Летняя школа «Современная математика»
Дубна, июль 2005

В. И. Арнольд

Экспериментальное наблюдение математических фактов

Москва
Издательство МЦНМО
2006

УДК 512.817.3
ББК 22.144
А84

Арнольд В. И.

А84 Экспериментальное наблюдение математических фактов. — М.: МЦНМО, 2006. — 120 с.

ISBN 978-5-94057-282-4

Книга содержит записи курсов лекций, прочитаных академиком В. И. Арнольдом в 2005 г., в Дубне, на летней школе «Современная математика». В книге рассказывается о нескольких новых направлениях математических исследований, основанных на численных экспериментах.

ББК 22.144

ISBN 978-5-94057-282-4

Оглавление

ПРЕДИСЛОВИЕ

Не достигнув желаемого, они делали вид,
что желали достигнутого

М. Монтень

В этом курсе лекций я расскажу о нескольких новых направлениях математических исследований. Все они основаны на численных экспериментах. Рассматривая примеры, вроде $5 \cdot 5 = 25$ и $6 \cdot 6 = 36$, мы догадываемся о гипотезах, вроде $7 \cdot 7 = 47$, а дальнейшие эксперименты либо подтверждают их, либо опровергают.

Например, гипотеза Ферма (о неразрешимости при целом $n > 2$ уравнения в натуральных числах $x^n + y^n = z^n$) была подмечена им при попытках найти решения. Эта гипотеза привела к созданию целой науки, но доказана она была только сотни лет спустя.

Бóльшая часть гипотез, к которым мы придем, пока не доказана (и не опровергнута). Я решился читать эти лекции именно потому, что надеюсь на участие слушателей в исследовании этих вопросов, хотя бы в проведении численных экспериментов (которые сам я провел без компьютера в ограниченной области чисел первого миллиона).

ЛЕКЦИЯ 1

СТАТИСТИКА ТОПОЛОГИИ И АЛГЕБРЫ

Я никогда не слыхал о таком математике: он ведь физик

Ландау о Пуанкаре

Главное не Шекспир, а примечания к нему

А. П. Чехов с слов Б. Л. Пастернака

Крупнейший математик нового времени Пуанкаре делил все проблемы на два класса: бинарные и интересные. Бинарная проблема — это проблема, допускающая ответ «да» или «нет» (как, например, вопрос Ферма).

А интересные проблемы — это те, в которых ответ «да» или «нет» недостаточен, в них нужно исследовать какой-либо вопрос, двигаясь вперед. Например, Пуанкаре интересовался, как можно изменить условия задачи (скажем, краевые условия для дифференциального уравнения), сохраняя существование и единственность решения, или как меняется число решений при других изменениях. Так он создал теорию бифуркаций.

За три года до проблем Гильберта Пуанкаре сформулировал основные, по его мнению, математические проблемы, которые девятнадцатый век оставляет двадцатому. Это — создание математической базы квантовой и релятивистской физики.

Сегодня некоторые думают, что релятивистской физики тогда, в 1897 году, еще не было, так как Эйнштейн опубликовал свою теорию относительности в 1905 году. Но Пуанкаре сформулировал принцип относительности уже в своей статье 1895 года «Об измерении времени», которую Эйнштейн и использовал (о чем он, впрочем, не писал до 1945 года). Точно так же при создании квантовой механики Шрёдингеру удалось добиться успеха только за счет использования предшествовавших математических работ Германа Вейля, о которых впоследствии никто не упоминает, хотя Шрёдингер на них и сослался (в своей первой книге).

§1. Шестнадцатая проблема Гильберта

Хотя я в основном соглашаюсь с Пуанкаре, сегодня я буду говорить о бинарной (или почти бинарной, почему я о ней и буду говорить) проблеме Гильберта, имеющей в его списке номер 16.

Задача эта гораздо старше Гильберта — это вообще одни из основных вопросов всей математической науки (и многих ее приложений).

Вот простейший пример: для алгебраического многочлена f от двух переменных x и y рассмотрим кривую, где он обращается в нуль:

$$\{(x,\, y) \in \mathbb{R}^2 \colon f(x,\, y) = 0\}.$$

Вопрос состоит в том, *как может быть устроена топологически эта кривая, если f — многочлен фиксированной степени n.*

Например, если $n = 2$, то, согласно древней теории конических сечений, кривая — либо эллипс, либо гипербола, либо парабола, либо пара прямых (быть может, сливающихся), либо вся плоскость (если многочлен — тождественный нуль).

Добавляя к плоскости бесконечно удаленные точки, мы превращаем ее в проективную плоскость, от чего задача становится проще (эллипс, гипербола и парабола на проективной плоскости устроены одинаково, различие — только в расположении этой «окружности» по отношению к бесконечно удаленной прямой, см. рис. 1).

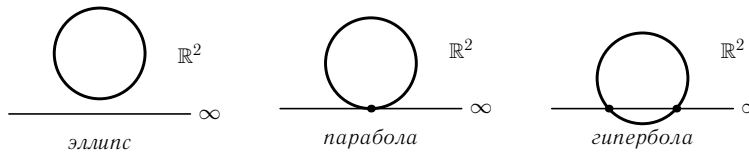


Рис. 1. Конические сечения на проективной плоскости

Для $n > 2$ вопрос более труден, но Декарт и Ньютон разобрали случаи $n = 3$ и $n = 4$. Гильберт утверждал, что он исследовал кривые степени $n = 6$, но его результат (доказательство которого он никогда не опубликовал) был ошибочным.

Кривая степени n состоит, согласно теореме Харнака, из не более чем $g + 1 = \dfrac{(n-1)(n-2)}{2} + 1$ связных компонент (где g — род соответствующей римановой поверхности, образованной комплексными решениями уравнения кривой в комплексной проективной плоскости $\mathbb{C}P^2$). Всякая замкнутая связная ориентируемая поверхность (согласно основной теореме топологии) представляет собой поверхность рода g, где g — число ручек, которые надо добавить к сфере, чтобы получить эту поверхность (см. рис. 2).

При $n = 6$ мы находим род римановой поверхности $g = 10$, так что вещественная кривая степени 6 имеет не более 11 компонент (называемых «овалами» и похожих на окружности, во всяком случае диффеоморфных окружности S^1).

Гильберт утверждает, что эти 11 овалов могут быть расположены на (проективной) плоскости $\mathbb{R}P^2$ *только двумя способами*.

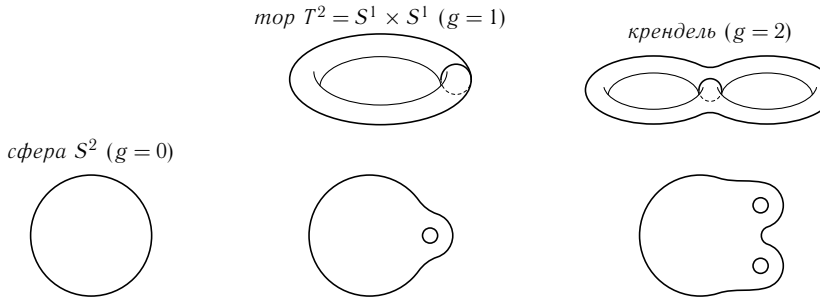


Рис. 2. Поверхности рода 0, рода 1 и рода 2

Каждый овал ограничивает «диск», диффеоморфный кругу (дополнение в $\mathbb{R}P^2$ к этому диску составляет лист Мёбиуса, из-за этого Мёбиусом и открытый).

И вот, Гильберт утверждал, что только один из этих дисков содержит внутри себя другие овалы, и что число этих внутренних овалов может принимать *только два значения*: 1 и 9 (рис. 3).

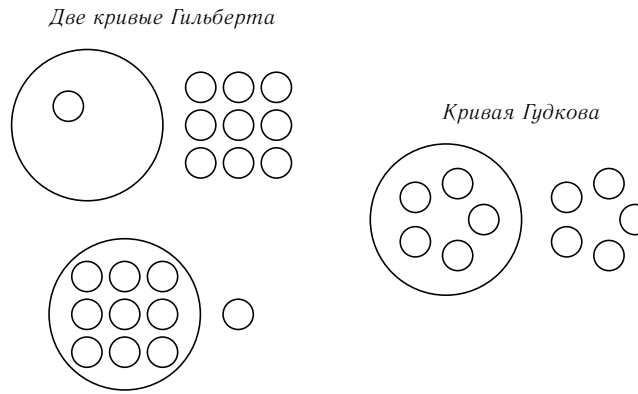


Рис. 3. Алгебраические кривые степени 6 с 11 овалами

Ошибка Гильберта состояла в том, что число внутренних овалов может равняться еще и пяти (это открыл нижегородский математик Дмитрий Андреевич Гудков около 1970 года).

Для кривых степени 8 вопрос Гильберта не решен и сегодня: 22 овала кривой 8 степени могли бы располагаться на плоскости миллиардами различных способов, но найденные сегодня ограничения уменьшают число

топологически разных расположений кривых, этих случаев остается менее 90. Число же построенных примеров, хотя и превосходит 70, пока еще не столь велико, как число допускаемых наукой возможностей.

Интересно, что, хотя вопрос и кажется относящимся к вычислительной математике, компьютеры до сих пор не внесли почти никакого вклада в его решение.

Если коэффициенты многочлена известны, то компьютер способен нарисовать расположение овалов соответствующей кривой. Но перечисление *всех* встречающихся возможностей (при всевозможных значениях коэффициентов) — гораздо более трудная задача.

Она тоже алгоритмически разрешима (в смысле математической логики), можно даже, в принципе, найти число связных областей, на которые делит пространство многочленов степени n бифуркационная диаграмма, вблизи которой тип кривой меняется. Но необходимые для этого вычисления столь велики, что никакой прогресс вычислительной техники не позволяет надеяться на компьютерное решение задачи о многочленах степени 8 в обозримое время.

Несколько отвлекаясь от темы сегодняшней лекции, я расскажу об единственном мне известном (и очень недавнем) успехе компьютерной техники в близкой задаче.

Рассмотрим график вещественного многочлена степени n от двух переменных как поверхность, $z = f(x, y)$, в трехмерном пространстве $\mathbb{R}^3$.

Около некоторых своих точек эта поверхность локально выпукла (такие точки называются *эллиптическими*), около других — локально седловая (такие точки называются *гиперболическими*, (см. рис. 4)).

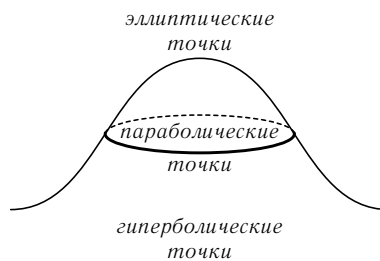


Рис. 4. Параболическая кривая на гладкой поверхности

Эллиптические и гиперболические точки поверхности разделяются линией *параболических точек*.

В терминах частных производных функции f кривая параболических точек задается уравнением

$$\det\begin{vmatrix} \partial^2 f/(\partial x)^2 & \partial^2 f/(\partial x \partial y) \\ \partial^2 f/(\partial y \partial x) & \partial^2 f/(\partial y)^2 \end{vmatrix} = 0,$$

то есть условием $f_{xx} f_{yy} = (f_{xy})^2$ обращения в нуль гессиана функции f.

Пусть f — многочлен степени n. Спрашивается, *из скольких замкнутых кривых* (*овалов*) *может состоять его параболическая кривая*?

Для многочлена f степени 4 гессиан — тоже многочлен степени 4, поэтому по теореме Харнака число овалов не превосходит $g + 1 = 4$.

Многочлен f степени 4, доставляющий параболическую кривую, состоящую из трех овалов, построить нетрудно (предоставляю это слушателям в виде задачи).

А вот вопрос о том, может ли параболическая кривая многочлена степени 4 состоять из четырех овалов, оказался очень трудным.

Его решила в Мексике в 2005 году Адриана Ортиц-Родригес, защитившая перед этим в Париже, как моя ученица, диссертацию (где для многочленов степени n число овалов параболической кривой оценивалось сверху числом an^2, а снизу числом bn^2, причем $a > b$).

