for free nilpotent algebras of class k, $k > 1$.

Among the most important classes of algebras, there remain only the commutative and anticommutative algebras. In the present paper, it is proved that for the free algebras of these two classes, the corresponding problem has a positive solution. For brevity and convenience of exposition, we will call commutative algebras C-algebras and anticommutative algebras AC-algebras.

Analogously to the definitions of A.I. Malcev [3] we call an algebra $\mathcal{A}$ over a field P a *free ε-algebra* where $\varepsilon = C$ or $\varepsilon = AC$ if it is defined by some set R of generators and by the identical relation

$$xy + \delta yx = 0, \tag{1}$$

where $\delta = -1$ for $\varepsilon = C$ and $\delta = +1$ for $\varepsilon = AC$, and also for $\varepsilon = AC$ we will assume[1] that the characteristic of P is not 2 since this case will be included in the case $\varepsilon = C$.

In the proof of Theorem 1 below, we use a method that is similar to Hall's method in [1], and in the proof of Theorem 2 below, we partially use the methods

Mat. Sbornik N.S. 34 (76), (1954), no. 1, 81–88.

[1]According to current terminology, an anticommutative algebra in characteristic 2 should also satisfy $x^2 = 0$ for all x, and this case is not included in the author's considerations. [Translators]

of A.G. Kurosh [2] and of the present author [4]. The present work can be studied independently of the above-mentioned papers, although it can be regarded as a sequel to the present author's work [4].

The present work was carried out under the supervision of A.G. Kurosh, to whom the author expresses his deep gratitude.

2. Let $R = \{a_\alpha\}$ be some set of symbols where α takes values in some non-empty set of indices.

Consider nonassociative words of various lengths formed from these symbols, in the sense of the definitions given in the work of A.G. Kurosh [2]; we will call them *R-words* or simply *words*.

Definition. Words of length 1 will be called *ε-regular* and ordered arbitrarily. Assuming that ε-regular words of length less than n, $n > 1$, have been already defined and ordered in such a way that words of smaller length precede words of greater length, a word w of length n will be called *ε-regular* if

1) $w = uv$ where u and v are ε-regular words;
2) $u \geq v$ for $\varepsilon = C$ and $u > v$ for $\varepsilon = AC$.

We order arbitrarily the ε-regular words of length n defined in this way, and declare them to be greater than regular words of smaller length.

The symbols $<, >, \leq, \geq$ as applied to ε-regular words in the above definition, as well as in the remainder of this paper, will be understood in the sense of the ordering of these words.

Theorem 1. *The collection of all ε-regular words for $\varepsilon = C, AC$ forms a basis of the free ε-algebra $\mathcal{A}$ with the system of free generators R.*

We demonstrate a method that allows us to assign uniquely, to each word w of the free ε-algebra $\mathcal{A}$, some element w^* of the same algebra such that

$$w^* = w \text{ in the algebra } \mathcal{A}, \tag{2}$$

where w^* is either an ε-regular word with coefficient $+1$ or -1, or 0. For words of length 1, we set $w^* = w$.

Suppose such a method is already defined for words of length less than n, and let w be a word of length n, $n > 1$. Then $w = uv$. We set

$$w^* = u^*v^*, \text{ if } u^* \geq v^* \text{ in case } \varepsilon = C, \text{ or } u^* > v^* \text{ in case } \varepsilon = AC;$$
$$w^* = 0 \text{ if } u^* = v^* \text{ in case } \varepsilon = AC;$$
$$w^* = -\delta v^*u^* \text{ if } u^* < v^*.$$

Obviously, all the conditions imposed on w^* are satisfied.

Each element $a \in \mathcal{A}$ has the form

$$a = \sum_{i=1}^{k} \alpha_i w_i, \tag{3}$$

where α_i are elements of the base field, and w_i are some words, not necessarily distinct. Clearly, the element a can be written in the form

$$a = \sum_{i=1}^{k} \alpha_i w_i^*, \tag{4}$$

from which it follows that every element of the free ε-algebra can be represented as a linear combination of ε-regular words.

The zero element of the algebra $\mathcal{A}$ only admits a representation of the form

$$0 = \sum_i \alpha_i a_{i1} a_{i2} \cdots a_{in_i} [c_i d_i + \delta d_i c_i] b_{i1} b_{i2} \cdots b_{im_i},$$

where $\alpha \in P$, and a, b, c, d are some words, and an appropriate arrangement of parentheses is assumed. One immediately sees that the ε-regular expression of the right-hand side obtained after applying distributivity and replacing each of the resulting words by the corresponding starred word, gives zero. Since for an ε-regular word w we have $w^* = w$, it follows that there do not exist two distinct ε-regular expressions for the same element, which is equivalent to the linear independence of ε-regular words. The theorem is proved.

The unique expression of an element a as a linear combination of ε-regular words will be denoted by a^*.

In the free ε-algebra $\mathcal{A}$, to each element a there corresponds a natural number $n(a)$, the *degree* of a, defined as the greatest length of the ε-regular words occurring in a^*.

The sum of the terms of the element a^*, i.e., ε-regular words with coefficients in P, whose degree is equal to $n(a)$, will be called the *highest part* of the element a.

3. The purpose of the present work is the proof of the following theorem.

Theorem 2. *Every subalgebra $\mathcal{B}$ of a free ε-algebra $\mathcal{A}$ (where $\varepsilon = C$ or $\varepsilon = AC$) is also free.*

Thanks to the existence of the concept of degree, we can use the method of A.G. Kurosh [2] to construct, for each subalgebra $\mathcal{B}$ of the free ε-algebra $\mathcal{A}$, a finite or countably infinite sequence of integers k_n and subalgebras $\mathcal{B}_n$ $(n = 1, 2, \ldots)$, where $k_0 = 0$, $\mathcal{B}_0 = 0$, k_n is the smallest degree of elements of the subalgebra $\mathcal{B}$ which have not been included in $\mathcal{B}_{n-1}$, and $\mathcal{B}_n$ is the subalgebra generated in $\mathcal{B}$ by the elements whose degree does not exceed k_n.

The set $\mathcal{K}_n$ of elements of $\mathcal{B}_n$ whose degree does not exceed k_n is a linear subspace, and the set $\mathcal{K}'_n$ of elements of $\mathcal{B}_{n-1}$ whose degree does not exceed k_n is a subspace of $\mathcal{K}_n$. We arbitrarily choose one representative from each coset in a basis of the linear space $\mathcal{K}_n/\mathcal{K}'_n$ and denote the resulting set by $\mathcal{M}_n$. Now let $\mathcal{M} = \bigcup_{n\geq 1} \mathcal{M}_n$. We prove that the set $\mathcal{M}$ has the following properties:

A. *The highest part of each element a, $a \in \mathcal{M}$, does not belong to the subalgebra generated by the highest parts of the elements of the set $\mathcal{M} \setminus \{a\}$;*

B. *The subalgebra $\mathcal{B}$ is generated by the set $\mathcal{M}$.*

Suppose, contrary to property A, that the highest part $\overline{a}$ of some element $a \in \mathcal{M}$ belongs to the subalgebra generated by the highest parts of the elements of the set $\mathcal{M} \setminus \{a\}$, i.e.,

$$\overline{a} = \sum_i \alpha_i \overline{a_i} + \sum_{i,j} \alpha_{ij} \overline{b_i}\,\overline{b_j} + \cdots + \sum_{i,j,\ldots,k} \alpha_{ij\ldots k} \overline{c_i}\,\overline{c_j} \cdots \overline{c_k},$$

where $\alpha \in P$, and parentheses are placed appropriately. Then all $\overline{b}, \ldots, \overline{c}$ that occur in the second and following summations can obviously be assumed to have degree less than the degree of $\overline{a}$, and all $\overline{a_i}$ that occur in the first summation can be assumed to be distinct from $\overline{a}$ and to have degree equal to the degree of $\overline{a}$, i.e., all corresponding elements a, a_i must belong to the same set $\mathcal{M}_n$. It follows that for the cosets A, A_i represented by the elements a, a_i there exists a linear dependence relation $A - \sum_i \alpha_i A_i = 0$, which contradicts the choice of these cosets. Thus, property A has been proved.

To prove property B, it suffices to observe that the subalgebra $\mathcal{B}_n$ is generated by the set $\bigcup_{k=1}^{n} \mathcal{M}_k$.

4. We will call any element of the free nonassociative algebra S, with the set $X = \{x_1, x_2, \ldots\}$ of free generators, a *nonassociative polynomial*. Let $\mathcal{S}$ be the free ε-algebra with generators $a_1, a_2, \ldots$ over the same field P. There exists a natural homomorphism of S onto $\mathcal{S}$ that sends the polynomial $f(x_{i_1}, x_{i_2}, \ldots)$ to the element $f(a_{i_1}, a_{i_2}, \ldots)$. We will call two polynomials S *equivalent* if their images in $\mathcal{S}$ are equal. We will call a polynomial $f(x_{i_1}, x_{i_2}, \ldots)$ *non-trivial* if its image is nonzero. If in $\mathcal{S}$, regular words are defined and $\varphi(a_{i_1}, a_{i_2}, \ldots)$ is the ε-regular form of the element $f(a_{i_1}, a_{i_2}, \ldots)$, then the polynomials $f(x_{i_1}, x_{i_2}, \ldots)$ and $\varphi(x_{i_1}, x_{i_2}, \ldots)$ are equivalent. In this case, we will call the polynomial $\varphi(x_{i_1}, x_{i_2}, \ldots)$ *ε-regular*, and then obviously any part of φ will also be ε-regular.

We now assume that there exists some non-trivial relation $f(b_1, b_2, \ldots, b_q) = 0$ for the elements of $\mathcal{M}$, i.e., $f(x_1, x_2, \ldots, x_q)$ is a non-trivial polynomial which can in fact be taken to be regular, and we derive a contradiction from this assumption.

Lemma. *If, in the free ε-algebra $\mathcal{A}$ there exists a finite set of elements b_i (where $i = 1, 2, \ldots, q$, and $n(b_i) \geq n(b_j)$ for $i > j$) which satisfy property A and some non-trivial relation $f(b_1, b_2, \ldots, b_q) = 0$, then the elements of the finite set $\mathcal{M} = \{c_i\}$ $(i = 1, 2, \ldots, q)$ that have the form $c_i = b_i + w_i$, where w_i is an element of the subalgebra generated by the elements b_k $(k < i)$ and either $w_i = 0$ or $n(w_i) = n(b_i)$, also satisfy property A and some non-trivial relation $f'(c_1, c_2, \ldots, c_q) = 0$.*

We will first prove that there exist expressions

$$b_i = c_i + w_i' \ (i = 1, 2, \ldots, q)$$

where w_i' is an element (possibly zero) of the subalgebra generated by the elements c_k $(k < i)$. Indeed, $b_1 = c_1$. Assuming that for all $i < k$ the desired expression has been found, we can replace, in the equation

$$b_k = c_k - w_k,$$

all b_j $(j < k)$ that occur in w_k by the already found expressions in terms of c_i, after which we obtain the desired expression for b_k.

We will prove that property A holds for the elements of the set $\overline{\mathcal{M}}$. If the highest part $\overline{c_j}$ of some element $c_j \in \mathcal{M}$ belongs to the subalgebra generated by the highest parts of the elements of $\mathcal{M} \setminus \{c_j\}$, then analogously to what was done in the proof of property A for the set $\mathcal{M}$, we may assume that $\overline{c_j}$ belongs to the subalgebra generated by $\overline{c_k}$, $k < j$. But then, using the obvious equations

$$\overline{c_i} = \overline{b_i} + \overline{w_i}\ (i = 1, 2, \ldots, q),$$

where $\overline{w_i}$ is the highest part of w_i, to replace all $\overline{c_i}$ by their expressions in terms of $\overline{b_i}$ $(i = 1, 2, \ldots, q)$, we obtain an expression of the element $\overline{b_j}$ in terms of $\overline{b_\ell}$, $\ell < j$, which contradicts the assumption.

Now we prove that the elements of the set $\overline{\mathcal{M}}$ satisfy some non-trivial relation. Separate, in the polynomial $f(x_1, x_2, \ldots, x_q)$, the highest part relative to x_q, i.e., the collection of the terms that contain the factor x_q the maximal number of times. Denote this part by f_q. Now separate in f_q the highest part f_{q-1} relative to x_{q-1}, and so on, and finally separate in f_2 the highest part f_1 relative to x_1. After this we have:

$$\begin{aligned} f(b_1, b_2, \ldots, b_q) &= f_1(b_1, b_2, \ldots, b_q) + f_1'(b_1, b_2, \ldots, b_q) \\ &= f_1(c_1 + w_1', \ldots, c_q + w_q') + f_1'(c_1 + w_1', \ldots, c_q + w_q') \\ &= f_1(c_1, \ldots, c_q) + \varphi(c_1, \ldots, c_q) \\ &= f'(c_1, \ldots, c_q) \\ &= 0. \end{aligned}$$

The polynomial $f'(c_1, \ldots, c_q)$ cannot be trivial because the polynomials f_1 and φ do not have terms of the same content and f_1 is a non-trivial polynomial. This completes the proof of the lemma.

Based on the lemma, one can easily prove that the assumption of the existence of a non-trivial relation among the elements of $\mathcal{M}$ implies the existence of a finite set $\mathcal{N}$ of elements, that satisfy property A and some non-trivial relation, such that in each of them any term included in the highest part is not the product of *leading terms* (relative to the fixed ordering of ε-regular words) of the other elements of the set.

To prove this statement, we enumerate the finite subset of elements of $\mathcal{M}$ that occur in the non-trivial relation $f = 0$ and denote the resulting finite set by $\mathcal{M}_1$ where $\mathcal{M}_1 = \{b_i\}$ $(i = 1, 2, \ldots, q)$. Without loss of generality we assume that for $i > j$ either $n(b_i) > n(b_j)$ or $\overline{b_i} \geq \overline{b_j}$ where $\overline{b_i}$, $\overline{b_j}$ are the ε-regular words of the leading terms of the elements b_i, b_j. We will show that using the lemma we can obtain $\overline{b_i} > \overline{b_j}$ for $i > j$. Separate in $\mathcal{M}_1$ the subset $\mathcal{M}_{1k}$ of elements of degree k. Consider in $\mathcal{M}_{1k}$ the subset $\overline{\mathcal{M}_{1k}}$ of elements with the greatest leading terms (up to a coefficient from the field P). Let

$$\overline{\mathcal{M}_{1k}} = \{b_i\}\ (i = i_k, i_k + 1, \ldots, i_k + q_k).$$

Replacing in $\mathcal{M}_{1k}$ the subset $\overline{\mathcal{M}_{1k}}$ by the set of elements

$$b_{i_k},\ b_{i_k+1}-\alpha_1 b_{i_k},\ b_{i_k+2}-\alpha_2 b_{i_k},\ \ldots,\ b_{i_k+q_k}-\alpha_{q_k} b_{i_k},$$

where α_s $(s=1,2,\ldots,q_k)$ are the elements of the field P chosen such that in the differences above the leading terms cancel, we obtain that in the set $\mathcal{M}_{1k}$ there will be only one element with leading term $\overline{b_{i_k}}$ and the leading terms of all other elements will be less than $\overline{b_{i_k}}$. Doing the same with the set $\{b_{i_1+s}-\alpha_s b_{i_k}\}$ $(s=1,2,\ldots,q_k)$ and so on, we transform the set $\mathcal{M}_{1k}$ into a set in which all leading terms are distinct. We perform the same transformations for all possible k. Obviously, these transformations conform to the requirements of the lemma.

If it now turns out that some term $\overline{\overline{w}}$ of the highest part of some element w of the resulting set can be represented as a product of leading terms of other elements of the set, then obviously the latter terms will not have greater degree. Therefore we can eliminate the term $\overline{\overline{w}}$ in w by subtracting from w the product of the corresponding elements with the appropriate arrangement of parentheses. We assume by induction that the elements of the set under consideration that precede the element w are such that the terms of their highest parts can no longer be represented as products of leading terms of other elements. It is clear that, even if after performing the subtraction we obtain new terms that can be represented as products of leading terms of other elements, then the number of such factors in such terms is strictly less than the corresponding number for the term $\overline{\overline{w}}$. The proof can now be completed by a straightforward induction.

Let us now prove that the properties satisfied by the set $\mathcal{N}$ are contradictory. Indeed, let $\overline{f}=0$ be a non-trivial relation satisfied by the elements of the set $\mathcal{N}$, and let

$$e=\alpha b_{i_1} b_{i_2}\cdots b_{i_s},\ \alpha\in P,\ b_{i_k}\in\mathcal{N}\ (k=1,2,\ldots,s),$$

where parentheses are arranged in a certain way, be one of the terms of the regular polynomial $\overline{f}$; this term is chosen from among the terms for which the number $n=\sum_{k=1}^{s} n(b_{i_k})$ is maximal, in such a way that the number s is maximal. We show that, when the word

$$\overline{e}=\overline{b_{i_1}}\,\overline{b_{i_2}}\cdots\overline{b_{i_s}},$$

where $\overline{b_{i_k}}$ is the leading term of the element b_{i_k} and parentheses are arranged in the same way as before, is rewritten in ε-regular form $\overline{e}^*$, there will be no such term among the other ε-regular words obtained by representing the left-hand side of the non-trivial relation $\overline{f}=0$ as a linear combination of ε-regular R-words. Indeed, such a word could only appear after rewriting some product of terms of highest parts of elements of $\mathcal{N}$ in ε-regular form. Assume that there is a term,

$$\overline{\overline{m}}=\beta\overline{\overline{b_{j_1}}}\,\overline{\overline{b_{j_2}}}\cdots\overline{\overline{b_{j_r}}},$$

where $\overline{\overline{b_{j_k}}}$ is some term in the highest part of the element b_{j_k}, such that $\overline{\overline{m}}^*$ and $\overline{e}^*$ are similar terms. Then, since all $\overline{\overline{b_{j_k}}}$ and $\overline{b_{i_k}}$ are assumed to be ε-regular, from the process of constructing w^* from w it follows that $\overline{\overline{m}}^*$ can be represented as a product of the same words $\overline{\overline{b_{j_1}}},\ \ldots,\ \overline{\overline{b_{j_r}}}$ with possibly a different order and a

different arrangement of parentheses. The same applies to the term $\overline{e}^*$. If some term $\overline{\overline{b_{j_k}}}$ is not in fact the leading term of the element b_{j_k}, then it cannot be represented as a product of leading terms of the elements of $\mathcal{N}$ (we recall that analogous statements are made up to a factor from the field P); therefore, from the similarity of $\overline{e}^*$ and $\overline{\overline{m}}^*$, it follows that b_{j_k} taken in a product with other terms $\overline{b_{j_t}}$ must give the leading term $\overline{b_{i_g}}$; but from here it follows that $r > s$ which is impossible. Therefore, all $\overline{\overline{b_{j_k}}}$ are in fact the leading terms of the corresponding elements. On the other hand, from the equation

$$(\alpha\overline{e} - \overline{\overline{m}})^* = 0,\ \alpha \in P,$$

it follows that

$$(\alpha_1 e - m)^* = 0,\ \alpha_1 \in P,$$

where $m = \beta b_{j_1} b_{j_2} \cdots b_{j_r}$ is the term of the polynomial $\overline{f}$ from which the term $\overline{\overline{m}}$ could be obtained. Therefore, since the polynomial $\overline{f}$ is ε-regular, e and m are similar terms, which is a contradiction. This completes the proof of Theorem 2.

References

[1] M. Hall, *A basis for free Lie rings and higher commutators in free groups*, Proc. Amer. Math. Soc. 1 (1950) 575–581.

[2] A.G. Kurosh, *Nonassociative free algebras and free products of algebras*, Mat. Sbornik N.S. 20 (1947) 239–262.

[3] A.I. Malcev, *On algebras defined by identical relations*, Mat. Sbornik N.S. 26 (1950) 19–33.

[4] A.I. Shirshov, *Subalgebras of free Lie algebras*, Mat. Sbornik N.S. 33 (1953) 441–45.

On Special J-rings

A.I. Shirshov

1. Introduction

A commutative ring such that for every pair of elements a and b the following equation holds,

$$J_0\{a,b\} \equiv (a^2b)a - a^2(ba) = 0, \tag{1}$$

is called a *Jordan ring*[1]. In the first four sections of this paper, we will consider Jordan algebras[2] over an arbitrary ring of coefficients Σ, assuming only that Σ is a unital ring and that for each element a in the Jordan algebra there exists a unique element b such that $2b = a$. Clearly, in this case the equation $2a = 0$ implies $a = 0$. In such Jordan algebras, i.e., Jordan algebras without elements of order 2 in the additive group, the following equations hold:

$$J_1\{x,y,z,t\} \equiv$$
$$[(yz)x]t + [(ty)x]z + [(zt)x]y - (yz)(xt) - (ty)(xz) - (zt)(xy) = 0, \tag{2}$$
$$J_2\{x,y,z,t\} \equiv$$
$$[(yz)x]t + [(ty)x]z + [(zt)x]y - [(xz)y]t - [(tx)y]z - [(zt)y]x = 0. \tag{3}$$

The validity of equation (2) follows from the relation

$$J_0\{y+z+t,x\} - J_0\{-y+z+t,x\} - J_0\{y-z+t,x\} - J_0\{y+z-t,x\}$$
$$= 8J_1\{x,y,z,t\},$$

which can be verified by direct computation, and then equation (3) follows from (2) using the relation

$$J_2\{x,y,z,t\} = J_1\{x,y,z,t\} - J_1\{y,x,z,t\}.$$

Mat. Sbornik N.S. 38 (80), (1956), no. 2, 149–166.

[1]Literally, "J-ring". We adopt the modern terminology, "Jordan ring". [Translators]
[2]Literally, "Jordan rings with an arbitrary ring Σ of operators". [Translators]

Jordan algebras also satisfy the relation

$$(ba^s)a^t = (ba^t)a^s, \tag{4}$$

which generalizes equation (1). Indeed, suppose that relation (4) holds for exponents s_1 and t_1 such that $s_1 + t_1 < s + t$. Then, from the equation

$$J_1\{a^{s+t-3}, a, a, ba\} - aJ_1\{a^{s+t-3}, a, a, b\} = a^{s+t-1}(ba) - (ba^{s+t-1})a = 0,$$

which is implied by the inductive hypothesis, the validity of equation (4) follows in the case when one of s or t equals 1. For $s = t$ there is nothing to prove. If $1 < s < t$, then from the equation

$$\begin{aligned} &J_1\{b, a, a^{s-1}, a^t\} \\ &= [(ba^s)a^t - (ba^t)a^s] + [(ba^{t+1})a^{s-1} - (ba^{s-1})a^{t+1}] + [(ba^{s+t-1})a - (ba)a^{s+t-1}] \\ &= 0 \end{aligned}$$

it easily follows that the proof can be completed by induction on $\min(s, t)$. From equation (4) it is easy to obtain associativity for the powers of one element.

It is known that if $\mathcal{A}$ is an associative algebra over Σ which admits unique division by 2, then introducing in $\mathcal{A}$ the new multiplication

$$a \cdot b = \frac{1}{2}(ab + ba),$$

we obtain a Jordan algebra $\mathcal{A}^{(+)}$ with the same additive group and new multiplication. A Jordan algebra I over Σ is called *special* if there exists an associative algebra $\mathcal{A}$ over Σ such that the Jordan algebra $\mathcal{A}^{(+)}$ contains a subalgebra isomorphic to I. Even in the case of algebras over a field, it is known [1] that not every Jordan algebra is special.

In the case of algebras over a field, Cohn [2] proved that a homomorphic image of a special Jordan algebra is not necessarily special. It follows that the class of special Jordan algebras cannot be defined by identical relations. In the present paper, it is proved that a Jordan algebra over Σ that has a finite or countably infinite set of generators is special if and only if it can be embedded into a Jordan algebra over Σ with two generators.

In the last section of this paper, we remove the requirement that for every element a there exists an element b such that $2b = a$. This condition will be replaced by the weaker condition that there are no elements of order 2 in the additive group. It is clear that in this case we will be forced to consider the operation $a \circ b = ab + ba$.

2. An embedding theorem

Consider the free associative algebra A over Σ (see the definition in [3]) with two generators a and b. We will assume that the ring Σ admits unique division by 2. Let A' be the subalgebra of A generated by the elements of the set $T = \{bab, ba^2b, \ldots, ba^nb, \ldots\}$.

Lemma 1. *The subalgebra A' is a free associative algebra over Σ with the set T of free generators.*

Proof. We introduce the notation: $ba^nb = c_n$. Let $f(x_1, x_2, \ldots, x_k)$ be an associative polynomial with coefficients in Σ (with all similar terms combined) such that $f(c_1, c_2, \ldots, c_k) = 0$. Clearly,

$$x_{i_1}^{n_1} x_{i_2}^{n_2} \cdots x_{i_s}^{n_s} \neq x_{j_1}^{m_1} x_{j_2}^{m_2} \cdots x_{j_t}^{m_t} \quad \text{implies} \quad c_{i_1}^{n_1} c_{i_2}^{n_2} \cdots c_{i_s}^{n_s} \neq c_{j_1}^{m_1} c_{j_2}^{m_2} \cdots c_{j_t}^{m_t}.$$

Hence it follows that $f \equiv 0$, and this proves the lemma. □

Let N' be an ideal of the algebra A' over Σ. Then N' generates in A some ideal N.

Lemma 2. *For any ideal N' of the algebra A', the following equality holds: $N \cap A' = N'$.*

Proof. Obviously, $N \cap A' \supseteq N'$. Let n be an element of the ideal N; then $n = \sum_i c_i n_i' d_i$ where $n_i' \in N'$ and c_i, d_i are monomials of A. Let $n \in A'$. This means that all terms that occur in the expression of n such that c_i or d_i does not belong to A' must cancel each other. This implies that $n \in N'$, and this proves the lemma. □

Theorem 1. *Every special Jordan algebra I over Σ that has a finite or countably infinite number of generators can be embedded into a special Jordan algebra over Σ with two generators.*

Proof. Obviously, the associative algebra B over Σ in which the algebra I can be represented can be assumed to have a finite or countably infinite number of generators. It is also clear that the algebra B is isomorphic to a quotient algebra $\overline{A'}$ of the algebra A' by some ideal N'. From Lemma 2 it follows that the quotient algebra $\overline{A}$, of the algebra A with respect to the corresponding ideal N, contains a subalgebra isomorphic to $\overline{A'}$, and therefore also isomorphic to B. Since $ba^kb = 2b \cdot (b \cdot a^k) - a^k \cdot b^2$ it follows that the algebra I is isomorphic to the subalgebra generated in $\overline{A}^{(+)}$ by the two generators $\overline{a}$ and $\overline{b}$ that are the images of the elements a and b. This proves the theorem. □

Clearly, as a byproduct we have reproved the theorem of A.I. Malcev [3] that states that any associative algebra over Σ with a finite or countably infinite number of generators can be embedded into an associative algebra over Σ with two generators.

Remark. From the proof of Theorem 1 it follows that the generators of the algebra I can be expressed in terms of the generators $\overline{a}$ and $\overline{b}$ using only the algebra product without scalar multiplication.

3. Main theorem

Consider the set S of associative words in two generators a and b. The degree of an associative word will be understood in the usual sense; in addition, we will introduce the notion of *height*. The heights of the associative words

$$a^{n_1}b^{m_1}a^{n_2}b^{m_2}\cdots a^{n_k}b^{m_k} \quad \text{and} \quad a^{n_1}b^{m_1}a^{n_2}b^{m_2}\cdots a^{n_k}b^{m_k}a^{n_{k+1}},$$

and also of the words obtained from these by interchanging a and b, will be respectively $2k$ and $2k+1$. The words of the form a^s and b^r have height 1, the words a^sb^r and b^ra^s have height 2, and so on.

We define a mapping $\alpha \to \overline{\alpha}$ of the set S onto itself as follows: for $\alpha \in S$ we set $\overline{\alpha} = \alpha$ if α has height 1, and $\overline{\alpha} = d^m\overline{c}$ if $\alpha = cd^m$ where d is one of the generators a and b.

To each associative word α in S we assign an element α^* of the free Jordan algebra I over Σ with two generators a and b as follows:

$$\alpha^* = \begin{cases} \alpha \text{ if } \alpha \text{ has height 1,} \\ a^s \circ b^r \text{ if } \alpha = a^sb^r \text{ or } \alpha = b^ra^s, \\ a^m \circ (ca^n)^* + (a^mc)^* \circ a^n - c^* \circ a^{m+n} \text{ if } \alpha = a^mca^n, \\ 2a^m \circ (b^n \circ c^*) + 2b^n \circ (a^m \circ c^*) - 2(a^m \circ b^n) \circ c^* - (b^nca^m)^* \\ \qquad \text{if } \alpha = a^mcb^n. \end{cases} \tag{5}$$

Interchanging a and b in the third and fourth cases, we obtain two more formulas. The symbol $\circ$ in the right-hand side means the multiplication in the free Jordan algebra; in the fourth case one should also take into account that the height of the word b^nca^m is smaller than the height of the word $\alpha = a^mcb^n$.

For two associative words α and β we introduce the operation

$$\alpha \circ \beta = \frac{1}{4}\big(\alpha\beta + \alpha\overline{\beta} + \beta\alpha + \overline{\beta}\alpha\big), \tag{6}$$

where in the right-hand side we have an element of the free associative algebra A over Σ on two generators a and b. The two meanings of the operation $\circ$ should not cause confusion, as can be seen from the Main Lemma stated below.

The next two formulas follow immediately from the definition:

$$(\overline{\alpha})^* = \alpha^* \quad \text{and} \quad (\alpha \circ \beta)^* = (\beta \circ \alpha)^*. \tag{7}$$

By straightforward computation, one can verify the equation

$$J_1(\alpha, \beta, \gamma, \delta) = 0, \tag{8}$$

where α, β, γ, δ are words from S and the multiplication is performed in the sense of the operation $\circ$. Clearly, from this it follows that

$$J_2(\alpha, \beta, \gamma, \delta) = 0. \tag{9}$$

The validity of equations (8) and (9) will also be clear from what follows.

We extend the operations $*, \overline{\ }, \circ$ linearly to the elements of the free associative algebra over Σ with generators a and b.

Main Lemma. *For associative words α and β in two generators a and b, the following equation holds:*

$$(\alpha \circ \beta)^* = \alpha^* \circ \beta^*.$$

The proof of the Main Lemma, owing to its complexity, will be given in the next section; now we will consider its consequences.

Let A be the free associative algebra over Σ with two generators a and b. In the Jordan algebra $A^{(+)}$, the elements a and b generate a subalgebra $A_0^{(+)}$.

Lemma 3. *Every element of the algebra $A_0^{(+)}$ can be represented as a linear combination (with coefficients from Σ) of elements of the form $\alpha + \overline{\alpha}$ where α is an associative word in a and b.*

Proof. Obviously, it suffices to prove Lemma 3 for monomials relative to the operation $\cdot$ of the algebra $A_0^{(+)}$. For monomials of the form a^r or b^s, the lemma is obvious. Suppose we have some monomial $M = N \cdot P$ of the algebra $A_0^{(+)}$ where N and P are monomials of lower degree for which we assume that Lemma 3 is valid. Then the validity of Lemma 3 follows from the equation:

$$\begin{aligned}&(\alpha + \overline{\alpha}) \cdot \left(\beta + \overline{\beta}\right)\\ &= \frac{1}{2}\left(\alpha\beta + \overline{\alpha\beta}\right) + \frac{1}{2}\left(\alpha\overline{\beta} + \beta\overline{\alpha}\right) + \frac{1}{2}\left(\overline{\alpha}\beta + \overline{\beta}\alpha\right) + \frac{1}{2}\left(\overline{\alpha}\,\overline{\beta} + \beta\alpha\right).\end{aligned}$$

This completes the proof. □

Theorem 2. *The algebra $A_0^{(+)}$ is isomorphic to the free Jordan algebra I over Σ with generators a and b.*

Proof. To each element of the form $(\alpha + \overline{\alpha})/2$ of the algebra $A_0^{(+)}$ where $\alpha \in S$, we assign the element α^* of I. By Lemma 3 this mapping can be extended to the entire additive group of the algebra $A_0^{(+)}$.

We show that this mapping is a homomorphism from $A_0^{(+)}$ to I. It follows from the definition that this mapping is Σ-linear (it preserves addition, and multiplication by elements of Σ). From the equation

$$\begin{aligned}&\left[\frac{1}{2}(\alpha + \overline{\alpha})\right] \cdot \left[\frac{1}{2}\left(\beta + \overline{\beta}\right)\right] = \frac{1}{8}\left(\alpha\beta + \alpha\overline{\beta} + \overline{\alpha}\beta + \overline{\alpha}\overline{\beta} + \beta\alpha + \overline{\beta}\alpha + \beta\overline{\alpha} + \overline{\beta}\overline{\alpha}\right)\\ &= \frac{1}{4}\left(\frac{\alpha\beta + \overline{\beta}\overline{\alpha}}{2}\right) + \frac{1}{4}\left(\frac{\alpha\overline{\beta} + \beta\overline{\alpha}}{2}\right) + \frac{1}{4}\left(\frac{\overline{\alpha}\beta + \overline{\beta}\alpha}{2}\right) + \frac{1}{4}\left(\frac{\overline{\alpha}\,\overline{\beta} + \beta\alpha}{2}\right),\end{aligned}$$

it follows that to the product of the elements $(\alpha + \overline{\alpha})/2$ and $(\beta + \overline{\beta})/2$ there corresponds the following element of I:

$$\frac{1}{4}\left[(\alpha\beta)^* + \left(\alpha\overline{\beta}\right)^* + (\beta\alpha)^* + \left(\overline{\beta}\alpha\right)^*\right] = (\alpha \circ \beta)^* = \alpha^* \circ \beta^*,$$

where we have used the Main Lemma and the bilinearity of the operation $*$. This implies that the image of the product equals the product of the images.

We now show that each element of the algebra I is the image of some element of the algebra $A_0^{(+)}$. Indeed, the elements of I of the form a^s and b^r have obvious pre-images; if we now assume the existence of pre-images for elements n and p of the algebra I, then clearly there exists a pre-image for $m = n \circ p$. The proof can now be completed by induction and the passage from monomials to polynomials.

Since the Jordan algebra I is free, it is clear that the mapping just constructed is an isomorphism, and this completes the proof. □

Every ideal I_1 of the algebra $A_0^{(+)}$, being a subset of the algebra A, generates in it some ideal $\overline{I_1}$.

Lemma 4. *For every ideal I_1 of the algebra $A_0^{(+)}$, the following equality holds:*

$$\overline{I_1} \cap A_0^{(+)} = I_1.$$

Proof. From Lemma 3 it follows that each element s of the algebra $A_0^{(+)}$, considered as an associative polynomial, satisfies the relation $s = \overline{s}$. Let v be an element of the intersection $\overline{I_1} \cap A_0^{(+)}$. Then v, being an element of the ideal $\overline{I_1}$, can be written as

$$v = \sum_k c_k i_k d_k,$$

where $i_k \in I_1$ and c_k, d_k are associative words. Since v is an element of the algebra $A_0^{(+)}$, we have

$$v = \frac{1}{2}(v + \overline{v}) = \sum_k \frac{c_k\, i_k\, d_k + \overline{d_k}\, i_k\, \overline{c_k}}{2}.$$

We show that each summand

$$e_k = \frac{c_k\, i_k\, d_k + \overline{d_k}\, i_k\, \overline{c_k}}{2}$$

belongs to the ideal I_1. We carry out an induction on the sum of the heights of the words c_k and d_k. If this sum is equal to 1, i.e., one of the words has the form a^r or b^s and the other is empty, then the statement is obvious. Suppose the statement has been proved for all smaller sums. We show that in this case the statement is true if both words c_k and d_k are nonempty. Then, up to interchanging the generators a and b, there are two possible cases:

(1) $e_k = \dfrac{a^m c a^n + a^n \overline{c} a^m}{2}$, and so

$$e_k = a^m \cdot \frac{c a^n + a^n \overline{c}}{2} + a^n \cdot \frac{a^m c + \overline{c} a^m}{2} - a^{m+n} \cdot \frac{c + \overline{c}}{2},$$

(2) $e_k = \dfrac{a^m c b^n + b^n \overline{c} a^m}{2}$, and so

$$e_k = 2a^m \cdot \left(b^n \cdot \frac{c + \overline{c}}{2}\right) + 2b^n \cdot \left(a^m \cdot \frac{c + \overline{c}}{2}\right) - 2\,(b^n \cdot a^m) \cdot \frac{c + \overline{c}}{2} - \frac{b^n c a^m + a^m \overline{c} b^n}{2}.$$

Clearly, in both cases, the conditions of the inductive hypothesis are satisfied, provided that

$$e_k \neq \frac{a^m i_k b^n + b^n i_k a^m}{2},$$

and in the remaining case,

$$e_k = a^m \cdot (b^n \cdot i_k) + b^n \cdot (a^m \cdot i_k) - (b^n \cdot a^m) \cdot i_k.$$

It remains to consider the case when

$$e_k = \frac{i_k CD + \overline{D}\,\overline{C}\,i_k}{2}.$$

This case can be reduced to a previous case using the inductive hypothesis and the equation

$$e_k = 2D \cdot \frac{i_k C + \overline{C} i_k}{2} - \frac{D i_k C + \overline{C} i_k D}{2},$$

if we assume that $D = \overline{D}$. We are permitted to make this assumption by separating as D a factor of height 1. Therefore, we have proved that $\overline{I_1} \cap A_0^{(+)} \subseteq I_1$. The reverse inclusion is obvious, and this completes the proof of the lemma. □

Theorem 3. *Every Jordan algebra $\mathcal{N}$ over Σ with two generators is special.*

Proof. From Theorem 1 it follows that the algebra $\mathcal{N}$ is isomorphic to a quotient algebra of $A_0^{(+)}$ by some ideal I_1. Lemma 4 implies that, in the quotient algebra $A/\overline{I_1}$, distinct elements of the algebra $A_0^{(+)}/I_1$ have distinct images. From here it follows that the Jordan algebra $\mathcal{N} \cong A_0^{(+)}/I_1$ is isomorphic to a subalgebra of the Jordan algebra $(A/\overline{I_1})^{(+)}$. This completes the proof. □

Remark. The statements of Lemmas 3 and 4 for algebras over a field are contained in the results of Cohn [2], where a special case of Theorem 3 is also proved, stating that a homomorphic image of a special Jordan algebra (over a field) with two generators is a special Jordan algebra.

4. Main lemma

We start the proof of the Main Lemma. In the course of the proof, for associative words α and β, we will assume the following inductive hypotheses:

1) *The lemma holds for pairs of words for which the sum of the heights is less than the corresponding sum for α and β.*
2) *The lemma holds for pairs of words for which the sum of the heights is equal to the corresponding sum for α and β, but the sum of the degrees is less than the corresponding sum for α and β.*

The basis of the induction is the obvious validity of the lemma when the sum of heights equals 2.

3) *The lemma holds for pairs of words for which the sum of the heights as well as the sum of the degrees are equal to the corresponding sums for α and β, but the smaller of the heights is less than the smaller of the heights of α and β.*

The basis of the induction for the last hypothesis will be justified below, where it will be shown that the lemma holds if one of the heights of α and β is less than 3.

Suppose now that β has height greater than 2, and the height of α is not less than the height of β. Then,

$$\beta^* = \sum_i \sigma_i \left(c_i^* \circ d_i^*\right) \circ e_i^* + \sum_j \sigma_j a^{k_j} \circ b^{s_j},$$

where σ_k are some coefficients. By the bilinearity of all the operations, to prove the lemma it suffices to consider in place of β^* the elements $\beta_1 = (c^* \circ d^*) \circ e^*$ and $\beta_2 = a^k \circ b^s$.

From inductive hypothesis 3) it follows that

$$\alpha^* \circ \beta_2^* = \alpha^* \circ \left(a^k \circ b^s\right) = \left[\alpha \circ \left(a^k \circ b^s\right)\right]^* = (\alpha \circ \beta_2)^*.$$

From equation (2) it follows that

$$\begin{aligned}
&\alpha^* \circ \beta_1^* = \alpha^* \circ [(c^* \circ d^*) \circ e^*] \\
&= J_1\{e^*, c^*, d^*, \alpha^*\} - [(\alpha^* \circ c^*) \circ e^*] \circ d^* - [(\alpha^* \circ d^*) \circ e^*] \circ c^* \\
&\quad + (\alpha^* \circ e^*) \circ (c^* \circ d^*) + (\alpha^* \circ c^*) \circ (d^* \circ e^*) + (\alpha^* \circ d^*) \circ (c^* \circ e^*),
\end{aligned}$$

but according to inductive hypothesis 3) and equation (8) we have

$$\alpha^* \circ \beta_1^* = [\alpha \circ \beta_1 - J_1\{e, c, d, a\}]^* = (\alpha \circ \beta_1)^*.$$

This completes the proof of the lemma.

It will be far more difficult to justify the basis for inductive hypothesis 3). Here the proof will consist of a number of cases.

4.1. Case 1: $\alpha = a^m b^s D b^r$, $\beta = a^n$

From the definitions of $*$ and $\circ$, and equation (2), it follows that

$$\begin{aligned}
&(\alpha \circ \beta)^* - \alpha^* \circ \beta^* = \frac{1}{2}\left(a^m b^s D b^r a^n\right)^* + \frac{1}{2}\left(a^{m+n} b^s D b^r\right)^* - \left(a^m b^s D b^r\right)^* \circ a^n \\
&= \frac{1}{2} a^m \circ \left(b^s D b^r a^n\right)^* - \frac{1}{2} a^n \circ \left(a^m b^s D b^r\right)^* - \frac{1}{2} a^{m+n} \circ \left(b^s D b^r\right)^* \\
&\quad + \frac{1}{2}\left(a^{m+n} b^s D b^r\right)^* \\
&= a^m \circ \{b^s \circ [a^n \circ (D b^r)^*]\} + a^m \circ \{a^n \circ [b^s \circ (D b^r)^*]\} \\
&\quad - a^m \circ \{(a^n \circ b^s) \circ (D b^r)^*]\} - \frac{1}{2} a^m \circ \left(a^n D b^{r+s}\right)^* - \frac{1}{2} a^n \circ \left(a^m b^s D b^r\right)^* \\
&\quad - \frac{1}{2} a^{m+n} \circ \left(b^s D b^r\right)^* + \frac{1}{2}\left(a^{m+n} b^s D b^r\right)^*
\end{aligned}$$

$$
\begin{aligned}
&= J_2\{b^s, (Db^r)^*, a^m, a^n\} - a^n \circ \{b^s \circ [a^m \circ (Db^r)^*]\} - (a^{m+n} \circ b^s) \circ (Db^r)^* \\
&\quad + b^s \circ [a^{m+n} \circ (Db^r)^*] + a^n \circ [(a^m \circ b^s) \circ (Db^r)^*] \\
&\quad + a^m \circ \{a^n \circ [b^s \circ (Db^r)^*]\} - \frac{1}{2} a^m \circ (a^n Db^{r+s})^* - \frac{1}{2} a^n \circ (a^m b^s Db^r)^* \\
&\quad - \frac{1}{2} a^{m+n} \circ (b^s Db^r)^* + \frac{1}{2} (a^{m+n} b^s Db^r)^*.
\end{aligned}
$$

Using inductive hypothesis 1), we can write the following equations:

$$
\begin{aligned}
&a^n \circ \{b^s \circ [a^m \circ (Db^r)^*]\} = \qquad (10)\\
&\frac{1}{4}\Big[a^n \circ (b^s a^m Db^r)^* + a^n \circ (b^s Db^r a^m)^* + a^n \circ (a^m Db^{r+s})^* + a^n \circ (Db^r a^m b^s)^*\Big],\\
&(a^{m+n} \circ b^s) \circ (Db^r)^* = \qquad (11)\\
&\frac{1}{4}\Big[(a^{m+n} b^s Db^r)^* + (b^s a^{m+n} Db^r)^* + (Db^r a^{m+n} b^s)^* + (Db^{r+s} a^{m+n})^*\Big],\\
&b^s \circ [a^{m+n} \circ (Db^r)^*] = \qquad (12)\\
&\frac{1}{4}\Big[(b^s a^{m+n} Db^r)^* + (b^s Db^r a^{m+n})^* + (a^{m+n} Db^{r+s})^* + (Db^r a^{m+n} b^s)^*\Big],\\
&a^n \circ [(a^m \circ b^s) \circ (Db^r)^*] = \qquad (13)\\
&\frac{1}{4}\Big[a^n \circ (a^m b^s Db^r)^* + a^n \circ (b^s a^m Db^r)^* + a^n \circ (Db^r a^m b^s)^* + a^n \circ (Db^{r+s} a^m)^*\Big],\\
&\frac{1}{2} a^m \circ (a^n Db^{r+s})^* = \frac{1}{4}\Big[(a^{m+n} Db^{r+s})^* + (a^n Db^{r+s} a^m)^*\Big], \qquad (14)\\
&\frac{1}{2} a^{m+n} \circ (b^s Db^r)^* = \frac{1}{4}\Big[(a^{m+n} b^s Db^r)^* + (b^s Db^r a^{m+n})^*\Big]. \qquad (15)
\end{aligned}
$$

Using equation (4), and inductive hypothesis 1), we obtain the equation

$$
\begin{aligned}
a^m \circ \{a^n \circ [b^s \circ (Db^r)^*]\} &= a^n \circ \{a^m \circ [b^s \circ (Db^r)^*]\} \qquad (16)\\
&= \frac{1}{4}\Big[a^n \circ (a^m b^s Db^r)^* + a^n \circ (a^m Db^{r+s})^* \\
&\qquad + a^n \circ (b^s Db^r a^m)^* + a^n \circ (Db^{r+s} a^m)^*\Big].
\end{aligned}
$$

Substituting the right-hand sides of equations (10–16) for the corresponding terms in the preceding expression for $(\alpha \circ \beta)^* - \alpha^* \circ \beta^*$, and combining like terms, we obtain:

$$
\begin{aligned}
&(\alpha \circ \beta)^* - \alpha^* \circ \beta^* = \\
&\frac{1}{2} a^n \circ \left(Db^{r+s} a^m\right)^* - \frac{1}{4}\left(Db^{r+s} a^{m+n}\right)^* - \frac{1}{4}\left(a^n Db^{r+s} a^m\right)^* = 0,
\end{aligned}
$$

by inductive hypothesis 1).

4.2. Case 2: $\alpha = a^s D a^m$, $\beta = a^t$

First we prove the validity of the lemma for $s = m$. In this case, it follows from the definitions of the operations, inductive hypothesis 1), and equation (4), that

$$\begin{aligned}(\alpha\circ\beta)^* &= \left(a^sDa^s\circ a^t\right)^* = \frac{1}{2}\left(a^sDa^{s+t}\right)^* + \frac{1}{2}\left(a^{s+t}Da^s\right)^* \\ &= \frac{1}{2}a^s\circ\left(Da^{s+t}\right)^* + \frac{1}{2}a^{s+t}\circ(a^sD)^* + \frac{1}{2}a^{s+t}\circ(Da^s)^* + \frac{1}{2}a^s\circ\left(a^{s+t}D\right)^* \\ &\quad - a^{2s+t}\circ D^* \\ &= a^s\circ\left(D^*\circ a^{s+t}\right) + a^{s+t}\circ(a^s\circ D^*) - a^{2s+t}\circ D^* \\ &= 2a^{s+t}\circ(a^s\circ D^*) - a^{2s+t}\circ D^*.\end{aligned} \tag{17}$$

By inductive hypothesis 2) we have:

$$\begin{aligned}\alpha^*\circ\beta^* &= (a^sDa^s)^*\circ a^t = a^t\circ[2a^s\circ(a^s\circ D) - a^{2s}\circ D]^* \\ &= 2a^t\circ[a^s\circ(a^s\circ D^*)] - a^t\circ(a^{2s}\circ D^*).\end{aligned} \tag{18}$$

From equations (17) and (18) it follows that:

$$\begin{aligned}&(\alpha\circ\beta)^* - \alpha^*\circ\beta^* \\ &= 2a^{s+t}\circ(a^s\circ D^*) + a^t\circ\left(a^{2s}\circ D^*\right) - 2a^t\circ[a^s\circ(a^s\circ D^*)] - a^{2s+r}\circ D^* \\ &= -J_1\{a^s, a^s, a^t, D^*\} = 0.\end{aligned}$$

In the proof of the general case, we will assume that $s > m$. We can do this without loss of generality, because in the contrary case, we can consider $\overline{\alpha}$ instead of α. Using hypothesis 2) we obtain:

$$\begin{aligned}\alpha^*\circ\beta^* &= (a^sDa^m)^*\circ a^t \\ &= a^t\circ\{2a^m\circ[a^m\circ(a^{s-m}D)] - a^{2m}\circ(a^{s-m}D)]\}^* \\ &= 2a^t\circ\{a^m\circ[a^m\circ(a^{s-m}D)^*]\} - a^t\circ[a^{2m}\circ(a^{s-m}D)^*] \\ &= J_2\{a^m, (a^{s-m}D)^*, a^t, a^m\} + 2a^{m+t}\circ[a^m\circ(a^{s-m}D)^*] - a^{2m+t}\circ(a^{s-m}D)^* \\ &= (a^sD)^*\circ a^{m+t} + (a^{s-m}Da^m)^*\circ a^{m+t} - a^{2m+t}\circ(a^{s-m}D)^*.\end{aligned} \tag{19}$$

From equation (19) it follows that the proof can be completed by induction on the degree of β, with constant sums of heights and degrees of α and β.

Assuming that the lemma holds for $t' > t$, we have:

$$(a^{s-m}Da^m)^*\circ a^{m+t} = \frac{1}{2}(a^{s+t}Da^m)^* + \frac{1}{2}(a^{s-m}Da^{2m+t})^*. \tag{20}$$

From inductive hypothesis 1) it follows that:

$$(a^sD)^*\circ a^{m+t} = \frac{1}{2}(a^{m+t+s}D)^* + \frac{1}{2}(a^sDa^{m+t})^*, \tag{21}$$

and

$$a^{2m+t}\circ(a^{s-m}D)^* = \frac{1}{2}(a^{m+t+s}D)^* + \frac{1}{2}(a^{s-m}Da^{2m+t})^*. \tag{22}$$

Using equations (20), (21) and (22), we obtain from (19):

$$\alpha^* \circ \beta^* = \frac{1}{2}(a^{s+t}Da^m)^* + \frac{1}{2}(a^sDa^{m+t})^* = (\alpha \circ \beta)^*,$$

as desired.

Remark. The sum of heights in Case 1 is odd, and in Case 2 is even. From transformations (17) through (22) it is clear that the proof in Case 2 reduces to Case 1 with the sum of heights being smaller by 1. Therefore the validity of the Main Lemma for Case 2 with the sum of the heights of α and β equal to 2ℓ can be assumed as soon as the validity is assumed for Case 1 with the corresponding sum equal to $2\ell - 1$. This remark will be needed in the proof of Case 5.

4.3. Case 3: $\alpha = a^mb^sDb^ra^t$**,** $\beta = b^n$

First we consider the easier special case when the word D is empty. Let $\alpha_1 = a^mb^sa^t$. Then from equations (2), (4), and the definitions of the operations, it follows that

$$\begin{aligned}
(\alpha_1 \circ \beta)^* &= \frac{1}{2}(a^mb^sa^tb^n)^* + \frac{1}{2}(b^na^mb^sa^t)^* \\
&= a^m \circ [b^n \circ (b^s \circ a^t)] + b^n \circ [a^m \circ (b^s \circ a^t)] - (b^n \circ a^m) \circ (b^s \circ a^t) \\
&\quad - \frac{1}{2}b^{n+s} \circ a^{m+t} + b^n \circ [a^t \circ (a^m \circ b^s)] + a^t \circ [b^n \circ (a^m \circ b^s)] \\
&\quad - (b^n \circ a^t) \circ (b^s \circ a^m) - \frac{1}{2}b^{n+s} \circ a^{m+t} \\
&= J_1\{b^n, b^s, a^t, a^m\} - b^s \circ (b^n \circ a^{t+m}) + 2b^n \circ [a^m \circ (b^s \circ a^t)] \\
&= 2[a^m \circ (b^s \circ a^t)] \circ b^n - (a^{t+m} \circ b^s) \circ b^n = \alpha_1^* \circ \beta^*,
\end{aligned}$$

as desired. In the general case, the proof is much longer.

Using inductive hypothesis 1), the definitions of the operations, and equation (4), we obtain

$$\begin{aligned}
(\alpha \circ \beta)^* - \alpha^* \circ \beta^* &= [(a^mb^sDb^ra^t) \circ b^n]^* - (a^mb^sDb^ra^t)^* \circ b^n \\
&= \frac{1}{2}(a^mb^sDb^ra^tb^n)^* + \frac{1}{2}(b^na^mb^sDb^ra^t)^* - (a^mb^sDb^ra^t)^* \circ b^n \\
&= a^m \circ [b^n \circ (b^sDb^ra^t)^*] + b^n \circ [a^m \circ (b^sDb^ra^t)^*] - (b^n \circ a^m) \circ (b^sDb^ra^t)^* \\
&\quad - \frac{1}{2}(b^{n+s}Db^ra^{t+m})^* + b^n \circ [a^t \circ (a^mb^sDb^r)^*] + a^t \circ [b^n \circ (a^mb^sDb^r)^*] \\
&\quad - (a^t \circ b^n) \circ (a^mb^sDb^r)^* - \frac{1}{2}(a^{m+t}b^sDb^{r+n})^* - b^n \circ [a^m \circ (b^sDb^ra^t)^*] \\
&\quad - b^n \circ [a^t \circ (a^mb^sDb^r)^*] + b^n \circ [a^{m+t} \circ (b^sDb^r)^*] \\
&= a^m \circ [b^n \circ (b^sDb^ra^t)^*] - (b^n \circ a^m) \circ (b^sDb^ra^t)^* + a^t \circ [b^n \circ (a^mb^sDb^r)^*] \\
&\quad - (a^t \circ b^n) \circ (a^mb^sDb^r)^* - \frac{1}{2}(b^{n+s}Db^ra^{t+m})^* - \frac{1}{2}(a^{m+t}b^sDb^{r+n})^* \\
&\quad + b^n \circ [a^{m+t} \circ (b^sDb^r)^*]
\end{aligned}$$

$$
\begin{aligned}
&= a^m \circ [b^n \circ (b^s D b^r a^t)^*] - (b^n \circ a^m) \circ (b^s D b^r a^t)^* \\
&\quad + 2a^t \circ \{b^n \circ [a^m \circ (b^s D b^r)^*]\} - a^t \circ [b^n \circ (b^s D b^r a^m)^*] \\
&\quad - 2(a^t \circ b^n) \circ [a^m \circ (b^s D b^r)^*] + (a^t \circ b^n) \circ (b^s D b^r a^m)^* \\
&\quad - \frac{1}{2}(b^{n+s} D b^r a^{t+m})^* - \frac{1}{2}(a^{m+t} b^s D b^{r+n})^* + b^n \circ [a^{m+t} \circ (b^s D b^r)^*] \\
&= 2a^m \circ \langle b^n \circ \{b^s \circ [a^t \circ (D b^r)^*]\}\rangle + 2a^m \circ \langle b^n \circ \{a^t \circ [b^s \circ (D b^r)^*]\}\rangle \\
&\quad - 2a^m \circ \{b^n \circ [(a^t \circ b^s) \circ (D b^r)^*]\} - 2(a^m \circ b^n) \circ \{b^s \circ [a^t \circ (D b^r)^*]\} \\
&\quad - 2(a^m \circ b^n) \circ \{a^t \circ [b^s \circ (D b^r)^*]\} + 2(a^m \circ b^n) \circ [(a^t \circ b^s) \circ (D b^r)^*] \\
&\quad - 2a^t \circ \langle b^n \circ \{b^s \circ [a^m \circ (D b^r)^*]\}\rangle - 2a^t \circ \langle b^n \circ \{a^m \circ [b^s \circ (D b^r)^*]\}\rangle \\
&\quad + 2a^t \circ \{b^n \circ [(a^m \circ b^s) \circ (D b^r)^*]\} + 2(a^t \circ b^n) \circ \{b^s \circ [a^m \circ (D b^r)^*]\} \\
&\quad + 2(a^t \circ b^n) \circ \{a^m \circ [b^s \circ (D b^r)^*]\} - 2(a^t \circ b^n) \circ [(a^m \circ b^s) \circ (D b^r)^*] \\
&\quad + 2a^t \circ \{b^n \circ [a^m \circ (b^s D b^r)^*]\} - 2(a^t \circ b^n) \circ [a^m \circ (b^s D b^r)^*] \\
&\quad - \frac{1}{2}(b^{n+s} D b^r a^{t+m})^* - \frac{1}{2}(a^{m+t} b^s D b^{r+n})^* + b^n \circ [a^{m+t} \circ (b^s D b^r)^*] \\
&\quad - a^m \circ [b^n \circ (a^t D b^{r+s})^*] + a^t \circ [b^n \circ (a^m D b^{r+s})^*] \\
&\quad + (a^m \circ b^n) \circ (a^t D b^{r+s})^* - (a^t \circ b^n) \circ (a^m D b^{r+s})^* \\
&= 2J_1\{b^n, a^m, b^s, a^t \circ (D b^r)^*\} - 2b^s \circ \langle b^n \circ \{a^m \circ [a^t \circ (D b^r)^*]\}\rangle \\
&\quad - 2[(a^m \circ b^s) \circ b^n] \circ [a^t \circ (D b^r)^*] + 2b^{n+s} \circ \{a^m \circ [a^t \circ (D b^r)^*]\} \\
&\quad + 2\{b^n \circ [a^t \circ (D b^r)^*]\} \circ (a^m \circ b^s) - 2a^t \circ \langle b^n \circ \{a^m \circ [b^s \circ (D b^r)^*]\} \\
&\quad + 2(a^t \circ b^n) \circ \{a^m \circ [b^s \circ (D b^r)^*]\} - 2J_1\{b^n, a^m, a^t \circ b^s, (D b^r)^*\} \\
&\quad + 2(a^t \circ b^s) \circ \{b^n \circ [a^m \circ (D b^r)^*]\} + 2\{b^n \circ [a^m \circ (a^t \circ b^s)]\} \circ (D b^r)^* \\
&\quad - 2[(a^t \circ b^s) \circ b^n] \circ [a^m \circ (D b^r)^*] - 2[a^m \circ (a^t \circ b^s)] \circ [b^n \circ (D b^r)^*] \\
&\quad - 2J_1\{b^n, a^t, b^s, a^m \circ (D b^r)^*\} + 2b^s \circ \langle b^n \circ \{a^t \circ [a^m \circ (D b^r)^*]\}\rangle \\
&\quad + 2[(a^t \circ b^s) \circ b^n] \circ [a^m \circ (D b^r)^*] - 2b^{n+s} \circ \{a^t \circ [a^m \circ (D b^r)^*]\} \\
&\quad - 2\{b^n \circ [a^m \circ (D b^r)^*]\} \circ (a^t \circ b^s) + 2a^m \circ \langle b^n \circ \{a^t \circ [b^s \circ (D b^r)^*]\}\rangle \\
&\quad - 2(a^m \circ b^n) \circ \{a^t \circ [b^s \circ (D b^r)^*]\} + 2J_1\{b^n, a^t, a^m \circ b^s, (D b^r)^*\} \\
&\quad - 2(a^m \circ b^s) \circ \{b^n \circ [a^t \circ (D b^r)^*]\} - 2\{b^n \circ [a^t \circ (a^m \circ b^s)]\} \circ (D b^r)^* \\
&\quad + 2[(a^m \circ b^s) \circ b^n] \circ [a^t \circ (D b^r)^*] + 2[a^t \circ (a^m \circ b^s)] \circ [b^n \circ (D b^r)^*] \\
&\quad + 2a^t \circ \{b^n \circ [a^m \circ (b^s D b^r)^*]\} - 2(a^t \circ b^n) \circ [a^m \circ (b^s D b^r)^*] \\
&\quad - \frac{1}{2}(b^{n+s} D b^r a^{t+m})^* - \frac{1}{2}(a^{m+t} b^s D b^{r+n})^* + b^n \circ [a^{m+t} \circ (b^s D b^r)^*] \\
&\quad - a^m \circ [b^n \circ (a^t D b^{r+s})^*] + (a^m \circ b^n) \circ (a^t D b^{r+s})^* \\
&\quad + a^t \circ [b^n \circ (a^m D b^{r+s})^*] - (a^t \circ b^n) \circ (a^m D b^{r+s})^*
\end{aligned}
$$

$$
\begin{aligned}
&= -2a^t \circ \langle b^n \circ \{a^m \circ [b^s \circ (Db^r)^*]\}\rangle + 2(a^t \circ b^n) \circ \{a^m \circ [b^s (Db^r)^*]\} \\
&\quad + 2a^m \circ \langle b^n \circ \{a^t \circ [b^s \circ (Db^r)^*]\}\rangle - 2(a^m \circ b^n) \circ \{a^t \circ [b^s \circ (Db^r)^*]\} \\
&\quad + 2a^t \circ \{b^n \circ [a^m \circ (b^s Db^r)^*]\} - 2(a^t \circ b^n) \circ [a^m \circ (b^s Db^r)^*] \\
&\quad - \frac{1}{2}(b^{n+s} Db^r a^{t+m})^* - \frac{1}{2}(a^{m+t} b^s Db^{r+n})^* + b^n \circ [a^{m+t} \circ (b^s Db^r)^*] \\
&\quad - a^m \circ [b^n \circ (a^t Db^{r+s})^*] + (a^m \circ b^n) \circ (a^t Db^{r+s})^* \\
&\quad + a^t \circ [b^n \circ (a^m Db^{r+s})^*] - (a^t \circ b^n) \circ (a^m Db^{r+s})^* \\
&= -2a^t \circ \langle b^n \circ \{a^m \circ [b^s \circ (Db^r)^*]\}\rangle + 2(a^t \circ b^n) \circ \{a^m \circ [b^s \circ (Db^r)^*]\} \\
&\quad + 2J_1\{b^n, a^t, a^m, b^s \circ (Db^r)^*\} - 2a^t \circ \langle b^n \circ \{a^m \circ [b^s \circ (Db^r)^*]\}\rangle \\
&\quad - 2(a^{m+t} \circ b^n) \circ [b^s \circ (Db^r)^*] + 2(a^t \circ b^n) \circ \{a^m \circ [b^s \circ (Db^r)^*]\} \\
&\quad + 2a^{m+t} \circ \{b^n \circ [b^s \circ (Db^r)^*]\} + 2a^t \circ \{b^n \circ [a^m \circ (b^s Db^r)^*]\} \\
&\quad - 2(a^t \circ b^n) \circ [a^m \circ (b^s Db^r)^*] - \frac{1}{2}(b^{n+s} Db^r a^{t+m})^* - \frac{1}{2}(a^{m+t} b^s Db^{r+n})^* \\
&\quad + b^n \circ [a^{m+t} \circ (b^s Db^r)^*] - a^m \circ [b^n \circ (a^t Db^{r+s})^*] + (a^m \circ b^n) \circ (a^t Db^{r+s})^* \\
&\quad + a^t \circ [b^n \circ (a^m Db^{r+s})^*] - (a^t \circ b^n) \circ (a^m Db^{r+s})^* \\
&= -2a^t \circ \{b^n \circ [a^m \circ (Db^{r+s})^*]\} + 2(a^t \circ b^n) \circ [a^m \circ (Db^{r+s})^*] \\
&\quad - 2(a^{m+t} \circ b^n) \circ [b^s \circ (Db^r)^*] + 2(a^{m+t} \circ \{b^n \circ [b^s \circ (Db^r)^*]\} \\
&\quad - \frac{1}{2}(b^{n+s} Db^r a^{t+m})^* - \frac{1}{2}(a^{m+t} b^s Db^{r+m})^* + b^n \circ [a^{m+t} \circ (b^s Db^r)^*] \\
&\quad - a^m \circ [b^n \circ (a^t Db^{r+s})^*] + (a^m \circ b^n) \circ (a^t Db^{r+s})^* \\
&\quad + a^t \circ [b^n \circ (a^m Db^{r+s})^*] - (a^t \circ b^n) \circ (a^m Db^{r+s})^* \\
&= \frac{1}{4}\Big[- (a^t b^n a^m Db^{r+s})^* - (a^t b^n Db^{s+r} a^m)^* - (a^{t+m} Db^{r+s+n})^* \\
&\qquad - (a^t Db^{r+s} a^m b^n)^* - (b^n a^m Db^{r+s} a^t)^* - (b^n Db^{r+s} a^{m+t})^* \\
&\qquad - (a^m Db^{r+s+n} a^t)^* - (Db^{r+s} a^m b^n a^t)^* + (a^t b^n a^m Db^{r+s})^* \\
&\qquad + (a^t b^n Db^{r+s} a^m)^* + (b^n a^{t+m} Db^{r+s})^* + (b^n a^t Db^{r+s} a^m)^* \\
&\qquad + (a^m Db^{r+s} a^t b^n)^* + (Db^{r+s} a^{m+t} b^n)^* + (a^m Db^{r+s+n} a^t)^* \\
&\qquad + (Db^{r+s} a^m b^n a^t)^* - (a^{m+t} b^{n+s} Db^r)^* - (b^n a^{m+t} b^s Db^r)^* \\
&\qquad - (a^{m+t} b^n Db^{r+s})^* - (b^n a^{m+t} Db^{r+s})^* - (b^s Db^r a^{m+t} b^n)^* \\
&\qquad - (b^s Db^{r+n} a^{m+t})^* - (Db^{r+s} a^{m+t} b^n)^* - (Db^{r+s+n} a^{m+t})^* \\
&\qquad + (a^{m+t} b^{n+s} Db^r)^* + (a^{m+t} b^n Db^{r+s})^* + (a^{m+t} b^s Db^{r+n})^* \\
&\qquad + (a^{m+t} Db^{r+s+n})^* + (b^{n+s} Db^r a^{m+t})^* + (b^n Db^{r+s} a^{m+t})^* \\
&\qquad + (b^s Db^{r+n} a^{m+t})^* + (Db^{r+s+n} a^{m+t})^* - 2(b^{n+s} Db^r a^{t+m})^* \\
&\qquad - 2(a^{m+t} b^s Db^{r+n})^* + (b^n a^{m+t} b^s Db^r)^* + (b^{n+s} Db^r a^{m+t})^*
\end{aligned}
$$

$$
\begin{aligned}
&+ (a^{m+t}b^sDb^{r+n})^* + (b^sDb^ra^{m+t}b^n)^* - (a^mb^na^tDb^{r+s})^* \\
&- (a^{m+t}Db^{r+s+n})^* - (b^na^tDb^{r+s}a^m)^* - (a^tDb^{r+s+n}a^m)^* \\
&+ (a^mb^na^tDb^{r+s})^* + (b^na^{m+t}Db^{r+s})^* + (a^tDb^{r+s}a^mb^n)^* \\
&+ (a^tDb^{r+s+n}a^m)^* + (a^tb^na^mDb^{r+s})^* + (a^{t+m}Db^{r+s+n})^* \\
&+ (b^na^mDb^{r+s}a^t)^* + (a^mDb^{r+s+n}a^t)^* - (a^tb^na^mDb^{r+s})^* \\
&- (b^na^{t+m}Db^{r+s})^* - (a^mDb^{r+s}a^tb^n)^* - (a^mDb^{r+s+n}a^t)^* \Big] \\
&= 0,
\end{aligned}
$$

as was to be established.

4.4. Case 4: $\alpha = a^mDb^n$, $\beta = a^tb^q$

From the definition of operation $*$ it follows that

$$
\begin{aligned}
(b^qa^mDb^na^t)^* = {} & 2b^q \circ [a^t \circ (a^mDb^n)^*] + 2a^t \circ [b^q \circ (a^mDb^n)^*] \\
& - 2(a^t \circ b^q) \circ (a^mDb^n)^* - (a^{m+t}Db^{n+q})^*.
\end{aligned}
$$

From this equation we have:

$$
\begin{aligned}
\alpha^* \circ \beta^* &= (a^mDb^n)^* \circ (a^t \circ b^q) \\
&= b^q \circ [a^t \circ (a^mDb^n)^*] + a^t \circ [b^q \circ (a^mDb^n)^*] \\
&\quad - \frac{1}{2}(a^{m+t}Db^{n+q})^* - \frac{1}{2}(b^qa^mDb^na^t)^*.
\end{aligned} \tag{23}
$$

Using inductive hypothesis 1), and the relation proved in Case 3, we obtain the equations

$$
\begin{aligned}
&b^q \circ [a^t \circ (a^mDb^n)^*] = b^q \circ (a^t \circ a^mDb^n)^* = [b^q \circ (a^t \circ a^mDb^n)]^* \qquad (24) \\
&= \frac{1}{4}(b^qa^{t+m}Db^n)^* + \frac{1}{4}(b^qa^mDb^na^t)^* + \frac{1}{4}(a^{t+m}Db^{n+q})^* + \frac{1}{4}(a^mDb^na^tb^q)^*, \\
&a^t \circ [b^q \circ (a^mDb^n)^*] = [a^t \circ (b^q \circ a^mDb^n)]^* \qquad (25) \\
&= \frac{1}{4}(a^tb^qa^mDb^n)^* + \frac{1}{4}(a^{t+m}Db^{n+q})^* + \frac{1}{4}(b^qa^mDb^na^t)^* + \frac{1}{4}(a^mDb^{n+q}a^t)^*.
\end{aligned}
$$

From equations (23), (24) and (25) it follows that

$$
\begin{aligned}
\alpha^* \circ \beta^* &= \frac{1}{4}\Big[(b^qa^{t+m}Db^n)^* + (a^mDb^na^tb^q)^* + (a^tb^qa^mDb^n)^* + (a^mDb^{n+q}a^t)^*\Big] \\
&= (\alpha \circ \beta)^*.
\end{aligned}
$$

Clearly, the same proof is valid if the word D is empty.

4.5. Case 5: $\alpha = a^mDa^n$, $\beta = a^pb^q$

4.5.1. Step 1. First of all we prove that the lemma holds if $m = n$. Using inductive hypothesis 1) and the relation proved in Case 1, we have:

$$
\alpha^* \circ \beta^* = (a^nDa^n)^* \circ (a^p \circ b^q) = [a^n \circ (a^nD + Da^n) - a^{2n} \circ D]^* \circ (a^p \circ b^q)
$$

$$
\begin{aligned}
&= [a^n \circ (a^n D + Da^n)^*] \circ (a^p \circ b^q) - (a^{2n} \circ D^*) \circ (a^p \circ b^q)\\
&= 2[a^n \circ (a^n \circ D^*)] \circ (a^p \circ b^q) - (a^{2n} \circ D^*) \circ (a^p \circ b^q)\\
&= 2J_2\{a^n, D^*, a^n, a^p \circ b^q\} - 2\{a^n \circ [a^n \circ (a^p \circ b^q)]\} \circ D^*\\
&\quad - 2\{a^n \circ [D^* \circ (a^p \circ b^q)]\} \circ a^n + 4\{D^* \circ [a^n \circ (a^p \circ b^q)]\} \circ a^n\\
&\quad + (a^{2n} \circ D^*) \circ (a^p \circ b^q)\\
&= -2\{a^n \circ [D^* \circ (a^p \circ b^q)]\} \circ a^n + 4\{D^* \circ [a^n \circ (a^p \circ b^q)]\} \circ a^n\\
&\quad + (a^{2n} \circ D^*) \circ (a^p \circ b^a) - J_1\{a^n, a^n, a^p, b^q\} \circ D^* + (b^q \circ a^{2n+p}) \circ D^*\\
&\quad - 2[a^{n+p} \circ (a^n \circ b^q)] \circ D^* - [a^{2n} \circ (a^p \circ b^q)] \circ D^*\\
&= \big[- 2\{a^n \circ [D \circ (a^p \circ b^q)] \circ a^n + 4\{D \circ [a^n \circ (a^p \circ b^q)]\} \circ a^n\\
&\qquad + (a^{2n} \circ D) \circ (a^p \circ b^q) - J_1\{a^n, a^n, a^p, b^q\} \circ D\\
&\qquad + (b^q \circ a^{2n+p}) \circ D - 2[a^{n+p} \circ (a^n \circ b^q)] \circ D - [a^{2n} \circ (a^p \circ b^q)] \circ D \big]^*\\
&= \big[- 2\{a^n \circ [a^n \circ (a^p \circ b^q)]\} \circ D - 2\{a^n \circ [D \circ (a^p \circ b^q)]\} \circ a^n\\
&\quad + 4\{D \circ [a^n \circ (a^p \circ b^q)]\} \circ a^n + (a^{2n} \circ D) \circ (a^p \circ b^q) \big]^*\\
&= \big[- 2J_2\{a^n, D, a^n, a^p \circ b^q\} + 2[a^n \circ (a^n \circ D)] \circ (a^p \circ b^q)\\
&\quad - (a^{2n} \circ D) \circ (a^p \circ b^q) \big]^*\\
&= \big\{ [2a^n \circ (a^n \circ D) - a^{2n} \circ D] \circ (a^p \circ b^q) \big\}^* = \big[a^n D a^n \circ (a^p \circ b^q)\big]^*\\
&= (\alpha \circ a^p b^q)^* = (\alpha \circ \beta)^*.
\end{aligned}
$$

4.5.2. Step 2. Now suppose that the lemma holds for some pair of words $\alpha_1 = a^t D a^r$, $\beta_1 = a^k b^s$. We will show that in this case the lemma also holds for the words $\alpha_2 = a^t D a^k$, $\beta_2 = a^r b^s$. Indeed,

$$
\begin{aligned}
&\alpha_2^* \circ \beta_2^* = (a^t D a^k)^* \circ (a^r \circ b^s)\\
&= 2[(a^t D)^* \circ a^k] \circ (a^r \circ b^s) - (a^{t+k} D)^* \circ (a^r \circ b^s)\\
&= -2J_1\{(a^t D)^*, a^k, a^r, b^s\} + 2[(a^t D)^* \circ a^{k+r}] \circ b^s + 2[(a^t D)^* \circ (a^k \circ a^s)] \circ a^r\\
&\quad + 2[(a^t D)^* \circ (a^r \circ b^s)] \circ a^k - 2[(a^t D)^* \circ b^s] \circ a^{k+r}\\
&\quad - 2[(a^t D)^* \circ a^r] \circ (a^k \circ b^s) - (a^{t+k} D)^* \circ (a^r \circ b^s)\\
&= \big[- 2J_1\{a^t D, a^k, a^r, b^s\} + 2(a^t D \circ a^{k+r}) \circ b^s + 2[a^t D \circ (a^k \circ b^s)] \circ a^r\\
&\quad + 2[a^t D \circ (a^r \circ b^s)] \circ a^k - 2(a^t D \circ b^s) \circ a^{k+r} - (a^{t+r} D) \circ (a^k \circ b^s)\\
&\quad - (a^{t+k} D) \circ (a^r \circ b^s) \big]^* - (a^t D a^r)^* \circ (a^k \circ b^s)\\
&= -\alpha_1^* \circ \beta_1^* + \big[2(a^t D \circ a^r) \circ (a^k \circ b^s) + 2(a^t D \circ a^k) \circ (a^r \circ b^s)\\
&\quad - (a^{t+r} D) \circ (a^k \circ b^s) - (a^{t+k} D) \circ (a^r \circ b^s) \big]^*\\
&= -\alpha_1^* \circ \beta_1^* + \big[a^t D a^r \circ (a^k \circ b^s) + a^t D a^k \circ (a^r \circ b^s)\big]^*\\
&= -\alpha_1^* \circ \beta_1^* + (\alpha_1 \circ \beta_1)^* + (\alpha_2 \circ \beta_2)^* = (\alpha_2 \circ \beta_2)^*.
\end{aligned}
$$

The claim is proved.

4.5.3. Step 3. Finally we prove that if the lemma holds for some words $\alpha_3 = a^{2s}Da^r$, $\beta_3 = a^pb^q$ then it also holds for the words $\alpha_4 = a^sDa^{r+s}$ and β_3. Indeed,

$$\begin{aligned}
\alpha_4^* \circ \beta_3^* &= (a^sDa^{r+s})^* \circ (a^p \circ b^q) = 2\{a^s \circ [a^s \circ (Da^r)^*]\} \circ (a^p \circ b^q) \qquad (26)\\
&= 2J_2\{a^s, a^p \circ b^q, a^s, (Da^r)^*\} - 2\{a^s \circ [(a^p \circ b^q) \circ (Da^r)^*]\} \circ a^s\\
&\quad - 2\{a^s \circ [a^s \circ (a^p \circ b^q)]\} \circ (Da^r)^* + 4\{(a^p \circ b^q) \circ [a^s \circ (Da^r)^*]\} \circ a^s\\
&\quad + 2[(a^p \circ b^q) \circ a^{2s}] \circ (Da^r)^* - [a^{2s} \circ (Da^r)^*] \circ (a^p \circ b^q)\\
&= -2\{a^s \circ [(a^p \circ b^q) \circ (Da^r)^*]\} \circ a^s + 4\{(a^p \circ b^q) \circ [a^s \circ (Da^r)^*]\} \circ a^s\\
&\quad - [a^{2s} \circ (Da^r)^*] \circ (a^p \circ b^q) - J_1\{a^s, a^s, a^p, b^q\} \circ (Da^r)^*\\
&\quad + (b^q \circ a^{2s+p}) \circ (Da^r)^* - 2[(a^s \circ b^q) \circ a^{s+p}] \circ (Da^r)^*\\
&\quad + [(a^p \circ b^q) \circ a^{2s}] \circ (Da^r)^*\\
&= -2\{a^s \circ [(a^p \circ b^q) \circ (Da^r)^*]\} \circ a^s + 4\{(a^p \circ b^q) \circ [a^s \circ (Da^r)^*]\} \circ a^s\\
&\quad - [a^{2s} \circ (Da^r)^*] \circ (a^p \circ b^q) + (b^q \circ a^{2s+p}) \circ (Da^r)^*\\
&\quad - 2J_2\{a^{s+p}, b^q, a^s, (Da^r)^*\} + J_2\{a^{2s}, b^q, a^p, (Da^r)^*\}\\
&\quad + 2\{a^{s+p} \circ [b^q \circ (Da^r)^*]\} \circ a^s + 2\{a^{s+p} \circ [a^s \circ (Da^r)^*]\} \circ b^q\\
&\quad - 2(b^q \circ a^{2s+p}) \circ (Da^r)^* - 2\{b^q \circ [a^{s+p} \circ (Da^r)^*]\} \circ a^s\\
&\quad - 2\{b^q \circ [a^s \circ (Da^r)^*]\} \circ a^{s+p} - \{a^{2s} \circ [b^q \circ (Da^r)^*]\} \circ a^p\\
&\quad - \{a^{2s} \circ [a^p \circ (Da^r)^*]\} \circ b^q + (b^q \circ a^{2s+p}) \circ (Da^r)^*\\
&\quad + \{b^q \circ [a^{2s} \circ (Da^r)^*]\} \circ a^p + \{b^q \circ [a^p \circ (Da^r)^*]\} \circ a^{2s}.
\end{aligned}$$

Using inductive hypothesis 2) we can move the symbol $*$ outside the braces in the first two monomials of the right-hand side of equation (26). As a result we obtain the monomials

$$\{a^s \circ [(a^p \circ b^q) \circ (Da^r)]\}^* \circ a^s \quad \text{and} \quad \{(a^p \circ b^q) \circ [a^s \circ (Da^r)]\}^* \circ a^s. \qquad (27)$$

After performing all the $\circ$ operations inside the braces, we obtain, either monomials whose heights will not exceed one less than the sum of the heights of α and β, or words whose heights are equal to the sum of the heights of the words α and β but which together with the word a^s form a pair of the form considered in Case 2. Using the cases already established, and the remark in Case 2, we conclude that the monomials (27) are equal respectively to the monomials

$$\langle\{a^s \circ [(a^p \circ b^q) \circ Da^r]\} \circ a^s\rangle^* \quad \text{and} \quad \langle\{(a^p \circ b^q) \circ [a^s \circ Da^r]\} \circ a^s\rangle^*. \qquad (28)$$

Since it is obvious that

$$\begin{aligned}
[a^{2s} \circ (Da^r)^*] \circ (a^p \circ b^q) &= [a^{2s} \circ Da^r]^* \circ (a^p \circ b^q)\\
&= \frac{1}{2}(a^{2s}Da^r)^* \circ (a^p \circ^q) + \frac{1}{2}(Da^{2s+r})^* \circ (a^p \circ b^q),
\end{aligned}$$

for the right-hand side of equation (26), using the cases established earlier or inductive hypothesis 1), we can move the operation $*$ outside the parentheses everywhere except in the term

$$\frac{1}{2}(a^{2s}Da^r)^* \circ (a^p b^q).$$

We do this, and then perform the transformations done in (26) in the reverse order:

$$\begin{aligned}
\alpha_4^* \circ \beta_3^* = \Big\langle & -2\{a^s \circ [(a^p \circ b^q) \circ (Da^r)]\} \circ a^s + 4\{(a^p \circ b^q) \circ [a^s \circ Da^r]\} \circ a^s \\
& - \frac{1}{2} Da^{2s+r} \circ (a^p \circ b^q) + (b^q \circ a^{2s+p}) \circ Da^r - 2J_2\{a^{s+p}, b^q, a^s, Da^r\} \\
& + J_2\{a^{2s}, b^q, a^p, Da^r\} + 2\{a^{s+p} \circ [b^q \circ Da^r]\} \circ a^s + 2\{a^{s+p} \circ [a^s \circ Da^r]\} \circ b^q \\
& - 2(b^q \circ a^{2s+p}) \circ Da^r - 2[b^q \circ (a^{s+p} \circ Da^r)] \circ a^s - 2[b^q \circ (a^s \circ Da^r)] \circ a^{s+p} \\
& - [a^{2s} \circ (b^q \circ Da^r)] \circ a^p - [a^{2s} \circ (a^p \circ Da^r)] \circ b^q + (b^q \circ a^{2s+p}) \circ Da^r \\
& + [b^q \circ (a^{2s} \circ Da^r)] \circ a^p + [b^q \circ (a^p \circ Da^r)] \circ a^{2s} \Big\rangle^* - \frac{1}{2}(a^{2s}Da^r)^* \circ (a^p \circ b^q) \\
= \Big\langle & -2\{a^s \circ [(a^p \circ b^q) \circ Da^r]\} \circ a^s + 4[(a^p \circ b^q) \circ (a^s \circ Da^r)] \circ a^s \\
& - \frac{1}{2} Da^{2s+r} \circ (a^p \circ b^q) + (b^q \circ a^{2s+p}) \circ Da^r - 2[a^{s+p} \circ (b^q \circ a^s)] \circ Da^r \\
& + [a^{2s} \circ (b^q \circ a^p)] \circ Da^r \Big\rangle^* - \frac{1}{2}\,\alpha_3^* \circ \beta_3^* \\
= \Big\langle & -2\{a^s \circ [(a^p \circ b^q) \circ Da^r]\} \circ a^s + 4[(a^p \circ b^q) \circ (a^s \circ Da^r)] \circ a^s \\
& - \frac{1}{2} Da^{2s+r} \circ (a^p \circ b^q) + J_1\{a^s, a^s, a^p, b^q\} \circ Da^r \\
& - 2\{a^s \circ [a^s \circ (a^p \circ b^q)]\} \circ Da^r + 2[a^{2s} \circ (b^q \circ a^p)] \circ Da^r \Big\rangle^* - \frac{1}{2}\,\alpha_3^* \circ \beta_3^* \\
= \Big\langle & -2J_2\{a^s, a^p \circ b^q, a^s, Da^r\} + 2[a^s \circ (a^s \circ Da^r)] \circ (a^p \circ b^q) \\
& - \frac{1}{2} Da^{2s+r} \circ (a^p \circ b^q) \Big\rangle^* - \frac{1}{2}\,\alpha_3^* \circ \beta_3^* \\
= & \frac{1}{2}[a^{2s}Da^r \circ (a^p \circ b^q)]^* + [a^s Da^{r+s} \circ (a^p \circ b^q)]^* - \frac{1}{2}\,\alpha_3^* \circ \beta_3^* \\
= & \frac{1}{2}(\alpha_3 \circ \beta_3)^* + (\alpha_4 \circ \beta_3)^* - \frac{1}{2}\,\alpha_3^* \circ \beta_3^* = (\alpha_4 \circ \beta_3)^*.
\end{aligned}$$

4.5.4. Step 4.

Definition. By the *δ-transformation* of a pair of natural numbers t and s, $t > s$, we will mean the transformation that replaces this pair by the pair $t - s, 2s$.

Lemma 5. *Starting with an arbitrary triple n, p, q of natural numbers, and using a finite number of δ-transformations, we can pass to another triple of natural numbers at least two of which are equal.*

Proof. Let γ be the greatest natural number such that at least one of the numbers $n+p$, $n+q$, $p+q$ is divisible by 2^γ. The proof will be complete if we can show that, in the case where n, p, q are pairwise distinct, the number γ can be increased by one by δ-transformations.

Obviously, $\gamma > 0$. First we show that if the sum of n and p is divisible by 2^γ then by δ-transformations we can replace them by a pair of numbers such that either they are equal or they are both are divisible by 2^γ. Let

$$n + p = 2^\gamma(2s-1), \quad n > p, \quad p = 2^\mu(2q-1), \quad \mu < \gamma.$$

Then after one δ-transformation both resulting numbers will be divisible by $2^{\mu+1}$. Clearly, for this argument we need simultaneously $n \neq p$ and $\mu < \gamma$. It is therefore obvious that after a finite number of steps we will arrive at a pair of numbers that are either equal or both divisible by 2^γ.

On the basis of the preceding argument, we may assume that two of the given three numbers, say n and p, are divisible by 2^γ. Clearly, one of the numbers n and p is divisible by $2^{\gamma+1}$ since otherwise their sum would be divisible by $2^{\gamma+1}$. Without loss of generality, we may assume that the greater of the numbers n and p is divisible by $2^{\gamma+1}$, since in the contrary case this can be easily obtained by doubling the smaller of the numbers n and p sufficiently many times at the expense of the greater. On the basis of the above arguments, we may assume that

$$n > p, \quad n + p = 2^\gamma(2s-1), \quad n = 2^{\gamma+1}k.$$

As to the number q, there are two possible cases:

1) $q > p$: Then, replacing the pair q, p by the pair $2p$, $q-p$ we see that the number $2p$ is divisible by $2^{\gamma+1}$ and therefore the sum $n+2p$ is also divisible by $2^{\gamma+1}$.
2) $q < p$: Then, the δ-transformations

$$(n,p,q) \longrightarrow (n,p-q,2q) \longrightarrow (n-p+q, 2p-2q, 2q)$$

lead to a triple of natural numbers, for two of which, $2p-2q$ and $2q$, the sum is divisible by $2^{\gamma+1}$.

This completes the proof of Lemma 5. □

Now let us complete the proof of Case 5.

The natural numbers m, n, p that occur in the expressions for α and β can be subjected to arbitrary permutations because of Step 2 and the possibility of replacing α by $\overline{\alpha}$. Step 3 allows us to perform δ-transformations on them. Because of Lemma 5, we can obtain after a finite number of steps the equality of two of these natural numbers. The proof can be completed by Step 1. Case 5 is finished.

The cases considered above together with the cases that can be obtained from them by interchanging a and b or by replacing α by $\overline{\alpha}$ justify the basis of inductive hypothesis 3). This completes the proof of the Main Lemma.

5. The operation $ab + ba$

In the preceding sections, we assumed everywhere that the associative ring Σ admits division by 2. Suppose that we have an associative algebra $\mathcal{B}$ over Σ such that for some elements a and b in $\mathcal{B}$ there does not exist an element c for which $2c = ab + ba$, but the characteristic of the algebra is not 2. Then in the algebra $\mathcal{B}$ we can introduce the operation $a \circ b = ab + ba$, relative to which the additive group of $\mathcal{B}$ will again be a Jordan algebra over Σ. We show that in this case also the main results of this article are valid.

Lemma 6. *Any associative ring Σ with characteristic different from 2 can be embedded in a ring $\overline{\Sigma}$ that admits unique division by 2.*

Proof. Consider the set $\overline{\Sigma}$ of pairs $(\sigma, 2^k)$ where $\sigma \in \Sigma$, and $k \geq 0$ is an integer. We will consider the pairs $(\sigma_1, 2^{k_1})$ and $(\sigma_2, 2^{k_2})$ to be *equivalent* if $2^{k_2}\sigma_1 = 2^{k_1}\sigma_2$. We define addition and multiplication of the pairs in the familiar way:

$$(\sigma_1, 2^{k_1}) + (\sigma_2, 2^{k_2}) = \left(2^{k_2}\sigma_1 + 2^{k_1}\sigma_2, 2^{k_1+k_2}\right),$$
$$(\sigma_1, 2^{k_1})(\sigma_2, 2^{k_2}) = \left(\sigma_1\sigma_2, 2^{k_1+k_2}\right).$$

Obviously, the ring $\overline{\Sigma}$ satisfies the requirements of Lemma 6. □

Suppose we have a Jordan algebra $\mathcal{M}$ over Σ with a finite or countably infinite set of generators, which is special in the new sense, i.e., there exists an associative algebra $\mathcal{B}$ over Σ whose Jordan algebra $\mathcal{B}^{(+)}_{(2)}$ with respect to the operation $\circ$ contains a subalgebra isomorphic to $\mathcal{M}$. Such algebras will be called *semispecial*.

We introduce a new multiplication $\times$ on $\mathcal{B}$ by the equation $a \times b = 2ab$ for $a, b \in \mathcal{B}$. The additive group of $\mathcal{B}$ relative to the old addition and the new multiplication $\times$ will be an associative algebra $\mathcal{B}^{(\times)}$ over Σ. This algebra can be embedded into the algebra $\overline{\mathcal{B}^{(\times)}}$ over $\overline{\Sigma}$ of pairs $(b, 2^k)$ in the way described for Σ. If the action of the elements of $\overline{\Sigma}$ is defined by $(\sigma, 2^s)(b, 2^k) = (\sigma b, 2^{k+s})$, then the subset of $\overline{\mathcal{B}^{(\times)}}$ consisting of the pairs $(m, 2^t)$, where m belongs to the subset of elements of the additive group of $\mathcal{B}$ that correspond to the elements of $\mathcal{M}$, will obviously be a special Jordan algebra over $\overline{\Sigma}$. By Theorem 1, it can be embedded into a special Jordan algebra over $\overline{\Sigma}$ with two generators.

Each element of the algebra $\overline{\mathcal{B}^{(\times)}}$ under this embedding can be expressed in terms of the generators using only the operation $\times$ without the action of $\overline{\Sigma}$, as can be seen from the remark to Theorem 1. If we now return from the multiplication $\times$ to the multiplication $ab = \frac{1}{2}a \times b$, then relative to this operation, the algebra $\mathcal{M}$ will be embedded into a semispecial Jordan algebra with two generators over $\overline{\Sigma}$, which can also be considered as an algebra over Σ. Clearly, the subalgebra over Σ generated by the two generators will be smaller than the corresponding subalgebra over $\overline{\Sigma}$, but it will still contain the algebra $\mathcal{M}$ as can be easily verified. Thus, we have proved the following theorem:

Theorem 4. *Every semispecial Jordan algebra $\mathcal{M}$ over Σ with a finite or countably infinite number of generators, and without elements of order 2 in the additive group, can be embedded into a semispecial Jordan algebra with two generators over Σ.*

Suppose we have a Jordan algebra $\mathcal{N}$ with two generators over Σ; regarding $\mathcal{N}$ we now assume only that its additive group does not have elements of order 2. Since the construction of Lemma 6 applies in this case, we may assume that the algebra $\mathcal{N}$ is embedded into the algebra $\overline{\mathcal{N}}$ of pairs of the form $(n, 2^k)$ for $n \in \mathcal{N}$ $(k = 0, 1, 2, \ldots)$ which is a Jordan algebra over $\overline{\Sigma}$. By Theorem 3, $\overline{\mathcal{N}}$ is a special Jordan algebra over $\overline{\Sigma}$. If we now introduce a new operation $a * b = \frac{1}{2}ab$ on the corresponding associative algebra $\mathcal{A}$ over $\overline{\Sigma}$, then it is obvious that with respect to the algebra $\mathcal{A}^{(*)}$ the Jordan algebra $\overline{\mathcal{N}}$ will be a semispecial Jordan algebra over $\overline{\Sigma}$. If we regard the algebra $\mathcal{A}^{(*)}$ as an algebra over Σ, then the subalgebra $\mathcal{N}$ (over Σ) of the algebra $\left(\mathcal{A}^{(*)}\right)^{(+)}_{(2)}$ will be semispecial. Thus, we have proved the following theorem:

Theorem 5. *Every Jordan algebra $\mathcal{N}$ over Σ with two generators and without elements of order 2 in the additive group is semispecial.*

References

[1] A.A. Albert, *A note on the exceptional Jordan algebra*, Proc. Nat. Acad. Sci. U.S.A. 36 (1950) 372–374.

[2] P.M. Cohn, *On homomorphic images of special Jordan algebras*, Canadian J. Math. 6 (1954) 253–264.

[3] A.I. Malcev, *On a representation of nonassociative rings*, Uspekhi Mat. Nauk N.S. 7 (1952) 181–185.

Some Theorems on Embedding of Rings

A.I. Shirshov

1. Introduction

The present work is a sequel to the author's article [4]. The main topic is special Jordan algebras over rings, but the methods employed also allow us to obtain new results for other classes of algebras. The main result of the present article concerning special Jordan algebras is a necessary and sufficient condition for speciality (or semispeciality) of a Jordan algebra formulated in terms of the algebra itself (Theorems 8 and 9). Other new theorems deal with the general theory of nonassociative rings (Theorems 2, 3, 4, 5).

2. Some embedding theorems

Suppose we have a commutative[1] associative ring Σ and a set Ω of (nonassociative) multilinear polynomials in the independent variables x, y, z, ... with coefficients in Σ. Then we can speak of Ω-*algebras over* Σ, i.e., algebras over Σ in which the polynomials in Ω vanish identically after substituting elements of the algebra for the variables x, y, z, ... In the usual sense we will speak of free Ω-algebras over Σ. Generally speaking, all algebras considered here will be nonassociative.

Definition 1. Let S_k be the free Ω-algebra over Σ with k generators. A countably infinite subset $\mathcal{N}$ of S_k, which is a free generating set for the subalgebra T it generates in S_k, will be called *distinguished* if any ideal I of T is the intersection of the ideal $\overline{I}$ generated by I in S_k with the subalgebra T.

Definition 2. The smallest natural number k (if it exists) for which S_k contains a distinguished subset will be called the *basis rank*[2] of the set Ω over Σ.

Mat. Sbornik N.S. 40 (82), (1956), no. 1, 65–72.

[1]The word "commutative" is omitted in the Russian. [Translators]
[2]Literally, the "dimension". [Translators]

Theorem 1. *If a set of identical relations Ω has basis rank k over the ring Σ, then any Ω-algebra R over Σ that has a finite or countably infinite set of generators can be isomorphically embedded into some Ω-algebra with k generators.*

Proof. Suppose that we have a distinguished subset $\mathcal{N}$ in the free Ω-algebra S_k, and as before let T be the Ω-subalgebra generated by $\mathcal{N}$ in S_k. Then the algebra R is isomorphic to the quotient of T by some ideal I. The ideal I generates in S_k an ideal $\overline{I}$ such that $I = \overline{I} \cap T$. Therefore the quotient algebra $S_k/\overline{I}$ contains a subalgebra isomorphic to T/I and hence to R. □

Theorem 2. *If some set Ω of identical relations has basis rank k over Σ, then any Ω-algebra K over Σ can be isomorphically embedded into an Ω-algebra N over Σ each of whose countable subsets is contained in a subalgebra generated by k elements.*

Proof. We will assume that the collection $\{B_\alpha\}$ of countably infinite subsets of K is well-ordered by the index variable α which ranges over some well-ordered set. Suppose we have already constructed an Ω-algebra K_α over Σ that is an extension of the algebra K such that each subset B_β, $\beta < \alpha$, is already contained in a subalgebra generated by k elements. If B_α itself lies in a subalgebra with k generators then we set $K_{\alpha+1} = K_\alpha$. Now suppose that B_α does not lie in any subalgebra of K_α with k generators. We extend the set B_α to some set Λ_α of generators of K_α. Consider the free Ω-algebra Q_α over Σ with the set of generators Λ'_α of the same cardinality as Λ_α. Then we arbitrarily select k elements a_1, a_2, $\ldots$, a_k in the set Λ'_α, and a distinguished set T_α in the free Ω-algebra $Q_{\alpha k}$ over Σ that is generated in the algebra Q_α by the elements a_i $(i = 1, 2, \ldots, k)$. Select a countably infinite subset $t_1^\alpha, \ldots, t_r^\alpha, \ldots$ of the set T_α that has a countably infinite complement in T_α. Clearly, the subalgebra $\overline{K}_\alpha$ generated in Q_α by the set

$$\overline{\Lambda}_\alpha = \{t_s^\alpha \mid s = 1, 2, \ldots\} \cup \big(\Lambda'_\alpha \setminus \{a_i \mid i = 1, \ldots, k\}\big),$$

is a free Ω-algebra, since from any relation which is non-trivial (i.e., not a consequence of Ω) we could obtain a non-trivial relation for the elements of the set T_α by replacing the generators from $\overline{\Lambda}_\alpha$ that are not in T_α by arbitrary elements of T_α. From this it follows that the algebra K_α is isomorphic to the quotient algebra of $\overline{K}_\alpha$ by some ideal I_α, where I_α can be chosen such that the images of the elements t_i^α $(i = 1, 2, \ldots, r, \ldots)$ correspond to the elements of B_α. The ideal I_α generates in Q_α some ideal $\overline{I}_\alpha$. We will prove that $\overline{I}_\alpha \cap \overline{K}_\alpha = I_\alpha$.

Let d be an arbitrary element of this intersection. Since $d \in I_\alpha$, it can be written as a (nonassociative) polynomial each of whose terms contains a factor from I_α; and since $d \in K_\alpha$, it can be written as a polynomial in elements of $\overline{\Lambda}_\alpha$. Comparing these two expressions, we obtain an equation in the free Ω-algebra Q_α over Σ, which obviously will still be valid if the free generators from the set $\Lambda'_\alpha \setminus \{a_i \mid i = 1, \ldots, k\}$ which occur in it are replaced by (distinct) elements of the set $T_\alpha \setminus \{t_s^\alpha \mid s = 1, 2, \ldots\}$. Let $\overline{I}_\alpha^0$ be the ideal of $Q_{\alpha k}$ generated by the finite set of elements obtained as a result of this replacement of the elements I_α that occur

in the expression for d, and let I_α^0 be the ideal generated by the same elements in the subalgebra generated by T_α. From the fact that T_α is a distinguished set, it follows that, after this replacement, the element d becomes an element of the ideal I_α^0. Using the fact that T_α is a set of free generators for the subalgebra it generates, we can perform the reverse replacement which gives us an expression for d as an element of I_α.

From the claim just proved it follows that the quotient algebra $Q_\alpha/\overline{I}_\alpha$ contains a subalgebra isomorphic to the algebra K_α, and under the natural embedding the set B_α is contained in a subring generated by k elements, namely the images of the elements of the set $\{a_i \mid i = 1, \ldots, k\}$. We extend the algebra K_α to the algebra isomorphic to $Q_\alpha/\overline{I}_\alpha$ and denote the extended algebra by $K_{\alpha+1}$.

If γ is a limit ordinal then by K_γ we denote the union of the increasing chain of algebras $\bigcup_{\delta<\gamma} K_\delta$.

By an obvious transfinite induction, it follows that the algebra K can be extended to an Ω-algebra K' over Σ such that every countably infinite subset of K is contained in a subalgebra of K' generated by k elements. Analogously, the algebra K' can be extended to K'' and so on. The union $N = \bigcup K^{(\gamma)}$ of this increasing chain of algebras will obviously satisfy the conditions of the theorem if γ ranges over all ordinals of the first two classes. This completes the proof. □

We now consider applications of Theorem 2 to some particular cases.

1) The set Ω is empty and Σ is an arbitrary ring.

Lemma 1. *The set* $T = \{aa^2, a^2a^2, (a^2a)a^2, [(a^2a)a]a^2, \ldots\}$ *is a distinguished subset of the free algebra* S *over* Σ *on one generator* a.

Proof. Let S_0 be the subring of S generated by T, let I_0 be an ideal of S_0, and let I be the ideal of S generated by I_0. If q is an element of the intersection $S_0 \cap I$, then q can be written as a polynomial, each of whose terms q_i is a product of an element of I_0 and some monomials in S. If at least one of the latter monomials does not belong to S_0, then from the definition of T it follows that none of the monomials obtained by expanding q_i belongs to S_0. Since $q \in S_0$, all such q_i must cancel each other, and thus $q \in I_0$. This completes the proof. □

2) The set Ω consists of the relation $xy - yx = 0$, and Σ is an arbitrary ring.

Lemma 2. *The set* $T_C = \{a^2a^2, (a^2a)a^2, [(a^2a)a]a^2, \ldots\}$ *is a distinguished subset of the free commutative algebra* S_C *over* Σ *on one generator* a.

Proof. From [3] it follows that the elements of S_C are linear combinations with coefficients from Σ of the so-called C-regular words. Taking this into account, we can complete the proof similarly to the proof of Lemma 1. □

3) The set Ω consists of the relation $xy + yx = 0$, and Σ is an arbitrary ring.

Lemma 3. *The set* $T_{AC} = \{[(ab)b](ab), \{[(ab)b]b\}(ab), \ldots\}$ *is a distinguished subset of the free anticommutative algebra* S_{AC} *over* Σ *on two generators* a *and* b.

The proof is similar, using the results of [3]. Obviously, in the general case the basis rank here is 2, but in the case when all elements of Σ have additive order 2 we will obtain a commutative algebra and thus the basis rank will be 1.

4) The set Ω consists of the relation $(xy)z - x(yz) = 0$, and Σ is an arbitrary ring.

Lemma 4. *The set $T_A = \{bab, ba^2b, ba^3b, \ldots\}$ is a distinguished subset of the free associative algebra S_A over Σ on two generators a and b.*

See the proof of this lemma in [4].

Theorem 2 and Lemmas 1–4 imply the following Theorems:

Theorem 3. *Every algebra over Σ can be embedded into an algebra over Σ in which every countably infinite subset is contained in a subalgebra generated by one element.*

This theorem generalizes a result of A.I. Zhukov [5].

Theorem 4. *Every commutative (resp. anticommutative) algebra over Σ can be embedded into a commutative (resp. anticommutative) algebra over Σ in which every countably infinite subset is contained in a subalgebra generated by one element (resp. by two elements).*

Theorem 5. *Every associative algebra over Σ can be embedded into an associative algebra over Σ in which every countably infinite subset is contained in a subalgebra generated by two elements.*

Theorem 5 generalizes a result of A.I. Malcev [2].

3. Applications to special Jordan algebras

Special Jordan rings cannot be immediately included into the scheme explained in Section 1 because there does not exist a set of identical relations defining this class of algebras (see [1], [4]). Suppose we have some algebra J over Σ. In the case when J is generated by a finite or countably infinite set is considered in [4], where it is shown that J can be embedded into a Jordan algebra with two generators. First we will assume that Σ admits division by 2.

Theorem 6. *Any special Jordan algebra J over Σ can be isomorphically embedded into a special Jordan algebra $\overline{J}$ each of whose countable subsets is contained in a subalgebra generated by two elements.*

Proof. Since the algebra J is special, it can be represented isomorphically in some associative algebra K over Σ. By Theorem 5, the algebra K is embeddable in an associative algebra N over Σ, each of whose countable subsets is contained in a subalgebra generated by two elements. Since the set T_A of Lemma 4 consists of the Jordan polynomials $t_s = ba^sb = 2b \circ (b \circ a^s) - a^s \circ b^2$, it is clear that, upon extending the algebra K, each countably infinite subset B_α that appears in the

proof of Theorem 2 will be contained in a special Jordan subalgebra generated by two elements, namely the images of the elements a and b. Obviously, we will also have to carry out this construction for the countably infinite subsets B_α that are contained in a subalgebra generated by two elements but do not consist of Jordan polynomials in these two elements. It follows that each countably infinite subset of the algebra N will be contained in a special Jordan subalgebra generated by two elements. To the associative algebra N there corresponds the special Jordan algebra $N^{(+)}$. This completes the proof. □

A question arises regarding the validity of the converse to Theorem 6: Is a Jordan algebra special if each of its countable subsets is contained in a subalgebra with two generators, that is, in a special Jordan algebra [4]? A positive answer to this question will be given in a somewhat more general form.

We will call a Jordan algebra *locally special* if each of its finitely generated subalgebras is special.

Theorem 7. *Any locally special Jordan algebra J is special.*

Proof. By $\overline{J}$ we will denote a set of elements that is in one-to-one correspondence with the algebra J. Consider the free associative algebra $\mathcal{A}$ over Σ with the set $\overline{J}$ of free generators, and its ideal $\overline{I}_1$ generated by all elements of the form[3]

$$\alpha_i = \overline{a}_{1i} + \overline{b}_{1i} - \overline{c}_{1i}, \qquad \beta_j = \frac{1}{2}\left(\overline{a}_{2j}\overline{b}_{2j} + \overline{b}_{2j}\overline{a}_{2j}\right) - \overline{c}_{2j}, \qquad \gamma_k = \sigma_k \overline{a}_{3k} - \overline{c}_{3k},$$

whenever the following equations hold in the algebra J:

$$a_{1i} + b_{1i} = c_{1i}, \qquad a_{2j} \circ b_{2j} = c_{2j}, \qquad \sigma_k a_{3k} = c_{3k}.$$

We will prove that the Jordan algebra J is isomorphically represented in the quotient algebra $\mathcal{A}/\overline{I}_1$. For this it suffices to show that $\overline{I}_1 \cap \overline{J}$ equals zero. Assume to the contrary that we have the following relation:

$$\sum_i d_i \alpha_i s_i + \sum_j r_j \beta_j t_j + \sum_k n_k \gamma_k m_k = q, \tag{1}$$

where $q \in \overline{J}$ and d, s, r, t, n, m may be absent[4]. Since (1) is a relation among the generators of a free associative algebra over Σ, it must hold in any associative algebra over Σ for arbitrary elements in one-to-one correspondence with the elements under consideration. However, to the finite set $\overline{T}$ of elements of $\overline{J}$ that occur in relation (1), there corresponds a finite set of elements T in J. By local speciality of J, there exists an associative algebra $\mathcal{A}_1$ that represents the subalgebra of J generated by the finite set T. For the elements of the set T in $\mathcal{A}_1$, the relation (1) must hold, but obviously the left-hand side of (1) vanishes in $\mathcal{A}_1$, which gives a contradiction. □

The results of Theorems 6 and 7 can be formulated as follows:

[3]It is implicit that $\sigma_k \in \Sigma$. [Translators]
[4]It is implicit that $q \neq 0$. [Translators]

Theorem 8. *If the ring Σ admits unique division by 2, then a Jordan algebra J over Σ is special if and only if it is isomorphically embeddable in a Jordan algebra each of whose countable subsets is contained in a subalgebra generated by two elements.*

If we restrict ourselves to the assumption that the additive group of the algebra J has no elements of order 2, then we can consider semispecial Jordan algebras (see [4]).

Theorem 9. *If the additive group of the Jordan algebra J has no elements of order 2, then J is semispecial if and only if J is isomorphically embeddable in a Jordan algebra each of whose countable subsets is contained in a subalgebra generated by two elements.*

Proof. From semispeciality of J it follows that there exists an associative algebra K_1 in which the algebra J is isomorphically represented by the operation $a \circ b = ab + ba$. We may assume that the additive group of the algebra K_1 does not have elements of order 2, since the collection of elements of order 2^s in the additive group of a ring is an ideal for which the corresponding quotient ring does not contain elements a satisfying $2^s a = 0$ for some natural number s. The intersection of this ideal with the set that corresponds in K_1 to the algebra J will be zero. The algebra K_1 can be extended to an algebra K with unique division by 2 ([4], Lemma 6). At the same time, the base ring Σ will be extended to $\overline{\Sigma}$ with unique division by 2.

We introduce in K a new associative operation $a \times b = 2ab$. Then the special algebra $K^{(+)}$ considered relative to the operation $a \cdot b = \frac{1}{2}(a \times b + b \times a)$ can be embedded into a Jordan algebra N over $\overline{\Sigma}$, each of whose countable subsets is contained in a subalgebra generated by two elements. Returning to the original operation $ab = \frac{1}{2}(a \times b)$, we convince ourselves that the algebra J is embedded in the desired way, since an algebra over $\overline{\Sigma}$ is also an algebra over Σ.

Conversely, suppose that J is embedded into a corresponding algebra N. Then in N ([4], Theorem 5) each finite subset is contained in a semispecial algebra. The proof can be completed by repeating the proof of Theorem 7 but replacing β_j by $\beta'_j = \overline{a}_{2j}\overline{b}_{2j} + \overline{b}_{2j}\overline{a}_{2j} - \overline{c}_{2j}$. □

4. Algebras of finite dimension

In this section we consider a refinement of the previous results for algebras of finite dimension.

Let $\mathcal{A}$ be an associative algebra of finite dimension over some field F. Obviously, $\mathcal{A}$ has a finite system of generators $\{c_i\}$ $(i = 1, 2, \ldots, m)$. Consider the free associative algebra $\mathcal{B}$ with two generators a and b over F. The subalgebra $\mathcal{B}'$ generated in $\mathcal{B}$ by the elements $c'_i = ba^i b$ $(i = 1, 2, \ldots, m, \ldots)$ is a free associative algebra with a countably infinite set of generators. Let I' be the kernel of the homomorphism of $\mathcal{B}'$ onto $\mathcal{A}$, and let I be the ideal generated by I' in $\mathcal{B}$. We may

assume[5] that $c'_{m+i} \in I'$ $(i = 1, 2, \ldots)$. Further, let I_1 be the ideal of $\mathcal{B}$ generated by the elements of the form aba, b^k, a^{m+k-2} $(k = 3, 4, \ldots)$. We will prove the equality

$$(I + I_1) \cap \mathcal{B}' = I'.$$

Suppose that the element $i + i_1$ in the sum of ideals $I + I_1$ is an element of $\mathcal{B}'$. From the obvious relation $I_1 \cap \mathcal{B}' \subseteq I$ it follows that $i_1 \in I$, and thus $i + i_1 \in I$. The proof can be completed using the equality $I \cap \mathcal{B}' = I'$, which is proved in the work [4]. The quotient algebra $\overline{\mathcal{B}} = \mathcal{B}/(I + I')$ therefore contains a subalgebra isomorphic to $\mathcal{A}$, and is clearly an algebra with two generators. In addition to the cosets corresponding to the elements of $\mathcal{A}$, the algebra $\overline{\mathcal{B}}$ contains only a finite number of linearly independent cosets, and thus is an algebra of finite dimension. Thus, we have proved the following result:

Theorem 10. *Any associative algebra $\mathcal{A}$ of finite dimension over a field F is isomorphic to a subalgebra of an associative algebra $\mathcal{B}$ with two generators and finite dimension over F.*

Now consider a special Jordan algebra J of finite dimension with basis a_1, a_2, …, a_n. Let A be an associative algebra in which J is represented. It is easy to show that the subalgebra of A generated by the elements $a_1, a_2, \ldots, a_n$ has finite dimension. Indeed, any product of elements a_i that contains more than n factors can be represented as a linear combination of such products with a smaller number of factors. This follows from the equation $a_i \circ a_j = \frac{1}{2}(a_i a_j + a_j a_i) = \sum c_{ij}^k a_k$, because in the product under consideration there will be at least two equal factors, which by a sequence of transpositions can be moved next to each other, and then it will become possible to decrease the number of factors. From this, using the proof of Theorem 10, the next result follows.

Theorem 11. *Any special Jordan algebra J of finite dimension over the field F of characteristic $\neq 2$ is isomorphic to a subalgebra of a special Jordan algebra $\overline{J}$ with two generators and finite dimension over F.*

Remark 1. Special Jordan algebras over fields of characteristic 2 are Lie algebras. It is known that a Lie algebra of finite dimension can be represented in an associative algebra of finite dimension; the question of the number of generators has not been considered for this case.

Remark 2. Without changing the methods, it is easy to obtain analogues of Theorems 10 and 11 for finite algebras, or algebras of finite rank over some ring of coefficients.

References

[1] P.M. Cohn, *On homomorphic images of special Jordan algebras*, Canadian J. Math. 6 (1954) 253–264.

[5] It is also implicit that c'_i maps to c_i for $i = 1, \ldots, m$. [Translators]

[2] A.I. Malcev, *On a representation of nonassociative rings*, Uspekhi Mat. Nauk N.S. 7 (1952) 181–185.

[3] A.I. Shirshov, *Subalgebras of free commutative and free anticommutative algebras*, Mat. Sbornik 34 (1954) 81–88.

[4] A.I. Shirshov, *On special J-rings*, Mat. Sbornik 38 (1956) 149–166.

[5] A.I. Zhukov, *Reduced systems of defining relations for nonassociative algebras*, Mat. Sbornik 27 (1950) 267–280.

On some Nonassociative Nil-rings and Algebraic Algebras

A.I. Shirshov

1. Introduction

In the works of Levitzki [5] and Jacobson [3] devoted to the solution of the problem of Kurosh [4], it is proved that any associative algebraic algebra of bounded degree is locally finite, and that every associative nil-ring of bounded index is locally nilpotent. The problem of Kurosh can be stated for any class of power associative algebras [1], but already Lie algebras give an example showing that the problem does not have a positive solution for arbitrary power associative algebras.

In the present paper, a positive solution is given for the analogous problem (also in the bounded case) for special Jordan algebras and for alternative algebras, under a natural restriction on the characteristic of the base field (Theorems 4 and 8). For nil rings, results generalizing the theorem of Levitzki in the associative case are also obtained (Theorems 2 and 7).

2. Preliminary results

Consider the associative words formed from the elements of some finite ordered set of symbols:

$$R = \{a_i\}\ (i = 1, 2, \ldots, k), \qquad a_i > a_j \ \text{ if } \ i > j.$$

Definition 1. A word of the form[1] $\alpha = a_k \cdots a_k a_{i_1} a_{i_2} \cdots a_{i_s}$ where $i_t \neq k$ $(t = 1, 2, \ldots, s)$, $s \geq 1$ will be called a_k-*irreducible*.

Definition 2. A representation of the word β (if possible) as the product of a number of a_k-irreducible words will be called an a_k-*factorization* of the word β.

Mat. Sbornik N.S. 41 (83), (1957), no. 3, 381–394.

[1]It is implicit that a_k occurs at least once. [Translators]

It is easy to see that for a word β there exists an a_k-factorization, which is moreover unique, if and only if β starts with the symbol a_k and ends with a symbol different from a_k. The following is an example of an a_3-factorization of a word on three symbols:

$$(a_3a_3a_2a_1a_1a_2a_1)(a_3a_1)(a_3a_3a_3a_1a_1a_2)(a_3a_1a_2).$$

On the set of all associative words in the elements of the set R, we introduce a partial order: for words α and β of the same length we declare $\alpha > \beta$ if this relation holds in the lexicographical sense. We order lexicographically the set T of all a_k-irreducible words; when the word α is the beginning of the word β (i.e., $\beta = \alpha a_{i_1}a_{i_2}\cdots a_{i_m}$, $i_t \neq k$, $t = 1, 2, \ldots, m$) we declare $\alpha > \beta$.

Definition 3. The associative word γ will be called *n-decomposable* if it can be represented as the product of n subwords in such a way that, for any non-identity permutation of these subwords, the resulting associative word is strictly less than γ.

For example, the word $a_3a_1a_2a_2a_1a_1a_2a_1a_1a_1$ is 3-decomposable and admits several 3-decompositions:

$$(a_3a_1)(a_2a_2a_1a_1)(a_2a_1a_1a_1), \qquad (a_3a_1a_2)(a_2a_1a_1)(a_2a_1a_1a_1),$$
$$(a_3)(a_1a_2a_2a_1a_1a_2)(a_1a_1a_1), \qquad \text{etc.};$$

the word $a_1a_2a_1a_3a_2a_1a_2a_3a_2$ is not 2-decomposable.

The words that admit a_k-factorization can be considered as words formed from the elements of the set T. In this case also, it makes sense to consider n-decomposable words. In the rest of the paper, where it could lead to confusion, we will speak about n_R-decomposable or n_T-decomposable words, specifying which set of symbols is to be regarded as generators.

When we consider words formed from elements of the set T, we will call them *T-words* (as opposed to *R-words*); analogously, we will use the terms *T-length* and *R-length*.

Lemma 1. *For any associative T-word α, n_T-decomposability implies n_R-decomposability.*

Proof. Let $\alpha = \alpha_1\alpha_2\cdots\alpha_n$ be an n_T-decomposition of α; then α, α_1, …, α_n admit an a_k-factorization. From Definition 3 it follows that $\alpha > \alpha_{i_1}\alpha_{i_2}\cdots\alpha_{i_n}$ in the sense of the set T whenever $(i_1, i_2, \ldots, i_n)$ is a non-identity permutation of the symbols $1, 2, \ldots, n$. It is easy to see that this relation also holds in the sense of the set R. Therefore, this n_T-decomposition is also an n_R-decomposition. The lemma has been proved. □

Lemma 2. *If the word α is $(n-1)_T$-decomposable, then the word αa_k is n_R-decomposable.*

Proof. Lemma 1 implies the existence of the following $(n-1)_R$-decomposition of the word α:

$$\alpha = (a_k a_{i_1}\cdots a'_{i_1})(a_k a_{i_2}\cdots a'_{i_2})\cdots(a_k a_{i_{n-1}}\cdots a'_{i_{n-1}}),$$

where $a, a' \in R$, $a'_{i_t} \neq a_k$ $(t = 1, 2, \ldots, n-1)$. We will prove that for the word αa_k we have the following n_R-decomposition:

$$\alpha a_k = (a_k)(a_{i_1} \cdots a'_{i_1} a_k)(a_{i_2} \cdots a'_{i_2} a_k) \cdots (a_{i_{n-1}} \cdots a'_{i_{n-1}} a_k).$$

Indeed, any permutation of the factors of αa_k that fixes the first factor a_k, transforms αa_k into $\alpha' a_k$ where α' is obtained by some permutation of the factors in the given $(n-1)_T$-decomposition of α. Therefore, $\alpha > \alpha'$ and $\alpha a_k > \alpha' a_k$. Now, if we consider permutations that move the symbol a_k from the first position, then it is obvious that the result of applying such a permutation to αa_k will start with a strictly smaller number of symbols a_k as compared to αa_k. Thus, it will be strictly less than αa_k. The lemma has been proved. □

Lemma 3. *For any three natural numbers k, s, n there exists a natural number $N(k, s, n)$ such that in any associative word of length $N(k, s, n)$ in k ordered symbols there exists either a subword repeated s times consecutively or an n-decomposable subword (or both).*

Proof. It is easy to see that the natural numbers $N(k, s, 1)$ and $N(1, s, n)$ satisfying the conditions of the lemma exist for any k, s, n. Suppose we are given some natural numbers k and n. We make the inductive assumption that there exist natural numbers $N(k-1, s, n)$ and $N(k, s, n-1)$ satisfying the conditions of the lemma for all natural numbers k and s.

Consider an arbitrary associative word α of length

$$[s + N(k-1, s, n)] \left[N\left(k^{N(k-1,s,n)+s}, s, n-1\right) + 1 \right],$$

in elements of our familiar set R. If, at the beginning of α there is a number of the symbols a_i other than a_k, and their number is not less than $N(k-1, s, n)$, then we can apply the inductive hypothesis to the subword α' that is at the beginning of the word α and depends only on $k-1$ symbols. Therefore, we may assume that the length of the word α' (if it exists) is less than $N(k-1, s, n)$. At the end of the word α there may be a subword $\alpha'' = a_k a_k \cdots a_k$. We may suppose that if α'' exists then its length is less than s, for in the contrary case the conclusion of the lemma would hold. Removing the words α' and α'' (if they exist) we obtain a subword α_1 whose length is greater than

$$[s + N(k-1, s, n)] N\left(k^{N(k-1,s,n)+s}, s, n-1\right).$$

Having performed a_k-factorization of the word α_1, we may assume in addition that the length of each a_k-irreducible word that occurs in this a_k-factorization is less than the number $s + N(k-1, s, n)$, for in the contrary case such a word would contain either s consecutive symbols a_k or a subword of length $N(k-1, s, n)$ not containing the symbol a_k. It is easy to see that there exist no more than $k^{N(k-1,s,n)+s}$ distinct a_k-irreducible words with the above-mentioned restriction on length. We will regard the word α_1 as a T-word. Since its T-length is strictly greater than $N(k^{N(k-1,s,n)+s}, s, n-1)$, in α_1 there exists either a subword repeated s times consecutively or an $(n-1)_T$-decomposable subword β.

If this second alternative holds, then by the strict inequality for the length of α_1 we may assume that the subword β is immediately followed by the symbol a_k. By Lemma 2 the subword βa_k is n_R-decomposable. In this case, as well as obviously in the case when the first alternative holds, the conclusion of the lemma is true. Therefore we set

$$N(k,s,n) = [s + N(k-1,s,n)]\left[N\left(k^{N(k-1,s,n)+s}, s, n-1\right) + 1\right].$$

This completes the proof of Lemma 3. □

Definition 4. An element b of the free associative ring $\mathcal{A}$ with the set R of generators will be called a *Jordan polynomial* if there exists a natural number t such that the element $2^t b$ can be represented as a polynomial in the elements of R with respect to addition and the Jordan multiplication $a \circ b = ab + ba$.

For example, the element $a_1a_2a_1$ is a Jordan polynomial in the sense of Definition 4, because $2^2 a_1a_2a_1 = 2a_1 \circ (a_1 \circ a_2) - (a_1 \circ a_1) \circ a_2$.

Definition 5. An associative word α in the elements of R will be called *special* if there exists a homogeneous Jordan polynomial b_α such that the highest word of b_α is α, and this occurs in b_α with coefficient of the form 2^t $(t = 0, 1, \ldots)$.

Lemma 4. *Every T-word α is special (relative to the set R).*

Proof. If the T-length of α equals 1, i.e.,

$$\alpha = a_k a_k \cdots a_k a_{i_1} a_{i_2} \cdots a_{i_m} \ (i_r \neq k; r = 1, 2, \ldots, m),$$

then $b_\alpha = [\cdots[\cdots(a_k \circ a_k) \circ \cdots \circ a_k] \circ a_{i_1}] \circ \cdots a_{i_m}$.