Когда она была еще студенткой (в университете Париж-Жюсьё), то, придя ко мне на семинар, попросила себе задачу. Я сказал, что, чтобы понимать мои задачи, надо решить сперва письменно 100 задач статьи «Математический Тривиум» (Успехи Мат. Наук. 1991. Т. 46, № 1, С. 225—232). Московские хорошие студенты решают их все за 3 часа.

Адриана принесла мне решения этих задач, но они все оказались неверными. Она попросила неделю на раздумье, и через неделю принесла 10 верно решенных задач. Через 10 недель она решила их все сто, и начала разбираться в математике.

Но когда я хотел сформулировать ей научную задачу для самостоятельных размышлений, то Адриана сказала: «нет, теперь я придумала себе задачу, в стиле Вашего семинара и работ Ваших учеников о лагранжевых особенностях в симплектической геометрии, сама» — и сформулировала обсуждавшийся выше вопрос о параболических кривых.

Я ответил, что убедился теперь, что и в Мексике учат математике так же плохо, как и в Париже (где я довольно хорошо знал, сколь низок уровень знаний студентов).

Неспособность решать задачи тривиума была у Адрианы следствием именно плохого обучения основам математики, которому она подверглась и в Мехико, и в Париже. Ведь и с сообразительностью, и с математическими способностями у нее все было в порядке (как показал ее дальнейший опыт

и с «тривиумом», и с параболическими кривыми): после того, как я всему ее обучил своей сотней задач, она стала отличным математиком.

Вопрос о росте числа овалов параболической кривой для многочлена степени n (о сближении постоянных a и b в асимптотических оценках $an^2\ldots$ и $bn^2\ldots$ сверху и снизу) остается открытым и сегодня (почему я включил его в эту лекцию, надеясь, что и здесь найду талантливых учеников).

Что же касается исходного случая $n=4$, то защитившая в Париже диссертацию Адриана, став профессором в Мехико, получила неограниченное компьютерное время. За год непрерывной работы ЦПУ ее компьютер рассмотрел 50 миллионов многочленов $f(x, y)$ степени 4. У трех из них оказалось по четыре овала в параболической кривой у каждого.

Когда коэффициенты многочлена известны, проверка того, сколько у него овалов в параболической кривой, занимает, даже без компьютера, считанные минуты. Так что из окончательных теорем компьютерный эксперимент можно было бы и выбросить.

Но *найти* эти замечательные многочлены без компьютера никак не удавалось, так что вклад этого компьютерного эксперимента в трудное решение описываемой задачи оказался решающим.

Я надеюсь, что и в обсуждаемых ниже задачах мои слушатели сумеют добиться аналогичных успехов.

Замечание. Прежде, чем двигаться дальше, я объясню несколько (использованных выше) вещей, тщательно скрываемых от обучающихся при традиционном псевдо-научном изложении математики.

Определитель («det») матрицы второго порядка $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ — это *площадь параллелограмма*, построенного на векторах — столбцах (a, c) и (b, d), (см. рис. 5), считаемая со знаком плюс, если векторы ориентируют плоскость так же, как первый и второй координатные орты (и со знаком минус в противном случае).

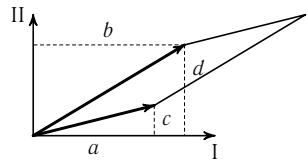


Рис. 5. Ориентированный положительно параллелограмм

Две пары линейно независимых векторов на плоскости *ориентируют ее одинаково*, если их можно соединить непрерывным путем в

пространстве упорядоченных пар линейно независимых векторов плоскости.

Разных ориентаций (классов эквивалентностей упорядоченных пар векторов на плоскости, упорядоченных реперов из n линейно независимых векторов в $\mathbb{R}^n$) — *ровно две* (при любом n). Этот важнейший естественно-научный факт (который только один и объясняет странное правило: «минус на минус дают плюс») обычно скрывают от обучающихся, заменяя всю эту геометрию постулируемой формулой

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc,$$

являющейся, на самом деле, легким *следствием* из приведенных выше топологических фактов (полезно еще отметить *линейную* зависимость определителя от каждого вектора столбца и его кососимметричность: смену знака при перестановке двух столбцов).

Вторые производные многочлена (или иной гладкой функции) в точке образуют матрицу (порядка m для функций от m переменных), $\partial^2 f/\partial x_i \partial x_j$.

Определитель этой *матрицы Гессе* функции f называется *гессианом* функции f. Полезно заметить, что знак гессиана функции f совпадают со знаком гауссовой кривизны графика функции f (и, между прочим, не зависит от выбора ориентации пространства, где функция определена). Я не останавливаюсь на этом замечании потому, что оно понятно только тем, кто знаком с гауссовой кривизной — а знакомые с ней легко докажут сделанные выше утверждения о связи гессиана с гауссовой кривизной графика.

Еще одно замечание — о роде g римановой поверхности алгебраической кривой степени n. Мы использовали выше «формулу Римана—Гурвица»,

$$g = \frac{(n-1)(n-2)}{2}.$$

Например, кривые степени $n = 1$ (прямая) и $n = 2$ (окружность) имеют род $g = 0$, т. е. вещественно диффеоморфны сфере S^2 (называемой также сферой Римана $\mathbb{C} \sqcup \{\infty\}$ или комплексной проективной прямой $\mathbb{C}P^1$).

Для прямой это ясно, а для окружности следует из ее рациональной параметризации «тангенсом половинного угла» $t = \operatorname{tg}\beta = y/(1+x)$:

$$x = \frac{1-t^2}{1+t^2}, \quad y = \frac{2t}{1+t^2} \quad \text{при} \quad x^2 + y^2 = 1. \tag{1}$$

Полезная задача — постараться понять топологическое строение «комплексной сферы», заданной в проективном пространстве, $\mathbb{C}P^3$, аффинным уравнением $x^2 + y^2 + z^2 = 1$. Ответ: это четырехмерное многообразие диффеоморфно прямому произведению двух обычных сфер, $S^2 \times S^2$.

Выписанные формулы (1) (доставляющие также «египетские прямоугольные треугольники», $3^2+4^2=5^2$, $5^2+12^2=13^2$, и т. п., а именно $a^2+b^2=c^2$ для $x=a/c$, $y=b/c$, где, согласно (1), при $t=u/v$, $a=v^2-u^2$, $b=2uv$, $c=u^2+v^2$) определяют диффеоморфизм комплексной окружности сфере S^2.

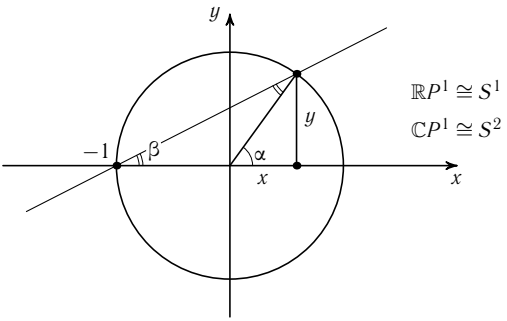


Рис. 6. Рациональная параметризация окружности

Формулы (1) выводятся так (рис. 6). Проведем через точку ($x=-1$, $y=0$) плоскости прямую $\{y=t(x+1)\}$. Подставляя это значение y в уравнение окружности $x^2+y^2=1$, мы получим для абсциссы x точки пересечения прямой с окружностью квадратное уравнение, один корень которого ($x=-1$) нам известен.

Для второго корня теорема Виета доставляет рациональное выражение через t, откуда и получается первая (а затем и вторая) формулы параметризации (1).

Вместо этой алгебры можно было бы воспользоваться геометрическим тождеством $\alpha=2\beta$ теоремы о внешнем угле (равнобедренного треугольника), ведь

$$x=\cos\alpha, \qquad y=\sin\alpha, \qquad t=\operatorname{tg}\beta.$$

Для знакомых с анализом слушателей отмечу еще, что из той же рациональной параметризации окружности следует явная вычислимость (в элементарных функциях) всех абелевых интегралов вдоль окружности:

$$I=\int\limits_{x^2+y^2=1} R(x,\,y)\,dx,$$

(где R — рациональная функция).

Действительно, рациональная параметризация (1) сводит вычисление интеграла I к интегрированию рациональной функции параметра t,

$$I=\int r(t)\,dt.$$

Абель доказал, что *такое элементарное интегрирование становится (для подходящей дроби R) невозможным, если вместо окружности абелев интеграл I берется вдоль кривой высшего рода* (с $g > 1$). Например, это так уже для эллиптических интегралов (вдоль кривой $y^2/2 + U(x) = 0$, где U — многочлен степени 3, хотя бы $U(x) = x^3 + ax + b$).

Доказательство этой топологической теоремы Абеля — тоже замечательное упражнение.

Топологической она является потому, что не представима конечной комбинацией элементарных функций не только функция[1]

$$t(X) = \int^{X} \frac{dx}{y}, \quad \text{где} \quad y^2/2 + U(x) = 0,$$

но и никакая топологически эквивалентная (многозначной) комплексной функции t функция, причем эта непредставимость имеет место и для обратных функций, эквивалентных «эллиптической функции» $X(t)$ (для невырожденных значений коэффициентов a и b).

Формула Римана—Гурвица ($g = (n-1)(n-2)/2$ для гладкой кривой степени n) проще всего доказывается следующим «итальянским» рассуждением.

Рассмотрим какое-нибудь естественное семейство алгебро-геометрических комплексных объектов (например, семейство многочленов степени n от одной переменной или семейство многочленов данной степени от двух переменных, задающих алгебраические кривые, или соответствующее семейство однородных многочленов фиксированной степени от трех переменных, задающих алгебраические кривые в комплексной проективной плоскости $\mathbb{C}P^2$).

«Итальянское» соображение состоит в том, что *все невырожденные объекты в семействе топологически одинаковы* (например, все многочлены степени n без кратных корней имеют одинаковое число корней в случае многочленов от одной переменной, все гладкие алгебраически кривые степени n в $\mathbb{C}P^2$ имеют одинаковый род $g(n)$, не зависящий от выбора конкретной кривой).

Доказательство этого соображения — топологическое. Дело в том, что вырожденность комплексного объекта задается комплексным уравнением (дискриминант равен нулю в случае многочленов от одной переменной и т. д.). А это комплексное условие на комплексные коэффициенты, выбор которых задает объект семейства, представляет собой *два веществен-*

[1] Выражающая время t движения до точки X под действием уравнения Ньютона, $\frac{d^2x}{dt^2} = -\frac{dU}{dx}$.

ных независимых уравнения (в ноль должны обращаться и вещественная, и мнимая части дискриминанта).

Поэтому, *алгебраическое многообразие всех вырожденных объектов имеет вещественную коразмерность два* (в рассматриваемом семействе многочленов и т. п.). Но подмногообразие вещественной коразмерности два не делит на части гладкое многообразие всех объектов семейства (как точка не делит плоскость, а прямая или кривая не делит на части трехмерное пространство).

Поэтому многообразие невырожденных объектов связно. А отсюда следует одинаковость топологического типа всех этих невырожденных объектов, так как при движении вдоль кривой в пространстве невырожденных объектов (например, многочленов без кратных корней) топологическая структура объекта (число корней уравнения в предыдущем примере) не меняется (по теореме о неявной функции).

Доказанный принцип показывает, что *для вычисления топологических характеристик всех невырожденных объектов комплексного семейства достаточно рассмотреть один пример и вычислить эти характеристики для него*: *для остальных невырожденных объектов характеристики будут такими же*.

Например, в качестве многочлена степени n достаточно взять многочлен

$$f(x) = (x-1)(x-2)\dots(x-n),$$

который очевидно имеет ровно n корней $x = 1, 2, \dots, n$ (кратности 1).

Согласно «итальянскому принципу», из этого следует «основная теорема алгебры»: *всякий многочлен степени n от одной переменной*, *не имеющий кратных корней*, *имеет ровно n комплексных корней*.

В случае плоских алгебраических кривых достаточно найти род одной (неособой) кривой степени n.

Начнем топологическое исследование с особой кривой степени n, распадающейся на n прямых, пересекающихся попарно в $n(n-1)/2$ различных точках (рис. 7).

Если уравнение этой кривой имеет вид $f_0 = 0$, где f_0 — произведение n линейных неоднородных функций $a_j x + b_j y + c_j$, то уравнение $f_\varepsilon = 0$ (где $f_\varepsilon = f_0 - \varepsilon$) задает при малых $\varepsilon \neq 0$ гладкую кривую степени n, род которой мы и будем теперь вычислять.

Это вычисление проводится так. Кривая $f_0 = 0$ состоит из n сфер S_j^2, пересекающихся попарно в $n(n-1)/2$ различных точках. При малом ε переход к кривой $f_\varepsilon = 0$ означает замену креста, образованного двумя пересекающимися трансверсально гладкими сферами около их точки пересечения,

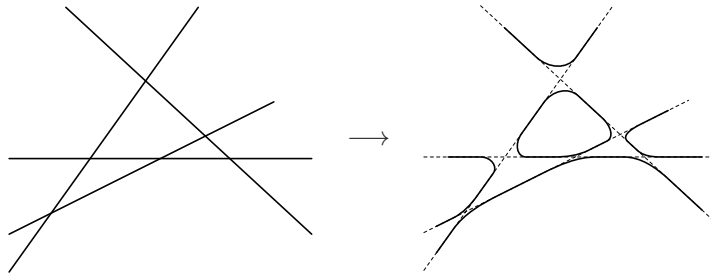


Рис. 7. Деформация распадающейся кривой степени n

на цилиндр, соединяющий дополнения к окрестностям точки пересечения на каждой из сфер (рис. 8).

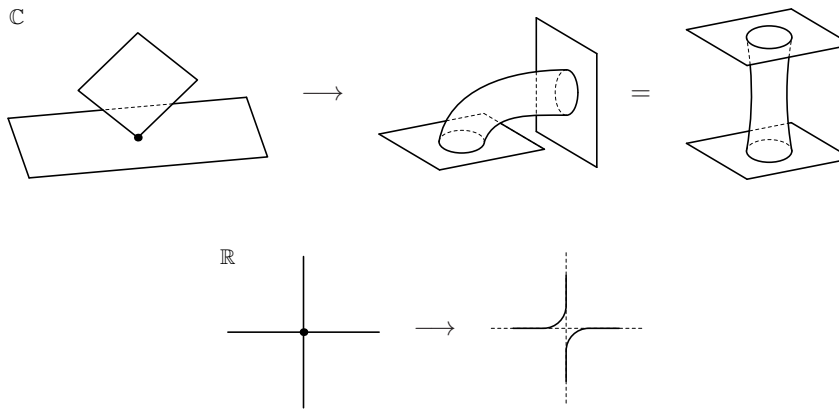


Рис. 8. Поведение двойной точки при деформации

Определим, сколько ручек получится после $n(n-1)/2$ таких происходящих независимо перестроек (около всех точек пересечения).

Из n сфер $S_1^2, \ldots, S_n^2$ выберем одну, S_1^2. Ее $n-1$ точка пересечения с остальными сферами после перестройки соединяет каждую из этих остальных сфер с первой, так что все вместе они образуют опять диффеоморфную сфере поверхность Σ^2, не считая лишь оставшихся $(n-2)+(n-3)+\ldots+1=(n-1)(n-2)/2$ точек пересечения остальных сфер между собой (рис. 9).

После замены $(n-1)(n-2)/2$ точек пересечения поверхности Σ^2 с собой таким же числом трубочек мы превратим «сферу» Σ^2 в (гладкую)

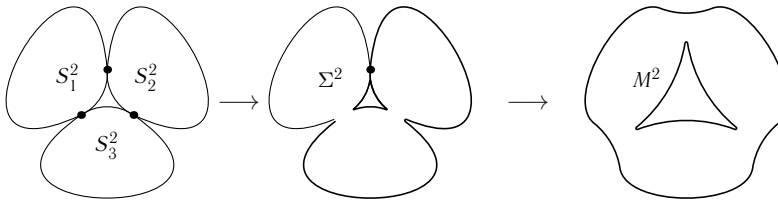


Рис. 9. Построение гладкой поверхности из набора n сфер

поверхность M^2 сферы с ручками, число ручек g — это число трубочек, заменивших точки самопересечения «сферы» Σ^2, так что $g=(n-1)(n-2)/2$.

На рисунке 9 выше (соответствующем $n=3$) получается $g=1$, так что поверхность M^2 рода 1 представляет собой тор.

Так мы получили *формулу Римана—Гурвица* $g=(n-1)(n-2)/2$.

Неравенство Харнака, утверждающее, что вещественная кривая рода g имеет не больше $g+1$ овала, является частным случаем *неравенства Смита*

$$\sum b_k(M_{\mathbb{R}}) \leqslant \sum b_k(M_{\mathbb{C}}). \tag{2}$$

Здесь $M_{\mathbb{C}}$ — комплексное алгебраическое многообразие (например, риманова поверхность кривой). Если это многообразие задается уравнением с вещественными коэффициентами, то на нем действует симметрия («инволюция») $\sigma\colon M_{\mathbb{C}}\to M_{\mathbb{C}}$ комплексного сопряжения (переводящая точку с комплексными координатами z_j в точку с комплексными координатами $\bar z_j=x_j-iy_j$ при $z_j=x_j+iy_j$). Очевидно, $\sigma^2=1$, и вещественное многообразие $M_{\mathbb{R}}$ состоит из неподвижных точек инволюции σ (овалов в случае кривой).

Числа b_k в неравенстве (2) — это «числа Бетти» для цепей с коэффициентами в группе $\mathbb{Z}_2$ из двух элементов.

Для окружности числа Бетти имеют вид $b_0=b_1=1$, $b_{k>1}=0$.

Для римановой поверхности рода g имеем

$$b_0=b_2=1,\qquad b_1=2g,\qquad b_{k>2}=0.$$

Здесь $2g$ одномерных циклов — это «параллели» и «меридианы» g ручек.

Неравенство Смита имеет поэтому в случае кривых рода g вид

$$2(\text{число овалов})\leqslant 2g+2,$$

то есть получается неравенство Харнака:

$$\text{число овалов}\leqslant g+1.$$

Само неравенство Смита доказать не очень трудно, рассматривая действие инволюции σ на всевозможные цепи (симметричной относительно инволюции σ триангуляции многообразия). В случае римановой поверхности вещественной кривой наиболее важное соображение теории Смита состоит в том, что между ее овалами может существовать *только одно* гомологическое соотношение (сумма овалов гомологична нулю) в одномерных гомологиях римановой поверхности, иначе эта поверхность не была бы связной, а распадалась бы на связные двумерные компоненты (образованные частями вида a и σa, где двумерная цепь a имеет границей левую часть соотношения между овалами).

К этим замечаниям из вещественной алгебраической геометрии добавлю еще, что, кроме графиков многочленов, Адриана Ортиц-Родригес рассматривала в своей диссертации и параболические кривые на любых алгебраических поверхностях степени n в трехмерном вещественном проективном пространстве $\mathbb{R}P^3$. В этом случае число параболических кривых оценено ею сверху и снизу величинами an^3 и bn^3, причем постоянная a больше постоянной b примерно в 10 раз.

Я формулирую здесь этот результат потому, что надеюсь на слушателей, которые захотели бы найти точную скорость роста числа параболических кривых, сблизив a и b.

Результаты Гудкова о кривых степени 6 послужили основой замечательной новой теории, связавшей вещественную алгебраическую геометрию 16-й проблемы Гильберта с квантовой теорией поля и с многомерной топологией.

Числа внутренних овалов в списке Гудкова кривых 6-й степени с 11 овалами (1, 5 или 9) недаром идут через 4. Переходя от кривой $f(x, y) = 0$ к ограниченной поверхности с краем M: $f(x, y) \geqslant 0$, мы приходим к следующим через 8 эйлеровым характеристикам.

Перебирая найденные в диссертации Гудкова вещественные проективные алгебраические кривые степени $n = 2k$, имеющие наибольшее возможное по теореме Х. Гарнака число овалов, я заметил, что для них эйлеровы характеристики поверхностей M удовлетворяют сравнениям

$$\chi(M) \equiv k^2 \pmod 8, \tag{3}$$

которые я назвал «*сравнениями Гудкова*».

Сравнения по модулю 8 (для сигнатур форм пересечений) являются стандартным в топологии четырехмерных замкнутых многообразий, поэтому я стал искать четырехмерное многообразие в топологии (*одномерных*) вещественных алгебраических кривых.

Таким многообразием оказалась *комплексификация поверхности с краем* M^2. Чтобы комплексифицировать поверхность, заданную неравен-

ством $f(x, y) \geqslant 0$, я записал это геометрическое неравенство в алгебраическом виде $f(x, y) = z^2$. Эта формула определяет (в комплексной области) четырехмерное в вещественном смысле многообразие: двулистное накрытие дополнения к римановой поверхности (комплексной кривой) $f(x, y) = 0$ в $\mathbb{C}P^2$, разветвленное вдоль этой римановой поверхности.

Применяя к этому четырехмерному многообразию топологические результаты о делимости сигнатур на 8, я доказал сравнение (3) по модулю 4, а затем Рохлин, применив более глубокие результаты дифференциально-топологического исследования гладких четырехмерных многообразий, доказал и само сравнение Гудкова (3).

Интересно, что сам Гудков, которому я сообщил об этом сравнении, когда писал отзыв на его диссертацию, считал его неверным, так как, якобы, располагал контрпримерами к нему (которые, однако, оказались столь же ошибочными, как и результат Гильберта о кривых степени 6, опровергавшийся этой диссертацией).

К настоящему времени сравнение (3) стало основой большого количества новых результатов и в вещественной алгебраической геометрии, и в дифференциальной топологии и даже в квантовой теории поля. Но, к сожалению, даже для классификации топологических структур кривых степени 8 в 16-й проблеме Гильберта этих результатов не хватает.

Возвращаясь к 16-й проблеме, замечу, что Гильберт, по-моему, пропустил в этой задаче самые главные вопросы.

Дело в то, что топологические структуры могут различаться не только у вещественных алгебраических кривых (данной степени) $\{f(x, y) = 0\}$, но и у многочленов $f : \mathbb{R}^2 \to \mathbb{R}$, которыми эти кривые задаются.

Гильберту следовало бы включить в формулировку своей проблемы не только вопрос о топологической классификации проективных вещественных плоских алгебраических кривых данной степени, но и вопрос о *топологической классификации самих многочленов, задающих эти кривые*.

Этот вопрос не решен, насколько мне известно, уже для кривых степени $n = 4$ (где кривые расклассифицировал еще Декарт). Я обсужу теперь этот вопрос о топологической классификации гладких функций и многочленов, где многое до сих пор неизвестно (отчасти по вине Гильберта).

§2. Статистика гладких функций

Чтобы описать топологическую структуру гладкой вещественной функции, сопоставим ей граф, точками которого являются связные компоненты гиперповерхностей уровня этой функции.

Для невырожденной «функции Морса» $f : S^n \to \mathbb{R}$, $n > 1$, такой граф оказывается деревом, имеющим T тройных точек ветвления, $K = T + 2$ кон-

цевых точек и $P = 2T + 1$ ребро, соединяющие $K + T = 2T + 2$ вершины графа.

Пример 1. Для горы Эльбрус функция (высота) имеет две точки локального максимума, A и B, и одну седловую точку C, так что граф имеет вид буквы Y (рис. 10).