Suppose the statement of the lemma has been proved for T-words whose T-length is strictly less than the T-length of α (which is greater than 1). Then,

$$\alpha = \beta a_k a_k \cdots a_k a_{i_1} a_{i_2} \cdots a_{i_m} \ (i_r \neq k; r = 1, 2, \ldots, m),$$

where β is a T-word to which the inductive hypothesis applies. Let b_β be a Jordan polynomial that corresponds to β. Then a simple calculation shows that we may take as b_α the Jordan polynomial

$$\begin{aligned}
&[\cdots[[b_\beta \circ [[\cdots(a_k \circ a_k) \circ \cdots \circ a_k] \circ a_{i_1}]] \circ a_{i_2}] \cdots] \circ a_{i_m} \\
+\ &[\cdots[[[b_\beta \circ [\cdots(a_k \circ a_k) \circ \cdots \circ a_k]] \circ a_{i_1}] \circ a_{i_2}] \cdots] \circ a_{i_m} \\
-\ &[\cdots[[(b_\beta \circ a_{i_1}) \circ [\cdots(a_k \circ a_k) \circ \cdots \circ a_k]] \circ a_{i_2}] \cdots] \circ a_{i_m}.
\end{aligned}$$

The lemma has been proved. □

Consider an arbitrary algebra K which will in general be nonassociative. Let Γ be a subsemigroup of the additive group of K, and suppose that the elements of Γ satisfy the following homogeneous identical relation (in K):

$$f(\gamma_1^{p_1}, \gamma_2^{p_2}, \ldots, \gamma_k^{p_k}) = 0, \text{ for all } \gamma_i \in \Gamma.$$

Here, $f(x_1^{p_1}, x_2^{p_2}, \ldots, x_k^{p_k})$ denotes a (nonassociative) homogeneous polynomial in the variables x_i $(i = 1, 2, \ldots, k)$, in each of whose monomials x_i occurs p_i times,

and whose coefficients can be taken to be elements of an arbitrary associative ring[2] Σ.

Definition 6. By the *multilinear polynomial*

$$\overline{f}(x_{11},\dots,x_{1p_1},x_{21},\dots,x_{2p_2},\dots,x_{k1},\dots,x_{kp_k}),$$

corresponding to the polynomial

$$f(x_1^{p_1},x_2^{p_2},\dots,x_k^{p_k}),$$

we mean the polynomial obtained from f by first replacing each variable x_i by one of the variables x_{is} so that in each monomial exactly one x_{is} occurs, and then summing over all permutations of the symbols $x_{i1},x_{i2},\dots,x_{ip_i}$ for all $i=1,2,\dots,k$.

For example, if

$$f(x_1^3,x_2^2)=[(x_1x_2)x_1](x_1x_2),$$

then

$$\begin{aligned}&\overline{f}(x_{11},x_{12},x_{13},x_{21},x_{22})=\\&[(x_{11}x_{21})x_{12}](x_{13}x_{22})+[(x_{11}x_{22})x_{12}](x_{13}x_{21})+[(x_{11}x_{21})x_{13}](x_{12}x_{22})\\&+[(x_{11}x_{22})x_{13}](x_{12}x_{21})+[(x_{12}x_{21})x_{11}](x_{13}x_{22})+[(x_{12}x_{22})x_{11}](x_{13}x_{21})\\&+[(x_{12}x_{21})x_{13}](x_{11}x_{22})+[(x_{12}x_{22})x_{13}](x_{11}x_{21})+[(x_{13}x_{21})x_{11}](x_{12}x_{22})\\&+[(x_{13}x_{22})x_{11}](x_{12}x_{21})+[(x_{13}x_{21})x_{12}](x_{11}x_{22})+[(x_{13}x_{22})x_{12}](x_{11}x_{21}).\end{aligned}$$

Lemma 5. *For arbitrary elements γ_{ij} $(i=1,2,\dots,k;j=1,2,\dots,p_i)$ of the semigroup Γ in K, the following relation holds:*

$$\overline{f}(\gamma_{11},\dots,\gamma_{1p_1},\gamma_{21},\dots,\gamma_{2p_2},\dots,\gamma_{k1},\dots,\gamma_{kp_k})=0.$$

Proof. Suppose that

$$p_1=p_2=\dots=p_{s-1}=1,\ p_s>1.$$

From the properties of the semigroup Γ it follows that the polynomial

$$\begin{aligned}&f\big(x_1,x_2,\dots,x_{s-1},(x_{s1}+x_{s2}+\dots+x_{sp_s})^{p_s},x_{s+1}^{p_{s+1}},\dots,x_k^{p_k}\big)\\&-\sum_{q=1}^{p_s}f\big(x_1,x_2,\dots,x_{s-1},\big[\big(\sum_{j=1}^{p_s}x_{sj}\big)-x_{sq}\big]^{p_s},x_{s+1}^{p_{s+1}},\dots,x_k^{p_k}\big)\\&+\sum_{p_s\ge q_1>q_2\ge 1}f\big(x_1,x_2,\dots,x_{s-1},\big[\big(\sum_{j=1}^{p_s}x_{sj}\big)-x_{sq_1}-x_{sq_2}\big]^{p_s},x_{s+1}^{p_{s+1}},\dots,x_k^{p_k}\big)\\&-\dots+(-1)^{p_s-1}\sum_{t=1}^{p_s}f\big(x_1,x_2,\dots,x_{s-1},x_t^{p_s},x_{s+1}^{p_{s+1}},\dots,x_k^{p_k}\big),\end{aligned}$$

vanishes, if the variables are replaced by arbitrary elements of Γ.

[2]It is implicit that Σ is also commutative. [Translators]

Now a simple calculation performed for each monomial of the polynomial f shows that the polynomial above is linear in each variable x_{si} $(i = 1, 2, \ldots, p_s)$, and can be obtained from f by first replacing each occurrence of the variable x_s in each term by one of the variables x_{si} so that in each monomial of f exactly one x_{si} occurs, and then summing over all permutations of the symbols $x_{s1}, x_{s2}, \ldots, x_{sp_s}$. Performing this construction consecutively for all s from 1 to k, we obtain the desired result. This completes the proof of the lemma. □

Remark. This rather simple statement, in a weaker formulation, has appeared many times already in algebraic papers, but usually it was proved for algebras over a field (see, for example, [6]) with some restrictions on the field.

3. Semispecial Jordan rings and algebras

Consider a semispecial Jordan ring I, i.e., a ring embeddable in some associative ring $A_0(I)$ such that the set of elements corresponding to the elements of I forms a Jordan ring isomorphic to I with respect to addition and the Jordan multiplication $a \circ b = ab + ba$. If, for some element c of I there exists a natural number $n(c)$ such that $c^{n(c)-1} \neq 0$ and $c^{n(c)} = 0$, then we will call c as usual a *nilpotent element of index* $n(c)$.

Definition 7. If all elements of the ring I are nilpotent, and their indices are uniformly bounded, then we will say that I is a *Jordan nil-ring of bounded index.*

Definition 8. An arbitrary ring S is called *nilpotent* if there exists a natural number $N(S)$ such that the product of any $N(S)$ elements of S, with any arrangement of brackets, is equal to zero.

Theorem 1. *Any semispecial Jordan nil-ring of bounded index with finitely many generators, and without elements of order* 2 *in the additive group, is nilpotent.*

Definition 9. The intersection of all subrings of $A_0(I)$ that contain I will be called an *enveloping associative ring* $A(I)$ of the semispecial Jordan ring I.

It is easy to see that the enveloping ring $A(I)$ is the subring generated in $A_0(I)$ by an arbitrary set of generators of I.

The validity of Theorem 1 will follow from Theorem 2, which generalizes a theorem of Levitzki [5] (generally speaking).

Theorem 2. *Any enveloping associative ring $A(I)$, without elements of order 2 in the additive group, of a semispecial Jordan nil-ring I of bounded index with a finite number of generators, is nilpotent.*

Proof. Suppose the ring I has $R = \{a_i\}$ $(i = 1, 2, \ldots, k)$ as a set of generators. We will regard the same set R as a set of generators of $A(I)$. We will carry out the proof of Theorem 2 by induction, assuming its validity in the case when the number of generators of I equals $k - 1$.

Consider an arbitrary R-word α of length $m[N(M,n,n)+2]$ where m is the maximal length of nonzero a_k-irreducible words (here we are using the inductive hypothesis), M is the number of such words, and n is a bound on the indices of the elements of I. Then, in the word α we can find a subword β that is a T-word and has T-length equal to $N(M,n,n)$.

By Lemma 3, in the word β there exists either a subword repeated n times consecutively, or an n-decomposable subword γ. We will consider both possibilities one after the other.

1. $\beta = \beta_1 \underbrace{\gamma\gamma\cdots\gamma}_{n \text{ times}} \beta_2$. By Lemma 4, the word γ is special. Thus, there exists a natural number p such that $2^p\gamma$ is the highest term of a Jordan polynomial b_γ. Since $b_\gamma^n = 0$, the element $2^{np}\beta$ can be written as a linear combination with integral coefficients of words of the same R-length that strictly precede the word β. Therefore $2^{np}\alpha$ can also be expressed in a similar way.

2. $\beta = \beta_1\gamma_1\gamma_2\cdots\gamma_n\beta_2$. The elements of I form a subgroup of the additive subgroup of $A(I)$. By Lemma 5, the relation $x_1^n = 0$ that holds in $A(I)$ for the elements of I, implies the relation $\sum_p x_{i_1}x_{i_2}\cdots x_{i_n} = 0$, where the summation extends over all permutations $(i_1, i_2, \ldots i_n)$ of the symbols $1, 2, \ldots, n$.

By Lemma 4, the elements γ_i are special, and thus, up to a factor of the form 2^s, each element γ_i is the highest term of a Jordan polynomial b_{γ_i}.

Using the definition of n-decomposition, and the defining property of the Jordan polynomials b_{γ_i}, we see that the relation $\sum_p b_{\gamma_{i_1}} b_{\gamma_{i_2}} \cdots b_{\gamma_{i_n}} = 0$ implies that the element $2^s\beta$, for some non-negative integer s, can be expressed as a linear combination with integral coefficients of words preceding β. Therefore, $2^s\alpha$ can also be expressed in a similar way.

Thus, we have arrived at the conclusion that either $\alpha = 0$ or the element $2^s\alpha$ can be expressed as a linear combination with integral coefficients of words that have the same R-length as α but precede α. Since a decreasing sequence of words of the same length must terminate, it follows that for some non-negative integer s_1, we have the equality $2^{s_1}\alpha = 0$; the absence of elements of order 2 in the additive group of $A(I)$ implies that $\alpha = 0$. Theorems 2 and 1 have been proved. □

Without changing the notation, we will now assume that I is a special algebraic Jordan algebra over a field F of characteristic different from 2 and that the degrees of the elements of I are bounded by n. In other words, each element of I is a root of some (associative) polynomial of degree n in one variable x with coefficients from F (compare [4]).

Let $P_t(x_1, x_2, \ldots, x_t) = \sum \pm x_{i_1}x_{i_2}\cdots x_{i_t}$ be the alternating sum of the $n!$ terms that are obtained from the product $x_1x_2\cdots x_t$ by all possible permutations of the factors; the sign of each term depends on the parity (even + or odd −) of the corresponding permutation.

Lemma 6. *For any elements $a, b_1, b_2, \ldots, b_{n-1}$ of the algebra I, the following equation holds in any enveloping associative algebra $A(I)$:*

$$P_{2n-1}(a, a^2, \ldots, a^n, b_1, b_2, \ldots, b_{n-1}) = 0.$$

Proof. It is easy to see that each alternating sum P_t of the above form equals zero if any two of the arguments are equal. On the other hand, by assumption, for each element $a \in I$ there exist elements $\delta_i(a) \in F$ such that

$$a^n = \delta_1(n)a^{n-1} + \delta_2(n)a^{n-2} + \cdots .$$

To complete the proof of Lemma 6, we substitute the above expression for a_n into the left-hand side of the desired equation. □

The equation just proved is not trivial, i.e., it does not hold in all associative algebras. Indeed, the term $ab_1a^2b_2 \cdots a^{n-1}b_{n-1}a^n$, for example, appears only once.

Theorem 3. *Any enveloping associative algebra $A(I)$ of a special algebraic Jordan algebra I of bounded degree over a field F of characteristic $\neq 2$, is locally finite, i.e., each finite subset of its elements generates a subalgebra of finite dimension.*

Proof. Any subalgebra $A_Q(I)$ of $A(I)$, that has a finite number of generators, is contained in a subalgebra $A_R(I)$ whose set of generators $R = \{a_i\}$ $(i = 1, 2, \ldots, k)$ consists of elements of I. We will prove the finiteness of the dimension of $A_R(I)$ by induction on k. Assume that subalgebras generated in $A(I)$ by $k-1$ elements of I have finite dimension. Then, there exists a natural number m such that each word of length $\geq m$ formed from elements of the set $R' = R \setminus \{a_k\}$ can be expressed as a linear combination of words of smaller length. Consider an R-word α of length

$$(m+n)\left[N\left(M, n, \frac{1}{2}(n^2+3n-2)\right)+1\right],$$

where M is the number of distinct a_k-irreducible words that cannot be represented as linear combinations of R-words of smaller R-length, and n is a bound on the degrees of the elements of I. Theorem 3 will be proved if we can show that α can be represented as a linear combination of words of smaller R-length.

If $\alpha = \alpha'\beta\alpha''$ where α' is an R'-word, β is a T-word, and $\alpha'' = a_k a_k \cdots a_k$, then we may assume that the R-lengths of the words α' and α'' are less than (respectively) m and n, because in the contrary case there would be nothing to prove. Then, the R-length of β is greater than

$$(m+n)N\left(M, n, \frac{1}{2}(n^2+3n-2)\right).$$

We may assume that each of the a_k-irreducible words that occur in β cannot be represented as a linear combination of R-words of smaller R-length. Each such word has R-length less than $m+n$. Thus, the T-length of β is greater than

$$N\left(M, n, \frac{1}{2}(n^2+3n-2)\right).$$

By Lemma 3 we can claim that in the word β there exists either a T-subword repeated n times consecutively or a $\frac{1}{2}(n^2+3n-2)$-decomposable T-subword. We will consider separately both possibilities.

1. $\beta = \beta_1\gamma_1\gamma_1\cdots\gamma_1\beta_2$. Using the algebraicity of the Jordan polynomial b_{γ_1} defined analogously to how it was done in the proof of Theorem 2, we obtain an expression of β (and thus also of α) as a linear combination of preceding words (in the sense of lexicographical order) and words of smaller R-length.

2. $\beta = \beta_1\gamma_1\gamma_2\cdots\gamma_{n'}\beta_2$ for $n' = \frac{1}{2}(n^2+3n-2)$. By Lemmas 6 and 5, for the Jordan polynomials $b_{\gamma_{i_s}}$ $(s = 1, 2, \ldots, n')$ we have a non-linear relation of degree $n' = (1+2+\cdots+n)+(n-1)$. From this it follows that the product $b_{\gamma_1}b_{\gamma_2}\cdots b_{\gamma_{n'}}$ can be expressed as a linear combination of products obtained from it by permuting the factors. As in the proof of Theorem 2, we conclude that it is possible to express the word β, and thus also α, as a linear combination of preceding words.

Applying the argument repeatedly to the words produced, we will obtain in the end an expression of α as a linear combination of words of smaller R-length. This completes the proof of the theorem. □

The following result is an obvious consequence of Theorem 3:

Theorem 4. *Any special algebraic Jordan algebra of bounded degree, over a field F of characteristic $\neq 2$, is locally finite.*

Remark 1. The question remains open, whether we can remove the hypotheses of semispeciality and speciality in Theorems 1 and 4. However, from the work of the present author [8] and Theorems 1 and 4, it follows that any two elements generate a nilpotent subring (respectively a subalgebra of finite dimension) in a Jordan nil-ring of bounded index (respectively in an algebraic Jordan algebra of bounded degree) under the same restriction on the additive group of the ring (respectively the characteristic of the base field of the algebra).

Remark 2. The restrictions on the additive group (respectively on the characteristic of the field) are essential, as shown already by the example of the free Lie algebra on two generators over a field of characteristic 2, which is a semispecial Jordan algebra by Birkhoff and Witt (see [2] and [9]) but has infinite dimension.

4. Right alternative and alternative rings and algebras

It is well known that a ring S is called *right alternative* (respectively, *left-alternative*) if for any two elements a and b we have $(ab)b = a(bb)$ (respectively, $b(ba) = (bb)a$). A ring that is simultaneously right and left alternative is called *alternative*. It is known [7] that, under the operation of addition and Jordan multiplication $a \circ b = ab + ba$, the elements of a right alternative ring form a semispecial Jordan ring. Lemma 5 implies the following well-known multilinear relation:

$$(ab)c + (ac)b = a(bc) + a(cb), \tag{1}$$

for all elements a, b and c of a right alternative ring S.

Fixing a set of generators R of the right alternative ring S and assuming that S has no elements of order 2 in its additive group, we transfer Definition 4 to the elements of the right alternative ring. For example, the element $(a_1a_2)a_1$ is a Jordan polynomial since using relation (1) we can easily verify that

$$2^2(a_1a_2)a_1 = 2a_1 \circ (a_2 \circ a_1) - (a_1 \circ a_1) \circ a_2;$$

on the other hand, the element $a_1(a_2a_1)$ is not a Jordan polynomial.

Definition 10. Let $a_{i_1}a_{i_2}\cdots a_{i_s}$ be an associative word in the elements of the set R. Then we set

$$\langle a_{i_1}a_{i_2}\cdots a_{i_s}\rangle = \{\cdots[(a_{i_1}a_{i_2})a_{i_3}]a_{i_4}\cdots\}a_{i_s}.$$

We extend the operation $\langle\ \rangle$ to nonassociative words by ignoring the existing arrangement of parentheses, and then to linear combinations of those words. For example,

$$\langle\,(ab)(cd) + m[n(pq)]\,\rangle = [(ab)c]d + [(mn)p]q.$$

We will indicate by a bar over some subword or element that this subword or element is considered as a generator and is not subjected to change. For example,

$$\langle\,[\,(ab)\,\overline{(cd)}\,]\,(mn)\,\rangle = \{[(ab)(cd)]m\}n = \langle\,\overline{[(ab)(cd)]}\,(mn)\,\rangle,$$

and

$$\langle\,[\,(ab)\,(cd)\,]\,\overline{(mn)}\,\rangle = \{[(ab)c]d\}(mn).$$

If an element q of the free nonassociative ring on the set of generators R lies in the ideal generated by the element of the form $(ab)b - a(bb)$, then obviously $\langle q\rangle = 0$.

Lemma 7. *If m is an element of a right alternative ring S that has no elements of order 2 in the additive group, and d is a Jordan polynomial, then we have the equation:*

$$md = \langle\overline{m}d\rangle.$$

Proof. By the last remark, it suffices to prove the lemma under the assumption that d is a Jordan monomial, i.e., it can be written as the Jordan product of some factors from R.

If d has degree 1, i.e., is an element of R, then there is nothing to prove. Suppose d has degree $n > 1$, and for lower degrees the statement has already been proved. Then $d = d_1 \circ d_2$, where d_1 and d_2 are Jordan monomials to which we can apply the inductive hypothesis. Then

$$\begin{aligned} md &= m(d_1d_2) + m(d_2d_1) = (md_1)d_2 + (md_2)d_1 = \langle\overline{m}d_1\rangle d_2 + \langle\overline{m}d_2\rangle d_1 \\ &= \langle(\overline{m}d_1)d_2\rangle + \langle(\overline{m}d_2)d_1\rangle = \langle\overline{m}(d_1 \circ d_2)\rangle = \langle\overline{m}\,d\rangle. \end{aligned}$$

Here we have used equation (1), the inductive hypothesis, and the linearity of the operation $\langle\ \rangle$. This completes the proof. □

Lemma 7 and the fact proved above that the element $(a_1a_2)a_1$ is a Jordan polynomial, imply under our assumptions the following well-known equation:

$$a[(bc)b] = [(ab)c]b, \tag{2}$$

for all elements a, b and c of S.

Lemma 8. *Under the assumptions of Lemma* 7*, we have* $d = \langle d \rangle$.

Proof. Using the method of the proof of Lemma 7 and the lemma itself, we obtain this series of equations:

$$d = d_1 \circ d_2 = d_1d_2 + d_2d_1 = \langle\, \overline{d_1}\, d_2 \,\rangle + \langle\, \overline{d_2}\, d_1 \,\rangle = \langle d_1d_2 \rangle + \langle d_2d_1 \rangle = \langle d \rangle,$$

which complete the proof. □

Definition 11. A monomial q of the free nonassociative ring with the set of generators $R = \{a_i\}$ $(i = 1, 2, \ldots, k)$ is called an r_1*-word* if $q = \langle q \rangle$. By induction we define an r_i*-word* to be an r_1-word in r_{i-1}-words. For example, the words

$$\{[(a_1a_2)a_3]a_4\}a_5 \quad \text{and} \quad (\{(a_1a_2)[(a_2a_1)a_3]\}a_4)(a_1a_3),$$

are respectively r_1- and r_2-words.

Lemma 9. *Every element* b*, of an alternative ring* C *that has no elements of order* 2 *in its additive group, can be represented as a linear combination with integer coefficients of* r_2*-words in any set* R *of generators of* C.

Proof. Obviously, it suffices to prove the lemma assuming that b is a monomial. For monomials of degree ≤ 3 the statement of the lemma is trivial. Suppose that the lemma has been proved for degrees $< n$, and the degree of the monomial b is n. Using this assumption we have:

$$b = b'b'' = \sum_i b_{2i}(c_{2i}d_{1i}), \text{ where } b_{ji}, c_{ji}, d_{ji} \text{ are } r_j\text{-words.}$$

It now suffices to prove the lemma for monomials of degree n of the form $b_{2i}(c_{2i}d_{1i})$. If the degree of the monomial $c_{2i}d_{1i}$ is less than or equal to 2, then our statement is trivially true. Let us perform a second induction, with the inductive hypothesis that the lemma is valid for monomials of degree n if the degree of the right factor is less than m.

Suppose now that the monomial $b_{2i}(c_{2i}d_{1i})$ has degree n, and that the monomial $c_{2i}d_{1i}$ has degree m where $3 \leq m < n$. First we assume that the monomial d_{1i} has degree 1. We will need these two well-known relations,

$$(ab)c + (ba)c = a(bc) + b(ac), \tag{3}$$

$$(ab)c + (cb)a = a(bc) + c(ba), \tag{4}$$

that hold for any elements a, b, c of an alternative ring. Relation (3) is the analogue of relation (1) and holds in any left alternative ring; relation (4) (flexibility) is an immediate consequence of the relations (1) and (3). Using consecutively relations

(4) and (1) we obtain:

$$b_{2i}(c_{2i}d_{1i}) = -d_{1i}(c_{2i}b_{2i}) + (b_{2i}c_{2i})d_{1i} + (d_{1i}c_{2i})b_{2i}$$
$$= d_{1i}(b_{2i}c_{2i}) - (d_{1i}b_{2i})c_{2i} + (b_{2i}c_{2i})d_{1i}.$$

We can apply the second inductive hypothesis to the monomials $(d_{1i}b_{2i})c_{2i}$ and $(b_{2i}c_{2i})d_{1i}$.

Consider the monomial $d_{1i}(b_{2i}c_{2i})$. Its factor $b_{2i}c_{2i}$ has degree $n-1$, so $b_{2i}c_{2i} = \sum_t \ell_{2t}p_{1t}$. Using consecutively relations (3) and (1), we have

$$d_{1i}(b_{2i}c_{2i}) = \sum_t d_{1i}(\ell_{2t}p_{1t}) = \sum_t [-\ell_{2t}(d_{1i}p_{1t}) + (d_{1i}\ell_{2t})p_{1t} + (\ell_{2t}d_{1i})p_{1t}]$$
$$= \sum_t [\ell_{2t}(p_{1t}d_{1i}) - (\ell_{2t}p_{1t})d_{1i} + (d_{1i}\ell_{2t})p_{1t}].$$

Since the monomial d_{1i} has degree 1, and ℓ_{jt} and p_{jt} are r_j-words, then obviously the inductive hypothesis applies to all the words that we obtain ($p_{1t}d_{1i}$ is an r_1-word, and $(d_{1i}\ell_{2t})p_{1t}$ can be expressed in terms of r_2-words, since when we right-multiply an r_2-word by an r_1-word we obtain an r_2-word by Definition 11).

Thus the lemma has been proved assuming the inductive hypotheses and making the additional assumption on d_{1i}. This provides the basis for a third induction with the hypothesis that the lemma is valid if we assume the second inductive hypothesis when the degree of d_{1i} is less than k. Suppose now that this degree is equal to k for $1 < k < m$.

Keeping the same meaning of the indices, we have

$$b_{2i}(c_{2i}d_{1i}) = b_{2i}[c_{2i}(d'_{1i}a_s)],$$

where the monomial d'_{1i} has degree $k-1$ and $a_s \in R$.

Applying Lemma 5 to relation (2), we obtain the relation

$$a[(b'c)b''] + a[(b''c)b'] = [(ab')c]b'' + [(ab'')c]b', \tag{5}$$

that holds for all elements a, b', b'', c of the ring C. Using consecutively relations (1) and (5) we obtain

$$b_{2i}[c_{2i}(d'_{1i}a_s)] = -b_{2i}[(d'_{1i}a_s)c_{2i}] + \omega_1 = b_{2i}[(c_{2i}a_s)d'_{i1}] + \omega_2 = \omega_3,$$

where the ω_i are linear combinations of monomials to which we can apply the inductive hypothesis. This completes the proof of Lemma 9. □

Definition 12. A ring S with a set of generators R is called *right nilpotent relative to* R if there exists a natural number m such that for any R-monomial d of degree $\geq m$ we have $\langle d \rangle = 0$.

Since any right alternative ring S is a power associative ring, Definition 7 of nil rings can be transferred without any change to right alternative rings.

Theorem 5. *Every right alternative nil-ring S of bounded index without elements of order 2 in the additive group is locally right-nilpotent relative to any set of generators.*

Theorem 5 is a consequence of the following result:

Theorem 6. *If the Jordan polynomials of a right alternative ring S, without elements of order 2 in the additive group and with a set of generators $R = \{a_i\}$ $(i = 1, 2, \ldots, k)$, are nilpotent of uniformly bounded index, then S is right-nilpotent relative to R.*

Proof. Let the number n be a bound on the indices of the Jordan polynomials of the ring S. In the free associative ring A with the set of generators R, we consider the ideal I_1 generated by the n-th powers of all Jordan polynomials. Theorem 2 implies that, for any monomial q of the ring A with degree $\geq m[N(M, n, n) + 2]$, there exists a natural number s_q such that $2^{s_q} q \in I_1$, i.e., $2^{s_q} q = \sum_r \ell_r c_r j_r^n d_r$, where the ℓ_r are integers, the c_r and d_r are monomials which may be absent, and the j_r are Jordan polynomials.

Since the last equation holds in the free associative ring, then in any (nonassociative) ring on the same set of generators, the following relation will be valid:

$$2^{s_q} \langle q \rangle = \sum_r \ell_r \langle c_r j_r^n d_r \rangle.$$

Using Lemma 7 and if necessary Lemma 8, we obtain that in the ring S,

$$2^{s_q} \langle q \rangle = \sum_r \ell_r \langle \langle c_r j_r^n \rangle d_r \rangle = \sum_r \ell_r \langle \langle c_r \rangle \overline{j_r^n} d_r \rangle = 0,$$

since any power of a Jordan polynomial is again a Jordan polynomial. This completes the proof of Theorems 5 and 6. □

Theorem 7. *Any alternative nil-ring S of bounded index, without elements of order 2 in the additive group, is locally nilpotent.*

Proof. Every finite set $R = \{a_i\}$ $(i = 1, 2, \ldots, k)$ in the ring S generates, by Theorem 6, a subring that is right-nilpotent relative to R. The finite set R_1 of nonzero r_1-words in S generates a subring S_1 that is right-nilpotent relative to R_1. From the proof of Lemma 9 it follows that any R-monomial q in the ring S can be written as a linear combination with integer coefficients of r_2-words of the same R-length. If q is an r_2-word of sufficiently large R-length, then from the right-nilpotency of the ring S_1 it follows that $q = 0$. This completes the proof. □

Lemma 10. *In any right alternative algebra S, which is algebraic of bounded degree over a field F of characteristic $\neq 2$ and is generated by a finite set R, there exist only finitely many linearly independent r_1-words (relative to R).*

Proof. For any Jordan polynomial j_s, the following equation holds:

$$f_s(j_s) = j_s^n + \delta_{s1} j_s^{n-1} + \delta_{s2} j_s^{n-2} + \cdots = 0,$$

where $\delta_{si} \in F$.

Consider the free associative algebra A over F with the set R of generators, and the ideal I_1 in A generated by all elements $f_s(j_s)$. From Theorem 3 it follows that the quotient algebra $\overline{A} = A/I_1$ is locally finite.

We will show that the linear dependence in $\overline{A}$ of the images of some words q_i $(i = 1, 2, \ldots, t)$ implies the linear dependence of the r_1-words $\langle q_i \rangle$ in the algebra S. The former linear dependence is equivalent to the following relation in the free associative algebra A:

$$\sum_{i=1}^{t} \mu_i q_i + \sum_s \rho_s c_s f_s(j_s) d_s = 0,$$

where $\mu, \rho \in F$ and c_s, d_s are some monomials. As in the proof of Theorem 6, we obtain that

$$\sum_{i=1}^{t} \mu_i \langle q_i \rangle + \sum_s \rho_s \langle \langle c_s \rangle \, \overline{f_s(j_s)} \, d_s \rangle = 0,$$

from which it follows that $\sum_{i=1}^{t} \mu_i \langle q_i \rangle = 0$ in the algebra A. This completes the proof. □

Theorem 8. *Any algebraic alternative algebra S of bounded degree over a field F of characteristic $\neq 2$ is locally finite.*

The validity of this theorem follows immediately from Lemmas 9 and 10.

References

[1] A.A. Albert, *Power associative rings*, Trans. Amer. Math. Soc. 64 (1948) 552–593.

[2] G. Birkhoff, *Representability of Lie algebras and Lie groups by matrices*, Annals of Math. 38 (1937) 526–632.

[3] N. Jacobson, *Structure theory for algebraic algebras of bounded degree*, Annals of Math. 46 (1945) 695–707.

[4] A.G. Kurosh, *Problems in the theory of rings related to the Burnside problem on periodic groups*, Izv. Akad. Nauk USSR, Ser. Mat. 5 (1941) 233–240.

[5] I. Levitzki, *On a problem of A. Kurosch*, Bull. Amer. Math. Soc. 52 (1946) 1033–1035.

[6] A.I. Malcev, *On algebras defined by identical relations*, Mat. Sbornik N.S. 26 (1950) 19–33.

[7] R.D. Schafer, *Representations of alternative algebras*, Trans. Amer. Math. Soc. 72 (1952) 1–17.

[8] A.I. Shirshov, *On special J-rings*, Mat. Sbornik 38 (1956) 149–166.

[9] E. Witt, *Treue Darstellung Liescher Ringe*, J. reine und angew. Math. 177 (1937) 152–160.

On Rings with Identical Relations

A.I. Shirshov

1. Introduction

The present work is a sequel to the author's paper [6]. In order to avoid repeating numerous definitions and explaining many notations, which would take up an unjustifiable amount of space, the author will limit himself to frequent references.

The first part of this work (§2) is devoted to associative rings with identical relations. In that section, Theorem 1 is proved, which establishes a local finiteness property for such rings, and consequences of that theorem which follow almost immediately are pointed out, including in particular the theorem of Kaplansky [2] on local finiteness of algebraic algebras with identical relations. In the last section (§3), a certain generalization of Kaplansky's theorem to the case of alternative rings (Theorem 5) is proved.

2. Associative rings with identical relations

First we consider associative algebras that satisfy one or more identical relations. Examples of such algebras are commutative algebras and algebras of finite dimension. These examples demonstrate the breadth of this class of associative algebras, and the importance of obtaining general theorems that hold for arbitrary associative algebras with identical relations.

It is known (see for example [4] and [6]) that one may assume that such an algebra satisfies a multilinear identity. Obviously, such an identity can always be written in the form

$$x_1 x_2 \cdots x_n = \sum_{(i_1,i_2,\ldots,i_n)} \alpha_{i_1 i_2 \cdots i_n} x_{i_1} x_{i_2} \cdots x_{i_n}, \tag{1}$$

Mat. Sbornik N.S. 43 (85), (1957), no. 2, 277–283.

where the summation on the right side is over all permutations $(i_1, i_2, \dots, i_n)$ of the symbols $1, 2, \dots, n$ other than the identity permutation $(1, 2, \dots, n)$, and all the coefficients $\alpha_{i_1 i_2 \dots i_n}$ belong to the base field.

Definition 1. If a word s in some symbols $v_1, v_2, \dots, v_r$ can be written in the form

$$s = v_{i_1}^{m_1} v_{i_2}^{m_2} \cdots v_{i_k}^{m_k} \text{ where } v_{i_\ell} \neq v_{i_{\ell+1}},$$

then the natural number k will be called the *height of the word s relative to the set $\{v_i\}$*.

Obviously, the word $s = v_1 v_1 v_2 v_1 v_1 v_2 v_1 v_1 v_2$, for example, has height 6 relative to the set $\{v_1, v_2\}$, and height 1 relative to the word $v_1 v_1 v_2$.

Definition 2. Let A be an algebra with a finite number of generators $a_1, a_2, \dots, a_\ell$. Suppose that there exists a set $t_1, t_2, \dots, t_k$ of elements which are homogeneous in each a_i such that each word s in the generators a_i is equal in A to some linear combination of words in the elements t_j that have the same content (relative to the set $\{a_i\}$) as the word s and have height (relative to the set $\{t_j\}$) less than or equal to some given number q. In this case, we will say that *the algebra A has bounded height*. If every finite subset of an algebra B generates a subalgebra of bounded height, then we will say that the algebra B has *locally bounded heights*.

Obviously, any commutative algebra has locally bounded heights.

Theorem 1. *Any associative algebra (over a field) which satisfies an identical relation of degree n has locally bounded heights (relative to some set of words whose degrees are less than n, with respect to any set of generators).*

Proof. Suppose the algebra A is generated by $a_1, a_2, \dots, a_k$. According to Lemma 3 of [6], there exists a natural number $N = N(n)$ such that for every word of length N in the generators a_i there exists either an n-decomposable subword [6, Definition 3] or a subword of the form b^{2^n}. First, we will prove that if b has length $m \geq n$ then some subword of the word b^{2^n} is itself n-decomposable; without loss of generality we may assume that the word b cannot be written in the form $\overline{b}^t$ for some $t > 1$. Using cyclic permutations of the generators, we can form m different words from the word b, namely $b = b_0$, b_1, …, b_{m-1}. We can order these words lexicographically: $b_{i_0} > b_{i_1} > \cdots > b_{i_{m-1}}$. Obviously, the word b^{2^n} can be written in the form $b^{2^n} = c b'_{i_0} b'_{i_1} \cdots b'_{i_{m-1}}$, where each of the words b'_i has b_i as an initial subword. Furthermore, it is obvious that the subword $b'_{i_0} b'_{i_1} \cdots b'_{i_{m-1}}$ is m-decomposable, and consequently, it contains an n-decomposable subword. By relation (1), every n-decomposable word can be represented as a linear combination of (lexicographically) smaller words. From this, it follows that every word of length N in the generators a_i is equal to a linear combination of words that have the same content relative to the generators but contain subwords of the form b^{2^n} where b has length $< n$.

The last remark implies that if the height of a word s is sufficiently large, and the word s does not contain n-decomposable subwords, then there exists a

subword s_1 of the form $s_1 = b^n b'$ where the lengths m, m' of the words b, b' satisfy the inequality $n > m \geq m'$ and the word b' is not an initial subword of b. Since the set of possibilities for the words b and b' is finite, it easily follows that there exists a sufficiently large natural number M such that every word $\overline{s}$ of height M relative to some set of words of length $< n$ is a linear combination of words of the same content which are lexicographically not greater than $\overline{s}$ and such that each of these words has n equal subwords of the form $s_1 = b^n b'$ (which are not necessarily consecutive). However, every such word has a subword that is n-decomposable in one of the following ways:

$$\alpha_0(b^n b'\alpha_1 b)(b^{n-1}b'\alpha_2 b^2)(b^{n-2}b'\alpha_3 b^3)\cdots(bb'\alpha_n),$$
$$\alpha_0 b^n(b'\alpha_1 b^{n-1})(bb'\alpha_2 b^{n-2})(b^2 b'\alpha_3 b^{n-3})\cdots(b^{n-1}b'\alpha_n),$$

depending on which of the words b and b' is lexicographically greater. Therefore, each word of height $\geq M$ can be written as a linear combination of words of smaller height. The proof is complete. □

Now we pass to associative algebras over an arbitrary commutative associative coefficient ring Σ.

Definition 3. An identical relation satisfied by an associative algebra over Σ will be called *admissible* if (after combining similar terms) at least one of the coefficients of a term of highest degree is equal to 1.

Theorem 2. *If an associative ring C over Σ satisfies an admissible identical relation of degree n, then the ring C has locally bounded heights (relative to some set of words whose degrees are less than n, with respect to any set of generators).*

Proof. Suppose that the distinguished term of degree n (with coefficient 1) of the identity involves the variables $x_1, x_2, \ldots, x_k$ to degrees $n_1, n_2, \ldots, n_k$ (respectively) with $\sum_{i=1}^{k} n_i = n$. Linearizing this term [6, Lemma 5] consecutively with respect to $x_1, x_2, \ldots, x_k$, we eliminate all the terms in which at least one of the x_i has degree less than n_i. (If $n_i = 1$ then we consider the relation

$$\phi(x_i, x_i') = f(x_i + x_i') - f(x_i') = 0,$$

where f is the left side of the original identity.) Since similar terms cannot appear as a result of this process, the multilinear identity thus obtained will have at least one term with coefficient 1. Performing, if necessary, a permutation of the variables, we obtain an identity of the form (1) which was used in the proof of Theorem 1. The remainder of the proof is a repetition of the proof of Theorem 1. □

We now consider some corollaries of the results already proved.

Theorem 3. *Let A be an associative ring over Σ with an admissible identical relation of degree n. If all products in A of fewer than n generators are nilpotent, then A is locally nilpotent.*

The proof of this theorem is obvious.

Corollary 1. *If all products of degree $\leq n$ of the generators of an associative algebra A of dimension n are nilpotent, then A is nilpotent.*

This statement follows from the known fact that any algebra of dimension n satisfies an identical relation of degree $n+1$: the alternating sum of all $(n+1)!$ distinct products of $n+1$ distinct variables is identically zero.

Definition 4. An element a of an associative ring A will be called *algebraic*[1] *over the subring Z_1 of the center Z of A* if there exist elements $z_i \in Z_1$ and a natural number m such that this equation holds:

$$a^m = \sum_{i=1}^{m-1} z_i a^{m-i}.$$

Definition 5. If the associative ring A has elements $b_1, b_2, \ldots, b_k$ such that for some natural number m every element $c \in A^m$ can be written in the form

$$c = \sum_{i=1}^{k} z_i b_i,$$

where the elements z_i belong to the subring Z_1 of the center Z, then A will be called *finite over Z_1*.

As in the case of algebras of finite dimension, it is obvious that a ring, which is finite over its center, satisfies an admissible identical relation. The following stronger result follows immediately from Theorem 2.

Theorem 4. *Let A be an associative ring with a finite number of generators and an admissible identical relation of degree n. If all products in A of fewer than n generators are algebraic over the subring Z_1 of the center of A, then A is finite over Z_1.*

In the special case where Z_1 is the zero subring, Theorem 4 includes Levitski's theorem [3]; it also contains a more general theorem of Kaplansky [2] (it suffices to adjoin a unit element).

3. Alternative and special Jordan rings with identical relations

In what follows, we will consider alternative rings without elements of order 2 in the additive group, satisfying some identical relation which is not a consequence of associativity.

Definition 6. An identical relation $f(x_1, x_2, \ldots, x_n) = 0$, satisfied by an alternative ring, will be called *essential* if at least one of the coefficients of the highest degree terms of the element $\langle f \rangle$ [6, Definition 9] of the free nonassociative ring in the generators x_i is equal to 1 (after combining similar terms).

[1] The current term is "integral". [Translators]

Definition 7. An identical relation $I = 0$ in a special Jordan ring will be called *admissible* if the relation $F = 0$ is admissible, where F is the associative polynomial obtained by expanding the Jordan polynomial I.

Lemma 1. *If an alternative ring K satisfies an essential identical relation, then the corresponding special Jordan ring K^+ satisfies an admissible identical relation.*

Proof. Let $f(x_1, x_2, \ldots, x_q) = 0$ be the identical relation that holds in K. If we substitute in f the monomial xy^i for x_i then we obtain an essential relation $\phi(x, y) = 0$. If $\overline{\phi}(x, y)$ is the polynomial obtained from ϕ by reversing the order of variables in each monomial, then [5, §3] the ring K satisfies the admissible essential identical relation $\psi(x, y) = \phi(x, y)\overline{\phi}(x, y) = 0$. However, the polynomial $\psi(x, y)$ is a Jordan polynomial, since $\overline{\psi} = \psi$ (see [1, 5]); we have also used the associativity of an alternative ring on two generators. Regarding $\psi(x, y) \equiv I(x, y)$ as a Jordan polynomial, we see that the Jordan polynomial $I(x, y)$ is identically zero in K^+. The proof of the lemma is complete. □

Definition 8. The *center* Z of a (nonassociative) ring T is the set of all elements $x \in T$ such that $xa = ax$ and $(xa)b = x(ab) = a(bx)$ for all elements $a, b \in T$.

It is easy to verify that Z is a subring.

Remark. When we consider the center of an alternative ring, we can, generally speaking, limit ourselves in Definition 8 to the condition $xa = ax$. We do not do this, because we do not wish to distract the reader from the main goal of this work by the details that arise.

We extend Definitions 4 and 5 to alternative rings.

Theorem 5. *Let K be an alternative ring with a finite number of generators and an essential identical relation. If Z_1 is any subring of the center for which all Jordan monomials in r_2-words of the generators are algebraic over Z_1, then K is finite over Z_1.*

Proof. Let λ be the maximal element of the set R of generators of the ring K. Consider an r_1-word w in the elements of the set R such that λ occurs consecutively fewer than $m(\lambda)$ times where $m(\lambda)$ is the degree of the element λ. We perform λ-factorization of the associative word $\overline{w}$ obtained from w by omitting parentheses, and denote by $d_\lambda(w)$ the number of λ-irreducible factors.

Suppose that in the word $\overline{w}$ there appear k distinct λ-irreducible words and $d_\lambda(w) > N(k, s, n)$ [6, Lemma 3], where n is the degree[2] of the identical relation that holds in K, and $s \geq 2n$ is the upper bound on the degrees of all r_1-words v that correspond to subwords $\overline{v}$ of $\overline{w}$ formed by λ-irreducible subwords for which $d_\lambda(v) < n$. From [6, Lemma 3] and the proof of Theorem 1 it follows that the word $\overline{w}$ has a subword u which has either the form $u = u_1 u_2 \cdots u_{n-1} u_n$ or the form

[2]The number n should denote not the degree of the identity in K but in K^+. In general it is not the same and much bigger. [Editors]

$u = (u')^s$ where u' and u_i are words formed by λ-irreducible words, $d_\lambda(u') < n$, and $u_1 u_2 \cdots u_{n-1} u_n$ is an n-decomposition of u.

In each of these cases, we can express the word w as a linear combination of r_1-words that are smaller than w, together with words with coefficients in the ring Z_1 such that the R-lengths of these latter words are strictly less than the R-length of w. Since the arguments are completely analogous, we will consider only the first case.

Let $\overline{w} = \alpha u \beta$ where $u = u_1 u_2 \cdots u_{n-1} u_n$. The words u_i (up to a scalar multiple of the form 2^t) are the maximal (associative) words of some Jordan monomials b_{u_i} [6, Lemma 4]. Thus the element

$$W = \langle \alpha b_{u_1} b_{u_2} \cdots b_{u_n} \beta \rangle = \langle \alpha \overline{b}_{u_1} \overline{b}_{u_2} \cdots \overline{b}_{u_n} \beta \rangle,$$

(recall the notation[3] from [6, §4]) has w as the leading term. According to the last lemma, there exists a Jordan polynomial $I(x_1, x_2, \ldots, x_n)$ that is identically zero in K and has the word $x_1 x_2 \cdots x_n$ as its maximal (associative) word. Since

$$\langle \overline{\alpha I(b_{u_1}, b_{u_2}, \ldots, b_{u_n})} \beta \rangle = 0, \quad \text{and}$$
$$\langle \overline{\alpha I(b_{u_1}, b_{u_2}, \ldots, b_{u_n})} \beta \rangle = \langle \alpha I(b_{u_1}, b_{u_2}, \ldots, b_{u_n}) \beta \rangle,$$

the word w is equal to a linear combination of (lexicographically) smaller r_1-words.

Using the above arguments, we carry out induction on the number of generators, and see that first, every sufficiently long λ-irreducible word is a linear combination of shorter λ-irreducible words, which justifies the introduction of the number k, and second, every sufficiently long r_1-word is a linear combination of shorter r_1-words with coefficients in Z_1. This statement immediately carries over to r_2-words which, by virtue of Lemma 9 of reference [6], completes the proof. □

Let us point out, for example, the following two corollaries of Theorem 5.

Corollary 2. *Any alternative algebraic algebra with an essential identical relation is locally finite.*

The proof of this statement follows from the sufficiently obvious fact that adjoining a unit element preserves algebraicity and also preserves the property of having an identical relation. (For example, the identical relation

$$f(x_1, x_2, \ldots, x_n) \sum (-1)^i \langle x_{i_1} x_{i_2} \cdots x_{i_n} \rangle = 0$$

holds[4], where $f(x_1, x_2, \ldots, x_n) = 0$ is the original relation and $i = (i_1, i_2, \ldots, i_n)$ ranges over all permutations of the symbols $1, 2, \ldots, n$, and $(-1)^i = \pm 1$ according to the parity of the permutation i.)

[3]The bar here has a different meaning from earlier in this proof, up to and including the first sentence of this paragraph. [Translators]

[4]In fact, this identical relation may not hold. For instance, if $f(x, y) = xxy = 0$ holds in A, then $xxy[x, y] = 0$ does not necessary hold in the algebra $A \oplus \mathbb{Z}1$ with an external unit element 1. We can consider instead the identical relation $f([x_1, y_1], \ldots, [x_n, y_n])$. [Editors]

Corollary 3. *The enveloping associative algebra, of an algebraic special Jordan algebra of characteristic* $\neq 2$ *with a finite number of generators and an identical relation, has finite dimension.*

The proof of this statement follows from the fact that the existence of an identical relation for the Jordan polynomials in the generators is sufficient to guarantee that the number of linearly independent r_1-words is finite.

References

[1] P.M. Cohn, *On homomorphic images of special Jordan algebras*, Canadian J. Math. 6 (1954) 253–264.

[2] I. Kaplansky, *Topological representation of algebras II*, Trans. Amer.Math. Soc. 68 (1950) 62–75.

[3] I. Levitski, *On a problem of A. Kurosch*, Bull. Amer. Math. Soc. 52 (1946) 1033–1035.

[4] A.I. Malcev, *On algebras defined by identical relations*, Mat. Sbornik N.S. 26 (1950) 19–33.

[5] A.I. Shirshov, *On special J-rings*, Mat. Sbornik 38 (1956) 149–166.

[6] A.I. Shirshov, *On some nonassociative nil-rings and algebraic algebras*, Mat. Sbornik 41 (1957) 381–394.

On Free Lie Rings

A.I. Shirshov

1. Introduction

Let Σ be a commutative associative ring with unit, let $R = \{a_\alpha\}$ be some set of symbols, and let $\mathfrak{A}_{\Sigma R}$ be the free associative algebra over Σ with free generating set R. In the ring $\mathfrak{A}_{\Sigma R}$, the set R generates a Lie subring $\mathfrak{A}_{\Sigma R}^{(-)}$ with respect to the operations $x \circ y = xy - yx$, addition, and scalar multiplication by elements of Σ.

If Σ is a field, then it is known that $\mathfrak{A}_{\Sigma R}^{(-)}$ is the free Lie algebra with free generating set R. This result can be derived as an immediate corollary of the Birkhoff-Witt theorem [1, 10]. Since the Birkhoff-Witt theorem cannot be generalized to algebras over an arbitrary coefficient ring [7], the question naturally arises whether the ring $\mathfrak{A}_{\Sigma R}^{(-)}$ is free for arbitrary Σ. In the present paper, a positive answer is given to this question.

For the cases when Σ is a field of characteristic 0 (and also for the case of the so-called restricted Lie algebras), a number of authors [2, 3, 4, 6, 9] concerned themselves with the problem of determining necessary and sufficient conditions under which a given element of $\mathfrak{A}_{\Sigma R}$ belongs to $\mathfrak{A}_{\Sigma R}^{(-)}$. In §3 this problem is resolved without any restrictions on the ring Σ.

Finally, in §4 it is proved that any Lie algebra over a field, with at most countable dimension, can be isomorphically embedded into a Lie algebra with two generators.

All the above-mentioned results are simultaneously proved for restricted Lie algebras.

2. Choice of basis in the ring $\mathfrak{A}_{\Sigma R}^{(-)}$

Consider the set $\mathfrak{R}$ of associative words generated by the elements of R.

Mat. Sbornik N.S. 45 (87), (1958), no. 2, 113–122.

Defining arbitrarily some total order on the set R, we partially order lexicographically the set $\mathfrak{R}$. The order relation will be undefined only for pairs of words in which one word is an initial subword of the other.

Definition 1. An associative word u is called *regular* if $u > u_2u_1$ for any factorization $u = u_1u_2$ where u_1 and u_2 are nonempty.

For example, the word $a_3a_3a_2a_3a_2a_1$ is regular because it is greater than the words $a_3a_2a_3a_2a_1a_3$, $a_2a_3a_2a_1a_3a_3$, $a_3a_2a_1a_3a_3a_2$, etc.

If u and v are regular words and $u = vv_1$, then we will define $v > u$.

Remark. If $u = vv_1$ is a regular word, then $u > v_1$ since v_1 cannot coincide with any initial subword of u.

Definition 2. A nonassociative R-word $[u]$ will be called *regular* if

(1) the associative word u, obtained by omitting the parentheses, is regular; and
(2) if $[u] = [v][w]$ then $[v]$ and $[w]$ are regular words; and
(3) if $[u] = [[v_1][v_2]][w]$ then $v_2 \leq w$.

It is easy to see that regular words are defined inductively, and that one can determine effectively whether a given nonassociative word is regular or not. We also remark that in Condition (2) it is implicit by Condition (1) that $v > w$.

Lemma 1. *In any regular associative word, one can place parentheses in one and only one way such that the resulting nonassociative word is regular.*

Proof. Suppose the lemma is proved for words whose lengths are less than n. Suppose that a given regular associative word u of length $n > 1$ contains an element $a_\beta \in R$ that is less than all the other elements of R occurring in u. Then it is obvious that the word u begins with an element of R that is greater than a_β. From Definition 2 it follows that for any placement of parentheses in the word u that results in a regular nonassociative word, only one placement of parenthesis is possible for subwords of the form $a_\gamma a_\beta a_\beta \cdots a_\beta$:

$$\{\cdots[(a_\gamma a_\beta)a_\beta]\cdots\}a_\beta \quad (a_\gamma > a_\beta).$$

Replacing in the word $[u]$ every subword of the form

$$[\cdots(a_\gamma \underbrace{a_\beta)a_\beta\cdots]a_\beta}_{k \text{ times}},$$

by the symbol a_γ^k, and setting $a_\gamma^k > a_\delta^\ell$ if either $a_\gamma > a_\delta$ or $\gamma = \delta$, $k < \ell$, we obtain a new regular word $[\overline{u}]$ in the symbols a_γ^k $(k = 0, 1, 2, \ldots)$, $a_\gamma^0 = a_\gamma$. If the conclusion of the lemma did not hold for the word $[u]$, then obviously it would not hold for the word $\overline{u}$ either. However, by the inductive hypothesis, this is impossible. The proof is complete. □

By virtue of the one-to-one correspondence between regular associative and nonassociative words that has just been established, we will retain the symbol $[u]$

to denote the regular nonassociative word corresponding to the regular associative word u.

We denote by $L_{\Sigma R}$ the free Lie algebra over Σ with free generating set R.

Lemma 2. *Every element of the free Lie algebra $L_{\Sigma R}$ over Σ with free generating set R can be represented as a linear combination of regular words with coefficients from Σ.*

Proof. Obviously, it suffices to prove the lemma only for words in the elements of the set R. Suppose a word v has length n, and that the lemma has been proved for words of smaller length. Then

$$v = uw = \sum_i \sum_k \sigma_{ik}[u_i][w_k],$$

where $\sigma_{ik} \in \Sigma$, and u_i and w_k are regular words with $u_i > w_k$. If

$$[u_i][w_k] = [[u_{i_1}][u_{i_2}]][w_k] \text{ and } u_{i_2} > w_k,$$

then obviously

$$[u_i][w_k] = [[u_{i_1}][w_k]][u_{i_2}] + [u_{i_1}][[u_{i_2}][w_k]].$$

If we now assume in addition that the associative words, obtained by omitting parentheses in the regular words that occur in the expressions for $[u_{i_1}][w_k]$ and $[u_{i_2}][w_k]$, are greater than w_k, then the proof will be complete by induction on the smallest factor. □

Lemma 3. *If we write the regular word $[v] \in \mathfrak{A}_{\Sigma R}^{(-)}$ as an element of the associative algebra $\mathfrak{A}_{\Sigma R}$, then in this expression v will appear with coefficient 1 and all other associative words that occur will be less than v.*

Proof. For words of length 1, Lemma 3 is trivially valid. Suppose it is valid for words of lengths less than n. If $[v]$ is a word of length $n > 1$ then $[v] = [u][w]$. If we denote by $\overline{t}$ the associative expression for a Lie word t, then obviously

$$\overline{[v]} = \overline{[u]}\,\overline{[w]} - \overline{[w]}\,\overline{[u]}.$$

Since $u > w$ and (by the inductive hypothesis) the maximal words in the expressions $\overline{[u]}\,\overline{[w]}$, $\overline{[w]}\,\overline{[u]}$ are equal respectively to the words uw, wu and have coefficient 1, it follows that the maximal word occurring in $\overline{[v]}$ is equal to $uw = v$ and has coefficient 1. The proof is complete. □

Theorem 1. *The rings $L_{\Sigma R}$ and $\mathfrak{A}_{\Sigma R}^{(-)}$ are isomorphic.*

Proof. Let the element $\ell \in L_{\Sigma R}$ be sent to the element $\overline{\ell}$, under the homomorphism φ of the ring $L_{\Sigma R}$ onto the ring $\mathfrak{A}_{\Sigma R}^{(-)}$ extending the correspondence between the generators. If $\ell \neq 0$, then by Lemma 2 we can assume that the element ℓ is written as a linear combination of regular words, and the coefficient σ of the maximal word $[\ell_1]$ is not zero. Then by Lemma 3, we have $\overline{\ell} \neq 0$ since the word ℓ_1 in the element $\overline{\ell}$ appears with the same coefficient σ. The proof is complete. □

In §4 we will need the following result.

Lemma 4. *Suppose that a regular associative R-word u has the form $u = \alpha\ell\beta$ where ℓ is a regular subword; the words α and β may be empty. Then in the placement of parentheses in the word $[u]$, one pair of parentheses will occur in the position $\alpha(\ell\beta_1)\beta_2$ where $\beta_1\beta_2 = \beta$ and each of the words β_1, β_2 may be empty. Furthermore, parentheses can be placed in the regular word $\ell\beta_1$ as follows:*

$$\{\cdots[(\ell\beta_1^{(1)})\beta_1^{(2)}]\cdots\}\beta_1^{(s)},$$

where $\beta_1^{(i)}$ are regular words with $\beta_1^{(1)} \leq \beta_1^{(2)} \leq \cdots \leq \beta_1^{(s)}$, and in each of the words ℓ, $\beta_1^{(i)}$ parentheses are placed in the unique way prescribed by Lemma 1, and[1] *the maximal (associative) word of the resulting expression*

$$\big(\alpha\{\cdots[(\ell\beta_1^{(1)})\beta_1^{(2)}]\cdots\}\beta_1^{(s)}\big)\beta_2,$$

is equal to u.

Proof. Let a_β be the smallest of the generators that occurs in u. If ℓ has length 1 then there is nothing to prove. Suppose that the lemma is valid if u has length less than n where $n > 1$.

If u has length n, then (as in the proof of Lemma 1) we represent it as a word in the symbols a_γ^k. Assuming that the length of ℓ is greater than 1, we note that it starts with a symbol other than a_β, and in the new representation it will be replaced by a new word ℓ_1 which, regarded as an R-word, can differ by several factors a_β appended on the right. It is easy to see that the R-word ℓ_1 will be regular. Considering the words u and ℓ_1 as words in the symbols a_γ^k, we find ourselves in a situation where we can apply the inductive hypothesis. The remainder of the argument is obvious, and this completes the proof. □

To conclude this section, we will show how Theorem 1 implies Witt's formula [10] for the rank of the homogeneous submodule of degree q in the free Lie algebra.

Definition 3. An associative word v is called *periodic* if it can be written as the product of k ($k > 1$) equal words. Two associative words u and w are called *cyclically comparable* if there exist representations $u = u_1u_2$, $w = w_1w_2$ such that $u_1 = w_2$, $u_2 = w_1$.

It is easy to see that the set of all associative words is partitioned into disjoint classes of cyclically comparable words. The following statements are trivial.

Lemma 5. *Each class of cyclically comparable non-periodic words contains one and only one regular word.*

Lemma 6. *No class of cyclically comparable periodic words contains a regular word.*

Let $\psi_q(n)$ be the rank of the submodule of homogeneous polynomials of degree q in the free Lie algebra on n generators. The number $\psi_q(n)$ coincides with

[1] The rest of this sentence has been added by the Editors.

the number of regular words of length q in n symbols. From Lemmas 5 and 6, we obtain the following equation:

$$n^q = q\,\psi_q(n) + d_1\psi_{d_1}(n) + \cdots + d_s\psi_{d_s}(n),$$

where n^q is the number of all associative words of length q in n symbols, and the d_i are the divisors of q (other than q itself). The Dedekind inversion principle[2] immediately gives Witt's formula:

$$\psi_n(q) = \frac{1}{q}\sum_{s|q}\mu(s)\,n^{q/s}$$

where $\mu(s)$ is the Möbius function.

3. Free restricted Lie rings

Suppose that the characteristic of the coefficient ring Σ is a prime number p. The associative ring $\mathfrak{A}_{\Sigma R}$ that was considered in §1 is obviously a ring of characteristic p. It is known [5] that in this case the element $(a+b)^p - a^p - b^p = \varphi(a,b)$ of the ring $\mathfrak{A}_{\Sigma R}$ is a Lie polynomial in the elements a and b.

Definition 4. A Lie algebra L over Σ in which a unary operation $x^{[p]}$ is defined is called a *restricted Lie algebra* if

$$(a+b)^{[p]} = a^{[p]} + b^{[p]} + \varphi(a,b), \quad a\cdot b^{[p]} = [\cdots(a\cdot\underbrace{b)\cdot b\cdots]b}_{p\text{ times}}, \quad (\sigma a)^{[p]} = \sigma^p a^{[p]},$$

for all elements $a, b \in L$ and $\sigma \in \Sigma$.

Obviously any associative algebra over Σ becomes a restricted Lie algebra with respect to addition and the operations $a \circ b = ab - ba$, $a^{[p]} = a^p$. In the free ring $\mathfrak{A}_{\Sigma R}$, the set R generates a restricted Lie algebra $\mathfrak{A}^{(p)}_{\Sigma R}$. We introduce the following notation:

$$x^{[p^k]} = [\cdots(x\underbrace{^{[p]})^{[p]}\cdots]^{[p]}}_{k\text{ times}}.$$

Lemma 7. *Every element of an arbitrary restricted Lie algebra A over Σ with generating set R can be written as a linear combination with coefficients in Σ of elements of the form $u^{[p^k]}$ $(k = 0, 1, 2, \ldots)$ where u is a regular nonassociative word.*

The proof of this result follows immediately from Definition 4 and Lemma 2.

Definition 5. An associative word v is called *p-regular* if it has the form u^{p^k} $(k = 0, 1, 2, \ldots)$ where u is a regular word.

[2]This is now usually called the Möbius inversion formula. [Translators]

Definition 6. Elements of the ring $\mathfrak{A}_{\Sigma R}^{(p)}$ that have the form $[u]^{p^k}$, where $[u]$ is a regular nonassociative R-word relative to the operation $a \circ b = ab - ba$, are called *p-regular elements.*

Lemma 8. *The set of p-regular elements of $\mathfrak{A}_{\Sigma R}^{(p)}$ is linearly independent over Σ.*

Proof. Obviously, the leading term of the polynomial which is the associative expansion of the p-regular element $[u]^{p^k}$ will be the p-regular associative word u^{p^k}. Therefore distinct p-regular elements correspond to distinct maximal words. From this the lemma follows. □

Lemmas 7 and 8 immediately imply the following result.

Theorem 2. *The algebra $\mathfrak{A}_{\Sigma R}^{(p)}$ is a free restricted Lie algebra over Σ with generating set R and a basis consisting of the p-regular elements.*

From the above constructions we immediately obtain an algorithm that allows us to determine, for a given element a of the algebra $\mathfrak{A}_{\Sigma R}$, whether or not it belongs to the algebras $\mathfrak{A}_{\Sigma R}^{(-)}$ or $\mathfrak{A}_{\Sigma R}^{(p)}$. For this determination, one should separate the lexicographically maximal monomial σu in the expression of the element a. If the word u is not regular (respectively, p-regular) then the corresponding membership question is answered in the negative. If the word is regular (respectively, p-regular) then subtracting from a the element $\sigma[u]$ (respectively, $\sigma[u_1]^{p^k}$) where $[u]$, $[u_1]$ are the corresponding regular nonassociative words, we obtain an element a_1 whose maximal monomial will be less than the monomial σu. After a finite number of steps this process will terminate. From this algorithm one can obtain the following criterion of Friedrichs [4].

Theorem 3. *An element $f(a_1, a_2, \ldots, a_s)$ of the algebra $\mathfrak{A}_{\Sigma R}$ belongs*[3] *to $\mathfrak{A}_{\Sigma R}^{(p)}$ if and only if the relations $a_i a'_j = a'_j a_i$ imply the equation*[4]

$$f(a_1 + a'_1, a_2 + a'_2, \ldots, a_s + a'_s) = f(a_1, a_2, \ldots, a_s) + f(a'_1, a'_2, \ldots, a'_s).$$

Proof. The proof of the necessity of the conditions is by induction and is almost trivial. Let us prove the sufficiency for the case of characteristic 0 (the proof of the general case is similar).

Let d be an element of $\mathfrak{A}_{\Sigma R}$ that does not belong to $\mathfrak{A}_{\Sigma R}^{(-)}$. Then, after a finite number of steps of the above-mentioned algorithm, we will obtain an element d_i whose leading term is σu_i where the word u_i is not regular. Then $u_i = vw$ where $wv \geq u_i$; and[5] wv is maximal among the words that are cyclically comparable with u_i. It is easy to see, however, that in the expression

$$d_i(a_1 + a'_1, a_2 + a'_2, \ldots, a_s + a'_s) - d_i(a_1, a_2, \ldots, a_s) - d_i(a'_1, a'_2, \ldots, a'_s),$$

[3] In the case of characteristic 0, one must replace $\mathfrak{A}_{\Sigma R}^{(p)}$ by $\mathfrak{A}_{\Sigma R}^{(-)}$. [Translators]
[4] Today this is expressed in terms of the coproduct on the algebra $\mathfrak{A}_{\Sigma R}$. [Translators]
[5] (without loss of generality). [Translators]

the element

$$v(a_1,\dots,a_s)\,w(a'_1,\dots,a'_s) = w(a'_1,\dots,a'_s)\,v(a_1,\dots,a_s),$$

occurs with coefficient $\sigma \neq 0$. The proof is complete. □

4. Theorems on embeddings of Lie algebras and restricted Lie algebras

In what follows we will denote by L a Lie algebra over an arbitrary field or a restricted Lie algebra over a field of characteristic $p > 0$. Our task is to demonstrate the possibility of embedding L into an appropriate algebra with two generators under certain assumptions of countability, and then to generalize this result. Let A be the free associative algebra on two generators a and b.

Lemma 9. *The elements*

$$d_k = [a \circ \{[\cdots(a \circ \underbrace{b) \circ b \cdots] \circ b\}]}_{k\ times} \circ (a \circ b) \qquad (k = 1, 2, \ldots),$$

of A generate (under the operations $a \circ b$ and $a^{[p]}$) a free Lie algebra (respectively a free restricted Lie algebra) $L(a,b)$, and constitute a set T of free generators.

Proof. We order the set T by setting $d_k > d_s$ for $k < s$. It is easy to verify that every regular nonassociative T-word (respectively, p-regular T-element) is a regular nonassociative R-word (respectively, p-regular R-element). From this follows the linear independence of regular nonassociative T-words (respectively, p-regular T-elements), and this proves the lemma. □

Lemma 10. *The set $T = \{d_k\}$ is distinguished (in the sense of Definition 1 of* [8]*).*

Proof. Let J be an ideal of $L(a,b)$ and J_1 the ideal generated by J in A. It suffices to prove the equality $J_1 \cap L(a,b) = J$ (in fact the definition of 'distinguished' demands that $J'_1 \cap L(a,b) = J$ where J'_1 is the ideal of the Lie algebra $A^{(-)}$ generated by J). Consider some element ℓ in J_1:

$$\ell = \sum_i m_i = \sum_i \alpha_i \ell_i \beta_i,$$

where the α_i and β_i are monomials (possibly empty) in a and b, and the ℓ_i are elements of J and thus of $L(a,b)$, that is, Lie polynomials in the elements of T. Let $\overline{\ell_i}$ be the maximal word among the words of highest degree that occur in the associative expansion of ℓ_i. Obviously, $\overline{\ell_i}$ is a regular (respectively p-regular) associative word. Among the words of the form $\alpha_i\overline{\ell_i}\beta_i$ that have highest degree, we choose one which is maximal in the lexicographical sense: $t = \alpha_j\overline{\ell_j}\beta_j$. Assume that $\ell \in L(a,b)$.

Case 1*:* Suppose that the word t does not occur[6] among the other associative words that occur in the expansion of ℓ. Then the word t is regular (respectively

[6]Literally, "does not have similars". [Translators]

p-regular) and contains a regular (respectively p-regular) subword $\overline{\ell_j}$. We place parentheses in the word t in the unique way which makes it a regular nonassociative word (p-regular element). By virtue of Lemma 4, one pair of these parentheses will be placed as follows[7]: $\alpha_j[\overline{\ell_j}\beta_{1j}]\beta_{2j}$ where $\beta_{1j}\beta_{2j} = \beta_j$ and each of the words β_{1j}, β_{2j} may be empty. Moreover, after the required placement of parentheses, the word $\overline{\ell_j}\beta_{1j}$ will take the form

$$\ell' = [\cdots(\overline{\ell_j} \circ \beta^{(1)}) \circ \beta^{(2)} \cdots] \circ \beta^{(k)},$$

where $\overline{\ell_j}$ and β_s are regular nonassociative words and $\beta^{(1)} \leq \beta^{(2)} \leq \cdots \leq \beta^{(k)}$.

The word t is the product of words of the form a^2b^kab $(k = 1, 2, \ldots)$ since otherwise, performing the algorithm of expressing ℓ as a linear combination of the basis elements of the algebra $A^{(-)}$ (respectively, the free restricted algebra $A^{(p)}$), we would obtain a leading term which is not a T-word; but it is obvious that every element of $L(a, b)$ contains only T-words in its expression in terms of the basis of regular words. From this it easily follows that each of the words $\beta^{(s)}$, as well as α_j and β_{2j}, is a T-word.

The element $\ell'' = \{\alpha_j[\overline{\ell_j}\beta_{1j}]\beta_{2j}\}$, in which the parentheses inside the square brackets are placed in the same way as in ℓ', and elsewhere in the way prescribed for regular words, will obviously be a nonassociative T-polynomial that belongs to the intersection $J_1 \cap L(a, b)$; from the proof of Lemma 4 it follows that its lexicographically maximal word, among the words of highest degree, coincides with t. Therefore, in the difference $\ell - \ell''$ the analogous word will be lexicographically smaller or will have lower degree.

Our argument does not apply only in the case when t is a p^k-th power of $\overline{\ell_j}$. Let

$$t = q^s\overline{\ell_j}q^{p^{k+k_1}-p^{k_1}-s},$$

where $\overline{\ell_j} = q^{p^{k_1}}$, and q is a regular word. The element $(\ell_j)^{p^k}$ of the ideal J can be written in the form

$$\left(\overline{\ell_j} + \omega\right)^{p^k} = \left(\overline{\ell_j} + \omega\right)^{p^k-1}\ell_j = q^{p^{k+k_1}-p^{k_1}}\ell_j + \sum_k \varepsilon_k\ell_j,$$

where ω stands for the terms smaller than $\overline{\ell_j}$, and the leading terms of the elements $\varepsilon_k\ell_j$ of the ideal J_1 are less than t. On the other hand, by virtue of the representation

$$\begin{aligned} q^s\ell_j q^{p^{k+k_1}-p^{k_1}-s} &= \sum_{t=1}^{p^{k+k_1}-p^{k_1}-s} q^{s+t-1}(\ell_j \circ q)q^{p^{k+k_1}-p^{k_1}-s-t} + q^{p^{k+k_1}-p^{k_1}}\ell_j \\ &= s_1 + q^{p^{k+k_1}-p^{k_1}}\ell_j = s_1 + (\ell_j)^{p^k} - \sum_k \varepsilon_k\ell_j, \end{aligned}$$

[7]We have added a bar over ℓ_j. [Translators]

where s_1 is also an element of J_1 with leading term less than t, we see that after subtracting from ℓ the element $(\ell_j)^{p^k}$ (that obviously belongs to the ideal J) we will obtain an element of J_1 with leading term less than t.

Case 2*:* Suppose that we have several maximal words: $t_1, t_2, \ldots, t_r$. Take any two of them:

$$t_1 = \alpha_j \overline{\ell_j} \beta_j, \qquad t_2 = \alpha_k \overline{\ell_k} \beta_k.$$

Again several subcases are possible.

(a) $t_1 = t_2 = \alpha_j \, \overline{\ell_j} \, \gamma_j \, \overline{\ell_k} \, \beta_k$. In this case m_j (the element to which the word t_1 belongs) can be written, up to a scalar coefficient, in the form

$$m_j = \alpha_j \, \ell_j \, \gamma_j \, \overline{\ell_k} \, \beta_k = \alpha_j \, \ell_j \, \gamma_j \, \ell_k \, \beta_k - \omega_j = m_k + \omega,$$

where ω and ω_j are polynomials whose terms are smaller than t_1 or have lower degree; and ω_j as well as ω belong to the ideal J_1. Combining similar terms reduces the number of distinct t_s.

(b) $t_1 = t_2 = \alpha_j \ell_{1j} \ell_{2j} \ell_{3j} \beta_k$ where $\ell_{1j} \ell_{2j} = \overline{\ell_j}$ and $\ell_{2j} \ell_{3j} = \overline{\ell_k}$; also $\overline{\ell_j}$ and $\overline{\ell_k}$ are regular words (one of the words ℓ_{1j} and ℓ_{3j} may be empty). From regularity of the words $\overline{\ell_j}$ and $\overline{\ell_k}$ and the Remark after Definition 1, it follows that the word $\ell_{1j} \ell_{2j} \ell_{3j}$ is regular. From Lemma 4 it follows that the placement of parentheses in the word $[\ell_{1j} \ell_{2j} \ell_{3j}]$ on the subword $\overline{\ell_k}$ coincides with the placement of parentheses in the word $[\overline{\ell_k}]$:

$$[\ell_{1j} \ell_{2j} \ell_{3j}] = \ell'_1 \{\ell'_2 \cdots (\ell'_s [\overline{\ell_k}]) \cdots \}.$$

The same lemma implies that we may place parentheses in the word $\ell_{1j} \ell_{2j} \ell_{3j}$ as follows:

$$\{([\overline{\ell_j}] \ell''_1) \ell''_2 \cdots \} \ell''_q,$$

where $\ell''_1 \le \ell''_2 \le \cdots \le \ell''_q$, and the ℓ''_r are regular words with the corresponding placement of parentheses. In this case each of the words ℓ_{tj} $(t = 1, 2, 3)$, ℓ'_p, ℓ''_r is a product of words of the form $a^2 b^k a b$ $(k = 1, 2, \ldots)$. Obviously, the element

$$\alpha_j \Big(\ell'_1 \circ \{\ell'_2 \circ \cdots (\ell'_s \circ \ell_k) \cdots \} + \{\cdots [(\ell_j \circ \ell''_1) \circ \ell''_2] \cdots \circ \ell''_q\} \Big) \beta_k,$$

of the ideal J_1, coincides up to some lower terms with the sum $m_j + m_k$. Therefore the words t_1 and t_2 in this case can be replaced by one word (or both can be omitted).

One can argue analogously using Lemma 4 in the case where $\overline{\ell_j} = \ell_{1j} \ell_{2j} \ell_{3j}$ and $\ell_{2j} = \overline{\ell_k}$.

(c) $t_1 = t_2 = \alpha_j \ell_{1j} \ell_{2j} \ell_{3j} \beta_k$ where $\ell_{1j} \ell_{2j} = \overline{\ell_j} = q^{p^r}$, $\ell_{2j} \ell_{3j} = \overline{\ell_k} = q^{p^s}$ $(s > r)$ and $\ell_{2j} = q^n$. Then we can reduce the number of the words t_s by using

the equations

$$\ell_j\ell_{3j} = \ell_j q^{p^s-n} = \sum_{t=0}^{p^s-n-1} q^t(\ell\circ q)q^{p^s-n-t-1} + q^{p^s-n}\ell_j = \omega_1 + q^{p^s-n}\ell_j$$
$$= \omega_1 + q^{p^s-n}\overline{\ell_j}^{p^{s-r}-1}\ell_j = \omega_1 + q^{p^r-n}\Big(\overline{\ell_j}^{p^{s-r}} - \sum_i \varepsilon_i\ell_j\Big)$$
$$= q^{p^r-n}\ell_j^{p^{s-r}} + \omega_2 = \ell_{1j}\ell_k + q^{p^r-n}\Big(\ell_j^{p^{s-r}} - \ell_k\Big) + \omega_2 = \ell_{1j}\ell_k + \omega_3,$$

where ω_i, as well as $\varepsilon_i\ell_j$, are elements of the ideal J_1 with smaller leading terms.

Considering the remaining possible cases, including those in which one of the words $\overline{\ell_j}$ or $\overline{\ell_k}$ is regular and the other is p-regular but not regular, by analogous arguments we can reduce the number of words t_s. Having reduced this number to 1, we will be under the conditions of Case 1. These arguments imply that after a finite number of steps we will express ℓ as an element of the ideal J. The proof is complete. □

Theorem 4. *Every Lie algebra or restricted Lie algebra of at most countable rank can be isomorphically embedded into an appropriate algebra with two generators over the same field.*

Theorem 5. *Any Lie algebra (respectively restricted Lie algebra) can be isomorphically embedded into a Lie algebra (respectively restricted Lie algebra) with the property that every subalgebra of countable rank is contained in a subalgebra with two generators.*

Theorems 4 and 5 are corollaries of Lemma 10 as well as Theorems 1 and 2 of [8]; it is necessary to remark that although the statements of the latter Theorems do not formally include the case of restricted Lie algebras, the given proofs also remain valid in this case without any changes.

It is easy to see that the algebras obtained here are automatically represented in an associative algebra. Therefore, the proof given here also contains a proof of the Birkhoff-Witt theorem [1], [10] and the theorem of Jacobson [5].

References

[1] G. Birkhoff: *Representability of Lie algebras and Lie groups by matrices.* Annals of Math. 38 (1937) 526–532.

[2] P.M. Cohn: *Sur le critère de Friedrichs pour les commutateurs dans une algèbre associative libre.* C. R. Acad. Sci. Paris 239 (1954) 743–745.

[3] E.B. Dynkin: *Evaluation of the coefficients in the Campbell-Hausdorff formula.* Doklady Akad. Nauk USSR 57 (1947) 323–326.

[4] K.O. Friedrichs: *Mathematical aspects of the quantum theory of fields, V.* Comm. Pure Appl. Math. 6 (1953) 1–72.

[5] N. Jacobson: *Restricted Lie algebras of characteristic p.* Trans. Amer. Math. Soc. 50 (1941) 15–25.

[6] R.C. Lyndon: *A theorem of Friedrichs*, Michigan Math. J. 3 (1955–56) 27–29.

[7] A.I. Shirshov: *On the representation of Lie rings in associative rings*. Uspekhi Mat. Nauk 8 (1953) 173–175.

[8] A.I. Shirshov: *Some theorems on embedding of rings*. Mat. Sbornik 40 (1956) 65–72.

[9] F. Wever: *Operatoren in Lieschen Ringen*. J. reine angew. Math. 189 (1947) 44–55.

[10] E. Witt: *Treue Darstellung Lieschen Ringen*. J. reine angew. Math. 177 (1937) 152–160.

On a Problem of Levitzki

A.I. Shirshov

An associative ring S is called a *nil-ring* if every element of S is nilpotent. Levitzki [4] posed the following problem: Is every nil-ring nilpotent? This problem was solved in the affirmative by Levitzki himself [5] for the case in which the elements of S have globally bounded indices of nilpotency. Later, Kaplansky [2], who was investigating the more general problem of Kurosh [3], extended the result of Levitzki to nil-rings with polynomial identities. In the present note an affirmative solution is given to Levitzki's problem for the wider class of rings introduced by Drazin [1].

Let $\Lambda = \{\lambda_i\}$, $i = 1, 2, \ldots, h$ be some set of variables, and let $\pi(\lambda) = \lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k}$ be some monomial in these variables. Denote by $T_\pi(\lambda)$ the set of all monomials in Λ of degree $\geq k$ and distinct from $\pi(\lambda)$. For any sequence of elements $\{x_i\}$, $i = 1, 2, \ldots, h$, of the ring S, we denote by $\pi(x)$ the element $x_{i_1}x_{i_2}\cdots x_{i_k}$ and by $T_\pi(x)$ the set of all elements of S obtained by replacing the variables λ_i by the corresponding elements x_i in each monomial of $T_\pi(\lambda)$.

If there exists a monomial $\pi(\lambda)$, such that for any collection of elements x_i, $i = 1, 2, \ldots, h$, of the ring S the element $\pi(x)$ belongs to the right ideal generated by $T_\pi(x)$, then the monomial $\pi(\lambda)$ is called a *strongly pivotal monomial* of S, and S is called a *ring with strongly pivotal monomial*. For brevity, we will call such rings SP-rings.

Drazin [1] has shown that the class of SP-rings contains the rings with minimum condition on right ideals and the rings with polynomial identity. In the same paper it was shown that for any SP-ring the monomial $\pi(\lambda)$ can be assumed to be linear in each variable λ_i. Under some strong restrictions, Drazin, using essentially the methods of Kaplansky, gave an affirmative solution to the problem of Kurosh for SP-algebras, i.e., he proved local finiteness of algebraic SP-algebras of a particular type. However, Drazin himself points out the difficulties that did not allow him to solve even the Levitzki problem for SP-rings without additional restrictions.

Doklady Akad. Nauk SSSR 120, (1958), no. 1, 41–42.

If a strongly pivotal monomial $\pi(\lambda)$, which in the sequel will be assumed linear in each variable λ_i, has degree t, then the SP-ring S will be called *an SP-ring of degree t.*

Lemma. *Let S be a nil SP-ring of degree t, and let I be the ideal generated by the elements a_i^t where a_i, $i = 1, 2, \ldots, n$, is some fixed set of elements of S. Then for any natural number $q > t$ there exists a natural number $k = k(q)$ such that the ideal I^k is contained in the ideal generated by the elements a_i^q.*

Proof. Suppose that there exists a natural number r such that the ideal I^r is contained in the ideal generated by the elements a_i^m, $i = 1, 2, \ldots, n$, for some fixed $m \geq t$. In order to prove the lemma we will show that there exists a natural number r_1 such that the ideal I^{r_1} is contained in the ideal generated by the elements a_i^{m+1}.

Every element of the ideal $I^{r(nt+1)}$ can be written as a sum of products of $nt+1$ elements of the form $\alpha\, a_j^m \beta$ where α and β are monomials in the generators of S. In each such product there exists an element a_j^m that occurs at least $t+1$ times. Therefore, each such product can be written in the form

$$\begin{aligned} c_1 D c_{t+2} &= c_1 d_1 d_2 \cdots d_t c_{t+2} \\ &= c_1 (a_j^m c_2 a_j)(a_j^{m-1} c_3 a_j^2)(a_j^{m-2} c_4 a_j^3) \cdots (a_j^{m-t+1} c_{t+1} a_j^t) c_{t+2}. \end{aligned}$$

By assumption, the monomial D belongs to the right ideal generated by all possible products of its factors d_1, d_2, …, d_t which are distinct from D itself and have total degree (with respect to the elements d_i) greater than or equal to that of D.

For any other monomial of degree t in the elements d_i, there exist two adjacent elements d_{j_1} and d_{j_2} with $j_1 \geq j_2$. In each such case, in the corresponding segment, there is a word

$$\begin{aligned} d_{j_1} d_{j_2} &= a_j^{m-j_1+1} c_{j_1+1} a_j^{j_1} a_j^{m-j_2+1} c_{j_2+1} a_j^{j_2} \\ &= a_j^{m-j_1+1} c_{j_1+1} a_j^{m+1+(j_1-j_2)} c_{j_2+1} a_j^{j_2}. \end{aligned}$$

It is easy to see that all such elements belong to the ideal generated by a_j^{m+1}. From this it follows that $c_1 D = \omega_1 + c_1 D q$ where ω_1 is an element of the ideal generated by the element a_j^{m+1}. But then

$$\begin{aligned} c_1 D &= \omega_1 + \omega_1 q + c_1 D q^2 = \omega_1 + \omega_1 q + \omega_1 q^2 + c_1 D q^3 = \cdots \\ &= \omega_1 + \omega_1 q + \omega_1 q^2 + \cdots + \omega_1 a^\ell + c_1 D q^{\ell+1}, \end{aligned}$$

for any ℓ. The lemma now follows from nilpotency of the element q. For the number r_1 we can take $r(nt+1)$. □

Theorem. *Any nil SP-ring is locally nilpotent.*

Proof. Let S be a nil SP-ring of degree t with a finite number of generators, and let J be the ideal generated by all possible elements a^t, $a \in S$. The quotient ring S/J is nilpotent by Levitzki's theorem [5], and this means that there exists a natural number M such that any element of the form $b_{i_1} b_{i_2} \cdots b_{i_M}$, where the b_{i_s} are the generators of S, belongs to the ideal J. Since there is only a finite number

of elements of the form $b_{i_1} b_{i_2} \cdots b_{i_M}$, the ideal S^M is contained in some ideal J_1 which is contained in J and is generated by some finite set of elements a_i^t. The lemma implies nilpotency of the ideal J_1, and hence of the ring S. □

References

[1] M.P. Drazin, *A generalization of polynomial identities in rings*, Proc. Amer. Math. Soc. 8 (1957) 352–361.

[2] I. Kaplansky, *Topological representation of algebras II*, Trans. Amer. Math. Soc. 68 (1950) 62–75.

[3] A.G. Kurosh, *Ringtheoretische Probleme, die mit dem Burnsideschen Problem über periodische Gruppen in Zusammenhang stehen*, Izvestiya Akad. Nauk USSR Ser. Mat. 5 (1941) 233–240.

[4] J. Levitski, *On the radical of a general ring*, Bull. Amer. Math. Soc. 49 (1943) 462–466.

[5] J. Levitski, *On a problem of A. Kurosch*, Bull. Amer. Math. Soc. 52 (1946) 1033–1035.

Some Problems in the Theory of Rings that are Nearly Associative

A.I. Shirshov

The words "some problems" in the title of this article mean primarily that the article considers absolutely no results about algebras of finite dimension. Among other questions that remain outside the scope of the article, we mention, for example, various theorems about decomposition of algebras (see for example [47, 70]) which are closely related to the theory of algebras of finite dimension.

The author is grateful to A.G. Kurosh and L.A. Skornyakov who got acquainted with the first draft of the manuscript and made a series of very valuable comments.

1. Introduction

1. Until recently the theory of rings and algebras was regarded exclusively as the theory of *associative* rings and algebras. This was a result of the fact that the first rings encountered in the course of the development of mathematics were associative (and commutative) rings of numbers and rings of functions, and also associative rings of endomorphisms of Abelian groups, in particular, rings of linear transformations of vector spaces.

In the survey article by A.G. Kurosh [40] he persuasively argued that the contemporary theory of associative rings is only a part of a general theory of rings, although it continues to play a very important role in mathematics. The present article, in contrast to the article of A.G. Kurosh, is dedicated to a survey of one part of the theory of rings: precisely, the theory of rings, which although nonassociative, are more or less connected with associative rings. More precise connections will be mentioned during the discussion of particular classes of rings.

Uspekhi Mat. Nauk 13, (1958), no. 6 (84), 3–20.

Because the classes of rings that are studied in this article were mentioned to some extent in the article of A.G. Kurosh, there is some intersection in the content of these two articles. In what follows, the author assumes that the following notions are understood: rings, algebras, ideals, quotient rings, rings with a domain Σ of operators (or Σ-operator rings[1]). These notions and also some other main notions of the theory of rings can be found in the same article by A.G. Kurosh.

2. We briefly describe the origins of the theory of nonassociative rings. Examples of such rings were known a long time ago. The nonassociativity of the vector product of 3-dimensional vectors was known in mechanics. With this operation and vector addition the collection of vectors is a Lie ring. Another very beautiful example is the algebra of so-called Cayley numbers, which have been used in different parts of mathematics.

The development of the theory of continuous groups in general and Lie groups in particular contributed to the study of Lie algebras of finite dimension, which are closely connected to Lie groups. Another connection between Lie algebras and groups which appears to be very fruitful has been studied in the works of W. Magnus [45], I.N. Sanov [50], A.I. Kostrikin [35] and others.

There is an interesting relationship between associative rings on the one hand and Lie rings and Jordan rings[2] on the other hand, constructed by the introduction of a new operation on an associative ring. This relationship, in addition to giving certain information about Lie rings and Jordan rings, allows us to study associative rings themselves from some new directions.

3. Because there are differences between the properties of rings in different classes, there are few results which have a universal character. We will describe some of them.

Let A be an associative ring, and let a be some element of the ring A. It is possible to connect with this element a new operation of "multiplication" which is defined by $x \cdot y = axy$. It is easy to check that the set of elements of the ring A forms, under this operation and addition, a ring (in general, already nonassociative), which we will denote by $A(a)$. In [48] A.I. Malcev proved that any ring is isomorphic to some subring of a ring of the form $A(a)$.

Let the additive group of an associative ring be decomposed into the direct sum of subgroups A_1 and A_2. Then every element $a \in A$ allows a unique representation of the form $a = a_1 + a_2$. Under the operations of "multiplication" $x \cdot y = (xy)_1$ and addition the set of elements of the ring A is a ring (in general, nonassociative). We denote this ring by A'. In [66] L.A. Skornyakov proved that any ring is isomorphic to some subring of a ring of the form A'.

The preceding results of Malcev and Skornyakov indicate the possibility of developing the entire theory of rings in terms of associative rings. However, nobody until now has been able to get any precise theorems about rings of some class based

[1]That is, an algebra over the commutative associative coefficient ring Σ. [Translators]
[2]Literally, "J-rings". [Translators]

on this method. Among the reasons for this is the fact that we cannot transfer the properties of A to $A(a)$ and A'. So, for example, if A is a Lie ring, then the rings $A(a)$ and A' may not be Lie rings.

The results and problems that correspond to different classes of rings are formulated very differently and require specific methods, and because of this it is difficult to imagine the development of the entire theory of rings from the theory of one specific, sufficiently studied class.

4. In the theory of rings, as in the theory of groups and other algebraic systems, free systems play an important role: free rings, free associative rings, free Lie rings, etc.

Let ν be a cardinal number. The free ring (free associative ring, free Lie ring, etc.) on ν generators is a ring (associative ring, Lie ring, etc.) which has a system S of generators of cardinality ν such that any mapping from S onto any system of generators of any ring (associative ring, Lie ring, etc.) can be extended to a homomorphism of rings. The free ring A_ν with the set S of generators of cardinality ν can be built constructively by the following steps.

We will call the elements of the set S *words of length* 1. If α and β are words of lengths m and n (respectively) then the symbol $(\alpha)(\beta)$ will be called a *word of length* $m+n$; furthermore, we will consider two words $(\alpha)(\beta)$ and $(\alpha_1)(\beta_1)$ to be equal if and only if $\alpha=\alpha_1$ and $\beta=\beta_1$. The collection of finite sums of the form $\sum_s k_s\gamma_s$ where k_s is an integer and γ_s is a word (we assume $\gamma_s\neq\gamma_t$ when $s\neq t$) becomes a ring, which we will denote by A_ν, when we define the operations as follows:

$$\sum_s k_s\gamma_s+\sum_s l_s\gamma_s=\sum_s(k_s+l_s)\gamma_s,$$
$$\sum_s k_s\gamma_s\cdot\sum_t l_t\gamma_t=\sum_{s,t}k_sl_t(\gamma_s)(\gamma_t).$$

It is easy to check that the ring A_ν satisfies the above-formulated definition, and that any ring that satisfies that definition is isomorphic to A_ν.

If the symbols k_s are allowed to come from some associative ring Σ and we define

$$k\sum_s k_s\gamma_s=\sum_s(kk_s)\gamma_s,\ k\in\Sigma,$$

then the ring A_ν will be a free Σ-operator ring with ν generators in the sense of Σ-operator homomorphisms. If, furthermore, Σ is a field, then A_ν is a free algebra with ν generators over the field Σ.

In the works of Kurosh [39, 41] it was proved that any subalgebra of a free algebra is again free, and some generalizations of this result to free sums of algebras were given. A.I. Zhukov [74] solved positively the word problem[3] for algebras[4] with

[3]Literally, the "problem of equality". [Translators]
[4]That is, nonassociative algebras. [Translators]

a finite number of generators and a finite number of defining relations which is analogous to the famous word problem in the theory of groups.

5. With additional axioms, or so-called identical relations, we may define various classes of rings. The general method applied to this problem is as follows.

Let A_ω be the free ring with a countably infinite number of generators x_i $(i = 1, 2, \ldots)$. In the ring A_ω we consider a subset Q. Any ring C which satisfies the condition that any substitution of any elements of C into the generators x_i in any element of the set Q gives zero, will be regarded as belonging to the class defined by the set Q, or simply to the class of Q-rings. If in some free ring A_ν we take the ideal J generated by the elements obtained by substituting all the elements of A_ν into the generators x_i in the elements of Q, then the quotient ring $D = A_\nu/J$ will be isomorphic to the free Q-ring in the sense given earlier. For example, if the set Q consists of the single element $(x_1x_2)x_3 - x_1(x_2x_3)$ then we obtain the class of associative rings. If the set Q consists of elements q_α, then it is sometimes said that the class of Q-rings is defined by the identical relations $q_\alpha = 0$. The same concepts can be defined in a very similar way for Σ-operator Q-rings.

For the case when the set Q is finite, Yu.I. Sorkin [69] showed that the corresponding class of rings can be given with the help of one ternary operation (that is, defined on ordered triples of elements) and one relation which this operation must satisfy.

2. Alternative rings

1. It is known that the field of complex numbers can be represented as the collection of pairs of real numbers with the natural addition and the familiar definition of multiplication. If on the Abelian group of ordered pairs (p, q) of complex numbers with coordinate-wise addition is defined an operation of multiplication by the formula

$$(p_1, q_1) \cdot (p_2, q_2) = (p_1p_2 - \overline{q_2}q_1, q_2p_1 + q_1\overline{p_2}), \tag{1}$$

where $\overline{p_2}$ and $\overline{q_2}$ are the complex conjugates of the complex numbers p_2 and q_2, then one can easily check that with respect to these operations the set we are considering is a ring. In this ring it happens that the equations $AX = B$ and $XC = D$ have a uniquely determined solution when $A \neq 0$, $C \neq 0$ and so this ring is the (associative but not commutative) division ring of real quaternions. If in equation (1) we replace the symbols p_i and q_i by real quaternions, and we understand $\overline{p}$ to be the quaternion conjugate of the quaternion $p = (a, b)$ – that is, $\overline{p} = (\overline{a}, -b)$ – then the pairs of quaternions become a ring with respect to these operations, which in this case is a nonassociative division ring. If for every real number α and pair (p, q) we define $\alpha(p, q) = (\alpha p, \alpha q)$, then the additive groups of the above division rings become vector spaces over the field of real numbers with corresponding dimensions 4 and 8, and the division rings become algebras over the

field of real numbers. The constructed nonassociative algebra of dimension 8 over the field of real numbers is called the *algebra of Cayley numbers*. In what follows we will denote it by R_8.

2. The *associator* of the elements a, b, c in any ring is defined to be the element

$$[a, b, c] = (ab)c - a(bc).$$

The algebra R_8 satisfies the following identical relations,

$$[x, y, y] = 0, \tag{2}$$

$$[x, x, y] = 0, \tag{3}$$

$$[x, y, x] = 0, \tag{4}$$

each of which is implied by the other two. Rings in which the identical relations (2)–(4) are satisfied are called *alternative*. A more general class of 8-dimensional alternative algebras was studied by Dickson. These algebras received the name Cayley-Dickson algebras.

In this and the following section (if this is not stated explicitly) for simplicity of language we will assume that the additive groups of the rings do not contain elements of order 2.

We next list some identical relations that hold in every alternative ring:

$$[(xy)z]y = x[(yz)y], \tag{5}$$

$$y[z(yx)] = [y(zy)]x, \tag{6}$$

$$(xy)(zx) = x[(yz)x]. \tag{7}$$

To prove relation (5) we notice that substitution of $y + z$ for y in equation (2) leads to the equation

$$[x, y, z] = -[x, z, y]. \tag{8}$$

Using equations (2) and (8) gives

$$\begin{aligned}
2x[(yz)y] &= x\big[\,2(yz)y + [z, y, y] - [y, z, y] - [y, y, z]\,\big] \\
&= x\big[\,(yz)y + (zy)y - zy^2 + y(zy) - y^2z + y(yz)\,\big] \\
&= [x(yz)]y + (xy)(yz) - [x, yz, y] - [x, y, yz] + [x(zy)]y + (xy)(zy) \\
&\quad - [x, zy, y] - [x, y, zy] + [x, z, y^2] + [x, y^2, z] - (xz)y^2 - (xy^2)z \\
&= \big[\,x(yz) + x(zy)\,\big]y + (xy)(yz + zy) - [(xz)y]z - [(xy)y]z \\
&= 2[(xy)z]y.
\end{aligned}$$

Thus equation (5) is proved, and for its proof we used only equation (2). From this it follows that equation (5) holds in any ring which satisfies equation (2), that is, in any so-called *right alternative* ring. The proofs of equations (6) and (7) are left to the reader.

3. Let us notice one property of alternative rings, which makes them close to associative rings. Let a and b be two elements of some alternative ring A, and let

D be the subring of the ring A generated by the elements a and b. It happens that the ring D is associative. To prove this proposition it is enough to show that any two elements of the ring D obtained by different parenthesizations of an associative monomial in a and b are equal.

Let c be some associative monomial as described. We denote by $\langle c\rangle$ the nonassociative monomial obtained from the monomial c by the following parenthesization: when $c = c_1a$ or $c = c_1b$ we let $\langle c\rangle = (\langle c_1\rangle)a$ or $\langle c\rangle = (\langle c_1\rangle)b$, respectively; and $\langle a\rangle = a$, $\langle b\rangle = b$. For example, $\langle a^2bab^2\rangle = ((((aa)b)a)b)b$. If d is a nonassociative monomial with some parenthesization, then we will denote by $\overline{d}$ the associative monomial obtained by removing the parentheses from d. The associativity of the ring D is equivalent to the equation $d = \langle\overline{d}\rangle$ holding where d is any nonassociative monomial in the generators a and b. The last equality, which is obvious if the degree of the monomial d in a and b is less than or equal to 3, will be proved by induction on the degree of d.

Let the degree of the monomial d be greater than 3: $d = d_1d_2$, $d_1 = a\langle\overline{d_3}\rangle$, and we assume that the equality to be proved holds for monomials with lower degree. Then we have the following cases:

$$\text{(i)}\qquad d_2 = \langle\overline{d_4}\rangle a, \quad d = (a\langle\overline{d_3}\rangle)(\langle\overline{d_4}\rangle a) = [a(\langle\overline{d_3}\rangle\langle\overline{d_4}\rangle)]a = \langle\overline{d}\rangle,$$

where we have used equation (7). If the monomial $\langle\overline{d_3}\rangle$ is empty, then the proof works using equation (4).

$$\text{(ii)}\qquad d_2 = (b\langle\overline{d_4}\rangle)b, \quad d = (a\langle\overline{d_3}\rangle)[(b\langle\overline{d_4}\rangle)b] = [(d_1b)\langle\overline{d_4}\rangle]b = \langle\overline{d}\rangle,$$

where equation (5) was used. Finally,

$$\begin{aligned}\text{(iii)}\quad d_2 &= (a\langle\overline{d_4}\rangle)b,\\ d &= (a\langle\overline{d_3}\rangle)[(a\langle\overline{d_4}\rangle)b]\\ &= -(a\langle\overline{d_3}\rangle)[b(a\langle\overline{d_4}\rangle)] + [(a\langle\overline{d_3}\rangle)(a\langle\overline{d_4}\rangle)]b + [(a\langle\overline{d_3}\rangle)b](a\langle\overline{d_4}\rangle)\\ &= -\langle\overline{d_5}\rangle + \langle\overline{d}\rangle + d_5,\end{aligned}$$

where we have used equation (8) and also the above-proved identities from cases (i) and (ii). Repeating (if necessary) the same transformation on d_5 and so on, we come in a finite number of steps to the identity which we are proving.

4. In spite of the noted closeness of alternative rings to associative rings, as of now there is no general method which allows us to prove identities in alternative rings. Each of the presently known such identities requires a separate and in some cases very difficult proof. This happens because as of now there is no known method to build constructively free alternative rings, so there is no known algorithm which solves the word problem in free alternative rings; that is, an algorithm which allows us, for every element of this ring written in terms of the generators, to determine if it is zero or not.

We mention the following interesting identity:

$$[\,(ab-ba)^2,\, c,\, d\,]\,(ab-ba)=0,$$

which was proved by Kleinfeld (see for example [67]) and which shows that in the free alternative ring there are zero divisors.

5. The study of alternative rings in general began with the study of alternative division rings, which in the theory of projective planes play the role of the so-called natural division rings of alternative planes (see [65]); that is, planes for which the little Desargues theorem holds.

In the works of L.A. Skornyakov [62, 63] a full description is given of alternative but not associative division rings. It happens that every such division ring is an algebra of dimension 8 over some field (a Cayley-Dickson algebra). Later and independently of Skornyakov this statement was proved by Bruck and Kleinfeld [8], but Kleinfeld [29] proved that even simplicity (that is, not having two-sided ideals) of an alternative but not associative ring implies that the ring is a Cayley-Dickson algebra.

If for an element a of some ring A there exists a natural number $n(a)$ such that $a^{n(a)}=0$ (with any parenthesization of the expression $a^{n(a)}$), then this element is called a *nilpotent* element. If all the elements in a ring (resp. ideal) are nilpotent, it is called a *nil-ring* (resp. *nil-ideal*).

Recently Kleinfeld [30] strengthened his results by proving that any alternative but not associative ring, in which the intersection of all the two-sided ideals is not a nil-ideal, is a Cayley-Dickson algebra over some field. Hence the class of alternative rings is much larger than the class of associative rings, but only outside the limits of the above-mentioned classes of rings.

6. Some attention has been given to right alternative rings (rings which satisfy identity (2)). Skornyakov [64] proved that every right alternative division ring is alternative. Kleinfeld [28] proved that for the alternativity of a right alternative ring it is sufficient that $[x,y,z]^2=0$ implies $[x,y,z]=0$. Smiley [68] analyzed the proof of Kleinfeld and noticed that it is sufficient to check only these cases: $x=y$, $x=yz-zy$, $x=(yz-zy)y$, $x=[y,y,z]$, or $z=wy$ and $x=[y,y,w]$ for some w. We know about the structure of free right alternative rings as little as we know about the structure of free alternative rings. The study of these rings is one of the main tasks of the theory of alternative rings.

It would be interesting to find out whether there are any identical relations which are not implied by (2)–(4) and are satisfied in the free alternative ring with three generators as, for example, the relation $(xy)z-x(yz)=0$ is satisfied by the free alternative ring with two generators.

Because alternative rings are close relatives of associative rings, we may ask of any statement which holds for associative rings whether it also holds for alternative rings. One such problem (the Kurosh problem) will be discussed in the next section.

San Soucie [51, 52] studied alternative and right alternative rings in characteristic 2 ($2x = 0$).

3. Jordan rings

1. Let A be an associative ring. If we set $a \circ b = ab + ba$, then with respect to addition and the operation $\circ$ the set of elements of the ring A becomes a ring which is in general nonassociative. We denote this ring by $A^{(+)}$. For an associative algebra B (or a Σ-operator ring) it is possible in a similar way to define an algebra $B^{(+)}$ over the same field (or a Σ-operator ring); for an algebra it is more convenient to use the operation $a \circ b = \frac{1}{2}(ab + ba)$. It is easy to check that in the ring $A^{(+)}$ the following identities hold:

$$a \circ b = b \circ a, \tag{9}$$

$$((a \circ a) \circ b) \circ a = (a \circ a) \circ (b \circ a). \tag{10}$$

Rings in which the multiplication satisfies (9) and (10) are called *J-rings* or *Jordan rings*.

It can happen that some subset of a ring, which is not a subring, becomes a Jordan ring under the operation $\circ$. As an example, consider the set of all real symmetric matrices of some fixed degree n. A Jordan ring which is isomorphic to a subring of some ring of the form $A^{(+)}$ is called a *special* Jordan ring. Special Jordan algebras can be defined in a similar way.

2. Not every Jordan ring and not every Jordan algebra is special. The classical example, that will be discussed below, of a non-special (often called exceptional) Jordan algebra of finite dimension belongs to Albert [5].

In the algebra R_8, which was discussed at the beginning of Section 2, for any element $s = (p, q)$ we set $\overline{s} = (\overline{p}, -q)$. In the set of all matrices of degree 3 with elements from the algebra R_8 we consider the subspace C_{27} of self-conjugate matrices (that is, matrices which do not change when the elements are conjugated and the matrix is transposed). It is possible to check that the set C_{27} with respect to addition, the usual multiplication of real numbers, and the operation $s \circ t = \frac{1}{2}(s \cdot t + t \cdot s)$ is a Jordan algebra of dimension 27 over the field of real numbers.

Let x be an element of the algebra R_8. Denote by x_{ij} the matrix S from the algebra C_{27} in which $s_{ij} = \overline{x}$ and $s_{ji} = x$ and all other entries are zero; by e denote the identity of the algebra R_8.

Assume that there exists an associative algebra $\mathfrak{A}$, such that the Jordan algebra $\mathfrak{A}^{(+)}$ has a subalgebra C'_{27} isomorphic to the algebra C_{27}. For simplicity in what follows we will identify the algebra C'_{27} with the algebra C_{27}. If $s, t \in C_{27}$ then it is obvious that $s \cdot t + t \cdot s = st + ts$ where st is the product of the elements s and t in the algebra $\mathfrak{A}$. The last observation allows us to easily verify the following equations:

$$e_{ij}^2 = e_{ij}e_{ij} = e_{ij} \cdot e_{ij} = e_{ii} + e_{jj}, \tag{11}$$

$$e_{ii}x_{ij} + x_{ij}e_{ii} = e_{jj}x_{ij} + x_{ij}e_{jj} = x_{ij}, \tag{12}$$
$$e_{kk}x_{ij} + x_{ij}e_{kk} = 0 \text{ (for } k \neq i, j), \tag{13}$$
$$x_{12}y_{23} + y_{23}x_{12} = (x \cdot y)_{13}, \tag{14}$$
$$x_{12}y_{13} + y_{13}x_{12} = (\overline{x} \cdot y)_{23}, \tag{15}$$
$$x_{13}y_{23} + y_{23}x_{13} = (x \cdot \overline{y})_{12}. \tag{16}$$

From equation (13) we have

$$e_{kk}(e_{kk}x_{ij} + x_{ij}e_{kk}) = (e_{kk}x_{ij} + x_{ij}e_{kk})e_{kk} = 0,$$

and because of $e_{kk}^2 = e_{kk}$, it easily follows that

$$e_{kk}x_{ij} = x_{ij}e_{kk} = 0 \ (k \neq i, j). \tag{17}$$

Setting $f_{ij} = e_{ii} + e_{jj}$, from the obvious equalities

$$f_{ij}x_{ij} + x_{ij}f_{ij} = 2x_{ij}, \quad 2f_{ij}x_{ij} = f_{ij}x_{ij} + f_{ij}x_{ij}f_{ij},$$

we easily obtain

$$f_{ij}x_{ij} = f_{ij}x_{ij}f_{ij} = x_{ij}f_{ij} = x_{ij}. \tag{18}$$

Finally,

$$e_{ii}y_{ij}e_{ii} = e_{jj}y_{ij}e_{jj} = 0, \tag{19}$$

because, for example,

$$e_{ii}y_{ij}e_{ii} = e_{ii}(y_{ij} - e_{ii}y_{ij}) = 0,$$

(equation (12)).

If $x \in R_8$ then we set $x' = e_{11}x_{12}e_{12}$. We show that the map $x \to x'$ is a homomorphism of the algebra R_8 into the algebra $\mathfrak{A}$. Clearly $(x + y)' = x' + y'$. From equations (14)–(17) it follows that

$$\begin{aligned}(x \cdot y)' &= e_{11}(x \cdot y)_{12}e_{12} = e_{11}(x_{13}\overline{y}_{23} + \overline{y}_{23}x_{13})e_{12} = e_{11}x_{13}\overline{y}_{23}e_{12}\\ &= e_{11}(x_{12}e_{23} + e_{23}x_{12})\overline{y}_{23}e_{12} = e_{11}x_{12}e_{23}\overline{y}_{23}e_{12}\\ &= e_{11}x_{12}e_{23}(y_{12}e_{13} + e_{13}y_{12})e_{12}.\end{aligned}$$

On the other hand,

$$\begin{aligned}y_{12}e_{13}e_{12} &= y_{12}e_{13}f_{13}e_{12} = y_{12}e_{13}e_{11}e_{12} = (\overline{y}_{23} - e_{13}y_{12})e_{11}e_{12}\\ &= -e_{13}y_{12}e_{11}e_{12} = -e_{13}f_{13}y_{12}e_{11}e_{12} = -e_{13}e_{11}y_{12}e_{11}e_{12} = 0,\\ e_{23}e_{13}y_{12} &= e_{23}e_{13}f_{12}y_{12} = e_{23}e_{13}e_{11}y_{12} = (e_{12} - e_{13}e_{23})e_{11}y_{12} = e_{12}e_{11}y_{12}.\end{aligned}$$

Making the corresponding substitution in the expression $(x \cdot y)'$ we get

$$(x \cdot y)' = e_{11}x_{12}e_{12}e_{11}y_{12}e_{12} = x'y'.$$

Because of the absence of proper ideals in the algebra R_8, and also because $e' = e_{11}e_{12}e_{12} = e_{11}f_{12} = e_{11} \neq 0$, we conclude that the algebra R_8 is isomorphic to a subalgebra of the associative algebra $\mathfrak{A}$, which contradicts the nonassociativity of the algebra R_8. This contradiction shows that there is no associative algebra $\mathfrak{A}$ with the required properties.

3. It would be natural to assume that special Jordan algebras satisfy some system of identities which do not follow from (9) and (10).

At the present time such identities have not been found. Moreover, every attempt to characterize special Jordan rings with the help of any system of identities must be completely unsuccessful, because Cohn [9] gave many examples of *non-special Jordan algebras which are homomorphic images of special Jordan algebras*. It was also shown by Cohn that *any homomorphic image of a special Jordan algebra with two generators is also a special Jordan algebra*.

Let $\mathfrak{B}$ be some Jordan ring. We define by the formula

$$\{a, b, c\} = (ab)c + (bc)a - (ca)b,$$

a ternary operation on the set of elements of the ring $\mathfrak{B}$. It is easy to check that if $\mathfrak{B}$ is a special Jordan ring then we have the identity

$$\{a, b, a\}^2 = \{a, \{b, a^2, b\}, a\}. \tag{20}$$

Hall [15] and Harper [17] independently proved that (20) holds for any Jordan ring. In the author's work [58] it was proved that *every Jordan ring on two generators is special*. From this result it easily follows that any identity which involves, like (20), only two variables and which holds in any special Jordan ring, also holds in any Jordan ring. This result was recently reproved by Jacobson and Paige [26].

At present it is still not known whether the identities

$$\{\{a, x, a\}, x, \{a, x, b\}\} = \{\{\{a, x, a\}, x, b\}, x, a\}, \tag{21}$$

$$\{\{x, b, x\}, a, \{x, b, x\}\} = \{x, \{b, \{x, a, x\}, b\}, x\}, \tag{22}$$

which hold in any special Jordan ring, also hold in any Jordan ring. These identities were pointed out by Jacobson; he proved in [27] that they hold in C_{27}.

Jacobson proposed the question: Does there exist a Jordan algebra which is not a homomorphic image of a special Jordan algebra?

Albert [6] proved that the algebra C_{27} is not a homomorphic image of any special Jordan algebra *of finite dimension*.

The above-mentioned problem is equivalent to the following: Is the free Jordan ring on more than two generators special or not? A positive answer would trivially imply the solution of the word problem for a free Jordan ring, but still it would not imply a solution of the problem of finding a basis for the free Jordan algebra on three or more generators (see Cohn [9]).

4. If, on the set of elements of a right-alternative ring T, we define the operation $a \circ b = ab + ba$, then it is easy to show that in this case the ring $T^{(+)}$ will be a Jordan ring. However, it turns out that the class of all Jordan rings that can be obtained in this way is equal to the class of all special Jordan rings. Indeed, the mapping $f\colon x \to R_x$ of elements of the ring T to the associative ring, generated in the ring T^* of all endomorphisms of the additive group of the ring T by right multiplications R_x ($aR_x = ax$), is a homomorphism of the ring $T^{(+)}$ onto some subring of the special Jordan ring $T^{*(+)}$. The mapping f will be an isomorphism

if we initially extend the ring T by an identity element (after which the extended ring remains right alternative).

The possibility of associating with every right alternative ring an associative ring (in general, not unique), through the corresponding (special) Jordan ring, turns out to be very useful in the study of right alternative rings, and so also in the study of alternative rings.

Using this method, the author proved in [59, 60] that all the results obtained as of the present towards solving the Kurosh problem [38] (or its special case, the Levitzky problem) for associative algebras (or rings) also hold for alternative algebras (or rings) and for special Jordan algebras (or Jordan rings). Let us formulate one of them:

An alternative ring D with a finite number of generators and the identical relation $x^n = 0$ is nilpotent, that is, there exists a natural number N such that any product of N elements of D is zero.

The closest generalization of Jordan rings are the so-called noncommutative Jordan rings, the study of which was started by Schafer. The natural place for them in the present article is in the last section.

4. Lie rings

1. A ring which satisfies the identical relations

$$x^2 = 0, \tag{23}$$

$$(xy)z + (yz)x + (zx)y = 0, \tag{24}$$

is called a *Lie ring*.

In this article we completely avoid the discussion of Lie algebras of finite dimension, an exposition of which would be more natural in connection with the theory of Lie groups.

If, in an associative ring A we define a new operation by the equation $a \cdot b = ab - ba$, then the set of elements of A will be a Lie ring with this operation and addition. We denote this new ring by $A^{(-)}$. Birkhoff [7] and Witt [71] independently proved that *every Lie algebra is isomorphic to a subalgebra of some algebra of the form $A^{(-)}$*. If we use the terminology of Jordan rings, then we can say that every Lie ring is special.

Lazard [42] and Witt [72] studied representations of Σ-operator Lie rings in Σ-operator associative rings. The existence of such a representation was proved by them in the case when Σ is a principal ideal domain, and in particular for Lie rings without operators. The example constructed by the author in [57] shows that there exist non-representable Σ-operator Lie rings which do not have elements of finite order in the additive group.

I.D. Ado [1, 2] proved that any finite-dimensional Lie algebra over the field of complex numbers can be represented in a finite-dimensional associative algebra.

Later Harish-Chandra [16] and Iwasawa [24] proved that Ado's theorem holds for any finite-dimensional Lie algebra.

We mention the cycle of works of Herstein [19]–[21], which, in essence, belong to the theory of associative rings and are dedicated to studying the ring $A^{(-)}$ under various assumptions on the ring A.

2. There are interesting relations between the theory of Lie rings and the theory of groups.

Let K be the ring of formal power series with rational coefficients in the noncommutative variables x_i $(i = 1, 2, \ldots)$. Magnus [45] proved that the elements $y_i = 1 + x_i$ of the ring K generate a free subgroup G of the multiplicative group of the ring K, and that every element of the subgroup G_n (the n-th commutator subgroup[5]) has the form $1+\ell_n+\omega$, where ℓ_n is some homogeneous Lie polynomial (with respect to the operations $a \cdot b$ and $a + b$) of degree n in the generators x_i, and ω is a formal power series in which all the terms have degree greater than n. Then because of known criteria [11, 12, 44] which allow us to determine whether a given polynomial is a Lie polynomial, the above mentioned representation of the free group allows us to determine whether any given element lies in one term or another of the lower central series.

The elements $z_i = e^{x_i}$ of the ring K also generate a free group [46] and if $e^x e^y = e^t$ then t is a power series, the terms of which are homogeneous Lie polynomials in x and y [18].

The relations which exist between the theory of groups and the theory of Lie rings allow us to obtain group-theoretical results from statements proved for Lie rings. For example, Higman [23] proved nilpotency (see the definition below) of any Lie ring which has an automorphism of prime order without nonzero fixed points. This statement allowed him to prove nilpotency of finite solvable groups which have an automorphism satisfying the analogous conditions.

Earlier Lazard [43] studied nilpotent groups using extensively the apparatus of Lie ring theory.

3. We consider one more circle of questions which are relevant to the theory of groups.

A Lie ring L is called a ring satisfying the *n-th Engel condition* if for any elements x and y we have the relation

$$\{\cdots[(x\underbrace{y)y]\cdots\}y}_{n\ y\text{'s}} = 0.$$

We introduce the following notation:

$$L = L^1 = L^{(1)}, \qquad L^k = L^{k-1}L, \qquad L^{(k)} = L^{(k-1)}L^{(k-1)}.$$

A Lie ring is called *nilpotent* (resp. *solvable*) if there exists a natural number m such that $L^m = 0$ (resp. $L^{(m)} = 0$).

[5]That is, the n-th term of the lower central series. [Translators]

With some restrictions on the additive group, Higgins [22] proved that solvable rings satisfying the n-th Engel condition are nilpotent. Then Cohn [10] constructed an example of a solvable Lie ring whose additive group is a p-group and which satisfies the p-th Engel condition, but is not nilpotent. For Lie rings with a finite number of generators and some restrictions on the additive group, A.I. Kostrikin [37] proved that the Engel condition implies nilpotency. This result is especially interesting because from it follows the positive solution of the group-theoretical restricted Burnside problem for p-groups with elements of prime order [35, 36].

An element a in a Lie algebra L is called *algebraic* if the endomorphism $R_a : x \mapsto xa$ generates a finite-dimensional subalgebra in the (associative) algebra of all endomorphisms of the additive group of the algebra L.

It is not known whether there exists a Lie algebra with a finite number of generators and infinite dimension in which every element is algebraic. This problem is analogous to the famous Kurosh problem [38] for associative algebras.

We mention one easier but unsolved problem. Let the Lie algebra L be such that any two elements belong to a subalgebra, the dimension of which does not exceed some fixed number. Does it follow from this that every finite subset of the algebra L belongs to some subalgebra of finite dimension?

4. An important role in the theory of Lie rings is played by free Lie rings. In contrast to free alternative rings and free Jordan rings, free Lie rings have been thoroughly studied. M. Hall [14] pointed out a method for constructing a basis of a free Lie algebra; E. Witt [71] found a formula for computing the rank of the homogeneous modules in a free Lie algebra on a finite number of generators.

We briefly describe one constructive method of building a free Lie ring. Let $\mathfrak{A}$ be a free associative Σ-operator ring with some set $R = \{a_i\}$ $(i = 1, 2, \ldots, k)$ as a set of free generators. It turns out that [61] the elements of the set R generate in the Lie ring $\mathfrak{A}^{(-)}$ a free Lie ring L for which they are free generators. We order the elements of the set R in some way, and then we order lexicographically every set of (associative) monomials of the same degree in the elements of the set R. Let W be the set of all monomials w such that $w = w_1 w_2 > w_2 w_1$, for any representation of the monomial w as a product of two monomials w_1 and w_2. Let $v \in W$ with $v = v_1 v_2$ where v_1 is a monomial from W of minimal degree such that $v_2 \in W$. We parenthesize the monomial v in the following way: $v = (v_1)(v_2)$, and we repeat this method of parenthesization on the monomials v_1 and v_2. The set of nonassociative monomials obtained from the elements of the set W by this method of parenthesization with the operation interpreted as $a \cdot b = ab - ba$ will be a basis of the ring L.

The author in [56] and independently Witt in [73] proved that *any subalgebra of a free Lie algebra is again free*. This theorem is analogous to the theorem of Kurosh mentioned in Section 1 for subalgebras of free algebras.

Using the above method of constructing a free Lie algebra allowed the author in [61] to prove that *any Lie algebra of finite or countable dimension can be embedded in a Lie algebra with two generators.*

Analogous theorems about embedding of arbitrary algebras and of associative rings were proved respectively by A.I. Zhukov [74] and A.I. Malcev [48].

5. The study of Lie algebras over fields of prime characteristic has led to the discussion of so-called restricted Lie algebras.

In a restricted Lie algebra over a field of characteristic $p > 0$ an additional unary operation is defined with some natural axioms which are typical of the usual (associative) p-th power. Jacobson [25] proved a theorem for restricted Lie algebras analogous to the Birkhoff-Witt theorem, which in this case already includes a theorem similar to Ado's theorem.

6. Recently A.I. Malcev [49] considered a class of binary-Lie rings, which are related to Lie rings in a way analogous to the way alternative rings are related to associative rings. A ring is called *binary-Lie* if every two elements lie in some Lie subring.

A.T. Gainov [13] proved that in the case of a ring without elements of order 2 in the additive group, for a ring to be binary-Lie it is sufficient that these identities hold:

$$x^2 = [(xy)y]x + [(yx)x]y = 0.$$

If, on the set of elements of some alternative ring D, we define the above described operation $a \cdot b = ab - ba$, then in the ring $D^{(-)}$, as was shown by A.I. Malcev [49], these relations hold identically:

$$x^2 = [(x \cdot y) \cdot z] \cdot x + [(y \cdot z) \cdot x] \cdot x + [(z \cdot x) \cdot x] \cdot y - (x \cdot y) \cdot (x \cdot z) = 0. \quad (25)$$

Rings satisfying the identities (25) are called by A.I. Malcev *Moufang-Lie* rings, and he also showed that the class of Moufang-Lie rings[6] without elements of additive order 6 is properly contained in the class of binary-Lie rings.

Recently Kleinfeld [31] proved that *a Moufang-Lie ring M without elements of additive order 2 which has an element a such that $aM = M$ is a Lie ring.* A corresponding result can clearly be formulated in the language of alternative rings.

The problem of the truth of a theorem, similar to the Birkhoff-Witt theorem, connecting the theory of Moufang-Lie rings with the theory of alternative rings remains open.

5. Some wider classes of rings

1. As was shown earlier, a ring is alternative if and only if every two elements lie in some associative subring.

Algebraists working in the theory of rings have been attracted for a long time to the wider class of rings with associative powers. A ring is called *power-associative*

[6]Now called Malcev rings. [Translators]

if every element lies in some associative subring. It is not difficult to check that all the classes of rings discussed in the present article are power-associative.

In the case of rings for which the additive group has no torsion, Albert [3] has shown that the identities $x^2x = xx^2$ and $(x^2x)x = x^2x^2$ are sufficient to guarantee power-associativity. This result was recently given another proof by A.T. Gainov [13]. Albert proved in [4] that if in the additive group of a ring there are no elements of order 30 then power-associativity follows from the identities

$$(xy)x = x(yx) \quad \text{and} \quad (x^2x)x = x^2x^2.$$

For rings of small characteristic some sufficient conditions for power-associativity were found by Kokoris [32, 33].

2. We mention one method for studying power-associative rings which has been used extensively in the works of Albert.

Let A be a commutative power-associative ring in which the equation $2x = a$ has a unique solution for every $a \in A$ and which contains an idempotent e ($e^2 = e$). Then it turns out that every element $b \in A$ has a unique representation in the form $b = b_0 + b_1 + b_{1/2}$ where $b_\lambda e = \lambda b_\lambda$; that is, the ring A can be represented as the direct sum of three modules $A = A_0 + A_1 + A_{1/2}$, the study of which gives some information about the ring A. If the ring A is noncommutative, then we can study the commutative ring $A^{(+)}$ which is obtained from the ring A with the help of the new multiplication $a \circ b = \frac{1}{2}(ab + ba)$. It is obvious that the subrings generated by a single element in the rings A and $A^{(+)}$ are the same. Therefore the ring $A^{(+)}$ is again power-associative.

Another very wide class of rings is the class of flexible rings; that is, rings which satisfy the identical relation (4). All the rings discussed in this article, except for right alternative rings, are from this class.

No significant results, which would go beyond the class of algebras of finite dimension, have been obtained for flexible rings.

3. It would be natural to expect a deeper study of flexible power-associative rings.

However, comparatively recently Schafer [53] began the study of the class of so-called noncommutative Jordan rings, defined by identities (4) and (10), which is slightly narrower than the class of flexible power-associative rings, but contains most of the rings mentioned above.

The study of this class of rings at the present time is restricted to the theory of algebras of finite dimension (see [54, 55, 34]); however, we can hope that in the future a sufficiently interesting theory of this class of rings will be constructed.

In conclusion, we mention one very wide class, the so-called *power-commutative rings*; that is, rings in which every element belongs to a commutative (but not necessarily associative) subring. This class includes not only the flexible rings, but also the power-associative rings. Unfortunately, at this point in time, we do not even know whether this class can be defined by a finite system of identities.

References

[1] I.D. Ado, *On representations of finite continuous groups using linear substitutions*, Izv. Kaz. fiz.-matem. ob-va 7 (1934–35) 1–43.

[2] I.D. Ado, *Representation of Lie algebras by matrices*, Uspekhi Mat. Nauk II (1947) 159–173.

[3] A.A. Albert, *On the power-associativity of rings*, Summa Brasil. Math. 2 (1948) 21–33.

[4] A.A. Albert, *Power-associative rings*, Trans. Amer. Math. Soc. 64 (1948) 552–593.

[5] A.A. Albert, *A note on the exceptional Jordan algebra*, Proc. Nat. Acad. Sci. USA 36 (1950) 372–374.

[6] A.A. Albert, *A property of special Jordan algebras*, Proc. Nat. Acad. Sci. USA 42 (1956) 624–625.

[7] G. Birkhoff, *Representability of Lie algebras and Lie groups by matrices*, Ann. of Math. 38 (1937) 526–532.

[8] R.N. Bruck, E. Kleinfeld, *The structure of alternative division rings*, Proc. Amer. Math. Soc. 2 (1951) 878–890.

[9] P.M. Cohn, *On homomorphic images of special Jordan algebras*, Canad. Journ. Math. 6 (1954) 253–264.

[10] P.M. Cohn, *A non-nilpotent Lie ring satisfying the Engel condition and a non-nilpotent Engel group*, Proc. Cambridge Philos. Soc. 51 (1955) 401–405.

[11] E.B. Dynkin, *Computation of the coefficients in the Campbell-Hausdorf formula*, Doklady Akad. Nauk USSR 57 (1947) 323–326.

[12] K.O. Friedrichs, *Mathematical aspects of the quantum theory of fields, V*, Comm. Pure Appl. Math. 6 (1953) 1–72.

[13] A.T. Gainov, *Identitical relations for binary-Lie rings*, Uspekhi Mat. Nauk XII (1957) 141–146.

[14] M. Hall, *A basis for free Lie rings and higher commutators in free groups*, Proc. Amer. Math. Soc. 1 (1950) 575–581.

[15] M. Hall, *An identity in Jordan rings*, Proc. Amer. Math. Soc. 42 (1956) 990–998.

[16] Harish-Chandra, *Faithful representations of Lie algebras*, Ann. of Math. 50 (1949) 68–76.

[17] L.R. Harper, *A proof of an identity for Jordan algebras*, Proc. Nat. Acad. Sci. USA 42 (1956) 137–139.

[18] E. Hausdorf, *Die simbolische Exponentialformel in der Gruppentheorie*, Bericht d. König. Sächs. Ges. d. wiss. Math.-Phis. Klasse 58 (1906) 19–48.

[19] I.N. Herstein, *On the Lie and Jordan rings of a simple associative ring*, Amer. Journ. Math. 77 (1955) 279–285.

[20] I.N. Herstein, *The Lie ring of a simple associative ring*, Duke Math. Journ. 22 (1955) 471–476.

[21] I.N. Herstein, *Lie and Jordan systems in simple rings with involution*, Amer. Journ. Math. 78 (1956) 629–649.

[22] P.J. Higgins, *Lie rings satisfying the Engel condition*, Proc. Cambridge Philos. Soc. 50 (1954) 8–15.

[23] G. Higman, *Groups and rings having automorphisms without non-trivial fixed elements*, Journ. London Math. Soc. 32 (1957) 321–332.

[24] K. Iwasawa, *On the representation of Lie algebras*, Jap. Journ. Math. 19 (1948) 405–426.

[25] N. Jacobson, *Restricted Lie algebras of characteristic p*, Trans. Amer. Math. Soc. 50 (1941) 15–25.

[26] N. Jacobson, L.J. Paige, *On Jordan algebras with two generators*, Journ. Math. and Mech. 6 (1957) 895–906.

[27] N. Jacobson, *Jordan algebras*, Report of a Conference on Linear Algebras (1957) 12–19.

[28] E. Kleinfeld, *Right alternative rings*, Proc. Amer. Math. Soc. 4 (1953) 939–944.

[29] E. Kleinfeld, *Simple alternative rings*, Ann. of Math. 58 (1953) 544–547.

[30] E. Kleinfeld, *Generalization of a theorem on simple alternative rings*, Portugal. Math. 14, 3–4 (1955) 91–94.

[31] E. Kleinfeld, *A note on Moufang-Lie rings*, Proc. Amer. Math. Soc. (1958) 72–74.

[32] L.A. Kokoris, *New results on power-associative algebras*, Trans. Amer. Math. Soc. 77 (1954) 363–373.

[33] L.A. Kokoris, *Power-associative rings of characteristic two*, Proc. Amer. Math. Soc. 6 (1955) 705–710.

[34] L.A. Kokoris, *Some nodal noncommutative Jordan algebras*, Proc. Amer. Math. Soc. 9 (1958) 164–166.

[35] A.I. Kostrikin, *On the relation between periodic groups and Lie rings*, Izv. Akad. Nauk USSR 21 (1957) 289–310.

[36] A.I. Kostrikin, *Lie rings satisfying the Engel condition*, Izv. Akad. Nauk USSR 21 (1957) 515–540.

[37] A.I. Kostrikin, *On the Burnside problem*, Doklady Akad. Nauk USSR 119 (1958) 1081–1084.

[38] A.G. Kurosh, *Problems in the theory of rings related to the problem of Burnside on periodic groups*, Izv. Akad. Nauk USSR 5 (1941) 233–247.

[39] A.G. Kurosh, *Free nonassociative algebras and free products of algebras*, Matem. Sb. 20, 62 (1947) 239–262.

[40] A.G. Kurosh, *The current state of the theory of rings and algebras*, Uspekhi Mat. Nauk VI, 2 (1951) 3–15.

[41] A.G. Kurosh, *Nonassociative free sums of algebras*, Mat. Sbornik 37, 79 (1955) 251–264.

[42] M. Lazard, *Sur les algèbres enveloppantes universelles des certaines algèbres de Lie*, Publ. Sci. de l'Univ. d'Alger, Ser. A 1 (1954) 281–294.

[43] M. Lazard, *Sur les groupes nilpotents et les anneaux de Lie*, Ann. Sci. Ec. Norm. Sup. 71 (1954) 101–190.

[44] R.C. Lyndon, *A theorem of Friedrichs*, Michigan Math. Journ. 3 (1955–56) 27–29.

[45] W. Magnus, *Über Beziehungen zwischen höheren Kommutatoren*, Journ. reine und angew. Math. 177 (1937) 105–115.

[46] W. Magnus, *Über Gruppen und zugeordnete Liesche Ringe*, Journ. reine und angew. Math. 182 (1940) 142–149.

[47] A.I. Malcev, *On the decomposition of an algebra into the direct sum of the radical and a semi-simple subalgebra*, Doklady Akad. Nauk USSR 36 (1942) 46–50.

[48] A.I. Malcev, *On a representation of nonassociative rings*, Uspekhi Mat. Nauk VII, 1 (1952) 181–185.

[49] A.I. Malcev, *Analytic loops*, Mat. Sbornik 36 (1955) 569–576.

[50] I.N. Sanov, *Investigations into the relation between periodic groups of prime period and Lie rings*, Izv. Akad. Nauk USSR 16 (1952) 23–58.

[51] R.L. San Soucie, *Right alternative division rings of characteristic 2*, Proc. Amer. Math. Soc. 6 (1955) 291–296.

[52] R.L. San Soucie, *Right alternative rings of characteristic two*, Proc. Amer. Math. Soc. 6 (1955) 716–719.

[53] R.D. Schafer, *Non-commutative Jordan algebras of characteristic zero*, Proc. Amer. Math. Soc. 6 (1955) 472–475.

[54] R.D. Schafer, *On non-commutative Jordan algebras*, Proc. Amer. Math. Soc. 9 (1958) 110–117.

[55] R.D. Schafer, *Restricted non-commutative Jordan algebras of characteristic p*, Proc. Amer. Math. Soc. 9 (1958) 141–144.

[56] A.I. Shirshov, *Subalgebras of free Lie algebras*, Mat. Sbornik 33 (1953) 441–452.

[57] A.I. Shirshov, *On the representation of Lie rings in associative rings*, Uspekhi Mat. Nauk VIII (1953) 173–175.

[58] A.I. Shirshov, *On special J-rings*, Mat. Sbornik 38 (1956) 149–166.

[59] A.I. Shirshov, *On some nonassociative nilrings and algebraic algebras*, Mat. Sbornik 41 (1957) 381–394.

[60] A.I. Shirshov, *On rings with identical relations*, Mat. Sbornik 43 (1957) 277–283.

[61] A.I. Shirshov, *On free Lie rings*, Mat. Sbornik 45 (1958) 113–122.

[62] L.A. Skornyakov, *Alternative division rings*, Ukr. Matem. Zhurn. 2, 1 (1950) 70–85.

[63] L.A. Skornyakov, *Alternative division rings of characteristic* 2 *and* 3, Ukr. Matem. Zhurn. 2, 3 (1950) 94–99.

[64] L.A. Skornyakov, *Right-alternative division rings*, Izv. Akad. Nauk USSR 15 (1951) 177–184.

[65] L.A. Skornyakov, *Projective planes*, Uspekhi Mat. Nauk VI, 6 (1951) 112–154.

[66] L.A. Skornyakov, *Representation of nonassociative rings in associative rings*, Doklady Akad. Nauk USSR 102, 1 (1955) 33–35.

[67] M.F. Smiley, *Kleinfeld's proof of the Bruck-Kleinfeld-Skornyakov theorem*, Math. Ann. 134 (1957) 53–57.

[68] M.F. Smiley, *Jordan homomorphisms and right alternative rings*, Proc. Amer. Math. Soc. 8 (1957) 668–671.

[69] Yu.I. Sorkin, *Rings as sets with one operation which satisfy only one relation*, Uspekhi Mat. Nauk XII, 4 (1957) 357–362.

[70] Liu-Shao Syue, *On decomposition of locally finite algebras*, Mat. Sbornik 39 (1956) 385–396.

[71] E. Witt, *Treue Darstellung Liescher Ringe*, Journ. reine und angew. Math. 177 (1937) 152–160.

[72] E. Witt, *Treue Darstellung beliebiger Liescher Ringe*, Collect. Math. 6 (1953) 107–114.

[73] E. Witt, *Die Unterringe der freien Liescher Ringe*, Mat. Zeitschr. 64 (1956) 195–216.

[74] A.I. Zhukov, *Complete systems of defining relations in nonassociative algebras*, Mat. Sbornik 27, 69 (1950) 267–280.

Some Algorithmic Problems for ε-algebras

A.I. Shirshov

Introduction

The word problem[1], stated relative to one or another algebraic system, has attracted the attention of many mathematicians. In the works of A.A. Markov [1] and E. Post [3] it was proved for the first time that there exist algebraic systems (semigroups) with undecidable word problem. The most significant achievement in this direction is the result of P.S. Novikov [2] that establishes undecidability of the word problem for groups. In 1950, A.I. Zhukov [5], while studying free nonassociative algebras, established that in the case in which one does not assume that the algebra satisfies any identical relation (for instance, associativity) the word problem (as well as some other algorithmic problems) is decidable. From the results obtained for semigroups, it easily follows that the word problem is undecidable for associative algebras.

The above-mentioned facts show that it is of interest to discover classes of algebras defined by identical relations for which the word problem or some other algorithmic problems are decidable. In the present work, the word problem is solved for commutative and anticommutative algebras (ε-algebras). Moreover, in these cases, the more general membership problem is solved, and a theorem is proved that is analogous to a known theorem on freeness in group theory.

1. The word problem

In the study of commutative and anticommutative algebras, we will for brevity use the terminology introduced in the work [4]. Hence, commutative and anticommutative algebras will be called respectively C-algebras and AC-algebras. The term ε-algebras with $\varepsilon = C$ or $\varepsilon = AC$ will be used when there is no need to distinguish

Sibirsk Mat. Zh. 3, (1962), no. 1, 132–137.

[1]Literally, "the problem of identity". [Translators]

the two cases. In [4] the definition of ε-regular words is given, and it is shown that they form a basis of the free ε-algebra. This result will also be used in the sequel without further mention.

Let E be the free ε-algebra over some field P (fixed once and for all), and let $R = \{a_\alpha\}$, $\alpha \in I$, be a set of free generators. We choose in E an arbitrary finite set of elements S and denote by $\langle S\rangle$ the ideal generated in E by S. To solve the word problem in this case means to provide an algorithm that allows us, for an arbitrary finite set S and an arbitrary element $a \in E$, to determine whether or not a belongs to $\langle S\rangle$.

The definition of ε-regular words requires that some ordering be fixed. In the sequel we will make the convention that, given two ε-regular words $u = u_1u_2$ and $v = v_1v_2$ of equal length ≥ 2, the greater word is either the one with greater first factor (u_1 or v_1), or if these are equal, then the one with greater second factor (u_2 or v_2). With respect to this ordering, we will speak of the leading term $\overline{a}$ of any element a in the algebra E.

The concept of a subword of a nonassociative word is sufficiently well known. Formally, it can be defined (by induction) on the word length, for example as follows.

Definition 1. Let u be a word of length ≥ 2 with $u = u_1u_2$. Then u, u_1, u_2 and the subwords of u_1 and u_2 are called *subwords* of u.

Definition 2. A set S of elements of E is called *reduced* if no element of S has a leading term which is a subword of the leading term of another element of S, and all the coefficients of the leading terms are equal to 1.

We now prove a few auxiliary results.

Lemma 1. *Let S be a finite set of elements of E. Then there exists a reduced finite set S' such that $\langle S'\rangle = \langle S\rangle$.*

Proof. In the expression of the elements of S, there occurs only a finite subset R' of elements of R. Let s_i, $i = 1, 2, \ldots, n$, be the elements of S, and let $\overline{s}_i$ be the leading term of s_i; then obviously we may assume that the coefficients of the leading terms of the elements of S are equal to 1. The symbol $\Sigma = (\overline{s}_1, \overline{s}_2, \ldots, \overline{s}_n)$ where $\overline{s}_i \leq \overline{s}_j$ if $i > j$ will be called the *type* of the set S, and the number n will be called the *length* of the type.

The set of all possible types that correspond to finite subsets of E, the expressions of whose elements only involve elements from R', will be ordered as follows: if the lengths of the types are equal then the order is lexicographical, and shorter types precede longer types.

Suppose that the finite set S is not reduced, i.e., the leading term $\overline{s}_j$ of some element $s_j \in S$ is a subword of the leading term $\overline{s}_i$ of an element $s_i \in S$, $i \neq j$. Then obviously there exists an element t_j of the ideal $\langle s_j\rangle$ such that $\overline{s}_i = \overline{t}_j$ and hence the leading term $\overline{d}_i$ of the element $d_i = s_i - t_j$ will be smaller than the word $\overline{s}_i$. Denote by S_1 the set obtained from S by replacing the element s_i by the

element d_i. Obviously $\langle S\rangle = \langle S_1\rangle$ and the type of S_1 is smaller than the type of S. The proof is complete since any decreasing sequence of types must terminate. □

Lemma 2. *An element $t \in E$ lies in the ideal $\langle S\rangle$, where $S = \{s_i\}$, $i = 1, 2, \ldots, n$, is a finite reduced set, only if at least one of the words $\overline{s}_i$, $i = 1, 2, \ldots, n$, is a subword of $\overline{t}$.*

Proof. If $t \in \langle S\rangle$ then obviously t can be represented as a linear combination of products d_i, $i = 1, 2, \ldots, m$, of one of the elements[2] s_{k_i} of S and some number of ε-regular words. Here we may assume that each $\overline{d}_i$ is an ε-regular word that has a subword $\overline{s}_{k_i}$, and replacing this subword by s_{k_i} turns $\overline{d}_i$ into d_i.

The last statement is obvious if $\varepsilon = C$, but it requires additional considerations if $\varepsilon = AC$. In this second case, one should look at products of the form $s_i\overline{s}_i$. But then, by virtue of the equation

$$s_i\overline{s}_i = s_i[s_i - (s_i - \overline{s}_i)] = -s_i(s_i - \overline{s}_i),$$

it is clear that in this case also the required representation is possible. In the more general case of the expression $\sigma_i\overline{\sigma}_i$, where $\overline{\sigma}_i$ is an AC-regular word with a distinguished subword $\overline{s}_{k_i}$ satisfying the above conditions, the argument is similar.

Among the ε-regular words $\overline{d}_i$, $i = 1, 2, \ldots, m$, we select the maximal. If this word is unique, then the lemma is proved. Assume now that the maximal word $\overline{d}_i$ is equal to the word $\overline{d}_j$. Since S is reduced, each of the subwords $\overline{s}_{k_i}$, $\overline{s}_{k_j}$ of the word $\overline{d}_j$ does not occur as a subword of the other in the expression of the word $\overline{d}_j$ (although they can be equal[3]). Therefore, without loss of generality, we may assume that

$$\overline{d}_j = b_1 b_2 \cdots b_{p_j}\overline{s}_{k_i} c_1 c_2 \cdots c_{q_j}\overline{s}_{k_j} f_1 f_2 \cdots f_{r_j},$$

where parentheses are placed in a certain way and all b, c, f are ε-regular words. By virtue of the equation

$$\begin{aligned} d_j = {} & b_1 b_2 \cdots b_{p_j} s_{k_i} c_1 c_2 \cdots c_{q_j}\overline{s}_{k_j} f_1 f_2 \cdots f_{r_j} \\ & + b_1 b_2 \cdots b_{p_j}\overline{s}_{k_i} c_1 c_2 \cdots c_{q_j}(s_{k_j} - \overline{s}_{k_j}) f_1 f_2 \cdots f_{r_j} \\ & + b_1 b_2 \cdots b_{p_j}(\overline{s}_{k_i} - s_{k_i}) c_1 c_2 \cdots c_{q_j} s_{k_j} f_1 f_2 \cdots f_{r_j}, \end{aligned}$$

it is obvious that the element d_j can be written as a linear combination of the element d_i and some other elements formed in a similar way to the elements d_k, $k = 1, 2, \ldots, m$, but having smaller leading terms. After combining like terms, the number of elements d_k whose leading terms coincide and are maximal is reduced by 1. The proof is complete by an obvious induction. □

From Lemmas 1 and 2 we easily obtain the following algorithm which solves the word problem for ε-algebras as stated at the beginning of this section:

(a) In a finite number of steps one performs replacement of the set S by the reduced set S' (Lemma 1).

[2]In this proof, we have replaced $s_{i_{k(i)}}$, $s_{j_{k(j)}}$ by s_{k_i}, s_{k_j} respectively. [Translators]

[3]The original says "although they can be subwords of each other". [Translators]

(b) If the word $\overline{a}$ does not contain a subword that coincides with the leading term of one of the elements S', then Lemma 2 implies that $a \notin \langle S \rangle$. If to the contrary such a subword is found, then it is easy to construct an element $a_1 \in \langle S \rangle$ such that $\overline{a} = \overline{a}_1$ and hence $\overline{a - a_1}$ will be less that $\overline{a}$.

It is easy to see that the element a lies in the ideal $\langle S \rangle$ if and only if the element $b = a - a_1$ lies in $\langle S \rangle$. The rest is obvious.

For ε-algebras, as well as nonassociative algebras [5], we have the following result.

Theorem. (Freeness Theorem) *Suppose that the expression of an element $c \in E$, in terms of the elements of the basis of ε-regular words, contains the generating element $a_\alpha \in R$. Then the images of the elements of the set $R \setminus \{a_\alpha\}$ generate a free ε-algebra in the quotient $E/\langle c \rangle$.*

Proof. In the construction of the basis of ε-regular words, we make the convention that for two such words the greater is the one whose expression contains the generator a_α more times, regardless of the degrees of the words. The words that contain the generator a_α the same number of times will be ordered in the usual way. Obviously, any subword v of the word u will be smaller than this word u, and any decreasing sequence of words that are ε-regular (in this sense) must terminate.

The proof of Theorem 1 of the work [4] can be applied to this situation without essential changes. The above way of ordering ε-regular words guarantees that the leading term $\overline{c}$ of an element c contains the generator a_α. Lemma 2, whose proof is still valid, states in our case that the maximal word $\overline{v}$, of any element v in the ideal $\langle c \rangle$, contains a subword that coincides with $\overline{c}$. From this it follows that the (free) subalgebra E' of E generated by the set $R \setminus \{a_\alpha\}$ has zero intersection with the ideal $\langle c \rangle$. This is equivalent to the statement of the theorem. □

2. The membership problem

The word problem is a special case of the so-called membership problem, which for the case considered in this paper has the following formulation:

> An arbitrary finite set $V = \{v_j\}$, $j = 1, 2, \ldots, k$, of elements of the algebra E generates a subalgebra $[V]$. It is necessary to find an algorithm that allows us to determine whether or not the image of an arbitrary element $t \in E$, under the natural homomorphism of E onto the quotient algebra $E' = E/\langle S \rangle$ where $S = \{s_i\}$, $i = 1, 2, \ldots, n$, belongs to the image $[V]'$ of $[V]$ under this homomorphism.

Obviously, when considering the membership problem for sets S and V and elements t, one can make the following assumptions without loss of generality:

(1) The set S is reduced.
(2) None of the words $\overline{t}$ and $\overline{v}_j$, $j = 1, 2, \ldots, k$, contains any of the words $\overline{s}_i$ as a subword.

(3) The coefficients of the leading terms of the elements t and v_j, $j = 1, 2, \ldots, k$, are equal to 1.
(4) Each element $\overline{v}_j$ does not belong to the subalgebra of E generated by the leading terms of the elements of the set $V \setminus \{v_j\}$.

One can achieve Conditions (1)–(4) in a finite number of steps without changing the ideal $\langle S\rangle$, the subalgebra V, or the image t' of the element t. The proof of this fact essentially repeats the argument given in the proof of Lemma 1. For example, if it happens that some element $\overline{v}_j$ belongs to the subalgebra generated by the elements $\overline{v}_i$, $i \neq j$, then instead of the element v_j one should consider the difference $v'_j = v_j - u_j$ where $u_j \in [V]$ and $\overline{v}'_j < \overline{v}_j$.

Let λ be the maximum of the degrees of the elements of S. We will describe a process for modifying the set V. Suppose some element $\overline{s}_i$ has the form $\overline{s}_i = \overline{v}_{i_1}\overline{v}_{i_2}\cdots\overline{v}_{i_q}$ with some placement of parentheses. Then to the set V we adjoin the element $v' = v_{i_1}v_{i_2}\cdots v_{i_q} - s_i$ with the same placement of parentheses. Note that the degrees of the elements v_{i_k} that appear in the expression of the element v' are less than λ. If necessary, to the set $V' = V \cup \{v'\}$ we apply the transformations which ensure Conditions (2)–(4). We repeat this entire process as many times as required.

Since, after each step, the set of words of degree $\leq \lambda$, which can be obtained by multiplying the leading terms of the elements of the corresponding $V^{(i)}$, can only increase, and the number of ε-regular words of degree $\leq \lambda$ that occur in this process is finite, the process will lead in the end to a set V_1 satisfying the following conditions:

– Conditions (2)–(4) above;
– the images of the algebras $[V_1]$ and $[V]$, under the natural homomorphism of the algebra E onto the quotient algebra $E/\langle S\rangle$, coincide;
– if for some placement of parentheses $\overline{s}_j = \overline{v}_{j_1}\overline{v}_{j_2}\cdots\overline{v}_{j_p}$ for some j, j_1, j_2, …, j_p, then the element $s_j - v_{j_1}v_{j_2}\cdots v_{j_p}$ can be represented as the sum $w + \tau$ where $w \in [V_1]$, $\tau \in \langle S\rangle$, $\overline{w} < \overline{s}_j$ and $\overline{\tau} < \overline{s}_j$.

The totality of these conditions imposed on the sets S and V_1 and the element t will be called for brevity *Condition* (5).

Lemma 3. *The image of an element $t \in E$ belongs to the image of the subalgebra $[V_1]$ under the natural homomorphism of E onto the quotient algebra $E/\langle S\rangle$ only if $t \in [\overline{V}_1]$ where $\overline{V}_1$ is the set of leading terms of the elements of V_1; here we assume that Condition* (5) *is satisfied.*

Proof. Indeed, suppose that $t = u + \sigma$ where $u \in [V_1]$ and $\sigma \in \langle S\rangle$. The leading term $\overline{\sigma}$ of σ contains some word $\overline{s}_p$ as a subword where $s_p \in S$ (Lemma 2). Obviously, $\overline{u} \in [\overline{V}_1]$. If $\overline{\sigma} \neq \overline{u}$ (ignoring the coefficients) then the lemma is proved, since $\overline{t} \neq \overline{\sigma}$ by Condition (2) and therefore $\overline{t} = \overline{u}$.

Now assume that $\overline{u} = \overline{\sigma}$. Then for the element $\overline{s}_p$ that is a subword of $\overline{\sigma}$ we have the representation $\overline{s}_p = \overline{v}_{p_1}\overline{v}_{p_2}\cdots\overline{v}_{p_\ell}$ with some placement of parentheses. By Condition (5) we have $s_p - v_{p_1}v_{p_2}\cdots v_{p_\ell} = u' + \sigma'$ where $u' \in [V_1]$, $\sigma' \in \langle S\rangle$,

$\overline{u}' < \overline{s}_p$ and $\overline{\sigma}' < \overline{s}_p$. Thus $s_p = u'' + \sigma'$, $u'' \in [V_1]$. The element σ can be written in the form $\sigma = \sigma_1 + \sigma_2$ where σ_1 is obtained by replacing the subword $\overline{s}_p$ in $\overline{\sigma}$ by the element s_p, and also σ_2 is in $\langle S\rangle$ with $\overline{\sigma}_2 < \overline{\sigma}_1$. Obviously, $\overline{\sigma}_1 = \overline{\sigma} = \overline{u}$. Replacing the factor s_p in σ_1 by the expression $u'' + \sigma'$, we obtain the following expression for the element t: $t = u + u_1 + \sigma_3$ where $u_1 \in [V_1]$ and $\sigma_3 \in \langle S\rangle$ with $\overline{\sigma}_3 < \overline{\sigma}$. The process of decreasing the leading terms of the summands in $\langle S\rangle$ that occur in the expression for t cannot continue indefinitely. The proof is completed by the obvious remark that the condition $u \in [V_1]$ implies $\overline{u} \in [\overline{V}_1]$. The lemma is proved. □

Lemma 3 implies the following algorithm for solving the membership problem for ε-algebras as stated above:

(a) Rewrite the element t, and the elements of the sets V and S, so that they satisfy Conditions (1)–(4).
(b) Extend the set V to the set V_1 satisfying Condition (5).
(c) If $\overline{t} \in [V_1]$ then instead of the element t consider the difference $t_1 = t - w$, where $w \in [V_1]$ and $\overline{w} = \overline{t}$, so that the leading term of t_1 is smaller than $\overline{t}$.
(d) If at any step the current difference equals zero, then the result concerning t is affirmative; if the process terminates with a nonzero element t_r, then the result is negative.

Remark 1. The above stated algorithm also applies of course to the case of nonassociative algebras considered in the work [5] by A.I. Zhukov. Therefore, the membership problem is decidable also for algebras without any identical relations.

Remark 2. In the same way as in [5], the finiteness problem is decidable for ε-algebras.

References

[1] A.A. Markov, *On the impossibility of certain algorithms in the theory of associative systems*, Doklady Akad. Nauk USSR 55 (1947), no. 7, 587–591.

[2] P.S. Novikov, *On the algorithmic unsolvability of the problem of identity*, Doklady Akad. Nauk USSR 85 (1952), no. 4, 709–712.

[3] E. Post, *Recursive unsolvability of a problem of Thue*, J. Symbolic Logic 12 (1947) 1–11.

[4] A.I. Shirshov, *Subalgebras of free commutative and free anticommutative algebras*, Mat. Sbornik 34 (1954), no. 1, 81–88.

[5] A.I. Zhukov, *Reduced systems of defining relations in nonassociative algebras*, Mat. Sbornik 27 (1950), no. 2, 267–280.

Some Algorithmic Problems for Lie Algebras

A.I. Shirshov

1. Introduction

In a previous work [2] the author considered some algorithmic problems in the theory of ε-algebras. The same paper mentioned some literature relevant to these problems.

In the present paper, we consider the analogous problems for Lie algebras. Unfortunately, we cannot obtain the solution of the word problem in this case. However, the word problem can be solved for Lie algebras with one defining relation, and for Lie algebras with a homogeneous system of defining relations.

Moreover, for Lie algebras we will prove a freeness theorem analogous to the corresponding theorem in group theory.

2. Definition of composition

Let L be the free Lie algebra over a field P with the set $R = \{a_\alpha\}$, $\alpha \in I$, of free generators. For brevity of exposition, in what follows we will use the definitions and results of the author's work [1] without particular explanation.

Having fixed once and for all an ordering on the set R, we define regular associative and regular nonassociative words formed by the elements of this set. In the work [1], it is shown that the regular nonassociative words form a basis of L. In what follows, unless otherwise indicated, when we speak of some element of L, we will mean its representation as a linear combination of the elements of this basis. The regular associative word that corresponds to the leading term of an element $b \in L$ (without coefficient) will be denoted by $\overline{b}$.

Sibirsk Mat. Zh. 3, (1962), no. 2, 292–296.

We choose in L two arbitrary elements b and c such that $\overline{b} = b_1b_2$ and $\overline{c} = c_1c_2$ with $b_2 = c_1$, where b_1, b_2, c_2 are (nonempty associative) words and the coefficients of the leading terms of the elements b and c equal 1.

Lemma 1. *The associative word $u = b_1b_2c_2 = b_1c_1c_2$ is regular.*

Proof. Suppose $u = w_1w_2$ and w_1 is a subword of $\overline{b}$. Then $\overline{b} = w_1v$, $\overline{b} > v$, and hence $w_1w_2 > w_2w_1$. In the case when w_2 is a subword of c_2, i.e., $c_2 = c_2'w_2$, the inequality $w_1w_2 > w_2w_1$ follows from the obvious inequalities $w_2 < c_2 < \overline{c} < u$. The proof is complete. □

According to Lemma 4 of [1], we form nonassociative words u_1 and u_2 by placing parentheses in the word u in two different ways[1]:

$$u_1 = \{\cdots[(\widetilde{\overline{b}}q_1)q_2]\cdots\}q_s,$$

where the q_i are regular nonassociative words and $\overline{q}_1\overline{q}_2\cdots\overline{q}_s = c_2$, with $q_1 \le q_2 \le \cdots \le q_s$; and

$$u_2 = r_1\{r_2\cdots[r_{t-1}(r_t\widetilde{\overline{c}})]\cdots\},$$

where the r_j are regular nonassociative words and $\overline{r}_1\overline{r}_2\cdots\overline{r}_t = b_1$. Let

$$u_1' = \{\cdots[(bq_1)q_2]\cdots\}q_s, \qquad u_2' = r_1\{r_2\cdots[r_{t-1}(r_tc)]\cdots\}.$$

Definition 1. The element $t = \alpha(u_1' - u_2')$, where $\alpha \in P$ is the inverse of the coefficient of the leading term of $u_1' - u_2'$, will be called *the composition $(b, c)_{c_1}$ of the elements b and c relative to the word c_1.*

Therefore, the notion of composition is defined for some but not all pairs b and c of elements of L, and essentially depends on the word c_1.

Lemma 2. *No composition can be formed for the pair (b, b).*

Proof. It suffices to show that there cannot be two representations $\overline{b} = b_1b_2 = b_2b_3$ where b_2 is a nonempty associative word. Suppose that $\overline{b} = b_1b_2 = b_2b_3$. From the definition of regularity it follows that $\overline{b} > b_3b_2$, i.e., $b_1 > b_3$; on the other hand, $\overline{b} > b_2b_1$, i.e., $b_3 > b_1$: an obvious contradiction. □

Note that if the composition $(b, c)_{c_1}$ is defined for some word c_1, then the composition $(c, b)_{b_1}$ of the elements c and b cannot be formed, since the assumption of the existence of the composition $(b, c)_{c_1}$ implies the inequality $\overline{b} > \overline{c}$.

[1]We have added tildes over $\overline{b}$ and $\overline{c}$ in the following equations for u_1 and u_2; the tilde means the regular nonassociative word corresponding to a given regular associative word. See the proof of Lemma 3. [Translators]

3. Some word problems

We consider some definitions necessary for what follows.

Definition 2. A finite set $S = \{s_i\}$, $i = 1, 2, \ldots, k$, of elements of the algebra L is called *reduced* if none of the associative words $\overline{s}_i$ is a subword of another word $\overline{s}_j$ $(s_i, s_j \in S)$ and the coefficients of the leading terms of the elements equal 1.

Let S be a reduced set of elements of L, and let S^* be the set of the leading terms of the elements of S and the elements obtained from S by all possible compositions (repeated any number of times).

Definition 3. A reduced set S of elements of L will be called *stable* if

(i) the degree of the composition $(s', s'')_c$ of two elements s' and s'', belonging to S or obtained from S by any number of compositions, is greater than the degree of each of the elements s' and s'', and
(ii) no element of S^* contains another element of S^* as a subword (in particular, the elements of S^* are distinct).

Theorem 1. *Let S be a stable set of elements of L. Then there exists an algorithm that allows us to determine, in a finite number of steps, whether or not an arbitrary element $t \in L$ belongs to the ideal $\langle S \rangle$ generated by S in L.*

We will obtain Theorem 1 from the following lemma.

Lemma 3. *An element $t \in L$ belongs to the ideal $\langle S \rangle$ generated in L by the elements of a stable set S, only if the word $\overline{t}$ contains one of the words of S^* as a subword.*

Proof. Suppose $t \in \langle S \rangle$. Then t can be written as a linear combination of elements d_i of the form

$$d_i = c_1 c_2 \cdots c_{k_i} s_{p_i} f_1 f_2 \cdots f_{\ell_i},$$

with some placement of parentheses, where $s_i \in S$ and c_j, f_j are regular words. Since for any regular associative words u and v, the greater of the words uv and vu is regular, we may assume without loss of generality that the following word is regular:

$$\overline{d}_i = \overline{c}_1 \overline{c}_2 \cdots \overline{c}_{k_i} \overline{s}_{p_i} \overline{f}_1 \overline{f}_2 \cdots \overline{f}_{\ell_i}.$$

The claim of the lemma is obvious if the word $\overline{d}_1$, which is the greatest of the words $\overline{d}_i$ of highest degree, does not occur among the other words[2] $\overline{d}_j$, $j \neq 1$, corresponding to the element t.

Now suppose that $\overline{d}_1 = \overline{d}_j$, $j \neq 1$. Consider the first and simplest case in which $\overline{s}_{p_j}$ is a subword of one of the words $\overline{c}_1 \overline{c}_2 \cdots \overline{c}_{k_1}$ or $\overline{f}_1 \overline{f}_2 \cdots \overline{f}_{\ell_1}$. Consider the former case (the latter is analogous). From the regularity of $\overline{d}_j$, $\overline{s}_{p_1}$, $\overline{s}_{p_j}$ it follows

[2]We have added a bar here, and twice in the first sentence of the next paragraph. [Translators]

(by Lemma 4 of [1]) that we can place parentheses in the word $\overline{d}_j$ in the following way[3]:

$$d' =$$

$$c_1c_2\cdots c_q[\cdots((\widetilde{\overline{s}}_{p_j}c'_{q+1})c'_{q+2})\cdots c'_r]\cdots c_{k_1}[\cdots((\widetilde{\overline{s}}_{p_1}f'_1)f'_2)\cdots f'_m]f_{m+1}\cdots f_{\ell_1},$$

where c'_ρ and f'_ν are regular words,

$$c'_{q+1}\le c'_{q+2}\le\cdots\le c'_r,\quad\text{and}\quad f'_1\le f'_2\le\cdots\le f'_m,$$

and the remaining parentheses are placed in the same way as in $\widetilde{\overline{d}}_j$, where the symbol $\widetilde{\ }$ means the regular nonassociative word corresponding to a given regular associative word. Furthermore, let d'_1 and d'_j denote the elements of L obtained from d' by replacing $\widetilde{\overline{s}}_{p_1}$ by s_{p_1} and $\widetilde{\overline{s}}_{p_j}$ by s_{p_j} respectively.

The differences $d_1-d'_1$ and $d_j-d'_j$ can obviously be written as linear combinations of elements similar to the elements d_i but having smaller leading terms than[4] d_1. As in the proof of Lemma 2 of [2], one can show that the difference $d'_j-d'_1$ can be written in an analogous way. From this, by virtue of the equation

$$d_j=d_1-(d_1-d'_1)+(d'_j-d'_1)+(d_j-d'_j),$$

it follows that the element d_j can be replaced by the sum of d_1 and some other similar elements with smaller leading terms. Combining like terms will either decrease the number of occurrences of the leading term or produce an expression with a smaller leading term. The induction is obvious.

One more case is possible: $\overline{s}_{p_1}=e_1e_2$, $\overline{s}_{p_j}=e_2e_3$. Then by Lemma 1, the subword $e_1e_2e_3$ of $\overline{d}_1$ is regular, and on $e=e_1e_2e_3$ parentheses can be placed in two ways as described in the definition of composition; we can then extend each of these placements of parentheses in a unique way to a complete placement of parentheses on $\overline{d}_1$. Let δ be the difference of the elements d''_1 and d''_j obtained from

[3]We have omitted the primes on $c_1,\ldots,c_q$. [Translators]

[4]Let us simplify and put $d'_1=d'_j=(c\,s_{p_j}c's_{p_1}f)$ where c,c',f are some associative words and $(\ldots)$ is the same placement of parentheses as in Shirshov's paper. (We shorten Shirshov's notation, and instead of two expressions d'_1,d'_j we use only one). Then, for example, $d_1-d'_1$ has the shorter form $(c\,\overline{s_{p_j}}c's_{p_1}f)-(c\,s_{p_j}c's_{p_1}f)$ where c,c',f are the same associative words, and the maximal associative words of each expression $\overline{d_1}$ and $\overline{d'_1}$ are equal to $\overline{d_1}$. Then we can rewrite d_1 as an associative expression $c\,\overline{s_{p_j}}c's_{p_1}f$ with maximal word $\overline{d_1}$ plus a linear combination of associative expressions $a_is_{p_1}b_i$ with maximal words less than $\overline{d_1}$. We can do the same with d'_1. The result is

$$D=d_1-d'_1=\sum_{1\le j\le k}\alpha_ja_js_{p_1}b_j,$$

with maximal words less than $\overline{d_1}$. Without loss of generality, we can assume $a_1\overline{s_{p_1}}b_1>a_2\overline{s_{p_1}}b_2>\ldots$, since the maximal word of s_{p_1} is a regular word, and any regular word has the property that its prefix cannot coincide with its suffix. Then $\overline{D}=a_1\overline{s_{p_1}}b_1$. By Lemma 4 of [1], one can place parentheses to obtain $(a_1s_{p_1}b_1)$ with the maximal word equal to $\overline{D}$. Then $D-\alpha_1(a_1s_{p_1}b_1)$ has the same form as D, but its maximal word is less than $\overline{D}$. The result now follows by induction on the maximal word. [Editors]

those described above by replacing the words $\widetilde{\overline{s}}_{p_1}, \widetilde{\overline{s}}_{p_j}$ by s_{p_1}, s_{p_j} respectively; then δ can be obtained from the word $\overline{d}_1$ by replacing the word e by the composition $(s_{p_1}, s_{p_j})_{e_2}$ and subsequently placing parentheses as on the words d_1'' and d_j''. As in the previous case, the proof is completed by considering the equation

$$d_j = d_1 - (d_1 - d_1'') + (d_j - d_j'') - \delta.$$

The lemma is proved. □

To prove Theorem 1 it suffices to verify that one can write down in a finite number of steps all the elements of the set S^* whose degrees do not exceed the degree of the element t. If the word $\overline{t}$ occurs in an element of S^* as a subword, then in the ideal $\langle S \rangle$ there can be found an element t_0 such that $\overline{t}_0 = \overline{t}$. Then instead of the element t, one should consider the difference $t - t_0$.

The theorem is proved.

Corollary 1. *There exists an algorithm that solves the word problem for Lie algebras with one defining relation.*

This follows from the obvious stability of a set that consists of one element.

Corollary 2. *There are no Lie algebras with one defining relation that have a finite dimension* ≥ 3.

This statement follows from the fact that in a Lie algebra with defining relation $s = 0$, all the distinct words v_i, such that $\overline{v}_i$ does not contain $\overline{s}$ as a subword, are linearly independent.

Theorem 2. *There exists an algorithm that solves the word problem for Lie algebras with a homogeneous set of defining relations.*

Proof. Suppose that in the algebra L some homogeneous set S has been selected. If S is not reduced, then it can be replaced by a reduced set S_1 such that $\langle S \rangle = \langle S_1 \rangle$. Indeed, if $\overline{s}_i$ ($s_i \in S$) is a subword of $\overline{s}_j$ ($s_j \in S$), then one constructs an element s_0 of the ideal $\langle s_i \rangle$ such that $\overline{s}_0 = \overline{s}_j$, and considers the element $s_j' = s_j - s_0$ instead of the element s_j.

The proof that this process of reduction will terminate in a finite number of steps coincides with the proof of Lemma 1 in [2]. Obviously, the resulting set S_1 will consist of homogeneous elements. Since the composition of homogeneous elements is homogeneous, the requirement on degrees in the definition of stability is satisfied. It is also obvious that after a finite number of steps one can write down all elements of S^* whose degrees do not exceed the degree of a given element $t \in L$; during this procedure it may be necessary to perform the reduction process on the sets obtained from S_1 by adjoining compositions of certain elements. The proof is completed as in Theorem 1. □

Theorem 3. (Freeness Theorem) *Let L_0 be a Lie algebra with a set R of generators and one defining relation $s = 0$ whose left side contains the generator a_α. Then the subalgebra L_0', generated in L_0 by the set $R \setminus \{a_\alpha\}$, is free.*

Proof. In addition to the natural ordering of the regular words that form a basis of the free Lie algebra L, we will consider the following ordering. A regular word u is considered to be greater than a regular word v if the generator a_α occurs in u more times than in v. If a_α occurs in u and v the same number of times, then these words are first compared by degree, and if the degrees are equal, then by the usual lexicographical comparison of the words $\overline{u}$ and $\overline{v}$. The associative word $\overline{\overline{s}}$ that corresponds to the leading term of an element s in the sense of the new ordering, may be different from the word $\overline{s}$. Repeating the arguments used in the proof of Lemma 3, and applying Lemma 2, we obtain the result that an element t belongs to the ideal $\langle s \rangle$ only if[5] the word $\overline{\overline{s}}$ is a subword of $\overline{\overline{t}}$. Since the generator a_α occurs in the expression of s, it follows that the subalgebra L_0' has zero intersection with the ideal $\langle s \rangle$. This is equivalent to the claim of the theorem. □

References

[1] A.I. Shirshov, *On free Lie rings*, Mat. Sbornik 45 (1958), no. 2, 113–122.

[2] A.I. Shirshov, *Some algorithmic problems on ε-algebras*, Sibirsk. Mat. Zh. 3, (1961), no. 1, 132–137.

[5] In the rest of this sentence, we have added double bars over s and t. [Translators]

On a Hypothesis in the Theory of Lie Algebras

A.I. Shirshov

1. Introduction

The concepts of a free group and the free product of groups, as well as the results related to these concepts, have their analogues in the theory of algebras. It is known, for example, that any subalgebra of a free Lie algebra is also free. This result is analogous to the well-known theorem of Nielsen-Schreier in group theory. The results of A.T. Gainov [1] on subalgebras of the free commutative and free anticommutative products of algebras, are analogous to the theorem of A.G. Kurosh [2] on subgroups of the free product of groups. Under the influence of this analogy, there existed a conjecture that subalgebras of the free Lie product of Lie algebras are described by a theorem analogous to the theorem of A.T. Gainov cited above. In the present note, we prove that this is not the case. Moreover, we give here a construction of interest in its own right, which it is natural to call the free Lie product of Lie algebras with an amalgamated subalgebra.

2. The free Lie product of Lie algebras with an amalgamated subalgebra

Let L_α $(\alpha \in I)$ be a family of Lie algebras over some fixed field P, each of which contains a subalgebra $L_{\alpha,0}$ which is isomorphic to a given algebra L_0. We construct a Lie algebra L with the following properties:

(1) L contains subalgebras L'_α which are isomorphic to the algebras L_α $(\alpha \in I)$ respectively;
(2) the intersection $L'_0 = \bigcap L'_\alpha$ of the algebras L'_α is a subalgebra isomorphic to L_0, and some fixed isomorphism of L_0 with L'_0 can be extended to isomorphisms of L_α with L'_α for all α;
(3) L is generated by the subalgebras L'_α $(\alpha \in I)$.

Sibirsk Mat. Zh. 3, (1962), no. 2, 297–301.

We choose an arbitrary basis of L_0, and for each $\alpha \in I$ we extend its isomorphic image in L_α to a basis of this latter algebra. As a result of this, we obtain a set S of elements of the algebras L_α, namely $S = \{e_{\alpha\gamma}\}$ ($\alpha \in I$, $\gamma \in J_\alpha$), where all the index sets J_α contain subsets J'_α of equal cardinality (which we will identify in what follows: $J'_\alpha = J'$) such that $\gamma \in J'_\alpha$ implies $e_{\alpha\gamma} \in L_{\alpha,0}$. Clearly, the symbols $e_{\alpha\gamma}$ and $e_{\beta\gamma}$ will not be distinguished if $\gamma \in J'$.

We take a set $R = \{f_{\alpha\gamma}\}$ ($\alpha \in I$, $\gamma \in J_\alpha$) in one-to-one correspondence with S, and make it into the set of free generators of the free Lie algebra $\overline{L}$. We choose a basis of $\overline{L}$ formed by regular words (see [3]), starting from some ordering of the sets I and J_α, where the ordering of J_α extends some ordering of J', and the conditions $\gamma \in J'$, $\delta \in J_\alpha$, $\delta \notin J'_\alpha$ imply that $\gamma < \delta$. That is, $f_{\alpha\gamma} < f_{\alpha'\gamma'}$ if either $\alpha < \alpha'$, or $\alpha = \alpha'$, $\gamma < \gamma'$. Consider the ideal Q of $\overline{L}$ generated by all elements of the form

$$q_{\alpha\gamma\delta} = f_{\alpha\gamma} f_{\alpha\delta} - \sum_\tau p^\tau_{\alpha\gamma\delta} f_{\alpha\tau} \quad (\gamma > \delta),$$

where the following equation holds in the algebra L_α:

$$e_{\alpha\gamma} e_{\alpha\delta} = \sum_\tau p^\tau_{\alpha\gamma\delta} e_{\alpha\tau} \quad (p^\tau_{\alpha\gamma\delta} \in P).$$

Definition. A basis word v of the algebra $\overline{L}$ will be called *special* if the corresponding regular associative word does not contain subwords of the form $f_{\alpha\beta} f_{\alpha\beta'}$, $\beta > \beta'$.

Clearly, a special word of length ≥ 2 can contain none of the symbols $f_{\alpha\beta}$ when $\beta \in J'$.

In what follows, by the *leading term* of an element $t \in \overline{L}$ we will mean the lexicographically maximal term among the terms of the highest degree.

Lemma 1. *An element $t \in \overline{L}$ belongs to the ideal Q only if its leading term is not special.*

Proof. Suppose that an element t of the algebra $\overline{L}$ belongs to the ideal Q, i.e., t can be represented as a linear combination of products of elements $q_{\alpha\gamma\delta}$ with elements of R. Obviously, the leading term of each of the elements $q_{\alpha\gamma\delta}$ corresponds to a regular associative word that contains a subword of the form $f_{\alpha\beta} f_{\alpha\beta'}$, $\beta > \beta'$.

If the greatest of these leading terms does not occur among the other leading terms, then the claim is proved. Otherwise, some of the leading terms are equal, and in view of the complete analogy with the proof of Lemma 3 of [4], it suffices to consider only the case in which the equal regular associative words, corresponding to the equal leading terms, have the form $c_1 c_2 \cdots c_s f_{\alpha\beta} f_{\alpha\gamma} f_{\alpha\delta} d_1 d_2 \cdots d_r$, and the products themselves have the form

$$v_1 = c_1 c_2 \cdots c_s \Big(f_{\alpha\beta} f_{\alpha\gamma} - \sum_\tau p^\tau_{\alpha\beta\gamma} f_{\alpha\tau} \Big) f_{\alpha\delta}\, d_1 d_2 \cdots d_r,$$

$$v_2 = c_1 c_2 \cdots c_s f_{\alpha\beta} \Big(f_{\alpha\gamma} f_{\alpha\delta} - \sum_\tau p^\tau_{\alpha\gamma\delta} f_{\alpha\tau} \Big) d_1 d_2 \cdots d_r,$$

where $\beta > \gamma > \delta$ and the parentheses on the products v_1 and v_2 are placed in the same way. By virtue of the known properties of the structure constants of a Lie algebra, the following equation holds:

$$v_1 = c_1 c_2 \cdots c_s \Big(f_{\alpha\beta} f_{\alpha\delta} - \sum_\tau p^\tau_{\alpha\beta\delta} f_{\alpha\tau} \Big) f_{\alpha\gamma} d_1 d_2 \cdots d_r$$
$$+ c_1 c_2 \cdots c_s f_{\alpha\beta} \Big(f_{\alpha\gamma} f_{\alpha\delta} - \sum_\tau p^\tau_{\alpha\gamma\delta} f_{\alpha\tau} \Big) d_1 d_2 \cdots d_r + w,$$

where the omitted parentheses are placed as on the element v_1, and the element w is a linear combination of the elements[1] of the form $q_{\alpha\gamma\delta}$, but of lower degree.

Having performed the corresponding substitution in the expression for the element t, and combined like terms (the second summand in the expression for v_1 coincides with v_2), we either decrease the number of the above-mentioned products with equal leading terms, or reduce the leading term itself. The proof is completed by induction on the leading term[2]. □

Now consider the quotient algebra $L = \overline{L}/Q$. Lemma 1 implies that the images of special words are linearly independent in L.

Theorem 1. *The images of the special words form a basis of the algebra L.*

Proof. By the remark preceding the statement of the theorem, it suffices to prove that the images of regular words can be represented as linear combinations of special words. A regular word is not special if it contains either

(1) a subword of the form $f_{\alpha\beta} f_{\alpha\gamma}$, $\beta > \gamma$, or
(2) a subword of the form $f_{\alpha\beta}(f_{\alpha\gamma} w)$, $\beta > \gamma$, where w is a regular word, or
(3) a subword of the form $f_{\alpha\beta}(uv)$ where the regular associative word that corresponds to u starts with $f_{\alpha\gamma}$, $\beta > \gamma$.

In the first case, since

$$f_{\alpha\beta} f_{\alpha\gamma} - \sum_\tau p^\tau_{\alpha\beta\gamma} f_{\alpha\tau} \equiv 0 \pmod{Q},$$

the word can be replaced by a linear combination of words of lower degree. In the second case, it follows from the equation

$$f_{\alpha\beta}(f_{\alpha\gamma} w) = (f_{\alpha\beta} f_{\alpha\gamma}) w + f_{\alpha\gamma}(f_{\alpha\beta} w),$$

that the given word can be replaced by a linear combination of words either of lower degree or smaller in the lexicographical sense. The argument in the third case is analogous. The rest is obvious. □

It is also obvious that the algebra L satisfies the required conditions stated at the beginning of this section. Furthermore, it is clear that L can be homomorphically mapped onto any Lie algebra satisfying the same list of conditions, and that the kernel of this homomorphism will have zero intersection with each of the

[1] Multiplied by the c_i and d_j. [Translators]
[2] And on the number of products with equal leading terms. [Translators]

algebras L'_α. From the latter remark it easily follows that L is uniquely determined up to isomorphism. Therefore, L does not depend on the choice of whichever ordering of the sets I and J_α is used in the construction of L. By analogy with the well-known definition in group theory, we will call L the *free Lie product of the Lie algebras L_α with an amalgamated subalgebra L_0*.

3. On subalgebras of free Lie products of Lie algebras

A special case of the construction considered above is the free Lie product of Lie algebras L_α, $\alpha \in I$, obtained with the assumption that $L_0 = 0$, i.e., J' is the empty set. As in the general case, the free Lie product is commutative, associative, and has many properties usually associated with free compositions. We point out only one of them.

Lemma 2. *Let L be the free Lie product of the Lie algebras L_α, $\alpha \in I$. Then the quotient $\overline{L}$ of L by the ideal S_β, generated in L by one of the factors L_β, is isomorphic to the free Lie product of the L_α, $\alpha \in \overline{I}$, $\overline{I} = I \setminus \{\beta\}$.*

The proof of this statement follows immediately from the fact that an element $c \in L$ belongs to the ideal S_β if and only if each basis element occurring in the expression of c contains at least one of the generators $f_{\beta\mu}$.

Theorem 2. *There exist Lie algebras such that their free Lie product has a subalgebra that is not free, is not isomorphic to any subalgebra of any of the factors, and cannot be decomposed as the free Lie product of any of its subalgebras.*

Proof. Let L_1 be a 1-dimensional Lie algebra with generator e_{11}, and let L_2 be the 2-dimensional Lie algebra with basis e_{21}, e_{22} such that $e_{22}e_{21} = e_{21}$. Let L be the free Lie product of L_1 and L_2. It follows from the above discussion that for a basis of L we can choose the collection of special words, i.e., regular nonassociative words whose corresponding associative words do not contain the subword $e_{22}e_{21}$. Note that here we assume the following ordering: $e_{22} > e_{21} > e_{11}$.

Consider the subalgebra L' of L generated by the elements e_{21}, e_{22}, $e_{21}e_{11}$, $e_{22}e_{11}$. First we show that L' is isomorphic to the Lie algebra L^* with four generators c_1, c_2, c_3, c_4 and two defining relations:

$$c_4c_1 + c_3c_2 - c_1 = 0, \qquad c_4c_2 - c_2 = 0.$$

We establish the following correspondence among the generators:

$$c_4 \to e_{22}, \qquad c_3 \to e_{22}e_{11}, \qquad c_2 \to e_{21}, \qquad c_1 \to e_{21}e_{11}.$$

Since the following relations hold in the algebra,

$$e_{22}(e_{21}e_{11}) + (e_{22}e_{11})e_{21} - e_{21}e_{11} = 0, \qquad e_{22}e_{21} - e_{21} = 0,$$

the correspondence above extends to a homomorphism of algebras from L^* onto L'. The algebra L^* is one of the algebras for which the word problem is decidable (see [4]).

As basis elements of L^* we can take the regular nonassociative words whose corresponding associative words do not contain the subwords c_4c_1 or c_4c_2. However, it is easy to see that every nonzero linear combination of such elements corresponds to a nonzero element of L'. Therefore, the above-mentioned homomorphism is an isomorphism.

In what follows, we will work with the algebra L^*. We will assume that L^* does not contain subalgebras of finite dimension, except for those of dimension 1 and the subalgebra generated by c_4 and c_2, since such a subalgebra would give the required example; for the same reason, we will assume that the free Lie product of Lie algebras which do not have finite-dimensional subalgebras except for those of dimension 1, does not contain such subalgebras either. From this it follows that, if L^* were decomposed as the free Lie product of algebras $\overline{L}_1$ and $\overline{L}_2$, then one of them, say $\overline{L}_1$, would contain the elements c_4 and c_2.

The ideal T generated by these elements contains the element c_1, and hence the quotient algebra L^*/T is 1-dimensional; however by Lemma 2 it is isomorphic to $\overline{L}_2$. The algebra L^* is not isomorphic to L, since it does not contain an element that together with c_4 and c_2 would generate L^*. Therefore, the algebra $\overline{L}_1$ is not generated by c_4 and c_2. It cannot be decomposed into the free Lie product of two Lie algebras, since otherwise it would follow from Lemma 2 that the quotient of L^* by the ideal T would not be 1-dimensional. Therefore, as the required example, we can take the subalgebra $\overline{L}_1$ of L. The theorem is proved. □

Remark. In fact, even the algebra L' cannot be decomposed as a free Lie product of its subalgebras. However, the proof of this fact is considerably more complicated than the proof given above.

The theorem just proved shows that subalgebras of the free Lie product of Lie algebras have a rather complicated structure, and the problem of their description is of great interest.

References

[1] A.T. Gainov, *Free commutative and free anti-commutative products of algebras*, Sibirsk. Mat. Zh. 3, (1962), no. 6, 805–833.

[2] A.G. Kurosh, *Die Untergruppen der freien Produkte von beliebigen Gruppen*, Math. Ann. 109, 1 (1934) 647–660.

[3] A.I. Shirshov, *On free Lie rings*, Mat. Sbornik 45, (1958), no. 2, 113–122.

[4] A.I. Shirshov, *Some algorithmic problems for Lie algebras*, Sibirsk Mat. Zh. 3, (1962), no. 2, 292–296.

On the Bases of a Free Lie Algebra

A.I. Shirshov

Introduction

In the work of M. Hall [1], a certain way of fixing a basis of a free Lie algebra is indicated. However, the concrete bases which one needs to construct, in order to solve certain problems, do not always fall into Hall's scheme. For instance, the basis of a free Lie algebra considered in the work [2] cannot be constructed using Hall's method. For this reason, in each such case it is necessary to reprove that a certain subset of a free Lie algebra is a basis. Below, we give a method that generalizes Hall's method for choosing a basis in a free Lie algebra.

A construction of a basis of a free Lie algebra

Let $R = \{a_\alpha\}$ be a set of symbols, where α ranges over a nonempty set of indices. The set of all nonassociative words that can be formed from the elements of R will be denoted by K.

Definition 1. Nonassociative words of length 1 in K will be called *regular* and ordered arbitrarily. Suppose regular words for all lengths less than n have already been defined and ordered by some relation $>$ such that for any regular words u, v and w the condition $w = uv$ implies $w > v$. Then a word t of length n, $n > 1$, will be called *regular* if

1) $t = rs$ where r and s are regular words with $r > s$, and
2) if $r = r_1 r_2$ then $r_2 \leq s$.

The regular words of length $\leq n$ defined in this way will be ordered arbitrarily, except that we preserve the existing ordering of the regular words of length less than n, and require as before that $w = uv$ implies $w > v$.

Algebra Logika 1, (1962), no. 1, 14–19.

The ordering described in this definition can be realized, for instance, by ordering words of the same length arbitrarily and declaring that words of smaller length precede words of greater length (see Hall [1]). This case, however, does not exhaust all the possibilities.

We indicate a method that assigns to every element w of K the unique formal expression,

$$w^* = \sum_{i=1}^{k(w)} n_i^{(w)} w_i,$$

where $k(w) \geq 0$, the $n_i^{(w)}$ are nonzero elements of the base field, and the w_i are distinct regular words. If the length of the word w equals 1, then we set $w^* = w$. Assume by induction that for words w whose length is less than n the following conditions hold:

i) the required method has been indicated,
ii) the words w_i obtained by this method have the same content relative to R as w (i.e., in each of these words every element of R occurs the same number of times as in w),
iii) if w is the product of two *distinct* regular words, $w = uv$, then all the w_i, $i = 1, 2, \ldots, k(w)$, are greater than the lesser of u and w (in the sense of the ordering of regular words), and
iv) if w is regular then $w^* = w$, i.e., $k(w) = 1$ and $n_1^{(w)} = 1$.

Suppose now that a word w has length $n > 1$. Then $w = uv$. By the inductive hypothesis it follows that the expressions,

$$u^* = \sum_{i=1}^{k(u)} n_i^{(u)} u_i, \qquad v^* = \sum_{j=1}^{k(v)} n_j^{(v)} v_j,$$

have already been defined. Let

$$w' = \sum_{i=1}^{k(u)} \sum_{j=1}^{k(v)} n_i^{(u)} n_j^{(v)} u_i v_j.$$

We delete in the expression w' the terms $n_j^{(v)} n_i^{(u)} u_i v_j$ in which $u_i = v_j$, and replace each term in which $u_i < v_j$ by the expression $-n_i^{(u)} n_j^{(v)} v_j u_i$.

After performing the formal combination of like terms, and removing the terms whose coefficients turn out to be zero, we denote the resulting expression by w''. In the expression w'', if it turns out that for some $mu_i v_j$ we have $u_i = u_i' u_i''$ with $u_i'' > v_j$, then we replace each such expression by $m(u_i' v_j) u_i'' + m u_i' (u_i'' v_j)$; we act analogously for the elements $m v_j u_i$ if it turns out that $v_j = v_j' v_j''$ with $v_j'' > u_i$. After this, we combine like terms and denote the resulting expression by w'''. For each monomial that occurs in w''' we perform all the transformations that have been done for w, each time replacing the monomial in w''' by the corresponding formal expression, multiplied by the original coefficient of the monomial; we then

combine like terms. Then for the resulting expression $w^{(4)}$ we perform the same transformations as were done for w''', and so forth. We will show that at some step this process must stabilize.

Indeed, by the inductive hypothesis for the element $u_i v_j$, with each passage to the sum of products of regular words, the lesser of the factors will be greater than the lesser of the words u_i and v_j. However, also by the inductive hypothesis, the content of the words will not change, but the number of words with the same content is finite. This stabilized expression obtained for w we will take as w^*.

Therefore, we have indicated a method which, to each word w of length n, assigns in a unique manner the formal expression w^*. Since at each step the content of the words is preserved, all the w_j will have the same content as w. If $w = uv$ is a product of two regular words, then all the w_i will be greater than the lesser of the words u and v, since at no step can a decrease of the lesser factor occur, and hence $w_i = u_i v_i$ where v_i is not less than the lesser of the words u and v, but $w_i > v_i$ by Definition 1. In the case that w is regular, it cannot undergo any changes, and hence $w^* = w$. Therefore, all the assumptions of the inductive hypothesis have been verified for words of length n.

Now consider the vector space $\mathfrak{A}$ over the base field with the basis of regular words. We make this space into an algebra $\mathfrak{S}$ by defining the product of basis elements as follows:

$$v_i \cdot v_j = (v_i v_j)^*.$$

Theorem. *The algebra $\mathfrak{S}$ is the free Lie algebra with the set of generators R.*

Proof. First we prove that $\mathfrak{S}$ is a Lie algebra. The construction for w^* explained above shows that

$$\Big(\sum_i \delta_i a_i + \sum_j \delta'_j u_j\Big) \cdot \Big(\sum_i \delta_i a_i + \sum_j \delta'_j u_j\Big) = 0,$$

which implies that the identical relation $x^2 = 0$ holds in $\mathfrak{S}$.

It is more difficult to prove that the Jacobi identity holds. By virtue of its multilinearity, it suffices to show that if u_i, u_j, u_k are regular words, then

$$[(u_i u_j)^* u_k]^* + [(u_j u_k)^* u_i]^* + [(u_k u_i)^* u_j]^* = 0. \tag{1}$$

If the sum of the lengths of u_i, u_j, u_k equals 3, then the validity of equation (1) follows from the definition of the operation w^*. Assume by induction that for any set R, and any way of defining regular R-words, i.e., words formed from the symbols of the set R, equation (1) holds if the sum of the lengths of u_i, u_j, u_k is less than n. Suppose now that u_i, u_j, u_k are such that the sum of their lengths equals n, $n > 3$. Let a_β be the lowest symbol (in terms of the ordering) of the set R, from among the symbols that occur in u_r $(r = i, j, k)$.

First, we assume that $u_r \neq a_\beta$ $(r = i, j, k)$. Consider the set of symbols $R' = \{a_\alpha^n\}$, $n = 0, 1, 2, \ldots$, where $a_\alpha \in R$ and $a_\alpha > a_\beta$. To each R'-word $\overline{w}$ we

assign an R-word w by replacing in $\overline{w}$ each symbol a_α^n by the monomial

$$[\cdots(a_\alpha \underbrace{a_\beta)a_\beta\cdots]a_\beta}_{n \text{ times}}.$$

We will say that the word $\overline{w}$ is *regular* if the corresponding word w is regular, and we will order regular R'-words using the already defined ordering of the corresponding regular R-words. Then R'-words of length 1 are regular, and R'-words of length n, $n > 1$, are regular if and only if they satisfy the conditions of Definition 1. Therefore, the definition of regular R'-words agrees with Definition 1. In our words u_r $(r = i, j, k)$ there occur symbols from R that are not less than a_β, and by Definition 1 the symbol a_β can occur only in words of the form $[\cdots(a_\alpha a_\beta)a_\beta\cdots]a_\beta$; hence we can find R'-words $\overline{u}_r$ $(r = i, j, k)$ which correspond, in the sense explained above, to the R-words u_r $(r = i, j, k)$. Since a_β occurs in at least one of the words u_r $(r = i, j, k)$, the sum of the lengths of the R'-words $\overline{u}_r$ $(r = i, j, k)$ is less than n. Hence by the inductive hypothesis it follows that

$$[(\overline{u}_i\overline{u}_j)^*\overline{u}_k]^* + [(\overline{u}_j\overline{u}_k)^*\overline{u}_i]^* + [(\overline{u}_k\overline{u}_i)^*\overline{u}_j]^* = 0.$$

But each transformation for $(\overline{u}_i\overline{u}_j)^*$ etc. corresponds to an analogous transformation for $(u_iu_j)^*$ etc., and as a result of performing these transformations, we obtain elements which correspond as explained above. Hence equation (1) holds in this case.

Now assume that a_β equals one of our words, for instance u_k. Then equation (1) takes the form

$$[(u_iu_j)^*a_\beta]^* + [(u_ja_\beta)^*u_i]^* + [(a_\beta u_i)^*u_j]^* = 0. \tag{2}$$

We also assume that $u_i \neq a_\beta$, $u_j \neq a_\beta$ and $u_i \neq a_j$, since otherwise equation (2) is obvious. Without loss of generality, we may assume that $u_i > u_j$, since otherwise, using the equation $(uv)^* = -(vu)^*$, we can reduce equation (2) to the equation

$$[(u_ju_i)^*a_\beta]^* + [(u_ia_\beta)^*u_j]^* + [(a_\beta u_j)^*u_i]^* = 0.$$

If it turns out that the following equation holds,

$$(u_iu_j)^* = u_iu_j, \tag{3}$$

then

$$\begin{aligned}&[(u_iu_j)^*a_\beta]^* + [(u_ja_\beta)^*u_i]^* + [(a_\beta u_i)^*u_j]^* = \\ &[(u_ia_\beta)^*u_j]^* + [u_i(u_ja_\beta)^*]^* + [(u_ja_\beta)^*u_i]^* + [(a_\beta u_i)^*u_j]^* = 0,\end{aligned}$$

by the definition of the operation w^*, and hence equation (2) holds. Equation (3) is valid if the length of u_i equals 1, or if u_i has the form $u_{i1}u_{i2}$ where $u_{i2} \leq u_j$. Hence we can exclude these cases and assume that $u_i = u_{i1}u_{i2}$ where $u_{i2} > u_j$. Consider the finite set of pairs u_i', u_j' of regular words which can be formed from the symbols that occur in u_i, u_j and for which $u_i' \geq u_j'$. Let u_j'' be the maximal

value taken on by u'_j. Then $(u''_i u''_j)^* = u''_i u''_j$ where u''_i is some value of the word u'_i that corresponds to the word u''_j. Indeed, if $u''_i = u''_{i1} u''_{i2}$ with $u''_{i2} > u''_j$, then

$$(w'')^* = (u''_i u''_j)^* = [(u''_{i1} u''_j)^* u''_{i2}]^* + [u''_{i1} (u''_{i2} u''_j)^*]^*.$$

During the application of the operation $*$, the content of the words does not change, and the lowest factor will not decrease, but each of the regular words in the expressions $(u''_{i1} u''_j)^*$, $(u''_{i2} u''_j)^*$, u''_{i2}, u''_{i1} is greater than u''_j. Hence, u''_j is not maximal. Therefore, equation (2) holds for the words u''_i, u''_j, a_β. We carry out induction on the finite number of values of the word u''_j in the pairs of words defined above.

Assume by induction that equation (2) holds for all possible values of the pairs u'_i, u'_j with $u'_j > u_j$. From the definition of the operation $*$ it follows that

$$\{[(u_{i1} u_{i2}) u_j]^* a_\beta\}^* = \{[(u_{i1} u_j)^* u_{i2}]^* a_\beta\}^* + \{[u_{i1} (u_{i2} u_j)^*]^* a_\beta\}^*. \tag{4}$$

Since each of the words u_{i1}, u_{i2}, u_j is distinct from a_β, from the case proved above the following equations hold:

$$[(u_j a_\beta)^* (u_{i1} u_{i2})]^* = \{[(u_j a_\beta)^* u_{i1}]^* u_{i2}\}^* + \{u_{i1} [(u_j a_\beta)^* u_{i2}]^*\}^*, \tag{5}$$

$$\{[(a_\beta u_{i1})^* u_{i2}]^* u_j\}^* = [(a_\beta u_{i1})^* (u_{i2} u_j)^*]^* + \{[(a_\beta u_{i1})^* u_j]^* u_{i2}\}^*, \tag{6}$$

$$\{[u_{i1} (a_\beta u_{i2})^*]^* u_j\}^* = [(u_{i1} u_j)^* (a_\beta u_{i2})^*]^* + \{u_{i1} [(a_\beta u_{i2})^* u_j]^*\}^*. \tag{7}$$

Finally, from the inductive hypothesis it follows that

$$\{[a_\beta (u_{i1} u_{i2})]^* u_j\}^* = \{[(a_\beta u_{i1})^* u_{i2}]^* u_j\}^* + \{[u_{i2} (a_\beta u_{i2})^*]^* u_j\}^*, \tag{8}$$

$$\{u_{i1} [(u_j a_\beta)^* u_{i2}]^*\}^* + \{u_{i1} [(u_{i2} u_j)^* a_\beta]^*\}^* + \{u_{i1} [(a_\beta u_{i2})^* u_j]^*\}^* = 0, \tag{9}$$

$$\{[(u_{i1} u_j)^* u_{i2}]^* a_\beta\}^* = \{[(u_{i1} u_j)^* a_\beta]^* u_{i2}\}^* + [(u_{i1} u_j)^* (u_{i2} a_\beta)^*]^*, \tag{10}$$

$$\{[u_{i1} (u_{i2} u_j)^*]^* a_\beta\}^* = [(u_{i1} a_\beta)^* (u_{i2} u_j)^*]^* + \{u_{i1} [(u_{i2} u_j)^* a_\beta]^*\}^*, \tag{11}$$

$$\{[(u_j a_\beta)^* u_{i1}]^* u_{i2}\}^* + \{[(a_\beta u_{i1})^* u_j]^* u_{i2}\}^* + \{[(u_{i1} u_j)^* a_\beta]^* u_{i2}\}^* = 0. \tag{12}$$

Adding separately the left and right sides of equations (4)–(12) with the corresponding sides of the obvious equations,

$$[(u_{i1} u_j)^* (u_{i2} a_\beta)^*]^* = -[(u_{i1} u_j)^* (a_\beta u_{i2})^*]^*, \tag{13}$$

$$[(u_{i1} a_\beta)^* (u_{i2} u_j)^*]^* = -[(a_\beta u_{i2})^* (u_{i2} u_j)^*]^*, \tag{14}$$

and comparing the results, we obtain

$$\{[(u_{i1} u_{i2}) u_j]^* a_\beta\}^* + [(u_j a_\beta)^* (u_{i1} u_{i2})]^* + \{[a_\beta (u_{i1} u_{i2})]^* u_j\}^* = 0,$$

which completes the proof that $\mathfrak{S}$ is a Lie algebra.

Now let S be any Lie algebra with R as the set of generators. To each element $h = \sum_i \delta_i a_i + \sum_j \delta'_j u_j$ of the algebra $\mathfrak{S}$ we assign the analogously written element $\overline{h}$ of S where δ_i, δ'_j are elements of the base field. Since the transformations that carry the word w to the element w^* can be performed in any Lie algebra, it follows that the above-mentioned correspondence is a homomorphism of $\mathfrak{S}$ onto S. □

Instead of coefficients from the base field one can consider integers. In this case, we will obtain a free Lie ring $\mathfrak{S}$. From what has been proved it follows that regular words form a basis of the free Lie ring, which is a generalization of the well-known theorem of Hall [1], since Definition 1 is broader than the corresponding definition given by Hall.

Obviously, this also implies a group-theoretic statement generalizing the corresponding result of Hall. To be specific, we introduce the notation $[x, y] = xyx^{-1}y^{-1}$ in the free group G with R as a set of free generators, and call a commutator product (i.e., an R-word with some placement of square brackets) *regular* if it is regular in the sense of Definition 1. Then from the well-known isomorphism of the group G^n/G^{n+1} with the subgroup generated by words of length n in the additive group of the free Lie ring with generating set R, it follows that *regular commutator products of length n form a basis of the Abelian group* G^n/G^{n+1}.

References

[1] M. Hall, *A basis for free Lie rings and higher commutators in free groups*, Proc. Amer. Math. Soc. 1 (1950) 575–581.

[2] A.I. Shirshov, *On free Lie rings*, Mat. Sbornik 45 (1958), no. 2, 113–122.

On Some Groups which are Nearly Engel

A.I. Shirshov

1. Introduction

In the present work, we give a certain modification of one of the possible definitions of an Engel group. As a consequence of this, we define a class of groups which in the finite case turns out to be wider than the class of Engel groups. For the finite case, we obtain a complete description of groups of this wider class (Section 3). In Section 4 we define a subclass of Engel groups that contains in particular the 3-Engel groups.

In this work we formulate several problems that can, in the author's opinion, attract the attention of mathematicians.

The author expresses his gratitude to M.I. Kargapolov, who looked over the manuscript and made a number of important remarks.

2. The definition of ν-group

Let G be a group. We introduce the following notation:

$$[a,b,1] = [a,b] = aba^{-1}b^{-1}; \qquad [a,b,k] = [\,[a,b,k-1],b\,] \ \ (k = 2,3,\ldots).$$

A group G in which, for any two elements a and b and some fixed number k, the equality $[a,b,k] = e$ holds, where e is the identity element of G, is called k-*Engel* or simply *Engel*. Obviously, any nilpotent group is Engel. However, it is not known up to now if there exist Engel groups which are not locally nilpotent. Local nilpotence has been proved only for 3-Engel groups [2].

The definition of Engel groups can be formulated in a slightly different way. Suppose that a variety M_1 of groups is determined by the equation

$$f_1(x,y) = \varphi_1(x,y),$$

Algebra Logika 2, (1963), no. 5, 5–18.

where f_1 and φ_1 are words in the variables x and y. Then setting

$$f_2(x,y) = f_1(xyx^{-1}, y), \qquad \varphi_2(x,y) = \varphi_1(xyx^{-1}, y),$$

we define a new variety M_2 by the equation

$$f_2(x,y) = \varphi(x,y).$$

Clearly $M_1 \subseteq M_2$. Analogously, we define the varieties M_3, M_4, and so on. Applying the above process to the variety E_1 of Abelian groups, i.e., to the relation $xy = yx$, we obtain the variety E_2 determined by the relation $xyx^{-1}y = yxyx^{-1}$, or equivalently by the relation $[x,y,2] = e$. Since $[xyx^{-1}, y, k] = [x,y,k+1]$, the variety E_k coincides with the variety of k-Engel groups.

The passage from M_1 to M_2 consists in replacing arbitrary elements x and y by a pair of conjugates xyx^{-1} and y. Taking into account that a pair of conjugate elements can always be written in the form xy and yx, we can define in a similar way another process of passing from a variety $M^{(1)}$ to another variety $M^{(2)}$. Suppose that the variety $M^{(1)}$ is determined by the relation

$$f^{(1)}(x,y) = \varphi^{(1)}(x,y).$$

Setting

$$f^{(2)}(x,y) = f^{(1)}(xy, yx), \qquad \varphi^{(2)}(x,y) = \varphi^{(1)}(xy, yx),$$

we define the variety $M^{(2)}$ by the relation

$$f^{(2)}(x,y) = \varphi^{(2)}(x,y).$$

Analogously, we define the varieties $M_{(k)}$, $k = 3, 4, \ldots$. Applying this process to the variety $E_1 = N^{(1)}$ of Abelian groups, i.e., again to the relation $xy = yx$, we obtain the variety $N^{(2)}$ determined by the relation $xy^2x = yx^2y$, the variety $N^{(3)}$ determined by the relation $xy^2xyx^2y = yx^2yxy^2x$, and so on.

Definition 1. The groups that belong to the variety $N^{(k)}$ will be called *ν_k-groups* or simply *ν-groups*.

The first question that arises in connection with the study of ν-groups is the question of the relation of this class with the class of Engel groups. Obviously, the varieties E_2 and $N^{(2)}$ coincide, since each of them is determined by the commutativity of any two conjugate elements. As the following theorem shows, starting with $k = 3$, the varieties E_k and $N^{(k)}$ no longer coincide.

Theorem 1. *A group G is 3-Engel if and only if it satisfies the following identical relations:*

$$xy^2xyx^2y = yx^2yxy^2x, \tag{1}$$

$$xy^2xyxyx^2y = yx^2yyxxy^2x. \tag{2}$$

The relations (1) *and* (2) *are independent.*

Proof. Let G be a 3-Engel group. Then by virtue of the easily and immediately verifiable relation,

$$\begin{aligned} &xy^2xyx^2yx^{-1}y^{-2}x^{-1}y^{-1}x^{-2}y^{-1} \\ &= [x, yx, 3]yx^2yxy^2x[x^{-1}, y^{-1}x^{-1}, 3]x^{-1}y^{-2}x^{-1}y^{-1}x^{-2}y^{-1}, \end{aligned}$$

it is obvious that G satisfies relation (1). In addition, since in the group G we have

$$e = [x, yx, 3] = xy^2xy^{-1}x^{-1}yx^2yx^{-1}y^{-1} \cdot y^{-1}x^{-2}y^{-1},$$

it follows that

$$\begin{aligned} e &= y^{-1}x^{-2}y^{-1} \cdot xy^2xy^{-1}x^{-1}yx^2yx^{-1}y^{-1} \\ &= y^{-1}x^{-2}y^{-1}xy^2xy^{-1}x^{-2}y^{-1}yxyx^2yx^{-1}y^{-1}. \end{aligned}$$

Using the already established relation (1), we obtain

$$e = y^{-1}x^{-2}y^{-1} \cdot y^{-1}x^{-2}y^{-1}xy^2x \cdot yx^2y \cdot x^{-1}y^{-1},$$

or equivalently $xy^2x \cdot yx \cdot yx^2y = yx^2y \cdot yx \cdot xy^2x$, i.e., relation (2).

If we now assume that G satisfies relations (1) and (2), then doing the just performed transformations in the reverse order, we obtain $[x, yx, 3] = e$, or equivalently $[x, y, 3] = e$.

To prove the independence of relations (1) and (2), it suffices to give examples of groups in which one of the relations holds but not the other.

It is easy to verify that the symmetric group S_3 of degree 3 satisfies relation (1). It is well known that S_3 is not Engel, and hence does not satisfy relation (2). Another group with the same properties is, for example, the free product of two groups of order 2.

Denote by Z_3 the collection of all pairs of the form (ε_i, t) where t is an arbitrary complex number, and ε_i is one of the three cube roots of unity. We define multiplication of elements of Z_3 by the formula $(\varepsilon_i, t_1)(\varepsilon_j, t_2) = (\varepsilon_i\varepsilon_j, t_1\varepsilon_j + t_2)$. It is easy to verify that the set Z_3, with the above operation, is a group that is isomorphic to the group of all rotations of the complex plane, by angles that are multiples of $2\pi/3$ around various points, and all translations. A direct computation shows that the identity (2) holds in Z_3. On the other hand, setting $\alpha = (\varepsilon_i, 1)$, $\beta = (1, 1)$, $\varepsilon_i \neq 1$, we convince ourselves that $\alpha\beta^2\alpha\beta\alpha^2\beta \neq \beta\alpha^2\beta\alpha\beta^2\alpha$. We note that the group Z_3 is a solvable group with Abelian commutator subgroup. □

Remark 1. The theorem just proved indicates the possibility of defining 3-Engel groups by relations that make sense for semigroups. Relations (1) and (2) can therefore be taken as the definition of a 3-Engel semigroup. From the results of the works [2] and [4] it follows that a 3-Engel semigroup with cancelation is locally nilpotent.

It would be interesting to find semigroup relations (if they exist) that define k-Engel groups for any k. The following two questions are also of interest:

1) Do there exist Engel groups that are not ν-groups?
2) Do there exist ν-groups that are not locally solvable?

Negative answers to both of these questions would give an affirmative solution to the problem of local nilpotence of Engel groups.

3. Finite ν-groups

In this section, we consider finite ν-groups in a little more detail. The example of the group S_3 shows that there exist finite ν-groups that are not nilpotent. On the other hand, the example of the alternating group A_4 shows that there exist finite solvable groups which are not ν-groups.

In Section 1, the ν_k-groups were defined by the equation

$$f^{(k)}(x,y) = \varphi^{(k)}(x,y), \text{ where } f^{(1)}(x,y) = xy,\ \varphi^{(1)}(x,y) = yx.$$

By induction we show that

$$f^{(k)}(x,y) = f^{(k-1)}(x,y)\,\varphi^{(k-1)}(x,y); \quad \varphi^{(k)}(x,y) = \varphi^{(k-1)}(x,y)\,f^{(k-1)}(x,y).$$

Indeed,

$$\begin{aligned} f^{(k)}(x,y) &= f^{(k-1)}(xy,yx) = f^{(k-2)}(xy,yx)\,\varphi^{(k-2)}(xy,yx) \\ &= f^{(k-1)}(x,y)\,\varphi^{(k-1)}(x,y), \end{aligned}$$

and the second equation is proved similarly.

Consider the set S of pairs (a,b) of elements of a group G. On the set S we define the mapping φ that sends each pair (a,b) to the pair (ab,ba); we write $(a,b)^{\varphi} = (ab,ba)$. The pairs of the form (a,a) will be called *trivial*. Obviously, G is a ν-group if and only if some power of the mapping φ sends every pair to a trivial pair. The group A_4 mentioned above is not a ν-group because

$$((1,2,3),(1,3,4))^{\varphi^2} = ((1,2,3),(1,3,4)).$$

Obviously, no power of the mapping φ can send $((1,2,3),(1,3,4))$ to a trivial pair.

It follows from the work of A.I. Malcev [4] that any nilpotent group G is a ν-group. A wider class of ν-groups is described by the following theorem.

Theorem 2. *A group G which is an extension of a nilpotent group, by a nilpotent group with an identical relation of the form $x^{2^k} = e$, is a ν-group.*

Proof. By assumption, G has a nilpotent normal subgroup N, such that the quotient group $\overline{G} \simeq G/N$ is a nilpotent group with the identical relation $x^{2^k} = e$. Therefore, any pair (a,b) of elements of G is sent by a power of φ to a pair of the form (cn,cm), $n,m \in N$, $c^{2^k} \in N$. Since

$$(cn,cm)^{\varphi} = (c^2 \cdot c^{-1}ncm, c^2 \cdot c^{-1}mcn) = (c^2 n_1, c^2 m_1), \quad n_1, m \in N,$$

it follows that φ^k sends the pair (cn,cm) to a pair of the form $(\overline{n},\overline{m})$, $\overline{n},\overline{m} \in N$ which, by nilpotence of the group N, will be sent by some power of φ to a trivial pair. The theorem is proved. □

The description of finite ν-groups of odd order is achieved by the following theorem.

Theorem 3. *A finite ν-group of odd order is nilpotent.*

For the proof we will need some auxiliary results.

Lemma 1. *Let s and t be natural numbers with $(s,t)=1$. Then the matrix $A_{(s,t)}$ of the form*

$$A_{(s,t)} = \left\| \begin{array}{ccccccccccc} 1 & 0 & 0 & \cdots & 0 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & -1 \\ -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 0 & 0 & 0 & \cdots & 1 \\ \underbrace{\qquad\qquad\qquad}_{s} & & & & & \underbrace{\qquad\qquad\qquad}_{t} & & & & \end{array} \right\|$$

$$= E_{s+t} - \left\| \begin{array}{cc} 0_{t\times s} & E_t \\ E_s & 0_{s\times t} \end{array} \right\|,$$

has rank $s+t-1$. (Here E_q and $0_{r\times p}$ are the identity and zero matrices of the indicated sizes.)

Proof. Without loss of generality, we may assume that $s > t$ since otherwise we could transpose the matrix $A_{(s,t)}$. We now add the first row of $A_{(s,t)}$ to the $(t+1)$-st row, the second row to the $(t+2)$-nd row, and so on, finally adding row t to row $2t$. As a result of these operations we obtain the new matrix $A^{(1)}_{(s,t)}$ of the form

$$A^{(1)}_{(s,t)} = \left\| \begin{array}{c|c|c} E_t & 0_{t\times(s-t)} & -E_t \\ \hline 0_{s\times t} & \multicolumn{2}{c}{A_{(s-t,t)}} \end{array} \right\|.$$

Let $r(C)$ denote the rank of a matrix C. Clearly,

$$r(A_{(s,t)}) = t + r(A_{(s-t,t)}).$$

Since $(s-t,t)=1$, the proof is completed by induction:

$$r(A_{(s,t)}) = t + (s-t) + t - 1 = t + s - 1.$$

The lemma is proved. □

Lemma 2. *Let G be a ν-group, let p and q be distinct odd primes, and let $a \in G$ be an element such that a^q belongs to the centralizer of an element b of order p lying in an Abelian normal subgroup N of G. Then a belongs to the centralizer of b.*

Proof. Let $b_s = a^{-s}ba^s$. Obviously,

$$(ab,a)^\varphi = (a^2b_1, a^2b_0), \quad (a^2b_1, a^2b_0)^\varphi = (a^4b_3b_0, a^4b_2b_1),$$

and so on. The indices of the elements b are reduced modulo q, and the powers of the elements b_r are reduced modulo p. A pair

$$C = \big(a^{2^i} b_0^{k_0} b_1^{k_1} \cdots b_{q-1}^{k_{q-1}},\ a^{2^i} b_0^{\ell_0} b_1^{\ell_1} \cdots b_{q-1}^{\ell_{q-1}} \big),$$

is sent by φ to the pair

$$C^\varphi = \big(a^{2^{i+1}} b_{2^i}^{k_0} b_{2^i+1}^{k_1} \cdots b_{2^i+q-1}^{k_{q-1}} b_0^{\ell_0} b_1^{\ell_1} \cdots b_{q-1}^{\ell_{q-1}},\ a^{2^{i+1}} b_{2^i}^{\ell_0} b_{2^i+1}^{\ell_1} \cdots b_{2^i+q-1}^{\ell_{q-1}} b_0^{k_0} b_1^{k_1} \cdots b_{q-1}^{k_{q-1}} \big).$$

Consider the q-dimensional vector space Q over the field of congruence classes modulo p. If $a = (\alpha_0, \alpha_1, \ldots, \alpha_{q-1}) \in Q$ then we set $a_{(1)} = (\alpha_{q-1}, \alpha_0, \alpha_1, \ldots, \alpha_{q-2})$, and by induction $a_{(t)} = (a_{(t-1)})_{(1)}$. To the pair C we assign the number 2^i and the vector $c \in Q$ of the form

$$c = (k_0 - \ell_0, k_1 - \ell_1, \ldots, k_{q-1} - \ell_{q-1}), \qquad C \mapsto (2^i, c).$$

Obviously, in this way, to the pair C^φ will be assigned the pair $(2^{i+1}, C_{(2^i)} - C)$:

$$C^\varphi \mapsto (2^{i+1}, C_{(2^i)} - C).$$

If we start transforming consecutively (with φ) the pair $(a^2 b_1, a^2 b_0)$, then all the vectors in Q that correspond to the resulting pairs will belong to the subspace Q' which consists of vectors for which the sum of the coordinates equals zero. Since the number of vectors in Q' is finite, and the numbers 2^i can be reduced modulo q, it follows that there will be repetitions in the indicated sequence of vectors for which the corresponding numbers are congruent modulo q.

On the other hand, if we wish to recover C from the known vector $C^{(2^i)} - C$, then we obtain a system of q linear equations with q unknowns, such that the coefficient matrix has the form $A_{(s,t)}$, $(s,t) = 1$, which by Lemma 1 is a matrix of rank $q-1$. Its columns, as well as the column of constant terms, are vectors in the subspace Q'. Clearly, this system of equations will have a unique solution in Q'. The above-mentioned sequence of vectors will therefore be periodic, and the vector $(-1, 1, 0, \ldots, 0)$ which begins the sequence will reoccur as far from the beginning as we wish. However, at a sufficient distance from the beginning of the sequence, all occurring pairs will be trivial, and hence the pair $(a^s b_1, a^s b_0)$ corresponding to our vector will also be trivial. Therefore $b_1 = b_0$ and $ab = ba$. The lemma is proved. □

Proof. (of Theorem 3) Assume that the claim of the theorem is valid for groups that have order less than that of G. Then all subgroups of G are supersolvable, and therefore the group G is solvable [3]; hence one of the commutator subgroups $G^{(s)}$ is an Abelian normal subgroup.

Let N_p be the primary component of $G^{(s)}$ relative to a prime number p, and let d be one of the elements of order p in N_p. The element d lies in a normal subgroup $\widetilde{N}_p \subseteq G$ all of whose elements have order p. By the inductive hypothesis, the quotient group $G/\widetilde{N}_p$ is nilpotent, and hence is the direct product of its Sylow subgroups. The subgroup $Q_p \subseteq G$, that corresponds to the Sylow p-subgroup of

$G/\widetilde{N}_p$, is the unique Sylow p-subgroup, and thus is a normal subgroup of G. Its center Z_p is also a normal subgroup of G. Let b be an element of order p in Z_p, and let a be an element of order m with $(m,p)=1$. If q is one of the prime factors of m, then by Lemma 2 the element $a^{m/q}$ lies in the centralizer of b. Considering the prime factorization of m/q, and repeating the argument as many times as required, we arrive at the statement that a lies in the centralizer of b.

Therefore, the element b is in the center Z of the group G. The quotient group G/Z of G by the (non-trivial) center Z is nilpotent by the inductive hypothesis. Therefore G is also nilpotent. Theorem 3 is proved. □

Below we will give a complete description of finite ν-groups. To this end, we first prove two more lemmas.

Lemma 3. *If a ν-group G has an Abelian normal subgroup N which is a 2-group of odd index, then N is a direct factor of G.*

Proof. Assume that the statement of the lemma is valid for groups whose order is smaller than that of G.

Let $a \in N$, $a^2 = e$, $b \in G$. Since $(aba, b)^\varphi = (a \cdot baba \cdot a, baba)$, it is clear that the mapping φ replaces b by $baba$. We form the sequence

$$b_1 = b, \quad b_2 = b_1 a b_1 a, \quad \ldots, \quad b_i = b_{i-1} a b_{i-1} a, \quad \ldots.$$

Since G is a ν-group, for some n we have $ab_n a = b_n$, i.e., $(ab_{n-1})^2 = (b_{n-1}a)^2$. On the other hand, for some q we have $(ab_{n-1})^{2q+1} \in N$. Hence

$$(ab_{n-1})^{2q+1} a = a(ab_{n-1})^{2q+1}.$$

But $(ab_{n-1})^{2q+1}a = a(b_{n-1}a)^{2q+1}$. Therefore $(ab_{n-1})^{2q+1} = (b_{n-1}a)^{2q+1}$, and hence $ab_{n-1} = b_{n-1}a$. From this we see that the equation $ab_n a = b_n$ implies $ab_{n-1}a = b_{n-1}$. It immediately follows that $ab = ba$. Therefore, the elements of order 2 of the group N form a subgroup Z of the center of G.

By the inductive hypothesis, the quotient $\overline{G} \simeq G/Z$ decomposes as a direct product $\overline{G} = \overline{P} \times \overline{N}$ where $\overline{P}$ is a subgroup of odd order, and the intersection of the corresponding subgroup $P \subseteq G$ with N is Z. By the inductive hypothesis, $P = P_1 \times Z$. Since the group P_1 coincides with the set of all elements of odd order in P, it follows that P_1 is a normal subgroup of G, and the elements of P_1 commute with the elements of N. Therefore, $G = P_1 \times N$. The lemma is proved. □

Lemma 4. *If a Sylow 2-subgroup N of a finite ν-group G is normal, then it is a direct factor.*

Proof. The center Z of N is a normal subgroup of G. Performing induction on the index of nilpotence of N, we consider the quotient group $\overline{G} \simeq G/Z$. Now the argument is completely identical to that concluding the proof of Lemma 3. □

Theorem 4. *The extensions of nilpotent groups of odd order by 2-groups are the only ν-groups among finite groups.*

Proof. According to Theorem 2, it suffices to prove that any finite ν-group is an extension of a nilpotent group of odd order by a 2-group.

We choose a Sylow 2-subgroup Q in the ν-group G, an arbitrary subgroup $S \subseteq Q$, and an element a of odd order in the normalizer of S. Then in the group $T = \langle a, S\rangle$ generated by a and S, the group S will be a normal Sylow 2-subgroup, and hence a direct factor by Lemma 4. Therefore, the element a is in the centralizer of the group T. Now appealing to the well-known result on the existence of p-complements [1, Theorem 14.4.7] we conclude that G has a normal subgroup of odd order for which the corresponding quotient group is a 2-group. The theorem is proved. □

Theorem 4 gives a complete description of finite ν-groups.

4. One subclass of ν-groups

With each element c of a semigroup G, we can associate a mapping φ_c from the set S of pairs of elements of G to itself that sends the pair (a, b) to the pair (acb, bca):

$$(a, b)^{\varphi_c} = (acb, bca).$$

If the semigroup G has an identity element e, then it is clear that the map φ_e coincides with the mapping φ considered earlier. All possible mappings φ_c generate a semigroup $[G]$, of self-mappings of the set S, which we will call *adjoint* to G.

In the theory of semigroups, two completely different concepts of nilpotence are used, brought to the theory of semigroups on the one hand from ring theory and on the other hand from group theory. We need to distinguish them.

Definition 2. A semigroup G with zero is called *r-nilpotent* if there exists a natural number n such that $a_1 a_2 \cdots a_n = 0$ for all elements a_i of G.

A.I. Malcev [4] gave the following definition in a slightly different form.

Definition 3. A semigroup G is called *g-nilpotent* if the adjoint semigroup $[G]$ is r-nilpotent. (The role of zero in $[G]$ is played by the mapping which sends any pair in S to a trivial pair.)

A.I. Malcev showed in the same work that if G is a group, then the concept of g-nilpotence coincides with the usual concept of nilpotence for groups.

Definition 4. A pair (a, b) in S will be called *m-central* and written $(a, b)_m$ if every element of the semigroup $[G]^m$ sends it to a trivial pair.

Lemma 5. *In a group G with generators c_1, c_2, …, c_k, the condition $(a, b)_n$ is equivalent to the conjunction of the conditions $(ab, ba)_{n-1}$ and $(ac_i b, bc_i a)_{n-1}$ for $i = 1, 2, \ldots, k$.*

Proof. By definition, we declare that $(a, b)_0$ means $a = b$. Obviously, the condition $(a, b)_n$ implies the indicated $k + 1$ conditions.

For the proof of the converse, we first take $n = 1$. Then we have $ab = ba$ and $ac_ib = bc_ia$ for $i = 1, 2, \ldots, k$. The following equations are obvious: $c_iba^{-1} = a^{-1}bc_i$ and $ba^{-1} = a^{-1}b$. Clearly, the element $a^{-1}b$ lies in the center. Therefore, for any d we have $adb = bda$, i.e., $(a, b)_1$. Now suppose that the lemma has been proved for all natural numbers less than n, and that the following conditions hold:

$$(ab, ba)_{n-1}, \quad (ac_ib, bc_ia)_{n-1}, \quad i = 1, 2, \ldots, k.$$

Obviously, the center Z of G is non-trivial. Denoting by $\overline{d}$ the image of an element $d \in G$ in the quotient group $\overline{G} \simeq G/Z$, we have

$$(\overline{a}\overline{b}, \overline{b}\overline{a})_{n-2}, \quad (\overline{a}\,\overline{c}_i\,\overline{b}, \overline{b}\,\overline{c}_i\,\overline{a})_{n-2}, \quad i = 1, 2, \ldots, k.$$

From the inductive hypothesis it follows that $(\overline{a}, \overline{b})_{n-1}$, i.e., any element from $[\overline{G}]^{n-1}$ sends the pair $(\overline{a}, \overline{b})$ to a pair of the form $(\overline{d}, \overline{d})$. This means that any element of $[G]^{n-1}$ sends the pair (a, b) to a pair of the form (d, dz), $z \in Z$, and hence any element of $[G]^n$ sends the pair (a, b) to a trivial pair, i.e., we have $(a, b)_n$. The lemma is proved. □

The lemma easily implies the following interesting property of 3-Engel groups.

Theorem 5. *Any two conjugate elements of a* 3*-Engel group generate a* 2*-Engel group.*

Proof. In the 3-Engel group G we choose any two conjugate elements a and b, which can always be written in the form $a = cd$ and $b = dc$ with $c, d \in G$. From equations (1) and (2) it follows that

$$ab^2a = ba^2b, \quad ab^3a = babab, \quad ba^3b = ababa.$$

From Lemma 5 it follows that the condition $(ab, ba)_1$ holds in the group G_1 generated by a and b. Now we prove the condition $(a, b)_2$ for which it suffices to verify the conditions $(a^2b, ba^2)_1$ and $(ab^2, b^2a)_1$. By the obvious symmetry we will prove only the condition $(a^2b, ba^2)_1$, which is equivalent to the equations:

$$a^2b^2a^2 = ba^4b, \quad a^2baba^2 = ba^5b, \quad a^2b^3a^2 = ba^2ba^2b.$$

However,

$$a^2b^2a^2 = aba^2ba = ba^4b,\ a^2baba^2 = aba^3ba = ba^5b,\ a^2b^3a^2 = abababa = ba^2ba^2b,$$

where we have used in every case the condition $(ab, ba)_1$, i.e., $abqba = baqab$ for all $q \in G_1$. The theorem is proved. □

The result of Theorem 5 suggests the following definition.

Definition 5. Abelian groups will be called σ_1*-groups*. Any group in which any two conjugate elements generate a σ_{i-1}-group will be called a σ_i*-group*. Finally, σ_i-groups, as i ranges over all natural numbers, will be called σ*-groups*.

Remark 2. We can also speak of σ-semigroups if by conjugate elements we understand elements of the form xy and yx.

Remark 3. Obviously, σ-groups lie in the intersection of the classes of ν-groups and Engel groups. It is not known to the author whether there exist Engel groups that are not σ-groups. Local nilpotence of σ_k-groups for $k > 3$ is also unclear.

Definition 6. A group G will be called *weakly nilpotent of bounded index* if there exists a natural number k such that any subgroup of G generated by two elements is nilpotent with nilpotence index less than or equal to k.

Theorem 6. *Every weakly nilpotent group of bounded index is a σ-group.*

Proof. Let G be a weakly nilpotent group of weak nilpotence index not exceeding k. We show that any two conjugate elements in G generate a nilpotent subgroup of nilpotence index not exceeding $k-1$. In what follows, by the symbol $(c_1, c_2, \ldots, c_s)$ we will understand the simple commutator in the sense of M. Hall's book [1]. Suppose elements x_1 and $x_2 = y^{-1}x_1y$ generate a subgroup Q in G. In any simple commutator $t = (x_{i1}, x_{i2}, \ldots, x_{ik})$ where i_s takes values 1 or 2, we have either $x_{i1} = x_{i2}$ and hence $t = e$, or the commutator (x_{i1}, x_{i2}) can be written as a triple commutator. Indeed,

$$(x_1, x_2) = (x_1, y^{-1}x_1y) = (y^{-1}x_1y, y^{-1}, y^{-1}x_1y),$$
$$(x_2, x_1) = (y^{-1}x_1y, x_1) = (x_1, y_1, x_1).$$

Therefore any commutator t of the indicated form is equal to the identity in G. The proof that the nilpotence index of Q does not exceed $k-1$, and hence the proof of Theorem 6 (by an obvious induction), follow immediately from the next lemma. □

Lemma 6. *Let the group F be generated by a_1, a_2, …, a_m. Then any normal subgroup that contains all simple commutators of the form $\alpha_i = (a_{i1}, a_{i2}, \ldots, a_{it})$, $i_s = 1, 2, \ldots, m$, where t is an arbitrary natural number, contains F_t (term t in the lower central series).*

Proof. For $t = 1$ the statement is trivial. Suppose now that $\tau = (q_1, q_2, \ldots, q_{t-1}, q_t)$ is an arbitrary simple commutator, $q_i \in F$. By the inductive hypothesis, we conclude that the commutator $\tau' = (q_1, q_2, \ldots, q_{t-1})$ lies in the normal subgroup generated by the commutators of the form $A_i = (a_{i1}, a_{i2}, \ldots, a_{i_{t-1}})$, i.e., $\tau' = \prod_i \ell_i^{-1} A_i \ell_i$. Using the well-known formulas relating commutators,

$$(xy, z) = y^{-1}(x, z)y(y, z), \qquad (x, yz) = (x, z)z^{-1}(x, y)z,$$
$$(x^{-1}, y) = x(x, y)^{-1}x^{-1}, \qquad (x, y^{-1}) = y(x, y)^{-1}y^{-1},$$

we convince ourselves that the commutator τ lies in any normal subgroup containing all commutators of the form $(\ell_i^{-1}A_i\ell_i, a_i)$, and hence in the normal subgroup generated by the commutators $(A_i, \ell_i a_i \ell_i^{-1})$. Using the stated formulas one more time, we arrive at the conclusion that τ is contained in the normal subgroup generated by all commutators of the form α_i. If N is the normal subgroup generated in

F by all commutators of the form α_i, then the quotient group $\overline{F} \simeq F/N$ satisfies the following identical relation:

$$(x_1, x_2, \ldots, x_t) = e,$$

which is equivalent to the statement of the lemma. The lemma, and hence Theorem 6, are proved. □

It is very probable that the converse of Theorem 6 is also true.

References

[1] M. Hall, *The Theory of Groups*, Macmillan, New York, 1959.

[2] H. Heineken, *Engelsche Elemente der Länge drei*, Illinois J. Math. 5, 4 (1961) 681–707.

[3] B. Huppert, *Normalteiler und maximale Untergruppen endlicher Gruppen*, Math. Z. 60 (1954) 409–434.

[4] A.I. Malcev, *Nilpotent semigroups*, Ivanov. Gos. Ped. Inst. Uch. Zap. Fiz.-Mat. Nauki 4 (1953) 107–111.

On Some Identical Relations for Algebras

A.I. Shirshov

1. In the work of A.I. Malcev [2], results of a general nature are applied in particular to the classification of identical relations of degree 3 for associative algebras. It is shown there that, under natural assumptions on the characteristic, any such identical relation is a linear combination of the following relations:

$$\sum_{(i_1,i_2,i_3)} x_{i_1}x_{i_2}x_{i_3} = 0, \tag{1}$$

$$\sum_{(i_1,i_2,i_3)} (-1)^{\sigma_i} x_{i_1}x_{i_2}x_{i_3} = 0, \tag{2}$$

$$x_1x_2x_3 + x_2x_1x_3 - x_2x_3x_1 - x_3x_2x_1 = 0, \tag{3}$$

$$x_1x_2x_3 + x_1x_3x_2 - x_3x_1x_2 - x_3x_2x_1 = 0, \tag{4}$$

where the summations in relations (1) and (2) are performed over all substitutions, and σ_i is the number of inversions in the permutation (i_1, i_2, i_3) of 1, 2, 3.

In the present note, we study algebras that satisfy one of the relations (3) and (4) and show that such algebras, in a sense to be made precise later, are close to commutative. In conclusion, we give and study a generalization of this closeness to commutativity.

2. For brevity, throughout this note, algebras with relations (3) and (4) will be called μ-algebras and μ'-algebras respectively.

Theorem 1. *Let S be a μ-algebra over a field P of characteristic $\neq 2$. Then the ideal S^3 lies in the center Z of S, and the ideal S^2 is commutative.*

Proof. It is easy to see that when relation (3) holds identically, these relations follow:

$$x_2x_4x_3x_1 + x_4x_2x_3x_1 - x_4x_3x_1x_2 - x_3x_1x_4x_2 = 0, \tag{5}$$

$$-x_3x_4x_2x_1 - x_4x_3x_2x_1 + x_4x_2x_1x_3 + x_2x_1x_4x_3 = 0, \tag{6}$$

Sibirsk Mat. Zh. 7, (1966), no. 4, 963–966

$$-x_2x_1x_4x_3 - x_1x_2x_4x_3 + x_1x_4x_3x_2 + x_4x_3x_1x_2 = 0, \tag{7}$$

$$-x_3x_4x_1x_2 - x_4x_3x_1x_2 + x_4x_1x_2x_3 + x_1x_2x_4x_3 = 0, \tag{8}$$

$$x_3x_4x_1x_2 + x_4x_3x_1x_2 - x_4x_1x_3x_2 - x_1x_4x_3x_2 = 0, \tag{9}$$

$$-x_2x_4x_3x_1 - x_4x_2x_3x_1 + x_4x_3x_2x_1 + x_3x_4x_2x_1 = 0, \tag{10}$$

$$-x_4x_2x_1x_3 - x_4x_1x_2x_3 + x_4x_1x_3x_2 + x_4x_3x_1x_2 = 0. \tag{11}$$

Adding equations (5)–(11) we obtain

$$x_4x_3x_1x_2 - x_3x_1x_4x_2 = 0. \tag{12}$$

Furthermore, the relations

$$x_4x_2x_1x_3 - x_2x_1x_4x_3 = 0, \tag{13}$$

$$x_4x_1x_2x_3 - x_1x_2x_4x_3 = 0, \tag{14}$$

are corollaries of relation (12). If we subtract the left side of relation (7) from the sum of the left sides of relations (11), (13), (14) then we obtain

$$x_4x_1x_3x_2 - x_1x_4x_3x_2 = 0. \tag{15}$$

Relations (12) and (15) together are equivalent to a system of relations of the form

$$x_1x_2x_3x_4 - x_{i_1}x_{i_2}x_{i_3}x_4 = 0, \tag{16}$$

where (i_1, i_2, i_3) is an arbitrary permutation of 1, 2, 3. Using relation (16) we rewrite relation (11) in the form

$$2x_4x_1x_3x_2 - 2x_2x_4x_1x_3 = 0,$$

or equivalently,

$$x_2(x_4x_1x_3) = (x_4x_1x_3)x_2. \tag{17}$$

Finally, repeated application of the last relation gives

$$(x_2x_4)(x_1x_3) = (x_1x_3)(x_2x_4). \tag{18}$$

The last two relations constitute the statement of the theorem. □

Remark 1. It follows from the proof that the restriction on characteristic is only essential in the derivations of relations (17) and (18); relation (16) is valid without restriction.

Remark 2. From the fact that an algebra which is anti-isomorphic to a μ-algebra is a μ'-algebra, it follows that Theorem 1 holds also for μ'-algebras.

3. Consider the following properties of an algebra A:

α) some power A^k of A is a commutative algebra;
β) some power A^t of A lies in the center.

Lemma 1. *Properties α and β are equivalent.*

Proof. Obviously, β implies α. Now, if A satisfies property α, then it is obvious that

$$\begin{aligned}(x_1x_2\cdots x_k)(x_{k+1}\cdots x_{2k}x_{2k+1}) &= (x_{k+1}\cdots x_{2k})(x_{2k+1}x_1\cdots x_k)\\ &= x_{2k+1}x_1\cdots x_kx_{k+1}\cdots x_{2k}.\end{aligned}$$

In other words, A satisfies property β for $t = 2k$. □

Definition 1. An algebra A that satisfies α and β will be called a *KD-algebra.*

The statement of Theorem 1 can be strengthened using the following lemma.

Lemma 2. *If every algebra A satisfying a multilinear identical relation*

$$F(x_1, x_2, \ldots, x_k) = 0,$$

over a field of characteristic zero is a KD-algebra, then every algebra B over the same field satisfying an identical relation of the form

$$F(x_1^{t_1}, x_2^{t_2}, \ldots, x_k^{t_k}) = 0,$$

where the t_i are arbitrary natural numbers, is also a KD-algebra.

Proof. Let $t = \max(t_1, t_2, \ldots, t_k)$. Then it was shown by Higman [1] that there exists a natural number $f(t)$ such that any element of the ideal $B^{f(t)}$ can be written as a linear combination of the s-th powers of elements of B, where s is any natural number less than or equal to t. From this it follows that the algebra $B^{f(t)}$ satisfies the identical relation $F(x_1, x_2, \ldots, x_k) = 0$, and thus by assumption $B^{f(t)}$ is a KD-algebra. Hence for some number q the algebra $[B^{f(t)}]^q = B^{qf(t)}$ is commutative. □

Theorem 2. *Any algebra A with identical relation*

$$x_1^{t_1}x_2^{t_2}x_3^{t_3} + x_2^{t_2}x_1^{t_1}x_3^{t_3} - x_2^{t_2}x_3^{t_3}x_1^{t_1} - x_3^{t_3}x_2^{t_2}x_1^{t_1} = 0, \tag{19}$$

over a field of characteristic zero, is a KD-algebra.

Remark 3. The result of Higman used above allows us to point out that for algebras over a field of characteristic zero, any identity of the form

$$x^py^q - y^qx^p = 0, \tag{20}$$

is equivalent to the definition of a KD-algebra.

4. The study of KD-algebras is of interest, if only for the reason that they include commutative and nilpotent algebras. But there is yet another reason.

Definition 2. An associative algebra is called *locally Noetherian* if every increasing chain of right ideals of every finitely generated subalgebra stabilizes after a finite number of steps.

Theorem 3. *Every KD-algebra is locally Noetherian.*

Proof. Obviously, a finitely generated subalgebra S of a KD-algebra is itself a KD-algebra, and hence is an extension of the commutative finitely generated algebra S^m by the nilpotent finite-dimensional algebra S/S^m; in both of these algebras, every increasing chain of right ideals must terminate. Hence, for any increasing chain of right ideals $J_1 \subset J_2 \subset \cdots \subset J_n \subset \cdots$ there exists a number k such that $S^m \cap J_k = S^m \cap J_{k+r}$ and $\overline{J}_k = \overline{J}_{k+r}$, where $\overline{J}_p$ is the pre-image of the ideal J_p under the natural homomorphism of S onto S/S^m, and r is any natural number. This immediately implies that $J_k = J_{k+r}$. □

Corollary. *An algebra with identical relation* (19) *over a field of characteristic zero is locally Noetherian.*

5. It is well known that the sum of any finite number of nilpotent ideals is a nilpotent ideal. On the other hand, it is not difficult to construct an example of an algebra in which the sum of two commutative ideals is not a commutative ideal. For this reason, the following result is of interest.

Theorem 4. *The sum of any finite number of KD-ideals* (*i.e., ideals that are KD-algebras*) *of an algebra A is again a KD-ideal.*

Proof. It suffices to prove the claim for two ideals. Let J_1 and J_2 be KD-ideals of the algebra A, and let Z_1 and Z_2 be their respective centers. Then

$$J_1^{t_1} \subset Z_1, \qquad J_2^{t_2} \subset Z_2, \qquad (J_1 + J_2)^{t_1+t_2-1} \subset Z_1 + Z_2.$$

If z_1, z_1', $z_1'' \in Z_1$ and z_2, z_2', $z_2'' \in Z_2$ then

$$(z_1 + z_2)(z_1' + z_2')(z_1'' + z_2'') = (z_1'' + z_2'')(z_1 + z_2)(z_1' + z_2'),$$

which can be easily and immediately verified. The commutativity of the ideal

$$[(J_1 + J_2)^{t_1+t_2-1}]^2 = (J_1 + J_2)^{2t_1+2t_2-2},$$

follows from this. □

Remark 4. The theorem just proved could be used in an obvious way to construct KD-radicals analogous to radicals based on nilpotency.

References

[1] G. Higman, *On a conjecture of Nagata*, Proc. Cambridge Philos. Soc. 52, 1 (1956) 1–4.

[2] A.I. Malcev, *On algebras with identical defining relations*, Mat. Sbornik 26, (1950), no. 1, 19–33.

On Some Positively Definable Varieties of Groups

A.I. Shirshov

1. A variety $\mathfrak{N}$ of groups will be called *positively definable* if it can be defined by identical relations that do not include variables with negative powers.

For example, the variety of Abelian groups is obviously positively definable. A.I. Malcev [2] proved positive definability of the varieties of nilpotent groups (with a given index of nilpotence) and showed that the varieties of solvable groups are not positively definable. In the author's work [3], it is also shown that the varieties of Engel groups for $n = 2$ and $n = 3$ are positively definable, and are determined respectively by the identities (A) and (B):

$$xy^2x = yx^2y, \tag{A}$$

$$xy^2xyx^2y = yx^2yxy^2x, \quad xy^2xyxy^2y = yx^2y^2x^2y^2x. \tag{B}$$

In the present note, we prove positive definability for a sufficiently broad class of varieties, which generalizes the class of n-nilpotent groups introduced by Baer [1].

Let $[a, b]_s = (ab)^s b^{-s} a^{-s}$ where a, b are elements of a group G and s is an integer, and let $(k) = (k_1, k_2, \ldots, k_t)$ be a t-tuple of integers.

Definition 1. A group G is called *nilpotent relative to the t-tuple* (k) if for any elements a_i, $i = 0, 1, 2, \ldots, t$, the following equality holds:

$$(a_0, a_1, \ldots, a_t)_{(k)} \stackrel{\text{def}}{=} [[\ldots[a_0, a_1]_{k_1}, \ldots]_{k_{t-1}}, a_t]_{k_t} = e. \tag{1}$$

Obviously, the collection of all groups that are nilpotent relative to a fixed t-tuple (k) is a variety. This variety will be denoted by

$$\mathfrak{N}_{(k)} = \mathfrak{N}_{(k_1, k_2, \ldots, k_t)}.$$

Special cases of varieties of the form $\mathfrak{N}_{(k)}$ are the n-nilpotent groups introduced by Baer [1]. The following statement holds.

Sibirsk Mat. Zh. 8 (1967), no. 5, 1190–1192

Theorem. *Any variety of the form $\mathfrak{N}_{(k)}$ is positively definable.*

2. We consider the so-called n-center of the group G.

Definition 2. The collection of all elements z in G, such that $[z,a]_n = e$ for all $a \in G$, is called the *n-center* of G and is denoted by $Z_n(G)$.

Obviously, $Z_{-1}(G)$ coincides with the center of G.

Lemma 1. *For every element $a \in G$ and every $z \in Z_n(G)$ we have $[a,z]_n = e$.*

Proof. The claim follows from the easily verified equation

$$[a,z]_n = a^n [z, z^{-1}a^{-1}]_n a^{-n},$$

and the definition of $Z_n(G)$. □

It is well known, and can be easily verified, that the n-center is a characteristic subgroup. The following statement can also be immediately verified.

Lemma 2. *We have $Z_n(G) = Z_{1-n}(G)$.*

3. We fix a t-tuple $(k) = (k_1, k_2, \ldots, k_t)$ and associate to it two sequences of recursively defined elements of the free group with generators x, y, z_1, z_2, …, z_t:

$$\left.\begin{array}{l} u_0 = x, \quad v_0 = y, \\ u_s = u_{s-1}^{k_s - 1}(v_{s-1}z_s)^{k_s-1}v_{s-1}, \quad v_s = v_{s-1}^{k_s}(z_s u_{s-1})^{k_s-1} \quad \text{for } k_s \geq 1, \\ u_s = u_{s-1}^{-k_s}(v_{s-1}z_s)^{-k_s}v_{s-1}, \quad v_s = v_{s-1}^{1-k_s}(z_s u_{s-1})^{-k_s} \quad \text{for } k_s < 1, \end{array}\right\} \quad (2)$$

for $s = 1, 2, \ldots t$.

Definition 3. A group G is called a *$\overline{(k)}$-group* if it satisfies the identical relation $u_t = v_t$.

Lemma 3. *A group G is a $\overline{(k)}$-group if and only if it is nilpotent relative to the t-tuple (k).*

Proof. We carry out induction on the length of (k), remarking that the case of length 1 is included in the general argument. We write $(k') = (k_1, k_2, \ldots, k_{t-1})$.

Assume it has been proved that any group nilpotent relative to (k') is a $\overline{(k')}$-group, and suppose that a group G is nilpotent relative to (k). Then

$$(a_0, a_1, \ldots, a_{t-1})_{(k')} \in Z_{k_t}(G) = Z_{1-k_t}(G).$$

For this reason, the group $\overline{G} = G/Z_{k_t}(G)$ is nilpotent relative to (k'), and hence by the inductive hypothesis it is a $\overline{(k')}$-group. Therefore,

$$u_{t-1}v_{t-1}^{-1} \in Z_{k_t}(G) = Z_{1-k_t}(G),$$

for any corresponding values of the words u_{t-1} and v_{t-1} in the group G. Hence for any $q \in G$ it follows that

$$(u_{t-1}v_{t-1}^{-1}v_{t-1}q)^\alpha = (u_{t-1}v_{t-1}^{-1})^\alpha (v_{t-1}q)^\alpha, \text{ where } \alpha = \max(k_t, 1-k_t).$$

In other words,

$$(u_{t-1}q)^{\alpha}(v_{t-1}q)^{-\alpha} = (u_{t-1}v_{t-1}^{-1})^{\alpha}. \tag{3}$$

Since the right side does not depend on q, it follows from (3) that

$$(u_{t-1}q)^{\alpha}(v_{t-1}q)^{-\alpha} = (u_{t-1}z_t)^{\alpha}(v_{t-1}z_t)^{-\alpha}, \tag{4}$$

which is valid for all values of z_t. Setting $q = v_{t-1}^{-1}u_{t-1}^{-1}$ in (4), and performing the obvious transformations, we obtain

$$(u_{t-1}v_{t-1}^{-1}u_{t-1}^{-1})^{\alpha}u_{t-1}^{\alpha} = (u_{t-1}z_t)^{\alpha}(v_{t-1}z_t)^{-\alpha},$$

$$u_{t-1}v_{t-1}^{-\alpha}u_{t-1}^{\alpha-1} = (u_{t-1}z_t)^{\alpha}(v_{t-1}z_t)^{-\alpha}, \tag{5}$$

$$u_{t-1}^{\alpha-1}(v_{t-1}z_t)^{\alpha-1}v_{t-1} = v_{t-1}^{\alpha}(z_tu_{t-1})^{\alpha-1}. \tag{6}$$

Therefore, the group G satisfies the identical relation $u_t = v_t$; i.e., it is a $\overline{(k)}$-group.

Conversely, suppose that G is a $\overline{(k)}$-group; i.e., it satisfies relation (6), and hence also (5). Since the left side of relation (5) does not depend on z_t, obviously relation (4) holds. If, in the latter relation, we set $z_t = v_{t-1}^{-1}p$, then we obtain

$$(u_{t-1}q)^{\alpha}(v_{t-1}q)^{-\alpha} = (u_{t-1}v_{t-1}^{-1}p)^{\alpha}p^{-\alpha}. \tag{7}$$

If, in relation (7), we make the two substitutions, $q = e$ and $q = p = e$, then we obtain respectively

$$u_{t-1}^{\alpha}v_{t-1}^{-\alpha} = (u_{t-1}v_{t-1}^{-1}p)^{\alpha}p^{-\alpha}, \tag{8}$$

$$u_{t-1}^{\alpha}v_{t-1}^{-\alpha} = (u_{t-1}v_{t-1}^{-1})^{\alpha}. \tag{9}$$

From these last relations it follows that

$$(u_{t-1}v_{t-1}^{-1}p)^{\alpha} = (u_{t-1}v_{t-1}^{-1})^{\alpha}p^{\alpha}, \tag{10}$$

for all $p \in G$; i.e., $u_{t-1}v_{t-1}^{-1} \in Z_{k_t}(G) = Z_{1-k_t}(G)$. By assumption, the group $\overline{G}$ is nilpotent relative to (k'), and hence G is nilpotent relative to (k). The lemma is proved. □

The statement of the theorem follows trivially from the lemma.

References

[1] R. Baer, *Factorization of n-soluble and n-nilpotent groups*, Proc. Amer. Math. Soc. 4 (1953), no. 1, 15–26.

[2] A.I. Malcev, *Nilpotent semigroups*, Ivanov. Gos. Ped. Inst. Uch. Zap. Fiz.-Mat. Nauki 4 (1953) 107–111.

[3] A.I. Shirshov, *On some groups which are nearly Engel*, Algebra Logika 2 (1963), no. 5, 5–18.

On the Definition of the Binary-Lie Property

A.I. Shirshov

In the present note we construct an example of an algebra over a field of characteristic 2 that satisfies the identical relations

$$x^2 = 0 \quad \text{and} \quad [(xy)y]x + [(yx)x]y = 0,$$

but is not binary-Lie. This example has been announced earlier [2]. For the necessary definitions and the history of the problem, see for example the work [1].

Over an arbitrary field P of characteristic 2, a 16-dimensional algebra A with basis a_i, $i = 1, 2, \ldots, 16$, is determined by the following multiplication table:

$a_i a_j = a_j a_i$ for $i, j = 1, 2, \ldots, 16$,

$a_1 a_2 = a_3$,	$a_1 a_3 = a_5$,	$a_1 a_4 = a_7$,	$a_1 a_5 = a_8$,
$a_1 a_6 = a_{10}$,	$a_1 a_7 = a_{12}$,	$a_1 a_9 = a_{13}$,	$a_1 a_{10} = a_{15}$,
$a_2 a_3 = a_4$,	$a_2 a_4 = a_6$,	$a_2 a_5 = a_7$,	$a_2 a_7 = a_9 + a_{10}$,
$a_2 a_8 = a_{11} + a_{12}$,	$a_2 a_{11} = a_{13}$,	$a_2 a_{12} = a_{15} + a_{16}$,	$a_3 a_4 = a_9$,
$a_3 a_5 = a_{11}$,	$a_3 a_7 = a_{14}$,	$a_4 a_5 = a_{16}$;	

all remaining products equal zero. It is easy to verify directly that the algebra A is generated by the elements a_1 and a_2; it is also obvious that $c^2 = 0$ for all $c \in A$.

Theorem. *The algebra A satisfies the identity*

$$[(xy)y]x + [(yx)x]y = 0, \tag{1}$$

but it is not a Lie algebra; i.e., it is not binary-Lie, since it is generated by two elements.