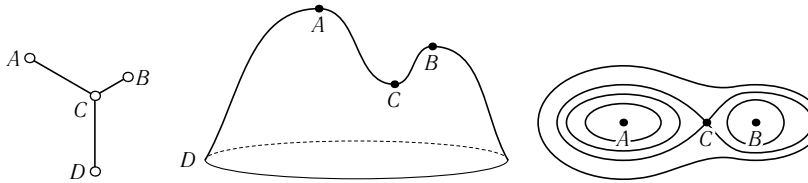


Рис. 10. Граф и линии уровня горы Эльбрус

Мы будем изучать функции $f\colon \mathbb{R}^n \to \mathbb{R}$, ведущие себя как $-r$ вдали от начала координат, продолжая их вблизи точки $\infty = S^n \setminus \mathbb{R}^n$ так, чтобы она была точкой локального минимума (предоставленного выше концевой вершиной D дерева).

Пример 2. Для горы Везувий получаем картинку (рис. 11).

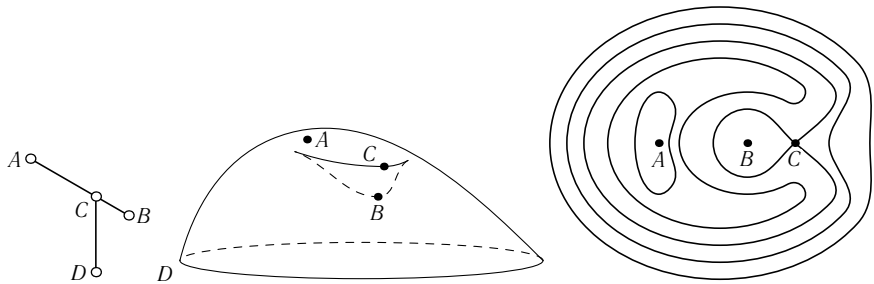


Рис. 11. Граф и линии уровня горы Везувий

В структуру графа функции мы будем включать упорядочение его вершин по высоте ($f(A) > f(B) > f(C) > f(D)$ для Эльбруса, $f(A) > f(C) > f(B) > f(D)$ для Везувия), различая графы примеров 1 и 2, хотя эти деревья и гомеоморфны.

Для простоты мы будем считать, что все $2T + 2$ критические значения функции f различны. Для функций на сфере S^2 из этих значений T соответствуют седлам, а $T + 2$ — максимумам и минимумам. Графы функций Морса на сферах S^n, $n > 2$, похожи на графы для $n = 2$ и тоже являются

деревьями, но мы применим позже развитую для изучения этих деревьев технику и к случаю функций на торе $T^2 = S^1 \times S^1$, где графы имеют циклы.

Нашей основной целью будет исследование статистики графов функции (с упорядоченными по высоте вершинами), имеющих T тройных точек: какие из этих упорядоченных деревьев реализуются многочленами соответствующей значению T степени?

Типичный многочлен степени n от двух переменных имеет не более $(n-1)^2$ критических точек на плоскости $\mathbb{R}^2$, это соответствует значениям $2T + 2 = (n-1)^2 + 1$, то есть $T = 2k(k-1)$ тройных вершин графа (седел функции) для многочленов степени $n = 2k$.

Теорема 1. *Числа* $\varphi(T)$ (*упорядоченных*) *графов функций* (*деревьев с* T *тройными точками при* $T \leqslant 4$) *суть*

T	1	2	3	4
$\varphi(T)$	2	19	428	17746

Упорядочения графов функций обладают следующим свойством *правильности*: среди трех соседей любой вершины ветвления есть и более высокие (1 или 2) и более низкие (2 или 1) вершины.

Это вытекает из того, что графы Эльбруса и Везувия (рис. 9 и рис. 10) упорядочены правильно, а топологическое строение функции около седловой критической точки всегда либо такое, как у Эльбруса, либо такое, как у Везувия.

Скорость роста числа $\varphi(T)$ правильно упорядоченных графов с ростом числа седел T оценивают следующие два результата.

Теорема 2. *Число* $\varphi(T)$ *правильно упорядоченных деревьев* (*графов функций*) *с* T *тройными вершинами не меньше, чем следующая оценка снизу*:

$$\varphi(T) \geqslant \frac{(T^2 + 5T + 5)(2T + 2)!}{(T + 4)!}.$$

При $2 \leqslant T \leqslant 4$ правая часть этого неравенства имеет, соответственно, значения 19, 232, 3690. Оценка снизу скорости роста φ снизу «по Стирлингу» дает величину $4(4/e)^T T^T > T^T$.

Теорема 3. *Число* $\varphi(T)$ *правильно упорядоченных деревьев с* T *тройными вершинами не больше, чем следующая оценка сверху*:

$$\varphi(T) \leqslant T^{2T}, \quad \textit{если } T > 2.$$

Основу доказательства теоремы 2 составляет прямой подсчет числа тех специальных упорядоченных деревьев, для которых T тройных точек

составляют монотонную A-цепь, с критическими значениями

$$f(A_1) > f(A_2) > \ldots > f(A_T)$$

в соседних вершинах графа: $A_1 — A_2 — \ldots — A_T$.

Здесь и всюду дальше мы будем считать функцию, граф которой мы изучаем, заданной и на этом графе: значение новой функции в каждой точке графа равно значению исходной функции в каждой точке той гиперповерхности уровня, компонентой связности которой эта точка графа является.

Теорема 4. *Число правильных упорядочений графа — дерева — с T тройными точками, образующими упорядоченную A-цепь, равно*

$$\psi(T) = \frac{(T^2 + 5T + 5)(2T + 2)!}{(T + 4)!}.$$

Теорема 2 вытекает из теоремы 4, так как число $\varphi(T)$ всех правильных упорядочений не меньше числа $\psi(T)$ тех правильных упорядочений, в которых тройные точки образуют монотонную A-цепь.

Замечание. Некоторые из наших правильно упорядоченных графов являются графами *многочленов* (степени $n = 2k$ при $T = 2k(k-1)$), а некоторые — не являются.

Было бы интересно узнать, будет ли число реализуемых многочленами правильных упорядочений мало по сравнению с числом всех реализуемых гладкими функциями упорядочений, или, может быть, относительно малым окажется, напротив, число нереализуемых многочленами правильных упорядочений (асимптотически, при $T \to \infty$).

Число топологически разных реализаций реализуемого графа также интересно. Здесь следовало бы рассмотреть и вопрос о классификации топологически разных реализующих гладких функциях Морса, и вопрос о числе компонент связности в пространстве реализующих многочленов данной степени (которое может оказаться бóльшим единицы даже и в том случае, когда все эти реализующие многочлены топологически друг другу эквивалентны).

Оба вопроса открыты, и я ожидаю достижений от слушателей.

Доказательство теоремы 4. Обозначим через a ту конечную вершину, соседнюю с тройной вершиной A_1 в графе, где критическое значение максимально, $f(a) > f(A_1)$.

Точно так же, обозначим через z ту конечную вершину, соседнюю с тройной вершиной A_T в графе, где критическое значение минимально, $f(z) < f(A_T)$.

Третью соседнюю конечную вершину тройной вершины A_1 в графе мы обозначим (рис. 12) через α (она отлична от a и от A_2). Точно также обо-

значим через ω третью конечную вершину соседнюю в графе с A_T (она отлична от z и от A_{T-1}).

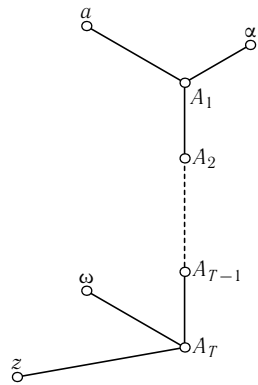


Рис. 12. Оснащение концевых вершин A_1 и A_T цепочки

Чтобы расклассифицировать оснащения тройных вершин $A_1, \ldots, A_T$ графа присоединяемыми в них ребрами, ведущими в конечные вершины графа, заметим прежде всего, что критическое значение $f(\alpha)$ принадлежит дополнению к следующему множеству из $(T+1)$-го вещественного числа, меньшего, чем $f(a)$:

$$\{f(A_1), \ldots, f(A_T); f(z)\}.$$

Следовательно, имеется $T+1$ топологически различный случай, когда $f(\alpha) > f(z)$, и еще один отличный топологически от них случай, когда $f(\alpha) < f(z)$.

Зная интервал, в котором расположено критическое значение $f(\alpha)$, рассмотрим критическое значение $f(\omega) > f(z)$. Оно должно отличаться от $T+2$ значений

$$\{f(A_1), \ldots, f(A_T); f(a), f(\alpha)\}$$

на луче $\{t > f(z)\}$, если $f(\alpha) > f(z)$ (будучи отличным от $T+1$ значения

$$\{f(A_1), \ldots, f(A_T); f(a)\}$$

на луче $\{t > f(z)\}$, если $f(\alpha) < f(z)$).

Итого мы находим

$$(T+1)(T+3) + 1(T+2) = T^2 + 5T + 5$$

топологически различных типов оснащений (α, ω) *тройных вершин* A_1 и A_T.

В конечной вершине a_2, соединенной в графе ребром с тройной вершиной A_2, критическое значение должно отличаться от $T+4$ уже выбранных значений,

$$\{f(A_1), \ldots, f(A_T); f(a), f(\alpha), f(z), f(\omega)\},$$

что подразделяет каждый из изученных выше случаев на $T+5$ подслучаев. Выбрав $f(a_2)$, мы получаем $T+5$ препятствий для выбора $f(a_3)$ и т. д., для выбора $f(a_i)$ число препятствий равно $T+2+i$: нужно избегать все значения

$$\{f(A_1), \ldots, f(A_T); f(a), f(z), f(\alpha), f(\omega); f(a_2), \ldots, f(a_{i-1})\}.$$

Эти препятствия разделяют вещественную ось на $T+3+i$ интервала, умножая число подслучаев в описываемой нами классификации на множитель $T+3+i$.

Повторяя это рассуждение $T-2$ раз (при $i=2, 3, \ldots, T-1$), мы подразделим каждый из T^2+5T+5 случаев оснащения концевых тройных вершин A_1 и A_T на много подслучаев, число которых равно произведению чисел интервалов на последовательных шагах нашей конструкции,

$$(T+5)(T+6)\ldots(T+3+T-1) = \frac{(2T+2)!}{(T+4)!}.$$

Все эти подклассы доставляют все топологически различные оснащения, и каждое встречается по одному разу, что и доказывает теорему 4. □

Чтобы доказать теорему 3, мы начнем со следующего (индуктивного) предложения.

Лемма. *Для любого $T \geqslant 2$ имеет место неравенство*

$$\varphi(T) \leqslant 4T^2\varphi(T-1).$$

Пример. При $T=2$, 3 и 4 мы находим (прямыми подсчетами)

$$(\varphi(2)=19) < (16\cdot 2 = 32);$$
$$(\varphi(3)=428) < (36\cdot 19 = 684);$$
$$(\varphi(4)=17746) < (64\cdot 428 = 27392).$$

В этих случаях утверждение леммы справедливо.

Доказательство леммы. Максимальное критическое значение достигается в одной из конечных вершин A связного графа с T тройными вершинами. Эта концевая вершина соединена ребром с одной из тройных вершин, B. Выкинув ребро AB, мы уменьшим исходный граф с T тройными вершинами до меньшего связного графа с $T-1$ тройной вершиной.

Таких меньших (правильно упорядоченных) графов $\varphi(T-1)$ штук. Чтобы восстановить исходный больший правильно упорядоченный граф, нужно выбрать в меньшем графе ребро, поместить на него новую тройную вершину B и соединить ее ребром с новой концевой вершиной A, находящейся выше всех остальных.

Для этих выборов мы располагаем $2(T-1)+1=2T-1$ ребрами меньшего графа, куда поместится вершина B. Значение в выбранной вершине B должно отличаться от $2T$ значений в вершинах меньшего графа, что доставляет $2T+1$ вариант (разных топологических типов).

Общее число вариантов обоих выборов равно $(2T-1)(2T+1)=$ $=4T^2-1<4T^2$, откуда и получается неравенство леммы 3. □

В действительности мы доказали больше, чем утверждение леммы 3, оценив сверху не только число графов функций, $\varphi(T)$, но и большее число, считающее и такие «неправильные» упорядочения вершин деревьев, у которых высота какой-либо тройной точки выше всех трех высот соседних в графе вершин (или ниже всех трех). Такого не может быть в упорядоченном графе функции, число которых, поэтому, меньше оцененного нами числа.