We remark that identity (1) is equivalent to the identity

$$J(xy, x, y) = 0, \tag{2}$$

where $J(x, y, z) \stackrel{\text{def}}{=} (xy)z + (yz)x + zx)y$ is the Jacobian of the elements x, y, z.

Algebra Logika 10, (1971), no. 1, 100–102.

Lemma. *If, for all distinct basis elements a_i, a_j, a_k, a_ℓ of the algebra A, the following equations hold,*

$$\Phi_1(a_i, a_j) \stackrel{\text{def}}{=} J(a_i a_j, a_i, a_j) = 0,$$

$$\Phi_2(a_i, a_j, a_k) \stackrel{\text{def}}{=} J(a_i a_j, a_i, a_k) + J(a_i a_k, a_i, a_j) = 0,$$

$$\Phi_3(a_i, a_j, a_k, a_\ell) \stackrel{\text{def}}{=}$$
$$J(a_i, a_j, a_k, a_\ell) + J(a_i a_\ell, a_k, a_j) + J(a_k a_j, a_i, a_\ell) + J(a_k a_\ell, a_i, a_j) = 0,$$

then A satisfies the identity (1)*: $\Phi_1(x, y) = 0$.*

Proof. It is easy to see (computing by hand if this is not clear) that

$$\Phi_1(a+b, c+d) = \Phi_1(a,c) + \Phi_1(a,d) + \Phi_1(b,c) + \Phi_1(b,d) + \Phi_2(a,c,d) + \Phi_2(b,c,d) + \Phi_2(c,a,b) + \Phi_2(d,a,b) + \Phi_3(a,c,b,d),$$

$$\Phi_2(a+b, c, a) = \Phi_2(a,c,d) + \Phi_2(b,c,d) + \Phi_3(a,c,b,d).$$

From the above equations, as well as the multilinearity of Φ_3, it follows that for all $u, v \in A$ the element $\Phi_1(u, v)$ can be written as a linear combination of the elements indicated in the statement of the lemma. In general, the indices i, j, k, ℓ occurring in this expression may coincide. But in this case, either the index of the corresponding Φ_s changes or we obviously obtain zero. The lemma is proved. □

Proof. (of the theorem) For the proof of the theorem, we assign to each basis element a_i the weights $p_j(a_i)$, $j = 0, 1, 2$, $i = 1, 2, \ldots, 16$, according to the following table:

$p_j(a_i)$	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}
p_0	1	1	2	3	3	4	4	4	5	5	5	5	6	6	6	6
p_1	1	0	1	1	2	1	2	3	2	2	3	3	3	3	3	3
p_2	0	1	1	2	1	3	2	1	3	3	2	2	3	3	3	3

It easy to verify that $p_s(a_i a_j) = p_s(a_i) + p_s(a_j)$, $s = 0, 1, 2$, if $a_i a_j \neq 0$. As a result of this remark and the lemma, it suffices for the proof of equation (1) to verify the vanishing only of those Φ for which the sum of all weights p_s, $s = 0, 1, 2$, of the arguments of the corresponding Jacobians, does not exceed the limit value; namely, $\Phi_1(a_1, a_2)$, $\Phi_2(a_1, a_2, a_3)$, $\Phi_2(a_1, a_2, a_4)$, $\Phi_2(a_2, a_1, a_3)$, $\Phi_2(a_2, a_1, a_5)$, $\Phi_2(a_3, a_1, a_2)$. The verification is obvious. On the other hand,

$$J(a_1, a_2, a_7) = a_{13} + a_{14} + a_{16} \neq 0.$$

The theorem is proved. □

Remark. Some quotients of the algebra A have the same property. For instance, it is easy to see that the subspace J of A with basis $\{a_6, a_8, a_9, a_{10}, a_{11}, a_{12}, a_{13}, a_{15}, a_{16}\}$ is an ideal, and that in the quotient $B = A/J$ we have $J(b_1, b_2, b_7) = b_{14} \neq 0$ where b_i is the image of a_i under the natural homomorphism of A onto B. Finally, if $P = GF(2)$, then B is a finite ring (with 128 elements) which satisfies the property indicated above.

References

[1] A.T. Gainov, *Binary-Lie algebras of characteristic* 2, Algebra Logika 8 (1969), no. 5, 505–522.

[2] A.I. Shirshov, *On a variety of rings*, Ninth All-Union Algebra Colloquium (Abstracts), Gomel, Belarus, 1968, pages 213–214.

On the Theory of Projective Planes

A.I. Shirshov and A.A. Nikitin

In 1976, in the special course on Projective Planes given at Novosibirsk State University, and later in 1977, in the report on Projective Planes given at the Fourteenth All-Union Algebra Conference, A.I. Shirshov presented the concept of a projective plane as a partial algebraic system. This approach allowed the formulation of a number of new problems, together with a new viewpoint on known results and problems in the theory of projective planes. In the present work, we discuss part of the results contained in the special course and in the report, and also some further developments.

In the study of projective planes, different authors starting with M. Hall [2] implicitly used a partial binary operation. Projective planes as a partial algebraic system were considered for the first time by Magari [7]. The works of Giovagnoli [1] and Kim and Roush [5] also follow this approach.

In [7] and [1], free and completely free projective planes were constructed as partial algebraic systems, in which every element was regarded as an equivalence class defined on the set of nonassociative words in the generators of the plane. In §2 of the present article, we give constructions of free and completely free projective planes as partial algebraic systems, in which each element is uniquely represented as a nonassociative word in the generators of the plane.

In the above-mentioned special course and report, A.I. Shirshov gave a construction of an embedding of the completely free projective plane with a finite number of generators into the completely free projective plane with four generators, and formulated the problem of constructing an embedding of the completely free projective plane with a countable number of generators into the completely free projective plane with a finite number of generators.

In 1972, Johnson [4] showed that every free projective plane with a finite number of generators is a homomorphic image of a completely free projective plane with four generators. In §4 of the present work, we show that in the completely free projective plane $\mathfrak{CF}(C_1)$ with four generators, there exists a countable

Algebra Logika 20, 3 (1981) 330–356.

subconfiguration $\mathfrak{C}_0$ such that $\mathfrak{CF}(C_1)$ is freely generated by $\mathfrak{C}_0$. Based on this result, we further prove that any finite or countably infinite projective plane is a homomorphic image of the completely free projective plane with four generators.

Theorems 1 and 2 were obtained by A.I. Shirshov, and Theorems 3 and 4 by A.A. Nikitin.

1. Preliminary definitions and results

1. Let A be an arbitrary nonempty set, and let A^0 and ${}^0A$ be subsets of A such that $A = A^0 \cup {}^0A$ and $A^0 \cap {}^0A = \emptyset$. In this case we will say that $(A^0, {}^0A)$ is a partition of A. Here one of the subsets A^0 and ${}^0A$ may be empty.

We now fix a partition $(A^0, {}^0A)$ of A. Elements a and b in A will be called *unitypical* relative to the partition $(A^0, {}^0A)$ if a and b belong to the same subset of the partition $(A^0, {}^0A)$. Otherwise, the elements a and b in A will be called *non-unitypical* relative to this partition.

Suppose now that on the set A, with a fixed partition $(A^0, {}^0A)$, a partial binary commutative operation $\cdot$ is defined, such that the following conditions hold:

1.1. If a and b are distinct unitypical elements of A relative to $(A^0, {}^0A)$, then the product $a \cdot b$ is defined.

1.2. If the product $a \cdot b$ is defined for elements a and b in A, then a and b are distinct unitypical elements, but a and $a \cdot b$ are non-unitypical relative to $(A^0, {}^0A)$.

1.3. If the products $a \cdot b$, $a \cdot c$ and $(a \cdot b) \cdot (a \cdot c)$ are defined for elements a, b and c in A, then we have

$$(a \cdot b) \cdot (a \cdot c) = a. \tag{1}$$

1.4. The set A contains pairwise distinct elements a, b, c and d such that the products $a \cdot b$, $b \cdot c$, $c \cdot d$ and $d \cdot a$ are defined and pairwise distinct.

Such a partial algebraic system $\langle A, (A^0, {}^0A), \cdot \rangle$ will be called a *projective plane*.

Suppose that a partial binary commutative operation $*$ is defined on the set A with a fixed partition $(A^0, {}^0A)$ such that Conditions 1.2 and 1.3 hold, as well as the condition

1.5. If the products $a * b$ and $a * c$ are defined for elements a, b and c in A, and $a * b \neq a * c$, then the product $(a * b) * (a * c)$ is also defined.

Here for the operation $*$ one or both of the Conditions 1.1 and 1.4 may not necessarily hold. A partial algebraic system with a partial binary commutative operation $*$ satisfying Conditions 1.2, 1.3 and 1.5 will occasionally be denoted by $\langle A, (A^0, {}^0A), * \rangle$.

Example 1. Let B be an arbitrary nonempty subset of elements of a projective plane $\mathfrak{P} = \langle A, (A^0, {}^0A), \cdot \rangle$. Then the partition $(A^0, {}^0A)$ of the set A of elements of the projective plane $\mathfrak{P}$ determines a partition $(B^0, {}^0B)$ of the set B where $B^0 = B \cap A^0$ and ${}^0B = B \cap {}^0A$. For the elements of B the concepts of unitypical

and non-unitypical elements are defined in the natural way relative to the partition $(B^0, {}^0B)$. The operation $\cdot$ defined in $\mathfrak{P}$ induces a partial binary commutative operation $\circ$ on the set B. For this operation $\circ$ Conditions 1.2, 1.3 and 1.5 hold.

Let $\mathfrak{A} = \langle A, (A^0, {}^0A), * \rangle$ be a partial algebraic system with a partial binary commutative operation $*$ satisfying Conditions 1.2, 1.3 and 1.5. We will say that an element a in A is a *divisor* of an element b in A if there exists an element c in A such that $b = a * c$. The set of all divisors of an element b in A will be denoted by $T_b^{\mathfrak{A}}$.

In what follows, the symbol of the operation $*$ defined in a partial algebraic system satisfying Conditions 1.2, 1.3 and 1.5 will be occasionally omitted if this does not lead to misunderstanding.

Proposition 1. *Let $\mathfrak{A} = \langle A, (A^0, {}^0A), * \rangle$ be a partial algebraic system with a partial binary commutative operation $*$ satisfying Conditions* 1.2, 1.3 *and* 1.5. *Then*

(a) *If the equation $ab = cd$ holds for elements a, b, c in A, and the product ac is defined, then we have $ab = ac$;*

(b) *If a, b, c and d are pairwise distinct elements of A^0 such that the elements ab, bc, cd and da are defined and pairwise distinct, then in the set ${}^0A$ there exist pairwise distinct elements $\overline{a}$, $\overline{b}$, $\overline{c}$ and $\overline{d}$ such that the elements $\overline{a}\overline{b}$, $\overline{b}\overline{c}$, $\overline{c}\overline{d}$ and $\overline{d}\overline{a}$ are defined and pairwise distinct.*

Proof. Indeed, if the product ac is defined and $ab \neq ac$, then from Condition 1.5 it follows that the products $(ab)(ac)$ and $(cd)(ac)$ are defined. From Condition 1.3 and the assumption of the proposition we obtain $a = (ab)(ac) = (cd)(ac) = c$. But this contradicts Condition 1.2. Therefore, $ab = ac$ and part (a) is proved.

For the proof of part (b) it suffices to set $\overline{a} = ab$, $\overline{b} = bc$ and $\overline{c} = cd$, $\overline{d} = da$ and then use the assumptions of the proposition, Conditions 1.3 and 1.5, and the statement of part (a). □

The following result holds:

Proposition 2. [1] *Let $\mathfrak{P} = \langle A, (A^0, {}^0A), * \rangle$ be a projective plane. Then, if the products ab, $(ab)c$ and $[(ab)c]a$ are defined for elements a, b and c, then the following equation holds:* $[(ab)c]a = ab$.

2. Now consider a set A with a fixed partition $(A^0, {}^0A)$ in a different situation. Suppose that a symmetric relation α is defined on the set A, such that α includes only pairs of elements which are non-unitypical relative to the partition $(A^0, {}^0A)$. If, for any elements a, b, c and d in A, the conditions $(a, c), (b, c), (a, d), (b, d) \in \alpha$ imply that at least one of the equations $a = b$, $c = d$ holds, then the relation α is called an *incidence relation relative to the partition* $(A^0, {}^0A)$. The system $\langle A, (A^0, {}^0A), \alpha \rangle$ thus obtained is sometimes called a *partial plane*.

[1] In this regard, see also [5].

If it does not lead to misunderstanding, then an incidence relation α relative to the partition $(A^0, {}^0A)$ will be called an *incidence relation*, and if a pair (a, b) belongs to α then we will sometimes say that the elements a and b are *incident*.

Remark 1. If the elements of A^0 are called 'points', and the elements of ${}^0A$ are called 'lines', and we declare that a point a is incident to a line b if and only if b passes through a, then as a result we obtain an interpretation of a partial plane. If we interchange the names of the elements of A^0 and ${}^0A$, then we obtain another interpretation which is sometimes called the 'dual' of the first interpretation.

Suppose now that on the set A there is a partition $(A^0, {}^0A)$, an incidence relation α relative to $(A^0, {}^0A)$, and a partial binary commutative operation $\cdot$ satisfying Conditions 1.2, 1.3 and 1.5 relative to $(A^0, {}^0A)$. We will say that the operation $\cdot$ and the relation α are *compatible* on A if the following conditions hold:

1.6. If the equation $a \cdot b = c$ holds for elements a, b and c in A, then we have $(a, c) \in \alpha$ and $(b, c) \in \alpha$.

1.7. If $(a, c) \in \alpha$ and $(b, c) \in \alpha$ and also $a \neq b$, then $a \cdot b$ is defined and we have $a \cdot b = c$.

In what follows, a partial algebraic system $\mathfrak{A} = \langle A, (A^0, {}^0A), \cdot, \alpha\rangle$, where the partial binary commutative operation $\cdot$ (satisfying Conditions 1.2, 1.3, 1.5) and the incidence relation α are compatible on A, will be called a *configuration*.

Remark 2. Condition 1.7 implies that the partial operation $\cdot$ in a configuration is uniquely determined by the incidence relation α.

Remark 3. Condition 1.6 implies that if, for each element a in a configuration $\mathfrak{A} = \langle A, (A^0, {}^0A), \cdot, \alpha\rangle$ either $T_a^{\mathfrak{A}} \neq \emptyset$ or there exists an element b in A such that $a \in T_b^{\mathfrak{A}}$, then the relation α is uniquely determined by the partial operation $\cdot$. If we define a symmetric relation $\widetilde{\alpha}$ on the set A in such way that $(a, b) \in \widetilde{\alpha}$ if and only if either $a \in T_b^{\mathfrak{A}}$ or $b \in T_a^{\mathfrak{A}}$, then we obtain the equation $\alpha = \widetilde{\alpha}$. In particular, it follows from this that the definition of projective plane given above is equivalent to the traditional definition, and that any partial plane can be considered as a configuration.

In what follows, we will sometimes omit the set of elements in the symbol for a configuration, and indicate only the partition. Thus, for example, $\mathfrak{A} = \langle (A^0, {}^0A), \cdot, \alpha\rangle$. Here we assume that $A = A^0 \cup {}^0A$ and $A^0 \cap {}^0A = \emptyset$.

Example 2. Let $\mathfrak{B} = \langle B, (B^0, {}^0B), \circ\rangle$ be one of the partial algebraic systems from Example 1, and let $\mathfrak{P} = \langle A, (A^0, {}^0A), \cdot\rangle$ be the projective plane containing $\mathfrak{B}$. Denote by α the incidence relation on A that is compatible with the operation $\cdot$ in $\mathfrak{P}$, and let β be the relation induced by α on the set B, namely $\beta = (B \times B) \cap \alpha$. Then β is an incidence relation compatible on B with the partial operation $\circ$ defined on $\mathfrak{B}$, and the partial algebraic system $\langle (B^0, {}^0B), \circ, \beta\rangle$ is a configuration.

3. At the present time in the theory of projective planes, it is traditional to use a number of definitions going back to [2, 3, 8, 9]. Below in this subsection, we give

the corresponding definitions and concepts in the form that is convenient for the rest of this paper.

A configuration $\mathfrak{B} = \langle (B^0, {}^0B), \circ, \beta \rangle$ will be called an *extension* of a configuration $\mathfrak{A} = \langle (A^0, {}^0A), \cdot, \alpha \rangle$ if $B^0 \supseteq A^0$, ${}^0B \supseteq {}^0A$ and $\beta \supseteq \alpha$.

Remark 4. It immediately follows from this definition, and the compatibility of the corresponding operations and incidence relations in the configurations $\mathfrak{A}$ and $\mathfrak{B}$, that if the product $a \cdot b$ is defined for elements a and b in $\mathfrak{A}$, then the product $a \circ b$ is also defined in $\mathfrak{B}$ and we have $a \cdot b = a \circ b$.

If $\mathfrak{B}$ is an extension of a configuration $\mathfrak{A}$, then $\mathfrak{A}$ will sometimes be called a *subconfiguration* of $\mathfrak{B}$; this will be written in the form $\mathfrak{B} \supseteq \mathfrak{A}$ or $\mathfrak{A} \subseteq \mathfrak{B}$.

An extension $\mathfrak{B}$ of a configuration $\mathfrak{A}$ will be called a *one-step extension*, if for any element a in $\mathfrak{B}$ that is not contained in $\mathfrak{A}$, there exist elements b and c in $\mathfrak{A}$ such that $a = bc$.

A one-step extension $\mathfrak{B}$ of a configuration $\mathfrak{A}$ will be called a *complete* one-step extension if for any two distinct unitypical elements a and b in $\mathfrak{A}$, there exists an element c in $\mathfrak{B}$ such that $ab = c$.

A one-step extension $\mathfrak{B}$ of a configuration $\mathfrak{A}$ will be called a *free* one-step extension if for any element a in $\mathfrak{B}$ that is not contained in $\mathfrak{A}$, there exist two and only two elements b and c in $\mathfrak{A}$ such that $a = bc$.

An extension $\mathfrak{B}$ of a configuration $\mathfrak{A}$ will be called a *complete free* one-step extension if this extension is simultaneously a complete one-step extension and a free one-step extension.

We will say that a configuration $\mathfrak{A} = \langle A, (A^0, {}^0A), \cdot, \alpha \rangle$ is *closed* if the operation $\cdot$ satisfies Condition 1.1.

A subconfiguration $\mathfrak{A}$ of a configuration $\mathfrak{B}$ will be called a *closed* subconfiguration if $\mathfrak{A}$ is closed as a configuration.

A closed subconfiguration $\mathfrak{A}$ of a projective plane $\mathfrak{P}$ will be called a *projective subplane* in $\mathfrak{P}$ if $\mathfrak{A}$ is a projective plane.

Let $\mathfrak{B}$ be a closed configuration and let $\mathfrak{A}$ be a subconfiguration of $\mathfrak{B}$. We denote by $\langle \mathfrak{A} \rangle_{\mathfrak{B}}$ the intersection of all closed subconfigurations in $\mathfrak{B}$ that contain $\mathfrak{A}$ as a subconfiguration.

For a subconfiguration $\mathfrak{A}$ in a closed configuration $\mathfrak{B}$, we set by definition $\mathfrak{A}^{[0]} = \mathfrak{A}$. Now, if for a natural number i the configuration $\mathfrak{A}^{[i-1]}$ is defined, then by $\mathfrak{A}^{[i]}$ we denote the complete one-step extension of the configuration $\mathfrak{A}^{[i-1]}$ in $\mathfrak{B}$. Then we have the following result.

Proposition 3. [2] *The configuration* $\langle \mathfrak{A} \rangle_{\mathfrak{B}}$ *is closed and we have* $\langle \mathfrak{A} \rangle_{\mathfrak{B}} = \bigcup_{i=0}^{\infty} \mathfrak{A}^{[i]}$.

We will say that a closed configuration $\mathfrak{B}$ is *generated by* a configuration $\mathfrak{A}$ if we have $\langle \mathfrak{A} \rangle_{\mathfrak{B}} = \mathfrak{B}$.

In the case when a closed configuration $\mathfrak{B}$ is generated by a configuration $\mathfrak{A}$, and for any natural number i the one-step extension $\mathfrak{A}^{[i]} \supseteq \mathfrak{A}^{[i-1]}$ is a complete free one-step extension, then we will say that $\mathfrak{B}$ is *freely* generated by $\mathfrak{A}$.

[2] See also for example Theorem 11.3 in [3].

A projective plane $\mathfrak{P}$ is called *free* if $\mathfrak{P}$ is freely generated by a configuration $\mathfrak{A} = \langle (A^0, {}^0A), \cdot, \alpha \rangle$ such that the set ${}^0A$ has only one element a, the set A^0 has at least four elements, and only two elements in the set A^0 are not incident to the element a in ${}^0A$.

A configuration $\mathfrak{A} = \langle (A^0, {}^0A), \cdot, \alpha \rangle$ will be called *a configuration without incidence* if $\alpha = \emptyset$.

A projective plane $\mathfrak{P}$ will be called *completely free* if $\mathfrak{P}$ is freely generated by a configuration $\mathfrak{A} = \langle (A^0, {}^0A), \cdot, \emptyset \rangle$ without incidence. In this case, we will also say that $\mathfrak{P}$ is *freely generated by the set* $A = A^0 \cup {}^0A$.

We will say that configurations $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are *freely equivalent* if there exists a natural number n and configurations $\mathfrak{B}_1$, $\mathfrak{B}_2$, ..., $\mathfrak{B}_n$ such that $\mathfrak{B}_1 = \mathfrak{A}_1$, $\mathfrak{B}_n = \mathfrak{A}_2$, and for any natural number i, $i = 1, 2, \ldots, n-1$, either $\mathfrak{B}_i$ is a free one-step extension of $\mathfrak{B}_{i+1}$ or, vice versa, $\mathfrak{B}_{i+1}$ is a free one-step extension of $\mathfrak{B}_i$.

In the case when a projective plane $\mathfrak{P}$ is freely generated by a configuration $\mathfrak{A} = \langle A, (A^0, {}^0A), \cdot, \alpha \rangle$ containing a finite number of elements, then following [2] we will call the number

$$r(\mathfrak{A}) = 2|A| - \frac{1}{2}|\alpha|,$$

the *rank of the configuration* $\mathfrak{A}$. The following result holds.

Remark 5. [2] If $\mathfrak{A}$ is a finite configuration and $\mathfrak{B}$ is a configuration that is freely equivalent to $\mathfrak{A}$, then their ranks are equal: $r(\mathfrak{A}) = r(\mathfrak{B})$.

By virtue of Remark 5, in the case when a projective plane $\mathfrak{P}$ is freely generated by a configuration $\mathfrak{A}$ of finite rank $r(\mathfrak{A})$, it is natural to call the number $r(\mathfrak{A})$ the *rank of the plane* $\mathfrak{P}$. In all other cases, the rank of a freely generated plane will be understood to be the cardinality of the set of its elements.

Remark 6. From the results of [6, §1] it follows that a projective plane $\mathfrak{P}$ is free if and only if either

(1) $\mathfrak{P}$ is a completely free plane of infinite rank, or
(2) $\mathfrak{P}$ is a completely free plane of finite rank, and in this case the rank $r(\mathfrak{P})$ of the plane $\mathfrak{P}$ is an even number, or
(3) $\mathfrak{P}$ is a free plane of finite rank which is an odd number, and in this case there exists a configuration $\mathfrak{A} = \langle (A^0, {}^0A), \cdot, \alpha \rangle$ such that the plane $\mathfrak{P}$ is freely generated by $\mathfrak{A}$, where
   (i) the set A^0 contains n elements, $n \geq 4$, i.e., $A^0 = \{t_1, t_2, \ldots, t_n\}$,
   (ii) the set ${}^0A$ contains one element p, i.e., ${}^0A = \{p\}$,
   (iii) the operation is nowhere defined in $\mathfrak{A}$, and
   (iv) $\alpha = \{(t_n, p), (p, t_n)\}$.

We will say that a plane $\mathfrak{P}$ is *freely generated by the set* A if $\mathfrak{P}$ is freely generated by a configuration of the form $\langle A, (A^0, {}^0A), \cdot, \emptyset \rangle$ where $A = A^0 \cup {}^0A$.

We have the following result:

Proposition 4. [3] *Let $\mathfrak{P}$ be a projective plane freely generated by a configuration $\mathfrak{B}$ without incidence. Then there exists a set A of unitypical elements in $\mathfrak{P}$ such that $\mathfrak{P}$ is freely generated by the configuration $\mathfrak{A} = \langle (A, \emptyset), \cdot, \emptyset \rangle$ and $\mathfrak{A}$ is freely equivalent to $\mathfrak{B}$.*

The following holds:

Proposition 5. [4] *Let a projective plane $\mathfrak{P}$ be freely generated by a configuration $\mathfrak{A}$, and let a configuration $\mathfrak{B}$ be freely equivalent to $\mathfrak{A}$. Then $\mathfrak{P}$ is also freely generated by $\mathfrak{B}$.*

2. Constructions of free and completely free projective planes

1. We give a construction for a completely free plane freely generated by a fixed set V of symbols where $|V| \geq 4$.

Construction 1. Fix a set of pairwise distinct symbols $V = \{v_i\}$ where i ranges over a well-ordered set I of indices and the cardinality of V is at least 4. We denote by $W(V)$ the set of all nonassociative words in the alphabet V. As usual, the number $d(w)$ of occurrences of elements of the set V in a word w in $W(V)$ will be called the *V-length* of w. If it does not lead to misunderstanding, the V-length will be called simply the length.

On the set $W(V)$ we define a lexicographical order as follows. For words u and w in $W(V)$ we set $u > w$ if either (i) the length of u is greater than the length of w, or (ii) the lengths of u and w equal 1 and the index of u is greater than the index of w, or (iii) the lengths of u and w are equal, $u = u_1u_2$, $w = w_1w_2$ and $u_1 > w_1$, or (iv) the lengths of u and w are equal, $u = u_1u_2$, $w = w_1w_2$, $u_1 = w_1$ and $u_2 > w_2$.

The words of length 1 in $W(V)$ will be called *regular words of the first type* (of length 1) relative to the set V. The words of length 2 in $W(V)$ of the form v_iv_j where $v_i > v_j$ will be called *regular words of the second type* (of length 2) relative to the set V.

A word w in $W(V)$ of length $3k + 1$ (respectively $3k + 2$) will be called a *regular word of the first type* (respectively *of the second type*) relative to the set V, if

(1) $w = w_1w_2$ where w_1 and w_2 are regular words of the second (respectively first) type, and w_1 is greater than w_2, and
(2) if $w = (w_1'w_1'')(w_2'w_2'')$ then the intersection of the sets $\{w_1', w_1''\}$ and $\{w_2', w_2''\}$ is empty, and
(3) if $w = ((w_1'w_1'')w_1''')w_2$ or $w = (w_1'''(w_1'w_1''))w_2$ then w_2 is not an element of the set $\{w_1', w_1''\}$.

[3] See for example Lemma 1 in [6].
[4] See for example Theorem 4.2 in [2].

If it does not lead to misunderstanding, then words which are regular relative to the set V will be called simply regular.

The set of all regular words of the first (respectively second) type contained in $W(V)$ will be denoted by W^0 (respectively ${}^0W$).

If, for elements w_1 and w_2 in $W(V)$, one of the words w_1w_2 and w_2w_1 is regular, then we will denote this regular word by $\overline{w_1w_2}$.

On the set $W^0 \cup {}^0W$ we define a partial binary commutative operation $\cdot$ in the following way. Given distinct unitypical regular words w_1 and w_2,

2.1. if one of the words w_1w_2 and w_2w_1 is regular, then $w_1 \cdot w_2 = \overline{w_1w_2}$,

2.2. if $w_1 = w_1'w_1''$, $w_2 = w_2'w_2''$ and the intersection $\{w_1', w_1''\} \cap \{w_2', w_2''\}$ contains an element w, then $w_1 \cdot w_2 = w$,

2.3. if $w_1 = \overline{\overline{(w_1'w_2)}w_1''}$ then $w_1 \cdot w_2 = \overline{w_1'w_2}$, and

2.4. in all other cases the operation $\cdot$ on the elements of $W^0 \cup {}^0W$ is undefined.

The partial algebraic system $\langle (W^0, {}^0W), \cdot \rangle$ obtained in this way will be regarded as *the result of Construction* 1 for the set V and denoted by $\mathfrak{CF}(V)$.

Lemma 1. *The partial algebraic system $\mathfrak{CF}(V)$ is a projective plane.*

Proof. We observe that for the operation $\cdot$ in $\mathfrak{CF}(V)$, Conditions 1.1, 1.2 and 1.4 follow immediately from the definition of this operation and the definition of regular words relative to the set V. To verify Condition 1.3 we need to prove that if w_1, w_2, w_3 are unitypical words such that $w_1 \neq w_2$, $w_1 \neq w_3$ and $w_1 \cdot w_2 \neq w_1 \cdot w_3$, then equation (1) holds. For this, it suffices to consider the following cases[5]:

(a) $w_1w_2 = \overline{w_1w_2}$, and either $w_1 \cdot w_3 = \overline{w_1w_3}$ or $w_3 = \overline{\overline{(w_1w_3')}w_3''}$ or $w_1 = \overline{\overline{(w_3w_1')}w_1''}$ or $w_1 = w_1'w_1''$, $w_3 = \overline{w_1w_3'}$;

(b) $w_1 = w_1'w_1''$, $w_2 = \overline{w_1'w_2'}$, and either $w_3 = \overline{w_1''w_3'}$ or $w_3 = \overline{\overline{(w_1w_3')}w_3''}$;

(c) $w_3 = \overline{\overline{(w_1w_3')}w_3''}$, and either $w_1 = \overline{\overline{(w_2w_1')}w_1''}$ or $w_2 = \overline{\overline{(w_1w_2')}w_2''}$;

(d) $w_1 = w_1'w_1''$, $w_2 = \overline{w_1'w_2'}$, and either $w_3 = \overline{w_1'w_3'}$ or $w_1 = \overline{\overline{(w_3w_1''')}w_1''''}$;

(e) $w_1 = \overline{\overline{(w_2w_1')}w_1''}$, $w_1 = \overline{\overline{(w_3w_1''')}w_1''''}$.

Cases (a)–(c) follow immediately from Conditions 2.1–2.3.

In case (d) the equation $w_3 = \overline{w_1'w_3'}$ and Condition 2.2 imply $w_1 \cdot w_2 = w_1 \cdot w_3$, which contradicts the assumption. Now let $w_1 = \overline{\overline{(w_3w_1''')}w_1''''}$. Then from $w_1 = \overline{w_1'w_1''}$ it follows that either $\overline{w_3w_1'''} = w_1'$, $w_1'''' = w_1''$ or $\overline{w_3w_1'''} = w_1''$, $w_1'''' = w_1'$. From Conditions 2.2 and 2.3, and from $w_1 \cdot w_2 \neq w_1 \cdot w_3$, we obtain $w_1'' = \overline{w_3w_1'''}$ and $w_1' = w_1''''$. Therefore $(w_1 \cdot w_2)(w_1 \cdot w_3) = w_1$.

In case (e), from Condition 2.3 we obtain $w_1 \cdot w_2 = \overline{w_2w_1'}$ and $w_1 \cdot w_3 = \overline{w_3w_1'''}$. From this, and from the equation $w_1 \cdot w_2 \neq w_1 \cdot w_3$ it follows that $\overline{w_2 \cdot w_1'} \neq \overline{w_3w_1'''}$. Hence $w_1'' = \overline{w_3w_1'''}$ and $w_1'''' = \overline{w_2w_1'}$. Therefore

$$(w_1 \cdot w_2) \cdot (w_1 \cdot w_3) = \overline{(w_2w_2')} \cdot \overline{(w_3w_1''')} = \overline{\overline{(w_2w_1')}\,\overline{(w_3w_1''')}} = w_1.$$

[5]In the rest of this proof, there are some typographical errors in the original text, especially regarding the superscripts. We have attempted to correct these errors. [Translators]

All the necessary cases have been considered. Thus the operation $\cdot$ in $\mathfrak{CF}(V)$ satisfies Conditions 1.1–1.4. Therefore the partial algebra system $\mathfrak{CF}(V)$ is a projective plane. □

Now we prove the following result:

Theorem 1. *Let V be a set containing at least four elements, and let $\mathfrak{CF}(V)$ be the partial algebraic system resulting from Construction 1 for the set V. Then $\mathfrak{CF}(V)$ is a completely free projective plane, freely generated by the set V of unitypical elements.*

Proof. From the construction of the plane $\mathfrak{CF}(V)$ it follows that we can choose in $\mathfrak{CF}(V)$ a subconfiguration of the form $\mathfrak{D} = \langle (V, \emptyset), \cdot, \emptyset \rangle$. Consider the sequence of configurations $\mathfrak{D}^{[0]} = \mathfrak{D}$, $\mathfrak{D}^{[1]}$, …, $\mathfrak{D}^{[i]}$.

From the definitions of the operation $\cdot$ in $\mathfrak{CF}(V)$ and the configuration $\mathfrak{D}^{[1]}$, it follows that for any $w \in \mathfrak{D}^{[1]}$, since $w = \overline{uv}$, we have $u, v \in \mathfrak{D}^{[0]}$, and if $w \notin \mathfrak{D}^{[0]}$ and $u_1, v_1 \in \mathfrak{D}^{[0]}$ are such that $w = u_1 v_1$, then $w = \overline{u_1 v_1}$. Thus $\mathfrak{D}^{[1]} \supset \mathfrak{D}^{[0]}$ is a free one-step extension. Therefore we have the basis of the induction.

Assume, for any natural number i not exceeding a natural number s, that $\mathfrak{D}^{[i]} \supset \mathfrak{D}^{[i-1]}$ is a free one-step extension.

Now choose arbitrarily an element $w \in \mathfrak{D}^{[s+1]}$. Then by definition of a complete one-step extension it follows that in $\mathfrak{D}^{[s]}$ there exist elements u and v such that $w = u \cdot v$. The definition of the operation $\cdot$ in $\mathfrak{CF}(V)$ implies that we have the following cases:

(a) $w = \overline{uv}$,
(b) $u = \overline{wu'}$, $v = \overline{wv'}$,
(c) $w = \overline{w_1 w_2}$ and either $u = \overline{wu'}$, $v = w_i$, $i \in \{1, 2\}$, or $v = \overline{wv'}$, $u = w_i$, $i \in \{1, 2\}$.

In cases (b) and (c) it follows from the inductive hypothesis that $w \in \mathfrak{D}^{[s]}$. This contradicts the choice of w in $\mathfrak{D}^{[s+1]}$. Therefore we have the equation $w = \overline{uv}$ and hence the inductive step from s to $s + 1$: that is, $\mathfrak{D}^{[s+1]} \subset \mathfrak{D}^{[s]}$ is a free one-step extension. Therefore, for any natural number i, $\mathfrak{D}^{[i]} \subset \mathfrak{D}^{[i-1]}$ is a free one-step extension, and

$$\mathfrak{CF}(V) = \bigcup_{i=1}^{\infty} \mathfrak{D}^{[i-1]} = \langle \mathfrak{D} \rangle_{\mathfrak{CF}(V)}.$$

From this and Lemma 1 we obtain that $\mathfrak{CF}(V)$ is a completely free projective plane, freely generated by the configuration $\mathfrak{D}$. □

2. Now we apply Construction 1 to build a free projective plane of odd rank.

Construction 2. Let $V_n = \{v_1, v_2, \ldots, v_n\}$ be a set consisting of n ($n \geq 5$) pairwise distinct symbols. We define the elements of V_n to be unitypical. Let $\mathfrak{CF}(V_n) = \langle (W_n^0, {}^0W_n), \cdot \rangle$ be the completely free projective plane freely generated by the set V_n, which is built according to Construction 1. Let W_n^0 and ${}^0W_n$ be the sets of all regular words relative to V_n of the first and second types respectively; let W_n' be

the subset of $W_n^0 \cup {}^0W_n$ consisting of the regular words that are formed from the elements of the set $V_{n-1} = V_n \setminus \{v_n\} = \{v_1, v_2, \ldots, v_{n-1}\}$; let W_n'' be the subset of $W_n^0 \cup {}^0W_n$ consisting of the regular words that have subwords of the form $v_n v_{n-1}$ but do not have subwords of the form $\overline{v_n v}$ where v is an arbitrary regular word of the first type distinct from v_{n-1}.

The operation $\cdot$ defined in the plane $\mathfrak{CF}(V_n)$ induces on the set $W_n' \cup W_n''$ a partial operation $\circ$. The partial algebraic system thus obtained,

$$\big\langle \big((W_n' \cup W_n'') \cap W_n^0, (W_n' \cup W_n'') \cap {}^0W_n\big), \circ \big\rangle,$$

will be denoted by $\mathfrak{F}(\widetilde{V}_n)$ where $\widetilde{V}_n = \{v_1, v_2, \ldots, v_{n-1}; v_n v_{n-1}\}$ and regarded as *the result of Construction* 2.

It immediately follows from the definition that $\mathfrak{F}(\widetilde{V}_n)$ is a closed subconfiguration in the projective plane $\mathfrak{CF}(V_n)$. From this, and from the construction of the configuration $\mathfrak{F}(\widetilde{V}_n)$, it follows that $\mathfrak{F}(\widetilde{V}_n)$ is a projective plane. From the definition of multiplication in the plane $\mathfrak{CF}(V_n)$ it follows that $\mathfrak{F}(\widetilde{V}_n)$ is generated by the configuration of the form

$$\mathfrak{D}_n = \big\langle \widetilde{V}_n, \big(\{v_1, v_2, \ldots, v_n\}, \{v_n v_{n-1}\}\big), *, \nu \big\rangle,$$

where $\nu = \{(v_{n-1}, v_n v_{n-1}), (v_n v_{n-1}, v_n)\}$, and for all elements in $\mathfrak{D}_n$ the operation $*$ is undefined.

If we apply, to the sequence of configurations $\mathfrak{D}_n^{[0]} = \mathfrak{D}_n$, $\mathfrak{D}_n^{[1]}$, …, $\mathfrak{D}_n^{[i]}$, …, arguments analogous to those done in the proof of Theorem 1 for the sequence $\mathfrak{D}$, $\mathfrak{D}^{[1]}$, …, $\mathfrak{D}^{[i]}$, …, then we obtain that for any natural number i, the complete one-step extension $\mathfrak{D}^{[i]} \supset \mathfrak{D}^{[i-1]}$ is a free one-step extension, and

$$\mathfrak{C}(\widetilde{V}_n) = \bigcup_{i=1}^{\infty} \mathfrak{D}_n^{[i-1]} = \langle \mathfrak{D} \rangle_{\mathfrak{C}(\widetilde{V}_n)}.$$

Thus we have the next result.

Proposition 6. *The partial algebraic system* $\mathfrak{F}(\widetilde{V}_n)$ *built in Construction* 2 *is a free projective plane of rank* $2n-1$ *where* $n \geq 5$.

From Propositions 4 and 6, Remark 6, Theorem 1, and Constructions 1 and 2, we obtain the following result.

Remark 7. Any free (including also completely free) projective plane can be regarded as a partial algebraic system in which every element has the form of a suitable regular word.

Remark 8. [6] In [1] and [7] are given constructions of free and completely free projective planes, but the elements of these planes are defined by the authors only up to a certain equivalence relation which is not always convenient for applications.

[6] With regard to completely free projective planes, see also [5].

3. On embeddings of projective planes

1. We have the following construction.

Construction 3. Let $C_1 = \{a_1, a_2, a_3, a_4\}$ be a fixed set of four elements. We define

$$\begin{aligned}
e_1 &= \big[\big((a_4a_2)(a_3a_1)\big)\big((a_4a_1)(a_3a_2)\big)\big](a_2a_1),\\
e_2 &= \big[\big((a_4a_2)(a_3a_1)\big)\big((a_4a_1)(a_3a_2)\big)\big](a_4a_3),\\
e_3 &= \big[\big((a_4a_3)(a_2a_1)\big)\big((a_4a_1)(a_3a_2)\big)\big](a_3a_1),\\
e_4 &= \big[\big((a_4a_3)(a_2a_1)\big)\big((a_4a_1)(a_3a_2)\big)\big](a_4a_2),\\
e_5 &= \big[\big((a_4a_3)(a_2a_1)\big)\big((a_4a_2)(a_3a_1)\big)\big](a_3a_2),\\
e_6 &= \big[\big((a_4a_3)(a_2a_1)\big)\big((a_4a_2)(a_3a_1)\big)\big](a_4a_2).
\end{aligned}$$

We further set

$$\begin{aligned}
&g_1 = (e_2a_2)(e_1a_4), \qquad g_2 = (e_3a_2)(e_2a_1), \qquad g_3 = (e_4a_1)(e_3a_4),\\
&g_4 = (e_5a_1)(e_4a_3), \qquad g_5 = (e_6a_2)(e_1a_3), \qquad g_6 = (e_6a_3)(e_5a_4).
\end{aligned}$$

We will regard the set $G = \{g_1, g_2, g_3, g_4, g_5, g_6\}$ as *the result of Construction* 3. It immediately follows from the definition of the elements of the set G that all of them are regular words relative to C_1. Hence, the set G is contained in the projective plane $\mathfrak{CF}(C_1)$ obtained from the set C_1 according to Construction 1. Now consider the subconfiguration $\langle (G, \emptyset), \cdot, \emptyset\rangle$ in $\mathfrak{CF}(C_1)$, where the operation $\cdot$ is undefined for all pairs of elements in the set G. This configuration will be denoted by $\mathfrak{G}$.

We have the following result.

Proposition 7. *In the projective plane $\mathfrak{CF}(C_1)$, the configuration $\widetilde{\mathfrak{G}} = \langle \mathfrak{G}\rangle_{\mathfrak{CF}(C_1)}$ is a completely free plane, freely generated by the set G which consists of six unitypical elements.*

Proof. For the proof of this statement, it suffices to observe that any word that is regular relative to the set G is also regular relative to the set C_1. Hence, for any natural number i, the complete one-step extension $\mathfrak{G}^{[i]} \supset \mathfrak{G}^{[i-1]}$ is also a free one-step extension within the plane $\mathfrak{CF}(C_1)$. The claim of the proposition follows from this and from the fact that $\widetilde{\mathfrak{G}} = \langle \mathfrak{G}\rangle_{\mathfrak{CF}(C_1)} = \langle \mathfrak{G}\rangle_{\widetilde{\mathfrak{G}}}$. □

Construction 4. Let $V_n = \{v_1, v_2, \ldots, v_n\}$ where $n \geq 6$, and let $\mathfrak{CF}(V_n)$ be the completely free projective plane obtained from the set V_n according to Construction 1. For any quadruple (i_1, i_2, i_3, i_4) of natural numbers such that $1 \leq i_1 < i_2 < i_3 < i_4 \leq n$, we will denote by $h(i_1, i_2, i_3, i_4)$ the word

$$\big((v_{i_4}v_{i_3})(v_{i_2}v_{i_1})\big)\big((v_{i_4}v_{i_2})(v_{i_3}v_{i_1})\big). \tag{2}$$

The set H of all words (2) formed from the elements of the set V_n will be regarded as *the result of Construction* 4. It is clear from the construction of the elements of H that all of them are regular words relative to the set V_n, and hence H is contained

in the projective plane $\mathfrak{CF}(V_n)$. Now consider the configuration $\langle (H, \emptyset), \cdot, \emptyset \rangle$ where the operation $\cdot$ is undefined for all pairs of elements in H. We will denote this configuration by $\mathfrak{H}$.

The proof of the next result is similar to that of Proposition 7.

Proposition 8. *In the projective plane $\mathfrak{CF}(V_n)$ the configuration $\widetilde{\mathfrak{H}} = \langle \mathfrak{H} \rangle_{\mathfrak{CF}(V_n)}$ is a completely free plane, freely generated by the set H of unitypical elements.*

Observe that since the cardinality $|V_n|$ of the set V_n equals n, the cardinality $|H|$ of the set H equals $\binom{n}{4}$. Hence if $n \geq 6$ then $|H| > |V_n|$. This, together with Propositions 7 and 8, implies the following result.

Theorem 2. *Let C_1 be a set which contains four symbols, and let $\mathfrak{CF}(C_1)$ be the completely free projective plane freely generated by the set C_1 of unitypical elements. Then for any natural number $n \geq 4$, there exists a projective subplane of $\mathfrak{CF}(C_1)$ which is a completely free projective plane freely generated by a set of n unitypical elements.*

It is easy to see that the rank of a free plane cannot be smaller than 8. Hence from Theorem 2 and Proposition 6 we obtain the following result.

Corollary 1. [2] *For any natural number $n \geq 8$, the completely free plane $\mathfrak{CF}(C_1)$ contains a projective subplane which is a free projective plane of rank n.*

2. In what follows we will need the next result.

Lemma 2. *Let $\mathfrak{CF}(V)$ be the completely free projective plane freely generated by the set V, and let U be a set of unitypical elements in $\mathfrak{CF}(V)$ such that*

(i) *for any two distinct words u_i and u_j in U, there exists*[7] *a word $\overline{u_i u_j}$ in $\mathfrak{CF}(V)$ which is regular relative to V, and*

(ii) *if, in the expression of an element u_k in U, there occurs a word u which is regular relative to V, then for any words of the form $\overline{uw_1}$ or $\overline{\overline{(uw_2)}w_3}$ which are regular relative to V, we have $u_k \neq \overline{uw_1}$ and $u_k \neq \overline{\overline{(uw_2)}w_3}$.*

Then any word which is regular relative to U is also regular relative to V.

Proof. For words of U-length 1 or 2 that are regular relative to U, the statement of the lemma follows immediately from the assumptions. Thus we have the basis of induction on the U-lengths of the words.

Assume that any word of U-length n which is regular relative to U is also regular relative to V, and consider an arbitrary word w that is regular relative to U and has U-length $n+1 \geq 2$. For the word w there exist words x_1 and x_2 that are regular relative to U such that $w = x_1 x_2$ and the U-lengths of both of these words are strictly less than $n+1$, and hence the inductive hypothesis applies to the words x_1 and x_2.

From the definition of regular words, it follows that if w fails to be regular relative to V, then we are in one of the following cases:

[7] That is, either $u_i u_j$ or $u_j u_i$ is regular relative to V, and hence $\overline{u_i u_j}$ is defined. [Translators]

(a) $x_1 = \overline{y'_1 y''_1}$, $x_2 = \overline{y'_2 y''_2}$, and the intersection $\{y'_1, y''_1\} \cap \{y'_2, y''_2\}$ is not empty;
(b) there exist elements z' and z'' such that for some index $i \in \{1, 2\}$ the elements x_i, x_{3-i}, z', z'' satisfy $x_{3-i} = \overline{\overline{(x_i z')} z''}$.

We consider the two cases separately.

(a) From the definition of regular words, it follows that the words y'_1, y''_1, y'_2, y''_2 cannot simultaneously be regular relative to the set U. Hence, without loss of generality, we can assume that y''_2 is not regular relative to U; but in this case $x_2 \in U$. If now at least one of the elements y'_1, y''_1 is not regular relative to U, then $x_1 \in U$ and hence by the conditions of the lemma there exists a word $\overline{x_1 x_2}$ that is regular relative to V, which contradicts the original assumption. Hence y'_1 and y''_1 are regular relative to U. From this, and from the fact that the intersection $\{y'_1, y''_1\} \cap \{y'_2, y''_2\}$ is not empty, we can assume without loss of generality that $y'_1 = y'_2$. Therefore the word w has the form $x_1 x_2 = \overline{(y'_1 y''_1)}\,\overline{(y'_1 y''_2)}$.

First consider x_1. This word has the form $\overline{y'_1 y''_1}$, where y'_1 and y''_1 are words that are regular relative to U, and hence the U-lengths of both of the words y'_1 and y''_1 are less than the U-length of x. From this, and the fact that $x_2 \in U$, it follows that y'_1 is a word of second type relative to U, and hence there exist words v' and v'', regular relative to U, such that the U-lengths of both of v' and v'' are strictly less than the U-length of y'_1 and we have $y'_1 = v'v''$. But then x_2 has the form $\overline{y'_1 y''_2} = \overline{\overline{(v'v'')} y''_2}$, which contradicts the assumptions of the lemma. For this reason, case (a) is impossible.

(b) From the definition of regular words, it follows that the elements z' and z'' cannot simultaneously be regular relative to U. If z'' is not regular relative to U, then the word $\overline{\overline{(x_i z')} z''}$ must be an element of U. But this contradicts the assumptions of the lemma. If z'' is regular relative to U, then z' is not regular relative to U, and in this case either $\overline{x_i z'} \in U$ or $\overline{\overline{(x_i z')} z''} \in U$, which also contradicts the assumptions of the lemma. Consequently case (b) is also impossible.

Therefore, the word w is regular relative to V, contrary to the assumption. Hence any word w which is regular relative to U and has U-length $n + 1$ is also regular relative to V. The induction is complete. □

Construction 5. Let $C_1 = \{a_1, a_2, a_3, a_4\}$ be a fixed set of symbols, and let $G = \{g_1, g_2, g_3, g_4, g_5, g_6\}$ be the set of words resulting from Construction 3. Take the elements g_1, g_2, g_3, g_4 from G and substitute them for the elements a_1, a_2, a_3, a_4 respectively into the words g_1, g_2, g_3, g_4, g_5, g_6. Denote the resulting words by $g_{1,1}$, $g_{1,2}$, $g_{1,3}$, $g_{1,4}$, $g_{1,5}$, $g_{1,6}$. Now suppose that for any natural number i the words $g_{i,1}$, $g_{i,2}$, $g_{i,3}$, $g_{i,4}$, $g_{i,5}$, $g_{i,6}$ have been constructed. Then we denote by $g_{i+1,1}$, $g_{i+1,2}$, $g_{i+1,3}$, $g_{i+1,4}$, $g_{i+1,5}$, $g_{i+1,6}$ the words resulting from substituting the elements $g_{i,1}$, $g_{i,2}$, $g_{i,3}$, $g_{i,4}$ for the symbols a_1, a_2, a_3, a_4 respectively into the words g_1, g_2, g_3, g_4, g_5, g_6. We denote the element g_6 by $g_{0,6}$. We define the set

$$\overline{G} = \{g_{0,6}, g_{1,6}, \ldots, g_{i,6}, \ldots\},$$

to be *the result of Construction* 5. It follows from the definition of the elements of $\overline{G}$, that for any natural number i, the element $g_{i-1,6}$ is a regular word relative to C_1. Hence the set $\overline{G}$ is contained in the projective plane $\mathfrak{CF}(C_1)$ obtained from the set C_1 according to Construction 1. Consider in $\mathfrak{CF}(C_1)$ the subconfiguration $\langle(\overline{G},\emptyset),\cdot,\emptyset\rangle$ where the operation $\cdot$ is undefined for all pairs of elements of $\overline{G}$. Denote this configuration by $\overline{\mathfrak{G}}$.

It follows from the construction of $\overline{G}$, that C_1 and $\overline{G}$ are sets of unitypical elements satisfying the conditions of Lemma 2. Thus any word that is regular relative to $\overline{G}$ is also regular relative to C_1. Hence for any natural number i, the complete one-step extension $\overline{\mathfrak{G}}^{[i]} \supset \overline{\mathfrak{G}}^{[i-1]}$ in the plane $\mathfrak{CF}(C_1)$ is a free one-step extension. For this reason the closed configuration $\langle\overline{\mathfrak{G}}\rangle_{\mathfrak{CF}(C_1)}$ is a completely free projective plane, freely generated by a countable set $\overline{G}$ of unitypical elements. Therefore we have the following result.

Theorem 3. *Let C_1 be a set of four elements, let $\mathfrak{CF}(C_1)$ be the completely free projective plane freely generated by the set C_1 of unitypical elements, and let $\overline{G}$ be the countable set of elements obtained according to Construction* 5. *Then $\mathfrak{CF}(C_1)$ contains a projective subplane which is a completely free projective plane freely generated by the set $\overline{G}$ of unitypical elements.*

4. On homomorphisms of projective planes

1. The following result holds.

Lemma 3. *Let $\mathfrak{CF}(V)$ be the completely free projective plane, freely generated by the set V of unitypical elements according to Construction* 1. *Let $\mathfrak{A}_1$, $\mathfrak{A}_2$, ..., $\mathfrak{A}_n$, ... be a sequence of subconfigurations in $\mathfrak{CF}(V)$ such that*

(i) $\mathfrak{A}_1 = \langle(V,\emptyset),\cdot,\emptyset\rangle$,

(ii) *for any natural number i the configuration $\mathfrak{A}_{i+1}$ is a free one-step extension of $\mathfrak{A}_i$, and*

(iii) $\mathfrak{A}_0 \stackrel{\text{def}}{=} \bigcup_{i=1}^{\infty} \mathfrak{A}_i \neq \mathfrak{CF}(V)$.

Then the following conditions hold:

(a) *for any natural number n, if $w \in \mathfrak{A}_{n+1}$ and $w \notin \mathfrak{A}_n$ then in $\mathfrak{A}_n$ there exist elements u and v such that $w = \overline{uv}$;*

(b) *the projective plane $\mathfrak{CF}(V)$ is freely generated by the configuration $\mathfrak{A}_0$.*

Proof. (a) In the case $n = 1$, the claim is obvious. Thus we have a basis for the induction. Suppose, for all natural numbers $k < n$, that from $w \in \mathfrak{A}_{k+1}$, $w \notin \mathfrak{A}_k$ it follows that there exist elements u and v in $\mathfrak{A}_k$ such that $w = \overline{uv}$. Now choose arbitrarily an element w in $\mathfrak{A}_{n+1}$ such that $w \notin \mathfrak{A}_n$, and let u and v be the elements in $\mathfrak{A}_n$ such that $w = u \cdot v$.

From the definition of the plane $\mathfrak{CF}(V)$, and Conditions 2.1–2.3 in the definition of the operation $\cdot$ in $\mathfrak{CF}(V)$, it follows that the equation $w = u \cdot v$ holds in $\mathfrak{CF}(V)$ in any of the following cases:

- $w = \overline{uv}$;
- $u = \overline{wu'}$, $v = \overline{wv'}$;
- $w = \overline{w_1 w_2}$ and either $u = \overline{wu'}$, $v = w_i$, $i \in \{1, 2\}$, or $v = \overline{wv'}$, $u = w_i$, $i \in \{1, 2\}$.

Assume that

$$u = \overline{wu'}. \tag{3}$$

Then $u \notin \mathfrak{A}_1$. From this, and from the assumptions of the lemma, it follows that there exists a smallest natural number s such that $1 < s \leq n$, $u \in \mathfrak{A}_s$, $u \notin \mathfrak{A}_{s-1}$. Hence from the inductive hypothesis, the definition of regular words in Construction 1, and equation (3), it follows that $w, u' \in \mathfrak{A}_{s-1}$. But this contradicts the choice of the element w. The case $v = \overline{wv'}$ is treated similarly. Therefore, $w = \overline{uv}$. The inductive step is complete. Part (a) of the lemma is proved.

(b) We show that for any natural number n, the complete one-step extension $\mathfrak{A}_0^{[n]} \supset \mathfrak{A}_0^{[n-1]}$ is a free one-step extension.

Let $n = 1$. We choose arbitrarily an element w in $\mathfrak{A}_0^{[1]}$ such that $w \notin \mathfrak{A}_0^{[0]} = \mathfrak{A}_0$. Then in $\mathfrak{A}_0$ there exist elements u and v such that $w = u \cdot v$. From the definition of the configuration $\mathfrak{A}_0$ it follows that there exists a smallest natural number s such that $u \in \mathfrak{A}_s$. Now, if $u > w$ then from the definition of the multiplication in the plane $\mathfrak{CF}(V)$ it follows that w occurs in the expression of the word u. Hence from part (a) of this lemma we obtain $w \in \mathfrak{A}_s \subset \mathfrak{A}_0$. This contradicts the choice of w. Consequently $u < v$. Similarly one shows that $v < w$. From this and the definition of multiplication in the plane $\mathfrak{CF}(V)$, it follows that $w = \overline{uv}$, i.e., for w there exist two and only two elements u and v in $\mathfrak{A}_0$ such that $w = u \cdot v$. For this reason the extension $\mathfrak{A}_0^{[1]} \supset \mathfrak{A}_0$ is a complete free one-step extension. Therefore we have the basis of induction.

Now suppose, for any natural number $k < n$, that the extension $\mathfrak{A}_0^{[k]} \supset \mathfrak{A}_0^{[k-1]}$ is a complete free one-step extension, and that for any element $w \in \mathfrak{A}_0^{[k]}$ with $w \notin \mathfrak{A}_0^{[k-1]}$ there exist elements u and v in $\mathfrak{A}_0^{[k-1]}$ for which $w = \overline{uv}$. Choose arbitrarily an element $w \in \mathfrak{A}_0^{[n]}$ such that $w \notin \mathfrak{A}_0^{[n-1]}$. Then for this element there exist elements u and v in $\mathfrak{A}_0^{[n-1]}$ such that $w = u \cdot v$. From this, the inductive hypothesis, and the definition of the operation $\cdot$ in the plane $\mathfrak{CF}(V)$, it follows that $w > u$ and $w > v$. Hence $w = \overline{uv}$, and the extension $\mathfrak{A}_0^{[n]} \supset \mathfrak{A}_0^{[n-1]}$ is a complete free one-step extension. The inductive step is complete. Therefore part (b) is also proved. □

2. We have the following construction.

Construction 6. Let $C_1 = \{a_1, a_2, a, a_4\}$ be a fixed set of four elements, and let $\mathfrak{CF}(C_1)$ be the completely free projective plane obtained from C_1 according to Construction 1. In the alphabet C_1, we introduce notation for words which will be needed in what follows: $b_{0,m}$, $c_{0,j}$, $d_{0,j}$, $e_{0,m}$, $f_{0,k}$, $g_{0,\ell}$ where $m = 1, 2, \ldots, 6$, $j = 1, 2, 3$, $k = 1, 2, \ldots, 12$, $\ell = 1, 2, 3, 4$, as displayed in Table 1; in each column, the definition of the element in the first row is given in the second row.

$b_{0,1}$	$b_{0,2}$	$b_{0,3}$	$b_{0,4}$	$b_{0,5}$	$b_{0,6}$	$c_{0,1}$	$c_{0,2}$
a_2a_1	a_3a_1	a_3a_2	a_4a_1	a_4a_2	a_4a_3	$b_{0,4}b_{0,3}$	$b_{0,5}b_{0,2}$

$c_{0,3}$	$d_{0,1}$	$d_{0,2}$	$d_{0,3}$	$e_{0,1}$	$e_{0,2}$	$e_{0,3}$	$e_{0,4}$
$b_{0,6}b_{0,1}$	$c_{0,2}c_{0,1}$	$c_{0,3}c_{0,1}$	$c_{0,3}c_{0,2}$	$d_{0,1}b_{0,1}$	$d_{0,1}b_{0,6}$	$d_{0,2}b_{0,2}$	$d_{0,2}b_{0,9}$

$e_{0,5}$	$e_{0,6}$	$f_{0,1}$	$f_{0,2}$	$f_{0,3}$	$f_{0,4}$	$f_{0,5}$	$f_{0,6}$
$d_{0,3}b_{0,3}$	$d_{0,3}b_{0,4}$	$e_{0,1}a_3$	$e_{0,1}a_4$	$e_{0,2}a_1$	$e_{0,2}a_2$	$e_{0,3}a_2$	$e_{0,3}a_4$

$f_{0,7}$	$f_{0,8}$	$f_{0,9}$	$f_{0,10}$	$f_{0,11}$	$f_{0,12}$	$g_{0,1}$	$g_{0,2}$
$e_{0,4}a_1$	$e_{0,4}a_3$	$e_{0,5}a_1$	$e_{0,5}a_4$	$e_{0,6}a_2$	$e_{0,6}a_3$	$f_{0,4}f_{0,2}$	$f_{0,5}f_{0,3}$

$g_{0,3}$	$g_{0,4}$
$f_{0,7}f_{0,6}$	$f_{0,9}f_{0,8}$

TABLE 1. Definition of the words $b_{0,m}$, $c_{0,j}$, $d_{0,j}$, $e_{0,m}$, $f_{0,k}$, $g_{0,\ell}$.

Now, if for a natural number i we have already defined the elements $b_{i-1,m}$, $c_{i-1,j}$, $d_{i-1,j}$, $e_{i-1,m}$, $f_{i-1,k}$, $g_{i-1,\ell}$ where $m = 1, 2, \ldots, 6$, $j = 1, 2, 3$, $k = 1, 2, \ldots, 12$, $\ell = 1, 2, 3, 4$, then we denote by $b_{i,m}$, $c_{i,j}$, $d_{i,j}$, $e_{i,m}$, $f_{i,k}$, $g_{i,\ell}$ respectively the words obtained by substituting the words $g_{i-1,1}$, $g_{i-1,2}$, $g_{i-1,3}$, $g_{i-1,4}$ for a_1, a_2, a_3, a_4 respectively in the expressions of the words $b_{0,m}$, $c_{0,j}$, $d_{0,j}$, $e_{0,m}$, $f_{0,k}$, $g_{0,\ell}$ where $m = 1, 2, \ldots, 6$, $j = 1, 2, 3$, $k = 1, 2, \ldots, 12$, $\ell = 1, 2, 3, 4$. We denote by C_0 the set of words

$$\{a_1, a_2, a_3, a_4\} \cup \{b_{i,m}, c_{i,j}, d_{i,j}, e_{i,m}, f_{i,k}, g_{i,\ell}\},$$

where $i = 0, 1, \ldots$ and $m = 1, 2, \ldots, 6$, $j = 1, 2, 3$, $k = 1, 2, \ldots, 12$, $\ell = 1, 2, 3, 4$.

Remark 9. Every element of the set C_0 is a regular word relative to C_1, and hence we have $C_0 \subset \mathfrak{CF}(C_1)$.

By definition $C_1 = \{a_1, a_2, a_3, a_4\}$. Now, if for some natural number i the set C_{6i-5} has been defined, then by $C_{6i-4}, C_{6i-3}, C_{6i-2}, C_{6i-1}, C_{6i}, C_{6i+1}$ respectively we denote the following sets:

$C_{6i-5} \cup \{b_{i-1,m}\}$, $m = 1, 2, \ldots, 6$; $\quad C_{6i-4} \cup \{c_{i-1,j}\}$, $j = 1, 2, 3$;

$C_{6i-3} \cup \{d_{i-1,j}\}$, $j = 1, 2, 3$; $\quad C_{6i-2} \cup \{e_{i-1,m}\}$, $m = 1, 2, \ldots, 6$;

$C_{6i-1} \cup \{f_{i-1,k}\}$, $k = 1, 2, \ldots, 12$; $\quad C_{6i} \cup \{g_{i-1,\ell}\}$, $\ell = 1, 2, 3, 4$.

On each of the sets C_i, $i = 0, 1, \ldots$ we define a partial binary commutative operation f_i (respectively) as follows:

4.1. If, for two distinct unitypical elements a and b in the set C_i, the product $a \cdot b$, using the operation $\cdot$ defined in the plane $\mathfrak{CF}(C_1)$, is also contained in C_i, then $f_i(a, b) = a \cdot b$, $i = 0, 1, \ldots$.

a_1	a_2	a_3	a_4	$b_{0,1}$	$b_{0,2}$	$b_{0,3}$	$b_{0,4}$
$b_{0,1}$	$b_{0,1}$	$b_{0,2}$	$b_{0,4}$	a_1	a_1	a_2	a_1
$b_{0,2}$	$b_{0,3}$	$b_{0,3}$	$b_{0,5}$	a_2	a_3	a_3	a_4
$b_{0,4}$	$b_{0,5}$	$b_{0,6}$	$b_{0,6}$	$c_{0,3}$	$c_{0,2}$	$c_{0,1}$	$c_{0,2}$
$f_{0,3}$	$f_{0,4}$	$f_{0,1}$	$f_{0,2}$	$e_{0,1}$	$e_{0,3}$	$e_{0,5}$	$e_{0,6}$
$f_{0,7}$	$f_{0,5}$	$f_{0,8}$	$f_{0,6}$				
$f_{0,9}$	$f_{0,11}$	$f_{0,12}$	$f_{0,10}$				

$b_{0,5}$	$b_{0,6}$	$c_{0,1}$	$c_{0,2}$	$c_{0,3}$	$d_{0,1}$	$d_{0,2}$	$d_{0,3}$
a_2	a_3	$b_{0,3}$	$b_{0,2}$	$b_{0,1}$	$c_{0,1}$	$c_{0,1}$	$c_{0,2}$
a_4	a_4	$b_{0,4}$	$b_{0,5}$	$b_{0,6}$	$c_{0,2}$	$c_{0,3}$	$c_{0,3}$
$c_{0,3}$	$c_{0,1}$	$d_{0,1}$	$d_{0,1}$	$d_{0,2}$	$e_{0,1}$	$e_{0,3}$	$e_{0,5}$
$e_{0,4}$	$e_{0,2}$	$d_{0,2}$	$d_{0,3}$	$d_{0,3}$	$e_{0,2}$	$e_{0,4}$	$e_{0,6}$

$e_{0,1}$	$e_{0,2}$	$e_{0,3}$	$e_{0,4}$	$e_{0,5}$	$e_{0,6}$	$f_{0,1}$	$f_{0,2}$
$b_{0,1}$	$b_{0,6}$	$b_{0,2}$	$b_{0,5}$	$b_{0,3}$	$b_{0,4}$	a_3	a_4
$d_{0,1}$	$d_{0,1}$	$d_{0,2}$	$d_{0,2}$	$d_{0,3}$	$d_{0,3}$	$e_{0,1}$	$e_{0,1}$
$f_{0,1}$	$f_{0,3}$	$f_{0,5}$	$f_{0,7}$	$f_{0,9}$	$f_{0,11}$		$g_{0,1}$
$f_{0,2}$	$f_{0,4}$	$f_{0,6}$	$f_{0,8}$	$f_{0,10}$	$f_{0,12}$		

$f_{0,3}$	$f_{0,4}$	$f_{0,5}$	$f_{0,6}$	$f_{0,7}$	$f_{0,8}$	$f_{0,9}$	$f_{0,10}$
a_1	a_2	a_2	a_4	a_1	a_3	a_1	a_4
$e_{0,2}$	$e_{0,2}$	$e_{0,3}$	$e_{0,3}$	$e_{0,4}$	$e_{0,4}$	$e_{0,5}$	$e_{0,5}$
$g_{0,2}$	$g_{0,1}$	$g_{0,2}$	$g_{0,3}$	$g_{0,3}$	$g_{0,4}$	$g_{0,4}$	

$f_{0,11}$	$f_{0,12}$	$g_{0,1}$	$g_{0,2}$	$g_{0,3}$	$g_{0,4}$
a_2	a_3	$f_{0,2}$	$f_{0,3}$	$f_{0,6}$	$f_{0,8}$
$e_{0,6}$	$e_{0,6}$	$f_{0,4}$	$f_{0,5}$	$f_{0,7}$	$f_{0,9}$
		$b_{1,1}$	$b_{1,1}$	$b_{1,2}$	$b_{1,4}$
		$b_{1,2}$	$b_{1,3}$	$b_{1,3}$	$b_{1,5}$
		$b_{1,4}$	$b_{1,5}$	$b_{1,6}$	$b_{1,6}$
		$f_{1,3}$	$f_{1,4}$	$f_{1,1}$	$f_{1,2}$
		$f_{1,7}$	$f_{1,5}$	$f_{1,8}$	$f_{1,6}$
		$f_{1,9}$	$f_{1,11}$	$f_{1,12}$	$f_{1,10}$

TABLE 2. The incidence relation α_0 (basis of induction).

4.2. In all other cases we declare that the operation f_i on C_i is undefined, $i = 0, 1, \ldots$.

Table 2 gives the incidence relation for elements of C_0: in each column, the first row gives an element $w \in C_0$, and the other rows give the elements of C_0

$b_{i,1}$	$b_{i,2}$	$b_{i,3}$	$b_{i,4}$
$g_{i-1,1}$	$g_{i-1,1}$	$g_{i-1,2}$	$g_{i-1,1}$
$g_{i-1,2}$	$g_{i-1,3}$	$g_{i-1,3}$	$g_{i-1,4}$
$c_{i,3}$	$c_{i,2}$	$c_{i,1}$	$c_{i,1}$
$e_{i,1}$	$e_{i,3}$	$e_{i,5}$	$e_{i,6}$

$b_{i,5}$	$b_{i,6}$	$c_{i,1}$	$c_{i,2}$	$c_{i,3}$	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$
$g_{i-1,2}$	$g_{i-1,3}$	$b_{i,3}$	$b_{i,2}$	$b_{i,1}$	$c_{i,2}$	$c_{i,1}$	$c_{i,2}$
$g_{i-1,4}$	$g_{i-1,4}$	$b_{i,4}$	$b_{i,5}$	$b_{i,6}$	$c_{i,3}$	$c_{i,3}$	$c_{i,3}$
$c_{i,2}$	$c_{i,3}$	$d_{i,1}$	$d_{i,1}$	$d_{i,2}$	$e_{i,1}$	$e_{i,3}$	$e_{i,5}$
$e_{i,4}$	$e_{i,2}$	$d_{i,2}$	$d_{i,3}$	$d_{i,3}$	$e_{i,2}$	$e_{i,4}$	$e_{i,6}$

$e_{i,1}$	$e_{i,2}$	$e_{i,3}$	$e_{i,4}$	$e_{i,5}$	$e_{i,6}$	$f_{i,1}$	$f_{i,2}$
$b_{i,1}$	$b_{i,6}$	$b_{i,2}$	$b_{i,5}$	$b_{i,3}$	$b_{i,4}$	$g_{i-1,3}$	$g_{i-1,4}$
$d_{i,1}$	$d_{i,1}$	$d_{i,2}$	$d_{i,2}$	$d_{i,3}$	$d_{i,3}$	$e_{i,1}$	$e_{i,1}$
$f_{i,1}$	$f_{i,3}$	$f_{i,5}$	$f_{i,7}$	$f_{i,9}$	$f_{i,11}$		$g_{i,1}$
$f_{i,2}$	$f_{i,4}$	$f_{i,6}$	$f_{i,8}$	$f_{i,10}$	$f_{i,12}$		

$f_{i,3}$	$f_{i,4}$	$f_{i,5}$	$f_{i,6}$	$f_{i,7}$	$f_{i,8}$	$f_{i,9}$	$f_{i,10}$
$g_{i-1,1}$	$g_{i-1,2}$	$g_{i-1,2}$	$g_{i-1,4}$	$g_{i-1,1}$	$g_{i-1,3}$	$g_{i-1,1}$	$g_{i-1,4}$
$e_{i,2}$	$e_{i,2}$	$e_{i,3}$	$e_{i,3}$	$e_{i,4}$	$e_{i,4}$	$e_{i,5}$	$e_{i,6}$
$g_{i,2}$	$g_{i,1}$	$g_{i,2}$	$g_{i,3}$	$g_{i,3}$	$g_{i,4}$	$g_{i,4}$	

$f_{i,11}$	$f_{i,12}$	$g_{i,1}$	$g_{i,2}$	$g_{i,3}$	$g_{i,4}$
$g_{i-1,2}$	$g_{i-1,3}$	$f_{i,2}$	$f_{i,3}$	$f_{i,6}$	$f_{i,8}$
$e_{i,1}$	$e_{i,6}$	$f_{i,4}$	$f_{i,5}$	$f_{i,7}$	$f_{i,9}$
		$b_{i+1,1}$	$b_{i+1,1}$	$b_{i+1,2}$	$b_{i+1,4}$
		$b_{i+1,2}$	$b_{i+1,3}$	$b_{i+1,3}$	$b_{i+1,5}$
		$b_{i+1,4}$	$b_{i+1,5}$	$b_{i+1,6}$	$b_{i+1,6}$
		$f_{i+1,3}$	$f_{i+1,4}$	$f_{i+1,1}$	$f_{i+1,2}$
		$f_{i+1,7}$	$f_{i+1,5}$	$f_{i+1,8}$	$f_{i+1,6}$
		$f_{i+1,9}$	$f_{i+1,11}$	$f_{i+1,12}$	$f_{i+1,10}$

TABLE 3. The incidence relation α_0 (inductive step).

which are incident with w in $\mathfrak{CF}(C_1)$. This incidence relation[8] will be denoted by α_0. We construct a sequence of configurations using Tables 2 and 3.

Let C^0 and ${}^0C$ be the sets of regular words of first and second type respectively relative to the set C_1 in the plane $\mathfrak{CF}(C_1)$. Choose arbitrarily an element w in $\mathfrak{CF}(C_1)$ such that for some elements u and v in $\mathfrak{CF}(C_1)$ we have $w = \overline{uv}$.

[8]Table 2 gives the basis of the induction defining α_0, and Table 3 gives the inductive step. [Translators]

Then the pairs of the form (w,u), (u,w), (w,v) and (v,w) will be called the *basic incidences* of the element w, and the set $\{(w,u),\ (u,w),\ (w,v),\ (v,w)\}$ will be denoted by $\mathfrak{O}_w$ and called the *full set of basic incidences* of w in $\mathfrak{CF}(C_1)$.

We define the sequence of sets α_1, α_2, ..., α_i, ... as follows:

4.3. Set by definition $\alpha_1 = \emptyset$.

4.4. If, for some natural number i the set α_i has already been defined, then we define α_{i+1} in this way: for all elements in $C_{i+1} \setminus C_i$ we denote by β_{i+1} the union of all full sets of basic incidences,

$$\beta_{i+1} = \bigcup_{w \in C_{i+1} \setminus C_i} \mathfrak{O}_w,$$

and define $\alpha_{i+1} = \alpha_i \cup \beta_{i+1}$.

For each natural number i, we will denote by $\mathfrak{C}_i$ the partial algebraic system

$$\big\langle (C_i \cap C^0, C_i \cap {}^0C), f_i, \alpha_i \big\rangle, \qquad i = 1, 2, \ldots,$$

where f_i is the partial binary commutative operation defined on C_i and satisfying Conditions 4.1 and 4.2, and the relation α_i is defined for each i according to Conditions 4.3 and 4.4.

It follows immediately from the definitions of f_i and α_i, and the construction of the sets C_i, $i = 1, 2, \ldots$, that α_i is an incidence relation relative to the partition $(C_i \cap C^0, C_i \cap {}^0C)$ such that α_i and f_i are compatible on the set. Therefore we have the following result.

Lemma 4. *For every natural number i, the partial algebraic system*

$$\mathfrak{C}_i = \big\langle (C_i \cap C^0, C_i \cap {}^0C), f_i, \alpha_i \big\rangle,$$

is a configuration, and the extension $\mathfrak{C}_{i+1} \supset \mathfrak{C}_i$ is free.

Denote by $\mathfrak{C}_0$ the configuration equal to the union of the configurations $\mathfrak{C}_i$, $i = 1, 2, \ldots$:

$$\mathfrak{C}_0 = \bigcup_{i=1}^{\infty} \mathfrak{C}_i.$$

The configuration $\mathfrak{C}_0$ will be regarded as *the result of Construction* 6.

Remark 10. It follows from the construction of the sequence of configurations $\mathfrak{C}_1$, $\mathfrak{C}_2$, ..., $\mathfrak{C}_i$, ... that the configuration $\mathfrak{C}_0$ can also be defined as follows:

$$\mathfrak{C}_0 = \big\langle (C_0 \cap C^0, C_0 \cap {}^0C), \alpha_0, f_0 \big\rangle,$$

where α_0 is defined by Tables 2 and 3, and f_0 is the partial binary commutative operation satisfying Conditions 4.1 and 4.2.

For the sequence of configurations $\mathfrak{C}_i$, $i = 1, 2, \ldots$, and the configuration $\mathfrak{C}_0$, by virtue of their constructions and Lemma 4, all the conditions of Lemma 3 are satisfied. Hence the following holds.

Lemma 5. *Let $\mathfrak{CF}(C_1)$ be the completely free projective plane freely generated by the set C_1 of unitypical elements according to Construction* 1*, and let $\mathfrak{C}_0$ be the configuration resulting from Construction* 6*. Then $\mathfrak{CF}(C_1)$ is freely generated by $\mathfrak{C}_0$.*

3. Let $\mathfrak{B}_1$ and $\mathfrak{B}_2$ be configurations contained respectively in the projective planes $\mathfrak{A}_1$ and $\mathfrak{A}_2$. A mapping φ of $\mathfrak{B}_1$ onto $\mathfrak{B}_2$ will be called a *homomorphism of configurations* if the incidence of the elements x and y in $\mathfrak{B}_1$ implies the incidence of the elements $\varphi(x)$ and $\varphi(y)$ in $\mathfrak{B}_2$. In the case $\mathfrak{A}_i = \mathfrak{B}_i$, $i = 1, 2$, we will say that φ is a *homomorphism of projective planes* from $\mathfrak{A}_1$ onto $\mathfrak{A}_2$.

Construction 7. Let $U^0 = \{u_1, u_2, \ldots, u_i, \ldots\}$ be a fixed countable set of symbols ordered according to the indices, and let $\mathfrak{CF}(U^0)$ be the completely free plane freely generated by U^0 according to Construction 1. Denote by ${}^0U$ the set of all words of the form $u_{i+1}u_i$ where $i = 1, 2, \ldots$. Denote by $\mathfrak{U}$ the subconfiguration $\langle (U^0, {}^0U), *, \alpha\rangle$ in the projective plane $\mathfrak{CF}(U^0)$, where α is the union of the full sets of basic incidences for all elements of ${}^0U$,

$$\alpha = \bigcup_{w \in {}^0U} \mathfrak{O}_w,$$

the partial binary commutative operation $*$ is defined on the set $U^0 \cap {}^0U$ as follows,

$$u_{i+1} * u_i = u_{i+1}u_i, \qquad (u_{i+2}u_{i+1}) * (u_{i+1}u_i) = u_{i+1}, \qquad i = 1, 2, \ldots,$$

and all remaining products in $\mathfrak{U}$ are undefined. The configuration $\mathfrak{U}$ just obtained will be regarded as *the result of Construction* 7.

We have the following result.

Lemma 6. *Let U^0 be a countable set of symbols, and let $\mathfrak{CF}(U^0)$ be the completely free projective plane obtained from U^0 according to Construction* 1*. Let C_1 be a set consisting of four symbols, and let $\mathfrak{CF}(C_1)$ be the completely free projective plane obtained from C_1 according to Construction* 1*. Then there exists a homomorphism $\overline{\theta}$ of projective planes from $\mathfrak{CF}(C_1)$ onto $\mathfrak{CF}(U^0)$.*

Proof. First, we construct a mapping θ of the configuration $\mathfrak{C}_0$, obtained as the result of Construction 6, onto the configuration $\mathfrak{U}$, obtained as the result of Construction 7:

$$\begin{aligned}
&\theta(a_\ell) = \theta(c_{0,j}) = \theta(e_{0,m}) = u_1,\\
&\theta(b_{0,m}) = \theta(d_{0,j}) = \theta(f_{0,k}) = u_2u_1,\\
&\theta(g_{0,\ell}) = u_2,\ \ldots,\ \theta(g_{i-1,\ell}) = \theta(c_{i,j}) = \theta(e_{i,m}) = u_{i+1},\\
&\theta(b_{i,m}) = \theta(d_{i,j}) = \theta(f_{i,k}) = u_{i+2}u_{i+1},
\end{aligned}$$

where $m = 1, 2, \ldots, 6$, $j = 1, 2, 3$, $k = 1, 2, \ldots, 12$, $\ell = 1, 2, 3, 4$, $i = 1, 2, \ldots$.

It is clear from inspection of Tables 2 and 3, and the definition of the mapping θ, that if elements a and b are incident in the configuration $\mathfrak{C}_0$ then the elements $\theta(a)$ and $\theta(b)$ are incident in the configuration $\mathfrak{U}$. Hence the mapping θ is a homomorphism of configurations from $\mathfrak{C}_0$ onto $\mathfrak{U}$.

The configuration $\mathfrak{U}$ is freely equivalent to the configuration $\langle U^0, (U^0, \emptyset), \cdot, \emptyset\rangle$, and hence by Proposition 5 the projective plane $\mathfrak{CF}(U^0)$ is freely generated by the configuration $\mathfrak{U}$.

In [3] it is shown that if the plane $\mathfrak{P}_1$ is freely generated by a configuration $\mathfrak{A}_1$, the plane $\mathfrak{P}_2$ is generated by a configuration $\mathfrak{A}_2$, and there exists a homomorphism τ of configurations from $\mathfrak{A}_1$ onto $\mathfrak{A}_2$, then there exists a homomorphism $\overline{\tau}$ of projective planes from $\mathfrak{P}_1$ onto $\mathfrak{P}_1$ such that $\overline{\tau}$ is an extension of τ.

From this, together with Lemma 5, the definition of the configurations $\mathfrak{C}_0$ and $\mathfrak{U}$, and the construction of the homomorphism θ, it follows that there exists a homomorphism $\overline{\theta}$ of projective planes from $\mathfrak{CF}(C_1)$ onto $\mathfrak{CF}(U^0)$ such that $\overline{\theta}$ is an extension of θ. The lemma is proved. □

For what follows we will need the next result.

Proposition 9. [10] *Let $\mathfrak{P} = \langle (P^0, {}^0P), \cdot\rangle$ be a projective plane where P^0 and ${}^0P$ are the sets of elements of the first and second types in P. Then there exists a completely free projective plane $\mathfrak{CF}(\overline{V})$ such that $\mathfrak{CF}(\overline{V})$ is freely generated by a set $\overline{V}$ of unitypical elements, where the cardinalities of the sets P^0 and $\overline{V}$ are equal and there exists a homomorphism of planes from $\mathfrak{CF}(V)$ onto $\mathfrak{P}$.*

Now we will prove the following result.

Theorem 4. *Any finite or countably infinite projective plane is a homomorphic image of a completely free projective plane freely generated by a set of four elements.*

Proof. Let $\mathfrak{P} = \langle (P^0, {}^0P), \cdot\rangle$ be an arbitrary finite or countable infinite projective plane. Then from Proposition 9 it follows that there exists a completely free projective plane $\mathfrak{P}_1 = \mathfrak{CF}(\overline{V})$ such that $\mathfrak{P}_1$ is freely generated by the set $\overline{V}$ of unitypical elements, where the cardinalities of P^0 and $\overline{V}$ are equal, and there exists a homomorphism τ_1 of planes from $\mathfrak{P}_1$ onto $\mathfrak{P}$. Observe that any completely free projective plane, freely generated by a finite or countably infinite set, consists of a countably infinite set of elements.

If U^0 is a countably infinite set of symbols, and $\mathfrak{CF}(U^0)$ is the completely free projective plane freely generated by the set U^0 according to Construction 1, then it follows from Proposition 9 that there exists a homomorphism τ of planes from $\mathfrak{CF}(U^0)$ onto $\mathfrak{P}_1$.

By Proposition 4 it follows that if a plane is freely generated by a set of four elements, then this plane can also be freely generated by a set of four unitypical elements. For this reason let $C_1 = \{a_1, a_2, a_3, a_4\}$ be a set of four elements, let $\mathfrak{CF}(C_1)$ be the completely free plane obtained from C_1 according to Construction 1, and let $\overline{\theta}$ be the homomorphism of planes from $\mathfrak{CF}(C_1)$ onto $\mathfrak{CF}(U^0)$ constructed in Lemma 6. Then we obtain the following sequence of homomorphisms:

$$\mathfrak{CF}(C_1) \xrightarrow{\overline{\theta}} \mathfrak{CF}(U^0) \xrightarrow{\tau_1} \mathfrak{P}_1 \xrightarrow{\tau} \mathfrak{P}.$$

The composition of these homomorphisms gives the required homomorphism of planes from $\mathfrak{CF}(C_1)$ onto $\mathfrak{P}$. □

Corollary 2. [4] *Any projective plane with a finite number of generators is a homomorphic image of a completely free projective plane freely generated by a set of four elements.*

References

[1] A. Giovagnoli, *Sulla rappresentazione di un piano libero mediante una classe di simboli*, Rend. Mat. e Appl. 25, 3–4 (1966) 427–438.

[2] M. Hall, *Projective planes*, Trans. Amer. Math. Soc. 54 (1943) 229–277.

[3] D.R. Hughes and F.C. Piper, *Projective Planes*, Graduate Texts in Mathematics 6, Springer-Verlag, New York-Berlin, 1973.

[4] N.L. Johnson, *Homomorphisms of free planes*, Math. Z. 125 (1972) 255–263.

[5] K.H. Kim and F.W. Roush, *A universal algebra approach to free projective planes*, Aequationes Math. 19, 1 (1979) 48–52.

[6] L.I. Kopeikina, *Free decompositions of projective planes*, Izv. Akad. Nauk SSSR Ser. Mat. 9, 1 (1945) 495–526.

[7] R. Magari, *Su una classe di simboli atta a rappresentare gli elementi di un piano grafico e su un teorema di riduzione a forma normale*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. 33, 1 (1962) 37–44.

[8] G. Pickert, *Projektive Ebenen*, Die Grundlehren der Mathematischen Wissenschaften LXXX, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1955.

[9] G. Pickert, *Projektive Ebenen*, Zweite Auflage, Die Grundlehren der Mathematischen Wissenschaften 80, Springer-Verlag, Berlin-New York, 1975.

[10] L.A. Skornyakov, *Projective planes*, Uspekhi Mat. Nauk 6, 6 (1951) 112–154.

Indication of Sources

[1] Подалгебры свободных алгебр Ли
(Subalgebras of free Lie algebras)
Mat. Sbornik N.S. 33 (75), (1953), no. 2, 441–452

[2] О представлении лиевых колец в ассоциативных кольцах
(On the representation of Lie rings in associative rings)
Uspekhi Mat. Nauk N.S. 8, (1953), no. 5 (57), 173–175

[3] Подалгебры свободных коммутативных и свободных антикоммутативных алгебр
(Subalgebras of free commutative and free anticommutative algebras)
Mat. Sbornik N.S. 34 (76), (1954), no. 1, 81–88

[4] О специальных J-кольцах
(On special J-rings)
Mat. Sbornik N.S. 38 (80), (1956), no. 2, 149–166

[5] Некоторые теоремы о вложении для колец
(Some theorems on embedding of rings)
Mat. Sbornik N.S. 40 (82), (1956), no. 1, 65–72

[6] О некоторых неассоциативных ниль-кольцах и алгебраических алгебрах
(On some nonassociative nil-rings and algebraic algebras)
Mat. Sbornik N.S. 41 (83), (1957), no. 3, 381–394

[7] О кольцах с тождественными соотношениями
(On rings with identical relations)
Mat. Sbornik N.S. 43 (85), (1957), no. 2, 277–283

[8] О свободных кольцах Ли
(On free Lie rings)
Mat. Sbornik N.S. 45 (87), (1958), no. 2, 113–122

[9] О проблеме Левицкого
(On a problem of Levitzki)
Doklady Akad. Nauk SSSR 120, (1958), no. 1, 41–42

[10] Некоторые вопросы теории колец, близких к ассоциативным
(Some problems in the theory of rings that are nearly associative)

Uspekhi Mat. Nauk 13, (1958), no. 6 (84), 3–20
Translated from the Russian original by Murray Bremner
and Natalia Fomenko.
Lect. Notes Pure Appl. Math., 246,
Non-associative algebra and its applications, 441–459,
Chapman & Hall/CRC, Boca Raton, FL, 2006

[11] Некоторые алгоритмические проблемы для ε-алгебр
(Some algorithmic problems for ε-algebras)
Sibirsk Mat. Zh. 3, (1962), no. 1, 132–137

[12] Некоторые алгоритмические проблемы для алгебр Ли
(Some algorithmic problems for Lie algebras)
Sibirsk Mat. Zh. 3, (1962), no. 2, 292–296

[13] Об одной гипотзе теории алгебр Ли
(On a hypothesis in the theory of Lie algebras)
Sibirsk Mat. Zh. 3, (1962), no. 2, 297–301

[14] О базах свободных алгебр Ли
(On the bases of a free Lie algebra)
Algebra Logika 1, (1962), no. 1, 14–19

[15] О некоторых группах, близких к энгелевым
(On some groups which are nearly Engel)
Algebra Logika 2, (1963), no. 5, 5–18

[16] О некоторых тождественних соотношениях в алгебрах
(On some identical relations for algebras)
Sibirsk Mat. Zh. 7, (1966), no. 4, 963–966

[17] О некоторых положительно определенных мнолообразиях групп
(On some positively definable varieties of groups)
Sibirsk Mat. Zh. 8, (1967) no. 5, 1190–1192

[18] К определению бинарной лиевости
(On the definition of the binary-Lie property)
Algebra Logika 10, (1971), no. 1, 100–102

[19] (с А.А. Никитиным) К теории проективных плоскостей
(with A.A. Nikitin. On the theory of projective planes)
Algebra Logika 20, (1981), no. 3, 330–356

COMMENTS ON SHIRSHOV'S HEIGHT THEOREM

ALEXANDER KEMER

In 1941 A.G. Kurosh [1] posed the problem: Is every finitely-generated algebraic associative algebra finite-dimensional? In 1964 E.S. Golod and I.R. Shafarevich [2, 3] constructed a counterexample: they presented an infinite-dimensional finitely-generated nil-algebra. This counterexample shows that in general finitely-generated algebraic associative algebras are very far from being finite-dimensional.

Every problem can be considered not only as an explicit problem but as a direction of research. In the case of Kurosh's problem such a direction can be formulated in the following way: Find the conditions which imply that a finitely generated algebra is finite-dimensional.

Before the counterexample of Golod-Shafarevich was constructed, many positive results on Kurosh's problem were obtained. In 1945 N. Jacobson [4] solved the problem of Kurosh for algebraic algebras of bounded index. In 1946 J. Levitzky [5] proved that for a finitely generated PI-algebra over a commutative ring, if each element is nilpotent then the algebra is nilpotent. Finally, in 1948 I. Kaplansky [6] solved Kurosh's problem for PI-algebras over a field. All of these results became classical and are included in textbooks on ring theory. The great role of these results in ring theory is well known. In fact, the structure theory of rings developed around the problem of A.G. Kurosh.

In 1957 A.I. Shirshov proved his famous theorem on height:

Theorem (A.I. Shirshov [7]). *For any finitely-generated associative PI-algebra A over a commutative ring R with* 1*, there exist a natural number h and elements $a_1, \dots, a_n \in A$ such that any element of A can be represented as an R-linear combination of elements of the form*

$$a_{i_1}^{\alpha_1} \cdots a_{i_k}^{\alpha_k},$$

where $k < h$.

We note that an algebra A over a commutative ring R with 1 is called a PI-algebra if A satisfies some polynomial identity $f = 0$ such that the ideal of the ring R generated by the coefficients of the highest-degree terms of the polynomial f contains 1.

The positive solution of Kurosh's problem for PI-algebras over a ring follows immediately from Shirshov's theorem. Indeed, since the elements $a_1, \dots, a_n \in A$ are algebraic (the elements $a_1, \dots, a_n$ are taken from the conclusion of the theorem on height), the degrees α_i are bounded. Hence the algebra A is a finitely-generated R-module.

Comparing the solutions of Kurosh's problem obtained by I. Kaplansky and A.I. Shirshov one notes that the solution of I. Kaplansky is based on the well-developed structure theory of rings, but makes little use of the PI-condition. In fact, the PI-condition is used in two statements: (1) The radical of a finitely-generated algebraic PI-algebra is nilpotent; (2) A matrix algebra of order n does not satisfy a polynomial identity of degree less than $2n$. These statements are quite easy from the contemporary point of view.

The solution of A.I. Shirshov does not use the structure theory at all. Moreover A.I. Shirshov also made little use of algebraicity. It follows from the above that it is sufficient to

require algebraicity only for some finite set of elements. But the most important merit of the theorem on height is that it was proved for algebras over a commutative ring. Many of the results in ring theory concerning PI-algebras would not have been obtained if the theorem on height were true only for algebras over fields.

With the first results about PI-algebras it became clear that the PI-condition is a peculiar finiteness condition. In 1957 S. Amitsur [8] proved a remarkable theorem: The radical of a finitely-generated PI-algebra is a nil-ideal. This theorem once again corroborated that the PI-condition is a finiteness condition, and allowed V.N. Latyshev at that time to formulate rather boldly the problem: Is the radical of a finitely-generated PI-algebra nilpotent? (See [9]). A great contribution to the solution of this problem was made by Yu.P. Razmyslov [10] who proved that the radical of finitely-generated PI-algebra over a field is nilpotent if and only if the algebra satisfies some standard identity. To prove this statement, Yu.P. Razmyslov constructed an embedding of certain algebras into algebras which are algebraic over the center and then applied the theorem on height. Yu.P. Razmyslov was the first algebraist to apply the theorem on height very often and deeply. For algebras over a field of characteristic zero, Latyshev's problem was solved by A.R. Kemer [11] who proved that every finitely-generated PI-algebra over a field of characteristic zero satisfies a standard identity of some order. Indeed this result and the theorem of Razmyslov mentioned above imply the positive solution of Latyshev's problem in the case of characteristic zero. In 1982, A. Braun [12] solved Latyshev's problem positively for algebras over a commutative Noetherian ring. At present the theorem on the nilpotency of the radical of a finitely-generated PI-algebra is known as the theorem of Braun-Kemer-Razmyslov.

In 1974, Yu. P. Razmyslov introduced a new concept of trace identity, and proved that each trace identity of the matrix algebra of order n over a field of characteristic 0 follows from the Cayley-Hamilton trace identity of degree n and the identity $\mathrm{Tr}(1) = n$ [13]. Little later C. Procesi [14] proved actually the same result in the terms of invariants.

The Cayley-Hamilton identity of degree n has the form

$$X_n(x) = x^n + b_1(x)x^{n-1} + \cdots + b_n(x) = 0,$$

where the coefficient $b_m(x)$ is a form of degree m. In the case of characteristic zero the coefficients $b_m(x)$ can be represented as linear combinations of trace monomials of the form

$$\mathrm{Tr}(x^{i_1})^{\alpha_1}\mathrm{Tr}(x^{i_2})^{\alpha_2}\cdots\mathrm{Tr}(x^{i_k})^{\alpha_k}.$$

Of course this theorem of Yu.P. Razmyslov does not concern the theorem on height directly, but the idea of trace identities gives a way of embedding (if possible) a finitely-generated PI-algebra over a field into a finite-dimensional algebra (a matrix algebra) over a larger field (such algebras are called representable). Indeed, let a finitely-generated algebra A over a field F be embeddable into the matrix algebra $M_n(K)$, $F \subseteq K$. Consider the F-subalgebra $C = SA$, where S is the F-subalgebra (with unity) of the field K generated by all the elements $b_m(a)$ $(a \in A)$ where the elements $b_m(a)$ are the coefficients of the Cayley-Hamilton identity of degree n. It follows from this that in the case of characteristic zero the algebra A is embeddable into the algebra

$$D = A \otimes T\langle A\rangle/J,$$

where $T\langle A\rangle$ is the commutative algebra generated by the symbols $\mathrm{Tr}(a), a \in A$, the trace on the algebra $A \otimes T\langle A\rangle$ is defined by the formula

$$\mathrm{Tr}\Big(\sum a_k \otimes t_k\Big) = \sum \mathrm{Tr}(a_k)t_k,$$

and the ideal J is generated by the elements $X_n(d)$ $(d \in A \otimes T\langle A\rangle)$. In the case of characteristic p the algebra $A \otimes T\langle A\rangle$ is generated by the symbols $b_m(a)$ $(a \in A)$. The forms $b_m(x)$ are defined in the same manner but with more complicated formulas.

Assume that the algebra A is embeddable into the algebra D. Then the algebra A is embeddable into the algebra

$$D' = A \otimes T'\langle A\rangle / J \cap A \otimes T'\langle A\rangle,$$

where $T'\langle A\rangle$ is the subalgebra of $A \otimes T\langle A\rangle$ generated by the elements $b_m(a_i)$ (the elements a_i are taken from the conclusion of the theorem on height). The algebra D' is finitely-generated and algebraic over the commutative algebra $T'\langle A\rangle$ because it satisfies the Cayley-Hamilton identity. By the theorem on height the algebra D' is a finitely-generated $T'\langle A\rangle$-module. Since the algebra $T'\langle A\rangle$ is noetherian, by a theorem of K. Beidar [15] the algebras D' and A are representable. In 1995 the theorem of Razmyslov in the case of characteristic p was proved by A.R. Kemer at the multilinear level [16] and little later A.N. Zubkov proved this theorem at the homogeneous level [17].

A very important problem in the theory of PI-algebras was posed by W. Specht [18] in 1950: Does every associative algebra over a field of characteristic zero have a finite basis of identities? The finite basis problem makes sense for algebras over any field, and even for rings, groups and arbitrary general algebraic systems. A positive solution of the finite basis problem for a given class of algebraic systems is a sort of classification of these algebraic systems in the language of identities.

A rather large number of papers have been devoted to Specht's problem for associative algebras over a field of characteristic zero. We note the most important results. In 1977 V.N. Latyshev [19] proved that any associative algebra over a field of characteristic zero satisfying a polynomial identity of the form

$$[x_1, \ldots, x_n] \cdots [y_1, \ldots, y_n] = 0,$$

has a finite basis of identities. This result was also obtained independently by G. Genov [20] and A. Popov [21].

In 1982 A.R. Kemer reduced the Specht problem to the finite basis problem for graded identities of finitely-generated associative PI-superalgebras [22] and in 1986 he solved the Specht problem positively [23]. The first proof of the theorem on the finite basis of identities was rather complicated. A little later in 1987 A.R. Kemer [24] proved that relatively free finitely-generated associative PI-superalgebras over a field of characteristic zero are representable. This theorem implies the theorem on the finite basis, and explains the reason why the Specht problem has a positive solution. This reason is that finite-generated PI-algebras over a field of characteristic zero cannot be distinguished in the language of identities from finite-dimensional algebras. More precisely, for every finitely-generated PI-algebra A there exists a finitely-dimensional algebra C such that the ideals of identities of these algebras are equal. In 1988 A.R. Kemer proved the same result for algebras over an infinite field of characteristic p [25].

The main idea of the proof of this theorem is to approach step-by-step the given T-ideal Γ by the ideals of identities of finite-dimensional algebras. At the first step there is constructed

a finite-dimensional algebra C_0 such that

$$T[C_0] \subseteq \Gamma.$$

The existence of this algebra follows from the theorem on nilpotency of Braun-Kemer-Razmyslov and the theorem of J. Lewin [33]. The most difficult part of the proof is the following statement: If $T[C] \subseteq \Gamma$, $T[C] \neq \Gamma$ (C is finite-dimensional) then there exists a finite-dimensional algebra C' such that

$$T[C] \subseteq T[C'] \subseteq \Gamma, \qquad T[C] \neq T[C'].$$

The proof of this statement uses identities with forms and the standard application of the theorem on height which was described above.

Examples of infinitely-based algebras in the case of characteristic p were constructed in 1999 by V.V. Schigolev [26] and A.Ya. Belov [27].

In 1998 A.Ya. Belov [28] announced a positive solution of the local finite basis problem for algebras over a commutative noetherian ring, and announced a result about the representability of the relatively free algebra over a commutative noetherian ring in some weak sense: The relatively free finitely-generated PI-algebra A over a commutative noetherian ring R is embeddable into some algebra A' over a commutative noetherian ring R' such that A' is a finitely-generated R'-module ($R \subseteq R'$). In other words the algebra A is embeddable into the algebra of endomorphisms of some finitely generated R'-module.

Regarding the methods of A.Ya. Belov we should note that most of the ideas of A.Ya. Belov are combinatorial, and come from the theorem on height and other results of A.I. Shirshov. A.Ya. Belov developed the combinatorial ideas of A.I. Shirshov which made it possible to consider more complicated combinatorial situations than in the theorem on height. In this sense one can call A.Ya. Belov a successor of A.I. Shirshov.

Another nice idea is applying Zariski closure. This idea was new for PI-theory. The algebras of endomorphisms of finitely generated modules over a ring have a more complicated structure than finite-dimensional algebras, but applying Zariski closure A.Ya. Belov proved that a finitely-generated PI-algebra A over a commutative noetherian ring R has the same identities as some algebra C over a commutative noetherian ring R', $R \subseteq R'$, satisfying the property that the radical of the algebra C splits off and is nilpotent, i.e., $C = P + \operatorname{Rad} C$, where the subalgebra C is semisimple. Applying Zariski closure A.Ya. Belov also obtained a lot of information about the semiprime part P. We note that the main results of A.Ya. Belov are not yet published.

We also mention the results devoted to the estimation of height in the theorem of A.I. Shirshov. The height $h(A)$ of an algebra A depends on the number of generators s and the minimal degree of identities $m = \deg(A)$. The estimate for the height which follows from the proof of the theorem on height is not satisfactory. In 1982 A.G. Kolotov [29] obtained the estimate

$$h(a) \leq s^{s^m}.$$

In [30] E.I. Zelmanov raised a question about the exponential estimate of the height. The positive answer was obtained by A. Ya. Belov in 1988 [31, 32].

References

[1] A.G. Kurosh, Ringtheoretische Probleme, die mit dem Burnsideschen Problem uber periodische Gruppen in Zusammenhang stehen, Izv. Akad. Nauk SSSR. Ser. Mat., v. 5(1941), 233–240. (russian)

[2] E.S. Golod, On nil-algebras and residually finite p-groups, Izv. Akad. Nauk SSSR. Ser. Mat., v. 28(1964), 273–276; English transl. in Amer. Math. Soc. Transl., v. 48(1965).
[3] E.S. Golod, I.R. Shafarevich, On class field towers, Izv. Akad. Nauk SSSR. Ser. Mat., v. 28(1964), 261-272; English transl. in Amer. Math. Soc. Transl., v 48(1965).
[4] N. Jacobson, Structure theory for algebraic algebras of bounded degree, Ann. of Math., v. 2(1945), 695-707.
[5] J. Levitzky, On a blem of A. Kurosh, Bull. Amer. Math. Soc., v. 52(1946), 1033-1035.
[6] I. Kaplansky, Rings with a polynomial identity, Bull. Amer. Math. Soc., v. 54(1948), 575-580.
[7] A.I. Shirshov, On rings with polynomial identities, Mat. Sb., v. 43(1957), 277-283; English transl. in Amer. Math. Soc. Transl., v. 119(1983).
[8] S.A. Amitsur, A generalization of Hilbert's Nullstellensatz, Proc. Amer. Math. Soc., v. 8(1957), 649-656.
[9] The Dniester notebook: unsolved problems in the theory of rings and modules, 3rd ed. (V.A. Andrunakievich, editor), Inst. Mat. Sibirsk. Otdel. Akad. Nauk SSSR, Novosibirsk, (1982)(Russian).
[10] Yu.P. Razmyslov, On the Jacobson radical in PI-algebras, Algebra i logika, v. 13(1974), 337-360; English transl. in Algebra and logic, 13(1974).
[11] A.R. Kemer, Capelli identities and the nilpotence of the radical of a finitely-generated PI-algebra, Dokl. Akad. Nauk SSSR, v. 255(1980), 793-797; English transl. in Soviet Math. Dokl., v. 22(1980).
[12] A. Braun, The radical in finitely-generated $P.I.$ algebra, Bull. (New Ser.) Amer. Math. Soc., v. 7(1982), 385-386.
[13] Yu.P. Razmyslov, Trace identities of full matrix algebras over field of characteristic zero, Izv. Akad. Nauk SSSR Ser. Mat., v. 38(1974), 723-756; English transl. in Math. USSR Izv. v. 8.(1974)
[14] C. Procesi, The invariant theory of $n \times n$- matrices, Adv. in Math., v. 19(1076), 306-381.
[15] K.I. Beidar, On theorems of A.I. Mal'tsev concerning matrix representations of algebras, Uspekhi Mat. Nauk, v. 41(1986), 161-162; English transl. in Ruusian Math. Sueveys., v. 41(1986).
[16] A.R. Kemer, Multilinear identities of the algebras over a field of characteristic p, Int. J. of Alg. and Comp., v. 5(1995), 189-197.
[17] A.N. Zubkov, On the generalization of the theorem of Procesi-Razmyslov, Algebra i Logika, v. 35(1996), 433-457. English transl. in Algebra and logic, 35(1996).
[18] W. Specht, Gesetze in Ringen.I, Math. Z., v. 52(1950), 557-589.
[19] V.N. Latyshev, On the finite basis property for the identities of certain rings, Uspekhi Mat. Nauk v. 32(1977), 259-260. (Russian)
[20] G.K. Genov, Some Specht varieties of associative algebras, Pliska Stud. Math. Bulgar v. 2(1981), 30-40. (Russian)
[21] A.P. Popov, On the Specht property for some varieties of assocuative algebras, Pliska Stud. Math. Bulgar v. 2(1981), 41-53. (Russian)
[22] A.R. Kemer, Varieties and Z_2-graded algebras, Izv. Akad. Nauk SSSR. Ser Mat., v. 48(1982), 1042-1059; English transl. in Math. USSR Izv. v. 25(1985).
[23] A.R. Kemer, Finite bases for identities of associative algebras, Algebra i Logika v. 26(1987), 597-641; English transl. in Algebra and logic, 26(1987).
[24] A.R. Kemer, Representation of relatively free algebras, Algebra i Logika v. 27(1988), 274-294; English transl. in Algebra and logic, 27(1988).
[25] A.R. Kemer, Identities of finitely generated algebras over an infinite field, Izv. Akad. Nauk SSSR. Ser Mat., v. 54(1990), 726-753; English transl. in Math. USSR Izv. v. 29(1990).
[26] V.V. Schigolev, Examples of infinitely-based T-ideals, Fund. and Appl. Math., v. 5(1999), 307-312.
[27] A.Ya. Belov, On non-spechtian varieties, Fund. and Appl. Math., v. 5(1999), 47-66.
[28] A.Ya. Belov, Local representability of relatively free associative algebras, Kurosh Algebraic conference - 98. Abstracts of talks. Ed. by Yu.A. Bahturin, A.I. Kostrikin, A. Yu. Ol'shansky. Moscow, 1998, 143-144.
[29] A.G. Kolotov, On the upper estimation of the height in finitely-generated PI-algebras, Sibirsk. Mat. Zh., v.23(1982), 187-189; English transl. in Siberian Math. J. v.23(1982).
[30] The Dniester notebook: unsolved problems in the theory of rings and modules, 3rd ed. (V.A. Andrunakievich, editor), Inst. Mat. Sibirsk. Otdel. Akad. Nauk SSSR, Novosibirsk, (1993)(Russian).
[31] A.Ya. Belov, Estimations for the height and Gelfand-Kirillov dimension of associative PI-algebras, Abstracts of Int. alg. Maltsev's conf. Novosibirsk, 1989. (Russian)

[32] A.Ya. Belov, Some estimations for nilpotence og nil-algebras over field of an arbitrary characteristic and height theorem, Comm. in algebra. v.20(1992), 2919-2922.
[33] J. Lewin, A matrix representation for associative algebras. I, II, Trans. Amer. Math. Soc., v. 188(1974), 293-308, 309-317.

КАК ПЕРЕСЕКАЮТСЯ В ПРОСТРАНСТВЕ КРИВОЛИНЕЙНЫЕ СФЕРЫ, ИЛИ ДВУМЕРНЫЕ МЕАНДРЫ [1]

С. Аввакумов, А. Бердников, А. Рухович и А. Скопенков [2]

1 Примеры и основные задачи

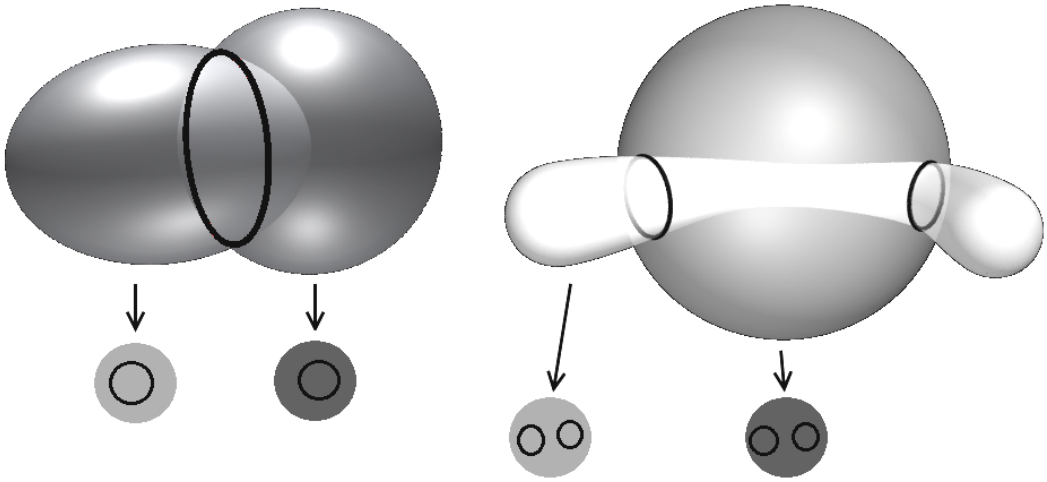


Рис. 1: Криволинейные сферы, пересекающиеся по окружности (слева) и по двум окружностям (справа)

Как могут пересекаться криволинейные сферы в трехмерном пространстве? На рисунках 1 и 2 изображены пары криволинейных сфер в трехмерном пространстве, пересекающиеся по набору окружностей.

Видимо, термин *криволинейная сфера* интуитивно понятен Вам. Если нет, то прочтите строгое определение ниже. Здесь мы только отметим, что в этом тексте криволинейные сферы подразумеваются *не имеющими самопересечений*.

Скорее всего, для решения задач Вам не понадобятся строгие определения. Пожалуйста, изображайте криволинейные сферы с помощью понятных (для членов жюри) рисунков, а не задавайте их формальными конструкциями. Если задача сформулирована как утверждение, то требуется его доказать. Если задача оказывается слишком сложной, попробуйте решить соседние задачи, они могут содержать подсказки. Нерешенные задачи отмечены звездочками.

1.1. Нарисуйте пару криволинейных сфер в трехмерном пространстве, пересекающихся по объединению трех непересекающихся окружностей так, что в каждой сфере эти окружности

(a) ограничивают 3 диска (как справа-снизу-слева на рис. 2).

(b) не ограничивают 3 дисков (как справа-снизу-справа на рис. 2).

1.2. Нарисуйте пару криволинейных сфер в трехмерном пространстве, пересекающихся по объединению четырех непересекающихся окружностей так, что в каждой сфере эти окружности

(a) ограничивают 4 диска (как на рис. 3.a).

(b) „параллельны“, или „одна внутри другой“(как на рис. 3.b).

(c) расположены как на рис. 3.c.

[1] Мы благодарим за полезные замечания и обсуждения С. Ландо и анонимного рецензента Московской Математической Конференции Школьников.

[2] Поддержан Грантом фонда Саймонса-НМУ

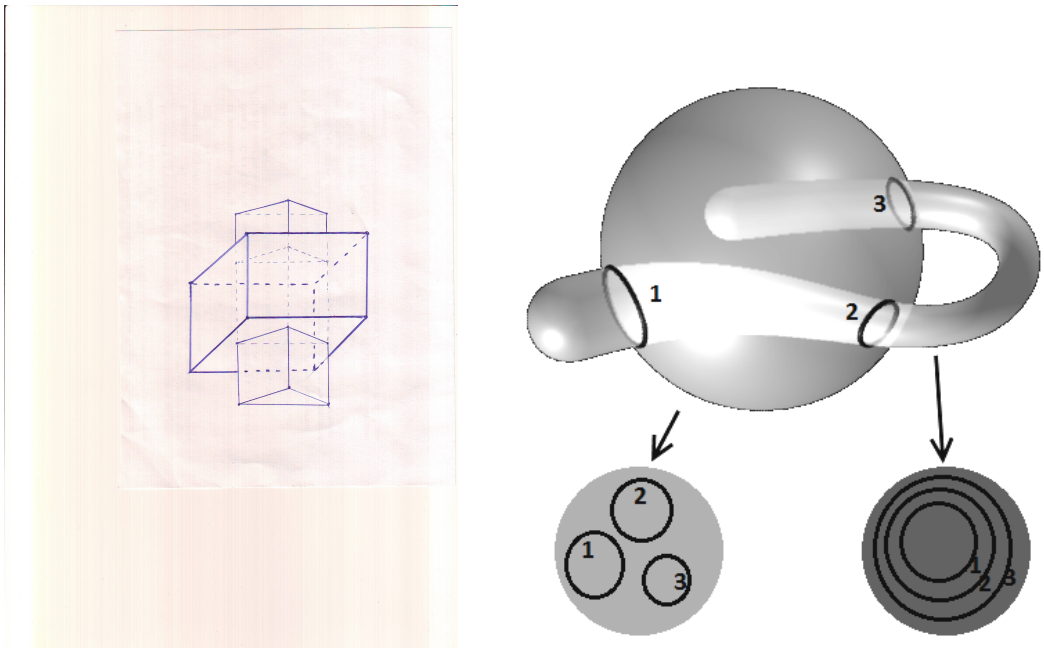


Рис. 2: Криволинейные сферы, пересекающиеся по двум окружностям (слева) и по трем окружностям (справа)

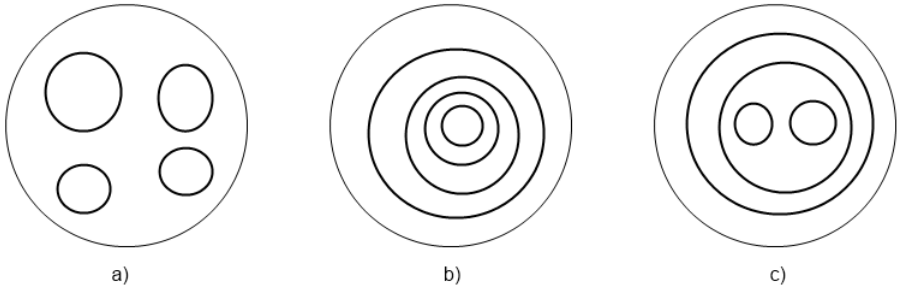


Рис. 3: Четыре окружности на сфере

Предположим, что M и N — наборы из одинакового числа окружностей на криволинейных сферах S и T. Тогда M *расположено в* S *так же, как* N *в* T, если есть биекция между связными компонентами дополнений $S - M$ и $T - N$, при которой две связные компоненты дополнения $S - M$ соседние тогда и только тогда, когда две соответствующие связные компоненты $T - N$ соседние. (Или, что то же самое, если пары (S, M) и (T, N) кусочно-линейно гомеоморфны.)

1.3. (ij), $i, j \in \{a, b, c\}$. Нарисуйте пару криволинейных сфер в трехмерном пространстве, пересекающихся по объединению четырех непересекающихся окружностей так, что в одной сфере эти окружности расположены как на рис. 3.i, а в другой — как на 3.j.

В этом тексте мы изучим две следующие проблемы и их обобщения. (Вполне возможно, вы в данный момент не сможете их решить, так что лучше отложите их и порешайте другие задачи.)

Криволинейные сферы пересекаются *трансверсально*, если в окрестности любой точки пересечения оно выглядит как пересечение двух плоскостей по прямой. (Строгое определение дано ниже в тексте.)

1.4. (a) **Проблема Ландо.** Пусть M и N — два набора одинакового числа окружностей на сфере. Существует ли пара криволинейных сфер в трехмерном пространстве, которые трансверсально пересекаются по конечному наюору непересекающихся окружностей, расположенных

в одной сфере как M, а в другой — как в N?

(b) Существует ли алгоритм, проверяющий существование таких двух криволинейных сфер? (См. „Связь с графами“ниже.)

(c)* Существует ли такой полиномиальный алгоритм?

На рис. 1, слева, две сферы пересекаются по окружности; каждая сфера делится окружностью на 2 связных компоненты, каждая из которых имеет одну соседнюю с ней компоненту (в той же сфере). На рис. 1, справа, (и на рис. 2 слева) две криволинейные сферы пересекаются по двум окружностям; каждая из сфер делится окружностями на 3 связные компоненты, из которых две имеют 1 соседнюю компоненту, а одна — 2 соседние компоненты. На рис. 2, справа, две криволинейные сферы пересекаются по трем окружностям; каждая сфера делится окружностями на 4 связные компоненты; в одной сфере количества соседей у компонент равны $3, 1, 1, 1$, а в другой — $1, 2, 2, 1$.

1.5. Проблема соседственных последовательностей. Даны последовательности $\vec{x} = (x_1, x_2, \ldots, x_n), \vec{y} = (y_1, y_2, \ldots, y_n)$ положительных целых чисел. Существуют ли две криволинейные сферы S и T в трехмерном пространстве, пересечение которых состоит из $n-1$ окружностей и делит

• S на n связных компонент, которые могут быть так пронумерованы, что у i-ой связной компоненты x_i соседних связных компонент в S, и

• T на n связных компонент, которые могут быть так пронумерованы, что у i-ой связной компоненты y_i соседних связных компонент в T?

Некоторые строгие определения.

Напомним, что Вы сможете решить все предложенные задачи фактически не используя приводимые строгие определения. Мы приводим удобные в рамках данного текста определения; они могут отличаться от общепринятых математический терминов.

В этом тексте *криволинейной окружностью*, или, сокращенно, *окружностью*, называется замкнутая несамопересекающаяся ломаная в трехмерном пространстве. Определение многогранника (несамопересекающегося, но не обязательно выпуклого) приведено в [D], см. также [W]. *Криволинейной сферой* называется многогранник в трехмерном пространстве (точнее, его двумерная поверхность), который разбивается на части любой лежащей на нем окружностью. См. рис. 1. (Такие многогранники называются топологически эквивалентными сфере. Это условие эквивалентно условию $V - E + F = 2$.)

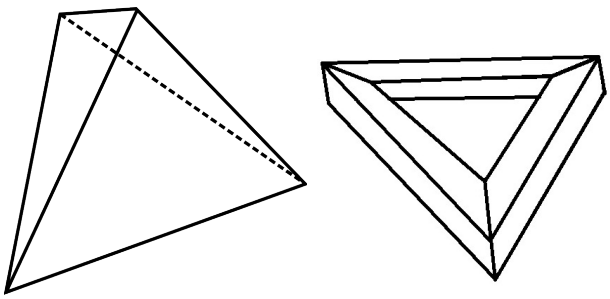


Рис. 4: Криволинейная сфера (слева) и фигура, не являющаяся криволинейной сферой (справа)

Для упрощения рисунков будем изображать вместо многогранника ‘близкую’ к нему *криволинейную* поверхность. Например, криволинейную сферу или ‘сосиску’, как на рисунке 2 справа. Вместо замкнутой ломаной будем изображать — ‘близкую’ к ней замкнутую *кривую*.

Подмножество X трехмерного пространства назовем *связным*, если любые две его точки можно соединить ломаной, лежащей в X. *Компонентой связности* или *связной компонентой* подмножества Y трехмерного пространства назовем максимальное связное подмножество в X,

т.е. такое связное подмножество $Y \subset X$, что не существует связного подмножества Z, для которого $Y \subset Z \subset X$ и $Y \neq Z \neq X$.

Пусть M — объединение непересекающихся окружностей в криволинейной сфере S. Две связные компоненты дополнения $S-M$ называются *соседями*, если их замыкания пересекаются.

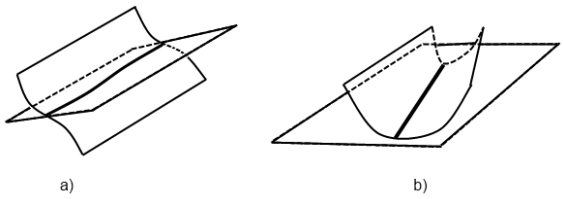


Рис. 5: Трансверсальное (слева) и не трансверсальное (справа) пересечения

Обозначим через $B(x,r) \subset \mathbb{R}^3$ шар радиуса r с центром в x. Пересечение двух криволинейных сфер $S, T \subset \mathbb{R}^3$ *трансверсально*, если для любой точки $x \in S \cap T$ существует $r > 0$ такое, что и $B(x,r) - S$, и $B(x,r) \cap (T - S)$ состоят из двух связных компонент, а каждая компонента $B(x,r) - S$ содержит компоненту $B(x,r) \cap (T - S)$.

Вы можете использовать следующую теорему и ее следствие без доказательства.

Теорема Жордана. *Криволинейная сфера делит трехмерное пространство ровно на две части. Две точки пространства, не принадлежащие сфере, лежат в одной части тогда и только тогда, когда их можно соединить ломаной, не пересекающей криволинейной сферы.*

Следствие. *Пусть S и T — криволинейные сферы, трансверсально пересекающиеся по конечному набору $S \cap T$ непересекающихся окружностей. Обозначим через B внутренность сферы S (точнее, ограниченную часть дополнения $\mathbb{R}^3 - S$). Пусть Q — связная компонента дополнения $T - S$, расположенная внутри S. Тогда Q делит B на две части. Две точки шара B, не лежащие на Q, лежат в одной части тогда и только тогда, когда их можно соединить ломаной, непересекающей Q.*

Связь с графами.

Пусть M — объединение непересекающихся окружностей в криволинейной сфере S. Определим (двойственный к M) граф $G = G(S, M)$ следующим образом. Вершины — связные компоненты $S - M$. Вершины соединяются ребром, если соответствующие компоненты — соседи.

На рис. 6 показаны графы для сфер S, T с рис. 2 справа и набора окружностей $S \cap T$. Аналогичным образом две криволинейные сферы, пересекающиеся по набору окружностей, определяют пару графов. Проблема Ландо требует описать получающиеся таким образом пары графов, а проблема соседственных последовательностей — получающиеся наборы степеней их вершин. [3]

Поблажки.

Команда получает „поблажку“ за каждое правильно ($\geq +.$) записанное решение. *Компьютерная программа, проходящая тесты, предложенные жюри, и имеющая понятную структуру расценивается как записанное решение. Большая понятная членам жюри картинка расценивается как записанное построение примера.* Жюри также может награждать „поблажками“ за красивые решения, решения сложных задач и за (некоторые) решения, записанные в

[3] Вот другая интерпретация, предложенная И.Н. Шнурниковым. Даны единичный квадрат на плоскости и (кусочно-линейная) функция на нем, строго положительная на границе квадрата. Диск соответствует одной криволинейной сфере (с проколом), график функции (над диском) — другой криволинейной сфере, множество нулей функции — пересечению криволинейных сфер.

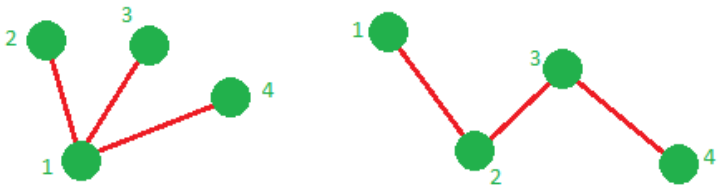


Рис. 6: Два графа для сфер с рис. 2 справа

TEX-е. „Поблажек“ у жюри бесконечно много. Можно также сдавать задачи устно, платя по „поблажке“ за каждую попытку.