Доказательство теоремы 3. При $T=3$ мы имеем

$$(\varphi(3)=428)<(3^6=729).$$

Если неравенство теоремы 3 справедливо для $T=S-1$, то мы получаем из леммы неравенство

$$\varphi(S)\leqslant 4S^2(S-1)^{2S-2}. \qquad (*)$$

Используя очевидное неравенство

$$\left(1-\frac{1}{S}\right)^S<\frac{1}{e},$$

мы находим оценку правой части неравенства $(*)$:

$$4S^2(S-1)^{2S-2}\leqslant\frac{4S^2}{(S-1)^2}\left(\frac{S-1}{S}\right)^{2S}S^{2S}\leqslant\frac{4}{e^2}\frac{S^2}{(S-1)^2}S^{2S}.$$

Коэффициент $4S^2/(e^2(S-1)^2)$ меньше 1 при $2S\leqslant e(S-1)$, что выполняется при $S\geqslant 4$.

Иначе, если $S\geqslant 4$, то неравенство теоремы 4 для $T=S-1$ вместе с неравенством $(*)$, доставляет неравенство $\varphi(S)\leqslant S^{2S}$ теоремы 3, которая, тем самым, последовательно доказывается для $T=4,5,\ldots$. □

Доказательство теоремы 1. Рассмотрим T тройных точек дерева с $2T+2$ вершинами. Они образуют множество вершин связного подграфа, получающегося из исходного графа с T тройными точками выкидыванием $T+2$ ребер, соединяющих их с его концевыми вершинами.

При $T = 1$, 2 или 3 остающиеся (упорядоченные) графы имеют вид, изображенный на рис. 13.

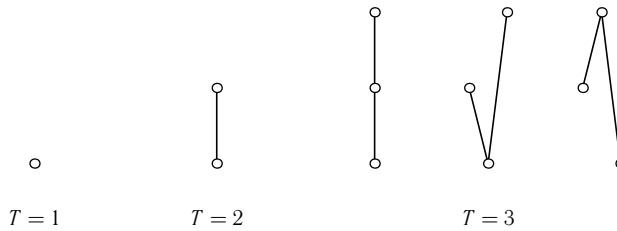


Рис. 13. Укороченные графы деревьев с $T \leqslant 3$ тройными вершинами

Оснащения пересчитываются, как в доказательстве теоремы 3 в первых трех случаях, доставляя

$$\varphi(1) = 2, \quad \varphi(2) = 4 + 5 \cdot 3 + 5 = 19,$$
$$\psi(3) = (9 + 5 \cdot 3 + 5) \cdot 8 = 232$$

оснащений соответственно.

Каждый из оставшихся двух случаев разбирается аналогично доказательству теоремы 4 выше, доставляя в каждом случае 98 оснащений.

Равенство чисел оснащений этих двух упорядоченных графов заранее очевидно из-за симметрии (отражение в горизонтальной оси), переводящий один из этих упорядоченных графов в другой.

В случае $T = 4$ тройных вершин приходится разбирать 9 существенно разных случаев (рис. 14).

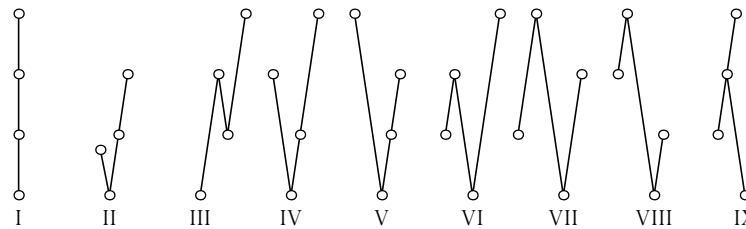


Рис. 14. Укороченные графы деревьев с $T = 4$ тройными вершинами

Случаи II, IV, V, VI, IX (каждый) представляют по два упорядоченных графа, симметричных относительно горизонтальной оси (подобно двум последним графам, рассмотренным выше, при $T = 3$).

Эти симметричные друг другу упорядоченные графы различны, но числа их оснащений одинаковы (из-за их симметрий, продолжаемых до симметрий оснащений симметричных графов).

Числа ψ оснащений в этих 9 случаях даются таблицей

случай	I	II	III	IV	V	VI	VII	VIII	IX
ψ	3690	1680	586	1360	1360	534	486	756	1180

Самый большой случай, I, исследован выше, в доказательстве теоремы 2. Подсчеты чисел оснащений в остальных случаях следуют тому же методу, но детали слишком многочисленны, чтобы все их помещать в эту лекцию: я предпочитаю рассматривать их как естественное упражнение к нашему курсу.

Общее число оснащений всех укороченных графов, с учетом симметричных пар случае, доставляется суммированием:

$$\varphi(4) = \psi(\mathrm{I}) + \psi(\mathrm{III}) + \psi(\mathrm{VII}) + \psi(\mathrm{VIII}) + \\ + 2(\psi(\mathrm{II}) + \psi(\mathrm{IV}) + \psi(\mathrm{V}) + \psi(\mathrm{VI}) + \psi(\mathrm{IX})) = 5518 + 2 \cdot 6114 = 17746,$$

что и доказывает теорему 1 (при $T = 4$). □

Будет ли аналогичный I случай наибольшим при любом $T > 4$, я не знаю.

Замечание. Было бы интересно исследовать, как именно растет на самом деле величина $\varphi(T)$ с ростом T: вероятно, она растет в основном как T^{cT} (в пренебрежении «логарифмическими» поправками вроде const^T). Постоянная c (заключенная между 1 и 2 по теоремам 2 и 3) кажется эмпирически более близкой к 2 (а может быть, и равна 2, в пренебрежении «логарифмическими» сомножителями)[1].

Приведенные выше вычисления функций ψ для девяти случаев ($T = 4$) подсказывают некоторую однородность распределения всех оснащенных графов между описанными девятью классами. Например, для специального типа подграфов с T тройными вершинами с критическими значениями

$$\mathrm{II}(T) = \{a_1 > a_2 > \ldots > a_{T-1};\ a_1 > b > a_2\}$$

в последовательных вершинах $(B, A_1, A_2, \ldots, A_{T-1})$ асимптотики чисел оснащений таковы, что

$$\frac{\psi(\mathrm{II}(T))}{\psi(\mathrm{I}(T))} \to \frac{1}{2} \quad \text{при} \quad T \to \infty, \quad \text{так что}$$

$$\psi(\mathrm{II}(T)) \sim \frac{T^2(2T+2)!}{(2(T+4)!)}.$$

[1] Эту гипотезу в 2006 доказал Л. Николаеску, ознакомившийся с настоящей лекцией 2005 года (Functional analysis and other mathematics. 2006. V. 1, № 1).

Интересно было бы вычислить подобные соотношения между асимптотиками чисел оснащений других укороченных графов: как они зависят от геометрии укороченного (оснащенного) графа и почему появляется отношение $1/2$ в случае пары I, II? Всегда ли подобные отношения будут стремится к рациональным числам? Меньше ли они единицы?

Я даже надеюсь, что из подобных соображений можно будет извлечь оценку снизу

$$\varphi(T) \geqslant BT^2\varphi(T-1),$$

с какой-либо постоянной B, и даже, может быть, с $B = 2$.

Эта надежда объясняется следующим (нестрогим) «физическим» рассуждением. Мы разобрали выше $4T^2 - 1$ способ добавить новое ребро, ведущее к самой высокой критической точке. Бóльшая часть этих способов действительно доставляет новые графы с одной новой тройной вершиной, произвольно выбранной на одном из $2T - 1$ ребер меньшего графа с $T - 1$ тройной вершиной. Трудность составляет только выбор критического значения в новой вершине. Для этого выбора мы располагаем $2T + 1$ интервалом, но при этом новое критическое значение w не должно оказаться меньшим, чем оба значения u и v в концевых вершинах того ребра, где выбрана новая вершина.

Беда в том, что (см. рис. 15) третий сосед новой тройной вершины — это самая высокая вершина нового упорядоченного графа, поэтому, если $u > w$ и $v > w$, то критическое значение w окажется меньшим, чем все три критических значения в соседних вершинах графа (чего в упорядоченном графе хорошей функции Морса никогда не бывает).

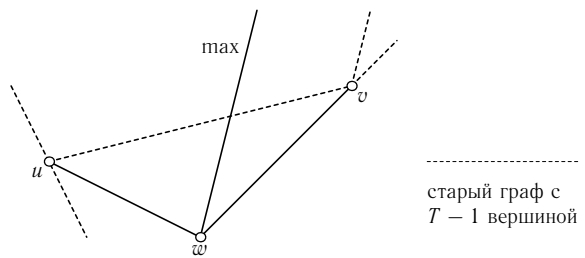


Рис. 15. Невозможный выбор нового критического значения w

Постараемся оценить, насколько это препятствие уменьшает число реализуемых бóльших упорядоченных графов, построенных из данного укороченного графа.

Эвристическое (нестрогое) рассуждение подсказывает «вероятность» порядка $1/4$ для события $\{w < u,\ w < v\}$, так как каждое из неравенств имеет, по-видимому, «вероятность» $1/2$ и они кажутся независимыми.

Если это так, то из общего числа $4T^2 - 1$ бóльших графов придется отбросить четверть, для которой не существует функций с такими графами. Использованная в этом нестрогом выводе асимптотическая «эргодичность» (распределения значений в вершинах случайного графа) — трудная гипотеза. Кроме доказательства, она может исследоваться эмпирически, путем экспериментального перечисления всех графов (скажем, с $T = 5, 6, 7, 8$ тройными точками), для чего потребовалось бы всего несколько часов работы компьютера.

Приняв «вероятность» $1/4$, мы заменим коэффициент $B = 4$ на $B = 3$. Но числа теоремы 1 подсказывают даже меньшее значение предела (при T стремящемся к бесконечности) отношения

$$\frac{\varphi(T)}{T^2\varphi(T-1)},$$

которое ближе к 2, чем к 3.

Это меньшее значение коэффициента B может быть объяснено (эвристически, по меньшей мере) следующим препятствием к построению большего упорядоченного графа (см. рис. 16).

Предположим, что новое значение w выше значений в обоих концах ребра, на котором выбрана новая тройная вершина: $w > u$, $w > v$.

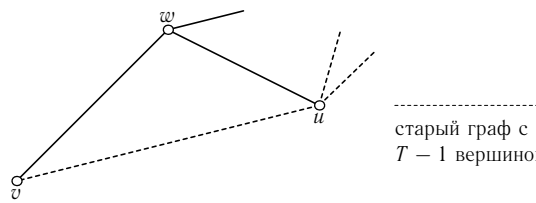


Рис. 16. Невозможный выбор нового критического значения w

Если при этом значения расположены в порядке

$$v < u < w,$$

то выбор значения w в новой тройной вершине может испортить старую тройную вершину, где достигается значение u: это случится, если в обеих соседних с ней в старом графе вершинах достигались бóльшие u значения. В такой ситуации в новом упорядоченном графе значение u (в старой тройной вершине нового графа) будет меньше значений во всех трех ее

соседних (в новом графе) вершинах, что невозможно для графа функции Морса.

Вычисляя опять «вероятность» нового препятствия, мы должны будем заменить коэффициент $B = 3$ на его три четверти, получая $B = 9/4$. Это (полуэмпирическое) предположение не слишком далеко от наблюденных значений: отношение

$$\frac{(\varphi(T = 4) = 17746)}{(\varphi(3) = 428)}$$

не сильно отличается от величины произведения

$$(B = 9/4)(T = 4)^2 = 36.$$

Разумеется, эти полуэмпирические выводы следовало бы подкрепить не только доказательством (которое может быть вовсе не простым) сформулированных выше гипотетических утверждений эргодической теории случайных графов (приводящих к $B = 9/4$), но и численным экспериментом (для которого нужно либо вычислить несколько следующих значений $\varphi(T)$, либо прямо перечислять или даже лишь предъявлять случайные упорядочения случайных деревьев, что должно бы требовать меньшего компьютерного времени).

Так или иначе, с какой бы постоянной B не получалось неравенство

$$\varphi(T) > BT^2\varphi(T - 1),$$

оно привело бы к росту величины $\varphi(T)$, подобному росту величины T^{2T} (пренебрегая зависящими от коэффициента B «логарифмически малыми» по сравнению с T^{2T}) множителями вроде const^T.

В действительности верхняя оценка теоремы 3 доказана выше для большего, чем $\varphi(T)$, числа упорядоченных графов, включая и не реализуемые как упорядоченные графы функций (не реализуемые графы допускают такие тройные вершины, которые выше всех трех соседних в графе вершин или ниже всех трех, чего в упорядоченных графах функций не бывает).

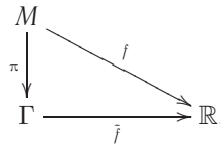
Чтобы извлечь из доказательств этих оценок сверху оценку снизу, нужно было бы знать «вероятности» описанных выше неприятностей, но их вычисление требует трудных результатов эргодической теории случайных упорядочений случайных графов, которых я не смог получить (ни доказать, ни подтвердить численным экспериментированием).

Именно в надежде на успехи слушателей в этой новой и неизведанной области я включил приведенную полуэмпирическую теорию в настоящую лекцию.

Кроме топологической классификации упорядоченных графов функций, интересен и вопрос о топологической классификации самих функций (с данным графом Γ) на фиксированном многообразии M.

В случае двумерного многообразия (например, сферы S^2) упорядоченный граф (с данными критическими значениями) определяет, по-видимому[1], топологический тип функции Морса. При большей размерности положение сложнее и соответствующий вариант нашей комбинаторной теории не построен: неясно, сколько типов функций Морса на S^3 соответствует зданному графу, даже если фиксировать индексы Морса и критические значения, соответствующие вершинам графа.

Функция f на M получается из функции $\tilde{f}$ на своем графе Γ при помощи естественного отображения $\pi\colon M \to \Gamma$, сопоставляющего каждой точке области определения функции компоненту множества уровня, содержащую эту точку области определения:



Топологические (например, гомотопические) инварианты естественного отображения π доставляют интересные топологические инварианты функций f (с данным графом Γ), и было бы интересно узнать, каковы они (и какие отображения π реализуются гладкими функциями f на данном многообразии M).

Например, этот вопрос, кажущийся легким для $M = S^2$, интересен для тригонометрических многочленов от двух переменных, $f\colon T^2 \to \mathbb{R}$. Соответствующий граф Γ имеет один цикл (g циклов для поверхности M^2 рода g). Доказать это свойство графа Γ — полезное упражнение на подсчет эйлеровой характеристики:

$$\chi(M^2) = 2 - 2g,$$

$$\chi(\Gamma) = (\text{число вершин}) - (\text{число ребер}) =$$
$$= (T + K) - (3T + K)/2 = 1 - (\text{число циклов в } \Gamma).$$

Для тригонометрического многочлена f данной степени n топологическая сложность отображения $\pi\colon T^2 \to \Gamma$ должна, видимо, оцениваться некоторой (пока не найденной) функцией от n. Было бы интересно узнать, сколько разных гомотопически классов отображений π реализуется тригонометрическими многочленами данной степени n (или с данным спектром, являющимся конечным подмножеством в решетке волновых векторов, $\mathbb{Z}^2$)[2].

[1] Строгое доказательство следовало бы опубликовать: оно, кажется, не опубликовано.

[2] Первые результаты в этом направлении опубликованы в 2006 году: Арнольд В. И. Статистика и классификация топологий, периодических функций и тригонометрических много-

§3. Статистика и топология периодических функций и тригонометрических многочленов

Функция Морса $f\colon T^2 \to \mathbb{R}$ с T седловыми критическим точками и K точками максимума и минимума имеет граф с T (тройными) точками ветвления и $K = T$ концевыми точками. Этот граф (см. рис. 17) имеет $p = 2T$ ребер и один цикл (оснащенный примыкающими к его точкам деревьями).

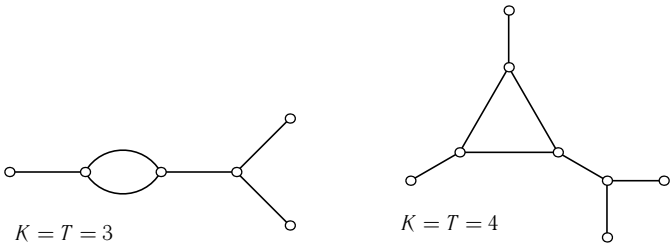


Рис. 17. Графы с одним циклом и T тройными вершинами

Пример. Рассмотрим четырехпараметрическое семейство тригонометрических многочленов

$$f_{A,B,C,D}(x,\, y) = A\sin x + B\sin y + C\sin(x+y) + D\cos(x+y). \qquad (1)$$

Мы будем считать их функциями на торе,

$$T^2 = \{x \pmod{2\pi},\ y \pmod{2\pi}\}.$$

Как мы увидим, максимальные числа их критических точек (для невырожденных функций Морса) суть

$$K = T = 4,$$

так что числа вершин и ребер графа суть

$$B = 8, \quad P = 8$$

(на самом деле максимумы меньше, $K = T = 3$).

Подсчет показывает, что *число упорядоченных графов гладких функций на торе с такими значениями параметров есть* 550, *причем многочлены* (1) *доставляют не более двенадцати из этих* 550 *правильно упорядоченных графов*.

членов // Труды Института Математики и Механики УрО РАН. 2006. Т. 12, № 1. 10 с. Arnold V. I. Topological Classification of trigonometric polynomials related to affine Coxeter group A_2. The Abdus Salm International Centre for Theoretical Physics, ICTP. IC/2006/039. 15 pp. `http://www.ictp.it/~pub_off`.

Гипотетически при повышении степеней тригонометрических многочленов реализуемая ими часть графов будет составлять все меньшую долю множества всех правильно упорядоченных графов с таким же числом вершин, и я надеюсь, что слушатели сумеют продвинуться в направлении доказательства этой гипотезы.

Наряду с классификацией упорядоченных графов тригонометрических многочленов данной степени интересен и вопрос о топологической классификации самих многочленов. Более того, даже топологически эквивалентные тригонометрические многочлены данной степени или с данным спектром могут образовывать несколько связных областей в пространстве всех таких тригонометрических многочленов, и изучить топологию этих областей (в частности, их число) — интересный вопрос, который я здесь упоминаю ради того, чтобы слушатели приняли участие в его решении.

При этом кроме пространства всех тригонометрических многочленов данной степени интересно также исследовать более общее пространство тригонометрических многочленов с любым данным спектром S

$$f(z) = \sum_{k \in S} f_k e^{i\langle k, x\rangle}, \quad f_k \in \mathbb{C},\ z \in \mathbb{C}^m.$$

Здесь «волновые векторы» $k \in S \subset \mathbb{Z}^m$ гармонических волн на m-мерном торе принадлежат конечному «спектру» S. Чтобы формула задавала вещественный тригонометрический многочлен $f\colon T^m \to \mathbb{R}$, коэффициенты должны удовлетворять условию вещественности ($f_{-k} = \bar{f}_k$). Напомню, что знаком $\langle \cdot, \cdot \rangle$ выше обозначено эрмитово скалярное произведение в пространстве $\mathbb{C}^m$ (заданное формулой $\langle k, z\rangle = \sum_{j=1}^{m} (k_j \bar{z}_j)$ для векторов k и z с компонентами k_j и z_j, где $\bar{z} = x - iy$ при $z = x + iy$, т. е. знак черты означает комплексное сопряжение).

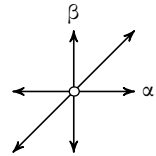
Я предполагаю, что слушатели знают *основы теории гармонических волн* из школьного курса физики (или из элементарного учебника Ландау и Лифшица), так что понимают, что предыдущие комплексные формулы можно переписать в вещественном виде

$$\sum_k (a_k \cos(k, x) + b_k \sin(k, x)),$$

где $x \in \mathbb{R}^m$, $k \in S$, $(k, x) = \sum (k_j x_j)$ — евклидово скалярное произведение.

В качестве спектров особенно интересных тригонометрических многочленов можно брать «расширенные системы корней аффинных групп отражений» (которые я здесь не буду описывать). Замечу только, что для простейшей группы A_2 (соответствующей симметриям правильного треуголь-

ника) соответствующий спектр состоит из 6 векторов ($\pm\alpha$, $\pm\beta$, $\pm(\alpha+\beta)$) двумерной плоскости.



Соответствующие этому спектру тригонометрические многочлены имеют вид (1). Они образуют не шестипараметрическое, а четырехпараметрическое семейство, так как на самом деле нужны были бы еще комбинации функций $\cos x$ и $\cos y$, но от них можно избавиться выбором в качестве начала координат нужной точки тора T^2, так что общий случай члена шестипараметрического семейства сводится к случаю (1) сдвигом начала координат.

Перенесение излагаемой ниже для случая простейшего семейства (1) теории на случай тригонометрических многочленов более высоких степеней (или связанных с более общими системами корней, или даже с любыми спектрами) — одна из цепей включения описываемой ниже элементарной теории в настоящую лекцию.

Здесь, несомненно, возможен быстрей прогресс (с интересными новыми научными результатами), не требующий каких-либо предварительных знаний. Это — задачи второго типа в классификации Пуанкаре.

Обратимся теперь к классификации графов функций Морса на торе T^2.

Определение. *Правильно упорядоченным графом* назовем граф с T тройными вершинами и K концевыми вершинами, которым приписаны различные значения так, что ни в какой тройной вершине значение не меньше, чем во всех трех соседних в графе вершинах и не больше, чем во всех трех соседних в графе вершинах. Мы будем называть упорядочивающие значения «высотами» вершин.

Рис. 18. Правильные и неправильные упорядочения вершин графа Y

Точные значения высот вершин в определение упорядочения графа не входят: два набора высот вершин считаются задающими одинаковые упорядочения, если одна вершина ниже другой в обоих наборах одновременно.

Теорема 1. *Число топологически разных правильных упорядочений графов с одним циклом и* $T = 4$ *тройными вершинами* ($K = 4$ *концевыми точками и* $P = 8$ *ребрами*) *равно* 550.

Доказательство. Такой граф имеет ровно один цикл, состоящий из 2, 3 или 4 ребер. Точки ветвления графа образуют одну из 11 конфигураций A—K рисунка 19. Здесь и далее мы обозначаем точки ветвления квадратиками, а концевые точки (отсутствующие на рис. 19) — маленькими окружностями.

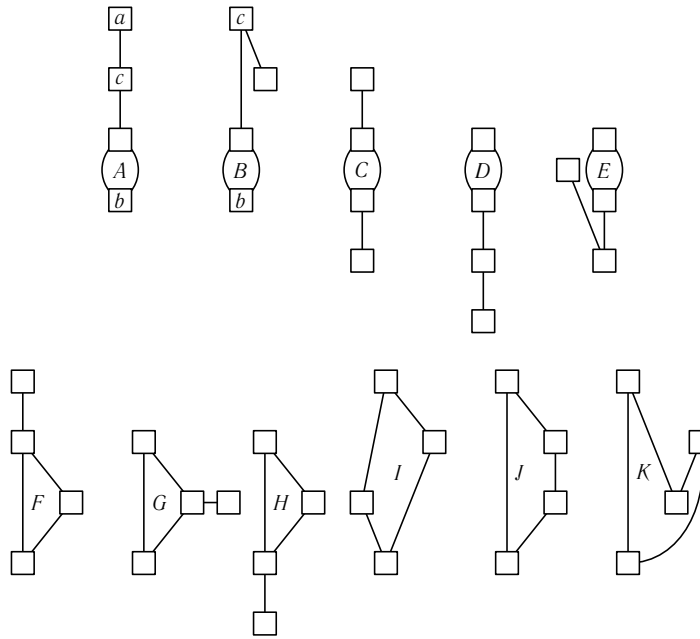


Рис. 19. Укороченные графы с одним циклом и 4 тройными вершинами

Здесь высоты изображены значениями ординат вершин, исключая только случай G, где ординаты двух вершин выбраны одинаковыми.