Мы приглашаем школьников, успешно продвинувшихся в решении задач и работающих над задачами для исследования, консультироваться по поводу возникающих вопросов и идей решения.

2 Проблема соседственных последовательностей

Пара $\vec{x} = (x_1, x_2, \ldots, x_n), \vec{y} = (y_1, y_2, \ldots, y_n)$ последовательностей положительных целых чисел называется *реализуемой*, если существуют две криволинейные сферы S и T в трехмерном пространстве, пересечение которых состоит из $n-1$ окружностей и делит

• S на n связных компонент, которые могут быть так занумерованы, что у i-ой связной компоненты x_i соседних связных компонент в S, и

• T на n связных компонент, которые могут быть так занумерованы, что у i-ой связной компоненты y_i соседних связных компонент в T?

Пара (S, T) сфер называется *реализацией* пары $(\vec{x}, \vec{y})$.

2.1. (n), $n \in \{2, 3, 4, 5\}$. Какие пары последовательностей из n положительных целых чисел реализуемы?

2.2. (a) Если пара $(\vec{x}, \vec{y})$ последовательностей реализуема, то $x_1 + x_2 + \ldots x_n = y_1 + y_2 + \cdots + y_n = 2n - 2$.

(b) Двойственный граф $G(S, M)$ к набору M непересекающихся окружностей в криволинейной сфере S является деревом.

Последовательность $\vec{x} = (x_1, x_2, \ldots, x_n)$ положительных целых чисел называется *деревянной*, если $x_1 + x_2 + \ldots x_n = 2n - 2$.

2.3. Если последовательность $\vec{x}$ деревянная, то в ней не менее x_1 единиц.

2.4. Пара $(\vec{x}, \vec{x})$ реализуема для любой деревянной последовательности $\vec{x}$.

2.5. Пусть $\vec{x}, \vec{y}$ — деревянные последовательности, в которых все единицы находятся в конце, и $x_1 \geq y_1$. Тогда последовательности $\vec{x}' := (x_1 - y_1 + 1, x_2, x_3, \ldots, x_{n-y_1+1})$ и $\vec{y}' = (y_2, y_3, \ldots, y_{n-y_1+2})$ — деревянные.

2.6. Какие пары деревянных последовательностей можно получить из $((1,1),(1,1))$ с помощью перестановок и замен пары векторов ($\vec{x} = (x_1, x_2, \ldots, x_n), \vec{y} = (y_1, y_2, \ldots, y_n)$) на пару:

(a) ($\vec{x'} = (a, x_1, x_2, \ldots, x_n, 1, 1, \ldots, 1)$, $\vec{y'} = (y_1 + a - 1, y_2, y_3, \ldots, y_n, 1, 1, \ldots, 1)$) (число новых единиц равно $a-2$ для $\vec{x}'$, и $a-1$ для y'; здесь a может быть различно для различных замен: $((1,1),(1,1)) \overset{a=3}{\to} ((3,1,1,1),(3,1,1,1)) \overset{a=4}{\to} ((4,3,1,1,1,1,1),(6,1,1,1,1,1,1))$).

(b) ($\vec{x'} = (x_1 + 1, x_2, x_3, \ldots, x_n, 1)$, $\vec{y'} = (y_1 + 1, y_2, y_3, \ldots, y_n, 1)$).

3 Проблема Ландо

Пара (M, N) наборов из одинакового числа непересекающихся окружностей на сфере, называется *реализуемой*, если существует пара криволинейных сфер в трехмерном пространстве, которые трансверсально пересекаются по конечному объединению непересекающихся окружностей, которые расположены в одной сфере так же, как M, а в другой — как в N.

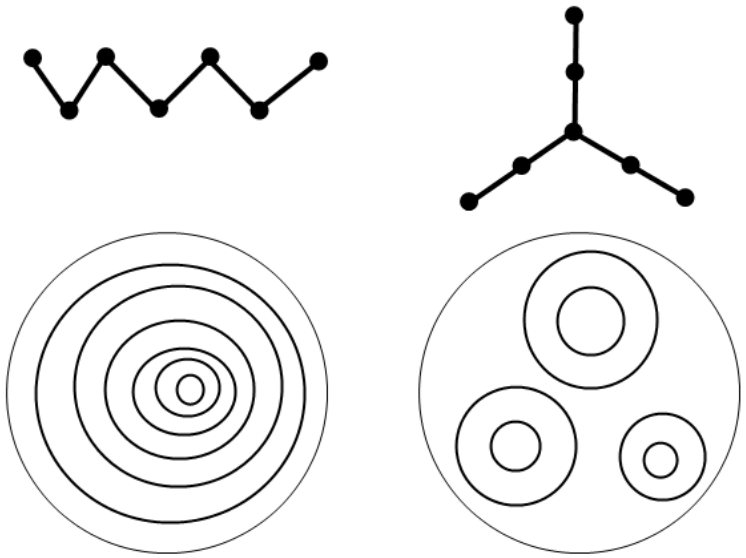


Рис. 7: Реализуема ли эта пара?

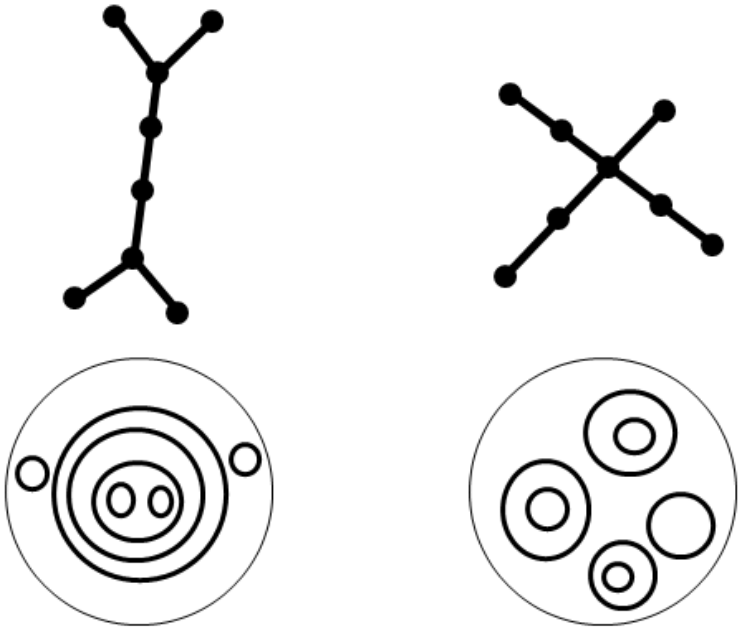


Рис. 8: Реализуема ли эта пара?

3.1. (a) Любая пара наборов, каждый из которых состоит из $n \leq 4$ непересекающихся окружностей, реализуема.

(b) Является ли пара на рис. 7 реализуемой? Один граф — цепочка из 6 ребер, другой — триод с 'лучами' из двух ребер.

(c) Является ли пара на рис. 8 реализуемой? Один граф — звезда с 4 'лучами', три 'луча' состоят из двух ребер, а один 'луч' — из одного ребра. Другой граф представляет собой букву 'H' с 'горизонтальной перекладиной' из трех ребер.

(d) Существует нереализуемая пара из двух наборов из одинакового числа непересекающихся окружностей.

3.2. Даны целые числа n и k.

(a) Какие наборы непересекающихся окружностей реализуемы в паре с набором n окружностей, ограничивающих n непересекающихся дисков? (Или, что то же самое, какие графы реализуемы в паре со звездой с n лучами?)

(b) Какие графы реализуемы в паре с графом, представляющим собой объединение по одному общему ребру звезды с n лучами и звезды с k лучами?

(c) * Какие наборы непересекающихся окружностей реализуемы в паре с набором n 'параллельных' окружностей? (Или, что то же самое, какие графы реализуемы в паре с цепочкой длины $n+1$?)

3.3. Пусть S и T — криволинейные сферы, трансверсально пересекающиеся по конечному набору $S \cap T$ непересекающихся окружностей. Тогда связные компоненты $S - T$ можно покрасить в черный и белый цвета так, чтобы любые две компоненты одного цвета не были соседями.

До конца этого параграфа M, N будут объединениями одинакового числа непересекающихся окружностей на криволинейных сферах S, T (Ни M, ни N не обязано совпадать с $S \cap T$.)

Для связной компоненты P множества $S - M$ обозначим за ∂P объединение граничных (краевых) окружностей компоненты P. Легко видеть, что связные компоненты P и Q дополнения $S - M$ являются соседями в том и только том случае, когда $\partial P \cap \partial Q \neq \emptyset$.

3.4. Незацепленные семейства окружностей.

Пусть S и T — криволинейные сферы, трансверсально пересекающиеся по конечному набору $S \cap T$ непересекающихся окружностей. Пусть P и Q — две связные компоненты дополнения $S - T$, расположенные внутри T.

(a) Если Q является криволинейным диском (то есть, если граница компоненты Q состоит из одной окружности), то ∂P находится целиком внутри одной компоненты дополнения $T - \partial Q$.

(b) Если Q является криволинейным цилиндром (то есть, если граница компоненты Q состоит из двух окружностей), то ∂P целиком содержится либо в кольцевой компоненте дополнения $T - \partial Q$ (то есть, в компоненте с двумя граничными окружностями), либо в объединении двух дисковых (то есть имеющих одну граничную окружность) компонент дополнения $T - \partial Q$.

(c) Покрасим связные компоненты дополнения $T - \partial Q$ в черный и белый цвета так, чтобы любые две соседние компоненты оказались разного цвета. Тогда ∂P целиком содержится в объединении одинаково окрашенных компонент дополнения $T - \partial Q$.

Знак $\sqcup$ обозначает объединение непересекающихся множеств.

3.5. Пусть S и T такие криволинейные сферы, что $S \cap T$ расположено на S как на рис. 9. Обозначим за A_i 'внешние' окружности, за B 'большую разделяющую' окружность, а за C — объединение 'внутренних' окружностей, см. рис. 9.

(a) Для каждого i объединение $B \cup C$ лежит по одну сторону от A_i в T.

(b) Объединение $B \cup C$ лежит в одной связной компоненте $T - \sqcup_i A_i$.

3.6. Проблема продолжения вложения. (a) Любые две непересекающиеся окружности в единичной сфере являются границами двух непересекающихся криволинейных дисков внутри сферы.

(b) Для каких трех непересекающихся окружностей p, q_1, q_2 на единичной сфере существуют непересекающиеся *криволинейный диск* P и *криволинейный цилиндр* Q внутри этой сферы, для которых $\partial P = p$ и $\partial Q = q_1 \sqcup q_2$? (Рис. 10 слева.)

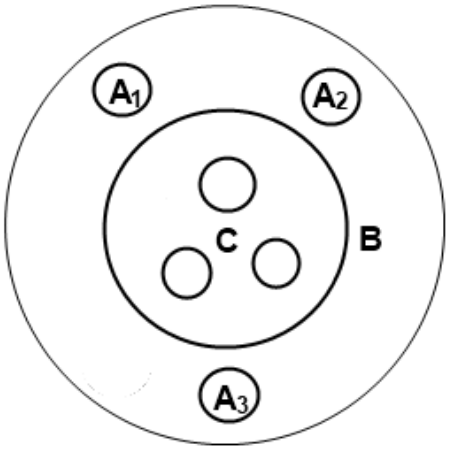


Рис. 9: Пересечение $S \cap T$ на сфере S

Рис. 10: Непересекающиеся криволинейный диск и криволинейный цилиндр вне шара (слева), непересекающиеся криволинейные цилиндры, один из которых заузлен, вне шара (справа)

(c) Для каких четырех непересекающихся окружностей p_1, p_2, q_1, q_2 на единичной сфере существуют непересекающиеся криволинейные цилиндры P и Q внутри этой сферы, для которых $\partial P = p_1 \sqcup p_2$ и $\partial Q = q_1 \sqcup q_2$? (Рис. 10 справа.)

(d) Для каких двух непересекающихся семейств p, q непересекающихся окружностей на единичной сфере существуют непересекающиеся криволинейные *сферы с дырками* P и Q внутри этой сферы, для которых $\partial P = p$ и $\partial Q = q$?

(e) Существуют ли три непересекающихся семейства p, q, r непересекающихся окружностей на единичной сфере, такие что

• каждая из трех пар (p, q), (q, r) и (p, r) может быть продолжена внутрь сферы (до непересекающихся криволинейных сфер с дырками) в смысле пункта (d);

• не существует непересекающихся криволинейных сфер с дырками P, Q и R внутри этой сферы, для которых $\partial P = p$, $\partial Q = q$ и $\partial R = r$? [4]

(f) Для каких m непересекающихся семейств $p_1, \ldots, p_m$ непересекающихся окружностей на единичной сфере существуют непересекающиеся криволинейные сферы с дырками $P_1, \ldots, P_m$ внутри этой сферы, такие что $\partial P_i = p_i$ для всех $i = 1, \ldots, m$?

Пусть S и T — криволинейные сферы, для которых все кроме одной компоненты $S - T$ имеют по одному соседу. (Оставшаяся компонента может иметь одного или более соседей.) Эта оставшаяся компонента называется *криволинейной сферой с дырками*. *Криволинейный диск* — это криволинейная сфера с 1 дыркой (= с одним соседом). *Криволинейный цилиндр* — это криволинейная сфера с 2 дырками (= с двумя соседями).

Пусть M и N — два набора непересекающихся окружностей в единичной сфере S. Покрасим связные компоненты $S - N$ в черный и белый цвета так, чтобы соседние компоненты были разных цветов. Набор M *лежит по одну сторону* (в сфере) от N, если M содержится в наборе компонент $S - N$, одинаково покрашенных. Наборы M и N *не зацеплены* (в сфере), если M лежит по одну сторону от N и N лежит по одну сторону от M. См. рис. 11.

3.7. (a) Существует два набора M и N непересекающихся окружностей в сфере, такие что M лежит по одну сторону от N, но N не лежит по одну сторону от M.

(b) Верно ли, что если пары M, N и N, P наборов окружностей на криволинейной сфере не зацеплены, то не зацеплена и пара M, P, то есть, транзитивно ли отношение незацепленности?

[4] Сравните с известным примером колец Борромео.

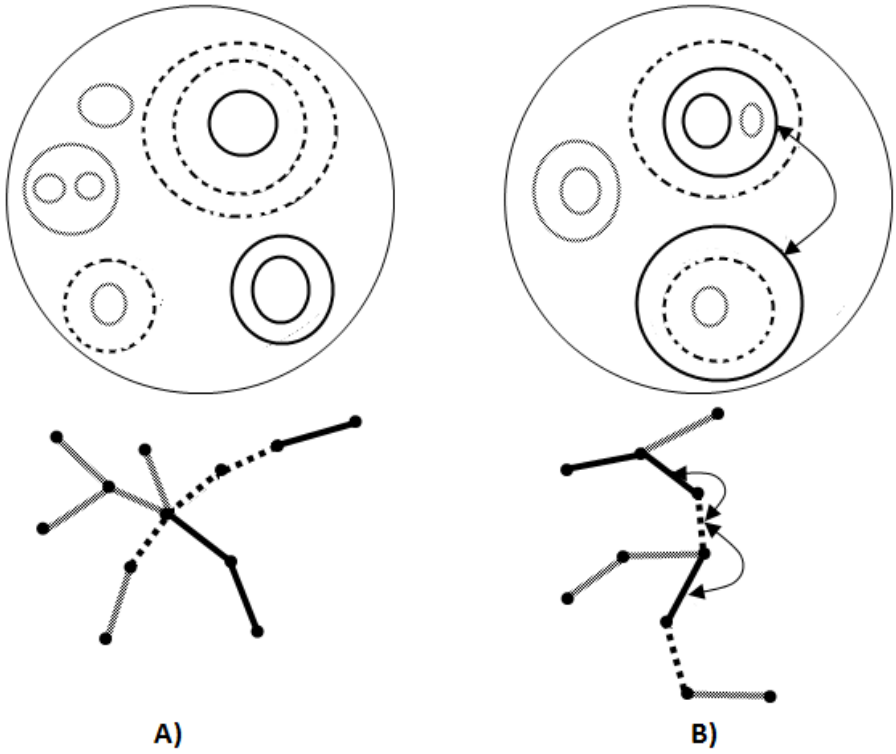


Рис. 11: (A): набор ‘пунктирных’ окружностей и набор ‘жирных’ окружностей не зацеплены (B): набор ‘пунктирных’ окружностей и набор ‘жирных’ окружностей зацеплены, поскольку выделенный стрелочками путь между двумя ‘жирными’ окружностями пересекает ‘пунктирные’ окружности по нечетному числу точек.

(c) Для набора M непересекающихся окружностей в единичной сфере S обозначим за $\mathring{M}$ объединение черных компонент связности $S - M$. два набора M и N не зацеплены если и только если для любых черно-белых раскрасок S относительно M и относительно N, таких что $\mathring{M} \cup \mathring{N} \neq S$, будет либо $\mathring{M} \subset \mathring{N}$, (Есть два способа выбрать $\mathring{p}$ для данного p; один из них — дополнение другого.) либо $\mathring{N} \subset \mathring{M}$, либо $\mathring{M} \cap \mathring{N} = \emptyset$.

4 Промежуточный финиш. Некоторые решения и новые задачи

1.1 и 1.2. Аналогично решению задачи 2.4.

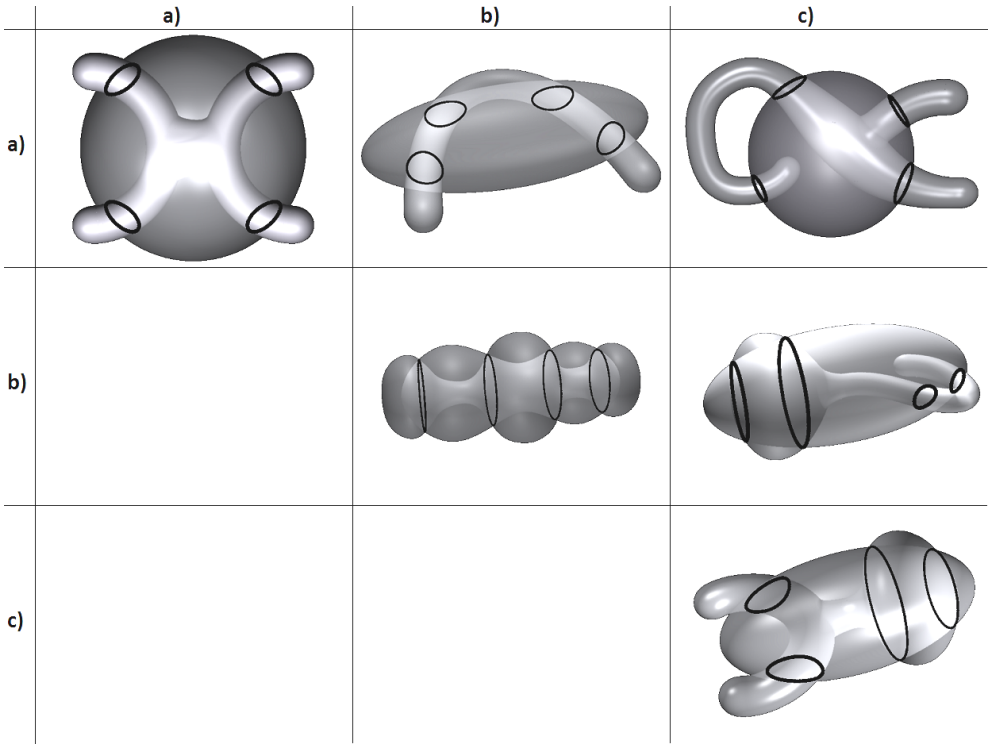


Рис. 12: К решению задачи 1.3.

1.3. Случай $i = j$ выводится аналогично задаче 2.4. Случаи ab, ac, bc см. на рис. 12.

1.4. (a) Ответ содержится в решении задачи 4.5.

(b) Такой алгоритм дается в ответе к 4.5. (Очевидно, он не полиномиален.)

1.5. Теорема 1. *Пусть n — положительное целое число, а $\vec{x} = (x_1, x_2, \dots, x_n)$, $\vec{y} = (y_1, y_2, \dots, y_n)$ — последовательности положительных целых чисел. Тогда условие $x_1 + \dots + x_n = y_1 + \dots + y_n = 2n - 2$ равносильно существованию криволинейных сфер S и T, пересечение которых состоит из $n - 1$ окружностей и делит*

- *S на n связных компонент, которые могут быть занумерованы, так что у i-ой связной компоненты x_i соседей в S, и*
- *T на n связных компонент, которые могут быть занумерованы, так что у i-ой связной компоненты y_i соседей в T.*

Это следует из Теоремы 1' (задача 4.3) ниже.

2.1. *Ответ.* Пары $(\vec{x}, \vec{x})$ реализуемы для $\vec{x}$, равного с точностью до перестановок одной из пар

$$(1,1), (2,1,1), (3,1,1,1), (2,2,1,1), (4,1,1,1,1), (3,2,1,1,1), (2,2,2,1,1).$$

Другие реализуемые пары равны с точностью до перестановок парам двух последовательностей равной длины из приведенного списка.

2.2. (a) Предположим, что криволинейные сферы S, T реализуют пару $(\vec{x}, \vec{y})$. Вспомним определение графа $G = G(S, S \cap T)$ из параграфа 1. Количество его вершин равно n. Из k-ой исходит x_k ребер. Поэтому количество ребер равно $(x_1 + \cdots + x_n)/2$. Очевидно, что G связен. По теореме Жордана о кривой[5] G делится любым своим ребром на несколько частей. Значит G — дерево. Поэтому число ребер равно $n-1 = (x_1 + \cdots + x_n)/2$. Аналогично $n-1 = (y_1 + \cdots + y_n)/2$.

Набросок альтернативного решения (a) Т. Новика. Индукция по числу окружностей. Утверждение верно для одной окружности (на каждой сфере только 2 диска, значит $n = 2$). Каждая следующая окружность делит одну из связных компонент на две части, и добавляет к ним по одной граничной окружности.

(b) Очевидно, G связен. По теореме Жордана о кривой G разделяется любым ребром. Значит, G — дерево.

2.3. Если количество единиц равно s, то $2n - 2 = x_1 + \cdots + x_n \geq x_1 + 2(n - 1 - s) + s = 2n - 2 + x_1 - s$. Значит, $s \geq x_1$.

2.4. Пусть S — единичный куб. Рассмотрим семейство M окружностей на S, 'реализующее' $\vec{x}$. (Существование такого семейства доказывается индукцией. При переходе используется удаление висячей вершины.) Покрасим в черный и белый цвета дополнение в S к этим окружностям так, что соседние компоненты будут разного цвета. Рассмотрим близкую к S криволинейную сферу T такую, что $S \cap T = M$, каждая черная компонента T лежит внутри S, и каждая белая компонента T лежит снаружи S. Тогда S и T реализуют $(\vec{x}, \vec{x})$.

2.5. Согласно задаче 2.3 $x_1 \leq s$. Тогда

$$x_{n-y_1+1} = x_{n-y_1+2} = \cdots = x_{n-y_1+1} = \cdots = x_n = y_{n-y_1+1} = y_{n-y_1+2} = \cdots = y_n = 1.$$

$$\text{Следовательно} \quad (\sum_{i=1}^{n-y_1+1} x_i) - y_1 + 1 = (\sum_{i=1}^{n} x_i) - y_1 + 1 - (y_1 - 1) = 2(n - y_1 + 1) - 2$$

$$\text{и} \quad (\sum_{i=2}^{n-y_1+2} y_i) = (\sum_{i=1}^{n} y_i) - y_1 - (y_1 - 2) = 2(n - y_1 + 1) - 2.$$

Значит новые последовательности деревянные.

2.6. *Ответ:* любая пара.

(a) Индукция по n. База индукции при $n = 2$ очевидна. Предположим, утверждение верно для всех $n < k$, докажем его для $n = k$.

Если последовательности $\vec{u} = (u_1, \ldots, u_k)$ и $\vec{v} = (v_1, \ldots, v_k)$ получаются друг из друга перестановками, то пара $(\vec{u}, \vec{v})$ реализуема по задаче 2.4. Если последовательности $\vec{u} = (u_1, \ldots, u_k)$ и $\vec{v} = (v_1, \ldots, v_k)$ не получаются друг из друга перестановками, можно так переупорядочить последовательности, что $u_1 \neq v_1$ и все единицы находятся в концах последовательностей. Без ограничения общности $u_1 < v_1$. Обозначим

$$a := u_1, \quad n := k - u_1 + 1, \quad x_i = u_{i+1} \text{ для } i = 1, \ldots, n, \quad y_1 := v_1 - u_1 + 1, \quad y_i = v_i \text{ для } i = 2, \ldots, n.$$

Согласно задаче 2.3 $u_i = 1$ для всех $i \geq n + 2$ и $v_i = 1$ для всех $i \geq n + 1$. Поэтому пара $(\vec{u}, \vec{v})$ получается из пары $(\vec{x}, \vec{y})$ применением данного в задаче преобразования. Имеем

$$x_1 + \cdots + x_n = u_2 + \cdots + u_{n+1} = 2k - 2 - u_1 - (u_1 - 2) = 2n - 2 \quad \text{и}$$

[5] **Теорема Жордана о кривой.** *Окружность на сфере делит сферу ровно на две части. Две точки сферы, не лежащие на этой окружности, лежат в одной части тогда и только тогда, когда их можно соединить некоторой сферической ломаной, не пересекающей окружности.*

$$y_1 + \cdots + y_n = v_1 + \cdots + v_n - (u_1 - 1) = 2k - 2 - (u_1 - 1) - (u_1 - 1) = 2n - 2.$$

Значит последовательности $\vec{x}$ и $\vec{y}$ деревянные. Из $u_1 > 1$ имеем $n < k$. Шаг индукции доказан.

(b) Индукция по n.

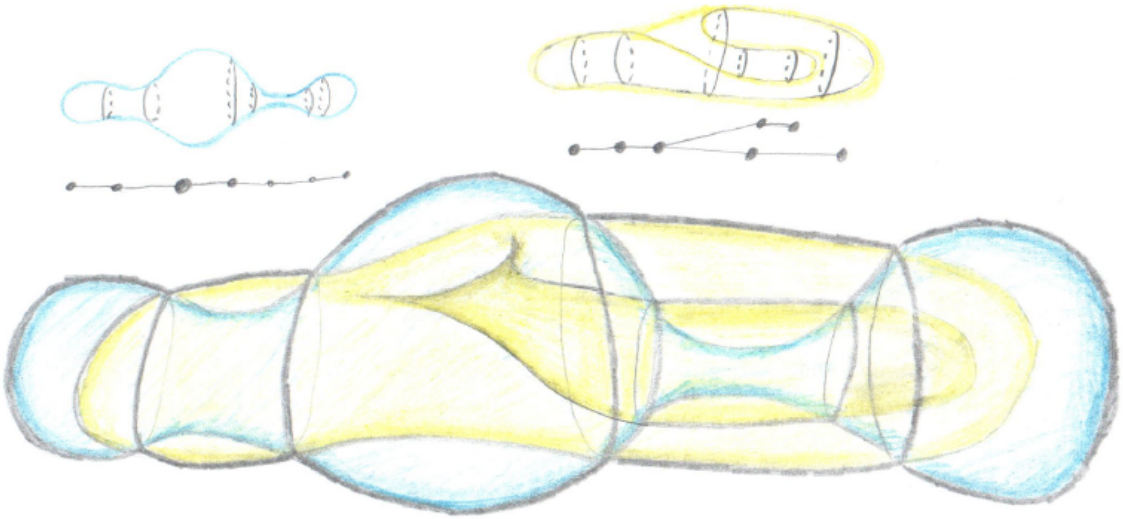


Рис. 13: Криволинейные сферы, реализующие пару с рисунка 7.

3.1. (a) Следует из задач 1.1 и 1.3.

(b) Да, см. рис. 13. Альтернативная конструкция. Пусть S и T' — криволинейные сферы с рисунка 2. Пусть T'' — сфера внутри T' 'близкая и параллельная' T'. Возьмем компоненту связности X дополнения $\mathbb{R}^3 - S - T'$ лежащую внутри T', такую что ее граница содержит связную компоненту $T' - S$, являющуюся диском. Пусть T — криволинейная сфера, полученная путем соединения T' и T'' тонкой трубкой в X. Тогда S и T будут требуемыми.

(c) Нет. Этот факт получен с помощью компьютерной программы, основанной на решении задачи 4.5.

(d) Контрпример изображен на рис. 14. Подсказка: используйте задачу 3.5 (или задачу 3.4 в форме 4.7 ниже).

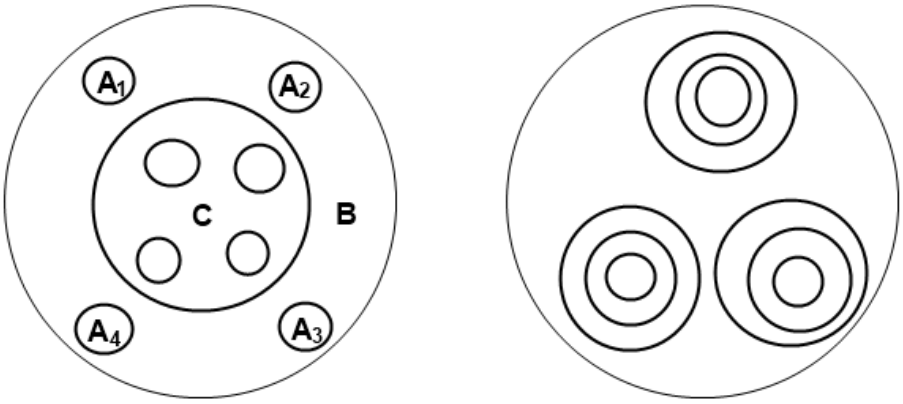


Рис. 14: девять окружностей (жирные), расположенные на сфере (тонкая) двумя разными способами (слева, справа).

3.2. (a) *Ответ:* Все наборы. *Подсказка.* Возьмем n непересекающихся сфер, пересекающих данную сферу S по n окружностям данного набора M. Соединим их $n - 1$ непересекающими-

ся трубками внутри S и получим криволинейную сферу T. Проверьте, что T удовлетворяет условию задачи.

(b) *Гипотеза.* Пара из такого графа и дерева реализуема тогда и только тогда, когда это дерево является объединением двух деревьев из n и k ребер, пересекающихся ровно по одному ребру.

(c) *Гипотеза.* Любой набор n окружностей реализуем в паре с набором n 'параллельных' окружностей.

3.3. Покрасим связные компоненты, лежащие внутри и вне T, в черный и белый цвета, соответственно.

3.4. (a) и (b) интуитивно очевидны и следуют из (c).

(c) Будем считать, что T — круглая сфера и окружности из ∂Q — круглые окружности, что никакая из них не является большим кругом. Для каждой окружности из ∂Q возьмем сферу, проходящую через эту окружность и центр сферы T. Объединение Q и частей таких сфер, лежащих снаружи T — криволинейная сфера, назовем ее Q'. Эта сфера делит $\mathbb{R}^3$ на две компоненты связности. Поскольку Q связно, пересечение обеих компонент связности с внутренностью T связно. Эти компоненты пересекают T по черным и белым частям соответственно. Поскольку P лежит в одной компоненте, то ∂P лежит либо в черной, либо в белой компоненте $T - \partial Q$.

3.5. (a) делается аналогично (b).

(b) Рассмотрим диски $\overline{A_1}, \overline{A_2}, \overline{A_3} \subset S$, ограниченные окружностями $A_1, A_2, A_3 \subset S$ и не содержащие других окружностей. Без ограничения общности считаем, что внутренности этих дисков лежат внутри T. Тогда внутренности компоненты дополнения $S - M$, ограниченной $B \cup C$, тоже лежит внутри T (поскольку пересечение $S \cap T$ трансверсально). Эта внутренность лежит в одной из компонент связности дополнения $\mathbb{R}^3 - (T \cup \overline{A_1} \cup \overline{A_2} \cup \overline{A_3})$. Значит, все 4 окружности объединения $B \cup C$ лежат в одной компоненте связности дополнения $T - (A_1 \cup A_2 \cup A_3)$.

Заметим, что это частный случай задачи 3.4.c (сравните с задачей 4.6.a).

3.7. (a) См. определение окружностей A_-, A_0, A_+ на рисунке 15. Семейство $\{A_0\}$ лежит по одну сторону от семейства $\{A_+, A_-\}$, но не наоборот.

(b) Нет.

Некоторые новые задачи по проблеме соседственных последовательностей.

Пара $\vec{x} = (x_1, x_2, \ldots, x_n)$, $\vec{y} = (y_1, y_2, \ldots, y_n)$ последовательностей положительных целых чисел называется *сильно реализуемой*, если существуют две такие криволинейные сферы S, T, что

(1) их пересечение состоит из $n-1$ окружностей и делит

- S на n связных компонент, которые можно занумеровать так, что у i-ой связной компоненты x_i соседей в S, играницей
- T на n связных компонент, которые можно занумеровать так, что у i-ой связной компоненты y_i соседей в T;

(2) в $S \cap T$ существует окружность, являющаяся границей некоторых диска и компоненты с x_1 соседями в $S - T$, а также границей некоторых диска и компоненты с y_1 соседями в $T - S$.

Пара (S, T) криволинейных сфер называется *сильной реализацией* пары $(\vec{x}, \vec{y})$.

4.1. Пусть $\vec{x}, \vec{y}$ — деревянные последовательности, у которых все единицы находятся в конце. Если пара последовательностей $\vec{x'} := (x_1 - y_1 + 1, x_2, x_3, \ldots, x_{n-y_1+1})$, $\vec{y'} := (y_2, y_3, \ldots, y_{n-y_1+2})$ сильно реализуема, то и пара $(\vec{x}, \vec{y})$ тоже.

4.2. Если пара $(\vec{x}, \vec{y})$ сильно реализуема, то для любого целого положительного a пары $(\vec{x'}, \vec{y'})$ сильно реализуемы для

(a) $\vec{x'} = (a, x_1, x_2, \ldots, x_n, 1, 1, \ldots, 1)$, $\vec{y'} = (y_1 + a - 1, y_2, y_3, \ldots, y_n, 1, 1, \ldots, 1)$ (в $\vec{y'}$ $a-1$ новых единиц, в $\vec{x'}$ их $a-2$; здесь a может быть различно для различных замен: $((1,1),(1,1)) \overset{a=3}{\to} ((3,1,1,1),(3,1,1,1)) \overset{a=4}{\to} ((4,3,1,1,1,1,1),(6,1,1,1,1,1,1))$).

(b) какого-либо выбора $\vec{x'} \in \{(1, x_1 + 1, x_2, x_3, \ldots, x_n), (x_1 + 1, x_2, x_3, \ldots, x_n, 1)\}$ и $\vec{y'} \in \{(1, y_1 + 1, y_2, y_3, \ldots, y_n), (y_1 + 1, y_2, y_3, \ldots, y_n, 1)\}$.

4.3. Пара последовательностей сильно реализуема тогда и только тогда, когда обе последовательности деревянные.

Некоторые новые задачи по проблеме Ландо.

4.4. Каждая пара объединений из
(5) 5; (6) 6;
непересекающихся окружностей, реализуема.

Проблема Ландо решается с помощью ее *занумерованного*, или *раскрашенного*, аналога. Дадим определения, необходимые для его формулировки.

В этом пункте M и N — два объединения непересекающихся окружностей на сферах S и T, соответственно, причем в каждом объединении окружности занумерованы числами $1, 2, \ldots, n$.

Пара (M, N) называется *реализуемой*, если существуют
(1) две криволинейные сферы S' и T', трансверсально пересекающиеся по конечному набору $S' \cap T'$ непересекающихся окружностей, и
(2) нумерация этих окружностей, для которой
- $S' \cap T'$ на сфере S' и M на сфере S занумерованно одинаковы;
- $S' \cap T'$ на сфере T' и N на сфере T занумерованно одинаковы.

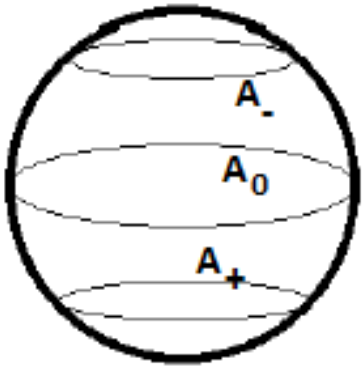


Рис. 15: К примеру

(Вы, по-видимому, сможете решить все задачи, не используя следующего строгого определения занумерованноодинаковости. Занумерованные наборы M на криволинейной сфере S и N на криволинейной сфере T называются *занумерованно одинаковыми*, если существует взаимно однозначное соответствие между связными компонентами дополнений $S - M$ и $T - N$, при котором две связные компоненты дополнения $S - M$ граничат по окружности из M тогда и только тогда, когда две соответствующие компоненты дополнения $T - N$ граничат по соответствующей окружности из N.)

Пример. На единичной сфере (или на поверхности Земли) возьмем экватор $A_0 = A_2$, параллель $A_+ = A_1$ шестидесяти градусов северной широты, параллель $A_- = A_3$ шестидесяти градусов южной широты. См. рисунок 15. Тогда
- незанумерованные (или неупорядоченные) наборы $\{A_-, A_0, A_+\}$ и $\{A_0, A_-, A_+\}$ одинаковы.

• занумерованные (или упорядоченные) наборы (A_-, A_0, A_+) и (A_+, A_0, A_-) занумерованно одинаковы.

• занумерованные наборы (A_-, A_0, A_+) и (A_0, A_-, A_+) не являются занумерованно одинаковыми.

• пара занумерованных наборов (A_-, A_0, A_+) и (A_0, A_-, A_+) не реализуема.

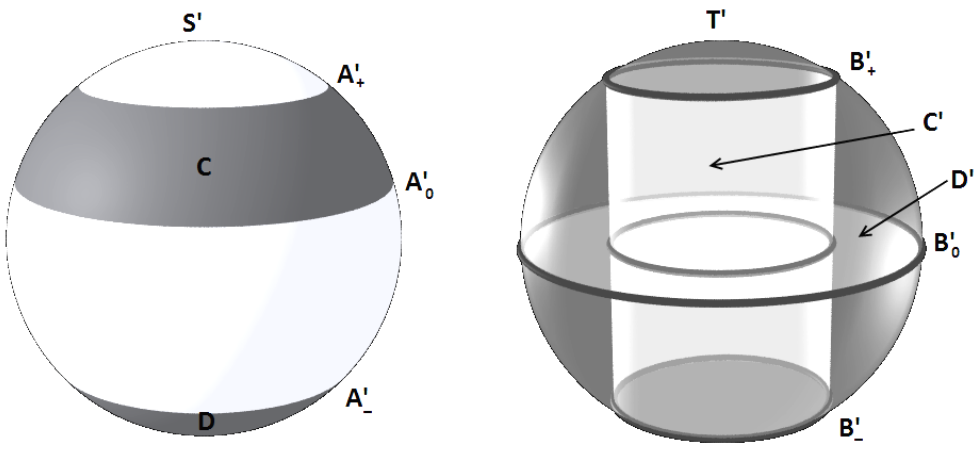


Рис. 16: Растащенные криволинейные сферы S' и T'

Доказательство последнего утверждения. Пусть, напротив, существуют криволинейные сферы S' и T', реализующие эту пару. Обозначим через

• B_k копию окружности A_k на копии T сферы S;

• A'_k окружность на криволинейной сфере S' соответствующую A_k;

• B'_k окружность на криволинейной сфере T' соответствующую B_k;

• $D' \subset S'$ диск в $S' - T'$ ограниченный A'_+;

• $C' \subset S'$ цилиндр в $S' - T'$ ограниченный A'_0 и A'_-.

Поскольку S' и T' реализуют пару $(A_+, A_0, A_-), (B_0, B_+, B_-)$, получаем $A'_+ = B'_0$, $A'_0 = B'_+$ и $A'_- = B'_-$.

Ясно, что C' и D' лежат в трехмерном пространстве по одну сторону от криволинейной сферы T'. (Сравните с задачей 3.3.) Имеем $\partial D = A'_+ = B'_0$. Граница $\partial C' = A'_0 \sqcup A'_- = B'_+ \sqcup B'_-$ не лежит в одной компоненте дополнения $T' - \partial D' = T' - B'_0$. Это противоречит утверждению задачи 3.4.a для $P = C'$ и $Q = D'$. QED

4.5. Занумерованная проблема Ландо. Какие пары занумерованных наборов на сферах реализуемы?

4.6. (a) Если пара (M, N) занумерованных наборов на сферах S и T реализуема, то связные компоненты дополнения $S - M$ можно покрасить в черный и белый цвета так, что для любых двух одноцветных компонент P и Q дополнения $S - M$ поднаборы в T, соответствующие ∂P и ∂Q, не зацеплены.

(b) Верно ли обратное к (a) утверждение?

Пусть p и q два подмножества множества ребер некоторого дерева. Множество p *лежит по одну сторону* (в этом дереве) от q, если $p \cap q = \emptyset$, и для любых двух концов ребер из p соединяющий их путь в дереве содержит четное число ребер из q. Наборы p и q *не зацеплены* (в этом дереве) если p лежит по одну сторону от q и q лежит по одну сторону от p.

Для вершины P графа обозначим через δP множество выходящих из P ребер.

Графы $G(S, M)$ и $G(T, N)$ определены в §1. Нумерация окружностей в M и окружностей в N задает нумерацию ребер в $G(S, M)$ и в $G(T, N)$.

4.7. (a) Если пара (M, N) занумерованных наборов на сферах S и T реализуема, то вершины графа $G(S, M)$ можно так покрасить в черный и белый цвета, что для любых двух одноцветных вершин P, Q графа $G(S, M)$ наборы ребер в $G(T, N)$, соответствующие δP и δQ, не зацеплены в $G(T, N)$.

(b)* Даны два дерева G и G' с одинаковым количеством ребер. Существует ли полиномиальный алгоритм для проверки наличия таких нумераций их ребер, при которой вершины графа G можно покрасить в черный и белый цвета так, что для любых двух одноцветных вершин P, Q графа G наборы ребер в G', соответствующие δP и δQ, не зацеплены в G'?

5 Больше сфер и сферы с ручками

Пусть n_1, n_2, n_3 — положительные целые числа. Тройка

$$\vec{x_1} = (x_{11}, x_{12}, \ldots, x_{1n_1}), \quad \vec{x_2} = (x_{21}, x_{22}, \ldots, x_{2n_2}), \quad \vec{x_3} = (x_{31}, x_{32}, \ldots, x_{3n_3})$$

последовательностей положительных целых чисел называется *реализуемой*, если существуют три криволинейных сферы S_1, S_2, S_3 в трехмерном пространстве, пересекающиеся попарно по окружностям так, что $S_1 \cap S_2 \cap S_3 = \varnothing$ и для всех $k = 1, 2, 3$ дополнение $S_k - S_{k+1} - S_{k+2}$ имеет n_k связных компонент, которые можно пронумеровать так, что у i-ой связной компоненты x_{ki} соседей в S_k для всех $i = 1, \ldots, n_k$.

В этом параграфе (и в соответствующих решениях) индексы $k, k+1, k+2$ рассматриваются по модулю 3.

Тройка (S_1, S_2, S_3) криволинейных сфер называется *реализацией* тройки $(\vec{x_1}, \vec{x_2}, \vec{x_3})$.

5.1. Проблема троек соседственных последовательностей. Какие тройки деревянных последовательностей реализуемы?

5.2. Какие тройки деревянных последовательностей, в каждой из которых не более четырех чисел, реализуемы?

5.3. Если тройка последовательностей длин n_1, n_2, n_3, реализуема, то
(a) $n_1 + n_2 + n_3$ нечетно;
(b) $n_k < n_{k+1} + n_{k+2}$ для всех $k = 1, 2, 3$.

5.4. Пусть $x_1 \geq x_2 \geq \cdots \geq x_n$ — деревянная последовательность, а p и q — такие положительные целые числа, что $p \geq q > 1$ и $p + q = n + 1$. Тогда существуют две такие деревянные последовательности $a_1, a_2, \ldots, a_p$ и $b_1, b_2, \ldots, b_q$, что $a_1 + b_1 = x_1$, и упорядоченные наборы $(a_2, a_3, \ldots, a_p, b_2, b_3, \ldots, b_q)$ и $(x_2, x_3, \ldots, x_n)$ одинаковы с точностью до перестановки.

Каковы аналоги характеризаций соседственных последовательностей (теорем 1 и 2) для пересечения *более, чем трех* многогранников?

5.5. * **Гипотеза.** Заданы натуральные числа $n_1, n_2, n_3, \ldots, n_s$ и s последовательностей

$$x_{11}, x_{12}, \ldots, x_{1n_1}, \quad x_{21}, x_{22}, \ldots, x_{2n_2}, \quad \ldots, \quad x_{s1}, x_{s2}, \ldots, x_{sn_s}$$

натуральных чисел. Тогда существуют s криволинейных сфер $S_1, S_2, \ldots, S_k$ в трехмерном пространстве, попарно пересекающихся по окружностям и таких, что

• никакие три из них не пересекаются в одной точке;

• для каждого $k = 1, \ldots, s$ и $j = 1, \ldots, n_k$ дополнение $S_k - S_{k+1} - S_{k+2} - \cdots - S_{k+s-1}$ имеет n_k компонент связности, j-тая из которых имеет x_{kj} соседей в S_k

тогда и только тогда когда все s последовательностей деревянные, и $n_1 + n_2 + \cdots + n_s - s$ есть четное число, не меньшее $2n_k$ для каждого $k = 1, \ldots, s$.

Для $s < 4$ эта гипотеза доказана (см. Теоремы 1 и 2), первый нетривиальный случай — $s = 4$.

Какими могут быть наборы количеств соседей у компонент, если разрешить тройные пересечения криволинейных сфер?

5.6. * **Гипотеза.** *Пусть n_1, n_2, n_3 — положительные целые числа, и*

$$x_{11}, x_{12}, \dots, x_{1n_1}, \quad x_{21}, x_{22}, \dots, x_{2n_2}, \quad x_{31}, x_{32}, \dots, x_{3n_3}$$

— последовательности положительных целых чисел. Тогда существование криволинейных сфер S_1, S_2, S_3,

- *попарно пересекающихся по окружностям,*
- *имеющих $2T$ точек пересечения всех трех сфер и*
- *таких, что для каждого $k = 1, 2, 3$ дополнение $S_k - S_{k+1} - S_{k+2}$ имеет n_k связных компонент, у i-ой из которых x_{ki} соседей в S_k для каждого $i = 1, \dots, k$*

равносильно тому, что $n_1 + n_2 + n_3 + T$ нечетно, $x_{k1} + x_{k2} + \dots + x_{kn_k} = 2n_k - 2 + 2T$ и $n_k + T < n_{k+1} + n_{k+2}$ для каждого $k = 1, 2, 3$.

Каковы аналоги теорем 1 и 2 для пересечения *криволинейных сфер с ручками*?

5.7. * **Гипотеза.** *Заданы целые числа $g_1, g_2, n > 0$ и две последовательности*

$$x_{11}, x_{12}, \dots, x_{1n}, \quad x_{21}, x_{22}, \dots, x_{2n}$$

натуральных чисел. Существуют криволинейные сфера с g_1 ручками S_1 и сфера с g_2 ручками S_2 в трехмерном пространстве, пересекающиеся по окружностям, разбивающим S_k на n компонент связности, j-тая из которых имеет x_{kj} соседей в S_k для каждого $k = 1, 2$ и $j = 1, \dots n$

тогда и только тогда когда $s := x_{11} + x_{12} + \dots + x_{1n} = x_{21} + x_{22} + \dots + x_{2n}$ четно и $2n - 2 \le s \le 2n - 2 + 2g_k$ для каждого $k = 1, 2$.

Было бы интересно решить аналогичные проблемы при наличии самопересечений. Интересны оба варианта — с точками троекратного самопересечения или без.

6 Финиш. Решения.

В этом параграфе криволинейные сферы называются просто *сферами*.

Проблема соседственных последовательностей

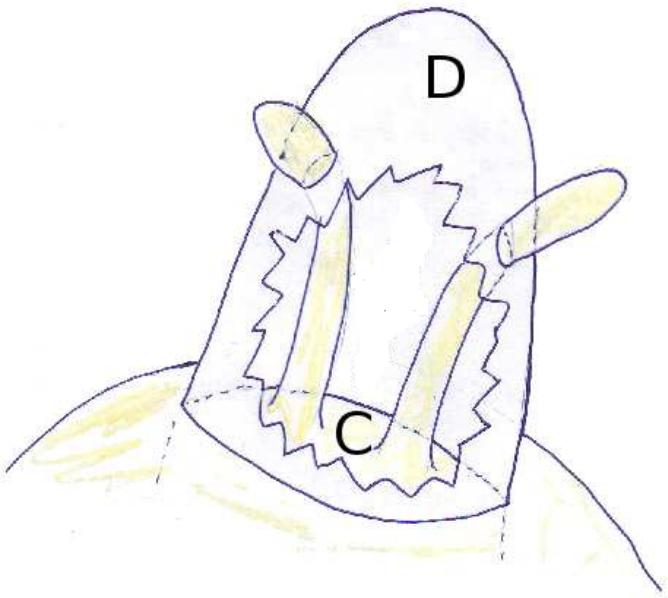


Рис. 17: Индуктивное построение

4.1. Можно считать, что $x_1 \geq y_1$. Рассмотрим сферы S', T', реализующие пару $(\vec{x}', \vec{y}')$ последовательностей. Рассмотрим окружность в пересечении $T' \cap S'$, из условия (2) в определении сильной реализуемости. Это окружность, ограничивающая

• в $S' - T'$ связную компоненту, назовем ее C, у которой $x_1 - y_1 + 1$ соседей,

• в $T' - S'$ — диск, назовем его D.

Изменим сферы S', T' соединением компонент C и D при помощи $y_1 - 1$ 'пальцев', см. рис. 17. Новые сферы обозначим через S, T. Докажем, что они реализуют пару $(\vec{x}, \vec{y})$ последовательностей.

Условие (1) выполнено для S, T, поскольку

• каждая компонента $S' - T'$, кроме C, также является и компонентой дополнения $S - T$,

• C делится $y_1 - 1$ окружностями из $(S \cap T) - (S' \cap T')$ на $y_1 - 1$ дисков и компоненту с $(x_1 - y_1 + 1) + (y_1 - 1) = x_1$ соседями,

и

• каждая компонента дополнения $T' - S'$, кроме D, также является и компонентой дополнения $T - S$,

• D разделено $y_1 - 1$ окружностями на $y_1 - 1$ дисков и компоненту с y_1 соседями.

Любая окружность из $(S \cap T) - (S' \cap T')$ удовлетворяет условию (2).

4.3. *Доказательство теоремы 1'.* Докажем индукцией по длине n последовательности. Для любой деревянной последовательности из n чисел имеем $n \geq 2$. База индукции ($n = 2$) очевидна.

Докажем шаг индукции. Пусть теорема 1' доказана для $2, 3, \ldots, n-1 \geq 2$. Докажем ее для n.

Переставим единицы в последовательностях в конец. Согласно задачам 2.5 и 4.1, по предположению индукции новые последовательности реализуемы. Рассмотрим сферы S, T, их реализующие. Сферы S, T удовлетворяют условию (1) из определения сильной реализуемости для старых последовательностей. Также

• если $x_1 \neq 1$, то условия (2) для старой и новой последовательностей $\vec{x}$ эквивалентны;

• если $x_1 = 1$, то окружность из условия (2) для новой последовательности $\vec{x}$ ограничивает диск, т.е. компоненту с $x_1 = 1$ соседом.

Это выполнено и с заменой $\vec{x}$ на $\vec{y}$. Итак, условие (2) выполнено для старых последовательностей. Значит, S, T сильно реализуют старые последовательности.

Проблема Ландо.

3.1. (d) (прямое решение) Пусть, напротив, существуют сферы S' и T', реализующие пару с рисунка 9. Обозначим связные компоненты дополнения $S' - T'$ как на рис. 14 слева.

Рассмотрим диски $A_1, \ldots, A_4 \subset S'$. Без ограничения общности предположим, что их внутренности лежат внутри T'. Тогда внутренность компоненты $C \subset S'$ также лежит внутри T' (так как пересечение $S' \cap T'$ трансверсально). Так как $C, A_1, \ldots, A_4$ не пересекаются, то C целиком лежит в одной связной компоненте дополнения $\mathbb{R}^3 - T' \cup \bigsqcup A_i$. Таким образом, все 5 окружностей границы ∂C лежат в одной связной компоненте дополнения $T' - \bigsqcup \partial A_i$. (Здесь мы используем тривиальный частный случай теоремы о продолжении вложения.)

Переформулируем предыдущее утверждение в терминах графа $G(T')$ (рис. 18). Обозначим за $G(C)$ объединение 5-и ребер графа $G(T')$, соответствующих окружностям границы ∂C. Тогда $G(C)$ целиком лежит в одной связной компоненте дополнения в $G(T')$ к 4-ем ребрам, соответствующим окружностям границы ∂A_i. Так как у $G(T')$ только 9 ребер, это значит, что $G(C)$ —поддерево $G(T')$. Обозначим через $G(B)$ объединение 5-и ребер графа $G(T')$, соответствующих окружностям границы ∂B. Аналогично, $G(B)$ — поддерево $G(T')$.

Так как $G(B) \cup G(C) = G(T')$, то хотя бы два из трех ребер a, b, c графа $G(T')$ (рис. 18) лежат в одном из поддеревьев $G(B)$ или $G(C)$. Без ограничения общности $a, b \in G(B)$. Но в любом поддереве $G(T')$, содержащем и a, и b хотя бы 6 ребер, в то время как в $G(B)$ лишь 5 ребер. Значит, исходное предположение противоречиво и неверно.

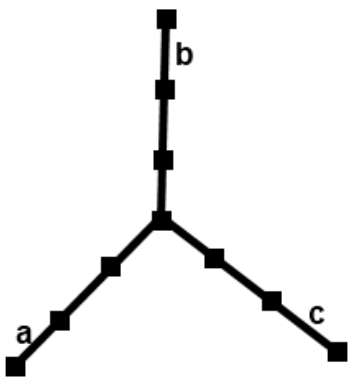


Рис. 18: Граф $G := G(T', N)$.

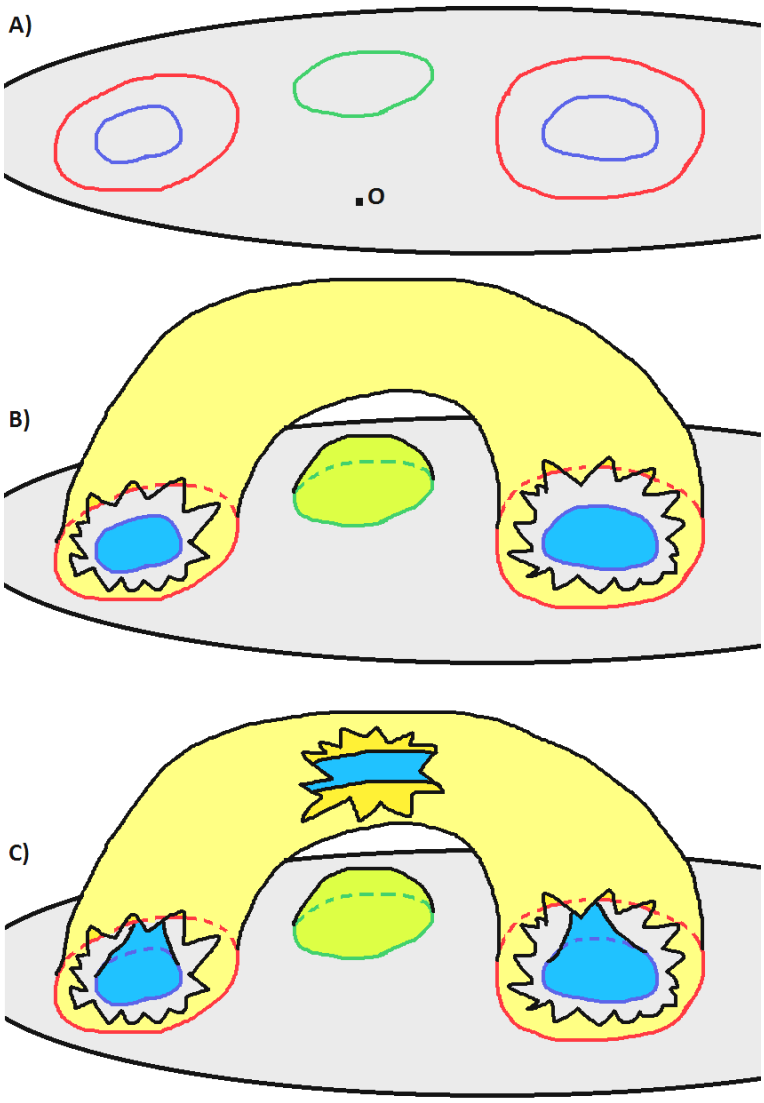


Рис. 19: К решению задачи 3.6.f. (A) Имеем S (серая), p_1 (красное), p_2 (зеленое), p_3 (синее). (B) Имеем, что $\mathring{p}_3$ (синее) — 'наименьшее'. Построим P_1 (желтое) и P_2 (зеленое) по индукции. (C) Связные компоненты $\mathring{p}_3$ (синее) можно соединить путем, не пересекающимся с $P_1 \cup P_2$. Тогда соединим их трубой и получим P_3 (синее).

3.6. (a) Очевидно.

(b) Это равносильно тому, что окружности q_1 и q_2 лежат по одну сторону от p.

(c) Это равносильно тому, что $p_1 \sqcup p_2$ и $q_1 \sqcup q_2$ не зацеплены.

(d) Это равносильно тому, что p и q не зацеплены. *Подсказка:* обобщите решение задачи 3.1.d. *Формальное решение* получается подстановкой $m = 2$ в решение для (f).

(e) Нет, согласно ответу на (f).

(f) Это равносильно тому, что p_i и p_j не зацеплены для любых $i \neq j$.

Теорема о продолжении вложения. *Наборы $p_1, \dots, p_m$ непересекающихся окружностей в единичной сфере S попарно не зацеплены в том и только том случае, когда существуют непересекающиеся сферы с дырками $P_1, \dots, P_m$, чьи внутренности лежат внутри S, и для которых $\partial P_i = p_i$ для всех $i = 1, \dots, m$.*

Доказательство. Необходимость легко следует из задачи 3.4.c. Достаточность докажем индукцией по m. База $m = 1$ по сути доказана при решении задачи 3.2.a. Докажем шаг индукции.

Возьмем точку $O \in S - \bigsqcup\limits_{i=1,\dots,m} p_i$. Для каждого i рассмотрим черно-белую раскраску дополнения $S - p_i$, при которой O белая. Напомним, что $\mathring{p}_i$ — объединение черных компонент дополнения $S - p_i$. Раз O белая, а p_i и p_j не зацеплены, по задаче 3.7.b для всех $i \neq j$ или $\mathring{p}_i \subset \mathring{p}_j$, или $\mathring{p}_j \subset \mathring{p}_i$, или $\mathring{p}_i \cap \mathring{p}_j = \emptyset$. Значит, есть 'наименьшее' $\mathring{p}_i$, то есть, такое $\mathring{p}_i$, что $\mathring{p}_j \not\subset \mathring{p}_i$ для всех $j \neq i$. Можно положить $i = m$. Тогда $\mathring{p}_m \cap \bigsqcup\limits_{i=1}^{m-1} p_i = \emptyset$, $\mathring{p}_m$ — набор сфер с дырками, $\partial \mathring{p}_m = p_m$ и $\mathring{p}_m \subset S$. Обозначим через Δ замкнутый шар, ограничиваемый S (то есть, 'внутреннюю часть' сферы S). По индукционному предположению существуют непересекающиеся сферы с дырками $P_1, \dots, P_{m-1} \subset \Delta$, такие что $\partial P_i = p_i$ для всех $i = 1, \dots, m-1$.

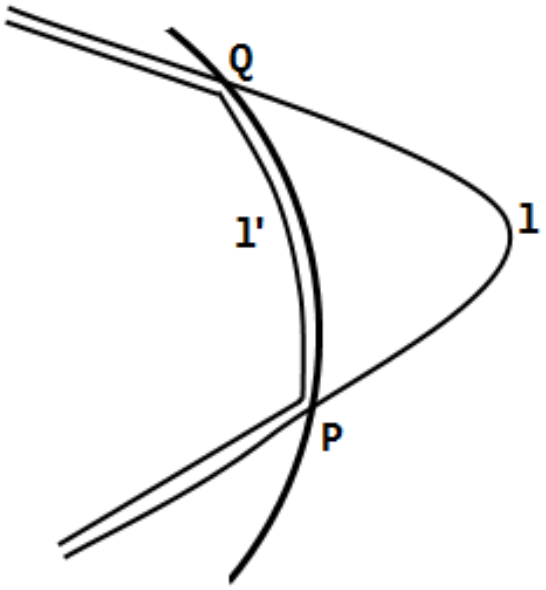


Рис. 20: Доказательство утверждения

Утверждение. *Набор p_m лежит в одной связной компоненте дополнения $\Delta - (P_1 \sqcup \dots \sqcup P_{m-1})$.* [6]

Доказательство утверждения. Предположим противное. Рассмотрим любые две точки $A, B \in p_m$ из разных связных компонент дополнения $\Delta - (P_1 \sqcup \dots \sqcup P_{m-1})$. Обозначим через l путь внутри S, соединяющий A и B, и при этом $\bar{l} := \#(l \cap \bigsqcup\limits_{i=1}^{m-1} P_i)$ минимально (минимум по всем таким l, объекты $A, B, p_m, S, P_1, \dots, P_{m-1}$ фиксированы). Тогда $\bar{l} > 0$ (иначе A и B лежали бы в разных компонентах). Так как p_m лежит по одну сторону от ∂P_i, точки A и B лежат в одной компоненте $\Delta - P_i$, и $\#(l \cap P_i)$ четно для всех i. (Если $m = 2$, можно даже получить $\#(l \cap P_1) = 0$ и остановиться.) Тогда $\#(l \cap P_i) \geq 2$ для некоторого i. Обозначим через Q и R последовательные (на l) точки $l \cap P_i$. Обозначим через Q' точку l чуть перед Q и за R' точку l чуть после R. Так как P_i связно, Q и R можно соединить путем P_i. Значит Q' и R' можно соединить путем l', достаточно близким к P_i, но не пересекающим P_i. Путь l' не пересекает ничего из $P_1, \dots, P_{m-1}$, так как он достаточно близок к P_i, а $P_1, \dots, P_{m-1}$ попарно не пересекаются. Заменим часть пути l от Q' до R' на l'. Полученный путь назовем l''. Теперь $\overline{l''} = \bar{l} - 2$. Это противоречит минимальности $\bar{l}$. Поэтому l не пересекает $P_1 \sqcup \dots \sqcup P_{m-1}$, значит A и B должны лежать в одной связной компоненте дополнения $\Delta - (P_1 \sqcup \dots \sqcup P_{m-1})$. ЧТД.

[6] Для $m = 2$ это утверждение — почти определение незацепленности (точнее, того, что 'p_2 лежит по одну сторону от ∂P_1'). Случай $m \geq 3$ интересней, ведь объединение двух подмножеств, вообще говоря, может разбить само множество, даже если каждое из этих подмножеств поодиночке исходное множество не разбивало.

Завершение доказательства теоремы о продолжении вложения. Пусть $\mathring{p}'_m$ — объединение непересекающихся сфер с дырками, полученных из $\mathring{p}_m$ небольшой деформацией, после которой внутренность $\mathring{p}'_m$ попадает в Δ и $\partial\mathring{p}'_m = \partial\mathring{p}_m = p_m$. Согласно Утверждению, любые две точки $\mathring{p}'_m$ можно соединить путем внутри S, не пересекающим $P_1, \dots, P_{m-1}$. Тогда можно соединить все связные компоненты множества $\mathring{p}'_m$ трубками внутри S, не пересекающими $P_1, \dots, P_{m-1}$. Количество трубок будет на 1 меньше количества компонент связности $\mathring{p}'_m$, значит, не будет 'циклов из трубок'. Значит, мы получили сферу с дырками. Обозначим ее через P_m. Имеем $\partial P_m = p_m$, $P_m \subset \Delta$, и P_m не пересекается с $P_1, \dots, P_{m-1}$. Шаг индукции доказан. ЧТД.

4.4. Этот факт получен с использованием компьютерной программы, основанной на теореме 3.

4.5. Ответ дается задачей 4.6, и он таков.

Теорема 3. *Пара (M, N) занумерованных наборов на сферах S и T реализуема тогда и только тогда, когда связные компоненты дополнения $S - M$ можно покрасить в черный и белый цвета так, что для любых двух одноцветных компонент P и Q дополнения $S - M$ поднаборы в N, соответствующие ∂P и ∂Q, не зацеплены.*

4.6. (a) Это переформулировка задачи 3.4.

(b) Да. Идея решения — доказать и использовать ответ к проблеме продолжения вложения 3.6.e.

Пусть T' — единичный куб. Нумерации дают взаимно-однозначное соответствие h между окружностями из M и окружностями из N.

Обозначим через $A_1, \dots, A_m$ белые связные компоненты дополнения $S - M$. По предположению $h(\partial A_1), \dots, h(\partial A_m)$ попарно не зацеплены в T. Согласно ответу к проблеме продолжения вложения 3.6.e существуют непересекающиеся криволинейные сферы с дырками $A'_1, \dots, A'_m$, внутренности которых лежат *внутри* T', и такие, что $\partial A'_i = h(\partial A_i)$ для всех $i = 1, \dots, m$.

Обозначим через $B_1, \dots, B_n$ черные связные компоненты дополнения $S - M$. Аналогично существуют непересекающиеся сферы с дырками $B'_1, \dots, B'_n$, внутренности которых лежат *снаружи* T', и такие, что $\partial B'_i = h(\partial B_i)$ для всех $i = 1, \dots, n$.

Пусть $S' := (A'_1 \cup \dots \cup A'_m) \cup (B'_1 \cup \dots \cup B'_n)$. Из построения следует, что S' несамопересекающаяся. У A'_i такое же число дырок, как и у A_i, и у B'_i такое же число дырок, как и у B_i. Так как $S = (A_1 \cup \dots \cup A_m) \cup (B_1 \cup \dots \cup B_n)$ — сфера, то и S' тоже. (Строгое доказательство можно получить, воспользовавшись эйлеровой характеристикой.) Очевидно, S' и T' реализуют пару M, N.

4.7. (a) Это — переформулировка задач 3.4 и 4.6.

Больше сфер и сферы с ручками

5.1. Теорема 2. *Пускай n_1, n_2, n_3 —положительные целые числа, а*

$$x_{11}, x_{12}, \dots, x_{1n_1}, \quad x_{21}, x_{22}, \dots, x_{2n_2}, \quad x_{31}, x_{32}, \dots, x_{3n_3}$$

— последовательности положительных целых чисел. Тогда существуют криволинейные сферы S_1, S_2, S_3 в трехмерном пространстве, пересекающиеся попарно по окружностям и такие, что

- *$S_1 \cap S_2 \cap S_3 = \varnothing$;*
- *У $S_k - S_{k+1} - S_{k+2}$ n_k связных компонент, которые можно так пронумеровать, что у i-ой компоненты x_{ki} соседей в S_k, для всех $k = 1, 2, 3$*

если и только если последовательности деревянные, $n_1 + n_2 + n_3$ нечетно, и $n_k < n_{k+1} + n_{k+2}$ для всех $k = 1, 2, 3$.

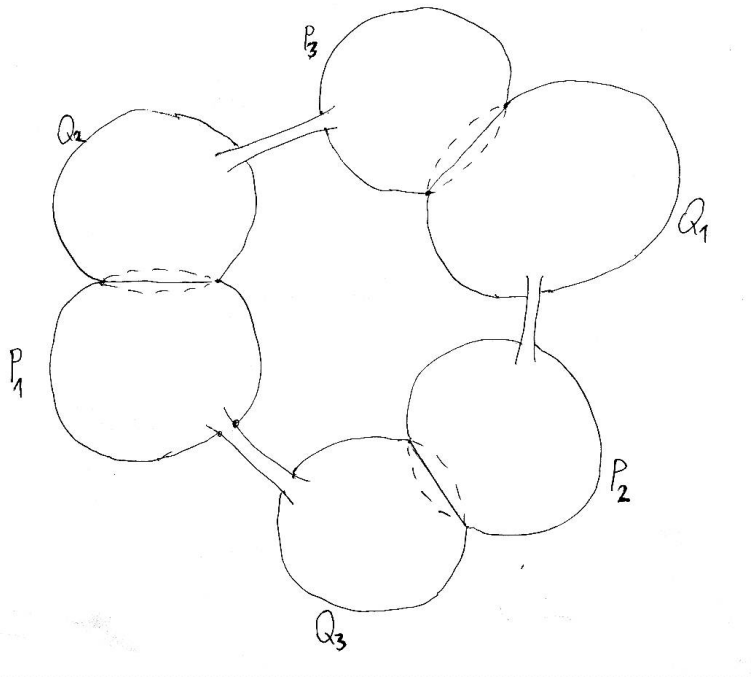


Рис. 21: Конструкция из трех сфер

Доказательство. Часть ‘только тогда’ следует из задач 2.2 и 5.3. Докажем часть ‘тогда’. Пусть

$$m_1 := (n_2 + n_3 - n_1 + 1)/2, \quad m_2 := (n_1 + n_3 - n_2 + 1)/2, \quad m_3 := (n_1 + n_2 - n_3 + 1)/2.$$

Тогда

$$m_1 + m_2 = n_3 + 1, \quad m_1 + m_3 = n_2 + 1, \quad m_2 + m_3 = n_1 + 1.$$

Согласно задаче 5.4 существуют последовательности

$$p_{11}, p_{12}, \dots, p_{1m_3}, \quad p_{21}, p_{22}, \dots, p_{2m_1}, \quad p_{31}, p_{32}, \dots, p_{3m_2},$$

$$q_{11}, q_{12}, \dots, q_{1m_2}, \quad q_{21}, q_{23}, \dots, q_{2m_3}, \quad q_{31}, q_{32}, \dots, q_{3m_1},$$

такие что $p_{k-1,1} + q_{k+1,1} = x_{k1}$ и упорядоченные множества

$$(p_{k-1,2}, p_{k-1,3}, \dots, p_{k-1,m_{k+1}}, q_{k+1,2}, q_{k+1,3}, \dots, q_{k+1,m_{k-1}}) \quad \text{и} \quad (x_{k2}, x_{k3}, \dots, x_{kn_k})$$

равны с точностью до перестановки. По теореме 1’ существуют сферы

$$Q_1, P_1, Q_2, P_2, Q_3, P_3 \subset \mathbb{R}^3 \quad \text{такие что} \quad Q_k \cap Q_{k+1} = \varnothing, \quad Q_k \cap P_l = \varnothing \quad \text{если} \quad l \neq k-1 \quad \text{и}$$

• $Q_k - P_{k-1}$ — объединение m_{k+1} непересекающихся связных компонент, у i-ой из которых q_{ki} соседей

• $P_{k-1} - Q_k$ — объединение m_{k+1} непересекающихся связных компонент, у i-ой из которых $p_{k-1,i}$ соседей

• граница некоторой связной компоненты $\mathbb{R}^3 - P_{k-1} - Q_k$ содержит компоненту $\widetilde{q}_k$ с q_{k1} соседями в Q_k и компоненту $\widetilde{p}_{k-1}$ с $p_{k-1,1}$ соседями в P_{k-1}.

Для $k = 1, 2, 3$ пусть S_k — связная сумма сфер Q_{k+1} и P_{k-1} и тонкой трубки, соединяющей две компоненты $\widetilde{q}_{k+1}$ и $\widetilde{p}_{k-1}$ из третьего условия, см. рис. 21. Это можно сделать без пересечений на этих трех трубках.

Тогда $S_k - S_{k+1} - S_{k+2}$ — такое, как и должно быть для каждого $k = 1, 2, 3$. ЧТД.

5.2. *Ответ:* реализуемы только следующие тройки.

$$\{(2,1,1),(2,1,1),(2,1,1)\},\quad \{(3,1,1,1),(3,1,1,1),(2,1,1)\},\quad \{(3,1,1,1),(2,2,1,1),(2,1,1)\},$$
$$\{(3,1,1,1),(2,1,1),(1,1)\},\quad \{(2,2,1,1),(2,2,1,1),(2,1,1)\},$$
$$\{(2,2,1,1),(2,1,1),(1,1)\},\quad \{(2,1,1),(1,1),(1,1)\}.$$

Доказательство. Существует только 4 деревянные последовательности длины не более 4:

$$(1,1),\ (2,1,1),\ (3,1,1,1),\ (2,2,1,1).$$

Согласно утверждению задачи 5.3 количество последовательностей нечетной длины в реализуемой тройке нечетно. Значит каждая реализуемая тройка последовательностей длины не более 4, кроме тройки $\{(2,1,1),(2,1,1),(2,1,1)\}$, состоит из последовательности $(2,1,1)$ и двух последовательностей четной длины. По задаче 5.1 все эти 7 троек реализуемы.

5.3. Пусть m_3, m_2, m_1 — количества окружностей в $f_1 \cap f_2$, $f_1 \cap f_3$ и $f_2 \cap f_3$. Тогда $n_1 = m_3 + m_2 + 1$, $n_2 = m_3 + m_1 + 1$, $n_3 = m_2 + m_1 + 1$.

Значит, $n_1 + n_2 + n_3 = 2(m_3 + m_2 + m_1) + 3$ нечетно.

Так как $2m_k + 1 > 0$, то $n_k < n_{k+1} + n_{k+2}$ для всех $k = 1, 2, 3$.

5.4. Пусть $r = r(\vec{x})$ — число членов x_i больших 1. Пусть $z_s = x_2 + x_3 + \cdots + x_s$. Для каждого $s \le r$ положим

$$a_1 = p - (z_s - s + 3) + 1,\quad a_i = x_i \quad \text{для} \quad 2 \le i \le s \quad \text{и} \quad a_i = 1 \quad \text{для} \quad s+1 \le i \le p,$$

$$b_1 = x_1 - a_1,\quad b_i = x_{i+s-1} \quad \text{для} \quad 2 \le i \le r - s + 1,\quad b_i = 1 \quad \text{для} \quad r - s + 2 \le i \le q = n + 1 - p.$$

Так как $s \le r$, последовательность $b_1, b_2, \ldots, b_q$ определена корректно. Для каждого i числа a_i и b_i зависят от s.

Получаем

$$a_1 + a_2 + \cdots + a_p = p - (z_s - s + 3) + 1 + z_s + p - s = 2p - 2,$$

то есть последовательность $a_1, a_2, \ldots, a_p$ деревянная. Также

$$b_1 + b_2 + \cdots + b_q = z_n - a_1 - a_2 - \cdots - a_p = 2n - 2 - 2p + 2 = 2q - 2,$$

то есть последовательность $b_1, b_2, \ldots, b_q$ деревянная.

Осталось доказать существование числа $s \le r$, при котором $1 \le a_1 \le x_1 - 1$. Для каждого $i < r$ имеем $x_1 \ge x_i$, значит

$$z_i - i + x_1 + 1 \ge (z_{i+1} - (i+1) + 3) - 1.$$

Другими словами,

$$\begin{gathered} 2 = z_1 - 1 + 3, \\ z_1 - 1 + x_1 + 1 \ge (z_2 - 2 + 3) - 1, \\ z_2 - 2 + x_1 + 1 \ge (z_3 - 3 + 3) - 1, \\ \ldots, \\ z_{r-1} - (r-1) + x_1 + 1 \ge (z_r - r + 3) - 1, \\ z_r - r + x_1 + 1 = n - 1. \end{gathered}$$

Здесь последнее равенство не аналогично предыдущим равенствам, но следует из того, что последовательность $x_1, x_2, \ldots, x_n$ деревянна, и $1 = x_{r+1} = \cdots = x_n$. Так как $2 \le p \le n - 1$, существует число $s \le r$, для которого

$$z_s - s + 3 \le p \le z_s - s + x_1 + 1 \quad \Leftrightarrow \quad 1 \le a_1 \le x_1 - 1.$$

Список литературы.

[A] S. Avvakumov, *A counterexample to the Lando conjecture on intersection of spheres in 3-space*, preprint, 2012.

[A12] I. Arzhantsev, V. Bogachev, A. Garber, A. Zaslavsky, V. Protasov and A. Skopenkov, *Students' mathematical olympiades at Moscow State University 2010-2011*, Mat. Prosveschenie, 16 (2012), 214-227.

[D] Н. П. Долбилин, *Жемчужины теории многогранников*, М.: МССМЕ, 2000.

[H10] T. Hirasa, *Dissecting the torus by immersions*, Geometriae Dedicata, 145:1 (2010), 33-41

[R] A. Rukhovich, *On intersection of two embedded spheres in 3-space*, http://arxiv.org/abs/1012.0925

[T07] T. Nowik, *Dissecting the 2-sphere by immersions*, Geometriae Dedicata 127, (2007), 37-41, http://arxiv.org/abs/math/0612796.

[W] http://ru.wikipedia.org/wiki/Многогранник

HOW DO CURVED SPHERES INTERSECT IN 3-SPACE, OR TWO-DIMENSIONAL MEANDRA [1]

S. Avvakumov, A. Berdnikov, A. Rukhovich and A. Skopenkov [2]

1 Examples and main problems

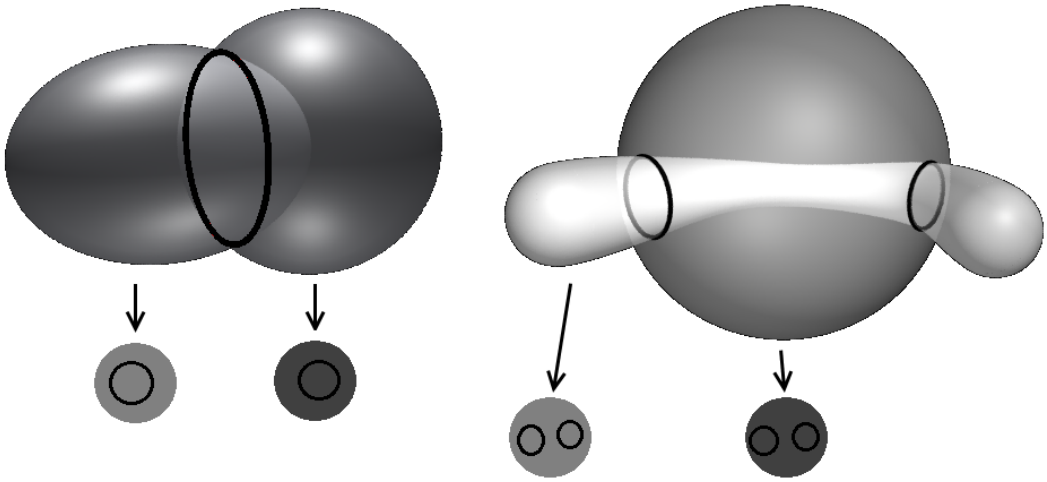


Figure 1: Curved spheres intersecting by a circle (left), by two circles (right)

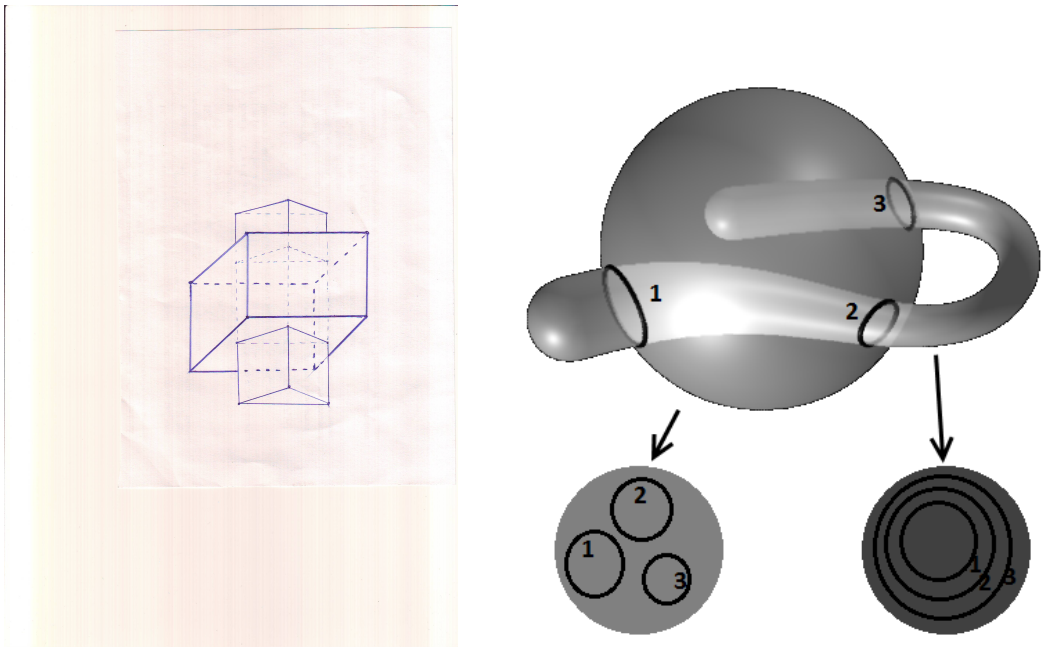


Figure 2: Curved spheres intersecting by two circles (left), by three circles (right)

How can two curved spheres intersect in 3-space? In figures 1 and 2 you see pairs of curved spheres in 3-space intersecting by a union of circles.

[1]We are grateful for useful remarks and discussions to G. Chelnokov, S. Lando and to an anonymous referee of Moscow Mathematical Conference of High-School Students.

[2]Supported by Simons-IUM Fellowship

The term *curved sphere* is possibly intuitively clear for you. If not, read rigorous definitions below. Here we only remark that in this text curved sphere *does not have self-intersections*.

You will possibly be able to solve all the problems below without really using those rigorous definitions. In your solutions you can represent curved spheres by large pictures clear to jury members, not only by formal constructions. If the formulation of a problem is a statement, it is required to prove this statement. If a problem looks like too difficult, try to solve the neighboring problems, they can contain hints. Unsolved problems are marked by stars.

1.1. Draw two curved spheres in 3-space intersecting by a disjoint union of 3 circles so that in each sphere these circles

(a) bound 3 disjoint disks (like in figure 2 left).

(b) do not bound 3 disjoint disks (like in figure 2 right).

1.2. Draw two curved spheres in 3-space intersecting by a disjoint union of 4 circles so that in each sphere these circles

(a) bound 4 disjoint disks (as in figure 3.a).

(b) are 'parallel', or 'one inside the other' (as in figure 3.b).

(c) are situated as in figure 3.c.

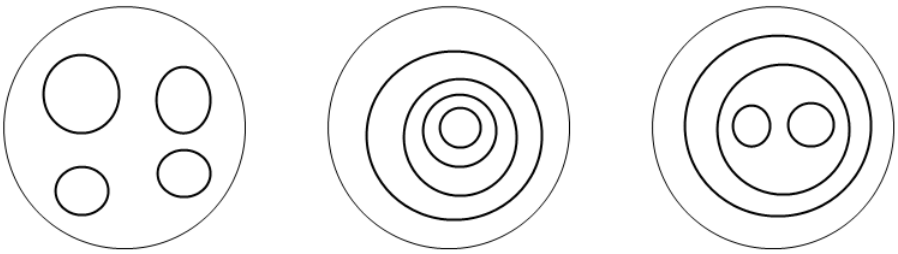


Figure 3: Four circles in a sphere

Suppose that M and N are collections of the same number of disjoint circles in curved spheres S and T. Then M *is situated in* S *as* N *in* T if there is a bijection between connected components (=connected parts) of $S - M$ and $T - N$ such that connected components of $S - M$ are neighbors if and only if the two corresponding connected components of $T - N$ are neighbors. (Or, equivalently, if pairs (S, M) and (T, N) are piecewise-linearly homeomorphic.)

1.3. (ij), $i, j \in \{a, b, c\}$. Draw two curved spheres in 3-space intersecting by a disjoint union of 4 circles so that in one sphere the circles are situated as in figure 3.i, and in the other as in figure 3.j.

In this text we study the following two problems and their generalizations. (You probably will not be able to solve the problems right away, so postpone them and try to solve other problems.)

The intersection of two curved spheres is *transversal* if near each intersection point it looks like the intersection of two planes having a common line. (See a rigorous definition below.)

1.4. (a) **The Lando Problem.** Let M and N be two unions of the same number of disjoint circles in a sphere. Do there exist two curved spheres in 3-space whose intersection is transversal and is a finite collection of disjoint circles that is situated as M in one sphere and as N in the other?

(b) Does there exist an algorithm for checking the existence of such two curved spheres? (Cf. 'relation to graphs' below.)

(c)* Does there exist such a polynomial algorithm?

In figure 1, left, two spheres intersect by a circle; each sphere is split by the circle into 2 connected components, and each connected component has one neighboring connected component (in the same sphere). In figure 1, right, (and in figure 2, left) two curved spheres intersect by 2 circles; each sphere

is split by the circles into 3 connected component, of which two have one neighboring connected component and one has two neighboring connected components (in the same sphere). In figure 2, right, two curved spheres intersect by 3 circles; each sphere is split by the circles into 4 connected components, in one sphere the numbers of neighbors of connected components are $3, 1, 1, 1$, in the other sphere those numbers are $1, 2, 2, 1$.

1.5. Neighbor Sequence Problem. Given sequences $\vec{x} = (x_1, x_2, \ldots, x_n)$, $\vec{y} = (y_1, y_2, \ldots, y_n)$ of positive integers, does there exist two curved spheres S, T in 3-space whose intersection consists of $n - 1$ circles and splits

• S into n connected components which can be numbered so that the i-th connected component has x_i neighbors in S, and

• T into n connected component which can be numbered so that the i-th connected component has y_i neighbors in T?

Some rigorous definitions.

You will probably be able to solve all the problems without really using these rigorous definitions.

We present definitions convenient in frame of this text; they could be different from common mathematical terms.

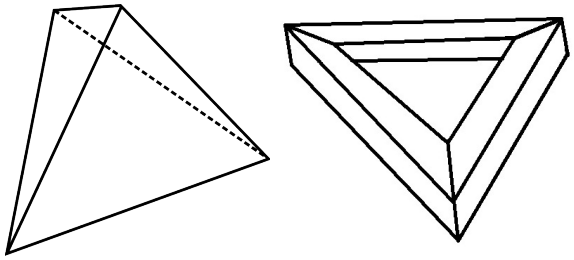


Figure 4: A curved sphere (left), not a curved sphere (right)

In this text a *curved circle* or, shortly, a *circle* is a closed broken line without self-intersections in 3-space. The definition of a polyhedron (without self-intersections, but possibly non-convex) is given in [D], see also [W]. A *curved sphere* is a polyhedron in 3-space (more precisely, 2-dimensional surface of the polyhedron), which is split into several parts by any circle lying on the polyhedron. See figure 1. (Such polyhedra are called topologically equivalent to sphere. This condition is equivalent to the condition $V - E + F = 2$.)

In order to simplify pictures, instead of a polyhedron we draw a *curved* surface 'close' to the polyhedron. For example, a curved sphere or a 'sausage' as in Figure 2. Instead of a broken line we draw a *curve* 'close' to the broken line.

A subset X of the 3-space is *connected* if each two points of X can be connected by a broken line in X. A *connected component* of a subset X of the 3-space is a maximal connected subset of X, i.e., a connected subset $Y \subset X$ such that there does not exist a connected subset $Z \subset X$ for which $Y \subset Z \subset X$ and $Y \neq Z \neq X$.

Suppose that M is a collection of disjoint circles in curved sphere S. Two connected components of the complement $S - M$ are *neighbors* if their closures intersect.

Denote by $B(x, r) \subset \mathbb{R}^3$ the ball of radius r centered at x. Intersection of two curved spheres $S, T \subset \mathbb{R}^3$ is *transversal* if for each point $x \in S \cap T$ there is $r > 0$ such that both $B(x, r) - S$ and $B(x, r) \cap (T - S)$ consist of two connected components, and each connected component of $B(x, r) - S$ contains a connected component of $B(x, r) \cap (T - S)$.

You can use the following theorem and corollary without proof.

Jordan Theorem. *A curved sphere splits 3-space into exactly two connected components.*

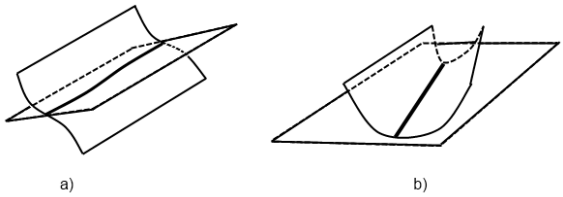


Figure 5: A transversal intersection (left), not a transversal intersection (right)

Corollary. *Suppose that S and T are curved spheres intersecting transversely by a finite union $S \cap T$ of disjoint circles. Denote by B the interior part of S (or, more precisely, the bounded part of $\mathbb{R}^3 - S$). Let Q be a connected component of $T - S$ which is situated inside S. Then Q splits B into exactly two connected components.*

Relation to graphs.

Suppose that M is a collection of disjoint circles in curved sphere S. Define ('dual to M') graph $G = G(S, M)$ as follows. The vertices are connected components of $S - M$. Two vertices are connected by an edge if the corresponding connected components are neighbors.

In figure 6 we show graphs for spheres S, T from figure 2 and collection $S \cap T$ of circles. Analogously two curved spheres intersecting by circles define a pair of graphs. Then the Lando Problem asks to describe such pairs of graphs, and Neighbor Sequence Problem asks to describe pairs of degree sequences of such pairs of graphs. [3]

Figure 6: Two graphs corresponding to Figure 2

Stars.

A team gets a star for each correct ($\geqslant +.$) written solution. *A large picture clear to jury members, or a well-structured computer program passing tests assigned by Jury, is recognized as an equivalent of a written solution.* Jury may also award stars for elegant solutions, for solutions of difficult problems and for (some) solutions written in TEX. The jury has infinite number of stars. A team may present a solution orally paying 1 star for each attempt.

We invite participants succeeding in solving these problems and working on unsolved problems to discuss their questions and ideas of solutions.

2 Neighbor Sequence Problem

A pair $\vec{x} = (x_1, x_2, \ldots, x_n)$, $\vec{y} = (y_1, y_2, \ldots, y_n)$ of sequences of positive integers is called *realizable* if there exist two curved spheres S, T in 3-space whose intersection consists of $n - 1$ circles and splits

• S into n connected components which can be numbered so that the i-th connected component has x_i neighbors in S, and

[3] Here is another interpretation suggested by I. N. Shnurnikov. Suppose that the unit square on the plane and (piecewise linear) function on the square are given. The function is strictly positive on the boundary of the square. The disk corresponds to the first curved sphere (with a hole), the graph of the function (above the disk) — to the second curved sphere, the zero set of the function — to the intersection of curved spheres.

• T into n connected components which can be numbered so that the i-th connected component has y_i neighbors in T.

Pair (S, T) of spheres is called a *realization* of pair $(\vec{x}, \vec{y})$.

2.1. (n), $n \in \{2, 3, 4, 5\}$. Which pairs of sequences of n positive integers are realizable?

2.2. (a) If pair $(\vec{x}, \vec{y})$ of sequences is realizable, then $x_1 + \cdots + x_n = y_1 + \cdots + y_n = 2n - 2$.

(b) The dual graph $G(S, M)$ to a collection M of disjoint circles in a curved sphere S is a tree.

A sequence $\vec{x} = (x_1, x_2, \ldots, x_n)$ of positive integers is called *tree-like* if $x_1 + \cdots + x_n = 2n - 2$.

2.3. If a sequence $\vec{x}$ is tree-like, then it has at least x_1 units.

2.4. Pair $(\vec{x}, \vec{x})$ is realizable for each tree-like $\vec{x}$.

2.5. Let $\vec{x}, \vec{y}$ be tree-like sequences in which all the units are situated at the end. If $x_1 \geq y_1$, then sequences $\vec{x}' := (x_1 - y_1 + 1, x_2, x_3, \ldots, x_{n-y_1+1})$ and $\vec{y}' := (y_2, y_3, \ldots, y_{n-y_1+2})$ are tree-like.

2.6. Which pairs of tree-like sequences could be obtained from pair $((1, 1), (1, 1))$ by reorderings and changes of pair ($\vec{x} = (x_1, x_2, \ldots, x_n)$, $\vec{y} = (y_1, y_2, \ldots, y_n)$) of vectors to pair:

(a) ($\vec{x'} = (a, x_1, x_2, \ldots, x_n, 1, 1, \ldots, 1)$, $\vec{y'} = (y_1 + a - 1, y_2, y_3, \ldots, y_n, 1, 1, \ldots, 1)$) (the number of new 1's is $a - 2$ for $\vec{x'}$ and is $a - 1$ for $\vec{y'}$; number a can be different for different changes, e.g. $((1, 1), (1, 1)) \overset{a=3}{\to} ((3, 1, 1, 1), (3, 1, 1, 1)) \overset{a=4}{\to} ((4, 3, 1, 1, 1, 1, 1), (6, 1, 1, 1, 1, 1, 1))$).

(b) ($\vec{x'} = (x_1 + 1, x_2, x_3, \ldots, x_n, 1)$, $\vec{y'} = (y_1 + 1, y_2, y_3, \ldots, y_n, 1)$).

3 The Lando Problem

A pair (M, N) of two unions of the same number of disjoint circles in a sphere is *realizable* if there exist two curved spheres in 3-space intersecting transversely by a finite union of disjoint circles which union is situated as M in one sphere and as N in the other sphere.

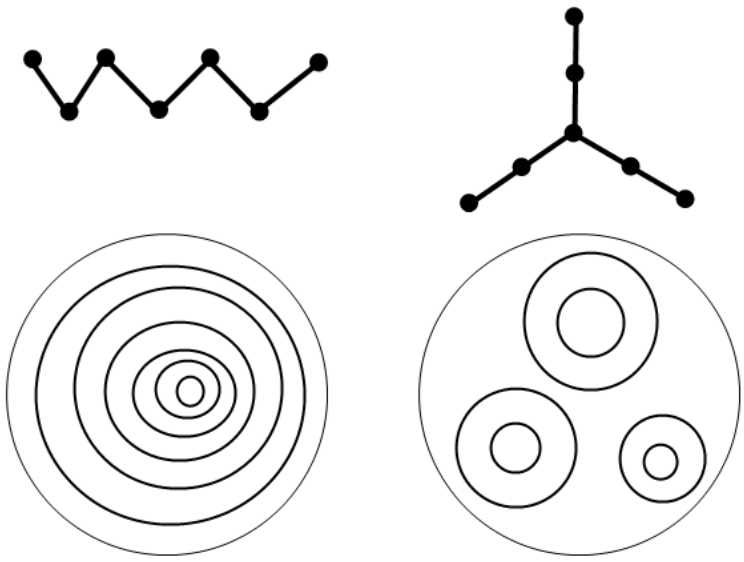


Figure 7: Is this pair realizable?

3.1. (a) Each pair of two unions of the same number $n \leq 4$ of disjoint circles is realizable.

(b) Is pair in figure 7 realizable? One graph is the path of 6 edges, the other is triod with 'rays' consisting of 2 edges.

(c) Is pair in figure 8 realizable? One graph is the star with 4 'rays', three 'rays' having two edges and one 'ray' having 1 edge. The other graph is letter 'H' with 'horizontal line' consisting of 3 edges.

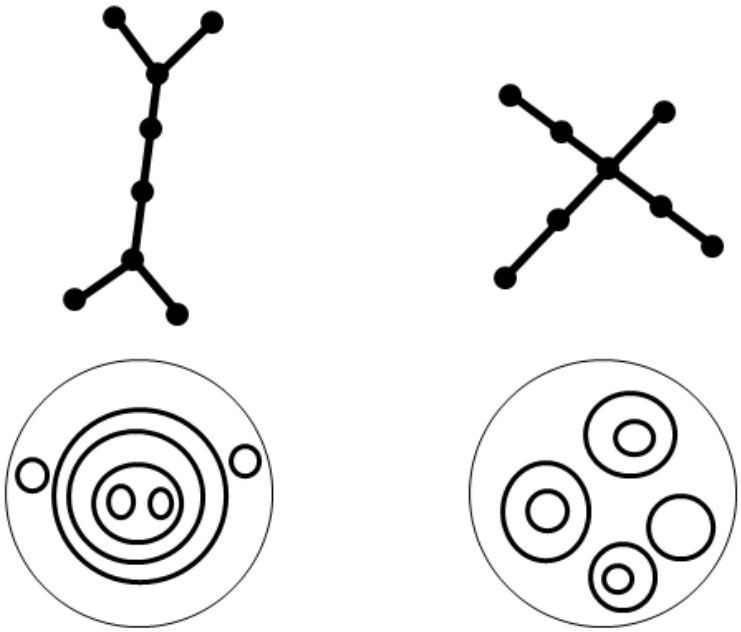


Figure 8: Is this pair realizable?

(d) There is a non-realizable pair of two unions of the same number of disjoint circles.

3.2. Suppose that n and k are given integers.

(a) Which collections of circles are realizable together with the collection of n circles bounding n disjoint disks? (Or, equivalently, which graphs are realizable together with the star of n rays?)

(b) Which graphs are realizable together with the graph that is a union, along a common edge, of the star of n rays and the star of k rays?

(c) * Which collections of circles are realizable together with the collection of n 'parallel' circles? (Or, equivalently, which graphs are realizable together with the path of length n?)

3.3. Suppose that S and T are curved spheres intersecting transversely by a finite union $S \cap T$ of disjoint circles. Then connected components of $S - T$ can be colored in black and white so that any two same colored components are not neighbors.

In the rest of this section M, N are unions of the same number of disjoint circles in curved spheres S, T. (Neither M nor N need to coincide with $S \cap T$.)

For a connected component P of $S - M$ denote by ∂P the union of boundary circles of P. Clearly, connected components P and Q of $S - M$ are neighbors if and only if $\partial P \cap \partial Q \neq \emptyset$.

3.4. Unlinked families of circles. Suppose that S and T are curved spheres intersecting transversely by a finite union $S \cap T$ of disjoint circles. Let P and Q be two connected components of $S - T$ which are situated inside T.

(a) If Q is a curved disk (i.e., if Q has one boundary circle), then ∂P is in one component of $T - \partial Q$.

(b) If Q is a curved cylinder (i.e., if Q has two boundary circles), then ∂P is contained either in the annulus component of $T - \partial Q$ (i.e., in the component with two boundary circle), or in the union of the two disk components of $T - \partial Q$ (i.e., of the components with one boundary circle).

(c) Colour connected components of $T - \partial Q$ in black and white so that adjacent components have different colours. Then ∂P is contained in the union of same coloured components of $T - \partial Q$.

The sign $\sqcup$ means a union of disjoint sets.

3.5. Suppose that S and T are curved spheres such that $S \cap T$ is situated in S as it is shown in figure 9. Denote by A_i the 'exterior' circles, by B the 'big splitting' circle and by C the union of the

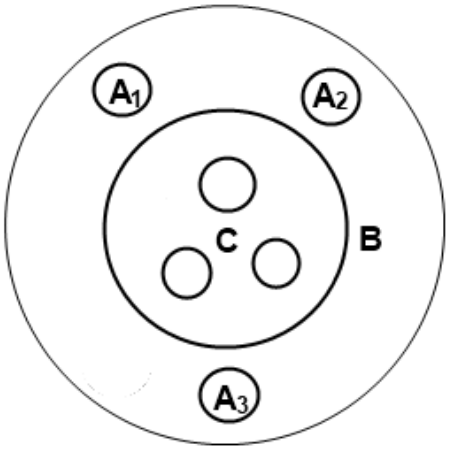


Figure 9: $S \cap T$ in S

'interior' circles, see figure 9.

(a) For each i the union $B \cup C$ is on the same side of A_i in T.

(b) The union $B \cup C$ is in the same connected component of $T - \sqcup_i A_i$.

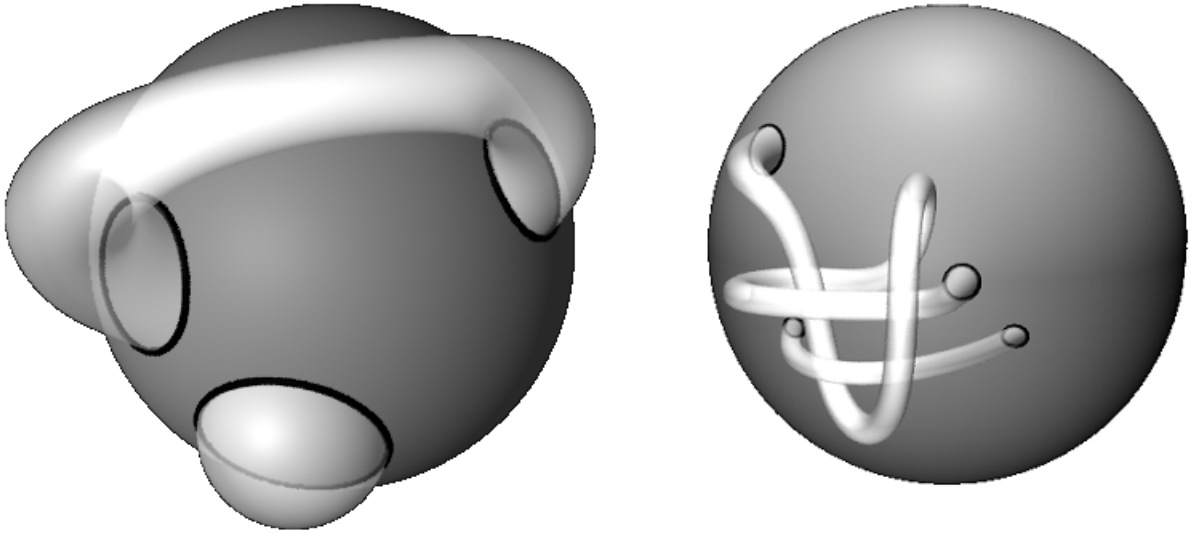


Figure 10: Disjoint curved disk and curved cylinder outside a ball (left), disjoint curved cylinders, one of them knotted, outside a ball (right)

3.6. Embedding Extension Problem. (a) Each two disjoint circles in the unit sphere bound disjoint disks inside this sphere.

(b) For which three disjoint circles p, q_1, q_2 in the unit sphere there exist disjoint *curved disks* P and *curved cylinder* Q inside this sphere such that $\partial P = p$ and $\partial Q = q_1 \sqcup q_2$? (Figure 10 left.)

(c) For which four disjoint circles p_1, p_2, q_1, q_2 in the unit sphere there exist disjoint curved cylinders P and Q inside this sphere such that $\partial P = p_1 \sqcup p_2$ and $\partial Q = q_1 \sqcup q_2$? (Figure 10 right.)

(d) For which two disjoint families p, q of disjoint circles in the unit sphere there exist disjoint curved *spheres with holes* P and Q inside this sphere such that $\partial P = p$ and $\partial Q = q$?

(e) Does there exist three disjoint families p, q, r of disjoint circles in the unit sphere such that

• each of the three pairs (p, q), (q, r) and (p, r) is extendable (to disjoint curved spheres with holes) in the sense of (d);

• there are no disjoint curved spheres with holes P, Q and R inside this sphere such that $\partial P = p$, $\partial Q = q$ and $\partial R = r$? [4]

(f) For which m disjoint families $p_1, \dots, p_m$ of disjoint circles in the unit sphere there exist disjoint curved spheres with holes $P_1, \dots, P_m$ inside this sphere such that $\partial P_i = p_i$ for each $i = 1, \dots, m$?

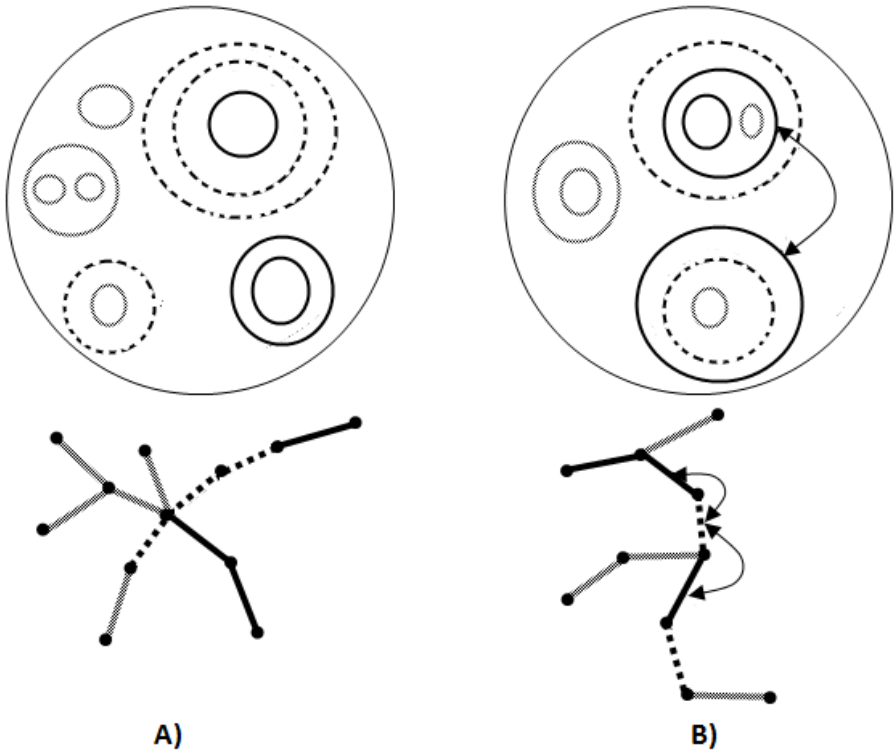


Figure 11: (A): the dotted and the bold unions of circles are unlinked.
(B): the dotted and the bold unions of circles are not unlinked because the arrowed path between two bold circles intersects the dotted circles in an odd number (one) of points.

Suppose that S and T are curved spheres such that each component of $S - T$ except one have one neighbor. (The 'exceptional' component may have one or more neighbors.) This 'exceptional' component is called *a curved sphere with holes*. A *curved disk* is a curved sphere with 1 hole (=with 1 neighbor). A *curved cylinder* is a curved sphere with 2 holes (=with 2 neighbors).

Let M and N be two unions of disjoint circles in the unit sphere S. Colour connected components of $S - q$ in black and white so that adjacent components have different colours. Union M is *on the same side the same side* (in this sphere) of N if M is contained in the union of same coloured components of $S - N$. Unions M and N are *unlinked* (in this sphere) if M is on the same side of N and N is on the same side of M. See figure 11.

3.7. (a) There are two unions M and N of disjoint circles in a sphere such that M is on the same side of N but N is not on the same side of M.

(b) Is unlinkedness transitive? That is, if M and N, N and P are unlinked, are then M and P necessarily unlinked?

(c) For a union M of disjoint circles in the unit sphere S denote by $\mathring{M}$ the unions of black connected components of $S - M$. (There are two choices of $\mathring{M}$ for given M; one of them is the complement to the other.) Two unions M and N are unlinked if and only if for each black and white colourings for M and for N such that $\mathring{M} \cup \mathring{N} \neq S$ we have either $\mathring{M} \subset \mathring{N}$ or $\mathring{N} \subset \mathring{M}$ or $\mathring{M} \cap \mathring{N} = \emptyset$.

[4] This should be compared with the well-known Borromean rings example.

4 Intermediate finish. Some solutions and new problems

1.1 and 1.2. Analogously to solution of Problem 2.4.

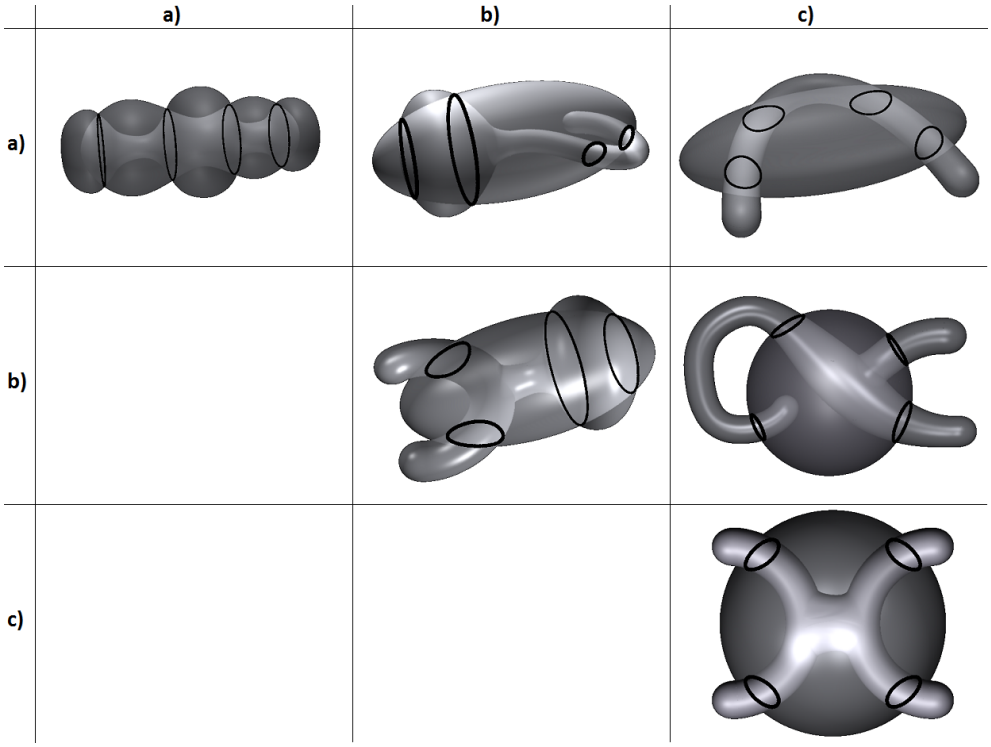


Figure 12: To the solution of Problem 1.3.

1.3. The case $i = j$ follows analogously to Problem 2.4. The cases ab, ac and bc are shown in figure 12.

1.4. (a) The answer is given by the answer to Problem 4.5.

(b) Such an algorithm is given by the answer to Problem 4.5. (Clearly, it is not polynomial.)

1.5. Theorem 1. *Let n be a positive integer and $\vec{x} = (x_1, x_2, \dots, x_n)$, $\vec{y} = (y_1, y_2, \dots, y_n)$ be sequences of positive integers. There exist curved spheres S, T in 3-space whose intersection consists of $n-1$ circles and splits*

- *S into n connected components which can be numbered so that the i-th connected component has x_i neighbors in S, and*
- *T into n connected components which can be numbered so that the i-th connected component has y_i neighbors in T*

if and only if $x_1 + \dots + x_n = y_1 + \dots + y_n = 2n - 2$.

This follows from Theorem 1' (Problem 4.3) below.

2.1. *Answer.* Pairs $(\vec{x}, \vec{x})$ are realizable for $\vec{x}$ up to reordering equal to

$$(1,1), (2,1,1), (3,1,1,1), (2,2,1,1), (4,1,1,1,1), (3,2,1,1,1), (2,2,2,1,1).$$

Other realizable pairs are up to reordering pairs of two sequences of the same number of elements from this list.

2.2. (a) Suppose that curved spheres S, T realize pair $(\vec{x}, \vec{y})$. Recall definition of a graph $G = G(S, S \cap T)$ from §1. The number of the vertices is n. Out of k-th vertex there issues x_k edges. Hence the number of the edges is $(x_1 + \cdots + x_n)/2$. It is obvious that G is connected. By the Jordan Curve Theorem [5] G is split by any edge. So G is a tree. Hence the number of edges is $n - 1 = (x_1 + \cdots + x_n)/2$. Analogously $n - 1 = (y_1 + \cdots + y_n)/2$. QED

Sketch of an alternative solution of (a) by T. Nowik. By induction on the number of circles. The statement is true for one circle (there are only 2 disks on each sphere hence $n = 2$). Each additional circle splits one connected component into two, and adds two boundary circles.

(b) Clealy G is connected. By the Jordan Curve Theorem G is split by any edge. So G is a tree.

2.3. If the number of units is s, then $2n-2 = x_1+\cdots+x_n \geq x_1+2(n-1-s)+s = 2n-2+x_1-s$. So $s \geq x_1$.

2.4. Let S be the unite cube. Take a family M of circles on S 'realizing' $\vec{x}$. (The existence of such a family is proved by induction; the inductive step is proved using deletion of a hanging vertex.) Color the complements in S to these circles into black and white so that neighboring components have different colors. Take a sphere T close to S and such that $S \cap T = M$, each black component of T is inside S, and each white component of T is outside S. Then S, T realize $(\vec{x}, \vec{x})$.

2.5. By Problem 2.3 $x_1 \leq s$. Then

$$x_{n-y_1+1} = x_{n-y_1+2} = \cdots = x_{n-y_1+1} = \cdots = x_n = y_{n-y_1+1} = y_{n-y_1+2} = \cdots = y_n = 1.$$

$$\text{Hence} \quad (\sum_{i=1}^{n-y_1+1} x_i) - y_1 + 1 = (\sum_{i=1}^{n} x_i) - y_1 + 1 - (y_1 - 1) = 2(n - y_1 + 1) - 2$$

$$\text{and} \quad (\sum_{i=2}^{n-y_1+2} y_i) = (\sum_{i=1}^{n} y_i) - y_1 - (y_1 - 2) = 2(n - y_1 + 1) - 2.$$

So the new sequences are tree-like.

2.6. *Answer:* each pair.

(a) Induction on n. The inductive base $n = 2$ is clear. Suppose that the statement is true for each $n < k$, let us prove it for $n = k$.

If sequences $\vec{u} = (u_1, \ldots, u_k)$ and $\vec{v} = (v_1, \ldots, v_k)$ are the same up to reordering, then pair $(\vec{u}, \vec{v})$ is realizable by Problem 2.4. If sequences $\vec{u}$ and $\vec{v}$ are not the same up to reordering, then reorder the sequences so that $u_1 \neq v_1$ and the units are at the end of the sequences. Without loss of generality $u_1 < v_1$. Denote

$$a := u_1, \quad n := k - u_1 + 1, \quad x_i = u_{i+1} \text{ for } i = 1, \ldots, n, \quad y_1 := v_1 - u_1 + 1, \quad y_i = v_i \text{ for } i = 2, \ldots, n.$$

By Problem 2.3 $u_i = 1$ for each $i \geq n + 2$ and $v_i = 1$ for each $i \geq n + 1$. Thus pair $(\vec{u}, \vec{v})$ is obtained from pair $(\vec{x}, \vec{y})$ by given transformation. We have

$$x_1 + \cdots + x_n = u_2 + \cdots + u_{n+1} = 2k - 2 - u_1 - (u_1 - 2) = 2n - 2 \quad \text{and}$$

$$y_1 + \cdots + y_n = v_1 + \cdots + v_n - (u_1 - 1) = 2k - 2 - (u_1 - 1) - (u_1 - 1) = 2n - 2.$$

So the sequences $\vec{x}$ and $\vec{y}$ are tree-like. Since $u_1 > 1$, we have $n < k$. Inductive step is proved.

(b) By induction on n.

[5] **Jordan Curve Theorem.** *A circle on a sphere splits the sphere into exactly two parts. Two points of the sphere not lying on the circle both lie in the same part if and only if they can be connected them by (spherical) broken line not intersecting the circle.*

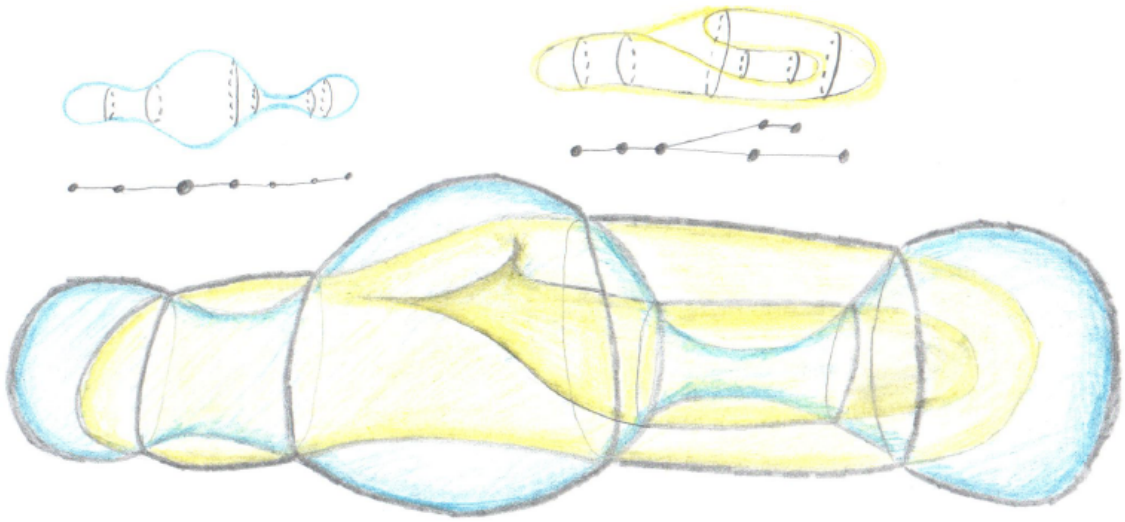


Figure 13: Two spheres realizing the pair in figure 13.

3.1. (a) Follows from Problems 1.1 and 1.3.

(b) Yes. See figure 13. An alternative construction is as follows. Let S and T' be the curved spheres from figure 2. Let T'' be a sphere inside T' 'close and parallel' to T'. Take the connected component X of $\mathbb{R}^3 - S - T'$ lying inside T' whose boundary contains a disk connected component of $T' - S$. Let T be a curved sphere obtained by joining T' and T'' by a tube in X. Then S and T are as required.

(c) No. This fact is obtained using a computer program based on solution of Problem 4.5.

(d) An example is shown in figure 14. Hint: use Problem 3.5 (or Problem 3.4 in the form 4.7 below).

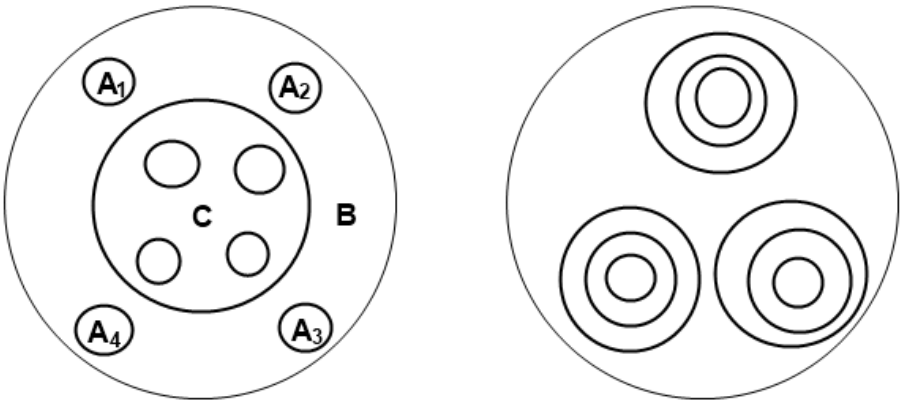


Figure 14: Nine circles (thick) situated in a sphere (thin) in two different ways (left, right).

3.2. (a) *Answer:* each collection is. *Hint.* Take n disjoint spheres intersecting given sphere S at the n circles of given collection M. Connect them by $n-1$ disjoint tubes ('along a tree') inside S to obtain sphere T. Check that T is as required.

(b) *Conjecture.* The pair of such a graph and a tree is realizable if and only if this tree is the union of two trees with n and k edges intersecting by exactly one edge.

(c) *Conjecture.* Any collection of n circles is realizable together with the collection of n 'parallel' circles.

3.3. Colour in black and white the connected components inside and outside T, respectively.

3.4. (a) and (b) are intuitively obvious and follow by (c).

(c) We may assume that T is a round sphere and that circles of ∂Q are round circles, none of them being an equator. For each circle of ∂Q take the round sphere passing through this circle and the center of T. The union of Q and the parts of such spheres lying outside T is a curved sphere, say, Q'. Sphere Q' splits $\mathbb{R}^3$ into two connected components. Since Q is connected, intersection of both connected components with the interior of T is connected. These connected components intersect T by black and white parts, respectively. Since P lies in one of the components, ∂P either lies in black part of $T - \partial Q$ or lies in white part of $T - \partial Q$.

3.5. (a) is analogous to (b).

(b) Consider disks $\overline{A_1}, \overline{A_2}, \overline{A_3} \subset S$ bounded by $A_1, A_2, A_3 \subset S$ and not containing other circles. Without loss of generality we may assume that the interiors of these disks lie inside T. Then the interior of the component of $S - M$ bounded by $B \cup C$ lies inside T as well (because the intersection $S \cap T$ is transversal). This interior lies in one of the connected components of $\mathbb{R}^3 - (T \cup \overline{A_1} \cup \overline{A_2} \cup \overline{A_3})$. So all the 4 circles of $B \cup C$ lie in the same connected component of $T - (A_1 \cup A_2 \cup A_3)$.

Note that this is a particular case of Problem 3.4.c (cf. Problem 4.6.a).

3.7. (a) See definitions of A_0, A_+ and A_- in figure 15 and the Example below in §4. Set $\{A_0\}$ is on the same side of set $\{A_+, A_-\}$ but not vice versa.

(b) No.

Some new problems on the Neighbor Sequence Problem.

A pair $\vec{x} = (x_1, x_2, \dots, x_n)$, $\vec{y} = (y_1, y_2, \dots, y_n)$ of sequences of positive integers is called *strongly realizable* if there exist two curved spheres S, T

(1) whose intersection consists of $n - 1$ circles and splits

• S into n connected components which can be numbered so that the i-th connected component has x_i neighbors in S, and

• T into n connected components which can be numbered so that the i-th connected component has y_i neighbors in T;

(2) there is a circle of $S \cap T$ that bounds a disk and a component with x_1 neighbors in $S - T$, as well as bounds a disk and a component with y_1 neighbors in $T - S$.

Pair (S, T) of spheres is called a *strong realization* of pair $(\vec{x}, \vec{y})$.

4.1. Let $\vec{x}, \vec{y}$ be tree-like sequences in which all the units are situated at the end. If pair of sequences $\vec{x}' := (x_1 - y_1 + 1, x_2, x_3, \dots, x_{n-y_1+1})$, $\vec{y}' := (y_2, y_3, \dots, y_{n-y_1+2})$ is strongly realizable, then pair $(\vec{x}, \vec{y})$ is strongly realizable.