Мы должны теперь сосчитать числа правильно упорядоченных графов каждого из этих 11 типов. Из симметрии (относительно горизонтальной оси) ясно, что эти числа удовлетворяют соотношениям

$$|A| = |D|, \quad |B| = |E|, \quad |F| = |H|,$$

так что существенно различны только 8 случаев, A, B, C, F, G, I, J, K, которые мы теперь и разберем.

Случай *A*. Из верхней точки ветвления, уровня a, обязано выходить (максимальное) ребро вверх, (aa'). Из нижней точки, уровня b, обязано выходить (минимальное) ребро вниз, (bb').

Остаются еще два ребра $(a\alpha)$ и $(c\beta)$, выходящие из точки ветвления уровня a и из точки ветвления уровня c (соответственно).

Граф определяется высотами вершин α и β. Четыре значения в точках ветвления и значение b' делят полуось значений, меньших значения a', на 6 частей, так что для выбора значения α есть 6 (топологически разных) возможностей.

После того, как значение α выбрано, для выбора значения β ось значений разделена четырьмя значениями в точках ветвления и выбранными уже значениями (a', b', α) на 8 частей.

Итого, получаем 48 (топологически разных) случаев, так что *число правильных неэквивалентных упорядоченных графов типа A равно* 48.

Замечание. Для дальнейшего анализа тригонометрических многочленов полезно выделять те случаи, когда число точек графа на каждой высоте не превосходит 2.

Это — только три случая α_1, α_2 и α_3 из шести на рис. 20.

При выборе высоты α_1 для выбора β имеется 2 возможности (выше или ниже высоты c). При выборе α_2 возможностей тоже две (обе ниже α_2). При выборе α_3 возможностей выбора высоты β нет. Итого, условию (не более двух точек графа на каждой высоте) удовлетворяют всего 4 (указанных выше) случая из 48 правильных упорядочений графа типа A.

Случай *B*. Значение высоты α в четвертой точке ветвления может находиться в четырех интервалах, обозначенных на рисунке 21 знаками α_1, α_2, α_3, α_4.

Если значение α выбрано, то для определения упорядоченного графа остается назначить значения в обоих соседних вершинах точки ветвления высоты α.

Обозначим меньшее из этих двух значений через β (оно меньше выбранного для α значения α_k). Для выбора второго значения, γ, остается столько интервалов, насколько делят интервал $\gamma > \beta$ уже выбранные значения. В случае выбора значения α типа α_1 мы получим числа интервалов N

выбор k в $\beta_{1,k}$	1	2	3	4
N	4	5	6	7

Совершеннотаким же образом выбор значения α типа α_2 приводит к интер-

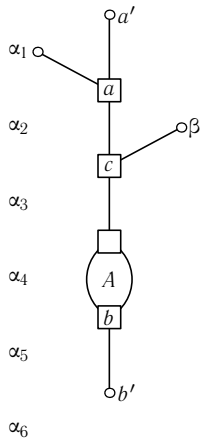


Рис. 20. Разные оснащения укороченного графа типа A

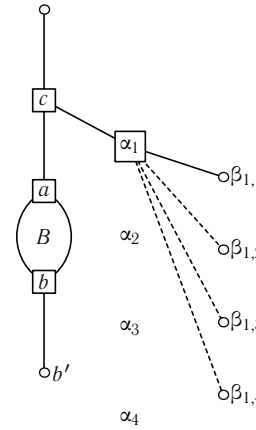


Рис. 21. Разные оснащения укороченного графа типа B

валам $\beta_{z,k}$ ($k = 1, 2, 3$) с числами ($N = 5, 6, 7$) интервалов для помещения значения γ.

В случае значения α типа α_3 получается два интервала $\beta_{3,k}$ ($k = 1, 2$) с числами ($N = 6, 7$) интервалов для помещения значения γ.

Наконец, для выбора значения α типа α_4 получаем единственный интервал $\beta_{4,1}$ (расположенный ниже уровня α_4), который поделен на $N = 7$ частей.

Суммируя все эти четыре подслучая, *мы находим* $(4 + 5 + 6 + 7) + (5 + 6 + 7) + (6 + 7) + 7 = 22 + 18 + 13 + 7 = 60$ *правильно упорядоченных графов типа* B.

Замечание. Ни в одном из них не отсутствуют горизонтали с более, чем двумя точками графа. Дело в том, что из вершины со значением α_k выходят либо вверх, либо вниз, два ребра, пересекающие одну горизонталь, а часть (bc) графа доставляет третью точку на той же горизонтали.

Случай *C*. Обозначим через a и c значения в верхней и в нижней точках ветвления. Выше уровня a заведомо есть концевые точки ребер, выходящих из a. Обозначим через a' максимальное из значений в их концевых вершинах. Аналогично, обозначим через c' минимальное из значений в концах ребер, выходящих из вершины уровня c.

Упорядочение определяется расположением значений α (во втором соседе вершины уровня a) и γ (во втором соседе вершины уровня c) — см. рис. 22.

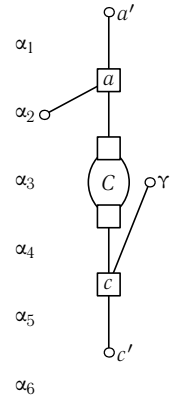


Рис. 22. Разные оснащения укороченного графа типа C.

При выбранном значении высоты α ось (меньших a') значений подразделена четырьмя значениями в точках ветвления и значением c' на 6 частей (α_k). При выборе значения γ ось значений (бóльших c') подразделена на 7 интервалов (четырьмя значениями в точках ветвления и выбранными значениями a' и α_k, если $k \leqslant 5$, в случае же $k = 6$ и интервалов только 6, так как $\alpha_6 < c'$).

Итак, общее число правильно упорядоченных графов типа C составляет $5 \cdot 7 + 6 = 41$.

Замечание. Отсутствие горизонтальных слоев с более, чем двумя точками ветвления, встречается только при $k = 1$ и 2 (в предыдущих обозначениях), иначе ребро ($\alpha_k a$) и цикл пересекли бы одну горизонталь трижды.

При этом для выбора значения γ тоже имеется только по 2 варианта (чтобы ребро (γc) было ниже цикла). Итак, *из* 41 *правильного графа типа C условию отсутствия трех точек графа на одной горизонтали удовлетворяют только* 4 *правильных графа*, описанные выше.

Случай *F*. Я оставляю в качестве задачи подсчет числа правильно упорядоченных графов типа F (их $6 \cdot 8 = 48$). Ни один из этих графов не удовлетворяет условию отсутствия трех точек графа на одной горизонтали.

Случай *G*. Таких правильно упорядоченных графов $2((7) + (7 + 6) + (7 + 6 + 5)) = 76$. Ни один из них не удовлетворяет условию отсутствия трех точек графа на одной высоте. Доказательства этих фактов оставляются слушателям в качестве задачи.

Случаи *I*, *J*, *K*. Числа правильно упорядоченных графов этих типов составляют, соответственно, $7 \cdot 8 = 56$, 56, $3 \cdot 3 = 9$. Ни один из этих упорядоченных графов не удовлетворяет условию отсутствия трех точек на одной горизонтали (задача).

Окончание доказательства теоремы 1. Суммируя числа правильно упорядоченных графов разных типов, указанные выше:

$$|A|=48,\quad |B|=60,\quad |C|=41,\quad |D|=48,\quad |E|=60,$$
$$|F|=48,\quad |G|=76,\quad |H|=48,\quad |I|=56,\quad |J|=56,\quad |K|=9,$$

мы получаем общую сумму, с учетом симметричных случаев,

$$2(48+60)+41+2(48+56)+76+9=216+41+208+85=257+293=550.$$

Замечание 1. Я перепроверял эти скучные вычисления до тех пор, пока совершенно разные способы перечисления правильно упорядоченных графов не стали давать одинаковые ответы, и после ряда ошибочных перечислений достиг уверенности в правильности окончательного ответа и всех промежуточных.

Но проводить аналогичные подсчеты, скажем, для $T=5$ или $T=6$, я не стал, хотя это и было бы очень полезно для открытия соответствующих общих гипотез (например, о росте числа правильно упорядоченных графов) с данным числом циклов (у нас 1) с ростом числа T тройных вершин.

Замечание 2. Число правильно упорядоченных графов всех типов, удовлетворяющих условию отсутствия трех точек графа на одной горизонтали, составляет

$$4(A)+0(B)+4(C)+4(D)+(0(E)+\ldots+0(K))=12.$$

Я предполагаю, что все эти 12 графов реализуются тригонометрическими многочленами

$$(x,\,y)=A\sin x+B\sin y+C\sin(x+y)+D\cos(x+y),$$

но не проверил этого, хотя можно надеяться реализовать все 12 случаев даже в окрестности точки

$$A=1,\quad B=1,\quad C=-1,\quad D=0$$

указанного четырехмерного пространства тригонометрических многочленов[1].

Бифуркационная диаграмма, образованная значениями $(A,\,B,\,C,\,D)\in\in\mathbb{R}^4$, соответствующими функциям с вырожденной (не морсовской) критической точкой или с кратными (принимаемыми в нескольких точках) критическими значениями, заслуживает явного топологического и алгебраического изучения (хотя бы с помощью компьютера, хотя можно сделать это и вручную). Эта трехмернаягиперповерхность в $\mathbb{R}^4$ является конусом

[1] В действительности для таких тригонометрических многочленов число седловых точек $T\leqslant 3$, так что упомянутые 12 графов не различаются, а различаются лишь нарисованные ниже 2 графа с $T=3$ (см. с. 47, где ответы написаны подробнее)

над своим сечением трехмерной плоскостью (например, гиперплоскостью $C = 1$), так что речь в этой задаче идет о рисовании двумерной (алгебраической) поверхности в обычном трехмерном пространстве.

При попытках компьютерного исследования таких задач я обнаружил, что делаю вручную примерно втрое меньше ошибок, чем компьютер (например, даже при простом перемножении сороказначных чисел). Притом, в то время, как непредсказуемые компьютерные ошибки происходят от каких-то космических частиц, мои ошибки оказались всегда одинаковыми и легко контролируемыми.

А именно, всякий раз, когда вычисление не умещается на одной странице и его приходится переносить и на следующую, некоторые (многозначные) числа приходится копировать. Именно это *переписывание* и вносит ошибки: я пишу мелким почерком похожие цифры 2 и 9, а также 3 и 5 — они-то и оказываются переписанными неверно, за ними-то и надо следить (возможно, не мне одному).

§4. Алгебраическая геометрия тригонометрических многочленов

Для сравнения графов специальных периодических функций

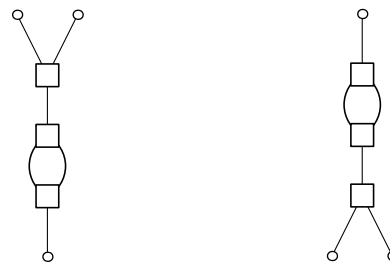
$$f_{A,B,C,D}(x,\, y) = A\sin x + B\sin y + C\sin(x+y) + D\cos(x+y) \qquad (1)$$

с графами общих функций Морса (с таким же числом критических точек) на двумерном торе я установил следующий удивительный для меня факт.

Теорема 1. *Функция Морса* (1) *имеет на торе не более* 8 *критических точек. Кривая* (*некритического*) *уровня этой функции — эллиптическая кривая* (*рода* $g = 1$), *ее вещественные точки образуют на торе не больше двух компонент связности.*

На самом деле критических точек у тригонометрического многочлена (1) не больше 6, но мы здесь не будем этого доказывать.