4.2. If pair $(\vec{x}, \vec{y})$ is strongly realizable, then for each positive integer a pairs $(\vec{x'}, \vec{y'})$ are strongly realizable for:

(a) $\vec{x'} = (a, x_1, x_2, \dots, x_n, 1, 1, \dots, 1)$, $\vec{y'} = (y_1 + a - 1, y_2, y_3, \dots, y_n, 1, 1, \dots, 1)$.

(The number of new 1's is $a - 1$ for $\vec{y'}$ and is $a - 2$ for $\vec{x'}$; number a can be different for different changes.)

(b) some choice of $\vec{x'} \in \{(1, x_1 + 1, x_2, x_3, \dots, x_n), (x_1 + 1, x_2, x_3, \dots, x_n, 1)\}$ and $\vec{y'} \in \{(1, y_1 + 1, y_2, y_3, \dots, y_n), (y_1 + 1, y_2, y_3, \dots, y_n, 1)\}$.

4.3. Theorem 1'. *A pair of sequences is strongly realizable if and only if both sequences are tree-like.*

Hint: Use Problems 2.5 and 4.1 (or, alternatively, Problems 2.6 and 4.2).

Some new problems on the Lando Problem.

4.4. Each pair of unions of

(5) 5; (6) 6;

disjoint circles is realizable.

The Lando problem is solved via solution of its *numbered*, or *colored*, analogue. Let us introduce definitions necessary to formulate the analogue.

Assume that M and N are two sets of disjoint circles in spheres S and T, and that in each set the circles are numbered by $1, 2, \dots, n$. Pair (M, N) is *realizable* if there exist two curved spheres S' and T' intersecting transversely by a finite union $S' \cap T'$ of disjoint circles, and a numbering of these circles such that

- $S' \cap T'$ in S' and M in S are numbered equivalent;
- $S' \cap T'$ in T' and N in T are numbered equivalent.

(Numbered sets M in S and N in T are *numbered equivalent* if there is a 1-1 correspondence between connected components of $S - M$ and of $T - N$ such that two connected components of $S - M$ are adjacent along a circle of M if and only if the two corresponding connected components of $T - N$ are adjacent along the corresponding circle of N.)

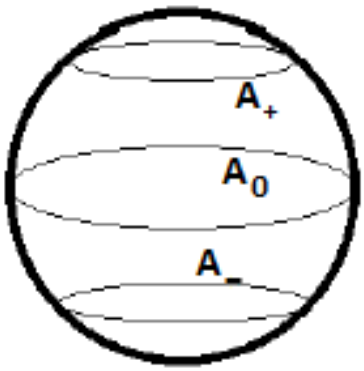


Figure 15: Numbered sets (A_-, A_0, A_+) and (A_0, A_-, A_+) are not numbered equivalent.

Example. On the unit sphere (or on the Earth sphere) let $A_0 = A_2$ be the equator, $A_+ = A_1$ the parallel of sixty degrees northern latitude, $A_- = A_3$ the parallel of sixty degrees southern latitude. See figure 15. Then

- unnumbered (or, equivalently, unordered) sets $\{A_+, A_0, A_-\}$ and $\{A_0, A_+, A_-\}$ are the same (or unnumbered equivalent).
- numbered (or, equivalently, ordered) sets (A_-, A_0, A_+) and (A_+, A_0, A_-) are numbered equivalent.
- numbered sets (A_+, A_0, A_-) and (A_0, A_+, A_-) are not numbered equivalent.
- pair $(A_+, A_0, A_-), (A_0, A_+, A_-)$ of numbered sets in the unit sphere S and in its copy T is non-realizable.

Proof of the last assertion. Suppose to the contrary that there are spheres S' and T' realizing given pair. Denote by

- B_k the copy A_k on the copy T of S;
- A'_k the circle on S' corresponding to A_k;
- B'_k the circle on T' corresponding to B_k;
- $D' \subset S'$ the disk in $S' - T'$ bounded by A'_+;
- $C' \subset S'$ the cylinder in $S' - T'$ bounded by A'_0 and A'_-.

See figure 16. Since S' and T' realize pair $(A_+, A_0, A_-), (B_0, B_+, B_-)$, we have $A'_+ = B'_0$, $A'_0 = B'_+$ and $A'_- = B'_-$.

Clearly, C' and D' lie in 3-space on the same side from sphere T'. (Cf. Problem 3.3.) We have $\partial D = A'_+ = B'_0$. The boundary $\partial C' = A'_0 \sqcup A'_- = B'_+ \sqcup B'_-$ does not lie in one component of $T' - \partial D' = T' - B'_0$. This contradicts to the assertion of Problem 3.4.a for $P = C'$ and $Q = D'$. QED

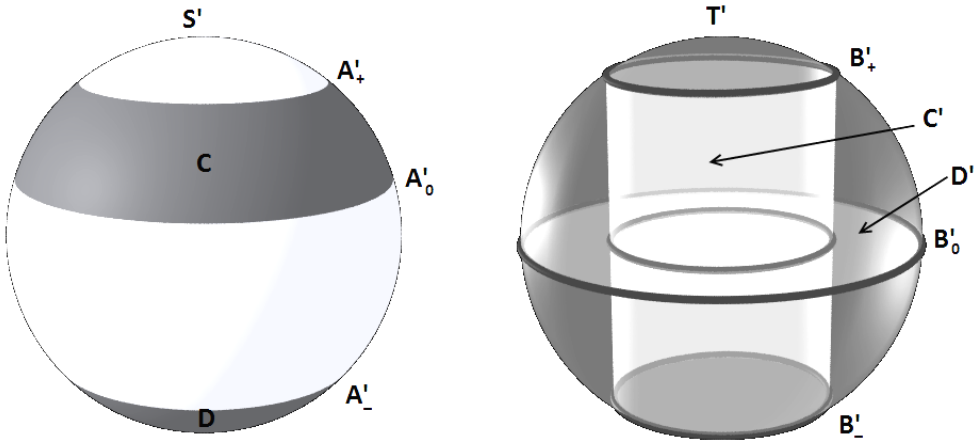


Figure 16: Curved spheres S' and T' drawn apart

4.5. The Numbered Lando Problem. Which pairs of disjoint unions of numbered circles are realizable?

4.6. (a) If pair (M, N) of disjoint unions of numbered circles in spheres S and T is realizable, then connected components of $S - M$ can be colored in black and white so that for each two same coloured components P and Q of $S - M$ unions in T corresponding to ∂P and ∂Q are unlinked in T.

(b) Does the converse to (a) hold?

Let p and q be two sets of edges of a tree G. Colour connected components of the complement in G to the interiors of edges of q. Set p is *on the same side* (in this tree G) of q if p is contained in the union of same-coloured connected components of $G - q$ (or, equivalently, if $p \cap q = \emptyset$ and for each two vertices of edges of p there is a path in the tree connecting these two points, and containing an even number of edges of q). Sets p and q are *unlinked* (in this tree) if p is on the same side of q and q is on the same side of p.

For a vertex P of a graph denote by δP the union of edges issuing out of P.

Graphs $G(S, M)$ and $G(T, N)$ are defined in §1. Numberings of circles in M and of circles in N give numberings of edges in $G(S, M)$ and of edges in $G(T, N)$.

4.7. (a) If pair (M, N) of disjoint unions of numbered circles in spheres S and T is realizable, then the vertices of $G(S, M)$ can be colored in black and white so that for each two same coloured vertices P, Q of $G(S, M)$ the unions in $G(T, N)$ corresponding to δP and δQ are unlinked in $G(T, N)$.

(b)* Given two trees G and G' having the same number of edges, is there a polynomial algorithm for checking the existence of numberings of their edges such that the vertices of G can be colored in black and white so that for each two same coloured vertices P, Q of G the unions in G' corresponding to δP and δQ are unlinked in G'?

5 More spheres and spheres with handles

Let n_1, n_2, n_3 be positive integers. A triple

$$\vec{x_1} = (x_{11}, x_{12}, \ldots, x_{1n_1}), \quad \vec{x_2} = (x_{21}, x_{22}, \ldots, x_{2n_2}), \quad \vec{x_3} = (x_{31}, x_{32}, \ldots, x_{3n_3})$$

of sequences of positive integers is called *realizable* if there exist three curved spheres S_1, S_2, S_3 in 3-space pairwise intersecting by circles and such that $S_1 \cap S_2 \cap S_3 = \varnothing$ and for each $k = 1, 2, 3$ the complement $S_k - S_{k+1} - S_{k+2}$ has n_k connected components which can be numbered so that the i-th connected component has x_{ki} neighbors in S_k for each $i = 1, \ldots, n_k$.

In this section (and in corresponding solutions) subscripts $k, k+1, k+2$ are considered mod 3.

Triple (S_1, S_2, S_3) of spheres is called a *realization* of triple $(\vec{x_1}, \vec{x_2}, \vec{x_3})$.

5.1. Triple Neighbor Sequence Problem. Which triples of tree-like sequences are realizable?

5.2. Which triples of tree-like sequences, each having at most 4 numbers, are realizable?

5.3. If a triple of sequences of lengths n_1, n_2, n_3 is realizable, then
(a) $n_1 + n_2 + n_3$ is odd;
(b) $n_k < n_{k+1} + n_{k+2}$ for each $k = 1, 2, 3$.

5.4. Let $x_1 \geq x_2 \geq \cdots \geq x_n$ be a tree-like sequence. Let p, q be positive integers such that $p \geq q > 1$ and $p + q = n + 1$. Then there exist two tree-like sequences $a_1, a_2, \ldots, a_p$ and $b_1, b_2, \ldots, b_q$ such that $a_1 + b_1 = x_1$ and ordered sets $(a_2, a_3, \ldots, a_p, b_2, b_3, \ldots, b_q)$ and $(x_2, x_3, \ldots, x_n)$ are the same up to reordering.

What are analogs of characterizations of neighbor sequences (of Theorems 1 and 2) for intersections of *more than three* curved spheres?

5.5. * **Conjecture.** Let $n_1, n_2, n_3, \ldots, n_s$ be positive integers and

$$x_{11}, x_{12}, \ldots, x_{1n_1}, \quad x_{21}, x_{22}, \ldots, x_{2n_2}, \quad \ldots, \quad x_{s1}, x_{s2}, \ldots, x_{sn_s}$$

sequences of positive integers. There exist s curved spheres $S_1, S_2, \ldots, S_s$ pairwise intersecting by circles and such that

• no three of them intersect;

• for each $k = 1, \ldots, s$ and $j = 1, \ldots, n_k$ the complement $S_k - S_{k+1} - S_{k+2} - \cdots - S_{k+s-1}$ has n_k connected components, of which the j-th has x_{kj} neighbors in S_k;

if and only if each of s sequences is tree-like, and $n_1 + n_2 + \cdots + n_s - s$ is an even number greater or equal to $2n_k$ for each $k = 1, \ldots, s$.

For $s < 4$ this conjecture is proved (see Theorems 1 and 2), the first unknown case is $s = 4$.

What can be neighbor sequences if there can be 'triple points', i.e. intersection points of of three spheres?

5.6. * **Conjecture.** *Let n_1, n_2, n_3 be positive integers and*

$$x_{11}, x_{12}, \ldots, x_{1n_1}, \quad x_{21}, x_{22}, \ldots, x_{2n_2}, \quad x_{31}, x_{32}, \ldots, x_{3n_3}$$

be sequences of positive integers. Then there exist curved spheres S_1, S_2, S_3 in 3-space

• *pairwise intersecting by circles,*

• *having $2T$ triple intersection points and*

• *such that for each $k = 1, 2, 3$ the complement $S_k - S_{k+1} - S_{k+2}$ has n_k connected components and the i-th connected component has x_{ki} neighbors in S_k for each $i = 1, \ldots, k$*

if and only if $n_1 + n_2 + n_3 + T$ is odd, $x_{k1} + x_{k2} + \cdots + x_{kn_k} = 2n_k - 2 + 2T$ and $n_k + T < n_{k+1} + n_{k+2}$ for each k.

What are analogs of Theorems 1 and 2 for intersections of *curved spheres with handles*?

5.7. * **Conjecture.** *Let g_1, g_2, n be positive integers and*

$$x_{11}, x_{12}, \ldots, x_{1n}, \quad x_{21}, x_{22}, \ldots, x_{2n}$$

two sequences of positive integers. There exist curved sphere with g_1 handles S_1 and curved sphere with g_2 handles S_2 such that they intersect by circles splitting S_k into n connected components, of which the j-th has x_{kj} neighbors in S_k for each $k = 1, 2$ and $j = 1, \ldots n$

if and only if $s := x_{11} + x_{12} + \cdots + x_{1n} = x_{21} + x_{22} + \cdots + x_{2n}$ is even and $2n - 2 \leq s \leq 2n - 2 + 2g_k$ for each $k = 1, 2$.

It would be interesting to solve analogous problems in case when self-intersections are allowed. Both cases are interesting — either with triple self-intersection points or without them.

6 Solutions after finish

In this section curved spheres are shortly called *spheres*.

Neighbor sequence problem

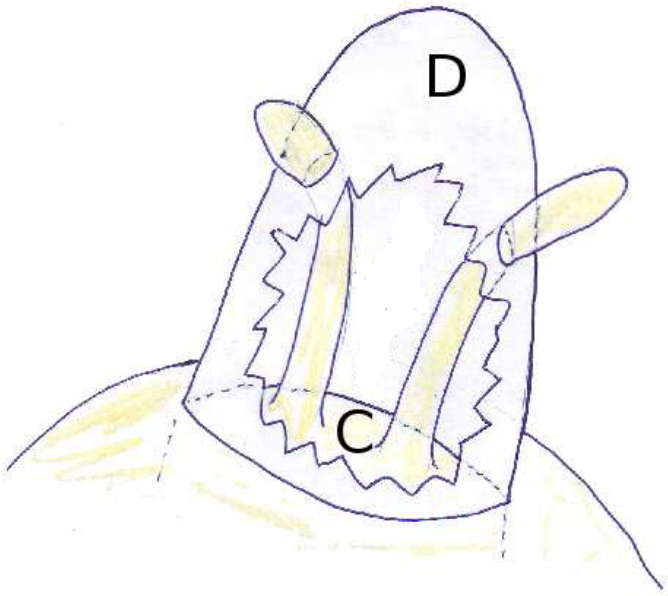


Figure 17: Inductive construction

4.1. We may assume that $x_1 \geq y_1$. Take spheres S', T' realizing pair $(\vec{x}', \vec{y}')$ of sequences. Take a circle of $S' \cap T'$ from condition (2). This circle that bounds

• in $S' - T'$ a connected component, say C, that has $x_1 - y_1 + 1$ neighbors,

• in $T' - S'$ a disk, say D.

We modify spheres S', T' by joining C and D by $y_1 - 1$ fingers, see Figure 17. Denote the new spheres by S and T. Let us prove that they realize pair $(\vec{x}, \vec{y})$ of sequences.

Condition (1) is satisfied for S, T because

• each component of $S' - T'$ except C is also a component of $S - T$,

• C is separated by $y_1 - 1$ circles of $(S \cap T) - (S' \cap T')$ into $y_1 - 1$ disks and a component with $(x_1 - y_1 + 1) + (y_1 - 1) = x_1$ neighbors.

and

• each component of $T' - S'$ except D is also a component of $T - S$,

• D is separated by $y_1 - 1$ circles into $y_1 - 1$ disks and a component with y_1 neighbors.

Any circle of $(S \cap T) - (S' \cap T')$ satisfies condition (2).

4.3. *Proof of Theorem 1'.* Proof by induction on the length n of the sequences. For each tree-like sequence of n numbers we have $n \geq 2$. The induction base is $n = 2$ and is clear.

Let us prove the induction step. Suppose Theorem 1' is proved for $2, 3, \ldots, n - 1 \geq 2$. Let us prove it for n.

Reorder our sequences so that the 1's will be at the end. By Problems 2.5 and 4.1 and by the induction hypothesis the new sequences are realizable. Take spheres S, T realizing the new sequences. So S, T satisfy condition (1) from the definition of the strong realizability for the old sequences. Also,

• if $x_1 \neq 1$, then conditions (2) for the new and for the old sequences $\vec{x}$ are equivalent;

• if $x_1 = 1$, then the circle from condition (2) for the new sequence $\vec{x}$ bounds a disk, so it bounds a component with $x_1 = 1$ neighbor.

Same holds for $\vec{x}$ replaced by $\vec{y}$. So condition (2) is also satisfied for the old sequences. Thus S, T strongly realize the old sequences.

Lando problem

3.1. (d) *A direct solution.* Assume to the contrary that there exist two spheres S' and T' realizing pair (M, N) from figure 14. Denote the connected components of $S' - T'$ as shown in fig. 14 left.

Without loss of generality we may assume that the interiors of disks $A_1, \dots, A_4 \subset S'$ lie inside T'. Then the interior of component $C \subset S'$ lies inside T' as well (because the intersection of S' and T' is transversal). Since $C, A_1, \dots, A_4$ are disjoint, C lies in one of the connected components of $\mathbb{R}^3 - T' \cup \bigsqcup A_i$. So all the 5 circles of ∂C lie in the same connected component of $T' - \bigsqcup \partial A_i$. (Here we use a trivial particular case of the Embedding Extension Theorem.)

Let us restate the previous statement in terms of graph $G := G(T', N)$ (fig. 18). Denote by $G(C)$ the union of 5 edges of G corresponding to the circles of ∂C. Then $G(C)$ lies completely in one of the connected components of the compliment of G to the 4 edges corresponding to the circles of $\bigsqcup \partial A_i$. Since G has only 9 edges, this means that $G(C)$ is a subtree of G. Denote by $G(B)$ the union of 5 edges of G corresponding to the circles of ∂B. Likewise, $G(B)$ is a subtree of G.

Since $G(B) \cup G(C) = G$, at least two of the three edges a, b, c of G (fig. 18) belong to one of subtrees $G(B)$ or $G(C)$. Without loss of generality we may assume that $a, b \in G(B)$. But any subtree of G containing both a and b has at least 6 edges while $G(B)$ has only 5 edges. Contradiction.

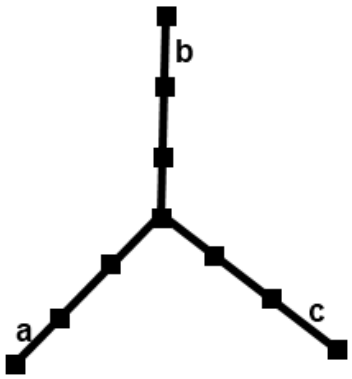


Figure 18: Graph $G := G(T', N)$.

3.6. (a) Clear.

(b) ... if and only if circles q_1 and q_2 are on the same side of p.

(c) ... if and only if $p_1 \sqcup p_2$ and $q_1 \sqcup q_2$ are unlinked.

(d) ... if and only if p and q are unlinked. *Hint:* generalize solution of Problem 3.6.d 3.2.a. *Formal solution* is obtain by taking $m = 2$ in the solution of (f).

(e) No, by the answer to (f).

(f) ... if and only if p_i and p_j are unlinked for each $i \neq j$.

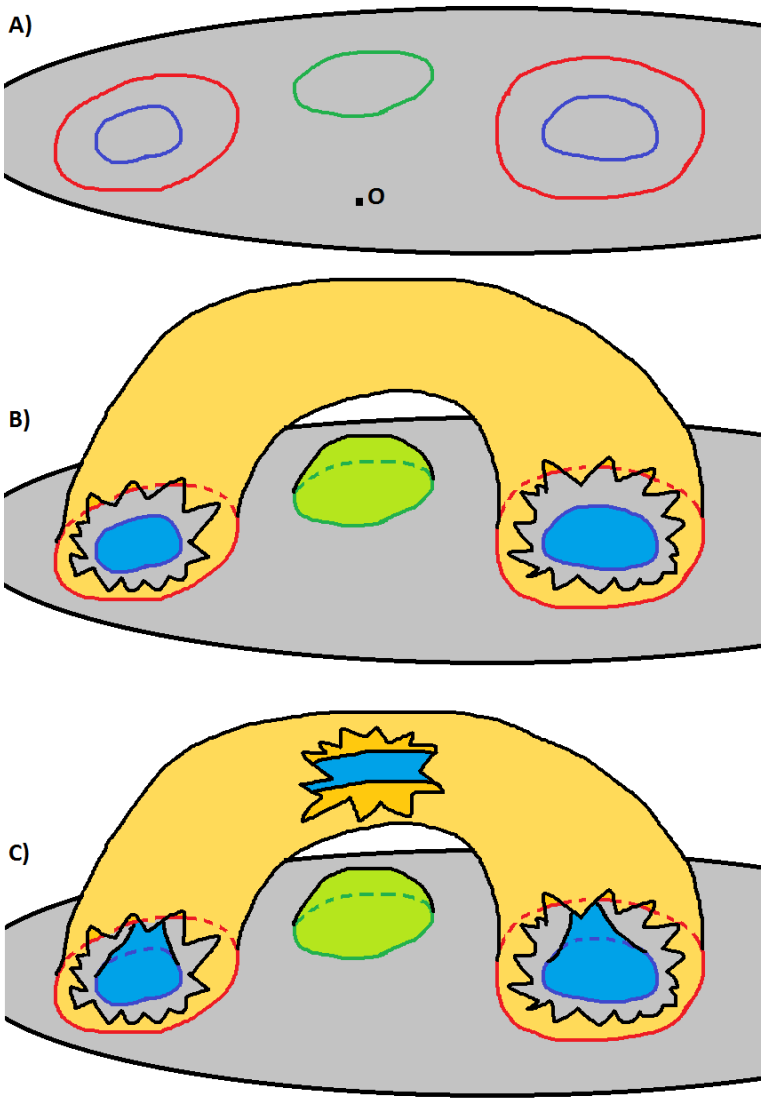


Figure 19: To the solution of Problem 3.6.f. (A) We have S (gray), p_1 (red), p_2 (green), p_3 (blue). (B) We have that $\mathring{p}_3$ (blue) is the 'smallest'. We construct P_1 (yellow) and P_2 (green) by induction. (C) Connected components of $\mathring{p}_3$ (blue) can be connected by a path disjoint with $P_1 \cup P_2$. So we connect them by a tube and obtain P_3 (blue).

Embedding Extension Theorem. *Unions $p_1, \dots, p_m$ of disjoint circles in the unit sphere S are pairwise unlinked if and only if there exist disjoint curved spheres with holes $P_1, \dots, P_m$ whose interiors are inside S and such that $\partial P_i = p_i$ for each $i = 1, \dots, m$.*

Proof. The necessity is essentially proved in Problem 3.4.c. The sufficiency is proved by induction on m. Base $m = 1$ is essentially proved in the solution of Problem 3.2.a. Let us prove the inductive step. Take a point $O \in S - \bigsqcup_{i=1,\dots,m} p_i$. For each i take a black and white colouring of $S - p_i$ such that O is white. Recall that $\mathring{p}_i$ is the union of black components of $S - p_i$. Since O is white and p_i and p_j are unlinked, by Problem 3.7.b for each $i \neq j$ either $\mathring{p}_i \subset \mathring{p}_j$ or $\mathring{p}_j \subset \mathring{p}_i$ or $\mathring{p}_i \cap \mathring{p}_j = \emptyset$. So there is a 'smallest' $\mathring{p}_i$, i.e. $\mathring{p}_i$ such that $\mathring{p}_j \not\subset \mathring{p}_i$ for each $j \neq i$. We may assume that $i = m$. Then $\mathring{p}_m \cap \bigsqcup_{i=1}^{m-1} p_i = \emptyset$, $\mathring{p}_m$ is a collection of curved spheres with holes, $\partial \mathring{p}_m = p_m$ and $\mathring{p}_m \subset S$. Denote by Δ the closed 3-ball bounded by S (i.e., 'the interior part' of S). By the inductive hypothesis there

are disjoint spheres with holes $P_1, \dots, P_{m-1} \subset \Delta$ such that $\partial P_i = p_i$ for each $i = 1, \dots, m-1$.

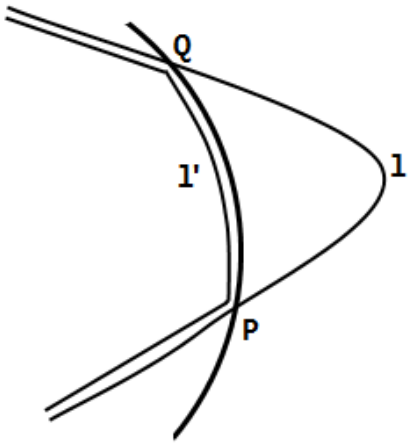


Figure 20: Proof of Claim

Claim. *Union p_m lies in one connected component of $\Delta - (P_1 \sqcup \dots \sqcup P_{m-1})$.* [6]

Proof of Claim. Take any two points $A, B \in p_m$. Denote by l a path inside S connecting A and B such that $\bar{l} := \#(l \cap \bigsqcup_{i=1}^{m-1} P_i)$ is minimal (minimal by l, objects $A, B, p_m, S, P_1, \dots, P_{m-1}$ are fixed). Assume to the contrary that l is not as required, i.e., $\bar{l} > 0$. Since p_m is on the same side of ∂P_i, points A and B are in the same connected component of $\Delta - P_i$, so $\#(l \cap P_i)$ is even for each i. (If $m = 2$, we may even obtain that $\#(l \cap P_1) = 0$ and stop here.) Then $\#(l \cap P_i) \geq 2$ for some i. Denote by Q and R two consecutive points of $l \cap P_i$. Denote by Q' the point of l slightly before Q and by R' the point of l slightly after R. Since P_i is connected, Q and R can be connected by a path in P_i. So Q' and R' can be connected by a path l' very close to P_i but not intersecting P_i. Path l' does not intersect any of $P_1, \dots, P_{m-1}$ because it is very close to P_i and $P_1, \dots, P_{m-1}$ are pairwise disjoint. Substitute the part of l between Q' and R' by l'. Denote the obtained path by l''. Then $\overline{l''} = \bar{l} - 2$. This contradicts to the minimality of $\bar{l}$. Thus l is as required. QED

Completion of the proof of Embedding Extension Theorem. Let $\mathring{p}'_m$ be a disjoint union of curved spheres with holes obtained from $\mathring{p}_m$ by a slight deformation so that the interior of $\mathring{p}'_m$ is inside the interior of Δ and $\partial \mathring{p}'_m = \partial \mathring{p}_m = p_m$. By Claim each two points of $\mathring{p}'_m$ can be connected by a path inside S disjoint with $P_1, \dots, P_{m-1}$. So we can connect all the connected components of $\mathring{p}'_m$ by tubes inside S disjoint with $P_1, \dots, P_{m-1}$. The number of the tubes is one less than the number of the connected components of $\mathring{p}'_m$, so that there are no 'cycles of tubes'. Then we obtain a sphere with holes. Denote it by P_m. We have $\partial P_m = p_m$, $P_m \subset \Delta$ and P_m is disjoint with $P_1, \dots, P_{m-1}$. The inductive step is proved. QED

4.4. This fact is obtained using a computer program based on Theorem 3 below.

4.5. The answer is given by Problem 4.6 and is as follows.

Theorem 3. *Pair (M, N) of disjoint unions of numbered circles in spheres S and T is realizable if and only if connected components of $S - M$ can be colored in black and white so that for each two same coloured components P and Q of $S - M$ unions in N corresponding to ∂P and ∂Q are unlinked in T.*

4.6. (a) This is a restatement of Problem 3.4.

[6] This assertion for $m = 2$ is essentially the definition of the comparability (or, rather, of 'p_2 is on the same side of ∂P_1'). This case $m \geq 3$ is interesting because in general the union of two subsets could split the ambient set even if each subset alone does not split the ambient set.

(b) Yes. The idea is to prove and use the answer to Embedding Extension Problem 3.6.e.

Let T' be the unit cube. Numberings give a 1–1 correspondence h between circles of M and circles of N.

Denote by $A_1, \ldots, A_m$ the white connected components of $S-M$. By the assumption $h(\partial A_1), \ldots, h(\partial A_m)$ are pairwise unlinked in T'. By the answer to Embedding Extension Problem 3.6.e there exist disjoint curved spheres with holes $A'_1, \ldots, A'_m$ whose interiors are *inside* T' and such that $\partial A'_i = h(\partial A_i)$ for each $i = 1, \ldots, m$.

Denote by $B_1, \ldots, B_n$ the black connected components of $S-M$. Analogously there exist disjoint curved spheres with holes $B'_1, \ldots, B'_n$ whose interiors are *outside* T' and such that $\partial B'_i = h(\partial B_i)$ for each $i = 1, \ldots, n$.

Let $S' := (A'_1 \cup \ldots \cup A'_m) \cup (B'_1 \cup \ldots \cup B'_n)$. By construction S' does not have self-intersections. We have that A'_i has the same number of holes as A_i, and B'_i has the same number of holes as B_i. Since $S = (A_1 \cup \ldots \cup A_m) \cup (B_1 \cup \ldots \cup B_n)$ is a curved sphere, S' is a curved sphere. (A rigorous proof is obtained using Euler characteristic.) Clearly, S' and T' realize given pair M, N.

4.7. (a) This is a restatement of Problems 3.4 and 4.6.

More spheres and spheres with handles

5.1. Theorem 2. *Let n_1, n_2, n_3 be positive integers and*

$$x_{11}, x_{12}, \ldots, x_{1n_1}, \quad x_{21}, x_{22}, \ldots, x_{2n_2}, \quad x_{31}, x_{32}, \ldots, x_{3n_3}$$

be sequences of positive integers. There exist curved spheres S_1, S_2, S_3 in 3-space pairwise intersecting by circles and such that

- *$S_1 \cap S_2 \cap S_3 = \varnothing$;*
- *$S_k - S_{k+1} - S_{k+2}$ has n_k connected components, which can be numbered so that the i-th component has x_{ki} neighbors in S_k, for each $k = 1, 2, 3$*

if and only if the sequences are tree-like, $n_1 + n_2 + n_3$ is odd and $n_k < n_{k+1} + n_{k+2}$ for each $k = 1, 2, 3$.

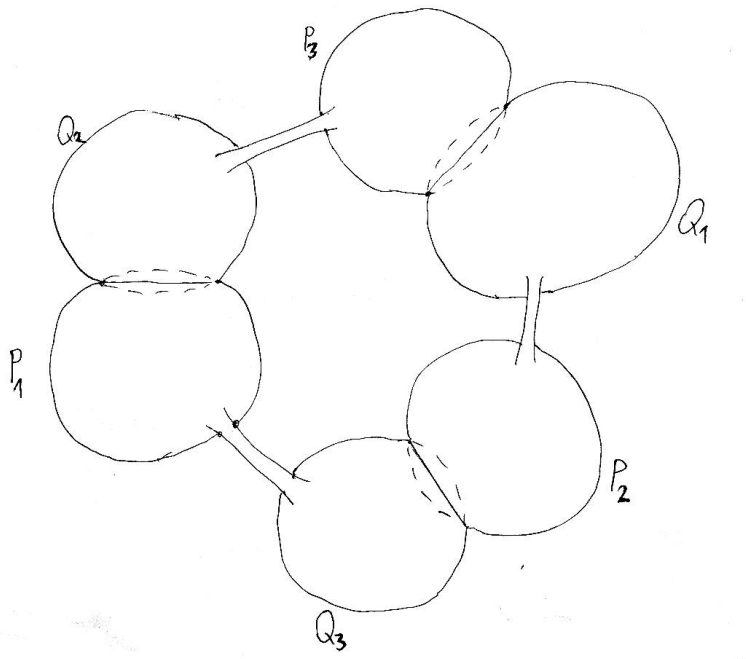


Figure 21: Construction of three spheres

Proof. The 'only if' part follows by Problems 2.2 and 5.3. Let us prove the 'if' part. Let

$$m_1 := (n_2 + n_3 - n_1 + 1)/2, \quad m_2 := (n_1 + n_3 - n_2 + 1)/2, \quad m_3 := (n_1 + n_2 - n_3 + 1)/2.$$

So

$$m_1 + m_2 = n_3 + 1, \quad m_1 + m_3 = n_2 + 1, \quad m_2 + m_3 = n_1 + 1.$$

By Problem 5.4 there exist sequences

$$p_{11}, p_{12}, \dots, p_{1m_3}, \quad p_{21}, p_{22}, \dots, p_{2m_1}, \quad p_{31}, p_{32}, \dots, p_{3m_2},$$

$$q_{11}, q_{12}, \dots, q_{1m_2}, \quad q_{21}, q_{23}, \dots, q_{2m_3}, \quad q_{31}, q_{32}, \dots, q_{3m_1},$$

such that $p_{k-1,1} + q_{k+1,1} = x_{k1}$ and ordered sets

$$(p_{k-1,2}, p_{k-1,3}, \dots, p_{k-1,m_{k+1}}, q_{k+1,2}, q_{k+1,3}, \dots, q_{k+1,m_{k-1}}) \quad \text{and} \quad (x_{k2}, x_{k3}, \dots, x_{kn_k})$$

are the same up to reordering. By Theorem 1' there exist spheres

$$Q_1, P_1, Q_2, P_2, Q_3, P_3 \subset \mathbb{R}^3 \quad \text{such that} \quad Q_k \cap Q_{k+1} = \varnothing, \quad Q_k \cap P_l = \varnothing \quad \text{if} \quad l \neq k-1 \quad \text{and}$$

- $Q_k - P_{k-1}$ is the disjoint union of m_{k+1} connected components, i-th one has q_{ki} neighbors
- $P_{k-1} - Q_k$ is the disjoint union of m_{k+1} connected components, i-th one has $p_{k-1,i}$ neighbors
- the boundary of some connected component of $\mathbb{R}^3 - P_{k-1} - Q_k$ contains a component $\widetilde{q}_k$ with q_{k1} neighbors on Q_k and a component $\widetilde{p}_{k-1}$ with $p_{k-1,1}$ neighbors on P_{k-1}.

For $k = 1, 2, 3$ let S_k be the connected sum of spheres Q_{k+1} and P_{k-1} along a small tube joining the two components $\widetilde{q}_{k+1}$ and $\widetilde{p}_{k-1}$ from the third condition, see Figure 21. This can be done without intersections of the three tubes.

Then $S_k - S_{k+1} - S_{k+2}$ is as required for each $k = 1, 2, 3$. QED.

5.2. *Answer:* these triples are

$$\{(2,1,1),(2,1,1),(2,1,1)\}, \quad \{(3,1,1,1),(3,1,1,1),(2,1,1)\}, \quad \{(3,1,1,1),(2,2,1,1),(2,1,1)\},$$

$$\{(3,1,1,1),(2,1,1),(1,1)\}, \quad \{(2,2,1,1),(2,2,1,1),(2,1,1)\},$$

$$\{(2,2,1,1),(2,1,1),(1,1)\}, \quad \{(2,1,1),(1,1),(1,1)\}.$$

Proof. There exist only 4 tree-like sequences of length at most 4. They are

$$(1,1), \ (2,1,1), \ (3,1,1,1), \ (2,2,1,1).$$

According to Problem 5.3 the number of odd length sequences in a realizable triple is odd. So in each realizable triple of sequences of length at most 4, except triple $\{(2,1,1),(2,1,1),(2,1,1)\}$, there are one sequence $(2,1,1)$ and two sequences of even length. According to the answer to Problem 5.1 all these 7 triples are realizable.

5.3. Let m_3, m_2, m_1 be the numbers of the circles in $f_1 \cap f_2$, $f_1 \cap f_3$ and $f_2 \cap f_3$. Then $n_1 = m_3 + m_2 + 1$, $n_2 = m_3 + m_1 + 1$, $n_3 = m_2 + m_1 + 1$.

So $n_1 + n_2 + n_3 = 2(m_3 + m_2 + m_1) + 3$ is odd.

Since $2m_k + 1 > 0$ we have $n_k < n_{k+1} + n_{k+2}$ for each $k = 1, 2, 3$.

5.4. Let $r = r(\vec{x})$ be the number of those x_i's that are greater than 1. Let $z_s = x_2 + x_3 + \dots + x_s$. For each $s \leq r$ let

$$a_1 = p - (z_s - s + 3) + 1, \quad a_i = x_i \quad \text{for} \quad 2 \leq i \leq s \quad \text{and} \quad a_i = 1 \quad \text{for} \quad s + 1 \leq i \leq p,$$

$$b_1 = x_1 - a_1, \quad b_i = x_{i+s-1} \quad \text{for} \quad 2 \le i \le r-s+1, \quad b_i = 1 \quad \text{for} \quad r-s+2 \le i \le q = n+1-p.$$

Since $s \le r$, the sequence $b_1, b_2, \dots, b_q$ is well-defined. For each i we have that a_i and b_i depend on s.

We have

$$a_1 + a_2 + \dots + a_p = p - (z_s - s + 3) + 1 + z_s + p - s = 2p - 2,$$

i.e. the sequence $a_1, a_2, \dots, a_p$ is tree-like. Also

$$b_1 + b_2 + \dots + b_q = z_n - a_1 - a_2 - \dots - a_p = 2n - 2 - 2p + 2 = 2q - 2,$$

i.e. the sequence $b_1, b_2, \dots, b_q$ is tree-like.

It remains to prove that there exists $s \le r$ such that $1 \le a_1 \le x_1 - 1$. For each $i < r$ we have $x_1 \ge x_i$, so

$$z_i - i + x_1 + 1 \ge (z_{i+1} - (i+1) + 3) - 1.$$

In other words,

$$\begin{gathered} 2 = z_1 - 1 + 3, \\ z_1 - 1 + x_1 + 1 \ge (z_2 - 2 + 3) - 1, \\ z_2 - 2 + x_1 + 1 \ge (z_3 - 3 + 3) - 1, \\ \dots, \\ z_{r-1} - (r-1) + x_1 + 1 \ge (z_r - r + 3) - 1, \\ z_r - r + x_1 + 1 = n - 1. \end{gathered}$$

Here the last equality is not analogous to the previous equalities but follows because sequence $x_1, x_2, \dots, x_n$ is tree-like and $1 = x_{r+1} = \dots = x_n$. Since $2 \le p \le n-1$, there exists $s \le r$ such that

$$z_s - s + 3 \le p \le z_s - s + x_1 + 1 \quad \Leftrightarrow \quad 1 \le a_1 \le x_1 - 1. \quad QED$$

References

[A] S. Avvakumov, *A counterexample to the Lando conjecture on intersection of spheres in 3-space*, preprint, 2012.

[AB] I. Arzhantsev, V. Bogachev, A. Garber, A. Zaslavsky, V. Protasov and A. Skopenkov, *Students' mathematical olympiades at Moscow State University 2010-2011*, Mat. Prosveschenie, 16 (2012), 214-227.

[D] N. P. Dolbilin, *Pearls of polyhedra theory*, M.: MCCME, 2000.

[H] T. Hirasa, *Dissecting the torus by immersions*, Geometriae Dedicata, 145:1 (2010), 33-41

[R] A. Rukhovich, *On intersection of two embedded spheres in 3-space*, http://arxiv.org/abs/1012.0925

[T] T. Nowik, *Dissecting the 2-sphere by immersions*, Geometriae Dedicata 127, (2007), 37-41, http://arxiv.org/abs/math/0612796.

[W] http://en.wikipedia.org/wiki/Polyhedron

HOW DO CURVED SPHERES INTERSECT IN 3-SPACE, OR TWO-DIMENSIONAL MEANDRA

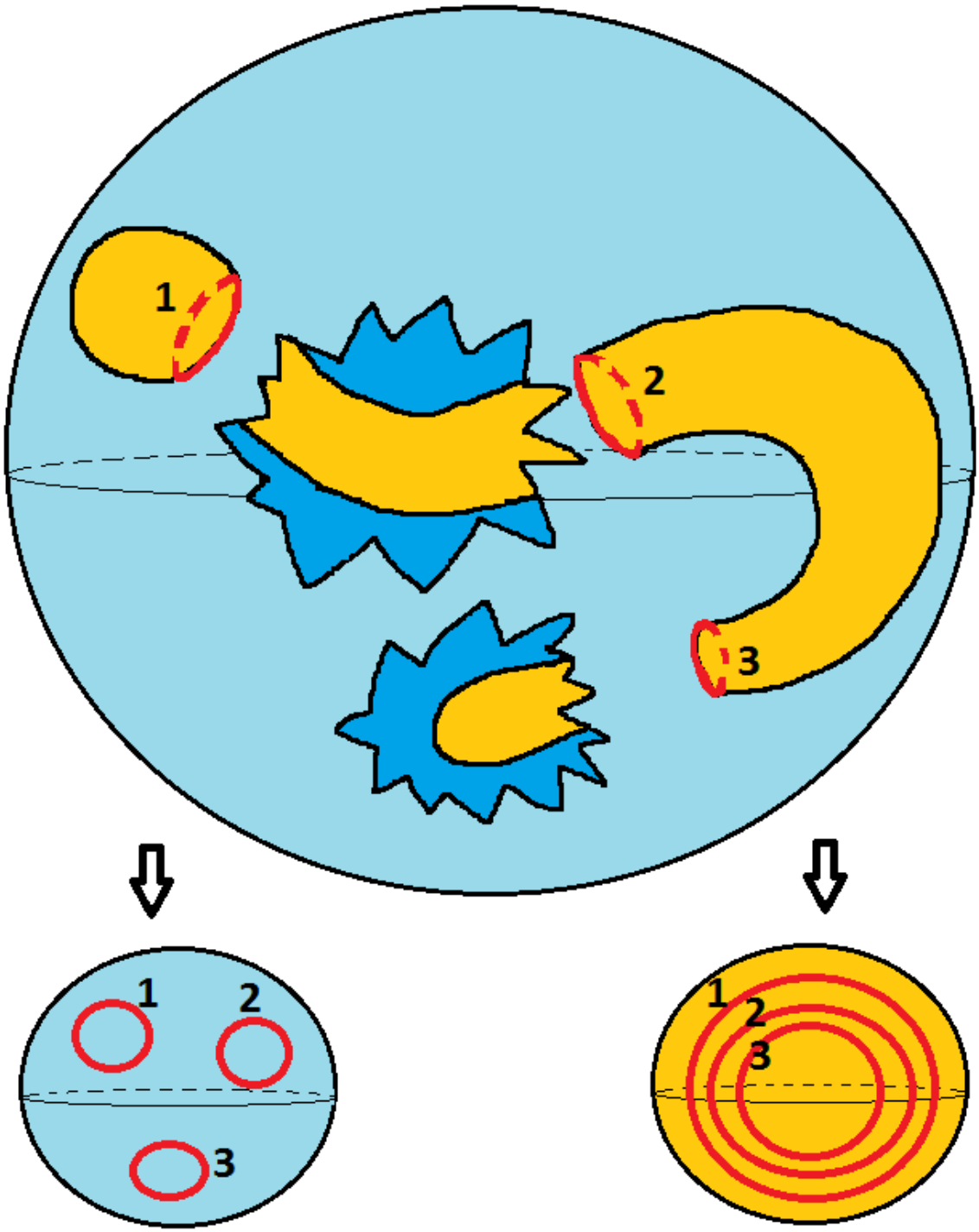


Fig. 2: curved spheres intersecting by three circles

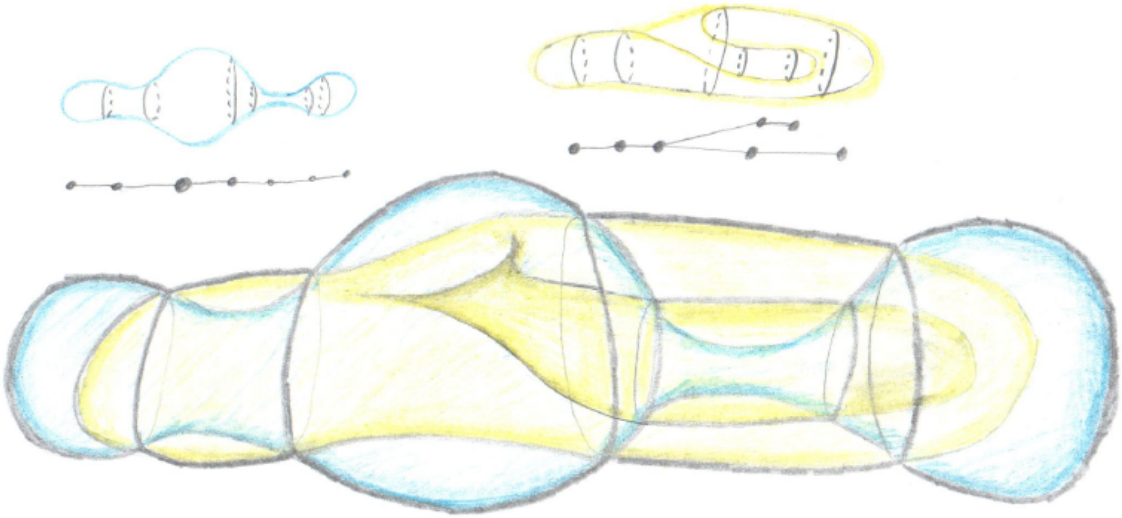


Fig. 13: two spheres realizing the pair in figure 7.

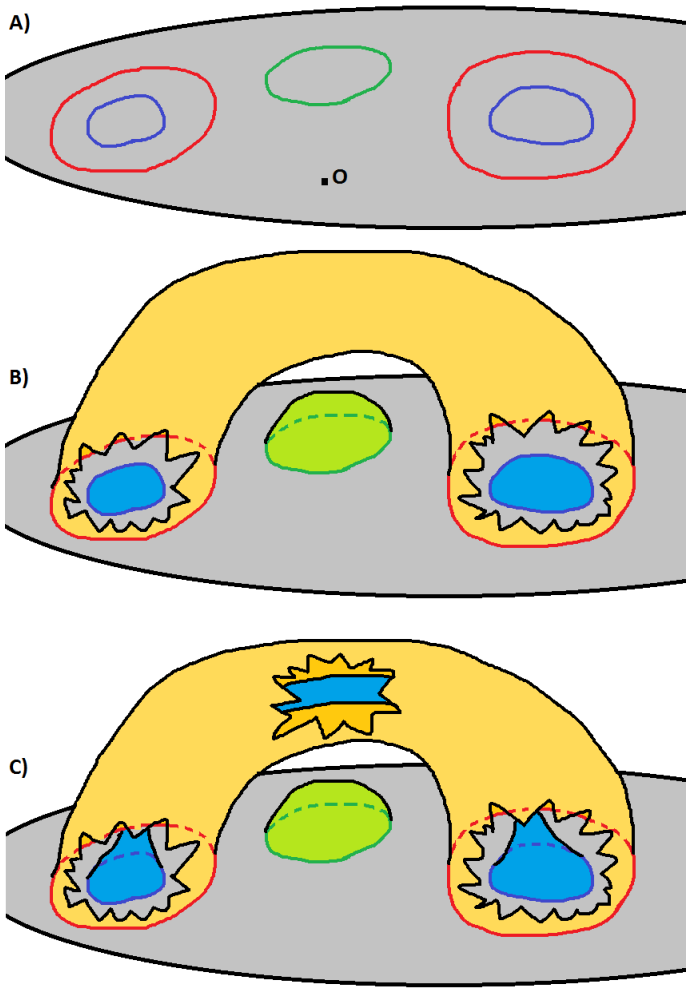


Fig. 18: to the solution of Problem 3.6.f. (A) We have S (gray), p_1 (red), p_2 (green), p_3 (blue). (B) We have that $\mathring{p}_3$ (blue) is the 'smallest'. We construct P_1 (yellow) and P_2 (green) by induction. (C) Connected components of $\mathring{p}_3$ (blue) can be connected by a path disjoint with $P_1 \cup P_2$. So we connect them by a tube and obtain P_3 (blue).

How do curved spheres intersect in 3-space?

Sergey Avvakumov, s.avvakumov@gmail.com

November 2, 2018

Abstract

The following problem was proposed in 2010 by S. Lando.

Let M and N be two unions of the same number of disjoint circles in a sphere. Do there always exist two spheres in 3-space such that their intersection is transversal and is a union of disjoint circles that is situated as M in one sphere and as N in the other? Union M' of disjoint circles is *situated* in one sphere as union M of disjoint circles in the other sphere if there is a homeomorphism between these two spheres which maps M' to M.

We prove (by giving an explicit example) that the answer to this problem is "no". We also prove a necessary and sufficient condition on M and N for existing of such intersecting spheres. This result can be restated in terms of graphs. Such restatement allows for a trivial brute-force algorithm checking the condition for any given M and N. It is an open question if a faster algorithm exist.

The Lando Problem

We work entirely in the piecewise-linear (PL) category [1].

Suppose M and M' are the unions of the same number of disjoint circles in spheres S and S'. Then M *is situated in* S *as* M' *in* S' if there is a homeomorphism $f : S \to S'$ such that $f(M) = M'$.

The following problem suggested by S. Lando was one of the (unsolved) problems at the Moscow State University mathematical tournament for students and young professors 2010 ([1], problem MB-8).

Let M and N be two unions of the same number of disjoint circles in a sphere. Do there exist two spheres in 3-space whose intersection is transversal and is a union of disjoint circles that is situated as M in one sphere and as N in the other?

This problem appeared in the discussion of related papers [3], [4], [5].

In this paper we prove that the answer to Lando problem is "no" by giving an explicit example.

[1] A *PL circle* or *circle* is a closed broken line (polygon) without self-intersections in 3-space. A *PL sphere* or *sphere* is a polyhedron in 3-space (more precisely, 2-dimensional surface of the polyhedron), which is split into several parts by any circle lying on the polyhedron, i.e. is a polyhedron homeomorphic to S^2.

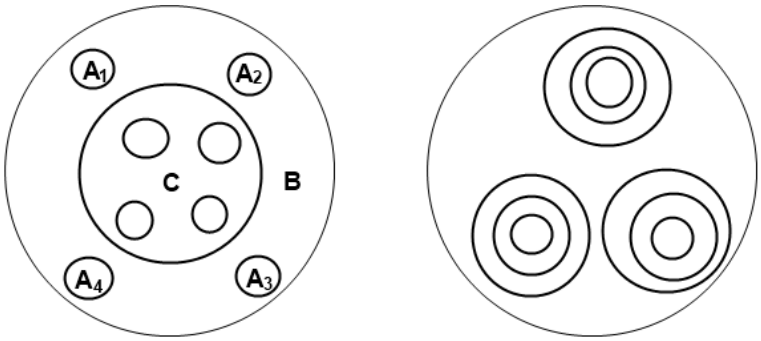


Figure 1: Two unions of M (left) and N (right) of 9 circles.

Theorem 1 (an example)**.** *Let M and N be two unions of 9 disjoint circles in S^2 shown in Fig. 1. Then there are no two spheres in 3-space whose intersection is transversal and is a union of 9 disjoint circles that is situated as M in one sphere and as N in the other.*

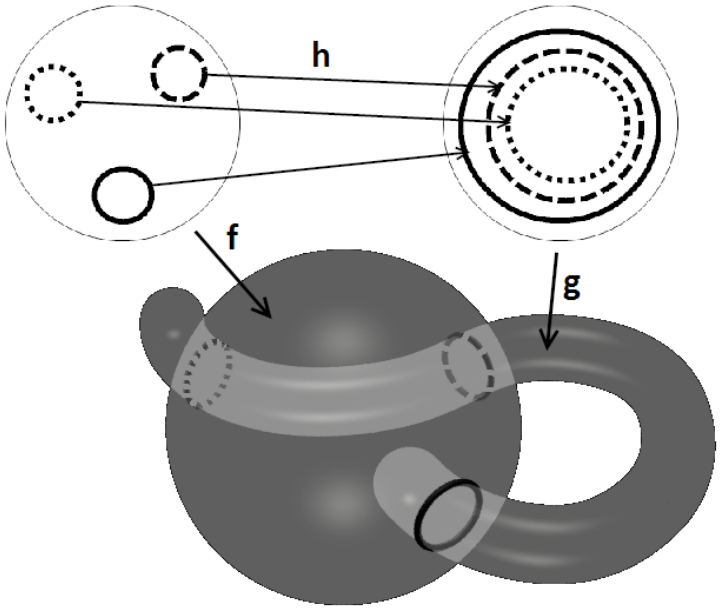


Figure 2: Bijection h between two sets of three circles is realized by PL embeddings f, g.

In Theorem 2 (see below) we describe all the collections of circles which can be realized by two intersecting spheres. The precise meaning of the word "realized" is defined in the following paragraph.

Assume that M and N are two unions of disjoint circles in sphere S^2. Suppose there exists PL embeddings[2] $f : S^2 \hookrightarrow \mathbb{R}^3$ and $g : S^2 \hookrightarrow \mathbb{R}^3$ such that intersection $f(S^2) \cap g(S^2)$ is transversal and $f(S^2) \cap g(S^2) = f(M) = g(N)$. These embeddings induce a bijection h between sets of circles of M and of N (for circles $m \subset M$ and $n \subset N$ let $h(m) = n$ if $f(m) = g(n)$). Equivalently we may number circles of M and of N by $1, \dots, k$ so that two circles corresponding to the same circle of $f(S^2) \cap g(S^2)$ have the same number. We say that f, g *realize* h (Fig. 2).

Theorem 2 (see below) gives a necessary and sufficient condition for the realizability of a bijection. In particular Theorem 2 can be used to prove Theorem 1. The following simple example shows that not every bijection is realizable.

[2]Map $f : A \to B$ is piecewise linear if f is a simplicial map for *some* simplicial decompositions of A and B

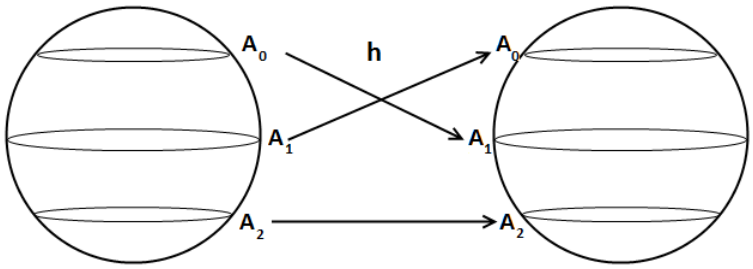


Figure 3: Circles A_0, A_1, A_2, bijection h.

Example 1. Let A_0, A_1, A_2 be the circles situated in S^2 as shown in the Fig. 3. Let $M = N = A_0 \cup A_1 \cup A_2$. Let h be a bijection defined by $h(A_0) = A_1$, $h(A_1) = A_0$ and $h(A_2) = A_2$. Then h is not realizable.

The proof of Example 1 (see "Proofs") demonstrates some of the ideas used in the proof of Theorem 2.

Let us introduce definitions necessary to state Theorem 2.

Let M and N be two unions (not necessary nonempty) of disjoint circles in sphere S^2. Color connected components of $S^2 - N$ in black and white so that adjacent components have different colors. Union M is *on one side* (in this sphere) of N if M is contained in the union of same colored components of $S^2 - N$. Unions M and N are *unlinked* (in this sphere) if M is on one side of N and N is on one side of M. Equivalently unions M and N are *unlinked* (in S^2) if $[M] = 0$ in $H_1(S^2 - N; \mathbb{Z}_2)$ and $[N] = 0$ in $H_1(S^2 - M; \mathbb{Z}_2)$.

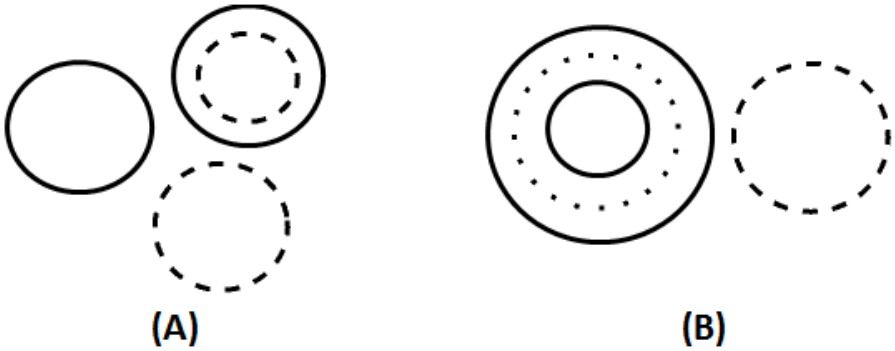


Figure 4: (A) M (solid) is on one side of N (dashed) while N is not on one side of M.
(B) M (solid) and N (dashed), N and P (dotted) are unlinked, but M and P are not unlinked.

Unions M and N are always unlinked if M or N is empty. If M is on one side of N then N is not necessary on one side of M (Fig. 4 left). Unlinkedness is not transitive. That is, if M and N, N and P are unlinked, then M and P are not necessarily unlinked (Fig. 4 right).

Let M be a union of disjoint circles in sphere S. Suppose A is a connected component of $S - M$. Denote by ∂A the boundary of the closure of A.

Theorem 2. *Let M and N be two unions of the same number of disjoint circles in S^2. Let h be a bijection between sets of circles of M and of N. Color connected components of $S^2 - M$ in two colors so that any two same colored components are not adjacent. Then h is realizable if and only if $h(\partial A)$ and $h(\partial B)$ are unlinked for each two same-colored components A and B of $S^2 - M$.*

We say that sphere with holes P is *properly embedded* in D^3 if $\partial P \subset \partial D^3$ and the interior of P lies in the interior of D^3. Theorem 2 is proved using the following:

Embedding Extension Theorem. *Let $M_1, \dots, M_m$ be unions of disjoint circles in the sphere $S^2 = \partial D^3$. Then there exist properly embedded in D^3 disjoint spheres with holes $P_1, \dots, P_m$ such that $\partial P_i = M_i$ for each $i = 1, \dots, m$ if and only if $M_1, \dots, M_m$ are pairwise unlinked.*

Embedding Extension Theorem immediately implies the following:

Corollary 1. Let $M_1, \dots, M_m$ be unions of disjoint circles in the sphere $S^2 = \partial D^3$. Suppose that for every i, j there exist properly embedded in D^3 disjoint spheres with holes P'_i, P'_j such that $\partial P'_i = M_i, \partial P'_j = M_j$. Then there exist properly embedded in D^3 disjoint spheres with holes $P_1, \dots, P_m$ such that $\partial P_i = M_i$ for each $i = 1, \dots, m$.

Note that analogous statement is false for all closed orientable 2-surfaces other than S^2. For instance:

Example 2. Let M_1, M_2, M_3 be unions of disjoint circles in the standard torus $T^2 \subset \mathbb{R}^3$. Let M_1 and M_2 be a single meridian each and let M_3 be a union of two meridians (Fig. 5). Then for every i, j there exist disjoint spheres with holes P'_i, P'_j whose interiors are inside T^2 and such that $\partial P'_i = M_i, \partial P'_j = M_j$. But there are no disjoint spheres with holes P_1, P_2, P_3 whose interiors are inside T^2 and such that $\partial P_1 = M_1, \partial P_2 = M_2, \partial P_3 = M_3$.

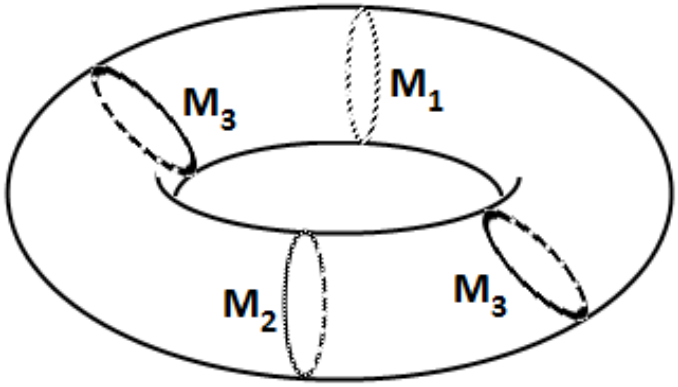


Figure 5: Unions M_1 and M_2 consists of one meridian each and M_3 consists of two meridians.

This example is similar to the famous Borromean rings example stated in the following way:

Borromean rings. Let S^1_1, S^1_2, S^1_3 be the Borromean rings in $S^3 = \partial D^4$. Then for every i, j there exist properly embedded in D^4 disjoint disks D'^2_i, D'^2_j such that $\partial D'^2_i = S^1_i, \partial D'^2_j = S^1_j$. But there are no properly embedded in D^4 disjoint disks D^2_1, D^2_2, D^2_3 such that $\partial D^2_1 = S^1_1, \partial D^2_2 = S^1_2, \partial D^2_3 = S^1_3$.

Relation to graphs

Suppose that M is a union of disjoint circles in sphere S^2. Define ("dual to M") graph $G = G(S^2, M)$ as follows. The vertices are the connected components of $S^2 - M$. Two vertices are connected by an edge if the corresponding connected components are neighbors.

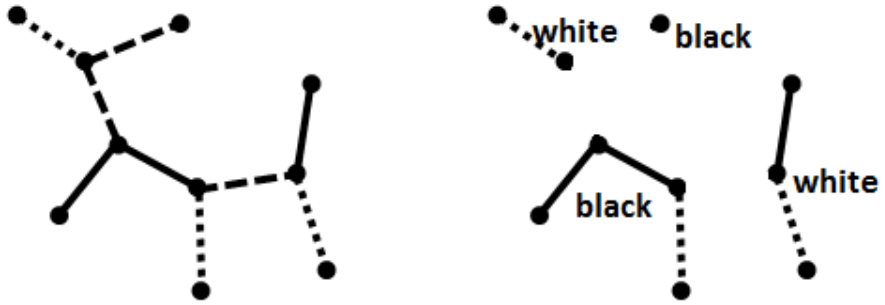


Figure 6: Graph G (left), p (solid edges), q (dashed edges), complement in G to the interiors of edges of q (right). Set p is not on one side of q since edges of p are both in black and white connected components. Thus p and q are not unlinked.

The definition of unlinked unions of circles can also be restated in terms of graphs. Let p and q be two sets of edges of a tree G. Color connected components of the complement in G to the interiors of edges of q in black and white so that adjacent components have different colors. The set p is *on the same side* of q (in this tree G) if p is contained in the union of same-colored connected components of $G - q$ (or, equivalently, if $p \cap q = \emptyset$ and for each two vertices of edges of p there is a path in the tree connecting these two points, and containing an even number of edges of q). Sets p and q are *unlinked* (in this tree) if p is on the same side of q and q is on the same side of p (for example see Fig. 6).

Let G and H be two trees with the same number of edges. Color vertices of G in two colors so that any two same colored vertices are not adjacent. Bijection h between the sets of edges of G and H is called *realizing* if $h(\delta A)$ [3] and $h(\delta B)$ are unlinked (in H) for each two same-colored vertices A and B of G.

Instead of a union of disjoint circles in a sphere let us consider its dual graph. Theorem 2 implies that a bijection between two sets of circles is realizable if and only if the corresponding bijection between the sets of edges of dual graphs is realizing.

Let G and H be two trees with k edges each. Given a bijection h between the sets of edges of G and H we can check algorithmically in at most $O(k^2)$ time if h is realizing. So, there is a brute-force algorithm which finds a realizing bijection (if any) in $O(k^2 k!)$ time. We don't know if the more efficient algorithm exists. More precisely there is the following open problem:

Open problem 1. Is there a "fast" algorithm, which takes as input two arbitrary trees G and H with k edges each and produces as output a realizing bijection (if any) between the sets of edges of G and H?

Open problem 2. Is there a tree G such that there is no realizing bijection between the sets of edges of G and H, where H is the path graph with the same number of edges as G?

Proofs

Proof of the Example 1. Assume to the contrary that there are PL embeddings $f : S^2 \hookrightarrow \mathbb{R}^3$ and $g : S^2 \hookrightarrow \mathbb{R}^3$ realizing h.

[3] δA is a set of all edges incident to A

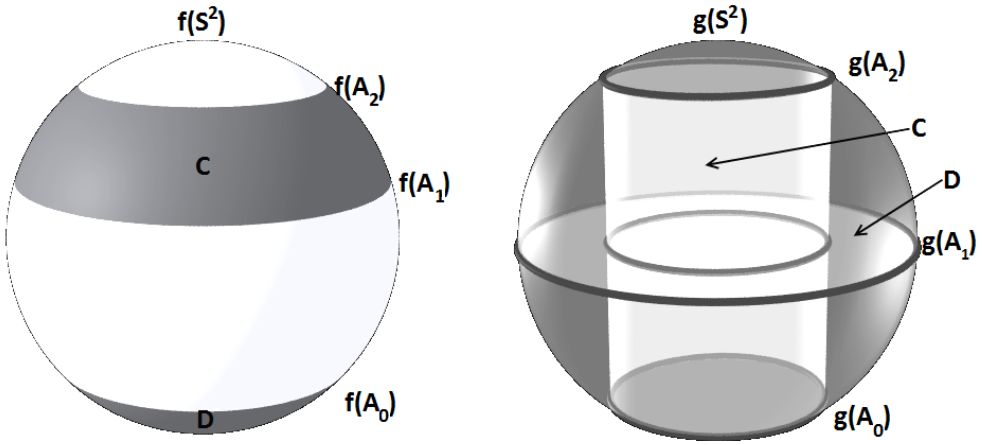


Figure 7: Circles $f(A_0) = g(A_1)$, $f(A_1) = g(A_0)$, $f(A_2) = g(A_2)$.

Denote by D the disk in $f(S^2) - g(S^2)$ bounded by $f(A_0)$ (Fig. 7). Denote by C the cylinder in $f(S^2) - g(S^2)$ bounded by $f(A_1)$ and $f(A_2)$. Clearly C and D lie in 3-space on the same side of sphere $g(S^2)$. Circles $f(A_1) = g(h(A_1)) = g(A_0)$ and $f(A_2) = g(h(A_2)) = g(A_2)$ lie in sphere $g(S^2)$ on the different sides of the circle $f(A_0) = g(h(A_0)) = g(A_1)$. So C intersects D. This contradicts to the assumption that f is an embedding. □

Proof of Theorem 1. Assume to the contrary that there is a bijection h between sets of circles of M and of N and PL embeddings $f : S^2 \hookrightarrow \mathbb{R}^3$ and $g : S^2 \hookrightarrow \mathbb{R}^3$ realizing h.

Denote the connected components of $f(S^2) - g(S^2)$ as shown in Fig. 1 (left).

Consider disks $A_1, \dots, A_4 \subset f(S^2)$. Without loss of generality we may assume that their interiors lie inside $g(S^2)$. Then the interior of component $C \subset f(S^2)$ lies inside $g(S^2)$ as well (since the intersection $f(S^2) \cap g(S^2)$ is transversal). Since $C, A_1, \dots, A_4$ are disjoint, C lies completely in one of the connected components of $\mathbb{R}^3 - g(S^2) \cup \bigsqcup A_i$. So all the 5 circles of ∂C lie in the same connected component of $g(S^2) - \bigsqcup \partial A_i$ (this argument is generalized in the proof of Claim 1 below).

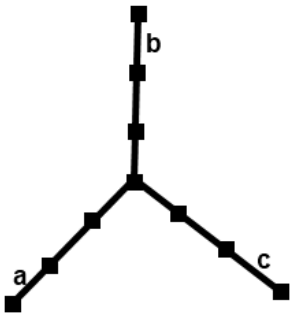


Figure 8: Graph $G(S^2, N)$.

Let us restate the previous statement in terms of graph $G(S^2, N)$ (Fig. 8). Denote by $G(C)$ the union of 5 edges of $G(S^2, N)$ corresponding to the circles of ∂C. Then $G(C)$ lies completely in one of the connected components of the compliment of $G(S^2, N)$ to the 4 edges corresponding to the circles of $\bigsqcup \partial A_i$. Since $G(S^2, N)$ has only 9 edges this means that $G(C)$ is a subtree of $G(S^2, N)$. Denote by $G(B)$ the union of 5 edges of $G(S^2, N)$ corresponding to the circles of ∂B. Likewise, $G(B)$ is a subtree of $G(S^2, N)$.

Since $G(B) \cup G(C) = G(S^2, N)$, at least two of the three edges a, b, c of $G(S^2, N)$ (Fig. 8) belong to one of subtrees $G(B)$ or $G(C)$. Without loss of generality we may assume that $a, b \in G(B)$. But any subtree of $G(S^2, N)$ containing both a and b has at least 6 edges while $G(B)$ has only 5 edges. This contradicts the initial assumption. □

Proof of the "only if" part of Theorem 2. Let A and B be two same colored components of $S^2 - M$. Then $f(A)$ and $f(B)$ lie on the same side of $g(S^2)$. So by the "only if" part of Embedding Extension Theorem $\partial f(A)$ and $\partial f(B)$ are unlinked in $g(S^2)$. Then $g^{-1}(\partial f(A)) = h(\partial A)$ and $g^{-1}(\partial f(B)) = h(\partial B)$ are unlinked in S^2. □

Proof of the "if" part of Theorem 2. Let $g : S^2 \hookrightarrow \mathbb{R}^3$ be any PL embedding. We define a PL embedding $f : S^2 \hookrightarrow \mathbb{R}^3$ such that f, g realize h by defining $f(A)$ for every connected component A of $S^2 - M$.

Color connected components of $S^2 - M$ in black and white so that any two same colored components are not adjacent.

Let $P_1, \dots, P_m$ be the white components of $S^2 - M$. By the assumption of the Theorem $h(\partial P_1), \dots, h(\partial P_m)$ are pairwise unlinked in S^2. So $g(h(\partial P_1)), \dots, g(h(\partial P_m))$ are pairwise unlinked in $g(S^2)$. By the "if" part of Embedding Extension Theorem there exist disjoint spheres with holes $P'_1, \dots, P'_m$ whose interiors are inside $g(S^2)$ and such that $\partial P'_i = g(h(\partial P_i))$ for each $i = 1, \dots, m$. Define $f(P_i) := P'_i$ for each i.

Likewise, let $Q_1, \dots, Q_n$ be the black components of $S^2 - M$. By the "only if" part of Embedding Extension Theorem there exist disjoint spheres with holes $Q'_1, \dots, Q'_n$ whose interiors are *outside* $g(S^2)$ and such that $\partial Q'_j = g(h(\partial Q_j))$ for each $j = 1, \dots, n$. Define $f(Q_j) := Q'_j$ for each j.

Image of f is the sphere $\bigsqcup P'_i \sqcup \bigsqcup Q'_j$. Clearly, f and g realize h. □

Proof of the "only if" part of Embedding Extension Theorem. Consider a properly embedded in D^3 sphere with holes P_i. Add a "cap" (homeomorphic to a disk) in $\mathbb{R}^3 - D^3$ to every circle of ∂P_i such that the union of P_i with these caps is a sphere $\hat{P}_i$. (In the smooth category we may assume that S^2 is a round sphere and that bounding circles of ∂P_i are round circles, none of them being an equator. Then for each circle of ∂P_i take the round sphere passing through this circle and the center of S^2. Take parts of such spheres lying in $\mathbb{R}^3 - D^3$ as these "caps". Analogous, albeit slightly more complicated, construction is possible in the PL category).

Clearly, all same colored connected components of $S^2 - M_i = S^2 - \partial P_i$ lie on the same side of $\hat{P}_i$. And since P_i and P_j are disjoint, $S^2 \cap P_j = M_j$ lie on one side of $\hat{P}_i$, i.e. in the union of same colored components of $S^2 - M_i$.

So M_j lie on one side of M_i by definition. Likewise, M_i lie on one side of M_j. Therefore M_i and M_j are unlinked. □

To prove the "if" part we require the following claim. Proof of the claim is postponed.

Claim 1. *Let $P_1, \dots, P_n$ be properly embedded in D^3 pairwise disjoint spheres with holes. Let M be a union of disjoint circles in $S^2 = \partial D^3$ such that M and ∂P_i are unlinked for every i. Then M lies in one connected component of $D^3 - (P_1 \sqcup \dots \sqcup P_n)$.*

Proof of the "if" part of Embedding Extension Theorem. This proof was suggested by A. Novikov. It is simpler than our original proof.

Use induction on number of circles in $M_1 \sqcup \dots \sqcup M_m$.

Let p be a circle of $M_1 \sqcup \dots \sqcup M_m$ bounding an open disk D in S^2 disjoint with $M_1 \sqcup \dots \sqcup M_m$ (p corresponds to an edge of $G(S^2, M_1 \sqcup \dots \sqcup M_m)$ issuing out of a leaf vertex). We may assume that $p \subset M_1$. Denote by M'_1 the union of circles $M_1 - p$ (note that M'_1 may be empty).

Unions $M'_1, M_2, \dots, M_m$ are pairwise unlinked. By the inductive hypothesis there are properly embedded in D^3 disjoint spheres with holes $P'_1, P_2, \dots, P_m$ such that $\partial P_i = M_i$ for each $i = 2, \dots, m$ and $\partial P'_1 = M'_1$. Let D' be a disk obtained from the closure of D by a slight deformation so that the interior of D' is in the interior of D^3 and $\partial D' = p$. By Claim 1 each two points of $M_1 = M'_1 \sqcup p$ can be connected by a path in the interior of D^3 disjoint with $P_2, \dots, P_m$. So we can connect D' with P'_1 by a tube in the interior of D^3 disjoint with $P_2, \dots, P_m$. Then we obtain a sphere with holes. Denote it by P_1. We have $\partial P_1 = p \sqcup \partial P'_1 = M_1$, P_1 is properly embedded in D^3 and P_1 is disjoint with $P_2, \dots, P_m$. The inductive step is proved. □

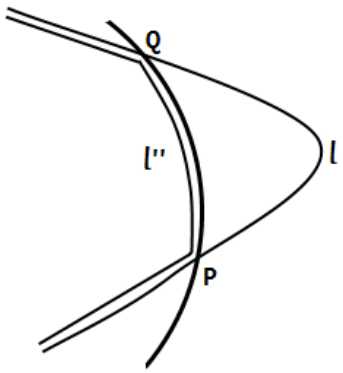


Figure 9: Paths l, l''.

Proof of Claim 1. Take any two points $A, B \in M$. Denote by l a path in D^3 connecting A and B such that $\bar{l} := \#(l \cap \bigsqcup_{i=1}^{n} P_i)$ is minimal (minimal by l, objects $A, B, M, D^3, P_1, \dots, P_n$ are fixed). Assume to the contrary that l is not as required, i.e. $\bar{l} > 0$. Since M is on one side of ∂P_i, number $\#(l \cap P_i)$ is even for each i. (If $m = 2$, we may even obtain that $\#(l \cap P_1) = 0$ and stop here.) Then $\#(l \cap P_i) \geq 2$ for some i. Denote by Q and R two consecutive points of $l \cap P_i$. Denote by Q' the point of l slightly before Q and by R' the point of l slightly after R (Fig. 9). Since P_i is connected, Q and R can be connected by a path in P_i. So Q' and R' can be connected by a path l' very close to P_i but not intersecting P_i. Path l' does not intersect any of $P_1, \dots, P_n$ because it is very close to P_i and $P_1, \dots, P_n$ are pairwise disjoint. Substitute the part of l between Q' and S' by l'. Denote the obtained path by l''. Then $\overline{l''} = \bar{l} - 2$. This contradicts to the minimality of $\bar{l}$. Thus l is as required. □

Acknowledgements

I thank A. Novikov for suggestion of a simpler proof of the "if" part of Embedding Extension Theorem. I also thank prof. A. Skopenkov for his useful suggestions and comments.

References

[1] I. Arzhantsev, V. Bogachev, A. Garber, A. Zaslavsky, V. Protasov and A. Skopenkov, *Students' mathematical olympiades at Moscow State University 2010-2011*, Mat. Prosveschenie, 16 (2012), 214-227, in russian, http://www.mccme.ru/free-books/matpros/mpg.pdf

[2] S. Avvakumov, A. Berdnikov, A. Rukhovich and A. Skopenkov, *How do curved spheres intersect in 3-space, or two-dimensial meandra*, http://www.turgor.ru/lktg/2012/3/3-1en_si.pdf

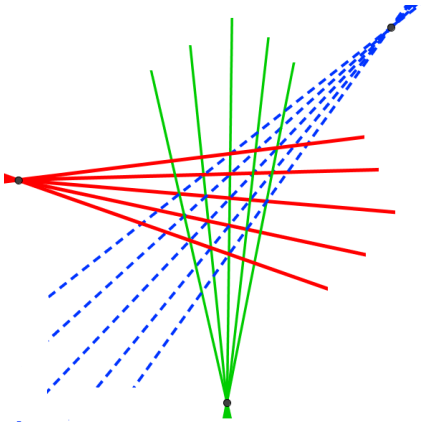
[3] A. Rukhovich, *On intersection of two embedded spheres in 3-space*, http://arxiv.org/abs/1012.0925

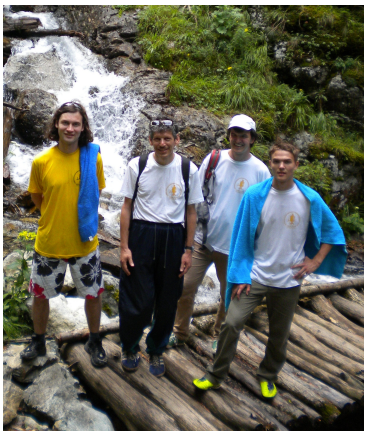
[4] T. Hirasa, *Dissecting the torus by immersions*, Geometriae Dedicata, 145:1 (2010), 33-41

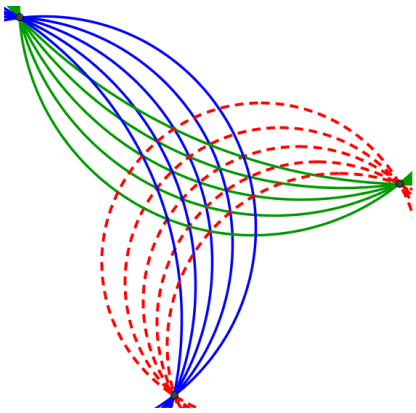
[5] T. Nowik, *Dissecting the 2-sphere by immersions*, Geometriae Dedicata 127, (2007), 37-41, http://arxiv.org/abs/math/0612796

Ткани из прямых и окружностей

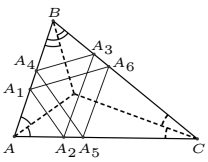
Алексей Заславский, Федор Нилов, Александр Полянский, Михаил Скопенков



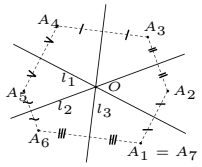




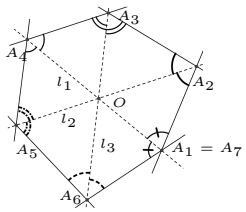
Замыкание с периодом 6

0.1. 

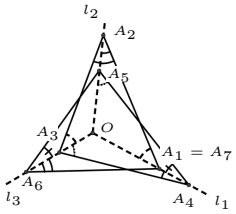
Через точку A_1 на стороне AB треугольника ABC провели прямую, перпендикулярную биссектрисе угла A. Она пересекла сторону AC в точке A_2. Через точку A_2 провели прямую, перпендикулярную биссектрисе угла C. Она пересекла сторону CB в точке A_3. Аналогично построили точки A_4, A_5, A_6, A_7. Докажите, что $A_7 = A_1$.

0.2. 

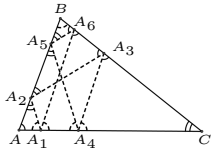
Три прямые l_1, l_2, l_3 пересекаются в одной точке. Точка A_1 выбирается произвольно. Точки A_2, A_3, A_4, A_5, A_6, A_7 получаются последовательным отражением точки A_1 относительно прямых l_1, l_2, l_3, а затем снова l_1, l_2, l_3. Докажите, что $A_7 = A_1$.

0.3. 

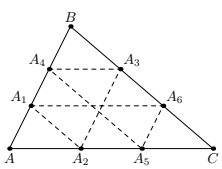
Три прямые l_1, l_2, l_3 пересекаются в одной точке. На прямых l_1 и l_2 выбираются произвольные точки A_1 и A_2, соответственно. Точка A_3 является пересечением прямой l_3 и прямой, симметричной A_1A_2 относительно l_2. Точка A_4 является пересечением прямой l_1 и прямой, симметричной A_2A_3 относительно l_3. Аналогично строятся точки A_5, A_6, A_7. Докажите, что $A_7 = A_1$.

0.4. 

Теорема Ф. Петрова. Три луча l_1, l_2, l_3 исходят из одной точки O. На лучах l_1 и l_2 выбираются произвольные точки A_1 и A_2, соответственно. На луче l_3 выбирается такая точка A_3, что угол между A_2A_3 и l_3 равен углу между A_1A_2 и l_1. Затем на луче l_1 выбирается такая точка A_4, что угол между A_3A_4 и l_1 равен углу между A_2A_3 и l_2. Аналогично стоятся точки A_5, A_6, A_7. Докажите, что $A_7 = A_1$.

0.5. 

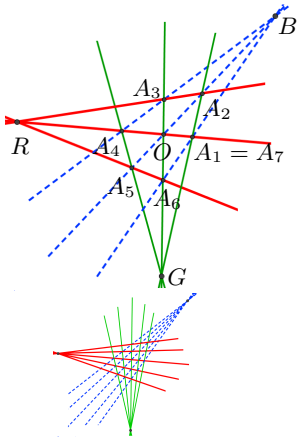
Бильярдный стол имеет форму треугольника ABC. Бильярдный шар выпустили из некоторой точки A_1 стороны AC под углом ABC к этой стороне. Обозначим через A_2, A_3, A_4, A_5, A_6, A_7 точки, в которых шар последовательно ударялся о края стола. Докажите, что $A_7 = A_1$.

0.6. 

Через точку A_1 на стороне AB треугольника ABC провели прямую параллельно BC. Она пересекла CA в точке A_2. Через A_2 провели прямую параллельно AB. Она пересекла BC в точке A_3. Аналогично построили точки A_4, A_5, A_6, A_7. Докажите, что $A_7 = A_1$.

[0]Летняя конференция Международного математического Турнира Городов, 2–10 августа 2012 г.

0.7.

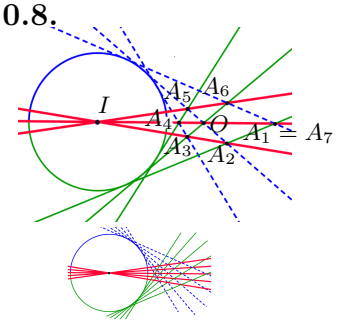


Теорема Паппа.

a) Пусть прямые, содержащие стороны невыпуклого шестиугольника, попеременно проходят через две фиксированные точки. Тогда прямые, содержашие его диагонали, пересекаются в одной точке или параллельны.

b) На плоскости отмечены красная (R), зеленая (G) и синяя (B) точки (см. рисунок слева). Любая прямая, проходящая ровно через одну из этих точек, окрашена в цвет этой точки. Возьмем произвольную точку O внутри треугольника RGB. Проведем через нее красную, зеленую и синюю прямые. На красной прямой возьмем произвольную точку A_1 внутри треугольника RGB. Проведем через нее зеленую прямую. Пусть она пересекла синюю прямую через точку O в точке A_2. Через точку A_2 уже проведены зеленая и синяя прямая; проведем через A_2 красную прямую. Точку пересечения полученной красной прямой с зеленой прямой через точку O обозначим A_3. Продолжая данное построение, получим точки A_4, A_5, A_6, A_7. Докажите, что $A_7 = A_1$.

0.8.

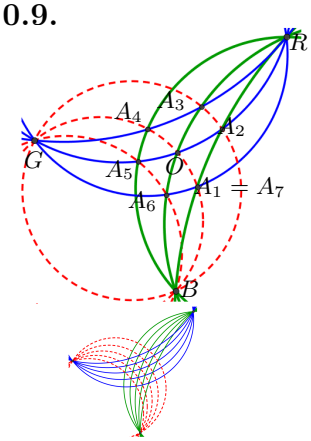


Теорема Брианшона.

a) Диагонали описанного шестиугольника пересекаются в одной точке.

b) Дана окружность с центром I и точка O вне нее. Прямые, проходящие через центр I, окрашены в красный цвет. Касательные к окружности окрашены в зеленый и синий цвета в зависимости от того, в какой полуплоскости относительно прямой OI расположена точка касания (см. рисунок слева). На прямой OI возьмем произвольную точку A_1. Проведем через нее зеленую прямую. Пусть она пересекла синюю прямую через точку O в точке A_2. Через точку A_2 уже проведены зеленая и синяя прямые; проведем красную. Точку пересечения этой красной прямой с зеленой прямой через точку O обозначим A_3. Аналогично получим точки A_4, A_5, A_6, A_7. Докажите, что $A_7 = A_1$.

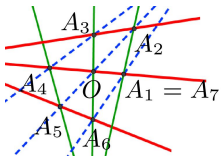
0.9.



Теорема Бляшке. На плоскости отмечены красная (R), зеленая (G) и синяя (B) точки (см. рисунок слева). Любая окружность, проходящая ровно через две из этих точек, окрашена в цвет третьей. Возьмем произвольную точку O внутри треугольника RGB. Проведем через нее красную, синюю и зеленую окружность. На красной окружности внутри треугольника RGB возьмем произвольную точку A_1. Проведем через нее зеленую окружность. Пусть она пересекла синюю окружность через точку O в точке A_2, отличной от R, G и B. Через точку A_2 уже проведены зеленая и синяя окружность; проведем красную. Точку пересечения полученной красной окружности с зеленой окружностью через точку O, отличную от R, G и B, обозначим A_3. Продолжая данное построение, получим точки A_4, A_5, A_6, A_7. Докажите, что $A_7 = A_1$.

Определение ткани

Определение.

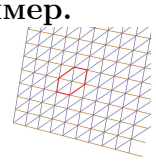


Пусть некоторые прямые на плоскости окрашены в красный, зеленый и синий цвета. Окрашенные прямые образуют (*гексагональную*) *ткань*, если для некоторого круга Ω на плоскости выполнены следующие 2 условия:

Условие слоения: Через каждую точку круга Ω проходит ровно одна прямая каждого цвета, причем прямые разного цвета не совпадают.

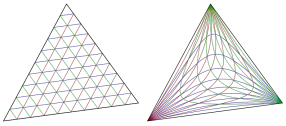
Условие замыкания (см. рисунок слева): Возьмем произвольную точку O внутри круга Ω. Проведем через нее красную, зеленую и синюю прямую. На красной прямой внутри круга Ω возьмем произвольную точку A_1. Проведем через нее зеленую прямую. Пусть она пересекла синюю прямую через точку O в точке A_2. Через точку A_2 уже проведены зеленая и синяя прямая; проведем красную. Точку пересечения этой красной прямой с зеленой прямой через точку O обозначим A_3. Продолжая это построение, получим точки A_4, A_5, A_6, A_7. *Условие замыкания* состоит в том, что если все указанные точки существуют и лежат внутри круга Ω, то $A_7 = A_1$.

Пример.



Три множества прямых, параллельных сторонам фиксированного треугольника, образуют ткань. Из прямых, образующих ткань, всегда можно получить "триангуляцию" некоторой части плоскости. На рисунках мы будем всегда изображать именно такие триангуляции; как, например, на нижних маленьких рисунках к задачам 0.7–0.9.

Замечание.



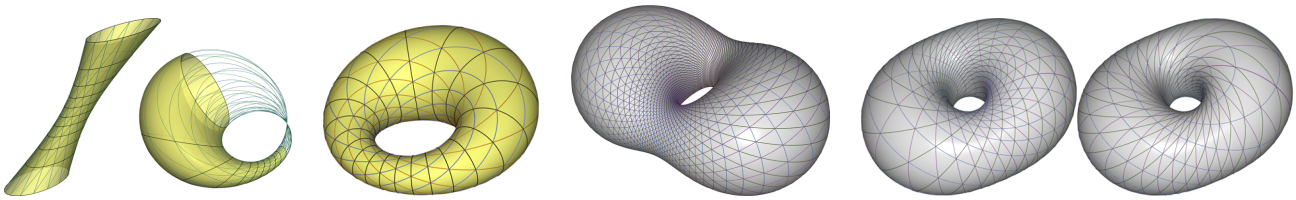
Любая ткань (из любых линий и на любой поверхности) получается из предыдущего примера с помощью подходящего непрерывного взаимно-однозначного отображения некоторой части плоскости на круг Ω.

Заменив в определении ткани слово "прямая" на слова "прямая или окружность", и потребовав в условии слоения, чтобы прямые и окружности не касались друг друга внути круга Ω, получим определение *ткани из прямых или окружностей*.

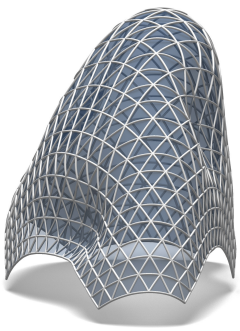
Проблема Бляшке (1920е). *Найти все ткани из окружностей на плоскости.*

Заменив в определении ткани "круг на плокости" на "пересечение поверхности с некоторым шаром", получим определение *ткани на поверхности*.

Замечание. Ткани из окружностей полностью описаны для всех поверхностей, за исключением плоскости и сферы [3]; некоторые примеры показаны на рисунке снизу.



Замечание.



Интерес к изучению тканей увеличился в последнее время в связи с возможными приложениями в архитектуре. Важная задача современной архитектуры состоит в *рационализации* поверхностей свободной формы, то есть разбиения их на относительно простые панели. Один из подходов к рационализации — это *конструкции из дуг окружностей*, то есть триангуляции поверхности дугами окружностей, полученные из некоторой ткани. На рисунке слева изображена подобная конструкция на Эйндховенском куполе, принадлежащая архитектору М. Фуксасу.

1 Ткани из прямых на плоскости

Какие из следующих троек множеств прямых образуют ткань? Подсказка: можно использовать программу Geogebra для экспериментов и рисования картинок.

1.1. (R) Прямые, паралельных оси Ox; (G) параллельные оси Oy; (B) проходящие через начало координат O.

1.2. Три множества прямых, проходящих через три фиксированные попарно различные точки плоскости.

Назовем *единичной полуокружностью* множество точек, координаты которых удовлетворяют условиям $x^2 + y^2 = 1$ и $x > 0$, а *дополнительной* к ней полуокружностью множество, заданное условиями $x^2 + y^2 = 1$ и $x < 0$.

1.3. (R) Прямые, касающиеся единичной полуокружности; (G) касающиеся дополнительной к ней полуокружности; (B) проходящие через начало координат O.

1.4. (R) Прямые, касающиеся единичной полуокружности; (G) касающиеся дополнительной к ней полуокружности; (B) проходящие через фиксированную точку.

1.5. (R) Прямые, паралельные оси Ox; (G) параллельные оси Oy; (B) касающиеся единичной полуокружности.

1.6. (R) Прямые, касающиеся единичной полуокружности; (G) проходящие через начало координат O; (B) паралельные оси Ox.

2 Ткани из окружностей на плоскости

2.1. Приведите пример ткани из окружностей (с доказательством).

Какие из следующих троек множеств прямых и окружностей образуют ткань?

2.2. (R) Прямые, касающиеся единичной полуокружности; (G) касающиеся дополнительной к ней полуокружности; (B) окружности с центром в начале координат.

2.3. (R) Прямые, касающиеся единичной полуокружности; (G) проходящие через начало координат; (B) окружности с центром в начале координат.

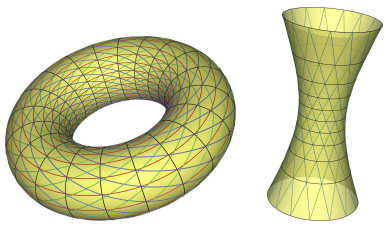
2.4. (R) Прямые, паралельные оси Ox; (G) параллельные оси Oy; (B) окружности, одновременно касающиеся отрезков $x = 0, 0 \le y \le 1$ и $x = 1, 0 \le y \le 1$.

2.5. (R) Прямые, проходящие через начало координат; (G) окружности, одновременно касающиеся отрезков $x = 0, 0 \le y \le 1$ и $y = 0, 0 \le x \le 1$; (B) окружности, одновременно касающиеся отрезков $x = 0, 2 \le y \le 4$ и $y = 0, 2 \le x \le 4$.

2.6. (R) Прямые, проходящие через начало координат; (G) окружности с центром в начале координат; (B) окружности, одновременно касающиеся отрезков $x = 0, 0 \le y \le 1$ и $y = 0, 0 \le x \le 1$.

3 3D

Тором называется результат вращения окружности вокруг прямой, лежащей в ее плоскости, но не пересекающей саму окружность; см. рисунок внизу слева. Окружности, получающиеся как траектории отдельных точек, называются *параллелями*. Исходная окружность и все окружности, которые получаются из нее вращением, называются *меридианами*. Через каждую точку тора проходят еще две окружности, целиком лежащие на нем; они называются *окружностями Вилларсо* (этим можно пользоваться без доказательства).



3.1. Параллели тора вместе с окружностями Вилларсо образуют ткань.

3.2. Меридианы тора вместе с окружностями Вилларсо образуют ткань.

Гиперболоидом вращения называется результат вращения прямой вокруг скрещивающейся с ней прямой; см. рисунок вверху справа. Через каждую точку гиперболоида вращения проходят две прямые, лежащие на нем (этим можно пользоваться без доказательства).

3.3. Прямые, лежащие на гиперболоиде вращения, вместе с его параллелями образуют ткань.

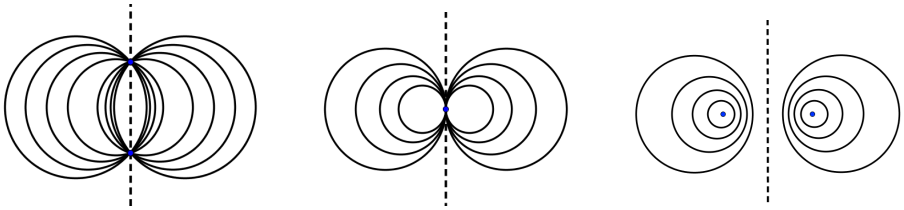
4 Ткани из окружностей: общие конструкции

Дополните данные множества красных и зеленых прямых до ткани, добавив некоторое множество синих **a)** прямых; **b)** окружностей (приведите как можно больше примеров таких множеств синих прямых или окружностей; постарайтесь найти все примеры и доказать, что других нет):

4.1. (R) Прямые, паралельных оси Ox; (G) параллельные оси Oy.

4.2. (R) Прямые, паралельных оси Ox; (G) проходящие через начало координат O.

Пучком прямых называется множество всех прямых, проходящих через фиксированную точку (*вершину* пучка) или параллельных фиксированной прямой. *Пучком окружностей* называется множество окружностей, имеющих общую радикальную ось (см. рисунок внизу). Если пучок содержит "окружности" нулевого радиуса, то они называются *предельными точками* пучка (жирные точки на рисунке).



4.3. Пучок окружностей с двумя предельными точками и два пучка прямых с вершинами в этих точках вместе образуют ткань.

Замечание. Все ткани, образованные тремя семействами окружностей, были найдены А. Шелеховым [4, Теорема 0.1].

В дальнейшем *обобщёнными окружностями* будем называть окружности и прямые. *Обобщёнными дугами* будем называть и дуги окружностей, окружности, отрезки, лучи и прямые. *Ткань* из обобщенных дуг определяется аналогично ткани из окружностей и прямых. Следующая задача может оказаться полезной для решения задач разделов 1–3.

4.4. *Построение тканей с помощью групп преобразований.* Пусть выполнены следующие условия:

- Для каждого $t \in \mathbb{R}$ задано взаимно-однозначное преобразование R_t плоскости, переводящее обобщенные окружности в обобщенные окружности. Для любых $t, s \in \mathbb{R}$ и для любой точки A выполняется, что $R_t(R_s(A)) = R_{t+s}(A)$. Для любой точки A множество $\gamma_A = \{R_t(A) : t \in \mathbb{R}\}$ представляет собой обобщённую дугу.
- Через некоторую точку проведены две различные обобщённые дуги γ_1, γ_2. Для любой точки $A \in \gamma_1$ дугу γ_A покрасим в красный цвет. Дуги $R_t(\gamma_1)$, где $t \in \mathbb{R}$, покрасим в зелёный цвет, дуги $R_t(\gamma_2)$, где $t \in \mathbb{R}$, покрасим в синий цвет. Известно, что цветные дуги имеют не более одной общей точки.
- Существует круг Ω, через каждую точку которого проходит ровно одна дуга каждого цвета.

Тогда красные, синие и зеленые дуги образуют ткань.

Следующая серия состоит из более трудных задач (за исключением самой первой).

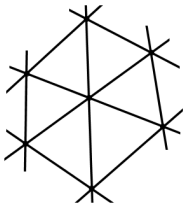
Если прямая на плоскости задается уравнением $px + qy = 1$, то назовем пару чисел (p, q) *координатами* этой прямой.

4.5. Все прямые, координаты которых удовлетворяют некоторому линейному уравнению, образуют пучок.

4.6. * Все прямые, координаты которых удовлетворяют некоторому уравнению второго порядка, либо касаются одной кривой, заданной уравнением второго порядка, либо образуют два пучка, либо один, либо пустое множество.

4.7. * *Обобщенная теорема Паскаля.* Три красные прямые пересекают три синие в 9 различных точках. Тогда если кривая, заданная уравнением третьего порядка, проходит через 8 из них, то она проходит и через девятую.

4.8. * *Теорема Шаля.* Девять прямых образуют шестиугольную конфигурацию как на рисунке снизу. Тогда если координаты 8 из этих прямых удовлетворяют уравнению третьего порядка, то и координаты девятой удовлетворяют тому же уравнению.



Множество неокрашенных обобщенных окружностей назовем *тканью*, если для некоторого круга Ω выполнены следующие условия:

- *Обобщенное условие слоения.* Через каждую точку круга Ω проходит ненулевое конечное число обобщенных окружностей из нашего множества.
- Часть обобщенных окружностей нашего множества можно окрасить в три цвета так, чтобы окрашенные обобщенные окружности образовали ткань в круге Ω .

4.9. * Если множество прямых на плоскости удовлетворяют обобщенному условию слоения в некотором круге, а их координаты удовлетворяют некоторому уравнению третьего порядка, то эти прямые образуют ткань.

4.10. * Нормали к параболе образуют ткань.

4.11. * Прямые Симсона произвольного треугольника образуют ткань.

В заключение приведем несколько трудных задач для исследования.

4.12. ** *Теорема Графа–Зауэра* ([2, 1]). Пусть множество прямых удовлетворяет обобщенному условию слоения в некотором круге. Тогда эти прямые образуют ткань тогда и только тогда, когда для некоторой декартовой системы координат на плоскости их координаты удовлетворяют одному уравнению 3-й степени.

Замечание. Этот результат позволяет найти все ткани из окружностей ортогональных фиксированной окружности [5].

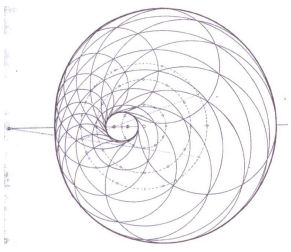
Циклика задается уравнением вида

$$\lambda(x^2+y^2)^2+(x^2+y^2)(\mu x+\nu y)+Q(x,y)=0,$$

где $\lambda,\mu,\nu\in\mathbb{R}$, а $Q(x,y)$ — многочлен степени не выше 2.

Следующая задача адресована тем, кто знает определение комплексных точек и непрерывного семейства окружностей.

4.13. ** *Теорема Вундерлиха* ([6], см. рисунок внизу). Три непрерывных семейства окружностей дважды касаются некоторой циклики (возможно, в комплексных точках). Если эти семейства удовлетворяют обобщенному условию слоения в некотором круге, то они образуют ткань.



4.14. *** Попробуйте придумать пример ткани из окружностей, отличный от приведенных выше.

Благодарности.

Авторы благодарны И. Богданову за полезные обсуждения.

Список литературы

[1] Wilhelm Blaschke and Gerrit Bol. *Geometrie der Gewebe.* Springer, 1938.

[2] H. Graf and R. Sauer. Über dreifache Geradensysteme in der Ebene, welche Dreiecksnetze bilden. *Sitz. Bayer. Akad. Math.-nat. Abt.*, pages 119–156, 1924.

[3] Helmut Pottmann, Ling Shi, and Mikhail Skopenkov. Darboux cyclides and webs from circles. *Computer Aided Geometric Design*, 29(1):77 – 97, 2012.

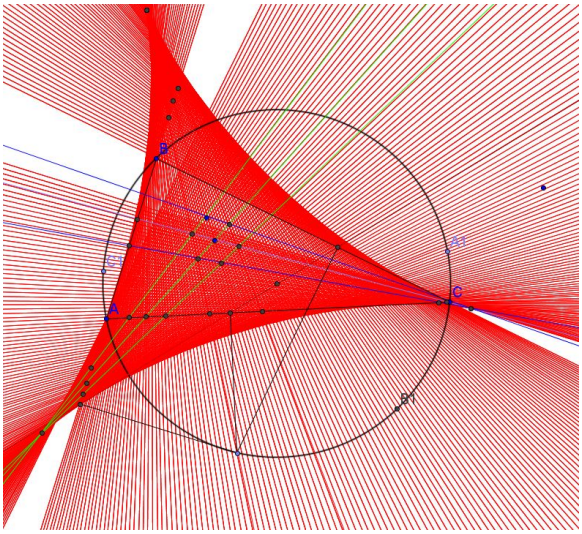
[4] A. M. Shelekhov. Classification of regular three-webs formed by pencils of circles. *J. Math. Sciences*, 143(6):3607–3629, 2007.

[5] K. Strubecker. Über ein Klasse spezieller Dreiecksnetze aus Kreisen. *Monaths. Math. Phys.*, 39:395–398, 1932.

[6] W. Wunderlich. Über ein besonderes Dreiecksnetz aus Kreisen. *Sitzungsber. Akad. Wiss. Wien*, 147:385–399, 1938.

Ткани из прямых и окружностей

Алексей Заславский, Федор Нилов, Александр Полянский, Михаил Скопенков



Решения задач.

Сформулируем лемму, которая поможет решить задачи 0.1, 0.5, 0.6.

Лемма 0. Пусть $A'B'C'$ — чевианный треугольник некоторой точки относительно треугольника ABC (то есть прямые AA', BB', CC' пересекаются в одной точке). Через произвольную точку M_1 на стороне AC проведем прямую, параллельную $A'B'$, и найдем точку M_2 ее пересечения с BC; через M_2 проведем прямую, параллельную $A'C'$ до пересечения с AB в точке M_3 и т.д. Тогда $M_1 = M_7$.

`Доказательство.`

Если точка M_1 совпадает с точкой B_1, то точки M_4 и M_7 тоже совпадут с B'. Иначе из теоремы Чевы получаем, что $\frac{\overline{AB'}}{\overline{B'C}}\frac{\overline{CA'}}{\overline{A'B}}\frac{\overline{BC'}}{\overline{C'A}} = 1$. А из теоремы Фалеса $\frac{\overline{B'C}}{\overline{CA'}} = \frac{\overline{B'M_1}}{\overline{M_2A'}}$, $\frac{\overline{A'B}}{\overline{BC'}} = \frac{\overline{M_2A'}}{\overline{C'M_3}}$, $\frac{\overline{C'A}}{\overline{AB'}} = \frac{\overline{C'M_3}}{\overline{M_4B'}}$. Подставляя последние три равенства в первое, получаем, что: $\frac{\overline{B'M_1}}{\overline{M_4B'}} = 1$. Следовательно, точки M_1 и M_4 симметричны относительно точки B'. Точно также доказывается, что M_4 и M_7 симметричны относительно точки B'. Следовательно, $M_1 = M_7$.

`Решение задач раздела 0.`

Решение задачи 0.1

Первое решение. Пусть a, b, c — это длины сторон BC, CA и AB соответственно. Пусть x — *направленная* длина отрезка AA_1 (то есть длина отрезка AA_1, взятая с положительным знаком, если вектора $\overrightarrow{AA_1}$, $\overrightarrow{AB}$ сонаправлены, или с отрицательным знаком, если эти вектора противоположно направлены). Поскольку прямая A_1A_2 перпендикулярна биссектрисе угла BAC, то отсюда получаем $AA_2 = AA_1 = x$ (с учетом знака). Аналогично мы получаем, что $CA_3 = CA_2 = b - x$, $BA_4 = a - b + x$, $AA_5 = c - a + b - x$, $CA_6 = a - c + x$, $AA_7 = x$ (с учетом знака). То есть $AA_7 = x = AA_1$ (с учетом знака), а это значит, что $A_7 = A_1$.

Второе решение. По лемме 0 для треугольника Жергонна.

Решение задачи 0.2

В решениях задач 0.2, 0.3 и 0.4 рассматриваются углы между направлениями (векторами).

[0]Летняя конференция Международного математического Турнира Городов, 2–10 августа 2012 г.

Пусть O — точка пересечения прямых l_1, l_2, l_3. Пусть $(l_1, l_2) = \varphi_{1,2}$,$(l_2, l_3) = \varphi_{2,3}$, $(OA_1, l_1) = \varphi$. Так как длины отрезков OA_i одинаковы (при симметрии сохраняется длина), поэтому достаточно показать, что $(OA_7, l_1) = (OA_1, l_1) = \varphi$.

Так как $(OA_2, l_1) = -(OA_1, l_1) = -\varphi$, то $(OA_2, l_2) = -\varphi + \varphi_{1,2}$. Поэтому $(OA_3, l_2) = \varphi - \varphi_{1,2}$. Отсюда получаем, что $(OA_3, l_3) = \varphi - \varphi_{1,2} + \varphi_{2,3}$. Поэтому $(OA_4, l_3) = -\varphi + \varphi_{1,2} - \varphi_{2,3}$. Отсюда получаем, что $(OA_4, l_1) = -\varphi + \varphi_{1,2} - \varphi_{2,3} - \varphi_{1,2} - \varphi_{2,3} = -\varphi - 2\varphi_{2,3}$. Действуя аналогично получаем, что $(OA_7, l_1) = -(OA_4, l_1) - 2\varphi_{2,3} = \varphi$.

Решение задачи 0.3

Пусть $(l_1, l_2) = \varphi_{1,2}$,$(l_2, l_3) = \varphi_{2,3}$ и $(l_{(1,2)}, l_1) = \varphi$, где $l_{(i,i+1)}$ — вектор, соединяющий точку O с основанием проекции из O на прямую A_iA_{i+1}. Так как расстояния от O до всех A_iA_{i+1} окажутся одинаковыми (при симметрии сохраняется длина), то достаточно показать, что $(l_{(7,8)}, l_1) = (l_{(1,2)}, l_1) = \varphi$.

Так как $(l_{(2,3)}, l_1) = -\varphi$. Поэтому $(l_{(2,3)}, l_2) = (l_{(2,3)}, l_1) + (l_1, l_2) = -\varphi + \varphi_{1,2}$. Аналогично $(l_{(3,4)}, l_2) = \varphi - \varphi_{1,2}$. Поэтому $(l_{(3,4)}, l_3) = (l_{(3,4)}, l_2) + (l_2, l_3) = \varphi - \varphi_{1,2} + \varphi_{2,3}$. Аналогично $(l_{(4,5)}, l_3) = -\varphi + \varphi_{1,2} - \varphi_{2,3}$. Поэтому $(l_{(4,5)}, l_1) = (l_{(4,5)}, l_3) + (l_1, l_3) = -\varphi + \varphi_{1,2} - \varphi_{2,3} + (-\varphi_{1,2} - \varphi_{2,3}) = -\varphi - 2\varphi_{2,3}$. Действуя аналогично получаем, что $(l_{(7,8)}, l_1) = -(l_{(4,5)}, l_1) - 2\varphi_{2,3} = \varphi$.

Решение задачи 0.4.

Пусть $(l_1, l_2) = \varphi_{1,2}$,$(l_2, l_3) = \varphi_{2,3}$, $(A_1A_2, l_1) = \varphi$. Так как радиусы описанных окружностей около треугольников OA_iA_{i+1} равны между собой (это следует из теоремы синусов), то достаточно показать, что $(A_7A_6, l_3) = -(A_1A_2, l_2) = -\varphi - \varphi_{1,2}$ (из обратной теоремы синусов).

Так как $(A_3A_2, l_3) = -(A_1A_2, l_1) = -\varphi$. Поэтому $(A_3A_2, l_2) = -\varphi - \varphi_{2,3}$. Так как $(A_3A_4, l_1) = -(A_3A_2, l_2) = \varphi + \varphi_{2,3}$. Поэтому $(A_3A_4, l_3) = \varphi + \varphi_{1,2} + \varphi_{2,3} + \varphi_{2,3}$. Так как $(A_5A_4, l_2) = -(A_3A_4, l_3) = -\varphi - \varphi_{1,2} - 2\varphi_{2,3}$. Поэтому $(A_5A_4, l_1) = -\varphi - \varphi_{1,2} - 2\varphi_{2,3} - \varphi_{1,2}$. Так как $(A_5A_6, l_3) = -(A_5A_4, l_1) = \varphi + 2\varphi_{1,2} + 2\varphi_{2,3}$. Поэтому $(A_5A_6, l_2) = \varphi + 2\varphi_{1,2} + 2\varphi_{2,3} - \varphi_{2,3}$. Так как $(A_7A_6, l_1) = -(A_5A_6, l_2) = -\varphi - 2\varphi_{1,2} - \varphi_{2,3}$. Поэтому $(A_7A_6, l_3) = -\varphi - 2\varphi_{1,2} - \varphi_{2,3} + \varphi_{1,2} + \varphi_{2,3} = -\varphi - \varphi_{1,2}$.

Решение задачи 0.5.

По лемме 0 для ортотреугольника.

Замечание. Условие задачи можно сформулировать следующим образом.

Пусть точка A_1 лежит на прямой AB. Окружность, описанная около треугольника A_1AC, пересекает прямую BC в точке A_2. Окружность, описанная около треугольника A_2BA, пересекает прямую CA в точке A_3 и т.д. Докажите, что $A_1 = A_7$.

Решение задачи 0.6

По лемме 0 для серединного треугольника.

Решение задачи 0.7

а) В несколько другой, эквивалентной, формулировке теорема Паппа доказывается в книге [1, Глава 1].

б) Рассмотрим шестиугольник $A_1A_2A_3A_6A_5A_4$: прямые A_1A_2, A_3A_6, A_5A_4 пересекаются в точке R, а прямые A_4A_1, A_2A_3, A_6A_5 в точке G. Следовательно "диагонали"A_2A_5, A_3A_4 (последние две прямые уже пересекаются в точке B) и A_6A_1 в одной точке (то есть в точке B). Следовательно $A_7 = A_1$.

Решение задачи 0.8

а) Теорема Брианшона доказывается в книге [1, Глава 1].

б) Несложно убедиться, что прямые A_4A_5, A_5A_6, A_6A_7 симметричны A_4A_3, A_3A_2, A_2A_1 относительно прямой OI. Следовательно $A_1 = A_7$.

Решение задачи 0.9

Первое решение (Д. Якутов). Давайте посчитаем угол $\angle GA_4R$:

$$\begin{aligned}\angle GA_4R &= \pi - \angle GA_4B - \angle BA_4R\\ &= \pi - \angle GOB - \angle BA_5R\\ &= \pi - \angle GOB - (\pi - \angle GA_5R - \angle GA_5B)\\ &= \angle GA_5R + \angle GA_5B - \angle GOB\\ &= \angle GOR - \angle GOB + \angle GA_6B\\ &= \angle GOR - \angle GOB + (\pi - \angle GA_6R - \angle BA_6R)\\ &= \angle GOR - \angle GOB + (\pi - \angle GA_7R - \angle BOR)\\ &= (\pi - \angle BOR - \angle GOB + \angle GOR) - \angle GA_7R.\end{aligned}$$

Следовательно, $\angle GA_4R + \angle GA_7R = \pi - \angle BOR - \angle GOB + \angle GOR$.

По аналогичным соображениям $\angle GA_1R + \angle GA_4R = \pi - \angle BOR - \angle GOB + \angle GOR$. А значит, $\angle GA_1R = \angle GA_7R$. Также $\angle GA_1B = \angle GOB = \angle GA_7B$. Тогда точки G, B, A_1, A_7 — на одной окружности, и точки G, R, A_1, A_7 — тоже на одной окружности. Но эти две окружности имеют не более двух точек пересечения, одна из которых G, при этом $A_1 \neq G$ и $A_7 \neq G$, откуда следует, что $A_1 = A_7$.

Второе решение. Совершим инверсию с центром в точке O и произвольным радиусом. В результате красная, синяя и зеленая окружности перейдут в прямые. Пусть красная точка перейдет в точку C, синяя в точку A, а зеленая в точку B. Теперь будем следить за точками A_i. Через точку A_1 ($\in AB$) мы проводим зеленую (то есть проходящую через A и C) окружность, которая пересекается в точке A_2 с синей "окружностью" — прямой BC. Через точку A_2 проводим красную окружность, которая пересекается в точке A_3 с зеленой "окружностью" — прямой AC. Через точку A_3 проводим синюю окружность до пересечения в A_4 с красной "окружностью" — прямой AB и т.д.

Таким образом, мы получили переформулировку задачи 0.5, описанную в замечании к решению той задачи. Значит $A_1 = A_7$.

`Решение задач раздела 1.`

Большинство решений задач разделов 1,2,3 опираются на результат задачи 4.4.

Решение задачи 1.1.

Будем использовать результат задачи 4.4.

- Для каждого $t \in \mathbb{R}$ зададим гомотетию $H_O^{2^t}$ с центром в начале координат O и коэффициентом 2^t. Несложно проверить, что для любой точки A выполняется $H_O^{2^{t+s}}(A) = H_O^{2^t}(H_O^{2^s}(A))$. Множество γ_A тогда представляет собой лучи, выходящие из начала координат.

- Проведем через точку $A(1,1)$ прямые $y = 1$ (это γ_1) и $x = 1$ (это γ_2). Через каждую точку $B \in \gamma_1$ проведем луч γ_B. Покрасим такие лучи в красный цвет. Теперь покрасим в зеленый и синие цвета прямые $H_O^{2^t}(\gamma_1)$ и $H_O^{2^t}(\gamma_2)$ соответственно, то есть будут прямые, параллельные осям Ox и Oy. Очевидно, что любые цветные лучи или прямые не пересекаются.

- Рассмотрим круг радиуса 1 с центром в точке (1;1). Несложно убедиться, что через каждую точку этого круга можно провести ровно одну прямую (или луч) каждого цвета.

Следовательно, из задачи 4.4 получаем, что данные лучи и прямые образуют ткань. Нетрудно убедиться, что и предложенные в условии задачи цветные прямые тоже образуют ткань.

Решение задачи 1.2.

Первое решение. Из решением задачи 0.7б) следует, что это ткань.

Второе решение. Сведём данную задачу к задаче 1.1. Для этого совершим проективное преобразование, переводящее прямую, проходящую через две фиксированные точки, в бесконечно удалённую прямую.

Решение задачи 1.3.

Из решения задачи 0.8б) следует, что это ткань.

Решение задачи 1.4.

Из решения задачи 0.8а) следует, что это ткань.

Решение задачи 1.5.

Указанные прямые не образуют ткань.

Решение задачи 1.6.

Указанные прямые не образуют ткань.

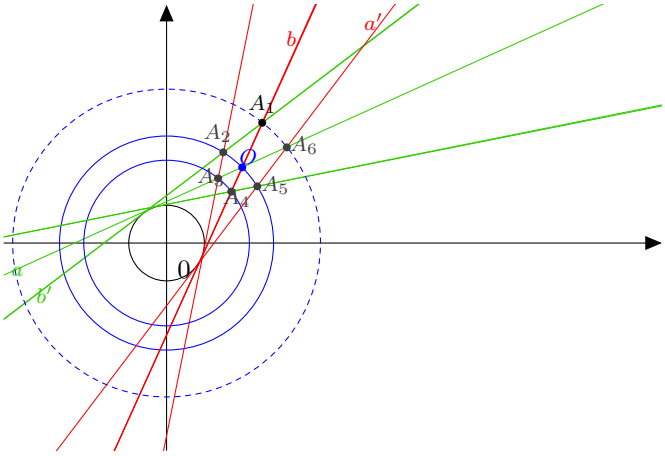
`Решение задач раздела 2.`

Решение задачи 2.1.

В качестве примера возьмём образ при инверсии ткани из прямых, параллельных сторонам треугольника.

Решение задачи 2.2.

Первое решение (Е. Стрельцова). Докажем, что данные прямые и окружности образуют ткань; см. рисунок снизу. Выберем круг в правой верхней четверти плоскости так, чтобы он не имел общих точек с единичным кругом и находился выше прямой $y = 1$. Радиус круга возьмем < 1. Теперь через каждую точку круга проходит ровно одна красная и ровно одна зеленая прямые,так как из любой точки можно провести ровно одну касательную каждого цвета. Через каждую точку T проходит ровно одна окружность с центром в начале координат (Z), потому что с фиксированным радиусом (ZT) и с фиксированным центром (Z) можно провести ровно одну окружность. Окружности с центром Z не могут совпасть с касательными к единичной окружности. А зеленые и красные прямые не могут совпасть, так как круг выше прямой $y = 1$. Концентрические окружности не могут касаться друг друга. А зеленые и красные прямые не могут касаться окружностей с центром Z, потому что у этих окружностей радиус > 1 и так как выбранный круг не имеет с единичным общих точек. Касательные пересекают эти окружности, так как проходят через точки внутри кругов (заключенных этими окружностями), которые полностью содержат единичный круг. Значит, условие слоения выполняется.



Зеленая (a) и красная (b) прямые через точку O симметричны относительно прямой ZO. Значит, $A_3 = S_{ZO}(A_4)$. Тогда красная прямая через A_3 симметрична зеленой прямой через A_4 относительно прямой ZO. Поэтому $A_2 = S_{ZO}(A_5)$. Получаем, что зеленая прямая через A_2 (b') симметрична красной прямой через A_5 (a').

Далее, $a = S_{ZO}(a'), b = S_{ZO}(b')$. Поэтому $A_6 = a \cap a' = S_{ZO}(b \cap b') = S_{ZO}(A_1)$. Значит, $A_6 = S_{ZO}(A_1)$. Тогда $ZA_6 = S_{ZO}(ZA_1)$. Следовательно, $ZA_6 = ZA_1$, то есть синяя окружность через точку A_6 проходит через точку A_1. Мы проверили, что выполняется условие замыкания.

Второе решение. Будем использовать результат задачи 4.4.

- Для каждого $t \in \mathbb{R}$ зададим поворот $R_O^{\pi t}$ вокруг начала координат O на угол πt. Несложно проверить, что для любой точки A выполняется $R_O^{\pi(t+s)}(A) = R_O^{\pi t}(R_O^{\pi s}(A))$, где $t, s \in \mathbb{R}$. Множество γ_A тогда представляет собой дуги окружностей с центром в начале координат.

- Проведем через точку $A(0, 2)$ лучи $y = \sqrt{3}x + 2, x > -\sqrt{3}/2$ (это γ_1) и $y = -\sqrt{3}x + 2, x < \sqrt{3}/2$ (это γ_2). Эти касательные к единичной полуокружности. Через каждую точку $B \in \gamma_1$ проведем

дугу γ_B окружности с началом в начале координат. Покрасим такие дуги в красный цвет. Теперь покрасим в зеленый и синие цвета лучи $R_O^{\pi t}(\gamma_1)$ и $R_O^{\pi t}(\gamma_2)$ соответственно, то есть это лучи, являющиеся касательными к единичными полуокружностям.

- Рассмотрим круг радиуса $1/2$ с центром в точке $(0;2)$. Несложно убедиться, что через каждую точку этого круга можно провести ровно одну дугу или луч каждого цвета.

Следовательно, из задачи 4.4 получаем, что данные дуги и лучи образуют ткань. Нетрудно убедиться, что и предложенные в условии задачи цветные прямые и окружности тоже образуют ткань.

Решение задачи 2.3.
Будем использовать результат задачи 4.4.

- Для каждого $t \in \mathbb{R}$ зададим поворот $R_O^{\pi t}$ вокруг начала координат O на угол πt. Несложно проверить, что для любой точки A выполняется $R_O^{\pi(t+s)}(A) = R_O^{\pi t}(R_O^{\pi s}(A))$, где $t, s \in \mathbb{R}$. Множество γ_A тогда представляет собой дугу окружности с центром в начале координат.

- Проведем через точку $A(0,2)$ лучи $y = \sqrt{3}x + 2, x > -\sqrt{3}/2$ (это γ_1) и $x = 0, y > 0$ (это γ_2). Это касательная к единичной полуокружности и луч, выходящий из начало координат. Через каждую точку $B \in \gamma_1$ проведем дугу γ_B окружности с началом в начале координат. Покрасим такие дуги в красный цвет. Теперь покрасим в зеленый и синие цвета лучи $R_O^{\pi t}(\gamma_1)$ и $R_O^{\pi t}(\gamma_2)$ соответственно.

- Рассмотрим круг радиуса $1/2$ с центром в точке $A(0;2)$. Несложно убедиться, что через каждую точку этого круга можно провести ровно одну прямую каждого цвета.

Следовательно, из задачи 4.4 получаем, что данные лучи и дуги образуют ткань. Нетрудно убедиться, что и предложенные в условии задачи цветные прямые и окружности тоже образуют ткань.

Решение задачи 2.4.
Будем использовать результат задачи 4.4.

- Для каждого $t \in \mathbb{R}$ зададим параллельный перенос $T_{(0,t)}$ на вектор $(0,t)$. Несложно проверить, что для любой точки A выполняется $T_{(0,t+s)}(A) = T_{(0,t)}(T_{(0,s)}(A))$, где $t, s \in \mathbb{R}$. Множество γ_A тогда представляет собой прямые, параллельные оси Oy.

- Проведем через точку $A(1/2 + 1/\sqrt{8}, 1/2 + 1/\sqrt{8})$ дугу окружности $(x - 1/2)^2 + (y - 1/2)^2 = 1/4, x > 1/2, y > 1/2$ (это γ_1) и прямую $y = 1/2 + 1/\sqrt{8}$ (это γ_2). Через каждую точку $B \in \gamma_1$ проведем прямые γ_B, параллельные оси Oy. Покрасим такие прямые в красный цвет. Теперь покрасим в зеленый и синие цвета дуги $T_{(0,t)}(\gamma_1)$ и прямые $T_{(0,t)}(\gamma_2)$ соответственно.

- Рассмотрим круг радиуса $1/2 - 1/\sqrt{8}$ с центром в точке $A(1/2 + 1/\sqrt{8}; 1/2 + 1/\sqrt{8})$. Несложно убедиться, что через каждую точку этого круга проходит ровно одна дуга обобщенной окружности каждого цвета.

Следовательно, из задачи 4.4 получаем, что данные дуги образуют ткань. Из этого следует, что и предложенные в условии задачи цветные прямые и окружности тоже образуют ткань.

Решение задачи 2.5.

Действуем в соответствии с задачей 4.4. В качестве преобразований рассмотрим гомотетии с центром в начале координат. Тогда красные прямые — прямые, проходящие через начало координат. Зеленые окружности — окружности, касающиеся первой пары отрезков. Синие окружности — окружности, касающиеся другой пары указанных отрезков.

Из задачи 4.4 получаем, что рассматриваемые прямые и подходящие дуги рассматриваемых окружностей образуют ткань.

Решение задачи 2.6.

Действуем в соответствии с задачей 4.4. В качестве преобразований рассмотрим гомотетии с центром в начале координат. Тогда красные прямые — прямые, проходящие через начало координат. Зеленые окружности — окружности, с центром в начале координат. Синие окружности — окружности, касающиеся пары указанных отрезков.

Из задачи 4.4 получаем, что рассматриваемые множества прямых и окружностей образуют ткань. Из этого следует, что и предложенные в условии задачи цветные прямые и окружности тоже образуют ткань.

`Решение задач раздела 3.`

Можно решить задачу, аналогичную задаче 4.4, и для тора и для гиперболоида вращения.

Решение задачи 3.1.

Действуем в соответствии с задачей 4.4. В качестве преобразований рассмотрим повороты вокруг оси вращения. Тогда красные окружности — параллели. Зеленые и синие окружности — окружности Вилларсо.

Из задачи 4.4 получаем, что рассматриваемые множества прямых и окружностей образуют ткань.

Решение задачи 3.2. Рассмотрим точку O. Проведем через неё меридиану γ_1 и окружности Вилларсо γ_2, γ_3. Все меридианы покрасим в красный цвет, окружности Вилларсо, получаемые из γ_2 поворотом, покрасим в зеленый цвет, окружности Вилларсо, получаемые из γ_3 поворотом, покрасим в синий цвет. Рассмотрим шар с центром в точке O и радиуса $R < \frac{r}{100}$ (r — расстояние между γ_1 и осью вращения). Внутри него две окружности Вилларсо пересекаются не более, чем в одной точке. Пересечение шара и тора обозначим через Ω.

Рассмотрим произвольную точку O', лежащую в $\mathbf{\Omega}$. Проведем через неё красный меридиан w_1, синюю окружность w_2 и зеленую окружность w_3. Пусть все точки A_i, которые будут получаться в результате постороения будут лежать в $\mathbf{\Omega}$. Пусть выбрана точка $A_1 \in w_1$. Проведём через неё зеленую окружность w_2'. Получим точку A_2 пересечения w_2' и w_3. Построим через точку A_2 красную окружность w_1'. Получаем точку A_3 пересечения w_1' и w_2. Построим через точку A_3 синюю окружность w_3'. Получим точку A_4 пересечения w_3' и w_1. Проведём через A_4 зеленую окружность w_2''. Получим точку A_5 пересечения w_2'' и w_3. Построим через точку A_5 красную окружность w_1''. Получаем точку A_5 пересечения w_1'' и w_2. Построим через точку A_6 синюю окружность w_3''. Получим точку A_7 пересечения w_3'' и w_1.

Пусть α плоскость, содержащая окружность w_1. Несложно убедиться, что окружности w_3',w_1', w_2' симметричны w_2'', w_1'', w_3'' относительно плоскости α. Следовательно, $A_1 = A_7$.

Решение задачи 3.3.

Действуем в соответствии с задачей 4.4. В качестве преобразований рассмотрим повороты вокруг оси вращения. Тогда красные окружности — параллели. Зеленые и синие окружности — прямые, лежащие в гиперболоиде вращения.

Из задачи 4.4 получаем, что рассматриваемые множества прямых и окружностей образуют ткань.

`Указания и решения задач раздела 4.`

Указание к задаче 4.1.

Перечислим несколько возможных примеров множеств синих обобщённых окружностей:

- (В) произвольный пучок прямых (Задачи 1.1 и 1.2);
- (В) концентрические окружности;
- (В) дуги окружностей, полученные из некоторой одной дуги с помощью параллельного переноса вдоль оси Ox или Oy (Задача 4.3);

Из теоремы Графа–Зауэра (Задача 4.12) следует, что не существует других примеров синих прямых. Из классификации Шелехова всех тканей, образованных пучками окружностей, [3, Теорема 0.1] следует, что не существует других примеров, для которых множество синих окружностей является пучком. Описание всех возможных примеров синих окружностей, не обязательно состоящих из пучков, является открытой проблемой.

Указание к задаче 4.2.

Перечислим несколько возможных примеров множеств синих обобщённых окружностей:

- (В) произвольный пучок прямых (Задачи 1.1 и 1.2);
- (В) пучок окружностей с предельной точкой в начале координат O и общей радикальной осью параллельной оси Ox.
- (В) дуги окружностей, полученные из некоторой одной дуги с помощью гомотетий с центром в начале координат (Задача 4.3).

Из теоремы Графа–Зауэра (Задача 4.12) следует, что не существует других примеров синих прямых. Из классификации Шелехова всех тканей, образованных пучками окружностей, [3, Теорема 0.1] следует, что не существует других примеров, для которых множество синих окружностей является пучком. Описание всех возможных примеров синих окружностей, не обязательно состоящих из пучков, является открытой проблемой.

Указание к задаче 4.3.

Рассмотрим инверсию с центром в одной из предельных точек. Полученные пучки обобщенных окружностей образуют ткань по задаче 4.4.

Решение задачи 4.4.

Условие слоения выполняется в соответствии с третьим условием задачи. Покажем, что выполняется условие замыкания.

Возьмём произвольную точку O внутри круга. Проведем через нее красную (w_1), зеленую (w_2) и синюю (w_3) дуги обобщённых окружностей. Пусть все точки A_i, которые будут получаться в результате построения будут лежать в $\mathbf{\Omega}$. Пусть точка $A_1 \in w_1$ и $t \in \mathbb{R}$ таково, что $R_t(O) = A_1$ (в соответствии с первым условием леммы такое t найдётся: пусть $w_1 = \gamma_X$ для некоторой точки $X \in \gamma_1$, тогда существуют такие $y, z \in \mathbb{R}$, что $R_y(X) = O$ и $R_z(X) = A_1$, тогда $R_{z-y}(O) = R_{z-y}(R_y(X)) = R_z(X) = A_1$, то есть $t = z - y$). Проведём через точку A_1 зеленую дугу w_2' обобщённой окружности. Получим точку A_2 пересечения w_2' и w_3. Построим через точку A_2 красную дугу w_1' обобщённой окружности. Получим точку A_3 пересечения w_1' и w_2. Построим через точку A_3 синюю дугу w_3' обобщённой окружности. Получим точку A_4 пересечения w_3' и w_1. И т.д.

Теперь покажем, что $R_t(O) = A_7$, отсюда будет следовать, что $A_1 = A_7$.

Нам известно, что $R_t(O) = A_1$, поэтому $R_t(w_2) = w_2'$ (это верно в связи с тем, что $R_t(w_2)$ — зеленая дуга, проходящая через A_1, а зеленых дуг, кроме w_2', проходящих через точку A_1, нет), следовательно, $R_t(A_3) \in w_2' \cap w_1' = A_2$. Так как $R_t(A_3) = A_2$, то $R_t(A_4) = O$ (из аналогичных соображений). Так как $R_t(A_4) = O$, то $R_t(A_5) = A_6$. Так как $R_t(A_5) = A_6$, то $R_t(O) = A_7$. Задача решена.

Указание к задачам 4.5–4.6.

Данные задачи рассматриваются в [4].

Указание к задаче 4.7.

Решение данной задачи дано в книге Прасолова и Соловьёва [2].

Указание. Пусть уравнения красных прямых — $a_1x+b_1y-1=0$, $a_2x+b_2y-1=0$, $a_3x+b_3y-1=0$, а уравнения синих — $c_1x + d_1y - 1 = 0$, $c_2x + d_2y - 1 = 0$, $c_3x + d_3y - 1 = 0$. Докажите, что тогда уравнение кривой можно записать в виде

$$p(a_1x + b_1y - 1)(a_2x + b_2y - 1)(a_3x + b_3y - 1) + q(c_1x + d_1y - 1)(c_2x + d_2y - 1)(c_3x + d_3y - 1) = 0$$

для некоторых действительных чисел p и q.

Указание к задаче 4.8.

Эта задача получается из предыдущей с помощью проективной двойственности.

Указание к задаче 4.9.

Использовать задачу 4.8.

Указание к задачам 4.10–4.11.

Использовать задачу 4.9. Рисунок к задаче 4.10, принадлежащий А. Ганеийану Себдани и Е. Ашуриуну, приводится на первой странице решений.

Список литературы

[1] Zaslavsky A.A. Akopyan A.V. *Geometry of conics.* AMS, 2007.

[2] V. Prasolov and Yu. Soloviev. *Elliptic functions and algebraic equations.* Moscow: Factorial, 1997.

[3] A. M. Shelekhov. Classification of regular three-webs formed by pencils of circles. *J. Math. Sciences*, 143(6):3607–3629, 2007.

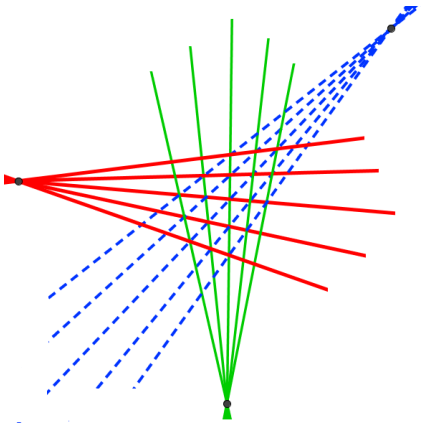
[4] S. L. Tabachnikov. Geometry of equations. *Kvant*, 10, 1988. in Russian.

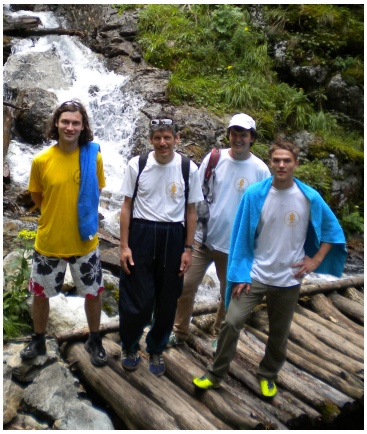
0.9 Теорема Бляшке (Вовченко Владислав)

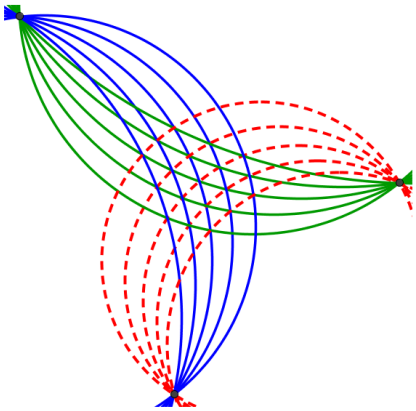
Пусть $R_1G_1B_1$ равносторонний треугольник с высотой 1. Тогда для каждой внутренней точки $R_1G_1B_1$ сумарное расстояние ее до сторон R_1G_1, G_1B_1 и R_1B_1 будет равно 1. Пусть функция f определена на множестве внутренних точек RGB такая, что для точки X внутри RGB (обозначем углы RXG, GXB и RXB α, β, μ соответственно) $f(X)$ точка внутри $R_1G_1B_1$ такая, что расстояние $f(X)$ до сторон R_1G_1, G_1B_1 и R_1B_1 будет $\frac{\alpha}{2\pi}$, $\frac{\beta}{2\pi}$ и $\frac{\mu}{2\pi}$ соответственно. Тогда зеленые дуги перейдут в паралельные отрезки и аналогично для красных и синих дуг. Тогда $f(A_1) = f(A_7)$, значит $\mathrm{A}_1 = \mathrm{A}_7$. Что и требовалось доказать.

Webs from lines and circles

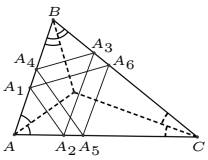
Alexey Zaslavskiy, Fedor Nilov, Alexander Polyanskiy, Mikhail Skopenkov



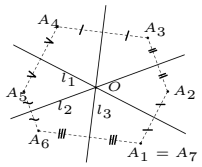




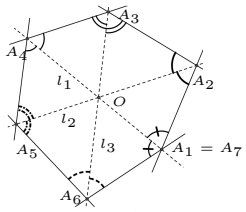
Closure with period 6

0.1. 

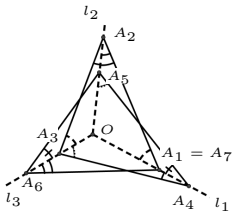
Through a point A_1 belonging to the side AB of a triangle ABC one draws a line orthogonal to the bissector of the angle A. The drawn line intersects the side AC at a point A_2. Through the point A_2 one draws a line orthogonal to the bissector of the angle C. The line intersects the side CB at a point A_3. Analogously one gets points A_4, A_5, A_6, A_7. Prove that $A_7 = A_1$.

0.2. 

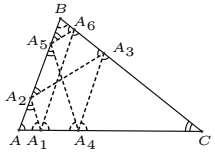
Three lines l_1, l_2, l_3 have a common point. Let A_1 be an arbitrary point of the plane. The points A_2, A_3, A_4, A_5, A_6, A_7 are obtained from A_1 by consecutive. reflections with respect to the lines l_1, l_2, l_3, and then l_1, l_2, l_3 again. Prove that $A_7 = A_1$.

0.3. 

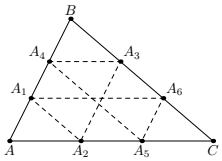
Three lines l_1, l_2, l_3 have a common point. On the lines l_1 and l_2 one takes arbitrary points A_1 and A_2, respectively. The point A_3 is the intersection of the line l_3 and the line symmetric to A_1A_2 with respect to l_2. The point A_4 is the intersection of the line l_1 and the line symmetric to A_2A_3 with respect to l_3. Analogously one gets the points A_5, A_6, A_7. Prove that $A_7 = A_1$.

0.4. 

The F. Petrov Theorem. Three rays l_1, l_2, l_3 have a common starting point O. Take arbitrary points A_1 and A_2 on the rays l_1 and l_2, respectively. Take a point A_3 on the ray l_3 such that the angle between A_2A_3 and l_3 equals the angle between A_1A_2 and l_1. Take a point A_4 on the ray l_1 such that the angle between A_3A_4 and l_1 equals the angle between A_2A_3 and l_2. Analogously get the points A_5, A_6, A_7. Prove that $A_7 = A_1$.

0.5. 

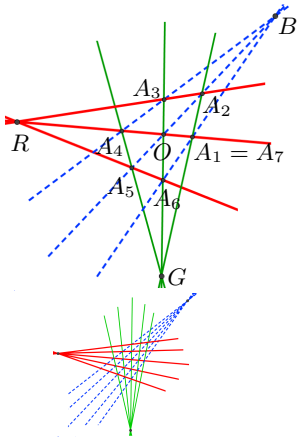
A billiard table has the shape of a triangle ABC. A billiard ball starts its movement from a point A_1 of the side AC under the angle ABC to the side. Denote by A_2, A_3, A_4, A_5, A_6, A_7 the points where the ball consecutively hits the borders of the table. Prove that $A_7 = A_1$.

0.6. 

Through a point A_1 on the side AB of a triangle ABC one draws a line parallel to BC. The line intersects CA at a point A_2. Through the point A_2 one draws a line parallel to AB. The drawn line intersects BC at a point A_3. Analogously one gets the points A_4, A_5, A_6, A_7. Prove that $A_7 = A_1$.

[0]Summer conference of the International mathematical Tournament of towns, August 2–10, 2012

0.7.

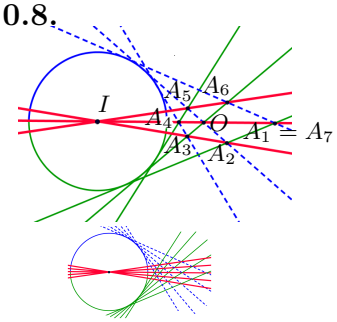


The Pappus Theorem.

a) Suppose that the lines containing the sides of a nonconvex hexagon pass alternately through two fixed points. Then the lines containg the diagoinals either have a common point or are parallel to each other.

b) A red (R), a green (G) and a blue (B) points are marked in the plane (see the figure to the left). Each line passing through exactly one of the marked point is paint the same color as the point. Take an arbitrary point O inside the triangle RGB. Draw the red, the green, and the blue line through the point. On the red line take an arbitrary point A_1 inside the triangle RGB. Draw the green line through the point. Suppose that the green line intersects the blue line through the point O at a point A_2. The green and the blue line through the point A_2 have already been drawn; draw the red line through A_2. The intersection point of the obtained red line with the green line through the point O is denoted by A_3. Continuing this construction we get the points A_4, A_5, A_6, A_7. Prove that $A_7 = A_1$.

0.8.

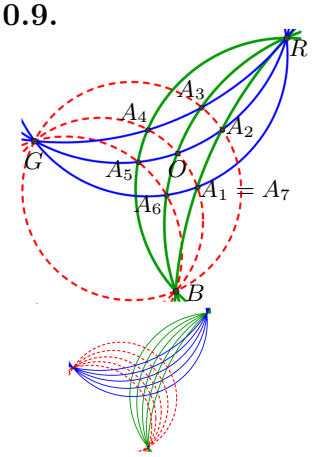


The Brianchon Theorem.

a) The diagonals of a circumscribed hexagon have a common point.

b) A circle with center I and a point O outside the circle are given. The lines passing through the center I are paint red. The tangent lines to the circle are paint either green or blue depending on the position of their common point with the circle with respect to the line OI (see the figure to the left). On the line OI take an arbitrary point A_1. Draw the green line through the point. Suppose that the green line intersects the blue line through the point O at a point A_2. The green and the blue line through the point A_2 have already been drawn; now draw the red one. The intersection point of the obtained red line with the green line through the point O is denoted by A_3. Continuing this construction we get the points A_4, A_5, A_6, A_7. Prove that $A_7 = A_1$.

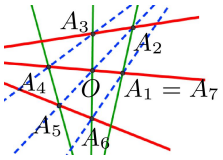
0.9.



The Blaschke Theorem. A red (R), green (G), and blue (B) point are marked in the plane (see the figure to the left). Each circle passing through exactly two of the points is paint the color of the remaining point. Take an arbitrary point O inside the triangle RGB. Draw the red, the green, and the blue circle through the point. On the red circle take an arbitrary point A_1 inside the triangle RGB. Draw the green circle through the point. Suppose that the green circle intersects the blue circle through the point O at a point A_2 distinct from R, G, and B. The green and the blue circles through the point A_2 have already been drawn; draw the red circle through A_2. The intersection point (distinct from R, G, B) of the obtained red circle with the green circle through the point O is denoted by A_3. Continuing this construction we get the points A_4, A_5, A_6, A_7. Prove that $A_7 = A_1$.

Definition of a web

Definition.

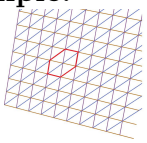


Suppose that some lines in the plane are paint red, green, and blue. The paint lines form a (*hexagonal*) *web* if there is a disk Ω satisfying the following 2 conditions:

Foliation condition: For each point A of the disk Ω there exists exactly one line of each colour passing through A; and the lines of different colours do not coincide.

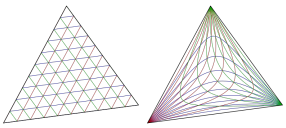
Closure condition (see figure to the left): Take an arbitrary point O inside the disk Ω. Draw the red, the green, and the blue line through the point. On the red line take an arbitrary point A_1 inside the disk Ω. Draw the green line through the point. Suppose that the green line intersects the blue line through the point O at a point A_2. The green and the blue line through the point A_2 have already been drawn; draw the red one. The intersection point of the obtained red line with the green line through the point O is denoted by A_3. Continuing this construction we get the points A_4, A_5, A_6, A_7. *Closure condition* asserts that if all the above points belong to the disk Ω then $A_7 = A_1$.

Example.



Three sets of lines parallel to the sides of a fixed triangle form a web. One can always obtain a triangulation of certain part of the plane from the lines of a web. In the figures of webs we always show such triangulations; e.g., see small bottom figures in Problems 0.7–0.9.

Remark.



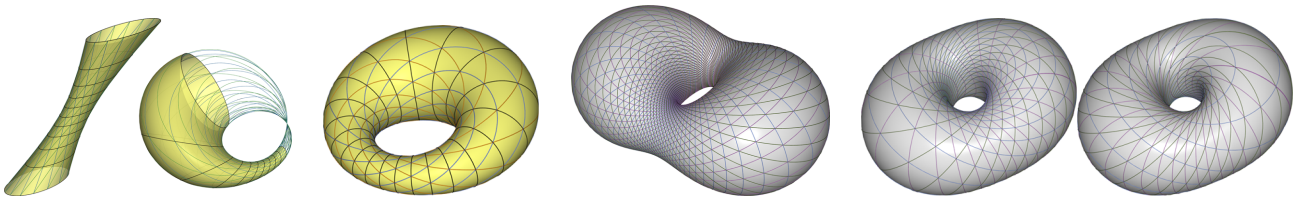
An arbitrary web (from arbitrary curves on arbitrary surface) can be obtained from the previous example by an appropriate continuous bijective map of a part of the plane onto the disk Ω.

Replacing the word "line" by the word "line or circle" in the definition of a web, and requiring that lines and circles pairwise do not touch each other, we get the definition of a *web from lines or circles*.

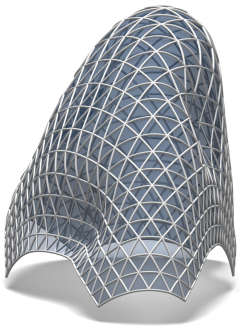
The Blaschke Problem (1920s). *Find all webs of circles in the plane.*

Replacing the word "disk" by the word "intersection of a surface with a ball" in the definition of a web we get the definition of a *web in the surface*.

Remark. Webs from circles are completely described for all the surfaces except the plane and the sphere [3]; some example are shown in the figure below.



Remark.



The interest to webs grows because of possible application in architecture. An important problem in modern architecutre is *rationalization* of freeform surfaces, i.e., their decomposition into relatively simple panels. One of the approaches to rationalization are *circular arc structures*, i.e., triangulations by circular arcs obtained from certain web. In the figure to the left one can see such a structure on Eidhoven Blob by architect M. Fuksas.

1 Webs from lines in the plane

Which of the following triples of sets of lines form a web? Hint: *Geogebra* software could be useful for experiments and drawing figures.

1.1. (R) Lines parallel to the Ox axis; (G) parallel to the Oy axis; (B) passing through the origin O.

1.2. Three sets of lines passing through three fixed pairwise distinct points in the plane.

By the *unit semicircle* we mean the set of points with coordinates satisfying the conditions $x^2+y^2=1$ and $x>0$. By the *complementary* semicircle we mean the set given by the conditions $x^2+y^2=1$ and $x<0$.

1.3. (R) Lines tangent to the unit semicircle; (G) tangent to its complementary semicircle; (B) passing through the origin O.

1.4. (R) Lines tangent to a unit semicircle; (G) tangent to its complementary semicircle; (B) passing through a fixed point.

1.5. (R) Lines parallel to the Ox axis; (G) parallel to the Oy axis; (B) tangent to the unit semicircle.

1.6. (R) Lines tangent to the unit semicircle; (G) passing through the origin O; (B) parallel to the Ox axis.

2 Webs from circles in the plane

2.1. Give an example of a web from circles in the plane (with a proof).

Which of the following triples of sets of lines and circles form a web?

2.2. (R) Lines tangent to the unit semicircle; (G) tangent to its complementary semicircle; (B) circles with centers at the origin.

2.3. (R) Lines tangent to the unit semicircle; (G) passing through the origin; (B) circles with centers at the origin.

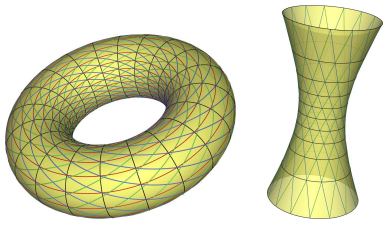
2.4. (R) Lines parallel to the Ox axis; (G) parallel to the Oy axis; (B) circles touching both segments $x = 0, 0 \le y \le 1$ and $x = 1, 0 \le y \le 1$ simultaneously.

2.5. (R) Lines passing through the origin; (G) circles touching both segments $x = 0, 0 \le y \le 1$ and $y = 0, 0 \le x \le 1$ simultaneously; (B) circles touching both segments $x = 0, 2 \le y \le 4$ and $y = 0, 2 \le x \le 4$ simultaneously.

2.6. (R) Lines passing through the origin; (G) circles with centers at the origin; (B) circles touching both segments $x = 0, 0 \le y \le 1$ and $y = 0, 0 \le x \le 1$ simultaneously.

3 3D

A *torus* is the result of rotation of a circle around a line lying in the plane of the circle but not intersecting the circle itself; see the left part of the figure below. Circles obtained as the trajectories of individual points are called *parallels*. The initial circle and all the circles obtained from it by the rotation are called *meridians*. Through each point of the torus one can draw two more circles lying on the torus; they are called the *Villarceau circles* (this can be used without proof in your solutions).



3.1. The parallels of a torus with the Villarceau circles form a web.

3.2. The meridians of a torus with the Villarceau circles form a web.

A *hyperboloid of revolution* is the result of rotation of a line around a skew line; see the right part of the figure above. Through each point of the hyperboloid one can draw two lines lying on the hyperboloid (this can be used without proof).

3.3. The lines lying on a hyperboloid of revolution with its parallels form a web.

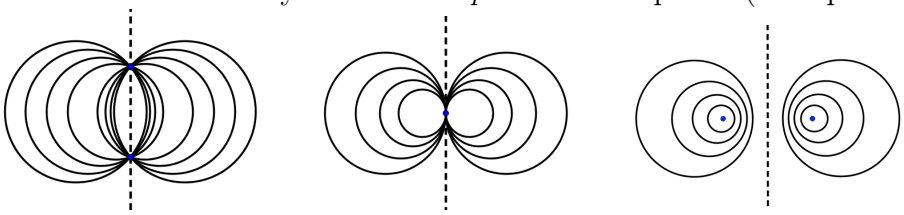
4 Webs from circles: general constructions

Supplement the given sets of red and green lines with a set of blue **a)** lines; **b)** circles to obtain a web (provide as many examples of such blue sets as possible; try to find all the examples and prove that there are no other ones):

4.1. (R) Lines parallel to the Ox axis; (G) parallel to the Oy axis.

4.2. (R) Lines parallel to the Ox axis; (G) passing through the origin O.

A *pencil of lines* is a set of all the lines passing through a fixed point (*vertex*) or parallel to a fixed line. A *pencil of circles* is a set of all the circles having a common radical line; see the figure below. If the pencil contains "circles" of zero radius then they called *limit points* of the pencil (bold points in the figure).



4.3. A pencil of circles with two limit points and two pencils of lines with vertices at these two points form a web.

Remark. All webs formed by three pencils of circles were found by A. Shelekhov [4, Theorem 0.1].

In what follows by *general circles* we mean circles and lines. By *general arcs* we mean circular arcs, circles, segments, rays, and lines. A *web* from general arcs is defined analogously to a web from circles or lines. The following problem can be useful for solution of problems from sections 1–3.

4.4. *Generation of webs using transformation groups.* Suppose that the following conditions hold:

- For each $t \in \mathbb{R}$ there is a map R_t of the plane taking general circles to general circles. For arbitrary $t, s \in \mathbb{R}$ and an arbitrary point A we have $R_t(R_s(A)) = R_{t+s}(A)$. Also for each point A the set $\gamma_A = \{R_t(A) : t \in \mathbb{R}\}$ is a general arc.
- Let γ_1, γ_2 be two distinct general arcs passing through some point. For each point $A \in \gamma_1$ paint the general circle γ_A red. Paint the arcs $R_t(\gamma_1)$, where $t \in \mathbb{R}$, and the arcs $R_t(\gamma_2)$, where $t \in \mathbb{R}$, green and blue, respectively. Suppose that any two colored arcs have at most one common point.
- There exists a disc Ω, such that exactly one arc of each color passes through each point of the disc.

Then the red, green and blue general arcs form a web.

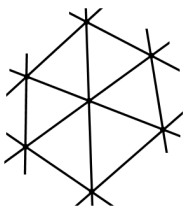
The following series of problems contains more complicated ones (except the very first one).
If a line in the plane has equation $px + qy = 1$ then the pair (p, q) is called the *coordinate* of the line.

4.5. All the lines whose coordinates satisfy a fixed linear equation form a pencil.

4.6. * All the lines whose coordinates satisfy a fixed equation of degree 2 either are tangent to one conic or form two pencils or form one pencil or form an empty set.

4.7. * *The generalized Pascal theorem.* Three red lines intersect three green lines at 9 distinct points. If a curve given by an equation of degree 3 is passing through 8 of these points then it is passing through the remaining point.

4.8. * *The Chasles theorem.* Let 9 lines form hexagonal configuration as in figure below. If the coordinates of 8 of these lines satisfy an equation of degree 3 then the coordinate of the remaining line satisfies the same equation.



A set of uncolored general circles is *a web*, if there is a disk Ω satisfying the following conditions:

- *Generalized foliation condition.* Through each point of the disk Ω there passes a finite nonzero number of general circles of the set.
- One can paint some of the general circles of the set red, green, and blue so that the paint ones form a web in the disk Ω.

4.9. * If a set of lines in the plane satisfies the generalized foliation condition in a disk and their coordinates satisfy a fixed equation of degree 3 then these lines form a web.

4.10. * Normal lines to a parabola form a web.

4.11. * Simson lines in a triangle form a web.

We conclude this list of problems by a few very hard ones.

4.12. ** *The Graf–Sauer Theorem* ([2, 1]). Suppose that a set of lines satisfies the generalized foliation condition. Then these lines form a web if and only if in some cartesian coordinate system their coordinates satisfy a fixed equation of degree 3.

Remark. This result allows to find all webs from circles orthogonal to a fixed circle [5].

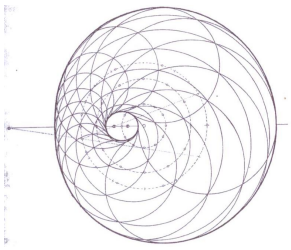
A *cyclic* is a curve given by equation of the form

$$\lambda(x^2+y^2)^2+(x^2+y^2)(\mu x+\nu y)+Q(x,y)=0,$$

where $\lambda,\mu,\nu\in\mathbb{R}$ and $Q(x,y)$ is a polynomial of degree at most 2.

The following problem is addressed to the readers who know the definitions of complex points and continuous families of circles.

4.13. ** *The Wunderlich Theorem* ([6]; see figure below). Three continuous families of circles are doubly tangent (possibly in complex points) to a cyclic. If these families satisfy the generalized foliation condition then they form a web.



4.14. *** Give an example of a web of circles different from the above.

5 Acknowlegdements

The authors are grateful to I. Bogdanov for useful discussions.

Список литературы

[1] Wilhelm Blaschke and Gerrit Bol. *Geometrie der Gewebe.* Springer, 1938.

[2] H. Graf and R. Sauer. Über dreifache Geradensysteme in der Ebene, welche Dreiecksnetze bilden. *Sitz. Bayer. Akad. Math.-nat. Abt.*, pages 119–156, 1924.

[3] Helmut Pottmann, Ling Shi, and Mikhail Skopenkov. Darboux cyclides and webs from circles. *Computer Aided Geometric Design*, 29(1):77 – 97, 2012.

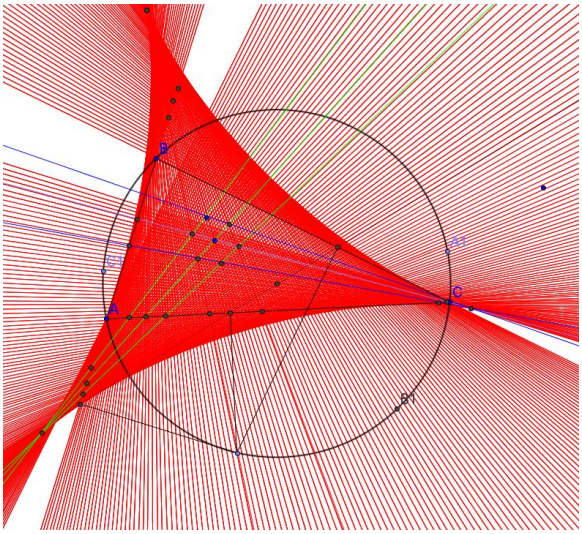
[4] A. M. Shelekhov. Classification of regular three-webs formed by pencils of circles. *J. Math. Sciences*, 143(6):3607–3629, 2007.

[5] K. Strubecker. Über ein Klasse spezieller Dreiecksnetze aus Kreisen. *Monaths. Math. Phys.*, 39:395–398, 1932.

[6] W. Wunderlich. Über ein besonderes Dreiecksnetz aus Kreisen. *Sitzungsber. Akad. Wiss. Wien*, 147:385–399, 1938.

Webs from lines and circles

Alexey Zaslavskiy, Fedor Nilov, Alexander Polyanskiy, Mikhail Skopenkov



Solutions of the problems

Next lemma allows to solve problems 0.1, 0.5, 0.6.

Lemma 0. Let $A'B'C'$ be a cevian triangle of some point wrt triangle ABC (i.e the lines AA', BB', CC' concur). The line passing through an arbitrary point M_1 of side AC and parallel to $A'B'$ intersect BC in point M_2; the line padding through M_2 and parallel to $A'C'$ intersect AB in point M_3 etc. Then $M_1 = M_7$.

`Proof.`

If M_1 coincide with B_1, then M_4 and M_7 also coincide with B'. Else by Ceva theorem $\frac{\overline{AB'}}{\overline{B'C}}\frac{\overline{CA'}}{\overline{A'B}}\frac{\overline{BC'}}{\overline{C'A}} = 1$. And by Thales theorem $\frac{\overline{B'C}}{\overline{CA'}} = \frac{\overline{B'M_1}}{\overline{M_2A'}}$, $\frac{\overline{A'B}}{\overline{BC'}} = \frac{\overline{M_2A'}}{\overline{C'M_3}}$, $\frac{\overline{C'A}}{\overline{AB'}} = \frac{\overline{C'M_3}}{\overline{M_4B'}}$. Placing three last equalities into the first one we obtain that: $\frac{\overline{B'M_1}}{\overline{M_4B'}} = 1$. Thus M_1 and M_4 are symmetric wrt B'. Similarly M_4 and M_7 are symmetric wrt B'. Therefore $M_1 = M_7$.

`Solutions of problems of part 0.`

Solution of problem 0.1

First solution. Let a, b, c be the side lengths of BC, CA, and AB, respectively. Let x be the *signed* length of AA_1 (i.e., the length of AA_1 taken with positive sign, if the vectors $\overrightarrow{AA_1}$ and $\overrightarrow{AB}$ have the same orientation, and with negative sign, if they have opposite orientation). Since A_1A_2 is orthogonal to the the bissector of the angle BAC it follows that $AA_2 = AA_1 = x$ (with signs). Analogously we find consecutively $CA_3 = CA_2 = b - x$, $BA_4 = a - b + x$, $AA_5 = c - a + b - x$, $CA_6 = a - c + x$, $AA_7 = x$ (with signs). Since $AA_7 = x = AA_1$ (with signs) it follows that $A_7 = A_1$.

Second solution. Use lemma 0 to Gergonne triangle.

Solution of problem 0.2

In problems 0.2, 0.3 and 0.4 we consider the angles between the directions (vectors).

Let O be the common point of l_1, l_2, l_3. Let $(l_1, l_2) = \varphi_{1,2}$,$(l_2, l_3) = \varphi_{2,3}$, $(OA_1, l_1) = \varphi$. Since the lengths of segments OA_i are equal (the symmetry conserve the length), it is sufficiently to prove that $(OA_7, l_1) = (OA_1, l_1) = \varphi$.

[0]Summer conference of the International mathematical Tournament of towns, August 2–10, 2012

Since $(OA_2, l_1) = -(OA_1, l_1) = -\varphi$, then $(OA_2, l_2) = -\varphi + \varphi_{1,2}$. Thus $(OA_3, l_2) = \varphi - \varphi_{1,2}$. From this we obtain that $(OA_3, l_3) = \varphi - \varphi_{1,2} + \varphi_{2,3}$. Therefore $(OA_4, l_3) = -\varphi + \varphi_{1,2} - \varphi_{2,3}$. This yields that $(OA_4, l_1) = -\varphi + \varphi_{1,2} - \varphi_{2,3} - \varphi_{1,2} - \varphi_{2,3} = -\varphi - 2\varphi_{2,3}$. Similarly we obtain that $(OA_7, l_1) = -(OA_4, l_1) - 2\varphi_{2,3} = \varphi$.

Solution of problem 0.3

Let $(l_1, l_2) = \varphi_{1,2}$,$(l_2, l_3) = \varphi_{2,3}$ and $(l_{(1,2)}, l_1) = \varphi$, where $l_{(i,i+1)}$ ix the vector from O to the projection of O to the line A_iA_{i+1}. Since the distances from O to all lines A_iA_{i+1} are equal (the symmetry conserve the length), it is sufficiently to prove that $(l_{(7,8)}, l_1) = (l_{(1,2)}, l_1) = \varphi$.

Since $(l_{(2,3)}, l_1) = -\varphi$ then $(l_{(2,3)}, l_2) = (l_{(2,3)}, l_1) + (l_1, l_2) = -\varphi + \varphi_{1,2}$. Similarly $(l_{(3,4)}, l_2) = \varphi - \varphi_{1,2}$. Thus $(l_{(3,4)}, l_3) = (l_{(3,4)}, l_2) + (l_2, l_3) = \varphi - \varphi_{1,2} + \varphi_{2,3}$. Also $(l_{(4,5)}, l_3) = -\varphi + \varphi_{1,2} - \varphi_{2,3}$. Therefore $(l_{(4,5)}, l_1) = (l_{(4,5)}, l_3) + (l_1, l_3) = -\varphi + \varphi_{1,2} - \varphi_{2,3} + (-\varphi_{1,2} - \varphi_{2,3}) = -\varphi - 2\varphi_{2,3}$. Similarly we obtain that $(l_{(7,8)}, l_1) = -(l_{(4,5)}, l_1) - 2\varphi_{2,3} = \varphi$.

Solution of problem 0.4.

Let $(l_1, l_2) = \varphi_{1,2}$,$(l_2, l_3) = \varphi_{2,3}$, $(A_1A_2, l_1) = \varphi$. Since the circumradii of triangles OA_iA_{i+1} are equal (by the sinus theorem), it is sufficiently to prove that $(A_7A_6, l_3) = -(A_1A_2, l_2) = -\varphi - \varphi_{1,2}$ (by the inverse sinus theorem).

Since $(A_3A_2, l_3) = -(A_1A_2, l_1) = -\varphi$ then $(A_3A_2, l_2) = -\varphi - \varphi_{2,3}$. Since $(A_3A_4, l_1) = -(A_3A_2, l_2) = \varphi + \varphi_{2,3}$ then $(A_3A_4, l_3) = \varphi + \varphi_{1,2} + \varphi_{2,3} + \varphi_{2,3}$. Since $(A_5A_4, l_2) = -(A_3A_4, l_3) = -\varphi - \varphi_{1,2} - 2\varphi_{2,3}$ then $(A_5A_4, l_1) = -\varphi - \varphi_{1,2} - 2\varphi_{2,3} - \varphi_{1,2}$. Since $(A_5A_6, l_3) = -(A_5A_4, l_1) = \varphi + 2\varphi_{1,2} + 2\varphi_{2,3}$ then $(A_5A_6, l_2) = \varphi + 2\varphi_{1,2} + 2\varphi_{2,3} - \varphi_{2,3}$. Since $(A_7A_6, l_1) = -(A_5A_6, l_2) = -\varphi - 2\varphi_{1,2} - \varphi_{2,3}$ then $(A_7A_6, l_3) = -\varphi - 2\varphi_{1,2} - \varphi_{2,3} + \varphi_{1,2} + \varphi_{2,3} = -\varphi - \varphi_{1,2}$.

Solution of problem 0.5.

Use lemma 0 to the orthotriangle.

NOTE. The problem can be reformulated in the next way.

Let point A_1 lies on AB. The circumcircle of triangle A_1AC secondary meets BC in point A_2. The circumcircle of triangle A_2BA secondary meets CA in point A_3 etc. Prove that $A_1 = A_7$.

Solution of problem 0.6

Use lemma 0 to the medial triangle.

Solution of problem 0.7

a) The Pappus theorem in an equivalent statement is proved in the book [1, Chapter 1].

b) Consider the hexagon $A_1A_2A_3A_6A_5A_4$: the lines A_1A_2, A_3A_6, A_5A_4 concur in point R, and the lines A_4A_1, A_2A_3, A_6A_5 concur in point G. Therefore "the diagonals" A_2A_5, A_3A_4 (these two lines pass through B) and A_6A_1 concur (in point B). Thus $A_7 = A_1$.

Solution of problem 0.8

a) The Brianchon theorem is proved in the book [1, Chapter 1].

b) It is clear that the lines A_4A_5, A_5A_6, A_6A_7 are the reflections of A_4A_3, A_3A_2, A_2A_1 in OI. Therefore $A_1 = A_7$.

Solution of problem 0.9

First solution (D. Yakutov). Let us compute the angle $\angle GA_4R$:

$$\begin{aligned}
\angle GA_4R &= \pi - \angle GA_4B - \angle BA_4R \\
&= \pi - \angle GOB - \angle BA_5R \\
&= \pi - \angle GOB - (\pi - \angle GA_5R - \angle GA_5B) \\
&= \angle GA_5R + \angle GA_5B - \angle GOB \\
&= \angle GOR - \angle GOB + \angle GA_6B \\
&= \angle GOR - \angle GOB + (\pi - \angle GA_6R - \angle BA_6R) \\
&= \angle GOR - \angle GOB + (\pi - \angle GA_7R - \angle BOR) \\
&= (\pi - \angle BOR - \angle GOB + \angle GOR) - \angle GA_7R.
\end{aligned}$$

Thus $\angle GA_4R + \angle GA_7R = \pi - \angle BOR - \angle GOB + \angle GOR$.

Similarly $\angle GA_1R+\angle GA_4R = \pi-\angle BOR-\angle GOB+\angle GOR$. Thus $\angle GA_1R = \angle GA_7R$. Also $\angle GA_1B = \angle GOB = \angle GA_7B$. Hence the points G, B, A_1, A_7 belong to one circle and the points G, R, A_1, A_7 also belong to one circle. But these two points have at most two common point, one of which is G. Since $A_1 \neq G$ and $A_7 \neq G$ it follows that $A_1 = A_7$.

Second solution. Consider an arbitrary inversion with center O. It transforms the red, green and blue circles to three lines. Let the image of the red point be C, the image of the blue point be A, and the image of the green point be B. Now look after points A_i. Through point A_1 ($\in AB$) we take a green (passing through A and C) circle, which secondary meets in A_2 the blue "circle" — line BC. Through A_2 we take a red circle secondary meeting in A_3 the green "circle" — line AC. Through A_3 we take a blue circle secondary meeting in A_4 the red "circle" — line AB etc.

Thus we obtained the reformulating of problem 0.5 given in the note to this problem. Therefore $A_1 = A_7$.

`Solutions of problems of part 1.`

Most of the solutions of problems from sections 1–3 are based on Problem 4.4.

Solution of problem 1.1.

Use problem 4.4.

- For each $t \in \mathbb{R}$ take a homothety $H_O^{2^t}$ with center O and coefficient 2^t. It is clear that for any point A $H_O^{2^{t+s}}(A) = H_O^{2^t}(H_O^{2^s}(A))$. The sets γ_A are a rays with origin O.
- Draw through point $A(1,1)$ the lines $y = 1$ (this is γ_1) and $x = 1$ (this is γ_2). Draw through an arbitrary point $B \in \gamma_1$ a ray γ_B. Paint all such rays red. Now paint green and blue respectively the lines $H_O^{2^t}(\gamma_1)$ ans $H_O^{2^t}(\gamma_2)$, i.e the lines parallel to Ox and Oy. It is evident that the rays or the lines of each color don't intersect.
- Consider a disc with radius 1 and center (1;1). It is clear that exactly one ray or line of each color passes through each point of this disc.

Therefore by problem 4.4 the constructed rays and lines form a web. It is clear that the lines given in the problem also form a web.

Solution of problem 1.2.

First solution. Follows from the assertion of problem 0.7b.

Second solution. This follows from problem 1.1 by a projective transformation taking the line through 2 points from the 3 given ones to the infinitely distant line.

Solution of problem 1.3.

Follows from the assertion of problem 0.8b).

Solution of problem 1.4.

Follows from the assertion of problem 0.8a).

Solution of problem 1.5.

These lines don't form a web.

Solution of problem 1.6.

These lines don't form a web.

`Solutions of problems of part 2.`

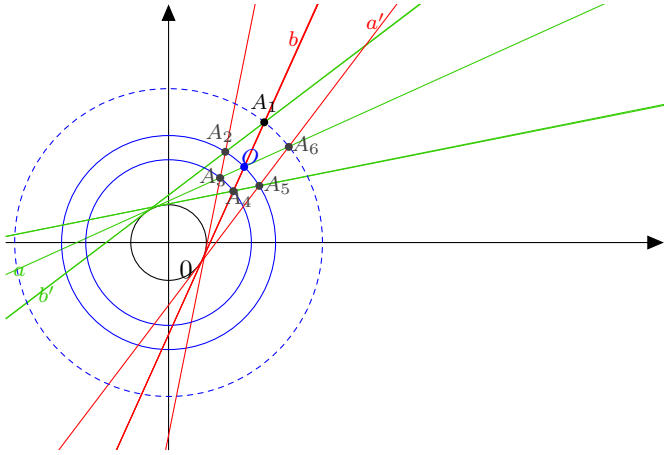
Solution of problem 2.1.

Apply for example an arbitrary inversion to the web formed by the lines parallel to the sidelines of some triangle.

Solution of problem 2.2.

First solution (E. Streltsova). Let us prove that these lines and circles form a wweb; see figure below. Take a disc in the first coordinate quarter above the line $y = 1$ so that it has no common points with the unit disk. Let the radius of the disk be < 1. Through each point of the disk there is exactly one red and

one green line because there is exactly one tangent line of each color from each point. Through each point T there is exactly one line with the center at the origin (Z) because the radius (ZT) and the center (Z) uniquely determine a circle. The circles with the center Z cannot coincide with the tangents to the unit circle. And the green and the red line cannot coincide because the disk is above the line $y = 1$. Concentric circles cannot touch each other. And the green and the red lines cannot touch the circles with the center Z because the latter have radius > 1 and our disk have no common points with the unit disk. The tangents intersect the blue circles because they contain points inside the circles. Thus the foliation condition holds.



The green (a) and the red (b) lines through the point O are symmetric with respect to the line ZO. Thus $A_3 = S_{ZO}(A_4)$. Thus the red line through A_3 is symmetric to the green line through A_4 with respect to the line ZO. Hence $A_2 = S_{ZO}(A_5)$. Hence the green line through A_2 (b') is symmetric to the red line through A_5 (a').

Further, $a = S_{ZO}(a'), b = S_{ZO}(b')$. Thus $A_6 = a \cap a' = S_{ZO}(b \cap b') = S_{ZO}(A_1)$. Thus $A_6 = S_{ZO}(A_1)$. Hence $ZA_6 = S_{ZO}(ZA_1)$. Therefore $ZA_6 = ZA_1$, hence the blue circle through the point A_6 passes through A_1. The closure condition has been checked.

Second solution. Use problem 4.4.

- For each $t \in \mathbb{R}$ take a rotation $R_O^{\pi t}$ around the origin O to the angle πt. It is clear that for any point A $R_O^{\pi(t+s)}(A) = R_O^{\pi t}(R_O^{\pi s}(A))$, if $t, s \in \mathbb{R}$. The sets γ_A are an arcs of the circles with center O.
- Draw through the point $A(0, 2)$ the rays $y = \sqrt{3}x + 2, x > -\sqrt{3}/2$ (this is γ_1) and $y = -\sqrt{3}x + 2, x < \sqrt{3}/2$ (this is γ_2). These rays touche the unit semicircle. Draw through each point $B \in \gamma_1$ an arc γ_B of the circle with the center in the origin. Paint all such arcs the red. Now paint the rays $R_O^{\pi t}(\gamma_1)$ and $R_O^{\pi t}(\gamma_2)$ green and blue respectively, these rays touche the unit semicercles.
- Consider a disc with radius $1/2$ and center $(0; 2)$. It is clear that exactly one arc or ray of each color passes through each point of this disc.

Therefore by problem 4.4 the constructed arcs and rays form a web. It is clear that the lines and the circles given in the problem also form a web.

Solution of problem 2.3.
Use problem 4.4.

- For each $t \in \mathbb{R}$ take a rotation $R_O^{\pi t}$ around the origin O to angle πt. It is clear that for any point A $R_O^{\pi(t+s)}(A) = R_O^{\pi t}(R_O^{\pi s}(A))$, if $t, s \in \mathbb{R}$. The sets γ_A are an arcs of the circles with center O.
- Draw through the point $A(0, 2)$ the rays $y = \sqrt{3}x + 2, x > -\sqrt{3}/2$ (this is γ_1) and $x = 0, y > 0$ (this is γ_2). One of these rays touches the unit semicircle and the other passes through the origin. Draw through each point $B \in \gamma_1$ an arc γ_B of the circle with center O. Paint all such arcs the red. Now paint the rays $R_O^{\pi t}(\gamma_1)$ and $R_O^{\pi t}(\gamma_2)$ green and blue respectively.
- Consider a disc with radius $1/2$ and center $A(0; 2)$. It is clear that exactly one line of each color passes through each point of this disc.

Therefore by problem 4.4 the constructed rays and arcs form a web. It is clear that the lines and the circles given in the problem also form a web.

Solution of problem 2.4.
Use problem 4.4.

- For each $t \in \mathbb{R}$ take a translation $T_{(0,t)}$ to the vector $(0,t)$. It is clear that for any point A $T_{(0,t+s)}(A) = T_{(0,t)}(T_{(0,s)}(A))$, if $t, s \in \mathbb{R}$. The sets γ_A are lines parallel to Oy.
- Draw through the point $A(1/2 + 1/\sqrt{8}, 1/2 + 1/\sqrt{8})$ the arc $(x-1/2)^2 + (y-1/2)^2 = 1/4, x > 1/2, y > 1/2$ (this is γ_1) and the line $y = 1/2 + 1/\sqrt{2}$ (this is γ_2). Through each point $B \in \gamma_1$ draw a line γ_B parallel to Oy. Paint all such lines red. Now paint the arcs $T_{(0,t)}(\gamma_1)$ and the lines $T_{(0,t)}(\gamma_2)$ green and blue respectively.
- Consider a disc with radius $1/2 - 1/\sqrt{8}$ and center $A(1/2 + 1/\sqrt{8}; 1/2 + 1/\sqrt{8})$. It is clear that exactly one arc of each color passes through each point of this disc.

Therefore by problem 4.4 the constructed arcs form a web. Then the generalized circles considered in the problem also form a web.

It is clear that the lines and the circles given in the problem also form a web.

Solution of problem 2.5.

Use problem 4.4. Consider as the maps the homotheties with center O. Then the red lines pass through the origin. The green circles touch both segments of the firs pair. The blue circles touch both segment of the second pair. By Problem 4.4 it follows that certain arcs of these circles form a web. Then the generalized circles considered in the problem also form a web.

By problem 4.4 these lines and circles form a web.

Solution of problem 2.6.

Use problem 4.4. Consider as the maps the homotheties with center O. Then the red lines pass through the origin. The green circles are the circles with center O. The blue circles touche two given segments.

By problem 4.4 these lines and circles form a web.

`Solutions of problems of part 3.`

The assertions the analogous of the problem 4.4 are true for the torus and the hyperboloid of revolution.

Solution of problem 3.1.

Use problem 4.4. Consider as the maps the rotations around the axis of the torus. Then the parallels are the red circles. The Villarceau circles are the green and blue circles.

By problem 4.4 these circles form a web.

Solution of problem 3.2. Take an arbitrary point O of the torus. Draw through it the meridean γ_1 and the Villarceau circles γ_2, γ_3. Paint the meridians red, paint the Villarceau circles obtained from the circle γ_2 by rotation green, paint the Villarceau circles obtained by rotation the circles γ_2 blue. Take a sphere with center O and radius $R = \frac{r}{100}$ (r is the distance between γ_1 and the axis of the torus). It is clear that any two the Villarceu circles have at most one common point inside the sphere. By Ω denote the intersection of the sphere and the torus.

Consider an arbitrary point O' inside $\mathbf{\Omega}$. Draw through it the red w_1, green w_2, and blue w_3 circles. Let all constructed points A_i lie inside Ω. Let $A_1 \in w_1$. Draw through A_1 the green circle w_2'. Let A_2 be the common point of w_2' and w_3. Draw through A_2 the red circle w_1'. Let A_3 be the common point of w_1' and w_2. Draw through A_3 the blue circle w_3'. Let A_4 be the common point of w_3' and w_1. Draw through A_4 the green circle w_2''. Let A_5 be the common point of w_2'' and w_3. Draw through A_5 the red circle w_1'. Let A_6 be the common point of w_1'' and w_2. Draw through A_6 the blue circle w_3'. Let A_7 be the common point of w_3'' and w_1. Let α be a plane such that w_1 lie in. It is clear that the circles w_3',w_1', w_2' are the reflections of w_2'', w_1'', w_3'' in α. Therefore $A_1 = A_7$.

Solution of problem 3.3.

Use problem 4.4. Consider as the maps the rotations around the axis of the torus. Then the parallels are the red circles. The line lying on the hyperboloid are the green and blue lines. By problem 4.4 these circles form a web.

`Hints and solutions of problems of part 4.`

Hint to problem 4.1.

Let us list a few possible examples of the sets of blue general circles:

- (B) an arbitrary pencil of lines (Problems 1.1 and 1.2);
- (B) circles with a common center;
- (B) circular arcs obtained from a given one by parallel translations along either the Ox or the Oy axis (Problem 4.3);

By the Graf–Sauer theorem (Problem 4.12) there are no other examples of sets of blue lines. By the Shelekhov classification of all webs from pencils of general circles [3, Theorem 0.1] it follows that there are no other examples in which the set of blue circles is a pencil. Description of all possible examples, not necessarily pencils, is an open problem.

Hint to problem 4.2.

Let us list a few possible examples of the sets of blue general circles:

- (B) an arbitrary pencil of lines (Problems 1.1 and 1.2);
- (B) pencil of circles with a limit point at the origin O and the common radical axis parallel to the Ox axis;
- (B) circular arcs obtained from a given one by homotheties with center at the origin (Problem 4.3).

By the Graf–Sauer theorem (Problem 4.12) there are no other examples of sets of blue lines. By the Shelekhov classification of all webs from pencils of general circles [3, Theorem 0.1] it follows that there are no other examples in which the set of blue circles is a pencil. Description of all possible examples, not necessarily pencils, is an open problem.

Hint to problem 4.3.

Perform an inversion with the center at one of the limit points. The obtained pencils of general circles form a web by Problem 4.4.

Solution of problem 4.4.

The foliation condition is true by the third condition of the problem. Let us show that the closure condition also is true.

Take an arbitrary point O inside the disc. Draw trough it the red (w_1), green (w_2) and blue (w_3) arcs of general circles. Let all constructed points A_i lie inside $\mathbf{\Omega}$. Let $A_1 \in w_1$ and $t \in \mathbb{R}$ is such that $R_t(O) = A_1$ (such t exists by the first condition of the lemma: if $w_1 = \gamma_X$, where $X \in \gamma_1$, then there exist such $y, z \in \mathbb{R}$, that $R_y(X) = O$ and $R_z(X) = A_1$, thus $R_{z-y}(O) = R_{z-y}(R_y(X)) = R_z(X) = A_1$. Therefore $t = z - y$.). Draw through A_1 the green arc w_2'. Let A_2 be the common point of w_2' and w_3. Draw through A_2 the red arc w_1'. Let A_3 be the common point of w_1' and w_2. Draw through A_3 the blue arc w_3'. Let A_4 be the common point of w_3' and w_1 etc.

Now let us show that $R_t(O) = A_7$, this yields that $A_1 = A_7$.

We know that $R_t(O) = A_1$, thus $R_t(w_2) = w_2'$ (it is true because $R_t(w_2)$ is the red arc passing through A_1, and the unique such arc is w_2'), therefore $R_t(A_3) \in w_2' \cap w_1' = A_2$. Since $R_t(A_3) = A_2$, then $R_t(A_4) = O$ (similarly). Since $R_t(A_4) = O$, then $R_t(A_5) = A_6$. Since $R_t(A_5) = A_6$, then $R_t(O) = A_7$. The assertion is proved.

Hints to problems 4.5–4.6.

These problems are discussed in [4].

Hints to problem 4.7.

The solution of this problem is given in the book by Prasolov and Solov'ev [2].

Hint. Let the equations of red lines be $a_1x + b_1y - 1 = 0$, $a_2x + b_2y - 1 = 0$, $a_3x + b_3y - 1 = 0$, and let the equation of blue ones be $c_1x + d_1y - 1 = 0$, $c_2x + d_2y - 1 = 0$, $c_3x + d_3y - 1 = 0$. Prove that the equation of the curve has the form

$$p(a_1x + b_1y - 1)(a_2x + b_2y - 1)(a_3x + b_3y - 1) + q(c_1x + d_1y - 1)(c_2x + d_2y - 1)(c_3x + d_3y - 1) = 0$$

for some real numbers p and q.

Hints to problem 4.8.
This problem is obtained from the previous one by the projective duality.
Hints to problem 4.9.
Use Problem 4.8.
Hints to problem 4.10–4.11.
Use Problem 4.9. A figure to Problem 4.10 by A. Ghaneiyan Sebdani and E. Ashourioun is shown at the first page of this document.

References

[1] Zaslavsky A.A. Akopyan A.V. *Geometry of conics*. AMS, 2007.

[2] V. Prasolov and Yu. Soloviev. *Elliptic functions and algebraic equations*. Moscow: Factorial, 1997.

[3] A. M. Shelekhov. Classification of regular three-webs formed by pencils of circles. *J. Math. Sciences*, 143(6):3607–3629, 2007.

[4] S. L. Tabachnikov. Geometry of equations. *Kvant*, 10, 1988. in Russian.

Честный раздел тортов

И.И. Богданов, К.А. Кноп, Г.Р. Челноков, И.Н. Шнурников

1 Общая постановка

Для начала приведём несколько задач, которые рассматриваются в этом проекте. В каждой задаче первый пункт — несложный. Однако сложность следующих быстро растёт, и, например, задача 1.3д) — очень сложна!

Внимание! Если у вас не получается некоторый пункт — перейдите к следующим задачам! Решения некоторых из них могут сильно помочь (например, очень полезно доказать Теорему о трети, см. задачу 3.3б).

1.1. а) Три пирожных, каждое из которых весит по 200 г, разрезали на куски. Известно, что их можно раздать четырём детям так, чтобы все получили поровну. Докажите, что какой-то из кусков весит не больше 50 г. Можно ли число 50 заменить на меньшее?

б) Четыре пирожных, каждое из которых весит по 210 г, разрезали на куски. Известно, что их можно раздать 7 детям так, чтобы все получили поровну. Докажите, что какой-то из кусков весит не больше 50 г. Можно ли число 50 заменить на меньшее?

в) Четыре торта, каждый из которых весит по 3 кг, разрезали на куски. Известно, что их можно раздать 25 детям так, чтобы все получили поровну. Докажите, что какой-то из кусков весит не больше 230 г. Можно ли число 230 заменить на меньшее?

1.2. а) Пять тортов массы 1 кг нужно разрезать на несколько частей так, чтобы можно было раздать их поровну семи людям. Найдите максимальный возможный вес минимального куска в таком разрезании.

б) Та же задача для 7 тортов и 9 человек.

1.3. а) Восемь тортов массы 1 кг нужно разрезать на несколько частей так, чтобы можно было раздать их поровну девяти людям. Найдите максимальный возможный вес минимального куска в таком разрезании.

б) Та же задача для 11 тортов и 14 человек.

в) Та же задача для 14 тортов и 17 человек.

г) Та же задача для 13 тортов и 16 человек.

д) Та же задача для 31 торта и 52 человек.

Все эти задачи является частным случаем следующей общей постановки. (Везде далее все параметры — натуральные числа.)

Мегазадача. Есть m одинаковых тортов единичного веса и n людей. Требуется разрезать торты на несколько частей и раздать их людям так, чтобы каждый получил куски одного и того же суммарного веса. При этом надо, чтобы минимальный вес куска был как можно больше. Требуется найти этот наибольший возможный вес минимального куска.

Определение. Обозначим ответ в Мегазадаче через $f(m, n)$.

Несмотря на простоту постановки, решить Мегазадачу в общем виде весьма сложно. Оказывается, что постепенными итерациями возможно найти ответ для «большинства» значений m и n, но всё время некоторые области значений оказываются неразобранными.

Видимо, ответа на Мегазадачу в замкнутой форме не существует. Основная цель данного проекта — как можно ближе подойти к построению алгоритма, решающего Мегазадачу при каждых m и n. Этот алгоритм **не сформулирован** в виде задачи, однако является главным маяком проекта.

Внимание! Если у вас появился общий алгоритм решения (или чёткие предположения о том, как он выглядит) — мы всегда готовы это обсудить. Также это можно сделать, если у вас готов такой алгоритм для достаточно большого интервала значений дроби m/n.

В дальнейшем мы всегда считаем, что переменные обозначают натуральные числа.

Сразу сформулируем три общих вопроса, один весьма лёгок, на другой авторы, как ни странно, не знают ответа; третий формально не связан с проектом, но его решение тоже может помочь.

1.4. Пусть известно значение $f(m,n)$. Найдите $f(n,m)$.

Замечание. В силу этой задачи, достаточно исследовать лишь случай $m < n$; поэтому в дальнейшем мы работаем лишь с этим случаем.

1.5*. Верно ли, что $f(tm,tn) = f(m,n)$?

Замечание. Авторы считают (и это хорошо подтверждено практикой), что ответ на последний вопрос положителен. Поэтому во многих последующих пунктах мы будем задавать не конкретные значения m и n, а их частное.

1.6. а) На какое минимальное число кусков надо разрезать m одинаковых тортов, чтобы эти куски можно было раздать поровну n людям?

б) Каким может быть размер наименьшего куска в таком разрезании на наименьшее число кусков?

В следующих разделах мы предлагаем найти ответы на Мегазадачу для некоторых последовательностей значений и, соответственно, для некоторых интервалов значений. В каждом из разделов задачи идут примерно по возрастанию сложности; мы рекомендуем решать оба раздела параллельно. С другой стороны, если вы прорешаете один из разделов (почти) до конца — мы всегда можем добавить более сложных задач!

2 Некоторые специальные последовательности значений

2.1. а) Найдите $f(3k-1,3k)$.
б) Найдите $f(3k+1,3k+2)$.
в) Найдите $f(3k,3k+1)$.

2.2. Докажите, что $f(m,2m-1) = \frac{m+1}{6m-3}$ при $m \geq 4$.

2.3. а) Найдите $f(3,n)$.
б) Найдите $f(4,n)$.
в) Найдите $f(5,n)$.

2.4. Найдите $f(m,2m+1)$.

2.5. Найдите $f(2k+1,3k+2)$.

2.6. Найдите $f(3k+1,4k+1)$.

2.7. Найдите $f(5k+2,8k+3)$.

2.8. Найдите $f(5k-1,9k-2)$.

2.9. Найдите $f(17k-4,21k-5)$.

3 Серийные результаты

Дальнейшие задачи являются лишь некоторыми вехами на пути к общему алгоритму. Если у вас появляются какие-то другие **серийные** результаты, сдавайте их тоже!

Напомним, что мы всегда полагаем $m < n$.

3.1. Пусть n не делится на m. Докажите, что $f(m,n) \leq \frac{m}{2n}$. При каких значениях m и n в этой оценке достигается равенство?

3.2. а) Пусть $\frac{3}{4} < \frac{m}{n} < 1$. Докажите, что $f(m,n) \le \frac{m}{n} - \frac{1}{2}$.

б) Пусть $\frac{1}{2} < \frac{m}{n} < 1$. Докажите, что $f(m,n) \le \frac{m}{n} - \frac{1}{3}$.

в) Пусть $\frac{2}{k+1} < \frac{m}{n} < 1$ при $k \ge 4$. Докажите, что $f(m,n) \le \frac{m}{n} - \frac{1}{k}$.

Внимание! Следующая оценка (пункт б)) очень важна!

3.3. а) Пусть $f(m,n) > \frac{m}{3n}$. Докажите, что в любом оптимальном примере для пары (m,n) каждому человеку досталось не более, чем по два куска.

б) (Теорема о трети) Докажите, что $f(m,n) \ge \frac{m}{3n}$.

в) Докажите, что при $\frac{2}{3} < \frac{m}{n} \le \frac{3}{4}$ в предыдущей оценке достигается равенство. (См. также задачу 3.11.)

3.4. а) Пусть $\frac{m}{n} < \frac{2}{3}$. Докажите, что $f(m,n) \le \frac{1}{4}$.

б) Найдите все значения (m,n) (такие, что $\frac{m}{n} < \frac{2}{3}$), для которых оценка предыдущего пункта достигается.

3.5. Пусть $\frac{2}{k+1} < \frac{m}{n} < \frac{2}{k}$. Найдите все такие пары (m,n), для которых $f(m,n) = \frac{1}{k+1}$.

3.6. а) Докажите, что $f(m,n) = \frac{m}{n} - \frac{1}{3}$ при $\frac{1}{2} < \frac{m}{n} \le \frac{5}{9}$.

б) Найдите все значения (m,n), для которых $f(m,n) = \frac{m}{n} - \frac{1}{3}$.

3.7. Пусть $\frac{2}{k+1} < \frac{m}{n} < \frac{2}{k}$. Найдите все такие пары (m,n), для которых $f(m,n) = \frac{m}{n} - \frac{1}{k}$ (при $k \ge 4$).

3.8. а) Найдите $f(m,n)$ при $\frac{7}{15} < \frac{m}{n} < \frac{1}{2}$.

б) При каких ещё значениях m/n получается такой же ответ?

3.9. а) Найдите $f(m,n)$ при $\frac{7}{12} < \frac{m}{n} < \frac{22}{37}$.

б) При каких ещё значениях m/n получается такой же ответ?

3.10. а) Найдите $f(m,n)$ при $\frac{14}{17} < \frac{m}{n} < \frac{5}{6}$.

б) При каких ещё значениях m/n получается такой же ответ?

3.11*. Найдите все пары (m,n), для которых $f(m,n) = \frac{m}{3n}$.

Честный раздел тортов

После промежуточного финиша

Добавочные задачи к предыдущим разделам

На всякий случай напомним задачу, добавленную на изначальной презентации.

1.6. а) На какое минимальное число кусков надо разрезать m одинаковых тортов, чтобы эти куски можно было раздать поровну n людям?

б) Каким может быть размер наименьшего куска в таком разрезании на наименьшее число кусков?

Следующие задачи — по сути добавка в третий раздел. Точнее, это ещё несколько оценок, аналогичных теореме о трети, но более точных.

3.12. Докажите, что $f(m,n) \geq \frac{2}{5} \cdot \frac{m}{n}$, если $\frac{5}{12} \leq \frac{m}{n} \leq \frac{1}{2}$.

3.13. а) Докажите, что $f(m,n) \geq \frac{3}{8} \cdot \frac{m}{n}$ при $\frac{m}{n} \leq \frac{1}{2}$.

б) Для каких ещё интервалов вы можете доказать это неравенство?

3.14. а) Докажите, что $f(m,n) \geq \frac{2}{5} \cdot \frac{m}{n}$, если $\frac{3}{5} \leq \frac{m}{n} \leq \frac{8}{13}$.

б) Попытайтесь доказать это неравенство для какого-нибудь интервала, смежного с $\left(\frac{3}{5}, \frac{8}{13}\right)$. Например, можно ли доказать его для $\frac{m}{n} \in \left(\frac{10}{17}, \frac{3}{5}\right)$? А для $\frac{m}{n} \in \left(\frac{8}{13}, \frac{5}{8}\right)$?

в) Для каких ещё интервалов (в других местах отрезка $[0,1]$) вы можете получить такую же оценку?

3.15. Для какого интервала внутри $\left(\frac{1}{2}, \frac{5}{8}\right)$ вы можете доказать оценку $f(m,n) \geq \frac{2}{3} \cdot \frac{m}{n} - \frac{1}{6}$?

Испытательный полигон.

Для тех, кому понадобятся дополнительные нетривиальные конкретные пары (m,n), здесь мы приводим несколько таких пар. **Внимание!** Мы можем проверить ответы и примеры для этих пар, но мы не станем проверять доказательство, если оно не содержит каких-то общих идей; поэтому эти пары не помещены в задачу.

Итак, вот эти пары (если будет надо, этот список может быть пополнен):

$$(17,29); \quad (31,70); \quad (17,47); \quad (117,133); \quad (27,61); \quad (566,643); \quad (3130,6813).$$

Удачи!

4 Вариации постановки

В этом разделе мы обобщаем исходную постановку разными способами. Решение этих задач может серьёзно помочь в решении исходной Мегазадачи.

4.1. а) Есть m тортов веса 1 и $n > m$ людей. Требуется разрезать торты и раздать их людям так, чтобы каждый получил поровну. При этом требуется, чтобы каждый человек получил не более двух кусков, а каждый торт был разрезан не более, чем на три части. При каких m, n это возможно?

б) Тот же вопрос, но каждый торт должен быть разрезан не более, чем на k частей.

в) Тот же вопрос, но каждый торт должен быть разрезан либо на $k-1$, либо на k частей.

Следующие несколько задач связаны с тем, что торты иногда бывают разными.

4.2. а) Два торта с весами 1 кг и 2 кг делятся между N людьми так, чтобы каждому досталось поровну. Чему равен наибольший возможный размер минимального куска?

б) Тот же вопрос для двух тортов весов 2 кг и 5 кг.

4.3. а) Пусть $k > 1$. Есть $3k$ тортов веса 3 каждый, $k-1$ торт веса 4 каждый, и $3k-1$ торт веса 7 каждый. Требуется разрезать каждый торт размера 3 на два куска, а каждый из остальных — на три куска, чтобы все куски можно было раздать нескольким людям по два каждому, чтобы все получили поровну. Каков максимальный возможный размер наименьшего куска?

б) Тот же вопрос для $3k$ тортов веса 3, $k+2$ тортов веса 4 и $3k+2$ тортов веса 7.

в) Тот же вопрос для $3k$ тортов веса 3, $2k-1$ тортов веса 4 и $4k-1$ тортов веса 7, где $k \geq 10$. А что будет, скажем, при $k = 7$?

4.4. а) Есть торт веса 59, торт веса 89 и два торта веса 41. Требуется разрезать первый торт на 4 куска, второй — на 6 кусков, а каждый из третьих — на 5 кусков так, чтобы их можно было раздать 10 людям поровну (по весу и по количеству кусков!). Каков максимальный возможный размер наименьшего куска?

б) Есть два торта веса 41, три торта веса 35, и 11 тортов веса 29. Требуется разрезать каждый из первых тортов на 5 кусков, каждый из вторых — на 4 куска, а каждый из третьих — на 2 куска так, чтобы их можно было раздать 22 людям поровну (по весу и по количеству кусков). Каков максимальный возможный размер наименьшего куска?

в) Найдите $f(23, 29)$.

Честный раздел тортов

Избранные решения

Если у вас появились какие-то идеи по этому проекту, не стесняйтесь написать нам:
Константин Кноп kostyaknop@gmail.com, Илья Богданов ilya.i.bogdanov@gmail.com

Решения, представленные здесь, устроены так. В разделе «Некоторые последовательности» мы находим значения функции f на некоторых последовательностях пар (m, n). Заметим, что большинство из них следуют также из более общих результатов из следующих разделов. Раздел «Серийные результаты» содержит решения (или их наброски) задач 3.13, а также 3.4–3.8; при этом как лемма используется задача 4.1. В разделе «Неравные торты» мы распространяем наши методы на случай различных тортов, что позволяет подступиться к задачам типа 3.9 (мы рекомендуем прочесть раздел о серийных результатах перед этим). Наконец, в разделе «Общий алгоритм» мы на примерах описываем идеи общего алгоритма решения Мегазадачи (в нём использованы идеи из предыдущих двух разделов).

Начнём же мы с задачи 1.6а).

1.6. а) **Ответ.** $m + n - \text{НОД}(m, n)$.

Построим пример. Рассмотрим отрезок длины m. Разделим его красными точками на m равных частей и синими точками на n равных частей (некоторые точки могут быть покрашены в оба цвета). Отрезки с красными соответствуют тортам. Разрежем торты по всем синим точкам. Докажем, что получено требуемое разрезание. Очевидно, что эти куски можно раздать людям: каждому человеку можно дать куски тортов между соседними синими точками. Имеем $m + 1$ красных точек, $n + 1$ синих точек и $\text{НОД}(m, n) + 1$ разноцветных точек. Таким образом всего $m + n - \text{НОД}(m, n) + 1$.

Осталось доказать, что количество кусков не может быть меньше, чем $m + n - \text{НОД}(m, n)$. Обозначим $d = \text{НОД}(m, n)$, $n = dn'$, $m = dm'$. Рассмотрим двудольный граф с m красными вершинами и n синими вершинами, торты и люди соответственно. Каждое ребро соответствует куску торта и соединяет человека, получившего этот кусок с тортом, от которого этот кусок отрезан. Рассмотрим какую-нибудь связную компоненту этого графа, пусть в ней r красных вершин и b синих вершин. Тогда b человек съели вместе r тортов, следовательно $b \cdot \frac{m}{n} = r$. Отсюда $\frac{r}{b} = \frac{m'}{n'}$ и следовательно $m' \mid r$. Следовательно количество связных компонент не более чем $\frac{m}{m'} = d$. С другой стороны количество рёбер на каждой компоненте не менее чем уменьшенное на 1 количество вершин. Поэтому общее количество рёбер в графе (то есть количество кусков, на которые порезаны торты) не менее $m + n - d$.

б) Решение оставляем читателю.

Некоторые последовательности

В этом разделе собраны некоторые решения задач из раздела 2. Многие из этих задач на самом деле следуют из задач раздела 3; однако мы приводим их, чтобы продемонстрировать более конкретные конструкции.

Общее замечание. Поскольку случай $n \vdots m$ тривиален, далее мы всегда предполагаем, что n не делится на m.

2.2. Из 3.2б) следует $f(m,2m-1)\le\frac{m+1}{6m-3}$. Пример дележа тортов веса $6m-3$:

$$2\times\Big(3\cdot(2m-1)\Big)+\Big(2\cdot(m+1)+(2m-4)+(2m-1)\Big)+2\times\Big(2\cdot(m+1)+(2m-2)+(2m-3)\Big)+\\ \Big((m+1+i)+(m+3+i)+(2m-4-i)+(2m-3-i)\Big)_{i=1,\dots,(m-5)}$$

2.3. б) **Ответ.** $\frac{4}{n}$, если $n\vdots 4$; $\frac{2}{n}$, если $n=4k+2$; $\frac{4}{n}-\frac{1}{[n/2]}$, если $n=2k+1$.

Если $n=4l+2$ четно, то $f(4,4l+2)=\frac{1}{2l+1}$ по 3.1. Пусть $n=2k+1$ нечетно. По 3.3а) мы можем считать, что каждый человек получил по два куска. Тогда всего есть $4k+2$ кусков и по принципу Дирихле есть торт с не более чем k кусками. Тогда найдется кусок не менее чем $\frac{1}{k}$, а дополнительный к нему кусок будет не более чем $\frac{4}{n}-\frac{1}{k}=\frac{4}{n}-\frac{2}{n-1}$. Пример для нечетных n:

$$2\times\left(\frac{n-1}{2}\cdot\frac{2}{n-1}\right)+2\times\left(\frac{n-1}{2}\cdot\left(\frac{4}{n}-\frac{2}{n-1}\right)+\frac{2}{n}\right)=\\ =(n-1)\times\left(\frac{2}{n-1}+\left(\frac{4}{n}-\frac{2}{n-1}\right)\right)+2\cdot\frac{2}{n}=n\cdot\frac{4}{n}.$$

в) **Ответ.** $\frac{5}{n}$, если $n\vdots 5$; $\frac{1}{\lceil 2n/5\rceil}$, если $n=5k+1\ge 16$ или $n=5k+3$; $\frac{5}{n}-\frac{1}{[2n/5]}$, если $n=5k+4$ или $n=5k+2\ge 12$; $f(5,11)=\frac{13}{66}$.

2.4. **Ответ.** $\frac{1}{5}$ для $m\ge 6$ и $m=2$; остальные ответы следуют из предыдущих задач: $f(1,3)=\frac{1}{3}$, $f(3,7)=\frac{5}{28}$, $f(4,9)=\frac{7}{36}$, $f(5,11)=\frac{13}{66}$.

2.5. **Ответ.** $\frac{1}{4}$ для $k\ge 1$.

Оценка следует из 3.4 а). Пример для тортов веса $12k+8$:

$$1\times\Big(4\cdot(3k+2)\Big)+2\times\Big((4k+2)+(5k+4-2i)+(3k+2+2i)\Big)_{i=1,\dots,k}=\\ =k\times\Big(2\cdot(4k+2)\Big)+2\times\Big((3k+2)+(5k+2)\Big)+2\times\Big((3k+4)+5k\Big)+\dots+2\times\Big((5k+2)+(3k+2)\Big).$$

Заметим, что задача следует из 3.4б).

2.6. **Ответ.** $\frac{2k+1}{2(4k+1)}$.

Оценка следует из 3.2 а). Пример для тортов веса $8k+2$:

$$(k+1)\times\Big(2\cdot(4k+1)\Big)+2\times\Big((2k+1)+(2k+i)+(4k+1-i)\Big)_{i=1,\dots,k}=\\ =(2k+2)\times\Big((2k+1)+(4k+1)\Big)+2\times\Big((4k+1-i)+(2k+1+i)\Big)_{i=1,\dots,(k-1)}+\Big(2\cdot(3k+1)\Big).$$

2.7. **Ответ.** $\frac{1}{4}$.

Оценка следует из 3.4а). Пример для тортов веса $32k+12$.

$$(5k+2)\times(32k+12)=\\ =k\times\Big(4\cdot(8k+3)\Big)+2\times\Big(2\cdot(10k+4)+(12k+4)\Big)+\Big((12k+5)+(8k+3+i)+(12k+4-i)\Big)_{i=1,\dots,4k}.$$

Заметим, что существование примера следует из 3.4б).

2.8. **Ответ.** $\frac{6k-1}{3(9k-2)}$.

Оценка следует из 3.2 б). Пример для тортов веса $27k-6$ и порций людей $15k-3$:

$$2k\times\Big(3\cdot(9k-2)\Big)+\Big((6k-1)+(6k-1)+(6k-2+i)+(9k-2-i)\Big)_{i=1,\dots,(3k-1)}=$$
$$=6k\times\Big((6k-1)+(9k-2)\Big)+\Big((6k-1+i)+(9k-2-i)\Big)_{i=1,\dots,3k-2}$$

2.9. **Ответ.** $\frac{18k-4}{63k-15}$.

Оценка вытекает из следующей леммы.

Лемма. Пусть $\frac{4}{5}<\frac{m}{n}<1$; тогда $f(m,n)\le 2\cdot\frac{m}{n}-\frac{4}{3}$.

Доказательство. Предположим противное. Легко проверить, что $2\cdot\frac{m}{n}-\frac{4}{3}\ge\frac{m}{3n}$ и $2\cdot\frac{m}{n}-\frac{4}{3}\ge\frac{1}{4}$, так что все торты содержат два или три куска и все люди имеют по два куска. Число двухкусочных тортов равно $3m-2n$, число трехкусочных тортов равно $2n-2m$. Возможны два варианта.

1) Допустим, что кому-то достались два куска с двухкусочных тортов. Тогда остающиеся два куска этих тортов весят в сумме $2-\frac{m}{n}$, а один из этих кусков не меньше чем $1-\frac{m}{2n}$. Поэтому дополнительный до порции кусок не больше чем

$$\frac{m}{n}-\left(1-\frac{m}{2n}\right)<2\cdot\frac{m}{n}-\frac{4}{3}.$$

2) Любой кусок из двухкусочного торта дополняется куском из трехкусочного торта. Пусть x — вес минимального куска. Тогда $\frac{m}{n}-x$ — это максимальный вес. Возьмем некоторый кусок A из двухкусочного торта, тогда $A\ge 1-\frac{m}{n}+x$. Поэтому дополнительный кусок можно оценить как $\frac{m}{n}-A\le 2\frac{m}{n}-1-x$. Из $\frac{m}{n}>\frac{4}{5}$ следует, что двухкусочных тортов больше, чем трехкусочных. Поэтому найдется трехкусочный торт, все куски которого дополняют до порций куски из двухкусочных тортов. Поэтому каждый из трех кусков тортов не больше чем $2\frac{m}{n}-1-x$ и

$$3\left(2\cdot\frac{m}{n}-1-x\right)\le 1\Leftrightarrow x\le 2\frac{m}{n}-\frac{4}{3}.$$

□

Пример. Торт весит $63k-15$, порция человека $51k-12$, минимальный вес куска $18k-4$.

$$(17k-4)\times(63k-15)=2k\times\Big(3\cdot(21k-5)\Big)+6k\times\Big((33k-8)+(30k-7)\Big)+$$
$$+\Big((33k-10-i)+(30k-7+i)\Big)_{i=1,\dots,3k-2}+2\times\Big((18k-3+i)+(27k-6-i)+(18k-4)\Big)_{i=1,\dots,3k-1}=$$
$$=6k\times\Big((21k-5)+(30k-7)\Big)+6k\times\Big((33k-8)+(18k-4)\Big)+$$
$$2\times\Big((18k-4+i)+(33k-8-i)\Big)_{i=1,\dots,3k-2}+\Big((24k-6+i)+(27k-6-i)\Big)_{i=1,\dots,3k-1}=$$
$$=(21k-5)\times(51k-12).$$

Серийные результаты

Для начала приведём оценку, аналогичную теореме о трети.

3.13. а) Если $n=2m$, то $f(m,2m)=\frac{1}{2}$; поэтому достаточно рассмотреть случай $\frac{m}{n}<\frac{1}{2}$. Пусть вес каждого торта — $8n$, тогда каждый человек получит по $8m$.

Рассмотрим отрезок длины $8mn$ и разделим его красными точками на m одинаковых отрезков (тортов). Мы будем по очереди отрезать куски с левого края оставшегося отрезка. Отрежем сначала несколько отрезков по $4m$, пока остаток торта не окажется (нестрого) между $6m$ и $10m$.

Далее, разделим оставшуюся часть на два равных отрезка (получим два куска размера от $3m$ до $5m$), а от начала следующего торта отрежем два отрезка, дополняющих только что отрезанные до $8m$. В результате, от следующего отрезка мы забрали не более $10m$, и остаток не меньше $8n - 10m \geq 6m$. Тогда мы можем повторять эту операцию, пока не придём к последнему торту. Поскольку все отрезанные куски группируются в пары общей длины $8m$, а длина всего отрезка равна $8mn$, последний торт закончится двумя кусками размера $4m$. Теперь распределение кусков между людьми очевидно.

Далее мы находим точные значения функции на некоторых интервалах. Для этого в качестве лемы нам понадобится задача 4.1.

4.1. а), б) **Ответ.** $\frac{m}{n} \in \left[\frac{1}{k-1}, 1\right) \cup \left\{\frac{v}{(k-1)v+1}\right\}_{v=1,2,\ldots}$.

Начнём с примера подтверждающего ответ. Будем действовать также как и в теореме о трети. Рассмотрим отрезок длины m. Разделим его красными точками на m равных частей и синими точками на n равных частей (некоторые точки могут быть покрашены в оба цвета). Отрезки с красными соответствуют тортам. Разрежем торты по всем синим точкам. Докажем, что получено требуемое разрезание. Очевидно, что эти куски можно раздать людям: каждому человеку можно дать куски тортов между соседними синими точками. Более того, каждый человек получил не более чем по два куска, так как каждый получившийся отрезок содержит не более одной красной точки. Докажем, что каждый торт разрезан не более чем на k частей.

Если $\frac{m}{n} > \frac{1}{k-1}$, то каждый торт разрезан не более чем на $k-2$ полных частей и не более чем на две части, меньшие $\frac{m}{n}$ — всего не более чем на k частей. Для $\frac{m}{n} = \frac{1}{k-1}$ утверждение очевидно. Теперь предположим, что $\frac{m}{n} = \frac{v}{(k-1)v+1}$. Рассмотрим k последовательных синих точек. Расстояние между первой и последней равно $\frac{(k-1)v}{(k-1)v+1}$ и их координаты — это дроби со знаменателем $(k-1)v+1$. Следовательно торт содержит все эти k точек только в том случае, если одна из них является его концом. Это означает в точности, что торт разрезан не более чем на k частей.

Докажем, что в других случаях требуемое разрезание невозможно. Будем называть человека *голодным*, если он получил 2 куска. Рассмотрим граф, в котором вершинам соответствуют торты, а рёбрам — голодные люди, причём каждое ребро соединяет два куска, которые получил соответствующий голодный человек. Рассмотрим произвольную связную компоненту этого графа; пусть v и e — количества вершин и рёбер соответственно. Тогда $e \geq v - 1$.

Эти v тортов содержат не более kv кусков, $2e$ из которых принадлежат голодным людям. Эти куски могут быть перераспределены в целые части (а именно, e частей из двух кусков каждая и, скажем, t частей из одного куска). Тогда $t \leq kv - 2e$. Сравнивая общий вес v тортов и $e+t$ людей получаем, что $\frac{n}{m} = \frac{e+t}{v} \leq \frac{kv-e}{v} = k - \frac{e}{v}$. Если $e \geq v$, то $\frac{n}{m} \leq k-1$. Иначе $e = v-1$ и следовательно $\frac{n}{m} = \frac{v(k-1)+1}{v}$.

Важное замечание. Если убрать условие $m < n$, ответ изменится не сильно. Во-первых необходимо включить в него 1; теперь предположим, что $m > n$. В этом случае все люди голодные, потому что часть торта, которая достаётся человеку больше самого торта. Каждая связная компонента графа соответствует делению v тортов между e людьми, следовательно $1 < \frac{m}{n} = \frac{v}{e}$. Последнее неравенство выполняется только в случае $e = v-1$. Для таких значений v и e пример строится аналогичным образом, в результате получаем такой ответ

$$\frac{m}{n} \in \left[\frac{1}{k-1}, 1\right] \cup \left\{\frac{v}{(k-1)v+1}\right\}_{v=1,2,\ldots} \cup \left\{\frac{e+1}{e}\right\}_{e=1,2,\ldots}.$$

в) **Ответ.** $\frac{m}{n} \in \left[\frac{1}{k-1}, \frac{2}{k-1}\right] \cup \left\{\frac{v}{(k-1)v+1}\right\}_{v=1,2,\ldots}$.

Решение оставляем читателю.

3.4. а) Как обычно, мы можем считать, что каждый человек получает не менее двух кусков. Тогда общее количество кусков не менее $2n > 3m$. Следовательно какой-то торт разрезается по крайней мере на четыре куска, один из которых не может быть больше $\frac{1}{4}$.

б) **Ответ.** $\frac{m}{n} \in \left[\frac{5}{8}, \frac{2}{3}\right) \cup \left\{\frac{5k+2}{8k+4}\right\}_{k=1,2,\ldots}$.

Предположим, что $f(m,n) = \frac{1}{4}$; тогда $f(m,n) > \frac{m}{3n}$, поэтому мы можем считать, что каждый человек получает ровно 2 куска. Далее, каждый кусок не менее $\frac{1}{4}$ и не более

$$d = \frac{m}{n} - \frac{1}{4}$$

(иначе другой кусок, принадлежащий человеку с куском веса d будет меньше, чем $\frac{1}{4}$). Следовательно каждый торт содержит не менее трёх кусков (иначе найдётся кусок с весом $\frac{1}{2} > d$) и не более 4 кусков (иначе найдётся кусок, меньший $\frac{1}{5}$). Итак у нас есть *богатые* торты, которые разрезаются на 4 части и *обычные* торты, которые разрезаются на 3 части. Количество богатых и обычных тортов равно

$$f = 2n - 3m \qquad \text{and} \qquad u = 4m - 2n$$

соответственно. Так как $u \geq 0$, получаем, что $\frac{m}{n} > \frac{1}{2}$.

Каждый богатый торт должен быть разрезан на равные части, которые отдаются $4f$ людям и второй кусок у каждого такого человека весит d. Все оставшиеся люди получают оба своих куска от обычных тортов; будем называть таких людей *обычными*. Получается

$$s = n - 4f = 12m - 7n$$

обычных людей.

Теперь рассмотрим вспомогательное деление **отрицательных** «тортов»; оно соответствует обычным тортам и людям. Давайте вычтем $\frac{1}{4}$ из каждого куска богатого торта и d из каждого куска обычного торта. Забудем на некоторое время про нулевые куски. Тогда в новом разделении все необычные люди и все не богатые торты исчезнут. Каждый обычный «торт» теперь разделен не более чем на три **отрицательных** «куска» одинакового общего веса, а каждый обычный человек имеет не более чем два неположительных «куска» одинакового общего веса. Заметим, что если взять все полученные веса с обратным знаком, то мы попадём в ситуацию задачи 4.1а) (без условия $m < n$).

Обратно, от деления этих *дефицитов* легко получить разделение исходных тортов. Будем резать настоящие торты на три части со следующими дефицитами (если новый «торт» был разделен на одну или две части, то оставшиеся куски не должны иметь дефицит, то есть они должны весить по d). Про людей: если кто-то имеет два дефицита, то мы даём ему два соответствующих куска; иначе он получает один кусок, соответствующий дефициту и еще один кусок с нулевым дефицитом. В результате все оставшиеся куски с нулевым дефицитом разбиваются на пары с кусками богатых тортов.

Итак, деление существует тогда и только тогда, когда $\frac{u}{s} = \frac{4m-2n}{12m-7n} \in \left[\frac{1}{2}, 1\right) \cup \left\{\frac{k}{2k+1}\right\} \cup \left\{\frac{k+1}{k}\right\} \cup \{\infty\}$ (мы используем замечание после 4.1б)); также мы должны добавить вырожденный случай $s = 0$). С учётом того, что $\frac{m}{n} < \frac{2}{3}$, получаем, что $\frac{m}{n} \in \left[\frac{5}{8}, \frac{2}{3}\right) \cup \left\{\frac{5k+2}{8k+4}\right\}_{k=1,2,\ldots}$.

3.5. **Ответ.** $\frac{m}{n} \in \left[\frac{2k-1}{k^2-1}, \frac{2}{k}\right) \cup \left\{\frac{d(2k-1)+2}{d(k^2-1)+k+1}\right\}_{d=1,2,\ldots}$.

Решение аналогично 3.4б).

3.6. а) По 3.2б), $f(m,n) \leq \frac{m}{n} - \frac{1}{3}$. Обратное неравенство доказано в пункте б).

б) **Ответ.** $\frac{m}{n} \in \left(\frac{1}{2}, \frac{5}{9}\right] \cup \left\{\frac{5k+2}{9k+3}\right\}_{k=0,1,2,\ldots}$.

Вначале заметим, что $f(m,n) \geq \frac{m}{3n} > \frac{m}{n} - \frac{1}{3}$ при $\frac{m}{n} < \frac{1}{2}$. С другой стороны, $f(m,n) \leq \frac{m}{2n} < \frac{m}{n} - \frac{1}{3}$ при $\frac{m}{n} > \frac{2}{3}$. Осталось разобраться с интервалом $\left(\frac{1}{2}, \frac{2}{3}\right)$ (левый конец интервала не удовлетворяет условию, а правый удовлетворяет).

Приведём тут набросок доказательства, оно похоже на доказательство задачи 3.4б). Докажем, что каждый человек получит 2 куска и веса кусков принадлежат отрезку $\left[\frac{1}{3}, d\right]$ где $d = \frac{m}{n} - \frac{1}{3}$. Имеем $f = 2n - 3m$ *богатых* тортов, каждый из которых разрезается на 4 части и $u = 4m - 2n$ *обычных* тортов, каждый из которых разрезается на 3 части. Каждый обычный торт должен разрезаться на одинаковые куски, которые достаются $3u$ людям, причём второй кусок у каждого такого человека весит d, а все остальные $s = n - 3u = 7n - 12m$ *обычные* люди получают оба куска из богатых тортов. Заметим, что $\frac{m}{n} > \frac{1}{2}$ влечёт $\frac{f}{s} \geq \frac{1}{2}$.

Теперь вычтем $\frac{1}{3}$ из каждого куска обычного торта и d из каждого из остальных кусков мы получим разбиение оставшихся f **неотрицательных** «тортов» между s обычными людьми. Единственное оставшееся условие, что каждый человек должен получить не более двух кусков, а каждый «торт» должен быть разрезан не более чем на 4 части. Это условие соблюдается и мы можем восстановить разбиение обычных тортов. Следовательно, по 4.1б) (вместе с замечанием после него), для $\frac{f}{s} > \frac{1}{2}$ требуемое разбиение существует тогда и только тогда, когда $\frac{f}{s} \in \left[\frac{1}{2}, 1\right] \cup \left\{\frac{v+1}{v}\right\} \cup \{\infty\}$ из чего следует ответ.

3.7. **Ответ.** $\frac{m}{n} \in \left(\frac{2}{k+1}, \frac{2k-1}{k^2}\right] \cup \left\{\frac{d(2k-1)+2}{dk^2+k}\right\}_{d=1,2,\ldots}$.

Решение аналогично 3.6б).

3.8. Частный случай задачи 3.5.

Идеи решения задачи 3.9 представлены в следующем разделе.

Неравные торты

Напомним, что через $\lfloor x \rfloor$ и $\lceil x \rceil$ обозначаются наибольшее целое число, не превосходящее x, и наименьшее целое число, не меньшее x, соответственно.

4.2. а) **Ответ.** $\frac{3}{N}$, если $3 \mid N$; $\max\left\{\frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}, \frac{1}{\lceil 2N/3 \rceil}\right\}$ в противном случае.

Если $3 \mid N$, тогда очевидно, что оптимальным будет разрезание тортов на куски по $\frac{3}{N}$ каждый.

Теперь предположим, что $3 \nmid N$. Тогда кто-то должен получить не менее двух кусков, поэтому ответ не превышает $\frac{3}{2N}$, и мы можем считать, что каждый человек получает по крайней мере два куска.

Далее, меньший торт разделен или на $\leq \lfloor 2N/3 \rfloor$ частей, или на $\geq \lceil 2N/3 \rceil$ частей (для $N = 2$ обязательно выполняется второй случай). В первом случае, один из этих кусков должен быть $\geq \frac{1}{\lfloor 2N/3 \rfloor}$, поэтому его дополнение (для человека) $\leq \frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}$. Во втором случае, меньший торт содержит кусок, который $\leq \frac{1}{\lceil 2N/3 \rceil}$. Итак, в любом случае минимальный вес не превышает хотя бы одного из чисел $\frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}$ и $\frac{1}{\lceil 2N/3 \rceil}$, т.е. не превышает $D = \max\left\{\frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}, \frac{1}{\lceil 2N/3 \rceil}\right\}$.

Осталось привести пример с минимальным весом D. Предположим сначала, что $D = \frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}$. Разделим меньший торт на куски веса $\frac{1}{\lfloor 2N/3 \rfloor} \geq D$, вырежем то же самое число кусков веса D каждый из большего торта и разделим все остальное на целые порции. Очевидно, такое разделение подходит. Во-втором случае, пример строится аналогично.

Замечание. Можно проверить, что $D = \frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}$ если $N = 3k + 2$, и $D = \frac{1}{\lceil 2N/3 \rceil}$ в противном случае.

б) **Ответ.** $\frac{7}{N}$, если $7 \mid N$; $\max\left\{\frac{7}{N} - \frac{2}{\lfloor 4N/7 \rfloor}, \frac{2}{\lceil 4N/7 \rceil}\right\}$ в противном случае.

Решение аналогично и предоставляется читателю.

Следующая задача подсказывает, как работает общий алгоритм. Нам нужно ввести некоторые

Определения и обозначения. Напомним, что *гиперграф* — это пара (V, E), где V — множество *вершин*, а E — множество *(гипер)ребер*, т.е. некоторых непустых подмножеств в

V. Гиперграф *однороден*, если все его ребра имеют одинаковое число элементов. Для любого гиперграфа $G = (V, E)$ мы можем построить его *подлежащий граф* $G' = (V, E')$ с тем же самым множеством вершин, в котором соединены ребром каждые две вершины, принадлежащие одному гиперребру G. Гиперграф *связен*, если его подлежащий граф связен.

В дальнейшем мы будем обозначать через $[b : c]$ следующую ситуацию: у нас есть торт веса b, который нужно разделить на c частей. Так, запись $2 \times [4 : 3] + 3 \times [7 : 4]$ будет обозначать набор из двух тортов веса 4, каждый из которых следует разделить на три части, и трех тортов веса 7, каждый из которых следует разделить на четыре части.

4.4. а) **Ответ.** $\frac{49}{6}$.

В наших обозначениях, имеем $[59 : 4] + [89 : 6] + 2 \times [41 : 5]$. Один из кусков в торте $[89 : 6]$ не меньше $\frac{89}{6}$; его дополнение не больше $\frac{49}{6}$. Осталось предъявить пример:

$$\left(4 \cdot \frac{59}{4}\right) + \left(6 \cdot \frac{89}{6}\right) + 2 \times \left(3 \cdot \frac{49}{6} + 2 \cdot \frac{33}{4}\right) = 4 \times \left(\frac{59}{4} + \frac{33}{4}\right) + 6 \times \left(\frac{89}{6} + \frac{49}{6}\right).$$

б) **Ответ.** $\frac{49}{6}$. В наших обозначениях, имеем $2 \times [41 : 5] + 3 \times [35 : 4] + 11 \times [29 : 2]$. Мы будем говорить, что торты веса 29 *маленькие*, а остальные *большие*. Заметим, что каждый человек должен получить два куска с общим весом 23. Предположим, что каждый кусок весит не менее чем $\frac{49}{6}$; тогда каждый кусок должен быть не более чем $23 - \frac{49}{6} = \frac{89}{6}$.

Если человек получает два куска из маленьких тортов (конечно, эти два торта различны), то средний вес оставшихся двух кусков в этих тортах $\frac{2 \cdot 29 - 23}{2} > \frac{89}{6}$, что невозможно. Поэтому все 22 куска маленьких тортов попадают к разным людям, и, следовательно, все куски из больших тортов также попадают к разным людям.

Теперь давайте назовем торты веса 41 *богатыми*, а торты веса 35 *обычными*. Построим гиперграф с маленькими тортами в качестве вершин; каждое ребро будет соответствовать обычному торту и состоять из всех маленьких тортов, содержащих дополнения (для людей) к кускам из этого обычного торта. Этот гиперграф содержит по крайней мере две связные компоненты.

Теперь давайте удалим все куски обычных тортов и их дополнения в маленьких тортах. Далее, мы склеим оставшиеся куски каждой связной компоненты в один новый «торт». Посчитаем число кусков и вес этого «торта».

Предположим, что компонента содержит v вершин и e ребер. Из-за каждого ребра мы удалили 4 куска общего веса $4 \cdot 23 - 35 = 57$; поэтому число кусков в нашей компоненте уменьшилось на $4e$, а их общий вес — $57e$. Таким образом, средний вес оставшихся кусков равен $\frac{29v - 57e}{2v - 4e}$, что должно быть $\leq \frac{89}{6}$, что переписывается как $2v \geq 7e$. С другой стороны, так как компонента связная, мы имеем $v \leq 3e + 1$. Два полученных неравенства выполняются только если пара (v, e) — или $(4, 1)$, или $(7, 2)$. Следовательно, наш гиперграф должен содержать одну компоненту типа $(4, 1)$ и одну типа $(7, 2)$. В последней компоненте, один из кусков будет не меньше, чем $\frac{7 \cdot 29 - 2 \cdot 57}{14 - 8} = \frac{89}{6}$, что дает желаемую оценку.

Но из этой конструкции можно получить и пример! Именно, из компоненты типа $(4, 1)$ мы получили «торт» из 4 кусков с общим весом $4 \cdot 29 - 57 = 59$, а из оставшейся компоненты мы получаем «торт» из 6 кусков с общим весом $7 \cdot 29 - 2 \cdot 57 = 89$. Также у нас осталось $2 \times [41 : 5]$. Таким образом мы пришли к ситуации 4.4а) и поэтому можем взять разбиение из того примера и найти веса удаленных кусков. Получается пример

$$\begin{aligned} 11 \times 29 + 2 \times 41 + 3 \times 35 = {} & 4 \times \left(\frac{59}{4} + \frac{57}{4}\right) + 6 \times \left(\frac{89}{6} + \frac{85}{6}\right) + \left(2 \cdot \frac{29}{2}\right) + \\ & + 2 \times \left(3 \cdot \frac{49}{6} + 2 \cdot \frac{33}{4}\right) + \left(4 \cdot \frac{35}{4}\right) + 2 \times \left(3 \cdot \frac{53}{6} + \frac{17}{2}\right) = \\ = {} & 4 \times \left(\frac{59}{4} + \frac{33}{4}\right) + 6 \times \left(\frac{89}{6} + \frac{49}{6}\right) + 4 \times \left(\frac{57}{4} + \frac{35}{4}\right) + 6 \times \left(\frac{85}{6} + \frac{53}{6}\right) + 2 \times \left(\frac{29}{2} + \frac{17}{2}\right). \end{aligned}$$

в) **Ответ.** $\frac{49}{138} = \frac{1}{23} \cdot \frac{49}{6}$ (могли бы догадаться?).

Умножим все веса на 29. Как обычно, мы будем считать, что каждый человек получает ровно два куска, и что каждый торт разделен или на две, или на три части. Тогда количества тортов обоих типов можно найти, и мы приходим к ситуации $12 \times [29 : 3] + 11 \times [29 : 2]$. Будем называть торты с тремя кусками *богатыми*, а остальные *обычными*. Предположим, что каждый кусок весит не менее чем $\frac{49}{6}$; тогда каждый кусок должен быть также не более чем $23 - \frac{49}{6} = \frac{89}{6}$.

По тем же причинам, что и выше, никто из людей не получит двух кусков из обычного торта. Поэтому все 22 куска обычных тортов достанутся различным людям, а их дополнения лежат в богатых тортах. Оставшиеся 14 кусков богатых тортов достанутся 7 оставшимся людям; будем называть этих людей *богатыми*. Теперь мы построим граф с богатыми тортами в качестве вершин; каждое ребро соответствует богатому человеку и соединяет два богатых торта, содержащих куски, принадлежащие этому человеку. Этот граф содержит по крайней мере 5 компонент связности.

Теперь удалим все куски богатых людей. Далее, мы склеим оставшиеся куски каждой связной компоненты в один новый «торт». Давайте найдем число кусков и вес этого «торта».

Предположим, что компонента содержит v вершин и e ребер (тогда $v \le e + 1$). Из-за каждого ребра мы удалили 2 куска с общим весом 23; поэтому число кусков, удаленных из нашей компоненты, равно $2e$, а их общий вес — $23e$. Таким образом, средний вес оставшихся кусков равен $\frac{29v-23e}{3v-2e}$, что должно быть $\ge \frac{49}{6}$; это неравенство переписывается как $27v \ge 40e$. Это невозможно, если $v \le e$, поэтому мы имеем $v = e + 1$ и, значит, $27 \le 13e$ или $e \le 2$. Таким образом, каждая связная компонента — дерево (следовательно, их ровно пять) и имеет максимум два ребра.

В самом «регулярном» случае есть две компоненты с двумя ребрами и три компоненты с одним ребром; тогда полученные новые «торты» будут иметь вид $2 \times [41 : 5] + 3 \times [35 : 4]$. По 4.4б), ответ будет не более чем $\frac{49}{6}$.

В любом другом случае появляется изолированная вершина, то есть все три куска этого торта спарены (в порции) с кусками из обычных тортов. Рассмотрим эти три дополнения и возьмем три обычных торта, содержащих их. Среднее трех оставшихся кусков этих тортов есть $\frac{4\cdot29-3\cdot23}{3} > \frac{89}{6}$, что невозможно. Оценка доказана.

Пример может быть снова получен из примера для 4.4б) добавлением удаленных кусков:

$$\begin{aligned}
11 \times 29 + 12 \times 29 = 4 \times \left(\frac{59}{4} + \frac{57}{4}\right) + 6 \times \left(\frac{89}{6} + \frac{85}{6}\right) + \left(2 \cdot \frac{29}{2}\right) + \\
+ 4 \times \left(\frac{49}{6} + \frac{33}{4} + \frac{151}{12}\right) + 2 \times \left(\frac{49}{6} + 2 \cdot \frac{125}{12}\right) + \\
+ 2 \times \left(2 \cdot \frac{35}{4} + \frac{23}{2}\right) + 2 \times \left(2 \cdot \frac{53}{6} + \frac{34}{3}\right) + 2 \times \left(\frac{53}{6} + \frac{17}{2} + \frac{35}{3}\right) = \\
= 4 \times \left(\frac{59}{4} + \frac{33}{4}\right) + 6 \times \left(\frac{89}{6} + \frac{49}{6}\right) + 4 \times \left(\frac{57}{4} + \frac{35}{4}\right) + 6 \times \left(\frac{85}{6} + \frac{53}{6}\right) + \\
+ 2 \times \left(\frac{29}{2} + \frac{17}{2}\right) + 4 \times \left(\frac{151}{12} + \frac{125}{12}\right) + \left(2 \cdot \frac{23}{2}\right) + 2 \times \left(\frac{34}{3} + \frac{35}{3}\right).
\end{aligned}$$

3.9. а) **Ответ.** $\frac{5}{4} \cdot \frac{m}{n} - \frac{1}{2}$.

Докажем сначала верхнюю оценку. Как обычно, мы считаем, что у каждого человека по два куска, каждый торт разбит на 3 или 4 части, и есть $u = 4m-2n$ *обычных* тортов по 3 части в каждом и $f = 2n-3m$ *богатых* тортов из 4 частей. Поскольку $4f < n$ (это следует из $\frac{m}{n} > \frac{7}{12}$), кто-то должен получить оба куска из обычных тортов. Рассмотрим два торта, содержащие его куски; средний вес остальных кусков этих тортов равен $d = \frac{1}{4}\left(2 - \frac{m}{n}\right)$, поэтому один из этих кусков весит не меньше d. А тогда его дополнение не тяжелее $\frac{m}{n} - d = \frac{5}{4} \cdot \frac{m}{n} - \frac{1}{2}$, что и требовалось.

Пример будет следовать из части б).

б) Мы будем исследовать лишь случай $\frac{m}{n} \in \left(\frac{7}{12}, \frac{2}{3}\right)$, когда проходит оценка из пункта а). В этом случае мы утверждаем, что $f(m,n) = \frac{5}{4} \cdot \frac{m}{n} - \frac{1}{2}$ тогда и только тогда, когда $\frac{m}{n} \in \left(\frac{7}{12}, \frac{22}{37}\right] \cup \left\{\frac{22d-3}{37d-2}\right\}_{d=1,2,\ldots}$; ниже приведён набросок решения.

Умножим все веса на $4n$; тогда вес минимального куска будет $d = 5m - 2n$, а вес максимального — $t = 2n - m$.

В нашем случае, каждый человек получает по два куска, есть $f = 2n - 3m$ *богатых* тортов по 4 куска и $u = 4m - 2n$ *обычных* тортов по три куска. Далее, легко видеть, что два куска из богатых тортов не могут попасть одному человеку, поэтому у $4f$ человек по одному куску из богатых тортов, а у остальных $s = n - 4f$ *обычных* людей по два куска из обычных тортов.

Построим граф G с обычными тортами в качестве вершин, где каждое ребро будет соединять два торта, из которых взяты куски одного обычного человека. Теперь удалим всех обычных людей со всеми их кусками, и объединим каждую компоненту связности в новый торт. Если в какой-то компоненте было более одного ребра, то, удаляя обычных людей, мы получаем куски со средним весом $> d$, что невозможно. Значит, мы должны получить s новых тортов веса $8n - 4m = 4t$, состоящие из 4 кусков каждый, и $u - 2s$ старых тортов веса $4n$, состоящих из трёх кусков каждый. Заметим, что теперь новые торты должны быть разделены на равные части.

Теперь можно действовать, как в 3.4б). Вычтем d из каждого оставшегося куска обычного торта, and и t — из каждого куска богатого торта. Тогда новые торты исчезают, все сохранившиеся обычные торты становятся положительными «тортами» из трёх или менее частей, а все богатые торты становятся отрицательными «тортами» из 4 или менее частей. Беря все куски последних по модулю, мы приходим к такой постановке:

Пусть есть $u-2s$ равных тортов, и нам надо разрезать каждый на не более чем 3 части и перераспределить полученные куски в f групп одинакового суммарного веса, причём в каждой из групп должно быть не больше 4 кусков

Более того, можно понять, что из разрезания новых «тортов» восстанавливается разрезание (и распределение) старых. Значит, остаётся выяснить, при каких параметрах новая постановка имеет решение. Это можно сделать аналогично 4.1.

Общий алгоритм

Наконец, мы продемонстрируем работу общего алгоритма на одном нетривиальном примере — а именно, мы найдём значение $f(31, 52)$.

1.3. д) Мы не будем, как в некоторых предыдущих задачах, сразу предъявлять ответ, а покажем, как можно его *найти*, с самого начала.

Часть I. Сперва мы проделаем некий странный процесс; сам по себе он не даёт ни примера, ни оценки. Однако он выдаёт точный ответ, про который затем просто доказывается, что он достижим и оптимален.

В процессе мы будем делать некоторые предположения о том, как должен выглядеть оптимальный пример. Так что после того, как пример будет построен, нам останется убедиться в том, что без выполнения одного из предположений получается худший ответ. Мы отмечаем эти предположения богатым шрифтом и нумеруем их.

Подготовка. Умножим все веса на 52. **Предположим(1)**, что каждый человек получил по два куска. Поскольку $\frac{1}{2} < \frac{m}{n} < \frac{2}{3}$, мы также **предположим (2)**, что каждый торт разделён на 3 или 4 части. Тогда мы получаем

$$11 \times [52 : 4] + 20 \times [52 : 3].$$

Назовём торты с 4 кусками *богатыми*, а остальные — *обычными*.

Начальный шаг. В богатых кусках всего 44 вершины, что меньше количества людей. Значит, мы **предположим (3)**, что все эти 44 куска распределены по 44 разным людям; значит, найдутся ровно 8 человек, у которых все куски лежат в обычных кусках. Рассмотрим графс обычными тортами в качестве вершин, рёбра которого соответствуют обычным людям. Он содержит 20 вершин м 8 рёбер, так что в нём хотя бы 12 компонент связности.

Теперь мы **предположим(4)** что (i) все компоненты — это деревья (и поэтому их ровно 12), и (ii) все рёбра распределены между компонентами почти равномерно (то есть разность числа пёбер в двух компонентах не превосходит 1). В нашем случае это означает, что появилось 5 компонент связности из 2 вершин, а также 4 изолированных вершины. Тогда, удаляя все куски, принадлежащие обычным людям, и склеивая каждую компоненту в один торт, мы приходим к ситуации

$$11 \times [52 : 4] + 8 \times [73 : 4] + 4 \times [52 : 3].$$

Регулярный шаг 1. Теперь у нас есть 44 *маленьких* куска в 11 богатых тортах и 44 *больших* куска в остальных тортах; любой человек должен получить по маленькому и большому куску. Заметим, что средний вес куска в торте $[52 : 3]$ меньше, чем в $[73 : 4]$. Неформально говоря, это значит, что нам надо разрезать торт $[73 : 4]$ настолько поровну, насколько это возможно. Поэтому мы отложим пока эти торты и займёмся остальными.

Рассмотрим гиперграф с богатыми тортами в качестве вершин и рёбрами, построены по тортам $[52 : 3]$ (каждое ребро состоит из богатых тортов, содержащих дополнения кусков одного торта $[52 : 3]$). Таким образом, в нашем гиперграфе 11 вершин и 4 ребра мощности ≤ 3. Тогда в нём не менее $11 - 4 \cdot (3 - 1) = 3$ компонент связности.

Как и раньше, мы **предполагаем(5)**, что в этом гиперграфе (i) в любой компоненте — наибольшее количество вершин, возможное при данном количестве рёбер (и поэтому компонент ровно три), и (ii) рёбра распределены по компонентам почти равномерно (т.е. в любых компонентах их количество различается максимум на 1). В нашем случае это означает, что получились две компоненты с одним ребром и тремя вершинами, а также одна с двумя рёбрами и пятью вершинами. Теперь, удаляя все куски тортов $[52 : 4]$ вместе с их дополнениями и склеивая каждую компоненту в один торт, мы приходим к

$$2 \times [105 : 9] + [178 : 14] + 8 \times [73 : 4].$$

Регулярный шаг 2. Остались 32 *больших* куска в тортах $8 \times [73 : 4]$ и 32 *маленьких* куска в остальных тортах; каждый должен получить по одному куску каждого типа. Заметим, что средний вес куска в $[105 : 9]$ больше, чем в $[178 : 14]$. Опять же, это означает, что надо резать $[178 : 14]$ как можно точнее, посему мы работаем с остальными.

Рассмотрим гиперграф на тортах $[73 : 4]$ как на вершинах, с рёбрами, соответствующими тортам $[105 : 9]$. В отличие от предыдущих случаев, этот гиперграф может оказаться связным. Тогда мы **предполагаем(6)**, что он связен и мы готовы завершать наш процесс. Действительно, делая граф связным, после стандартного выкидывания и склеивания получаем ситуацию

$$[178 : 14] + [256 : 14],$$

которая тривиальна. Действительно, минимальный кусок не превосходит $\frac{178}{14} = \frac{89}{7}$, и добиться такого значения легко: разрезать каждый торт на равные куски и раздать их всем людям поровну. Заметьте, что наша последняя объявленная цель (разрезать $[178 : 14]$ как можно точнее) с блеском выполнена.

Итак, если все наши предположения выполнены, то минимальный кусок не превосходит $d = \frac{89}{7}$.

Часть II. Теперь мы построим пример, показывающий, что минимальный кусок может быть равен d. Для этого мы пойдём по нашему процессу в обратном порядке. Напоминаем, что в конце мы построили пример

$$[178 : 14] + [256 : 14] = \left(14 \cdot \frac{89}{7}\right) + \left(14 \cdot \frac{128}{7}\right).$$

Регулярный шаг 2. Торт $[256 : 14]$ получился из $8 \times [73 : 4]$ удалением дополнений к кускам из $2 \times [105 : 9]$. Восстановим эти куски — например, с помощью «метода отрезков»; сейчас этот метод легко применить:

$$8 \times [73 : 4] + 2 \times [105 : 9] = 8 \times \left(3 \cdot \frac{128}{7} + \frac{127}{7}\right) + 2 \cdot \left(5 \cdot \frac{89}{7} + 4 \cdot \frac{90}{7}\right).$$

Регулярный шаг 1. Торты $2 \times [105 : 9]$ получились из $3 \times [52 : 4] + 3 \times [52 : 4]$ выкидыванием дополнений к кускам из $[52 : 3] + [52 : 3]$; аналогично, торт $[178 : 14]$ получен из других $4 \times [52 : 3]$ выкидыванием дополнений к кускам из $2 \times [52 : 3]$. Теперь мы восстановим эти торты; это делается автоматически, после того, как мы скажем, как куски из $[105 : 9]$ и $[178 : 14]$ разбиваются на куски в тортах $[52 : 4]$. Сделав это произвольным образом, получаем

$$11 \times [52 : 4] + 4 \times [52 : 3] = 4 \times \left(2 \cdot \frac{89}{7} + \frac{90}{7} + \frac{96}{7}\right) + 2 \times \left(\frac{89}{7} + 2 \cdot \frac{90}{7} + \frac{95}{7}\right) + \\ + 4 \times \left(3 \cdot \frac{89}{7} + \frac{97}{7}\right) + 2 \times \left(2 \cdot \frac{89}{7} + 2 \cdot \frac{93}{7}\right) + \\ + 2 \times \left(2 \cdot \frac{121}{7} + \frac{122}{7}\right) + 2 \times \left(2 \cdot \frac{120}{7} + \frac{124}{7}\right).$$

Начальный шаг. Нам осталось восстановить последние торты $16 \times [52 : 3]$ из $8 \times [73 : 4]$ добавление долей обычных людей; это также происходит автоматически:

$$16 \times [52 : 3] = 8 \times \left(2 \cdot \frac{128}{7} + \frac{108}{7}\right) + 8 \times \left(\frac{128}{7} + \frac{127}{7} + \frac{109}{7}\right).$$

Таким образом, пример построен:

$$11 \times [52 : 4] + 20 \times [52 : 3] = 4 \times \left(2 \cdot \frac{89}{7} + \frac{90}{7} + \frac{96}{7}\right) + 2 \times \left(\frac{89}{7} + 2 \cdot \frac{90}{7} + \frac{95}{7}\right) + \\ + 4 \times \left(3 \cdot \frac{89}{7} + \frac{97}{7}\right) + 2 \times \left(2 \cdot \frac{89}{7} + 2 \cdot \frac{93}{7}\right) + \\ + 2 \times \left(2 \cdot \frac{121}{7} + \frac{122}{7}\right) + 2 \times \left(2 \cdot \frac{120}{7} + \frac{124}{7}\right) + \\ + 8 \times \left(2 \cdot \frac{128}{7} + \frac{108}{7}\right) + 8 \times \left(\frac{128}{7} + \frac{127}{7} + \frac{109}{7}\right).$$

Часть III. Нам осталось проверить, что все наши предположения должны были выполняться. Это несложно делается теми же методами, что и в предыдущих разделах. Обозначим $d = \frac{89}{6}$, $t = 31 - d = \frac{128}{7}$. Если $f(31, 52) > d$, то размеры всех кусков находятся в интервале (d, t).

Предположение (1) выполнено, иначе наименьший кусок не превосходит $\frac{31}{3} < d$.

Предположение (2) необходимо, иначе мы получаем кусок либо размера $\leq \frac{52}{5} < d$, либо размера $\geq 31 - \frac{52}{2} > t$.

Предположение (3) верно: иначе, рассматривая человека с двумя кусками в двух богатых тортах, получаем, что средний вес оставшихся 6 кусков в этих тортах равен $\frac{52 \cdot 2 - 31}{6} < d$.

Предположение (4): пусть оно не выполняется. Тогда два ребра имеют общую вершину. Рассмотрим три торта, участвующие в этих рёбрах; средний вес 5 их кусков, не соответствующих нашим рёбрам, равен $\frac{3 \cdot 52 - 2 \cdot 31}{5} > t$.

Предположение (5) проверяется так же, как и в 4.4б).

Наконец, **Предположение (6)** можно не проверять: к этому момкету у нас уже получился торт $[178 : 14]$, а значит, минимальный кусок не может превосходить $\frac{178}{14} = d$.

Готово!

Предлагаем вам попытаться применить алгоритм к парам с Испытательного полигона, чтобы увидеть, как он работает!

Честный раздел тортов

Решения после промежуточного финиша

Несколько общих результатов

Мы начнём с разбора задач 3.1–3.3, поскольку их решение упрощает остальные решения.

3.1. Если n не делится на m, то торты невозможно разрезать на куски веса $\frac{m}{n}$. Значит, найдётся кусок меньшего веса. Тогда кто-нибудь получит хотя бы два куска, и наименьший из них будет весить не более $\frac{m}{2n}$.

Пусть теперь $f(m,n) = \frac{m}{2n}$. Тогда каждый кусок весит либо $\frac{m}{2n}$, либо $\frac{m}{n}$. Каждый кусок веса $\frac{m}{n}$ можно разбить на два куска веса $\frac{m}{n}$. Итак, каждый торт состоит из кусков веса $\frac{m}{2n}$, что может случиться тогда и только тогда, когда $2m$ делится на n (а m не делится).

3.3. а) Рассмотрим любое оптимальное разбиение. Если у кого-то хотя бы три куска, то минимальный из них не превосходит $\frac{m}{3n}$; поэтому и $f(m,n) \le \frac{m}{3n}$.

3.2. а) Предположим, что $f(m,n) > \frac{m}{n} - \frac{1}{2}$. Поскольку $\frac{m}{n} > \frac{3}{4}$, имеем $\frac{m}{n} - \frac{1}{2} > \frac{m}{3n}$. Значит, по 3.3а) каждый человек получает не более двух кусков. Если есть кусок размера $\frac{m}{n}$, разделим его на два равных; полученное разбиение по-прежнему оптимально ввиду 3.1. Итак, всего есть ровно $2n$ кусков. Поскольку $2n < 3m$, найдётся торт, разделённый на две части (каждый торт должен быть порезан!); одна из этих частей не меньше $\frac{1}{2}$. Тогда у человека, получивший эту часть, оставшийся кусок не превосходит $\frac{m}{n} - \frac{1}{2}$. Противоречие.

б), в) Пусть $k \ge 3$; предположим, что $\frac{2}{k+1} < \frac{m}{n} < 1$, но $f(m,n) > \frac{m}{n} - \frac{1}{k}$. Поскольку $\frac{m}{n} - \frac{1}{k} > \frac{m}{3n}$, по 3.1 и 3.3а) мы опять можем считать, что у каждого человека ровно по два куска, и общее число кусков — $2n$. Так как $2n < (m+1)k$, найдётся торт с не более чем k кусками, а тогда один из них не меньше $\frac{1}{k}$. У человека, получившего этот кусок, вторая часть не будет превосходить $\frac{m}{n} - \frac{1}{k}$. Противоречие.

3.3. б) Рассмотрим отрезок длины m. Разделим его красными точками на m равных частей, а синими точками — на $3n$ равных частей (концы отрезка будут иметь оба цвета). Полученные отрезки с красными концами обозначают торты.

Теперь одновременно удалим все синие точки, одна из соседних с которыми точек — красная (но не разноцветная!). Затем разрежем торты по оставшимся синим точкам. Мы утверждаем, что теперь полученные части можно раздать n людям поровну.

Поскольку $m < n$, каждый торт содержал хотя бы три синих точки, поэтому хотя бы одна из них осталась. Далее, каждый полученный кусок, не находящийся на границе торта, имеет длину $\frac{m}{3n}$. То же верно для куска с разноцветным концом. Все же остальные куски разбиваются на пары, имеющие общий красный конец; общая длина двух кусков пары тогда будет равна $3 \cdot \frac{m}{3n} = 1$. Тогда каждую такую пару мы можем отдать одному человеку. Все остальные куски имеют длину $\frac{m}{3n}$, и их можно раздать по три оставшимся людям.

в) От противного, пусть $\frac{m}{n} \in \left(\frac{2}{3}, \frac{3}{4}\right]$, но $f(m,n) > \frac{1}{3}$. По 3.3а) и 3.1, можно предположить, что каждый человек получает ровно два куска. Тогда общее число кусков равно $2n < 3m$,

поэтому существует торт, разбитый на два куска. Один из этих кусков не меньше $\frac{1}{2}$. У человека, получившего его, второй кусок не больше $\frac{m}{n} - \frac{1}{2}$, что не превосходит $\frac{m}{3n}$? поскольку $\frac{m}{n} \leq \frac{3}{4}$. Противоречие.

Наконец, мы приведём также решение 1.4.

1.4. Рассмотрим такой набор кусков, что его можно распределить как по n равным тортам веса m, так и по m равным тортам веса n. Пусть x — максимальный возможный наименьший кусок в таком наборе. Тогда

$$nf(m,n) = x = mf(n,m),$$

откуда $f(m,n) = \frac{m}{n} f(n,m)$.

Некоторые значения функции f

Обозначение. При построении примеров мы будем обозначать через $(i_1 \cdot a_1 + i_2 \cdot a_2 + \cdots + i_l \cdot a_l)$ разрез торта на $i_1 + i_2 + \cdots + i_l$ кусков, среди которых i_1 кусков веса a_1, i_2 кусков a_2 и т.д. Так же будет обозначаться состав доли одного человека.

1.1. а) Частный случай 3.3в).

б) Оценка следует из 3.2б). Пример:

$$4 \times 210g = 2 \times (3 \cdot 70g) + 2 \times (3 \cdot 50g + 60g)$$

в) Оценка следует из 3.2в) при $k = 12$. Пример:

$$4 \times 3kg = 2 \times (12 \cdot 250g) + 2 \times (240g + 12 \cdot 230g) = 24 \times (230g + 250g) + (2 \cdot 240g) = 25 \times 480g.$$

1.2. а) **Ответ.** $\frac{5}{21}$.

Частный случай 3.3в).

б) **Ответ.** $\frac{5}{18}$.

Оценка следует из 3.2а). Пример:

$$7 \times 1 = 3 \times \left(2 \cdot \frac{1}{2}\right) + 2 \times \left(\frac{5}{18} + \frac{6}{18} + \frac{7}{18}\right) + 2 \times \left(2 \times \frac{5}{18} + \frac{8}{18}\right) =$$
$$= 6 \times \left(\frac{5}{18} + \frac{1}{2}\right) + \left(2 \cdot \frac{7}{18}\right) + 2 \times \left(\frac{6}{18} + \frac{8}{18}\right) = 9 \times \frac{7}{9}.$$

1.3. а) **Ответ.** $\frac{1}{3}$.

Как обычно, можно предположить, что у каждого человека по два куска. Тогда найдётся торт с тремя или более частями, одна из которых не превосходит $\frac{1}{3}$, поэтому $f(8,9) \leq \frac{1}{3}$. Пример:

$$8 \times 1 = 2 \times \left(3 \cdot \frac{1}{3}\right) + 6 \times \left(\frac{4}{9} + \frac{5}{9}\right) = 6 \times \left(\frac{1}{3} + \frac{5}{9}\right) + 3 \times \left(2 \cdot \frac{4}{9}\right) = 9 \times \frac{8}{9}.$$

б) **Ответ.** $\frac{2}{7}$.

Оценка $f(11,14) \geq \frac{2}{7}$ следует из 3.2а). Пример:

$$11 \times 1 = 5 \times \left(2 \cdot \frac{1}{2}\right) + 4 \times \left(2 \cdot \frac{2}{7} + \frac{3}{7}\right) + 2 \times \left(\frac{2}{7} + 2 \cdot \frac{5}{14}\right) =$$
$$= 10 \times \left(\frac{1}{2} + \frac{2}{7}\right) + 4 \times \left(\frac{3}{7} + \frac{5}{14}\right) = 14 \times \frac{11}{14}$$

в) **Ответ.** $\frac{5}{17}$.

Как обычно, можно предположить, что у каждого человека по два куска. Тогда всего есть 34 куска. Заметим, что каждый торт разрезан на два или три куска: целого торта быть не может, а если какой-то торт разрезан хотя бы на 4 куска, то один из них не превосходит $\frac{1}{4} < \frac{5}{17}$. Тогда есть 6 *богатых* тортов с тремя кусками и 8 *обычных* тортов с двумя кусками; в богатых тортах всего 18 кусков, а в обычных — 16 кусков. Поэтому найдётся человек, у которого оба куска — из богатых тортов (если они из одного торта, то выберем ещё один богатый произвольно). Суммарный вес остальных 4 кусков в этих тортах не превосходит $2 - \frac{14}{17} = \frac{20}{17}$. Поэтому один из них не тяжелее $\frac{5}{17}$.

Пример:

$$14 \times 1 = 8 \times \left(\frac{9}{17} + \frac{8}{17}\right) + 4 \times \left(2 \cdot \frac{6}{17} + \frac{5}{17}\right) + 2 \times \left(\frac{7}{17} + 2 \cdot \frac{5}{17}\right) =$$
$$= 8 \times \left(\frac{5}{17} + \frac{9}{17}\right) + 8 \times \left(\frac{6}{17} + \frac{8}{17}\right) + \left(2 \cdot \frac{7}{17}\right) = 17 \times \frac{14}{17}.$$

2.1. а) **Ответ.** $\frac{1}{3}$.

Пусть $s = f(3k-1, 3k)$. Докажем сначала, что $s \le \frac{1}{3}$. Пусть это не так. Тогда $s > \frac{1}{3}$, и согласно 3.3а) можно считать, что у любого человека ровно два куска. Тогда найдётся торт, разрезанный хотя бы на 3 куска, и наименьший из них не превосходит $\frac{1}{3}$. Противоречие.

Для построения примера умножим все веса на $3k$. Имеем:

$$(3k-1) \times 3k = 2 \times \big(3 \cdot k\big) +$$
$$+ 3 \times \big((2k-1) + (k+1)\big) + 3 \times \big((2k-2) + (k+2)\big) + \cdots + 3 \times \big((k+1) + (2k-1)\big) =$$
$$= 3 \times \big(k + (2k-1)\big) + 3 \times \big((k+1) + (2k-2)\big) + \cdots + 3 \times \big((2k-1) + k = (3k+2) \times (6k+2).$$

б) **Ответ.** $\frac{2k+1}{2(3k+2)}$.

Обозначим $s = f(3k+1, 3k+2)$, $t = \frac{2k+1}{2(3k+2)}$. Сначала покажем, что $s \le t$. Пусть это не так. Опять же, из $s > t \ge \frac{3k+1}{3(3k+2)}$ можно предположить, что у каждого человека по два куска. Если какой-то торт разрезан хотя бы на 4 части, то наименьшая из них не превосходит $\frac{1}{4} \le \frac{2k+1}{2(3k+2)}$, что невозможно. Поэтому каждый торт разрезан на 2 или 3 куска, и легко понять, что есть ровно два торта, разделённых на 3 куска.

Рассмотрим теперь следующий граф. Вершинами являются торты, а каждому человеку сопоставлено ребро, соединяющее торты, из которых он получил свои куски. В этом графе есть две *выделенных* вершины степени 3, а все остальные имеют степень 2. Поэтому он состоит из трёх путей, соединяющих выделенные вершины (возможно, некоторые из них — циклы). Длина одного из этих путей не меньше $k+1$; рассмотрим этот путь $v_0, v_1, \ldots, v_{k+1}$.

Обозначим долю человека, соответствующего ребру (v_i, v_{i+1}), через (p_i, q_i), где кусок p_i взят из торта v_i, а q_i — из v_{i+1}. Тогда $p_i + q_i = \frac{3k+1}{3k+2}$ при $i = 0, 1, \ldots, k$, и $q_i + p_{i+1} = 1$ при $i = 1, 2, \ldots, k$. Поэтому $p_{i+1} - p_i = \frac{1}{3k+2}$, а значит, $p_k \ge \frac{k}{3k+2} + p_0 \ge \frac{k}{3k+2} + t = \frac{4k+1}{2(3k+2)}$. Наконец, получаем $q_k = \frac{3k+1}{3k+2} - p_k \le t$. Противоречие.

В решении содержится также идея примера: должны получиться два пути длины $k+1$ и один — длины k. length k. Для удобства умножим все веса на $2(3k+2)$. Тогда пример выглядит так:

$$(3k+1) \times (6k+4) = 2 \times \big(2 \cdot (2k+1) + (2k+2)\big) +$$
$$+ 2 \times \big((4k+1) + (2k+3)\big) + 2 \times \big((4k-1) + (2k+5)\big) + \cdots + 2 \times \big((2k+3) + (4k+1)\big) +$$
$$+ \big(4k + (2k+4)\big) + \big((4k-2) + (2k+6)\big) + \cdots + \big((2k+4) + 4k\big) =$$
$$= 2 \times \big((2k+1) + (4k+1)\big) + 2 \times \big((2k+3) + (4k-1)\big) + \cdots + 2 \times \big((4k+1) + (2k+1)\big) +$$
$$+ \big((2k+2) + 4k\big) + \big((2k+4) + (4k-2)\big) + \cdots + \big(4k + (2k+2)\big) = (3k+2) \times (6k+2).$$

в) **Ответ.** $\frac{k}{3k+1}$.

Обозначим $s = f(3k, 3k+1), t = \frac{k}{3k+1}$. Опять же, предполагая, что $s > t$, мы можем считать, что у каждого человека по два куска, есть ровно два торта с тремя кусками, а остальные содержат по два куска. Далее, строя аналогичный граф, мы получаем, что один из путей длины хотя бы $k+1$. Действуя как и выше, мы находим, что $p_k - p_0 = \frac{k}{3k+1}$, поэтому $p_k \geq \frac{2k}{3k+1}$ и $q_k = \frac{3k}{3k+1} - p_k \leq t$. Противоречие.

Для построения примера умножим все веса на $3k+1$. Пример выглядит так:

$$\begin{aligned}
3k\times(3k+1) &= 2\times\big(2\cdot k+(k+1)\big)+\\
&\quad+2\times\big(2k+(k+1)\big)+2\times\big((2k-1)+(k+2)\big)+\cdots+2\times\big((k+1)+2k\big)+\\
&\quad+\big((2k-1)+(k+2)\big)+\big((2k-2)+(k+3)\big)+\cdots+\big((k+2)+(2k-1)\big)=\\
&=2\times\big(k+2k\big)+2\times\big((k+1)+(2k-1)\big)+\cdots+2\times\big(2k+k\big)+\\
&\quad+\big((k+1)+(2k-1)\big)+\big((k+2)+(2k-2)\big)+\cdots+\big((2k-1)+(k+1)\big)=(3k+1)\times 3k.
\end{aligned}$$

2.3. а) Возможны три случая, в зависимости от остатка от деления n на 3.

1) $n = 3k$. Очевидно, $f(3, 3k) = \frac{1}{k}$.

2) $n = 3k+1$. Если $n = 1$, то $f(3,1)$=1. Если же $n = 3k+1 \geq 4$, то $f(3, 3k+1) = \frac{3k-1}{2k(3k+1)}$. Оценка следует из 3.2в) (при $k' = 2k$). Пример:

$$\begin{aligned}
3\times 1 &= \left(2k\cdot\frac{1}{2k}\right)+2\times\left(k\cdot\frac{3k-1}{2k(3k+1)}+(k+1)\cdot\frac{3}{6k+2}\right)=\\
&=2k\times\left(\frac{3k-1}{2k(3k+1)}+\frac{1}{2k}\right)+(k+1)\times\left(2\cdot\frac{3}{6k+2}\right).
\end{aligned}$$

3) $n = 3k+2$. В этом случае $f(3, 3k+2) = \frac{1}{2k+2}$.

Можно предположить, что у каждого человека по два куска, и общее число кусков равно $6k+4$. Тогда найдётся торт с хотя бы $2k+2$ кусками, один из которых не превосходит $\frac{1}{2k+2}$. Пример:

$$\begin{aligned}
3\times 1 &= \left((2k+2)\cdot\frac{1}{2k+2}\right)+2\times\left((k+1)\cdot\frac{3k+4}{(3k+2)(2k+2)}+k\cdot\frac{3}{6k+4}\right)=\\
&=(2k+2)\times\left(\frac{1}{2k+2}+\frac{3k+4}{(3k+2)(2k+2)}\right)+k\times\left(2\cdot\frac{3}{6k+4}\right).
\end{aligned}$$

Fair cake division

I.I. Bogdanov, G.R. Chelnokov, K.A. Knop, I.N. Shnurnikov

1 General setting

Firstly, we present several model problems of this project. In every problem, a first question is easy; but then the difficulty grows fast. For instance, problem 1.3e) is very hard!

Warning! If you are stuck with some question for a long time, we advice to switch to another problems of the project. Perhaps you will find there some clues or hints. For instance, you may find that problem 3.3b) is very useful.

1.1. a) Three small cakes, each of weight 200 g, are divided into some pieces. It happens that one can distribute all the pieces to 4 children so that each gets the pieces of the same total weight. Prove that the minimal piece weight is not more than 50 g. Is it possible to replace the number 50 by a smaller one?

b) Four small cakes, each of weight 210 g, are divided into some pieces. It happens that one can distribute all the pieces to 7 children so that each gets the pieces of the same total weight. Prove again that the minimal piece weight is not more than 50 g. Is it possible to replace the number 50 by a smaller one?

c) Now we have four large cakes, each of weight 3 kg, and we divide them into pieces so that it is possible to distribute all the pieces to 25 children so that each gets the same total weight. Prove that the minimal piece weighs not more than 230 g. Is it possible to replace the number 230 by a smaller one?

1.2. a) We have five cakes of 1 kg each, and we need to cut them and to distribute the pieces to seven people. We need the minimal piece weight to be the largest possible. Find this largest possible minimal weight.

b) The same problem for 7 cakes and 9 people.

1.3. a) We need to cut eight cakes of 1 kg each and to distribute the pieces to 9 people. Find the largest possible weight of the smallest piece in this division.

b) The same problem for 11 cakes and 14 people.

c) The same problem for 14 cakes and 17 people.

d) The same problem for 13 cakes and 16 people.

e) The same problem for 31 cakes and 52 people.

Surely, all these problems are the particular cases of the following general setting. (We always assume that the variables denote some positive integers.)

Megaproblem. There are m cakes (of weight 1 kg each) and n people. We need to divide the cakes into pieces and to distribute them all to the people so that each gets the same total weight of pieces. We aim to maximize the minimal piece weight. So we need to find this largest possible weight of the minimal piece.

Definition. Denote by $f(m,n)$ the answer to the Megaproblem.

Although the problem seems to be quite innocent, it is very hard to solve it in this general setting. It happens that in several iterations one can find the answers for "most" values of m and n, but at every step there still remain pairs (m,n) for which the answer is unknown.

It seems that there is no answer in a closed form. Thus our main purpose is to construct an algorithm of solving Megaproblem for every particular case. We do not formulate the search of this algorithm as a separate problem, but please keep in mind that this is the guiding star of this project.

Attention! If you think that you have invented the general algorithm (or even some well-formulated conjectures of how should it look like) — we are always open to discuss that! Moreover, it also concerns to the algorithm which works for some (large enough) interval of values of the ratio m/n.

We finish this introduction with formulating three general problems. The first one is quite easy; surprisingly, the answer for the second one is yet unknown. Formally speaking, the third one is not connected with the project, but its solution may be helpful.

1.4. Given the value of $f(m,n)$, determine $f(n,m)$.

Remark. In view of this problem, it is enough to investigate only the case $m < n$. Hence, further we deal with this case only.

1.5*. Determine whether the relation $f(tm,tn) = f(m,n)$ always holds.

Remark. The authors are almost sure that the answer is affirmative; this is confirmed in all known particular cases. So, in many further problems we deal with the ratio m/n instead of the pair (m,n).

1.6. a) We need to divide m equal cakes and to distribute them among n people so that each gets the same total weight of pieces. Find the minimal number of pieces in such division.

b) For such a division with a minimal number of pieces, find all possible weights of the minimal piece.

In the next two sections we collect the questions mainly concerning the particular cases of Megaproblem: in Section 2 — for some sequences of values of m/n, and in Section 3 — (mostly) for some intervals of values. In each Section, the problems are arranged (more or less) by the difficulty in an increasing order. We recommend to try to solve the problems from all Sections simultaneously.

On the other hand, if you will solve one of the Sections (almost) up to the end, we are always ready to add some more difficult problems.

2 Some special sequences of values

2.1. a) Determine $f(3k-1,3k)$.
b) Determine $f(3k+1,3k+2)$.
c) Determine $f(3k,3k+1)$.

2.2. Prove that $f(m,2m-1) = \frac{m+1}{6m-3}$ for all $m \geq 4$.

2.3. a) Determine $f(3,n)$.
b) Determine $f(4,n)$.
c) And also $f(5,n)$.

2.4. Determine $f(m,2m+1)$.

2.5. Determine $f(2k+1,3k+2)$.

2.6. Determine $f(3k+1,4k+1)$.

2.7. Determine $f(5k+2,8k+3)$.

2.8. Determine $f(5k-1,9k-2)$.

2.9. Determine $f(17k-4,21k-5)$.

3 Serial results

We consider the problems of this Section as some steps towards the general algorithm (but surely not all of them!). So, if you get some different **serial** results — do not hesitate to submit them!

Recall that we always assume $m < n$.

3.1. Given that m does not divide n, prove that $f(m,n) \le \frac{m}{2n}$. Determine also all the pairs (m,n) for which the equality is achieved.

3.2. a) Assume that $\frac{3}{4} < \frac{m}{n} < 1$. Prove that $f(m,n) \le \frac{m}{n} - \frac{1}{2}$.

b) Assume that $\frac{1}{2} < \frac{m}{n} < 1$. Prove that $f(m,n) \le \frac{m}{n} - \frac{1}{3}$.

c) Assume that $\frac{2}{k+1} < \frac{m}{n} < 1$ with $k \ge 4$. Prove that $f(m,n) \le \frac{m}{n} - \frac{1}{k}$.

Attention! Part b) of the next problem is very important!

3.3. a) Suppose that $f(m,n) > \frac{m}{3n}$. Prove that in every optimal distribution for the pair (m,n) each man gets not more than 2 pieces.

b) (Theorem on One Third) Prove that $f(m,n) \ge \frac{m}{3n}$.

c) Given that $\frac{2}{3} < \frac{m}{n} \le \frac{3}{4}$, prove that $f(m,n) = \frac{m}{3n}$. (See also problem 3.11.)

3.4. a) Given that $\frac{m}{n} < \frac{2}{3}$, prove that $f(m,n) \le \frac{1}{4}$.

b) Determine all pairs (m,n) (with $\frac{m}{n} < \frac{2}{3}$) such that $f(m,n) = \frac{1}{4}$.

3.5. Determine all pairs (m,n) such that $\frac{2}{k+1} < \frac{m}{n} < \frac{2}{k}$ and $f(m,n) = \frac{1}{k+1}$.

3.6. a) Prove that $f(m,n) = \frac{m}{n} - \frac{1}{3}$ if $\frac{1}{2} < \frac{m}{n} \le \frac{5}{9}$.

b) Determine all pairs (m,n) such that $f(m,n) = \frac{m}{n} - \frac{1}{3}$.

3.7. Determine all pairs (m,n) such that $\frac{2}{k+1} < \frac{m}{n} < \frac{2}{k}$ and $f(m,n) = \frac{m}{n} - \frac{1}{k}$ (for $k \ge 4$).

3.8. a) Given that $\frac{7}{15} < \frac{m}{n} < \frac{1}{2}$, find $f(m,n)$.

b) Find all values of m/n for which $f(m,n)$ has the same form.

3.9. a) Given that $\frac{7}{12} < \frac{m}{n} < \frac{22}{37}$, find $f(m,n)$.

b) Find all values of m/n for which $f(m,n)$ has the same form.

3.10. a) Given that $\frac{14}{17} < \frac{m}{n} < \frac{5}{6}$, find $f(m,n)$.

b) Find all values of m/n for which $f(m,n)$ has the same form.

3.11*. Determine all pairs (m,n) such that $f(m,n) = \frac{m}{3n}$.

Fair cake division

After semifinal

Additional problems to the previous sections

Just in case, we remind the problem added on the first presentation.

1.6. a) We need to divide m equal cakes and to distribute them among n people so that each gets the same total weight of pieces. Find the minimal number of pieces in such division.

b) For such a division with a minimal number of pieces, find all possible weights of the minimal piece.

Next problems form the addition to the third section. Namely, these are some bounds analogous to the theorem on one third, but more exact ones.

3.12. Prove that $f(m,n) \geq \frac{2}{5} \cdot \frac{m}{n}$, if $\frac{5}{12} \leq \frac{m}{n} \leq \frac{1}{2}$.

3.13. a) Prove that $f(m,n) \geq \frac{3}{8} \cdot \frac{m}{n}$ if $\frac{m}{n} \leq \frac{1}{2}$.

b) Find more intervals on which this inequality is true.

3.14. a) Prove that $f(m,n) \geq \frac{2}{5} \cdot \frac{m}{n}$, if $\frac{3}{5} \leq \frac{m}{n} \leq \frac{8}{13}$.

b) Try to prove this inequality for another interval, adjacent to the $\left(\frac{3}{5}, \frac{8}{13}\right)$. For example, is it true for $\frac{m}{n} \in \left(\frac{10}{17}, \frac{3}{5}\right)$? Or for $\frac{m}{n} \in \left(\frac{8}{13}, \frac{5}{8}\right)$?

c) Find more intervals (in other places of the segment $[0,1]$) where this bound is true.

3.15. For which intervals inside $\left(\frac{1}{2}, \frac{5}{8}\right)$ you can prove the bound $f(m,n) \geq \frac{2}{3} \cdot \frac{m}{n} - \frac{1}{6}$?

Testing area.

This section is useful for everyone who wants to have some nontrivial pairs (m,n), here are some of these pairs. **Attention!** We can check answers and examples for these pairs, but we will not check a proof, unless it includes some general ideas; so this pairs are not formed as a problem.

Here are these pairs (This list can be replenished):

$$(17,29); \quad (31,70); \quad (17,47); \quad (117,133); \quad (27,61); \quad (566,643); \quad (3130,6813).$$

Good luck!

4 Variations of the problems setting

In this section we generalize original setting in different ways. Solutions of these problems can be very useful in solving Megaproblem.

4.1. a) We have m cakes with weight 1 and $n > m$ people. We have to cut cakes and give it to the people so that each gets the same total weight of pieces. And herewith each gets at most two pieces and each cake is cut into at most three parts. Find all such pairs (m,n).

b) The same problem, but each cake is cut into at most k parts.

c) The same problem, but each cake is cut into either $k-1$ or k parts.

Next problems deal with the situation when cakes can be different.

4.2. a) Two cakes with weights 1 kg and 2 kg are divided between N people so that each gets the same total weight of pieces. What is the maximal weight of the minimal piece?

b) The same question for two cakes with weights 2 kg and 5 kg.

4.3. a) Let $k > 1$. There are $3k$ cakes (each of weight 3), $k-1$ cakes (each of weight 4), and $3k-1$ cakes (each of weight 7). We have to cut each cake of the weight 3 into two pieces, and the each of the others — into three pieces so that the pieces may be distributed among several people and each of them will have two pieces with the same total weight. What is the maximal possible size of the minimal piece?

b) The same question for $3k$ cakes of the weight 3, $k+2$ cakes of the weight 4, and $3k+2$ cakes of the weight 7.

c) The same question for $3k$ cakes of the weight 3, $2k-1$ cakes of the weight 4, and $4k-1$ cakes of the weight 7, where $k \geq 10$. What can you say for other values of k (for example, $k = 7$)?

4.4. a) We have a cake with weight 59, a cake with weight 89 and two cakes of weight 41. We have to cut the first cake into 4 pieces, the second cake into 6 pieces, and the each of the last into 5 pieces so that they may be distributed among 10 persons so that all persons will have the same number of pieces and the same total weight of their pieces. What is the maximal value of the minimal piece?

b) There are two cakes with weight 41, three cakes with weight 35, and 11 cakes with weight 29. Each cake of the first group should be divided into 5 pieces, each cake of the second group — into 4 pieces, and the each cake of the third group — into 2 pieces. The pieces should be distributed among 22 persons so that all persons have the same number of pieces and the same total weight of their pieces. What is the maximal possible value of the minimal piece?

c) Find $f(23, 29)$.

Fair cake division

Selected solutions

If you have any ideas on this project, please do not hesitate to contact us by the e-mail:
Konstantin Knop `kostyaknop@gmail.com`, Ilya Bogdanov `ilya.i.bogdanov@gmail.com`

The presented solutions are arranged as follows. In section "Some sequences" we find the values of f on some sequences of pairs (m, n); notice that plenty of them also follow from more general results from the next sections. Section "Serial results" contains the solutions (or their outlines) for problems 3.4–3.8 (with problem 4.1 as a useful lemma) and bound 3.13. In section "Nonequal cakes" we extend our methods; it allows us to approach to problems 3.9 and 3.10 (it is recommended to read section on serial results before). Finally, in section "General algorithm" we describe on concrete examples the ideas of a general algorithm of solving Megaproblem (it involves some ideas from the previous two sections).

We start with the solution of problem 1.6.

1.6. a) **Answer.** $m + n - \gcd(m, n)$.

To construct an example, consider a segment of length m. Divide it by red points into m equal segments, and by blue points into n equal segments (some points will be multicolored). The segments with the red endpoints represent the cakes. We cut the cakes by all the blue endpoints. We claim that a desired division is obtained. Obviously, these pieces may be distributed among the people: it suffices to give to each person pieces between some neighboring blue points. There are $m + 1$ red points, $n + 1$ blue points, and $\gcd(m, n) + 1$ multicolored points. So the total number of points is $m + n - \gcd(m, n) + 1$ and we have the required number of pieces.

We are left to show that the number of pieces should be at least $m + n - \gcd(m, n)$. Denote $d = \gcd(m, n)$, $n = dn'$, $m = dm'$. Consider a bipartite graph with m red vertices and n blue vertices, corresponding to cakes and people. Each edge corresponds to a piece, and it connects the person getting the piece with the cake it is taken from. Consider a connected component of this graph, let it have r red vertices and b blue vertices. Then b persons eat together r cakes, which means that $b \cdot \frac{m}{n} = r$. So $\frac{r}{b} = \frac{m'}{n'}$ and hence $m' \mid r$. So, the number of connected components is at most $\frac{m}{m'} = d$. On the other hand, in each component the number of edges is at least the number of vertices decreased by 1. So the total number of edges in the graph (which is the number of pieces) is at least $m + n - d$.

b) The solution is left to the reader.

Some sequences

Here we present the solutions for some problems from section 2. Many of them follow also from more general results from section 3; nevertheless, we have put them here to show more concrete constructions.

General remark. Since the case $m \mid n$ is trivial, further we always assume that $m \nmid n$.

2.2. By 3.2b) it follows that $f(m, 2m-1) \le \frac{m+1}{6m-3}$. Example of decomposition of cakes of weights $6m - 3$:

$$2 \times \Big(3 \cdot (2m-1)\Big) + \Big(2 \cdot (m+1) + (2m-4) + (2m-1)\Big) + 2 \times \Big(2 \cdot (m+1) + (2m-2) + (2m-3)\Big) + \Big((m+1+i) + (m+3+i) + (2m-4-i) + (2m-3-i)\Big)_{i=1,\dots,(m-5)}$$

2.3. b) **Answer.** $\frac{4}{n}$ if $4 \mid n$; $\frac{2}{n}$ if $n = 4k+2$; $\frac{4}{n} - \frac{2}{n-1}$ if n is odd.

If $4 \mid n$ then the answer $f(4,n) = \frac{4}{n}$ is trivial. If $n = 4l+2$ is even then $f(4, 4l+2) = \frac{1}{2l+1}$ by 3.1.

Let $n = 2k+1$ is odd. By 3.3a) we may assume that every person gets exactly two pieces. So we have $4k+2$ pieces, and by the pigeonhole principle there exists a cake with not more than k pieces. So there is a piece not less than $\frac{1}{k}$, and its complement (at a person) is at most $\frac{4}{n} - \frac{1}{k} = \frac{4}{n} - \frac{2}{n-1}$; hence $f(4,n) \le \frac{4}{n} - \frac{2}{n-1}$. An example for odd n:

$$2 \times \left(\frac{n-1}{2} \cdot \frac{2}{n-1}\right) + 2 \times \left(\frac{n-1}{2} \cdot \left(\frac{4}{n} - \frac{2}{n-1}\right) + \frac{2}{n}\right) =$$
$$= (n-1) \times \left(\frac{2}{n-1} + \left(\frac{4}{n} - \frac{2}{n-1}\right)\right) + \left(2 \cdot \frac{2}{n}\right) = n \cdot \frac{4}{n}.$$

c) **Answer.** $\frac{5}{n}$ if $n \vdots 5$; $\frac{1}{\lceil 2n/5 \rceil}$ if $n = 5k+1 \ge 16$ or $n = 5k+3$; $\frac{5}{n} - \frac{1}{\lfloor 2n/5 \rfloor}$ if $n = 5k+4$ or $n = 5k+2 \ge 12$; $f(5,11) = \frac{13}{66}$. The other examples follow from the previous problems.

2.4. **Answer.** $\frac{1}{5}$ for $m \ge 6$ and $m = 2$; the other answers follow from the previous problems: $f(1,3) = \frac{1}{3}$, $f(3,7) = \frac{5}{28}$, $f(4,9) = \frac{7}{36}$, $f(5,11) = \frac{13}{66}$.

2.5. **Answer.** $\frac{1}{4}$ for $k \ge 1$. The bound follows from 3.4 a). Example for cake weight $12k+8$:

$$1 \times \Big(4 \cdot (3k+2)\Big) + 2 \times \Big((4k+2) + (5k+4-2i) + (3k+2+2i)\Big)_{i=1,\dots,k} =$$
$$= k \times \Big(2 \cdot (4k+2)\Big) + 2 \times \Big((3k+2) + (5k+2)\Big) + 2 \times \Big((3k+4) + 5k\Big) + \dots + 2 \times \Big((5k+2) + (3k+2)\Big).$$

Note that the problem follows from 3.4b).

2.6. **Answer.** $\frac{2k+1}{2(4k+1)}$. The bound follows from 3.2 a). Example for cake weight $8k+2$:

$$(k+1) \times \Big(2 \cdot (4k+1)\Big) + 2 \times \Big((2k+1) + (2k+i) + (4k+1-i)\Big)_{i=1,\dots,k} =$$
$$= (2k+2) \times \Big((2k+1) + (4k+1)\Big) + 2 \times \Big((4k+1-i) + (2k+1+i)\Big)_{i=1,\dots,(k-1)} + \Big(2 \cdot (3k+1)\Big).$$

2.7. **Answer.** $\frac{1}{4}$. Bound follows from 3.4a). Example for cake with weight $32k+12$.

$$(5k+2) \times (32k+12) =$$
$$= k \times \Big(4 \cdot (8k+3)\Big) + 2 \times \Big(2 \cdot (10k+4) + (12k+4)\Big) + \Big((12k+5) + (8k+3+i) + (12k+4-i)\Big)_{i=1,\dots,4k}.$$

Note that the existence of example follows from 3.4b).

2.8. **Answer.** $\frac{6k-1}{3(9k-2)}$. The bound follows from 3.2b). Example with cake weights $27k-6$ and portions of people $15k-3$:

$$2k \times \Big(3 \cdot (9k-2)\Big) + \Big((6k-1) + (6k-1) + (6k-2+i) + (9k-2-i)\Big)_{i=1,\dots,(3k-1)} =$$
$$= 6k \times \Big((6k-1) + (9k-2)\Big) + \Big((6k-1+i) + (9k-2-i)\Big)_{i=1,\dots,3k-2}$$

2.9. **Answer.** $\frac{18k-4}{63k-15}$. The bound follows from the lemma.

Lemma. If $\frac{4}{5} < \frac{m}{n} < 1$, then $f(m,n) \le 2 \cdot \frac{m}{n} - \frac{4}{3}$.

Proof. Suppose the contrary. It is easy to check that $2\cdot\frac{m}{n}-\frac{4}{3}\ge\frac{m}{3n}$ and $2\cdot\frac{m}{n}-\frac{4}{3}\ge\frac{1}{4}$, so all the cakes contain two or three pieces and all people have two pieces. The number of two-piece cakes is $3m-2n$, the number of three-piece cakes is $2n-2m$. Two cases are possible.

1) Assume that some person gets both pieces from two-piece cakes. Then the remaining two pieces from these two cakes weigh in total $2-\frac{m}{n}$, and one of pieces is at least $1-\frac{m}{2n}$. Hence the completing piece is at most

$$\frac{m}{n}-\left(1-\frac{m}{2n}\right)<2\cdot\frac{m}{n}-\frac{4}{3}.$$

2) Every piece of two-piece cake is completed by a piece of a three-piece cake. Let x be the minimal piece weight. Then $\frac{m}{n}-x$ is the maximal weight. Let A be some piece of a two-piece cake; then $A\ge1-\frac{m}{n}+x$. Hence the completing piece is $\frac{m}{n}-A\le2\frac{m}{n}-1-x$. From $\frac{m}{n}>\frac{4}{5}$ it follows that there are more two-piece cakes than three-piece ones. So there is a three-piece cake, all pieces of which are complementary to pieces of two-piece cakes. Hence all three pieces are at most $2\frac{m}{n}-1-x$ and

$$3\left(2\cdot\frac{m}{n}-1-x\right)\le1\Leftrightarrow x\le2\cdot\frac{m}{n}-\frac{4}{3}.$$

□

Example. Cake weight is $63k-15$, portion of a person is $51k-12$, the minimal weight is $18k-4$.

$$(17k-4)\times(63k-15)=2k\times\Big(3\cdot(21k-5)\Big)+6k\times\Big((33k-8)+(30k-7)\Big)+$$
$$+\Big((33k-10-i)+(30k-7+i)\Big)_{i=1,\dots,3k-2}+2\times\Big((18k-3+i)+(27k-6-i)+(18k-4)\Big)_{i=1,\dots,3k-1}=$$
$$=6k\times\Big((21k-5)+(30k-7)\Big)+6k\times\Big((33k-8)+(18k-4)\Big)+$$
$$2\times\Big((18k-4+i)+(33k-8-i)\Big)_{i=1,\dots,3k-2}+\Big((24k-6+i)+(27k-6-i)\Big)_{i=1,\dots,3k-1}=$$
$$=(21k-5)\times(51k-12).$$

Serial results

First, we present an estimate analogous to the Theorem on One Third.

3.13. a) If $n=2m$, then $f(m,2m)=\frac{1}{2}$, so we assume $\frac{m}{n}<\frac{1}{2}$. Let $8n$ be the weight of every cake, then each person receives $8m$.

Consider the segment of length $8nm$ and divide it by red points into m equal segments (cakes). We will cut off the pieces consequently from the left end of the remaining segment. Cut off several pieces of $4m$ until the remainder will be between $6m$ and $10m$. Next, we divide the remaining part into two equal pieces of length between $3m$ and $5m$. We complete both of these two pieces to $8m$ by two pieces from the next cake. So we use at most $10m$ from the next cake, and the remainder is at least $8n-10m\ge6m$. Thus we may continue cutting until the last cake. Since we have combined extracted pieces to pairs of weight $8m$, and the total segment equals to $8mn$, then in the last cake will be ended by two pieces of $4m$. The distribution of pieces among people is also constructed.

Next, we present some exact values for several intervals. We start with problem 4.1, which happens to be very helpful.

4.1. a), b) **Answer.** $\frac{m}{n}\in\left[\frac{1}{k-1},1\right)\cup\left\{\frac{v}{(k-1)v+1}\right\}_{v=1,2,\dots}$.

Firstly, we present an example showing that the answer fits. We act as in Theorem on One Third. Consider a segment of length m. Divide it by red points into m equal segments, and by blue points into n equal segments (some points will be multicolored). The obtained segments with the red endpoints represent the cakes. Now we cut the cakes by all the blue endpoints.

We claim that a desired division is obtained. Obviously, these pieces may be distributed among the people: it suffices to give to each person pieces between some neighboring blue points. Moreover, each person gets no more than two pieces since each such segment may contain at most one red point. We are left to show that each cake is divided into at most k parts.

If $\frac{m}{n} > \frac{1}{k-1}$, then each cake contains at most $k-2$ whole portions and at most two pieces less than $\frac{m}{n}$ — thus at most k pieces at all. For $\frac{m}{n} = \frac{1}{k-1}$ the claim is obvious. Now assume that $\frac{m}{n} = \frac{v}{(k-1)v+1}$. Consider now k consecutive blue points; the distance between the first and the last of them is $\frac{(k-1)v}{(k-1)v+1}$, and their coordinates are the fractions with denominator $(k-1)v+1$. This implies that a cake can contain all these k points only if one of them is its endpoint. This means exactly that a cake is cut into no more than k pieces. The example is justified.

We are left to prove that in other cases the desired division is impossible. Let us call a person *angry* if he gets two pieces. Let us construct a graph having cakes as vertices, with each edge corresponding to an angry person and connecting two cakes he gets his pieces from. Consider any connected component of this graph; let v and e be the numbers of its vertices and edges, respectively. Then $e \geq v-1$.

These v cakes contain at most kv pieces, $2e$ of which belong to angry people. Since there are no edges from our component outside it, these pieces can be rearranged into whole portions (namely, e portions of two pieces each and, say, t of portions of one piece). Then $t \leq kv-2e$. Next, comparing the total weight at v cakes and $e+t$ people, we get $\frac{n}{m} = \frac{e+t}{v} \leq \frac{kv-e}{v} = k - \frac{e}{v}$. If $e \geq v$, then we get $\frac{n}{m} \leq k-1$, otherwise $e = v-1$, and we have $\frac{n}{m} = \frac{v(k-1)+1}{v}$, as desired.

Important remark. Had we omitted the condition $m < n$, the answer would expand a bit. Surely it will include 1; now assume that $m > n$. In this case, all people are angry since a person's portion is greater than a cake. Now, each connected component of a graph corresponds to the division of v cakes between e people, hence $1 < \frac{m}{n} = \frac{v}{e}$ which may happen only if $e = v-1$. For such values an example can be constructed in the same way; hence the answer becomes

$$\frac{m}{n} \in \left[\frac{1}{k-1}, 1\right] \cup \left\{\frac{v}{(k-1)v+1}\right\}_{v=1,2,\ldots} \cup \left\{\frac{e+1}{e}\right\}_{e=1,2,\ldots}.$$

c) **Answer.** $\frac{m}{n} \in \left[\frac{1}{k-1}, \frac{2}{k-1}\right] \cup \left\{\frac{v}{(k-1)v+1}\right\}_{v=1,2,\ldots}$.

The solution is left to the reader.

3.4. a) As usual, we may assume that each person receives at least two pieces. Then the total number of pieces is at least $2n > 3m$. Hence some cake contains at least four pieces, one of which should not exceed $\frac{1}{4}$.

b) **Answer.** $\frac{m}{n} \in \left[\frac{5}{8}, \frac{2}{3}\right) \cup \left\{\frac{5k+2}{8k+4}\right\}_{k=1,2,\ldots}$.

Suppose that $f(m,n) = \frac{1}{4}$; then $f(m,n) > \frac{m}{3n}$, so we may assume that each person gets exactly two pieces. Next, each piece is at least $\frac{1}{4}$ and at most

$$d = \frac{m}{n} - \frac{1}{4}$$

(otherwise the other piece at the person having our one is less than $\frac{1}{4}$). Hence each cake contains at least three pieces (otherwise there exists a piece of at least $\frac{1}{2} > d$) as well as at most 4 pieces (otherwise there exists a piece not exceeding $\frac{1}{5}$). Thus, we have *fat* cakes with 4 pieces each and *usual* cakes with 3 pieces each, and the numbers of fat and usual cakes are

$$f = 2n - 3m \qquad \text{and} \qquad u = 4m - 2n$$

respectively. Since $u \geq 0$, we obtain $\frac{m}{n} > \frac{1}{2}$.

Next, each fat cake should be split into equal parts, these parts belong to $4f$ people, and the second piece at each such person weighs d. All the remaining people get both their pieces from the usual cakes; let us call these people *usual*. Then there are

$$s = n - 4f = 12m - 7n$$

usual people.

Now we will consider some auxiliary decomposition of **negative** "cakes"; it corresponds to the division of the usual cakes and people. Let us subtract $\frac{1}{4}$ from each piece of a fat cake, and d from each piece of a usual cake. Let us forget for a while about zero pieces. Then in a new decomposition all non-usual people and all fat cakes vanish (and we forget about them too), each usual "cake" now contains not more than three **negative** "pieces" of the same total weight, while each usual person gets at most two **nonpositive** "pieces" of the same total weight. So, taking the opposites of all the obtained weights, we arrive to the situation of 4.1a) (without a condition $m < n$).

Conversely, from the division of these *deficiencies* it is easy to pass to the distribution of the original cakes. Let us cut the real cakes into three pieces with the corresponding deficiencies (if a new "cake" is divided into one or two parts, then the remaining pieces should have zero deficiencies, i.e. they should be equal to d). For the people, if someone had got two deficiencies, then we give him the two corresponding pieces; otherwise he gets one corresponding piece together with the piece of zero deficiency. Finally, all the remaining pieces with zero deficiency are paired with the pieces from the fat cakes.

Thus, we obtain that the desired decomposition exists if and only if $\frac{u}{s} = \frac{4m-2n}{12m-7n} \in \left[\frac{1}{2}, 1\right) \cup \left\{\frac{k}{2k+1}\right\} \cup \left\{\frac{k+1}{k}\right\} \cup \{\infty\}$ (see the Remark after 4.1b); we have also added the degenerate case $s = 0$). With the use of $\frac{m}{n} < \frac{2}{3}$, this leads to the condition $\frac{m}{n} \in \left[\frac{5}{8}, \frac{2}{3}\right) \cup \left\{\frac{5k+2}{8k+4}\right\}_{k=1,2,\ldots}$.

Remark. In the further solutions following the same scheme we will omit repeating details, leaving them to the reader.

3.5. **Answer.** $\frac{m}{n} \in \left[\frac{2k-1}{k^2-1}, \frac{2}{k}\right) \cup \left\{\frac{d(2k-1)+2}{d(k^2-1)+k+1}\right\}_{d=1,2,\ldots}$.

The solution is analogous to 3.4b).

3.6. a) By 3.2b), $f(m,n) \le \frac{m}{n} - \frac{1}{3}$. The converse inequality is proved in part b).

b) **Answer.** $\frac{m}{n} \in \left(\frac{1}{2}, \frac{5}{9}\right] \cup \left\{\frac{5k+2}{9k+3}\right\}_{k=0,1,2,\ldots}$.

First, let us note that $f(m,n) \ge \frac{m}{3n} > \frac{m}{n} - \frac{1}{3}$ for $\frac{m}{n} < \frac{1}{2}$. On the other hand, $f(m,n) \le \frac{m}{2n} < \frac{m}{n} - \frac{1}{3}$ for $\frac{m}{n} > \frac{2}{3}$. So we are left to investigate the interval $\left(\frac{1}{2}, \frac{2}{3}\right)$ (the left end of the interval does not satisfy the condition, while the right end does).

We present an outline of the further solution which is similar to 3.4b). We obtain that each person gets two pieces, the sizes of pieces belong to a segment $\left[\frac{1}{3}, d\right]$ where $d = \frac{m}{n} - \frac{1}{3}$. Next, we have $f = 2n - 3m$ *fat* cakes with 4 pieces each and $u = 4m - 2n$ *usual* cakes with 3 pieces each. Each usual cake should be split into equal parts, these parts belong to $3u$ people, the second piece at each such person weighs d, and all the other $s = n - 3u = 7n - 12m$ *usual* people get both pieces from fat cakes. Notice that $\frac{m}{n} > \frac{1}{2}$ implies $\frac{f}{s} \ge \frac{1}{2}$.

Now, subtracting $\frac{1}{3}$ from each piece of a usual cake and d from each other piece we obtain the distribution of the remaining f **nonnegative** "cakes" over s usual people. The only condition remained is that each person should get not more than two pieces, while each "cake" should contain at most four pieces; this condition being satisfied, we can recover the division of the usual cakes. Hence by 4.1b) (together with the remark after it), for $\frac{f}{s} > \frac{1}{2}$ the desired division exists if and only if $\frac{f}{s} \in \left[\frac{1}{2}, 1\right] \cup \left\{\frac{v+1}{v}\right\} \cup \{\infty\}$. The answer follows.

3.7. **Answer.** $\frac{m}{n} \in \left(\frac{2}{k+1}, \frac{2k-1}{k^2}\right] \cup \left\{\frac{d(2k-1)+2}{dk^2+k}\right\}_{d=1,2,\ldots}$.

The solution is analogous to 3.6b).

3.8. A particular case of 3.5.

The ideas of solution of problem 3.9 is presented in the next section.

Different cakes

Recall that $\lfloor x \rfloor$ and $\lceil x \rceil$ are the largest integer not exceeding x and the smallest integer not less than x, respectively.

4.2. a) **Answer.** $\frac{3}{N}$, if $3 \mid N$; $\max\left\{\frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}, \frac{1}{\lceil 2N/3 \rceil}\right\}$ otherwise.

If $3 \mid N$, then obviously the optimal way is to cut the cakes into the pieces of $\frac{3}{N}$ each.

Now assume that $3 \nmid N$. Then some person should get at least two pieces, hence the answer does not exceed $\frac{3}{2N}$, and we may assume that each person gets at least two pieces.

Next, the smaller cake is divided either into $\le \lfloor 2N/3 \rfloor$ parts, or into $\ge \lceil 2N/3 \rceil$ parts (for $N = 2$, the second case necessarily holds). In the first case, one of these pieces should be $\ge \frac{1}{\lfloor 2N/3 \rfloor}$, so its complement (at a person) is $\le \frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}$. In the second case, the smaller cake contains a piece which is $\le \frac{1}{\lceil 2N/3 \rceil}$. So, in any case the minimal weight does not exceed one of the numbers $\frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}$ and $\frac{1}{\lceil 2N/3 \rceil}$, i.e. it does not exceed $D = \max\left\{\frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}, \frac{1}{\lceil 2N/3 \rceil}\right\}$.

We are left to present an example with D as the minimal piece weight. Assume that $D = \frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}$. Let us divide the smaller cake into the pieces of $\frac{1}{\lfloor 2N/3 \rfloor} \ge D$, cut away the same number of pieces of D each from the larger cake, and divide the rest into the whole portions. Obviously, this division fits. In the second case, the example is constructed analogously.

Remark. One may check that $D = \frac{3}{N} - \frac{1}{\lfloor 2N/3 \rfloor}$ if $N = 3k + 2$, and $D = \frac{1}{\lceil 2N/3 \rceil}$ otherwise.

b) **Answer.** $\frac{7}{N}$, if $7 \mid N$; $\max\left\{\frac{7}{N} - \frac{2}{\lfloor 4N/7 \rfloor}, \frac{2}{\lceil 4N/7 \rceil}\right\}$ otherwise.

The solution is completely analogous and is left to the reader.

The next problem gives a hint of how does the general algorithm work. We need to introduce some

Definitions and notation. Recall that a *hypergraph* is a pair (V, E) where V is the set of *vertices*, and E is the set of *(hyper)edges* which are some nonempty subsets of V. A hypergraph is *homogeneous* if all its edges have the same cardinality. For any hypergraph $G = (V, E)$ we can construct its *underlying graph* $G' = (V, E')$ with the same set of vertices, connecting every two vertices belonging to one hyperedge of G. A hypergraph is *connected* if its underlying graph is connected.

Further we will denote by $[b : c]$ the following situation: we have a cake of weight b which should be divided into c parts. So, the notation $2 \times [4 : 3] + 3 \times [7 : 4]$ will denote the collection of two cakes of weight 4 which should be divided into three parts each together with three cakes of weight 7 which should be divided into four parts each.

4.4. a) **Answer.** $\frac{49}{6}$.

In our notation, we have $[59 : 4] + [89 : 6] + 2 \times [41 : 5]$. One of the pieces in $[89 : 6]$ is at least $\frac{89}{6}$, and its complement is $\le \frac{49}{6}$, as desired. It remains to provide an example:

$$\left(4 \cdot \frac{59}{4}\right) + \left(6 \cdot \frac{89}{6}\right) + 2 \times \left(3 \cdot \frac{49}{6} + 2 \cdot \frac{33}{4}\right) = 4 \times \left(\frac{59}{4} + \frac{33}{4}\right) + 6 \times \left(\frac{89}{6} + \frac{49}{6}\right).$$

b) **Answer.** $\frac{49}{6}$.

In our notation, we have $2 \times [41 : 5] + 3 \times [35 : 4] + 11 \times [29 : 2]$. We say that the cakes of weight 29 are *small*, and the others are *large*. Notice that each person should get two pieces of total weight 23. Assume that each piece weighs at least than $\frac{49}{6}$; then each piece should also be at most than $23 - \frac{49}{6} = \frac{89}{6}$.

Assume that a person gets two pieces from small cakes (surely these two cakes are distinct), then the average weight of the remaining two pieces in these cakes is $\frac{2 \cdot 29 - 23}{2} > \frac{89}{6}$ which is impossible.

Hence all 22 pieces of small cakes come to different people, and therefore all the pieces from the large cakes also come to different people.

Now let us call the cakes of weight 41 *fat*, and the cakes of weight 35 *usual*. Construct a hypergraph with small cakes as vertices; each edge will correspond to a usual cake and consist of all the small cakes containing the people's complements of the pieces of this usual cake. This hypergraph contains at least two connected components.

Now let us remove all the pieces of the usual cakes, as well as their complements in small cakes. Next, we glue the remaining pieces of each connected component into one new "cake". Let us calculate a number of pieces and a weight of this "cake".

Assume that a component contains v vertices and e edges. Due to each edge, we have removed 4 pieces of total weight $4 \cdot 23 - 35 = 57$; hence the number of the pieces removed from our component is $4e$, while their total weight is $57e$. Thus the average weight of the remaining pieces is $\frac{29v-57e}{2v-4e}$ which should be $\le \frac{89}{6}$, which rewrites as $2v \ge 7e$. On the other hand, since the component is connected, we have $v \le 3e+1$. The two obtained inequalities hold only if the pair (v,e) is either $(4,1)$ or $(7,2)$. Hence our hypergraph should contain one component of type $(4,1)$ and one of type $(7,2)$. In the latter component, one of the pieces will be at least $\frac{7\cdot 29-2\cdot 57}{14-8} = \frac{89}{6}$ which provides the desired estimate.

But we can also get the example from this construction! Namely, from the component of type $(4,1)$ we have obtained a "cake" of 4 pieces and total weight $4\cdot 29-57 = 59$, while from the remaining component we get a "cake" of 6 pieces and total weight $7 \cdot 29 - 2 \cdot 57 = 89$. Also we have $2 \times [41:5]$ remained. Thus we come to the situation of 4.4a), so we may take the division from that example and then find the weights of the removed pieces. The resulting example is

$$11 \times 29 + 2 \times 41 + 3 \times 35 = 4 \times \left(\frac{59}{4} + \frac{57}{4}\right) + 6 \times \left(\frac{89}{6} + \frac{85}{6}\right) + \left(2 \cdot \frac{29}{2}\right) +$$
$$+ 2 \times \left(3 \cdot \frac{49}{6} + 2 \cdot \frac{33}{4}\right) + \left(4 \cdot \frac{35}{4}\right) + 2 \times \left(3 \cdot \frac{53}{6} + \frac{17}{2}\right) =$$
$$= 4 \times \left(\frac{59}{4} + \frac{33}{4}\right) + 6 \times \left(\frac{89}{6} + \frac{49}{6}\right) + 4 \times \left(\frac{57}{4} + \frac{35}{4}\right) + 6 \times \left(\frac{85}{6} + \frac{53}{6}\right) + 2 \times \left(\frac{29}{2} + \frac{17}{2}\right).$$

c) **Answer.** $\frac{49}{138} = \frac{1}{23} \cdot \frac{49}{6}$ (could you guess it?).

Let us multiply all the weights by 29. As usual, we may assume that each person gets exactly two pieces, and each cake is divided into either two or three parts. Then the numbers of cakes of both types can be found, and we arrive to the situation $12 \times [29:3] + 11 \times [29:2]$. We say that the cakes with three pieces are *fat*, and the others are *usual*. Assume that each piece weighs at least than $\frac{49}{6}$; then each piece should also be at most than $23 - \frac{49}{6} = \frac{89}{6}$.

By the same reasons as above, none of the people gets two pieces from a usual cake. Hence all 22 pieces of usual cakes come to different people, and their complements belong to fat cakes. The remaining 14 pieces of fat cakes come to 7 remaining people; let us call these people *fat*. Now construct a graph with fat cakes as vertices; each edge will correspond to a fat person and connect two fat cakes containing the cakes containing the pieces of this person. This graph contains at least five connected components.

Now let us remove all the pieces of fat people. Next, we glue the remaining pieces of each connected component into one new "cake". Let us calculate a number of pieces and a weight of this "cake".

Assume that a component contains v vertices and e edges (then $v \le e+1$). Due to each edge, we have removed 2 pieces of total weight 23; hence the number of the pieces removed from our component is $2e$, while their total weight is $23e$. Thus the average weight of the remaining pieces is $\frac{29v-23e}{3v-2e}$ which should be $\ge \frac{49}{6}$, which rewrites as $27v \ge 40e$. This is impossible if $v \le e$, so we get $v = e+1$ and hence $27 \le 13e$, or $e \le 2$. Thus, each component is a tree (so there are exactly five of them) and has at most two edges.

The most "regular" case is when there are two components with two edges and three components with one edge; so the obtained new "cakes" will look as $2 \times [41:5] + 3 \times [35:4]$. So by 4.4b) the answer will be at most $\frac{49}{6}$.

In any other case, an isolated vertex appears; this means that all three pieces of this cake are paired up (in portions) with the pieces from usual cakes. Consider these three complements, and take three usual cakes containing them. The average of three remaining pieces of these cakes is $\frac{4\cdot 29-3\cdot 23}{3} > \frac{89}{6}$, which is impossible. Hence the estimate is established.

The example again can be obtained from the example for 4.4b) by filling up the removed pieces:

$$\begin{aligned}11\times 29+12\times 29 = 4\times\left(\frac{59}{4}+\frac{57}{4}\right)+6\times\left(\frac{89}{6}+\frac{85}{6}\right)+\left(2\cdot\frac{29}{2}\right)+\\ +4\times\left(\frac{49}{6}+\frac{33}{4}+\frac{151}{12}\right)+2\times\left(\frac{49}{6}+2\cdot\frac{125}{12}\right)+\\ +2\times\left(2\cdot\frac{35}{4}+\frac{23}{2}\right)+2\times\left(2\cdot\frac{53}{6}+\frac{34}{3}\right)+2\times\left(\frac{53}{6}+\frac{17}{2}+\frac{35}{3}\right)=\\ =4\times\left(\frac{59}{4}+\frac{33}{4}\right)+6\times\left(\frac{89}{6}+\frac{49}{6}\right)+4\times\left(\frac{57}{4}+\frac{35}{4}\right)+6\times\left(\frac{85}{6}+\frac{53}{6}\right)+\\ +2\times\left(\frac{29}{2}+\frac{17}{2}\right)+4\times\left(\frac{151}{12}+\frac{125}{12}\right)+\left(2\cdot\frac{23}{2}\right)+2\times\left(\frac{34}{3}+\frac{35}{3}\right).\end{aligned}$$

3.9. a) **Answer.** $\frac{5}{4}\cdot\frac{m}{n}-\frac{1}{2}$.

First, let us prove the upper bound. As usual, we may assume that each person gets two pieces, all the cakes are divided into 3 or 4 parts, hence there are $u = 4m-2n$ *usual* cakes of three parts and $f = 2n-3m$ *fat* cakes of four parts. Since $4f < n$ (provided by $\frac{m}{n} > \frac{7}{12}$), some person should get both pieces from the usual cakes. Consider two cakes containing these pieces; the average weight of the rest four pieces in them will be $t = \frac{1}{4}\left(2-\frac{m}{n}\right)$, so one of these pieces weighs at least t. So its complement weighs at most $d = \frac{m}{n}-t = \frac{5}{4}\cdot\frac{m}{n}-\frac{1}{2}$, as desired.

The example will follow from part b).

b) We investigate only the case $\frac{m}{n}\in\left(\frac{7}{12},\frac{2}{3}\right)$ where the upper bound from part a) holds. We claim that on this interval $f(m,n) = \frac{5}{4}\cdot\frac{m}{n}-\frac{1}{2}$ if and only if $\frac{m}{n}\in\left(\frac{7}{12},\frac{22}{37}\right]\cup\left\{\frac{22d-3}{37d-2}\right\}_{d=1,2,\ldots}$; here is an outline of the proof.

We multiply all the weights by $4n$; so the weight of the minimal piece should be $d = 5m-2n$, while the maximal weight should be $t = 2n-m$.

In our case, each person should get two pieces, there are $f = 2n-3m$ *fat* cakes with four pieces each and $u = 4m-2n$ *usual* cakes with three pieces each. Next, it is easy to see that no person gets two pieces from the fat cakes, so there are $4f$ persons getting a piece from the fat cake and $s = n-4f$ *usual* persons with two pieces from usual cakes.

Construct a graph G having usual cakes as vertices, with the edges formed by the two pieces of a usual person. Remove all the usual people's pieces and glue each component into a new cake. If some connected component contains more than one edge, then deleting all the usual people's pieces from this component we get some pieces of average weight $> d$ which is impossible. Hence we get s new cakes of weight $8n-4m = 4t$ consisting of four pieces and $u-2s$ old cakes of weight $4n$ consisting of three pieces. Notice that new cakes should be divided into four equal parts each.

Now we act as in 3.4b). Subtract d from each remaining piece of usual cake, and t from each piece of a fat cake. Then the new cakes vanish, all remaining usual cakes turn into positive "cakes" of three (or less) pieces, and all fat cakes turn into negative "cakes" of four (or less) pieces. Taking the absolute values of negative pieces, we come to the following situation:

We have $u-2s$ equal cakes, and we need to cut each of them into at most three parts and redistribute into f groups of equal weight having at most four pieces each.

Moreover, one can see that if the desired division of the "cakes" is possible then one may recover the division of the initial cakes. Hence it remains to determine when the new problem has a solution. This can be made in the same way as in 4.1.

General algorithm

Finally, we show how the general algorithm works on some nontrivial example — that is, we will find $f(31, 52)$.

1.3. e) Unlike in the other problems, we do not start with an answer, but we wish to see how to *find* it from the very beginning.

Part I. Firstly, we will perform some strange process which provides neither an example nor the bound. But we will get an answer; and then we will check that this answer is achievable and optimal.

During the process, we will make some assumptions on how should the optimal example look like. So, after the example is constructed we will need to check that lacking these assumptions we will obtain a worse division. We mark these assumprions by a bold font and number them consecutively.

Preliminaries. Multiply all the weights by 52. We will **assume(1)** that each person has two pieces, and since $\frac{1}{2} < \frac{m}{n} < \frac{2}{3}$ we will **assume(2)** that each cake is divided into three of four parts. Then we have

$$11 \times [52 : 4] + 20 \times [52 : 3].$$

Let us call the cakes with four pieces *fat*, and the other cakes *usual*.

Initial step. Now we have 44 pieces in fat cakes, which is smaller than the number of people. We **assume(3)** that all of them come to different persons. Hence there will be exactly 8 *usual* persons with both pieces in usual cakes. So we construct a graph on usual cakes as vertices, where each usual person induces an edge. This graph contains 20 vertices and 8 edges, so it has at least 12 connected components.

In such a situation we **assume(4)** that (i) all components are trees (so there are exactly 12 of them), and (ii) the edges are distributed between the components almost uniformly (that is, the numbers of edges in any two components differ by at most 1). In our case, this means that there are 8 components with one edge and 4 isolated vertices. Now, removing the pieces belonging to usual people and gluing the pieces of one component, we come to a situation

$$11 \times [52 : 4] + 8 \times [73 : 4] + 4 \times [52 : 3].$$

Regular step 1. Now we have 44 *small* pieces in 11 fat cakes, and 44 *large* pieces in remaining cakes; each person should get one piece of each type. Notice that the average weight of a piece in $[52 : 3]$ is smaller than that in $[73 : 4]$. Informally speaking, this means that we need to cut the latter cakes as uniformly as we can. So we postpone them and deal with the remaining ones.

Consider a hypergraph on the fat cakes as vertices, with each $[52 : 3]$ cake inducing an edge (this edge consists of the fat cakes containing the complements of the pieces of our $[52 : 3]$ cake). Thus we have a hypergraph on 11 vertices with 4 edges of cardinality ≤ 3. Such hypergraph has at least $11 - 4 \cdot (3 - 1) = 3$ connected components.

As before, we **assume(5)** that (i) each component has the maximal possible number of vertices for its number of edges (so there are exactly three of them), and (ii) the edges are distributed between the components almost uniformly (that is, the numbers of edges in any two components differ by at most 1). In our case, this means that there are two components with three vertices and one edge, as well as one component with two edges and five vertices. Now, removing all the pieces of $[52 : 3]$ cakes together with their complements, and gluing the pieces of one component, we come to a situation

$$2 \times [105 : 9] + [178 : 14] + 8 \times [73 : 4].$$

Regular step 2. Now we have 32 *large* pieces in $[73 : 4]$ cakes, and 32 *small* pieces in remaining cakes; each person should get one piece of each type. Notice that the average weight of a piece in $[105 : 9]$ is larger than that in $[178 : 14]$. Again, this means that we need to cut the latter cakes as uniformly as we can, so we deal with $[105 : 9]$ cakes.

Consider a hypergraph on the $[73 : 4]$ cakes as vertices, with each $[105 : 9]$ cake inducing an edge. Unlike the previous cases, this hypergraph may be connected; so we **assume(6)** that it is connected, and we are ready to finish. We make this graph connected, hence by the standard removing we arrive to the situation

$$[178 : 14] + [256 : 14].$$

But this situation is trivial, and the maximal possible smallest piece is $\frac{178}{14}$: it is enough to divide each cake into equal parts and give to each person one piece from each cake. Notice that our last aim (to cut $[178:14]$ with the maximal uniformity) is completely reached.

Thus, under all our assumptions we obtain that the minimal piece is at most $d = \frac{89}{7}$.

Part II. Now we are to construct an example, moving backwards in our process. Recall that in our example we have

$$[178:14] + [256:14] = \left(14 \cdot \frac{89}{7}\right) + \left(14 \cdot \frac{128}{7}\right).$$

Regular step 2. The cake $[256:14]$ is obtained from $8 \times [73:4]$ by removing the pieces complementary to the ones from $2 \times [105:9]$. Now we reconstruct the division of these cakes — for instance, with the help of intervals; now it is easy to apply:

$$8 \times [73:4] + 2 \times [105:9] = 8 \times \left(3 \cdot \frac{128}{7} + \frac{127}{7}\right) + 2 \cdot \left(5 \cdot \frac{89}{7} + 4 \cdot \frac{90}{7}\right).$$

Regular step 1. The cakes $2 \times [105:9]$ were obtained from $3 \times [52:4] + 3 \times [52:4]$ by removing the complements of the pieces from $[52:3] + [52:3]$; analogously, the cake $[178:14]$ was obtained from another $4 \times [52:3]$ by removing the complements of the pieces from $2 \times [52:3]$. Now we will reconstruct them; it is immediate after we split the pieces in $[105:9]$ and $[178:14]$ into the $[52:4]$ cakes they were taken from. Making this arbitrarily we get

$$\begin{aligned} 11 \times [52:4] + 4 \times [52:3] = {} & 4 \times \left(2 \cdot \frac{89}{7} + \frac{90}{7} + \frac{96}{7}\right) + 2 \times \left(\frac{89}{7} + 2 \cdot \frac{90}{7} + \frac{95}{7}\right) + \\ & + 4 \times \left(3 \cdot \frac{89}{7} + \frac{97}{7}\right) + 2 \times \left(2 \cdot \frac{89}{7} + 2 \cdot \frac{93}{7}\right) + \\ & + 2 \times \left(2 \cdot \frac{121}{7} + \frac{122}{7}\right) + 2 \times \left(2 \cdot \frac{120}{7} + \frac{124}{7}\right). \end{aligned}$$

Initial step. We are left to reconstruct the last cakes $16 \times [52:3]$ from $8 \times [73:4]$ by adding the usual people; it also goes automatically:

$$16 \times [52:3] = 8 \times \left(2 \cdot \frac{128}{7} + \frac{108}{7}\right) + 8 \times \left(\frac{128}{7} + \frac{127}{7} + \frac{109}{7}\right).$$

So the whole example is reconstructed:

$$\begin{aligned} 11 \times [52:4] + 20 \times [52:3] = {} & 4 \times \left(2 \cdot \frac{89}{7} + \frac{90}{7} + \frac{96}{7}\right) + 2 \times \left(\frac{89}{7} + 2 \cdot \frac{90}{7} + \frac{95}{7}\right) + \\ & + 4 \times \left(3 \cdot \frac{89}{7} + \frac{97}{7}\right) + 2 \times \left(2 \cdot \frac{89}{7} + 2 \cdot \frac{93}{7}\right) + \\ & + 2 \times \left(2 \cdot \frac{121}{7} + \frac{122}{7}\right) + 2 \times \left(2 \cdot \frac{120}{7} + \frac{124}{7}\right) + \\ & + 8 \times \left(2 \cdot \frac{128}{7} + \frac{108}{7}\right) + 8 \times \left(\frac{128}{7} + \frac{127}{7} + \frac{109}{7}\right). \end{aligned}$$

Part III. We are left to check that all our assumptions were necessary. It can be done easily within the methods involved in the previous sections. Denote $d = \frac{89}{6}$, $t = 31 - d = \frac{128}{7}$. If $f(31, 52) > d$ then all the piece weights should belong to the interval (d, t).

Assumption (1) should hold since otherwise the least piece is at most $\frac{31}{3} < d$.

Assumption (2) is necessary since otherwise we get a piece of $\leq \frac{52}{5} < d$ or $\geq 31 - \frac{52}{2} > t$.

Assumption (3) should hold, otherwise let us consider a person with two pieces in two fat cakes. Then the average weight of the remaining six pieces in these cakes is $\frac{52 \cdot 2 - 31}{6} < d$.

Assumption (4): suppose it fails; then two edges have a common vertex. Consider three cakes participating in these edges; the average of 5 their pieces not corresponding to our edges is $\frac{3 \cdot 52 - 2 \cdot 31}{5} > t$.

Assumption (5) may be checked in the same way as in 4.4b).

Finally, we do not need to check the last **Assumption (6)** at all: to this step, we already have a $[178 : 14]$ cake, so the minimal piece cannot exceed $\frac{178}{14} = d$.

We are done!

We suggest you to apply this algorithm to the pairs from the Testing area to see how it works!

Fair cake division

Solutions after semifinal

Some general results

We start with problems 3.1–3.3, since their solutions allow to simplify the further text.

3.1. If $m \nmid n$, then it is impossible to split a cake into equal pieces which weigh $\frac{m}{n}$. So there will be some piece less than $\frac{m}{n}$. That means that some person will receive at least two pieces, and one of them weigh at most $\frac{m}{2n}$ by the pigeonhole principle.

If $f(m,n) = \frac{m}{2n}$, then every piece weighs $\frac{m}{2n}$ or $\frac{m}{n}$ (if its weight is between these numbers, then the person receiving it has all other pieces smaller than $\frac{m}{2n}$). We may split all the pieces of $\frac{m}{n}$ into two pieces which weigh $\frac{m}{2n}$ which may happen if and only if $m \mid 2n$ (while $m \nmid n$).

3.3. a) Consider any optimal distribution. If someone gets at least three pieces, then the minimal of them is at most $\frac{1}{3} \cdot \frac{m}{n}$; hence $f(m,n) \le \frac{m}{3n}$.

3.2. a) Suppose that $f(m,n) > \frac{m}{n} - \frac{1}{2}$. For $\frac{m}{n} > \frac{3}{4}$ we have $\frac{m}{n} - \frac{1}{2} > \frac{m}{3n}$. So by 3.3a) every person has not more than two pieces. If there is a piece of weight $\frac{m}{n}$, then we cut it into two equal pieces and, and we still have the optimal decomposition of cakes because of 3.1. So the total of pieces in decomposition is $2n$. Since $2n < 3m$, then by the pigeonhole principle there exists a cake with not more than two pieces. A cake with one piece is impossible, so we have a cake with two pieces, one of which is at least $\frac{1}{2}$. A person who has this piece also gets one more piece, the weight of which is not more than $\frac{m}{n} - \frac{1}{2}$. A contradiction.

b), c) Let $k \ge 3$. Suppose that $\frac{2}{k+1} < \frac{m}{n} < 1$ but $f(m,n) > \frac{m}{n} - \frac{1}{k}$. Since $\frac{m}{n} - \frac{1}{k} > \frac{m}{3n}$ for $\frac{m}{n} > \frac{2}{k+1}$, then by 3.1 and 3.3a) we may assume that each person has got two pieces and so the total number of pieces is $2n$. Since $2n < (m+1)k$, then there is a cake with not more than k pieces, one of which by the pigeonhole principle is at least $\frac{1}{k}$. A person with this piece has one more piece with weight not more than $\frac{m}{n} - \frac{1}{k}$. A contradiction.

3.3. b) Consider a segment of length m. Divide it by red points into m equal segments, and by blue points into $3n$ equal segments (the endpoints will be multicolored). The obtained segments with the red endpoints represent the cakes.

Now we simultaneously erase all blue points adjacent to a red point which is not blue. Next, cut the cakes by the remaining blue points. We claim that we have obtained a desired division.

Since $m < n$, each cake contained at least three blue points, hence at least one of them is not erased. Next, each piece not adjacent to the border of a cake is of length $\frac{m}{3n}$. The same holds for the segments having a multicolored endpoint. All the other segments are grouped into pairs sharing a common red endpoint, and the total length of each pair is $3 \cdot \frac{m}{3n} = 1$. Hence we may give each such pair to one man. All the remaining pieces are $\frac{m}{3n}$ in length, so we can give to every remaining man three such pieces.

c) Suppose, to the contrary, that $\frac{m}{n} \in \left(\frac{2}{3}, \frac{3}{4}\right]$ but $f(m,n) > \frac{1}{3}$. By 3.1 and 3.3a), we may assume that each man gets exactly two pieces. Then the total number of pieces is $2n < 3m$, hence by the

pigeonhole principle there exists a cake consisting of no more than two pieces. Hence one of these pieces is at least $\frac{1}{2}$. Finally, the man getting this piece should get some other piece of weight at most $\frac{m}{n} - \frac{1}{2}$ which does not exceed $\frac{m}{3n}$ since $\frac{m}{n} \leq \frac{3}{4}$. A contradiction.

Finally, we present the solution of 1.4.

1.4. Consider a set of pieces such that we may arrange it into n equal cakes with weight m, as well as into m equal cakes with weight n. Suppose that x is the maximal weight of the minimal piece of sets with given property. Then

$$nf(m,n) = x = mf(n,m),$$

and $f(m,n) = \frac{m}{n} f(n,m)$.

Some concrete values of f

Notation. In the examples we shall denote by $(i_1 \cdot a_1 + i_2 \cdot a_2 + \cdots + i_l \cdot a_l)$ the decomposition of a cake into $i_1 + i_2 + \cdots + i_l$ pieces, among which i_1 pieces weigh a_1, i_2 pieces weigh a_2, etc. The composition of a man's portion will be denoted in the same manner.

1.1. a) A particular case of 3.3c).

b) The bound follows from 3.2b). Example:

$$4 \times 210g = 2 \times (3 \cdot 70g) + 2 \times (3 \cdot 50g + 60g)$$

c) The bound follows from 3.2c) for $k = 12$. Example:

$$4 \times 3kg = 2 \times (12 \cdot 250g) + 2 \times (240g + 12 \cdot 230g) = 24 \times (230g + 250g) + (2 \cdot 240g) = 25 \times 480g.$$

1.2. a) **Answer.** $\frac{5}{21}$.

A particular case of 3.3c).

b) **Answer.** $\frac{5}{18}$.

The bound follows from 3.2a). Example:

$$7 \times 1 = 3 \times \left(2 \cdot \frac{1}{2}\right) + 2 \times \left(\frac{5}{18} + \frac{6}{18} + \frac{7}{18}\right) + 2 \times \left(2 \times \frac{5}{18} + \frac{8}{18}\right) =$$
$$= 6 \times \left(\frac{5}{18} + \frac{1}{2}\right) + \left(2 \cdot \frac{7}{18}\right) + 2 \times \left(\frac{6}{18} + \frac{8}{18}\right) = 9 \times \frac{7}{9}.$$

1.3. a) **Answer.** $\frac{1}{3}$.

We may assume that every person has two pieces, so there is a cake with three (or more) pieces, one of which is not more than $\frac{1}{3}$, and $f(8,9) \leq \frac{1}{3}$.

Example:

$$8 \times 1 = 2 \times \left(3 \cdot \frac{1}{3}\right) + 6 \times \left(\frac{4}{9} + \frac{5}{9}\right) = 6 \times \left(\frac{1}{3} + \frac{5}{9}\right) + 3 \times \left(2 \cdot \frac{4}{9}\right) = 9 \times \frac{8}{9}.$$

b) **Answer.** $\frac{2}{7}$.

The bound $f(11,14) \geq \frac{2}{7}$ follows from 3.2a). Example:

$$11 \times 1 = 5 \times \left(2 \cdot \frac{1}{2}\right) + 4 \times \left(2 \cdot \frac{2}{7} + \frac{3}{7}\right) + 2 \times \left(\frac{2}{7} + 2 \cdot \frac{5}{14}\right) =$$
$$= 10 \times \left(\frac{1}{2} + \frac{2}{7}\right) + 4 \times \left(\frac{3}{7} + \frac{5}{14}\right) = 14 \times \frac{11}{14}.$$

c) **Answer.** $\frac{5}{17}$.

As usual, we may assume that each person has got two pieces. Then there are 34 pieces in total. Note that each cake contains two or three pieces: a cake could not contain only one piece, and if it contains at least 4 pieces, then one of them is not more than $\frac{1}{4} < \frac{5}{17}$. So we have 6 *fat* cakes with three pieces and 8 *usual* cakes with two pieces. There are 18 pieces in fat cakes and 16 pieces in usual ones. So there is a person getting both his pieces from the fat cakes (if it is the same cake, we choose a second fat cake arbitrarily). The rest four pieces of these two cakes weigh in total $2 - \frac{14}{17} = \frac{20}{17}$. So one of them is not more than $\frac{5}{17}$.

Example:

$$14 \times 1 = 8 \times \left(\frac{9}{17} + \frac{8}{17}\right) + 4 \times \left(2 \cdot \frac{6}{17} + \frac{5}{17}\right) + 2 \times \left(\frac{7}{17} + 2 \cdot \frac{5}{17}\right) =$$
$$= 8 \times \left(\frac{5}{17} + \frac{9}{17}\right) + 8 \times \left(\frac{6}{17} + \frac{8}{17}\right) + \left(2 \cdot \frac{7}{17}\right) = 17 \times \frac{14}{17}.$$

2.1. a) **Answer.** $\frac{1}{3}$.

Denote $s = f(3k-1, 3k)$. Firstly we prove that $s \le \frac{1}{3}$. Assume the contrary. Then $s > \frac{1}{3}$, hence by 3.1 and 3.3a) each man gets two pieces, and as usual we may assume that he gets exactly to pieces. Hence there exists a cake split into at least three parts, and the minimal of them is at most $\frac{1}{3}$. A contradiction.

To construct an example, let us multiply all the weights by $3k$. We have

$$(3k-1) \times 3k = 2 \times \big(3 \cdot k\big) +$$
$$+ 3 \times \big((2k-1) + (k+1)\big) + 3 \times \big((2k-2) + (k+2)\big) + \cdots + 3 \times \big((k+1) + (2k-1)\big) =$$
$$= 3 \times \big(k + (2k-1)\big) + 3 \times \big((k+1) + (2k-2)\big) + \cdots + 3 \times \big((2k-1) + k = (3k+2) \times (6k+2).$$

b) **Answer.** $\frac{2k+1}{2(3k+2)}$.

Denote $s = f(3k+1, 3k+2)$, $t = \frac{2k+1}{2(3k+2)}$. Firstly, let us prove that $s \le t$. Assume the contrary. As in the previous problem, by $s > t \ge \frac{3k+1}{3(3k+2)}$ we may assume that each man gets exactly two pieces. If some cake is divided into at least four parts then the least of them is at most $\frac{1}{4} \le \frac{2k+1}{2(3k+2)}$ which is impossible. Hence each cake is split into two or three pieces, and it is easy to see that there are exactly two cakes split into three parts.

Now consider the following graph. Its vertices are the cakes, and to each man corresponds an edge connecting two cakes the pieces of this man are taken from. This graph has two *distinguished* vertices of degree 3 and all other vertices of degree 2, hence it consists of three paths connecting distinguished vertices (some of them may be circuits). The length of one of these paths is at least $k+1$; consider this path $v_0, v_1, \ldots, v_{k+1}$.

Denote the portion of the man corresponding to the edge (v_i, v_{i+1}) as (p_i, q_i) so that p_i is taken from cake v_i, and q_i is taken from v_{i+1}. Then $p_i + q_i = \frac{3k+1}{3k+2}$ for $i = 0, 1, \ldots, k$, while $q_i + p_{i+1} = 1$ for $i = 1, 2, \ldots, k$. Hence $p_{i+1} - p_i = \frac{1}{3k+2}$, and hence $p_k \ge \frac{k}{3k+2} + p_0 \ge \frac{k}{3k+2} + t = \frac{4k+1}{2(3k+2)}$. Finally, we get $q_k = \frac{3k+1}{3k+2} - p_k \le t$. A contradiction.

This proof also provides a hint of how to construct an example: there should be two paths of length $k+1$ and one path of length k. To construct such an example, let us multiply all the weights by $2(3k+2)$. We have

$$(3k+1) \times (6k+4) = 2 \times \big(2 \cdot (2k+1) + (2k+2)\big) +$$
$$+ 2 \times \big((4k+1) + (2k+3)\big) + 2 \times \big((4k-1) + (2k+5)\big) + \cdots + 2 \times \big((2k+3) + (4k+1)\big) +$$
$$+ \big(4k + (2k+4)\big) + \big((4k-2) + (2k+6)\big) + \cdots + \big((2k+4) + 4k\big) =$$
$$= 2 \times \big((2k+1) + (4k+1)\big) + 2 \times \big((2k+3) + (4k-1)\big) + \cdots + 2 \times \big((4k+1) + (2k+1)\big) +$$
$$+ \big((2k+2) + 4k\big) + \big((2k+4) + (4k-2)\big) + \cdots + \big(4k + (2k+2)\big) = (3k+2) \times (6k+2).$$

c) **Answer.** $\frac{k}{3k+1}$.

Denote $s = f(3k, 3k+1)$, $t = \frac{k}{3k+1}$. Again, supposing that $s > t$, we may assume that each man has two pieces, there are exactly two cakes split into three parts, and each other cake consists of two parts. Next, constructing a graph as above, we get that it contains a path of length at least $k+1$. Again, acting as above we find that $p_k - p_0 = \frac{k}{3k+1}$, so $p_k \geq \frac{2k}{3k+1}$, and $q_k = \frac{3k}{3k+1} - p_k \leq t$. A contradiction.

To construct an example, let us multiply all the weights by $3k+1$. We have

$$\begin{aligned}
3k\times(3k+1) &= 2 \times \big(2\cdot k + (k+1)\big) + \\
&\quad + 2 \times \big(2k + (k+1)\big) + 2 \times \big((2k-1) + (k+2)\big) + \cdots + 2 \times \big((k+1) + 2k\big) + \\
&\quad + \big((2k-1) + (k+2)\big) + \big((2k-2) + (k+3)\big) + \cdots + \big((k+2) + (2k-1)\big) = \\
&= 2\times\big(k + 2k\big) + 2 \times \big((k+1) + (2k-1)\big) + \cdots + 2 \times \big(2k + k\big) + \\
&\quad + \big((k+1) + (2k-1)\big) + \big((k+2) + (2k-2)\big) + \cdots + \big((2k-1) + (k+1)\big) = (3k+1) \times 3k.
\end{aligned}$$

2.3. a) There are three cases, depending on the residue of n modulo 3.

1) $n = 3k$. Obviously, $f(3, 3k) = \frac{1}{k}$.

2) $n = 3k+1$. If $n = 1$, then $f(3,1) = 1$. For $n = 3k+1 \geq 4$ we have $f(3, 3k+1) = \frac{3k-1}{2k(3k+1)}$. The bound follows from 3.2c) for $k' = 2k$. Example:

$$\begin{aligned}
3 \times 1 &= \left(2k \cdot \frac{1}{2k}\right) + 2 \times \left(k \cdot \frac{3k-1}{2k(3k+1)} + (k+1) \cdot \frac{3}{6k+2}\right) = \\
&= 2k \times \left(\frac{3k-1}{2k(3k+1)} + \frac{1}{2k}\right) + (k+1) \times \left(2 \cdot \frac{3}{6k+2}\right).
\end{aligned}$$

3) $n = 3k+2$. Here we have $f(3, 3k+2) = \frac{1}{2k+2}$.

We may assume that each person has got two pieces and that the total number of pieces is $6k+4$. Then there is a cake with at least $2k+2$ pieces, one of them is not more than $\frac{1}{2k+2}$. Example:

$$\begin{aligned}
3 \times 1 &= \left((2k+2) \cdot \frac{1}{2k+2}\right) + 2 \times \left((k+1) \cdot \frac{3k+4}{(3k+2)(2k+2)} + k \cdot \frac{3}{6k+4}\right) = \\
&= (2k+2) \times \left(\frac{1}{2k+2} + \frac{3k+4}{(3k+2)(2k+2)}\right) + k \times \left(2 \cdot \frac{3}{6k+4}\right).
\end{aligned}$$

Равновесия Нэша

Владимир Гурвич

1. Предварительное обсуждение

Что общего у шашек, шахмат, го и крестиков-ноликов на ограниченной доске? Это все конечные позиционные игры с полной информацией. С полной информацией — значит, что вся информация известна всем игрокам (в картах это не так) и нет скрытой информации (нерозданных карт или непредсказуемых бросков костей). Позиционные — значит ходы ведут из позиции в позицию. Конечные — значит, число позиций конечно, игра начинается с некоторой начальной позиции, и заканчивается в какой-то из конечных позиций (например, в шахматах — позицией с патом, матом, голыми королями или повторением позиции). Эта конечная позиция и определяет, кто выиграл и с каким счетом (каждую игру можно сделать игрой на счет, например, начисляя 1 за победу, 0 за ничью и -1 за поражение). Кроме того, все эти игры — антагонистические для двух участников. Антагонистические — значит выигрыш одного является проигрышем другого, и если один сколько-то очков выигрывает, то другой — проигрывает. Анализом с конца легко доказать, что при этих условиях у каждого из игроков есть оптимальная стратегия и оптимальный результат. Игра по оптимальной стратегии гарантирует результат не хуже оптимального. При этом нет стратегии, добивающейся большего с гарантией, то есть при любой игре противника. В антагонистических играх оптимальные результаты противников противоположны, в игре на счет их сумма равна нулю. Игра обоих по оптимальной стратегии создает «равновесие»: отклонение одного из игроков от этой стратегии не может принести ему выгоды. Ситуация заметно усложняется, если игроков больше двух. Может оказаться, что каждому не гарантирован никакой результат, кроме наихудшего ввиду невозможности противостоять сговору остальных.

Упражнение. На столе лежат 10 спичек. Трое игроков берут спички по очереди, от 1 до 5 за ход. Взявший последнюю спичку идет мыть посуду. Докажите, что сговорившись, любые двое могут послать мыть посуду третьего.

Давайте за мытье посуды начислять -2, а остальным по $+1$. Сумма в любой партии равна 0, однако сумма оптимальных результатов равна -6. Хочется, однако, устроить какой-то аналог равновесия и для троих: предложить игрокам три такие стратегии, которым они захотят следовать. Для этого достаточно, чтобы отклонившийся был наказан: если он не следует указанной стратегии, а остальные двое следуют, то он получит меньше (или не больше).

Упражнение. Придумайте три такие стратегии для данной игры. Набор стратегий, где единственный отклонившийся не выигрывает называется равновесием Нэша (это определение годится и для неантагонистических игр). Наша задача состоит в том, чтобы разобраться, для каких игр равновесия Нэша есть (и какие), а для каких — нет.

Равновесия Нэша можно описать как правила, о соблюдении которых можно договориться даже без механизма внешнего принуждения.

Упражнение. «Встреча в супермаркете». Два (или три) человека потерялись в супермаркете, мобильные телефоны сеть не ловят... Они могут встречаться у одного из трёх выходов, каждый выбирает куда идти независимо и не зная выбор остальных. Если все встретились, каждый получает выигрыш $+1$, иначе каждый получает -1. Какие равновесия Нэша в этой игре?

Примечание. Эта игра задана в так называемой нормальной форме (все игроки одновременно делают выбор, не зная выбор друг друга; после этого по выбору всех игроков определяются выигрыши).

Примечание 2. В некоторых супермаркетах вешают большие таблички «потерявшимся встречаться у первой кассы».

Упражнение. Есть ли равновесия Нэша в крестиках-ноликах 3 на 3? Опишите их.

Оказывается, что для ациклических игр (то есть игр, где позиции не повторяются), хотя бы одно равновесие Нэша всегда есть (даже если участников не два, а больше). Вы сами сможете это легко доказать. Давайте свяжем с игрой ориентированный граф: позиции будут вершинами, а ходы — ориентированными ребрами (стрелками). Позиции, из которых нет ходов — конечные, каждой приписан набор очков, который получат игроки при ходе в эту позицию. Остальные позиции поделены между игроками, для каждой известно, кто из нее должен ходить. Пусть из позиции П все ходы ведут в конечные, и игрок выберает наиболее выгодный для него ход в некоторую позицию Т. Очки из Т можно перенести в П. Теперь и П стала определенной, как бы конечной. Так анализом с конца делаем все позиции определенными, в том числе начальную. Оптимальная стратегия состоит в ходе игрока в такую позицию, где его выплата максимальна.

Упражнение. Докажите, что для любой ациклической игры указанные оптимальные стратегии образуют равновесие Нэша.

Давайте построим теперь граф игры в шахматах. Мы уже понимаем, что позиция — это не просто расстановка фигур, она должна включать в себя еще и очередь хода. Кроме того, полезно знать, есть ли право рокировки, взятия на проходе, повторялась ли позиция раньше. Один из выходов снабдить позицию нужной информацией — запоминать ее вместе с предысторией. Тогда повторяющихся позиций точно не будет, граф ациклический, и нем для обоих игроков есть оптимальная стратегия. Но такое понимание позиции для нас неинтересно. Правило троекратного повторения позиции подразумевает под позицией нечто другое. А именно, расстановка фигур, очередь хода, есть ли право рокировки, взятия на проходе. Тогда в графе есть ориентированные циклы, и анализ с конца уже не действует. Уточним понятие стратегии. Назовем стационарной стратегией данного игрока правило, выбирающее один определенный ход в каждой позиции с его ходом. Например, в игре в спички жадная стратегия предписывает каждый раз брать максимально возможное количество спичек. Заметим, что стационарная стратегия не зависит от предыстории, поэтому если каждый играет по стационарной стратегии, то при повторении позиции игра зацикливается. Понятно, что в шахматах зацикливание означает ничью. Можно, однако договорится, что циклы тоже имеют свою цену (например, при зацикливании все проигрывают). В играх с циклами про равновесие Нэша мало что известно.

Упражнение. Есть ли равновесие Нэша в шахматах? А если убрать правила про ничью при троекратном повторе позиции или после 50 ходов без взятий и передвижений пешек?

До сих пор мы рассматривали только терминальные игры, когда результат игры (платежи) определяется только конечной позицией или циклом. В некоторых играх, помимо этого, игрок получает или платит еще и за каждый ход, а окончательный результат игры определяется для него суммой всех платежей.

Упражнение. На столе лежат 5 спичек. Трое игроков берут спички по очереди 1 или 2 спички. Взявший последнюю спичку получает премию из 3 спичек. Число заработанных очков равно числу спичек. Постройте граф игры и найдите для всех оптимальные стратегии. Если игра заканчивается циклом, то мы предполагаем, что цикл проходится бесконечно много раз. Тогда результат игрока будет конечен, только если сумма его платежей по циклу равна 0. Иначе результат равен плюс или минус бесконечности.

В этой игре очевидно, сумма оптимальных результатов равна -3.

Упражнение. 100 кровожадных отморозков ограбили банк на миллион долларов и уселись в ряд за стол делить деньги. Сначала первый предлагает, кому сколько: мне столько-то, второму столько-то и т.д. (каждому —- целое число долларов), и все 100 голосуют. Если «за» не менее половины, то предложение принимается, каждый получает предложенную долю, и все расходятся. Если более половины голосуют «против», первого убивают, и тогда уже второй отморозок предлагает на тех же условиях кому сколько, и т.д. Каждый отморозок руководствуется в первую очередь желанием выжить, во вторую (если жизнь вне опасности) — получить побольше денег, в третью (если на жизнь и сумму это не влияет) —- убить как можно больше (а то ведь подстерегут в темном переулке!). Как распределятся деньги, если все отморозки будут действовать и рассуждать абсолютно логически? (то есть, найдите равновесие Нэша)

2. Введение «без строгих определений»

Мы рассмотрим следующий вопрос: Какие конечные позиционные игры с полной информацией имеют равновеся Нэша в чистых стационарных стратегиях? В некоторых случаях ответ хорошо известен. Дадим сначала небольшой обзор, откладывая точные определения до следующего параграфа. Равновесие Нэша существует для таких классов:

А. Ациклические игры. В них позиции не могут повторяться. В этом случае равновесие всегда имеется. Однако, уже в Шахматах, или даже в Го, повторения позиции возможны.

Б. Антагонистические игры двух лиц. Этот класс включает и Го, и Шахматы. Но что если интересы двух игроков непротивоположны? или число игроков больше двух?

В. Если ходы игроков могут зависеть от предыстории. Мы, однако, ограничиваем себя (и игроков) чистыми стационарными стратегиями. Иными словами, ход может зависеть только от текущей позиции, но не от предшествующих, и выбирается он детерминировано, без всякой рандомизации. Например, Нарды, исключаются.

Заметим, впрочем, что в рассмотренных случаях А, Б, и В равновесие существует даже и при наличии случайных ходов.

Известно, однако, что равновесий может не быть в играх с неполной информацией (карточные игры или Домино). Но мы таких игр не рассматриваем и даже определять их не будем.

Резюмируем:

Мы ограничимся играми с полной информацией, без случайных ходов и чистыми стационарными стратегиями. При этом игроков может быть более двух, и даже если два, их интересы не обязательно противоположны.

Удивительно, но в этом случае мало что известно. Есть несколько концепций решения, простейшей из которых является несомненно равновесие Нэша. (Мы определим его ниже.) Хотя за работы о равновесиях Нэша было выдано пять нобелевских премий по экономике, но мне кажется, что наиболее «простые» и естественные математические вопросы до сих пор открыты. Здесь я предлагаю два таких вопроса-гипотезы. Они проверены, с помощью компьютера, для достаточно (но не чрезмерно) больших примеров. Я надеюсь на положительные ответы, но не удивлюсь и контрпримерам.

У этих гипотез есть относительно простые частные случаи, на которых можно будет упражняться. Впрочем, другие частные случаи довольно сложны, а для некоторых ответ неизвестен, как и в общем случае.

3. Основные определения

Мне кажется, что большинство из них интуитивно очевидно. Однако, формализм может и «напугать» кого-то. Если так, то пропустите этот параграф при первом чтении и используйте его потом как словарь или справочник.

Граф игры, позиции и ходы.

Дан конечный ориентированный граф (орграф) $G = (V, E)$. Каждая его вершины $v \in V$ — позизия игры, а ориентированное ребро $e = (v, v')$ — возможный ход в позиции v. Позизии, $V_T \subset V$, в которых вообще нет ходов, называются терминальными. Выберем также начальную позицию $v_0 \in V \setminus V_T$.

Каждой нетерминальной позиции $v \in V \setminus V_T$ поставим в соответствие игрока $i \in I = \{1, \dots, n\}$, который выбирает ход в позиции v. Будем говорить, что i контролирует v и писать $i = \phi(v)$; иными словами, отображение $\phi : V \setminus V_T \to I$ распределяет нетерминальные позиции по игрокам.

Тройка $\{G, \phi, v_0\}$ называется позиционной структурой.

Стратегии и ситуации.

Стратегия x_i игрока $i \in I$ — это план, выбирающий ход $e = (v, v')$ в любой позиции $v \in \phi^{-1}(i)$, контролируемой i, иными словами, — отображение x_i, ставит в соответствие каждой позиции $v \in \phi^{-1}(i)$ некоторый ход $e = (v, v')$ из v. Это — так называемые *чистые стационарные стратегии*. Как уже говорилось, других мы ни рассматривать, ни даже определять, не будем.

Партии. Пусть каждый игрок i выберет стратегию x_i. Полученный набор $x = (x_1, \dots, x_n)$ называется *профилем стратегий* или *ситуацией*. Каждая такая ситуация однозначно определяет партию $p(x)$, поскольку каждый игрок $i \in I$ в каждой своей позиции $v \in \phi^{-1}(i)$ знает, какой ход ему делать (тот, который предписывает его стратегия x_i). Партия $p(x)$ начинается в v_0 и либо заканчивается в одном из терминалов $v \in V_T$ либо "зацикливается", т. е. возникает ориентированный цикл C, который затем повторяется бесконечно. (Партия $p(x)$ уйти с C не может, так как все стратегии стационарны.)

Таким образом, мы получаем отображение $g : X \to P$, которое каждой ситуации $x \in X$ ставит в соответствие партию $p \in P$. Такие отображения называются игровыми формами

Функции стоимости.

Каждый игрок $i \in I$ за каждый ход $e \in E$ платит $c(i, e) \in \mathcal{R}$. Это вещественное число называется *локальной стоимостью*. (Если $c(i, e) < 0$, то i не платит, а наоборот, получает $|c(i, e)|$.)

Позиционная структура и локальный платеж определяют игру в позиционной форме.

Эффективная стоимость партии $p = p(x)$ определяется для каждого игрока $i \in I$ так. Если p заканчивается в терминале $v \in V_T$, то ее стоимость $c(i, p) = \sum_{e \in p} c(i, e)$ аддитивна, т.е. равна сумме стоимостей всех ходов p. Если же p зацикливается, то надо вычислить стоимость $c(i, C) = \sum_{e \in C} c(i, e)$ соответствующего цикла C для i. Если $c(i, C) \geq 0$, то $c(i, p) = \infty$ и $c(i, p) = -\infty$, если $c(i, C) < 0$.

Такое определение естественно, поскольку цикл проходится неограниченное число раз, а локальные стоимости суммируются. Однако, для если партия p зацикливается на «нулевом цикле», $c(i, C) = 0$, мы всё равно полагаем $c(i, p) = \infty$. Это всего лишь удобное соглашение.

Игровая форма g и эффективная стоимость c определяют игру (g, c) в нормальной форме.

Естественно, каждый игрок i пытается *минимизировать* свою эффективную стоимость $c(i, p)$.

Терминальные игры. Ход $e = (v, v')$ называется *терминальным*, если $v' \in V_T$ — терминальна. Заметим, что терминальный ход не может принадлежать никакому циклу. Функция стоимости c (и сама игра) называется *терминальной*, если $c(i, e) \equiv 0$ для любого игрока i и *нетерминального* хода e. В этом случае, стоимость партии p зависит только от её терминальной позиции. Если же партия p зацикливается, то её стоимость по определению равна $+$ или $-\infty$.

Игры с нулевой суммой. Говорят, что функция стоимости c (и сама игра) имеют *нулевую сумму*, если $\sum_{i \in I} c(i, e) = 0$ для любого хода $e \in E$. Игры двух лиц, $n = 2$, с нулевой суммой играют очень важную роль, как исторически, так и по существу.

Любую игру n лиц легко превратить в игру $n + 1$ лиц с нулевой суммой. Достаточно ввести $(n + 1)$-го игрока-болвана (который не контролирует ни одной позиции) и определить локальную стоимость его ходов

формулой $c(n+1,e) = -\sum_{i=1}^{n} c(i,e)$.

Игры в нормальной форме; общее определение.

Итак, пусть $I = \{1,\ldots,n\}$ множество игроков, X_i — конечное множество стратегий игрока $i \in I$, а $X = X_1 \times \ldots \times X_n$ — их прямое произведение, т.е. множество ситуаций.

Далее, пусть P обозначает произвольное множество исходов игры (в нашем случае — партий). Произвольное отображение $g : X \to P$ называется *игровой формой*.

Наконец, пусть дана произвольная функция стоимости $c : I \times P \to \mathcal{R}$. Её вещественные значения $c(i,p)$ показывают, сколько должен платить игрок $i \in I$ за партию $p \in P$.

Пара (g,c) определяет *игру в нормальной форме*.

Равновесие Нэша и седловая точка.

Ситуация $x = (x_1,\ldots,x_n) \in X_1 \times \ldots \times X_n = X$ называется *равновесием Нэша*, если изменение стратегии любым игроком $i \in I$ (но только одним) не приносит ему выгоды, т.е. не уменьшает стоимости для него. Формально, это можно записать так: $c(i,g(x)) \le c(i,g(x'))$ для любого игрока $i \in I$ и для любой ситуации $x' \in X$ такой, что все её координаты (стратегии) те же, что и в x, за исключением, быть может, координаты i, иными словами, только x'_i может отличаться от x_i.

Это понятие было введено Джоном Нэшем в 1950 году. В случае игр двух лиц с нулевой суммой равновесие Нэша носит название *седловая точка*. Это понятие лет на 200 старше.

В отличие от седловой точки, концепция Нэша весьма уязвима для критики. Зачастую, два игрока могут изменить одновременно свои стратегии и оба выгадать. Более того, то же могут сделать иногда и все n игроков.

Ситуаций равновесия (в чистых стратегиях) может вообще не быть. А если есть, то их может быть много. Более того, не только равновесий, но и равновесных платежей может быть много.

Седловая точка лишена большинства этих недостатков. Однако, критика Нэша не является нашей целью. (Вспомним также о пяти нобелевских премиях :-)

Однородное равновесие Нэша Ситуация $x \in X$ называется *однородным* равновесием Нэша, если она является равновесием не только при данной начальной позиции $v_0 \in V$, но и при любой другой начальной позиции $v'_0 \in V$.

4. Задачи и гипотезы

Мы будем интересоваться теоремами существования равновесия Нэша (т.е. разрешимостью по Нэшу) позиционных игр, определенных выше.

Сложность проблемы я оцениваю числом очков, данным в скобках.

Гипотеза 1 (500). Верно ли, что любая позиционная игра двух лиц разрешима по Нэшу.

Задача 1 (10). Покажите, что без нарушения общности можно предположить отсутствие «нулевых циклов», точнее, ориентированных циклов с нулевой суммой локальных стоимостей. Иными словами, можно без нарушения общности предположить, что $\sum_{e\in C} c(i,e) \ne 0$ для любого ориентированного цикла C и игрока $i \in I = \{1,2\}$.

Напомним, что, эффективная стоимость любой зацикливающейся партии равна $+$ или $-\infty$.

Это совсем новая гипотеза. Владимир Удалов написал программу, которая подтвердила её для многих орграфов с 10 – 18 вершинами.

Задача 2 (25). На случай трёх игроков Гипотеза 1 не обобщается. Постройте пример.

Для игр двух лиц с нулевой суммой гипотеза верна, но доказательство сложное. Более того, в этом случае можно так ввести конечную эффективную стоимость каждой партии p, заканчивающейся «нулевым циклом» C, что седловая точка всегда будет существовать. (Напомним, что мы определили $c(i,p) = +\infty$ в этом случае.) Однако, такое переопределение непросто.

Задача 3 (70). Попробуйте его найти и доказать разрешимость. Покажите, что «очевидные попытки» не проходят. Например, если положить $c(i,p) = 0$ или $c(i,p) = \sum_{e\in p} c(i,e)$, то седловой точки может и не быть. Постройте примеры.

Гипотеза 2 (500). Верно ли, что любая позиционная игра n лиц, в которой все локальные стоимости неотрицательны, разрешима по Нэшу?

Задача 3а (5). Докажите, что достаточно рассмотреть строго положительные локальные стоимости.

Гипотеза не доказана даже в следующих «очень частных» случаях.

Гипотеза 2а (300). Терминальный платеж.

При этом эффективная стоимость любой зацикливающейся партии для любого игрока равна $+\infty$.

Гипотеза 2б (400). Терминальный платеж. При этом по-прежнему все циклы образуют один и тот же исход, НО не обязательно наихудший для всех игроков. Вместо этого, мы теперь предполагаем, что каждый из игроков ранжирует все терминалы и циклический исход произвольно.

Гипотеза 2в (300). Случай двух игроков, $n = 2$. В этом случае мы обьединяем предположения Гипотез 1 и 2.

Задача 4 (100). Докажите, что в случае двух игроков и терминальной функции стоимости Гипотеза 2 всё же верна.

Этот результат можно вывести из одной моей старой теоремы, 1975 года.

По определению, общая игровая форма n лиц $g : X_1 \times \ldots \times X_n \to P$ разрешима по Нэшу, если соответствующая игра (g, c) имеет хотя бы одно равновесие Нэша при любой функции стоимости $c : I \times P \to \mathcal{R}$. Здесь $c(i, p)$ — стоимость исхода $p \in P$ для игрока $i \in I$.

Для случая двух игроков, $I = \{1, 2\}$, наряду с общим определением разрешимости рассмотрим следующие два более слабых свойства:

Игровая форма двух лиц g называется *антагонистически разрешимой*, если она разрешима в классе игр с нулевой суммой. Наконец, g называется ± 1 *разрешимой*, если она разрешима в классе игр двух лиц с нулевой суммой, причем функция стоимости принимает только два значения: $+1$ и -1.

Задача 5 (100). Докажите, что все эти три свойства (разрешимость, антагонистическая разрешимость и ± 1 разрешимость) эквивалентны.

Эквивалентность последних двух свойств я доказал немного раньше, в 1973 году, но ещё раньше, в 1970, то же сделали Джек Эдмондс и Дэлберт Рэй Фалкерсон. Edmonds, J.; Fulkerson, D. R. (1970), "Bottleneck extrema", Journal of Combinatorial Theory 8:3 (1970) 299-306.

К сожалению, утверждение задачи 5 на игры трех лиц уже не обобщается. Сформулируем это точнее. Каждой игровой форме n лиц, $I = \{1, \ldots, n\}$, можно поставить в соответствие n игровых форм двух лиц, в которых i играет против $I \setminus \{i\}$, где $i \in I$.

Задача 5а (50). Приведите пример неразрешимой по Нэшу игровой формы трёх лиц, такой что все три соответствующих ей игровые формы двух лиц разрешимы.

Задача 5б (20) Приведите обратный пример, разрешимой по Нэшу игровой формы трёх лиц, такой что все три соответствующих ей формы двух лиц неразрешимы.

Задача 6 (20). Покажите, что Задача 4 сводится к Задаче 5.

Задача 7 (15). Докажите, что равновесие Нэша существует, если граф G ациклический (не имеет ориентированных циклов).

Подсказка: Примените динамичаское программирование. В теории позиционных игр это называется "обратная индукция".

Этот результат был получен Гарольдом Куном (1952) и Давидом Гейлом (1953) вскоре после того, как Нэш ввел своё понятие равновесия.

Задача 7а (20). Докажите, что для ациклических игр равновесие Нэша существует, даже если разрешить позиции случая (в которых задано распределение вероятностей).

Разумеется, за решение обеих этих задач можно получить максимум 20 очков, но не 35.

Задача 8 (40). Докажите, что равновесие Нэша (седловая точка) существует для позиционных игр двух лиц с нулевой суммой.

В частности, для шахмат или го. Этот результат Эрнст Цермело доложил на 5-м международном конгрессе математиков в 1912 году. Его доклад назывался: «О применении теории множеств к шахматной игре».

Замечу, что и в этом случае результат можно обобщить, разрешив позиции случая. Однако, это увело бы нас далеко в сторону (стохастических игр). Поэтому отложим это направление на будущее.

Задача 9 (10). Договоримся заканчивать игру при первом же повторении позиции. При этом исходом игры будет считаться полученный цикл. Покажите, что при этом конечный граф может быть заменен конечным деревом (в котором нет не только ориентированных, но и вообще никаких циклов). Почему же Гипотезы 1 и 2 не вытекают из Задачи 7?

Задача 10 (15) Приведите пример терминальной игры двух лиц в которой имеется всего один цикл и нет однородного равновесия. (При этом цикл не обязательно худший исход для обоих игроков. Каждый из них ранжирует терминалы и цикл произвольно.)

Задача 11 (100) Приведите пример терминальной игры двух лиц в которой нет однородного равновесия и при этом имеется всего один цикл, который является худшим исходом для обоих игроков.

Задача 12 (25) Приведите аналогичный пример терминальной игры трех лиц: в ней нет однородного равновесия и при этом имеется всего один цикл, который является худшим исходом для всех трёх игроков.

Такие примеры были построены не так давно: к Задаче 11 в 2003, а к Задаче 12 в 2008 годах.

Разумеется, во всех трех случаях (Задачи 10,11 и 12) относительно любой фиксированной начальной позиции, равновесие существует. Иначе, Гипотеза 2 была бы опровергнута.

Nash Equilibria

Vladimir Gurvich

1 Preliminary talks

What is common in Go, Chess, Checkers, and Gomoku? They all are finite positional games with perfect information and without positions of chance. The last means that "all players know everything" and hence, they know the same. This is not the case with cards or domino. A player does (or at least supposed to) not see the opponents' cards. In Backgammon, there are positions of chance, when the dices are tossed. Word "positional" means that moves lead between positions. Word "finite" means that there are finitely many positions.

Any configuration of stones is a position in Go. There are very (but finitely) many of them. A play starts with some initial position and terminates in a final one (which is called a terminal). For example, a mate or a stale mate are terminals in Chess. It is important to notice that positions can be repeated and a play may cycle.

All games considered above are zero-sum two-person games. The word "zero-sum" means that the winning of one is the losing of the other. And if one gets some number of points then the other loses the same number. A pair of strategies is optimal if they form an equilibrium, that is, the corresponding result can be improved by neither of the players.

Yet, everything becomes much more complicated when there are more than two players (or two but the game is not zero-sum). It may happen that each player can guarantee only a very poor result, because it is difficult to fight against the coalition of all other players.

An exercise: There are 10 matches, 3 players take them in a cyclic order, 1,2,3,4, or 5 for one move. One that take the last one should do the dishes. Prove that any two can always force the third one to do the dishes.

Let doing the dishes costs 2, while the winners get +1 each (that is, their cost is (-1) for each). The game is zero-sum. We want to define an equilibrium in this case too, that is, to suggest 3 strategies such that if one player changes his strategy, he cannot profit whenever two opponents keep their old strategies. This concept was introduced by John Nash in 1950 and it is called Nash equilibrium.

An exercise: Suggest some Nash equilibria in the considered game.

Our main problem is to understand which games are Nash-solvable (that is, always have Nash equilibria) and which may have none.

First, let us demonstrate that any acyclic case (in which no position can be repeated) is Nash-solvable.

Let us assign to a game a directed graph (digraph), whose vertices are the positions and the directed edges (arcs) are moves. The positions with no moves are called terminals. Let us assign to each terminal and player a number which this player gets in this terminal. For any non-terminal position we define a player who makes a move in it. Let all moves from a position v lead to terminals. Then the corresponding player chooses a best his move, say to a terminal v'. Then we transfer to v all points defined for v'. Now v "becomes terminal", etc. This procedure is called Backward Induction. It defines strategies of all players.

An exercise: Prove that these strategies form a Nash equilibrium for any acyclic games.

Let us define the concept of a strategy more accurately. A stationary strategy of a player is a rule choosing his move in each his position. For example in the above "match-game" the greedy strategy suggests to take the maximum number of matches each time. Let us note that a stationary strategy does not depend on history, that is, on the preceding positions or moves. In particular, the play must cycle whenever a position appears twice. In Chess this means a draw. However, cycles may have some other costs, for example, all players lose.

Up to now, we considered only the terminal cost functions, that is, the result depended only on the terminal (or cycle). More generally, each player may pay for each (not only for his own) move and the final result is defined as the sum of all these local costs.

An exercise: There are 5 matches. 3 players take them in a cyclic order. One who takes the last match gets a bonus: 3 matches. The result of a player is the number of collected matches. Construct the graph and find an equilibrium.

If a play cycles then we assume that the cycle is passed infinitely many times. In other words, the cost of such a play is either $+$ or $-\infty$ For convenience we will assume that it is $+\infty$ whenever the sum of all local costs ≥ 0.

2 A mini-survey "before accurate definitions"

We will consider the following question:
Which positional games (with perfect information and without moves of chance) have Nash equilibria (in pure stationary strategies)? In some cases, the answer is well known. We start with a mini-survey postponing the precise definitions till the next section. Nash equilibria (NE) exist for the following classes of games:

A. Acyclic games, in which no position can appear twice. In this case, NE always exist. Yet, already in Chess, or even Go, positions can be repeated.

B. Zero-sum two-person games. Both Go and Chess are included. Yet, what if the conflict is not antagonistic? or there are more than two participants?

C. NE exist if players' moves may depend on the preceding moves.
Yet, we will restrict ourselves (and the players) to the pure stationary strategies. In other words, a move in the present position depends only on it, not on preceding positions or moves. Furthermore, this move is chosen deterministically, without any randomization. For example, Backgammon will not be considered.

Let us note, however, that in all of the above cases, A,B, and C, a NE exists even if we allow positions of chance. It is also known that NE may fail to exist in games with imperfect information (like card games or Domino). Yet, we will not consider (and will not even define) them.

Let us summarize:
We will restrict ourselves to games with perfect information, without random moves, and to pure stationary strategies. Yet, the number of players may be greater than two and, even a two-person game, may be not zero-sum.

Somewhat surprisingly, not much is known about the considered games. There are several solution-concepts, among which the NE is certainly most popular. (It will be defined in the next section.) Although, five Nobel Prizes in Economics were granted for works on NE, but in my opinion, the "simplest" and most challenging related mathematical problems are open yet.

Here we will consider two such problems-conjectures. By computers, they are verified for sufficiently (but not too) large examples. So, I hope for affirmative solutions but a counterexample would not be too surprising to me, either.

These conjectures are verified in some special cases, which will serve as exercises. In contrast, some other special cases are difficult, or even open.

3 Main definitions

They all are pretty natural but still so much formalities may scare some.
In this case, I recommend to skip this section during the first reading and use it as a dictionary, only when necessary.

Graphs, positions, and moves. Given a finite directed graph (digraph) $G = (V, E)$, a vertex $v \in V$ is interpreted as a *position*, while a directed edge $e = (v, v')$ is a move from the position v in the corresponding game.

The positions $V_T \subset V$ without any moves are called *terminal*.

Let us choose also an initial position $v_0 \in V \setminus V_T$.

Players. Let us assign to each non-terminal position $v \in V \setminus V$ a *player* $i \in I = \{1, \dots, n\}$ who will make a move in v, say that v is controlled by i, and write $i = \phi(v)$. In other words, the mapping $\phi : V \setminus V_T \to I$ distributes the non-terminal positions among players.

Triplet $\{G, \phi, v_0\}$ is called a *positional structure*.

Strategies and situations. A *strategy* x_i of player $i \in I$ is a plan choosing a move $e = (v, v')$ in every position $v \in \phi^{-1}(i)$ controlled by i. In other words, the mapping x_i assigns a move $e = (v, v')$ from v to v' for each position $v \in \phi^{-1}(i)$.

These are the so-called *pure stationary strategies*. As we already mentioned, we will not consider, nor even define, any others.

Let each player i choose a strategy x_i. The obtained n-tuple $x = (x_1, \dots, x_n)$ will be called a strategy profile or a *situation*.

Plays. Eeach situation x uniquely defines a play $p(x)$, since each player i knows what to do in every position $v \in \phi^{-1}(i)$ (to move according to x_i).

The play $p(x)$ begins in v_0 and either it terminates an a $v \in V_T$ or "cycles", that is, passes a directed cycle C infinitely many times. (Let us notice that $p(x)$ cannot leave C, since all strategies of x are stationary.)

Thus, we obtain a mapping $g : X \to P$ that assigns a play $p = p(x) \in P$ to every situation $x \in X$. Such a mapping is called a *game form*.

Local and effective cost functions. Each player $i \in I$ pays the value $c(i, e) \in \mathbb{R}$ for each (not only for his own) move $e \in E$. This real number is the *local cost*. (Of course, if $c(i, e) < 0$ then i gets $|c(i, e)|$ rather than pays it.)

A positional structure and a local cost function define a positional game.

The *effective cost* $c(i, p)$ of a play $p = p(x)$ for a player $i \in I$ is defined as follows. If p ends in a terminal position $v \in V$ then $c(i, p) = \sum_{\in p} c(i,)$, that is, the cost is additive and is equal to the sum of the costs of all moves of p for i.

If p cycles on C, we have to compute the cost $c(i, C) = \sum_{\in C} c(i, e)$ of C first. If $c(i, C) \geq 0$ then $c(i, p) = \infty$ and $c(i, p) = -\infty$ when $c(i, C) < 0$.

Such definition is natural. Indeed, play p repeats C infinitely and the local costs are summed up. Yet, when C is a "zero-cycle" for i, that is, $c(i, C) = 0$, we still set $c(i, p) = \infty$. This is just a helpful convention.

A game form $g : X \to P$ together with an effective cost function $c : I \times P \to \mathbb{R}$ define a game (g, c) in the normal form.

Naturally, each player $i \in I$ is trying to *minimize* his effective cost $c(i, p)$.

In particular, all players should avoid non-negative cycles. We will see, however, that it might be not that easy to do.

Terminal moves, costs, and games. A move $e = (v, v')$ is called *terminal* if $v' \in V_T$ is a terminal position. Let us notice that a terminal move cannot belong to a directed cycle. A local cost function c (and the obtained game) are called *terminal* if $c(i, e) \equiv 0$ for each player i and every *non-terminal* move e. In this case, the effective cost of a terminal play p depends only on its last move and if p cycles then its cost is $+$ or $-\infty$, by definition.

Zero-sum games. We say that a local cost function c (as well as the corresponding game) are *zero-sum* if $\sum_{i \in I} c(i, e) = 0$ for every move $e \in E$. Zero-sum two-person, $n = 2$, games play very important role.

Any n-person game can be easily converted into a zero-sum $(n + 1)$-person game. Let us introduce a new $(n + 1)$st player who will be a dummy (in control of no positions) and define his local cost function $c(n + 1, e) = -\sum_{i=1}^{n} c(i, e)$.

Games in normal form; general definition. Let $I = \{1, \dots, n\}$ be a set of players, $_i$ be a finite set of strategies of $i \in I$, and $X = X_1 \times \ldots \times X_n$ be the direct product of these n sets, that is, X is the set of situations.

Furthermore, let P denote an arbitrary set of outcomes (plays, in our case). An arbitrary mapping $g : X \to P$ is called a *game form*.

Finally, given an arbitrary cost function $c : I \times P \to \mathbb{R}$, its real values $c(i, p)$ show how much player $i \in I$ must pay for the play $p \in P$.

Finally, the pair (g, c) is called a *game in normal form*.

Nash equilibria and saddle points.

A situation $x = (x_1, \dots, x_n) \in X_1 \times \ldots \times X_n = X$ is called a *Nash equilibrium* (NE), if no player $i \in I$ can profit by replacing his strategy (x_i by x'_i) provided all other players keep their old strategies. Formally, we can write this as follows:

$c(i, g(x)) \leq c(i, g(x'))$ for every player $i \in I$ and for each situation $x' \in X$ such that all components (strategies) of x' are the same as in x, except for, maybe, the ith one; that is, only x'_i may differ from x_i, while $x'_j = x_j$ for all $j \in I \setminus \{i\}$.

This concept was introduced by John Nash in 1950. For the two-person zero-sum games, NE are *saddle points*, which concept is about 200 years older.

In contrast to saddle points, the concept of NE is vulnerable for criticism. Indeed, two players might change simultaneously their strategies and profit both. Moreover, sometimes all n players may do the same. (In other words, NE are not necessarily Pareto-optimal.) Furthermore, NE (in pure strategies) may fail to exist (as well as the saddle points, yet). Then, NE may be numerous, moreover NE costs might be not unique, either. However, we are here not for criticizing but for studying the concept of NE. (Remember also about five Nobel Prizes :-)

Uniform Nash equilibria. A situation $x \in X$ is called a *uniform* NE , if it is a NE not only with respect to the given initial position v_0 but with respect to any other initial position $v_0' \in V$, as well.

4 Conjectures and problems

We are interested in the NE existence theorems, or in other words in *Nash-solvability* (NS), of the positional games defined above.

The number of points in parentheses measure the complexity of the problems.

Conjecture 1 (500). Prove or disprove that any two-person positional game is Nash-solvable (NS).

Problem 1 (10). Show that, proving Conjecture 1, one can assume, without loss of generality, that the considered game contains no "zero-cycles"; more precisely, $\sum_{e\in C} c(i,e) \neq 0$ for every directed cycle C and player $i \in I = \{1,2\}$.

Let us recall that the effective cost of any cycling play is either $+$ or $-\infty$.

This is a brand new conjecture. It was verified by a recent computer code written by Vladimir Oudalov for many digraphs with 10-18 vertices.

Problem 2 (25). Give an example showing that Conjecture 1 cannot be extended for the three-person case, $n = 3$.

The conjecture does hold for the two-person zero-sum case but all known proofs are difficult.

Moreover, in this case we can introduce a *finite* effective cost function for all plays resulting in zero-cycles such that a saddle point always exists. (Let us recall that we have set $c(i,p) = +\infty$ in the considered case.)

Problem 3 (70). Show that the following "natural redefinings" fail:
$c(i,p) = 0$ or $c(i,p) = \sum_{e\in p} c(i,e)$. In both cases there may be no saddle points. Give examples and try to find the "right" definition.

Conjecture 2 (500). Prove (or disprove) that an n-person positional game is Nash-solvable whenever all its local costs are non-negative.

Problem 4 (5). Show that proving Conjecture 2, without loss of generality, one can restrict himself to the strictly positive costs.

This conjecture remains unproved even in the following very special cases:

Conjecture 2a (300). Does Conjecture 2 hold for the games with a terminal cost function? (In this case the effective cost of every cycling play is $+\infty$.)

Conjecture 2b (400). The same question but now we do not assume that every cycling play is the worst outcome for all n players. Instead, we assume that all such plays form the same outcome but each player can rank this cyclic and all terminal outcomes arbitrarily. Does NS hold?

Conjecture 2c (200). Does conjecture 2 hold for two-person games?

In other words, here we unite the conditions of Conjectures 1 and 2.

Problem 5 (100). Prove that Conjecture 2 holds, yet, for the case of two players and terminal cost functions.

This result can be derived from my old theorem of 1975. By definition, a general n-person game form $g : X_1 \times \ldots \times X_n \to P$ is NS when the corresponding game (g, c) has at least one NE for every cost function $c : I \times P \to \mathbb{R}$.

Here $c(i,p)$ is the cost of the outcome $p \in P$ for the player $i \in I$.

For the two-person case, $I = \{1, 2\}$, let us introduce the next two relaxations of the above condition. A two-person game form g will be called:

zero-sum solvable if it is solvable in the class of the zero-sum games;

± 1 *solvable* if it is solvable in the class of the zero-sum games whose cost function takes the values $+1$ and -1 only.

Problem 5a (100). Prove that all three above properties (solvability, zero-sum solvability, and ± 1 solvability) of the two-person game forms are equivalent.

The equivalence of the last two properties is not difficult to show and I proved it a bit earlier, in 1973. Yet, even earlier this was demonstrated by Jack Edmonds and Delbert Ray Fulkerson, in their paper "Bottleneck extrema", Journal of Combinatorial Theory 8:3 (1970) 299–306.

Unfortunately, the statement of Problem 5a cannot be extended to the three-person case, already. More precisely, let us assign to each n-person game form n two-person game forms in which i is playing against the complementary coalition $I \setminus \{i\}$, for all n players $i \in I$.

Problem 5b (50). Construct a three-person game form that is not NS, while all three corresponding two-person game forms are NS.

Problem 5c (20). Construct an "inverse" example: a three-person NS game form such that all three corresponding two-person game forms are not NS.

Problem 6 (20). Reduce problem 4 to Problem 5.

Problem 7 (15). Prove Nash-solvability for the games on acyclic digraphs.
(A digraph G is acyclic if it has no directed cycles.)
Hint: Make use of the dynamic programming (which is called "backward induction" in game theory.) This result was obtained by Harold Kuhn in 1951 and David Gale in 1953, soon after Nash coined his concept of equilibrium.

Problem 7a (20). Prove that NS holds for the acyclic case, even in the presence of positions of chance (for each of which a probabilistic distribution among possible moves is given).

Of course, solution of the last two problems costs 20 points rather than 35.

Problem 8 (40). Prove that a NE (that is, a saddle point) always exists in the zero-sum two-person positional games.

In particular, for Chess and Go. This theorem belongs to Ernst Zermelo:
"On an Application of Set Theory to the Theory of the Game of Chess",
Proceedings of the Fifth Congress of Mathematicians at Cambridge, 1912.

Let us note that this result also can be extended to allow positions of chance. Yet, this direction would lead us too far to the stochastic game theory. So, we will postpone it for the future.

Problem 9 (10). Let us agree to finish the game as soon as a position is repeated, the obtained cycling play being the outcome. Then, any finite digraph is reduced to a tree (that have no cycles at all, directed or not). Then, why Conjectures 1 and 2 do not result from Problem 7?

Problem 10 (15). Give an example of a terminal two-person game that has a unique directed cycle and no uniform NE. (Here, we do not assume that the cycle is the worst outcome for both players. Each of them may rank it and the terminals arbitrarily.)

Problem 11 (100). Give an example of a terminal two-person game that has a unique directed cycle and no uniform NE, now assuming that the cycle is the worst outcome for both players.

Problem 12 (25). Provide a similar three-person example with a unique directed cycle that is the worst outcome for all players and without uniform NE.

Such examples were obtained just recently: in 2003 for Problem 11 and in 2008 for Problem 12. Of course, in all three cases (of Problems 10, 11, and 12), a NE exists with respect to any fixed initial position. Otherwise, Conjecture 2 would be disproved.