Упорядоченный граф тригонометрического многочлена (1) *с* 6 *критическими точками есть один из следующих двух графов* (*упорядочение вершин — по высоте*):



Остальные 14 упорядоченных графов гладких функций Морса на двумерном торе, имеющих 6 критических точек не встречаются ни у одного тригонометрического многочлена (1).

Все функции Морса на торе с одинаковыми упорядоченными 6*-вершинными графами, имеющие критические значения* $\{1, 2, \ldots, 6\}$, *переводятся друг в друга диффеоморфизмами тора.*

Существует бесконечное множество гладких функций Морса на двумерном торе с 6 *критическими точками и критическими значениями* $\{1, 2, \ldots, 6\}$, *упорядоченные графы которых все одинаковы, но ни одна из них не переводится в другую гомотопным тождеству диффеоморфизмом тора.*

Из этого бесконечного множества попарно несводимых функций Морса с упорядоченным графом, указанным выше, только три функции сводятся к тригонометрическим многочленам (1) *посредством диффеоморфизмов тора, принадлежащих связной компоненте единицы в группе диффеоморфизмов двумерного тора.*

Точка (общего положения) на цикле графа функции на двумерном торе изображает нестягиваемую на торе простую замкнутую кривую. Эта кривая, для подходящей гладкой функции Морса с 6 критическими точками, может быть любой замкнутой кривой на торе (отчего и получается бесконечное количество не сводимых друг к другу функций Морса).

Для тригонометрического многочлена (1) *простая замкнутая кривая на торе описанная выше, принадлежит к одному из трех типов*: *это либо параллель, либо меридиан, либо диагональ* ($x = \mathrm{const}$, $y = \mathrm{const}$ *или* $x + y = \mathrm{const}$), *отчего и получается три класса функций, упомянутые выше.*

Доказательство. Воспользуемся рациональностью окружности, т. е. используем обычные координаты $t \in \mathbb{R}P^1$ и $\tau \in \mathbb{R}P^1$ на окружностях $\{x \pmod{2\pi}\}$, $\{y \pmod{2\pi}\}$:

$$\cos x = \frac{1-t^2}{1+t^2}, \quad \sin x = \frac{2t}{1+t^2},$$
$$\cos y = \frac{1-\tau^2}{1+\tau^2}, \quad \sin y = \frac{2\tau}{1+\tau^2}.$$

Критические точки функции (1) определяются системой уравнений

$$\begin{cases} A\cos x + C\cos(x+y) - D\sin(x+y) = 0, \\ B\cos y + C\cos(x+y) - D\sin(x+y) = 0. \end{cases}$$

Из этих двух уравнений мы находим

$$A\frac{1-t^2}{1+t^2} = B\frac{1-\tau^2}{1+\tau^2}, \tag{2}$$

откуда следует, что

$$\tau^2 = \frac{A(t^2-1)+B(t^2+1)}{-A(t^2-1)+B(t^2+1)} = \frac{P_2(t)}{Q_2(t)},$$

где P_2 и Q_2 — многочлены второй степени,

$$P_2 = (B-A)+t^2(A+B), \quad Q_2(t) = (A+B)+t^2(B-A).$$

Нам потребуются и вытекающие из этого формулы

$$1+\tau^2 = 2B(1+t^2)/Q_2, \quad 1-\tau^2 = 2A(1-t^2)/Q_2.$$

В частности, имеет место удивительный факт:

$$\text{если } t^2=-1, \quad \text{то } \tau^2=-1 \quad (\text{так что } 1-\tau^2=2).$$

Мы использовали одно из двух уравнений критических точек. Второе уравнение

$$A\cos x + C\cos(x+y) - D\sin(x+y) = 0$$

переписывается в обозначениях t и τ в виде

$$A(1-t^2)(1+\tau^2) + C[(1-\tau^2)(1-t^2) - 4t\tau] + 2D[t(1-\tau^2)+\tau(1-t^2)] = 0.$$

Иными словами, выполняется квадратное уравнение

$$\tau^2 U + \tau V + W = 0 \tag{3}$$

с коэффициентами

$$U = (A-C)(1-t^2)+2Dt, \quad V = -4tC + 2D(t^2-1),$$
$$W = (A+C)(1-t^2) - 2Dt.$$

Подставляя вместо τ^2 указанную выше дробь P_2/Q_2, мы получаем из уравнения (3) решение

$$\tau = \frac{p_4(t)}{q_4(t)}, \quad p_4 = UP_2 + WQ_2, \quad q_4 = -VQ_2. \tag{4}$$

Теперь уравнение $\tau^2 = P_2/Q_2$ принимает вид

$$p_4^2 Q_2 = q_4^2 P_2,$$

то есть вид

$$Q_2(p_4^2 - V^2 P_2 Q_2) = 0. \tag{5}$$

Это уравнение относительно переменной t имеет степень 10, так что мы получаем из него 10 комплексных критических точек (t, τ).

Однако, две из них заведомо не вещественны: это $(i, -i)$ и $(-i, i)$, где

$$t^2 = -1, \quad \tau^2 = -1, \quad t\tau = 1, \quad t + \tau = 0. \tag{6}$$

Действительно, при $t^2 + 1 = 0$ мы получаем

$$P_2 = -2A, \quad Q_2 = 2A, \quad V = -4(tC + D),$$
$$U = 2(A - C + tD), \quad W = 2(A + C - tD),$$

так что левая часть соотношения (5) приобретает множитель

$$[4A(A - C + tD) + 4A(A + C - tD)]^2 + 16(tC + D)^2 4A^2 =$$
$$= 16A^2(2C - 2tD)^2 + 64(tC + D)^2A^2 =$$
$$= 64A^2(C^2 + t^2D^2 - 2tCD + t^2C^2 + D^2 + 2tCD) = 64A^2(C^2 + D^2)(1 + t^2),$$

обращающийся в нуль при $1 + t^2 = 0$.

Итак, *уравнение* (5) *имеет не более восьми вещественных корней* t, *доставляющих, в силу соотношения* (4), *не более* 8 *вещественных критических точек функции* (1) *на торе*. □

Для исследования комплексного множества уровня $\{(x, y\colon f(x, y) = c\}$, используем (аффинные) координаты t и τ на комплексных проективных прямых — сомножителях произведения $\mathbb{C}P^1 \times \mathbb{C}P^1$ (комплексифицирующего исходный тор).

Уравнение множества уровня имеет вид

$$A\frac{2t}{1+t^2} + B\frac{2\tau}{1+\tau^2} + C\Big(\frac{2t}{1+t^2}\frac{1-\tau^2}{1+\tau^2} + \frac{2\tau}{1+\tau^2}\frac{1-t^2}{1+t^2}\Big) +$$
$$+ D\Big(\frac{1-t^2}{1+t^2}\frac{1-\tau^2}{1+\tau^2} - \frac{4t\tau}{(1+t^2)(1+\tau^2)}\Big) = c.$$

Иными словами, уравнение линии уровня имеет вид

$$A2t(1+\tau^2) + B2\tau(1+t^2) + C[2t(1-\tau^2) + 2\tau(1-t^2)] +$$
$$+ D[(1-t^2)(1-\tau^2) - 4t\tau] = c(1+t^2)(1+\tau^2). \tag{7}$$

При фиксированном значении t это — квадратное уравнение относительно τ. Поэтому комплексная линия уровня $\{f(x, y) = c\}$ при проектировании

$$\pi\colon \mathbb{C}P^1 \times \mathbb{C}P^1 \to \mathbb{C}P^1,$$

заданном формулой $\pi(t, \tau) = t$, двулистно (разветвленно) накрывает сферу Римана $\mathbb{C}P^1$ с аффинной координатой t.

Точки ветвления t определяются условием $\Delta(t) = 0$, где Δ — дискриминант квадратного уравнения (7) относительно неизвестной τ.

Из формулы (7) видно, что этот дискриминант — многочлен степени 4 относительно t. Для типичного множества уровня типичного тригонометрического многочлена (1) дискриминант имеет 4 различных корня (а больше он не имеет никогда).

Двулистное накрытие сферы, разветвленное в 4 точках, накрывает сферу поверхностью тора. В этом можно убедиться, например, сосчитав эйлерову характеристику, или же «итальянским» методом, задав накрытие формулой

$$w^2 = (z^2 - 1)^2 - \varepsilon \quad (\text{где } \pi(z, w) = z).$$

При $\varepsilon = 0$ проектируемая комплексная кривая (рис. 23) представляет собой две сферы Римана ($w = z^2 - 1$ и $w = 1 - z^2$), пересекающиеся (трансверсально) в двух точках ($z = \pm 1$, $w = 0$).

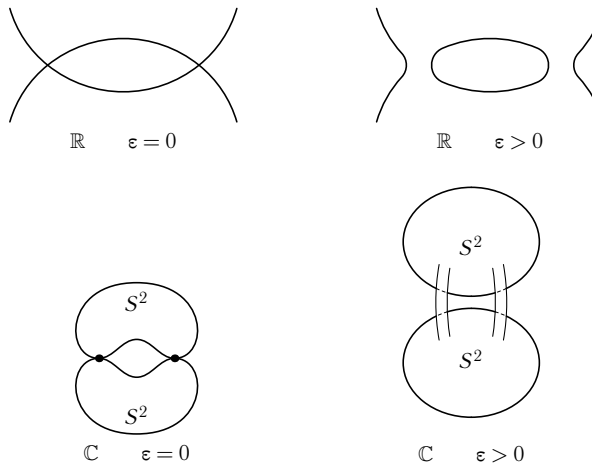


Рис. 23. Итальянский способ исследования двулистного накрытия сферы с четырьмя точкам ветвления

Переход к $\varepsilon \neq 0$ заменяет пару точек пересечения парой соединяющих две сферы трубочек и пара сфер превращается в тор (поверхность рода $g = 1$).

Достаточно также рассмотреть эллиптическую кривую, заданную уравнением рис. 24,

$$w^2 = z(z - 1)(z - 2)(z - 3),$$

и ее проекцию на ось z параллельно оси w (являющуюся двулистным накрытием, разветвленным над точками $z = 0$, 1, 2 и 3). Разрезы [0, 1] и

[2, 3] плоскости переменной z доставляют два листа w_1 и w_2 накрытия, склеенные вдоль разрезов, как указано, склейки и доставляют тор.

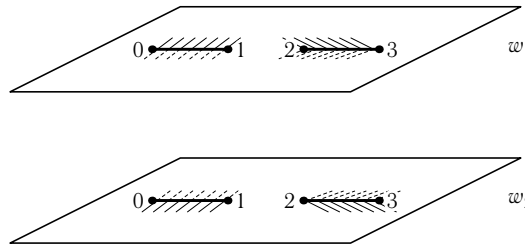


Рис. 24. Склеивание римановой поверхности эллиптической кривой из двух сфер Римана с парой разрезов на каждой из них

Из всего этого следует, что (неособая) линия уровня функции (1) — эллиптическая кривая (с торической римановой поверхностью рода $g=1$).

Но вещественная эллиптическая кривая не может иметь больше $g+1=$ $=2$ компонент связности (по теореме Харнака, обсуждавшейся выше, в § 2).

Поэтому вещественная линия уровня тригонометрического многочлена (1) не может иметь на торе

$$\{x \pmod{2\pi},\ y \pmod{2\pi}\}$$

больше двух компонент связности.

Из этого следует, что в соответствующем графе на одной горизонтали не может быть больше двух точек.

По доказанному в § 3, из 550 правильно упорядоченных графов функций Морса на торе с $T=4$ седловыми точками имеют не больше двух точек на всех горизонталях всего 12 графов.

Значит только такими могут быть графы тригонометрических многочленов (1): ими реализуются не более 12 из всех 550 топологически возможных графов с $T=4$. Я предполагаю, что все эти 12 графов реализуются тригонометрическими многочленами (1), и то, что сам я не сумел их реализовать, послужило причиной рассказать и о тригонометрических многочленах в этой лекции о 16-й проблеме Гильберта, посвященной вещественной алгебраической геометрии обычных многочленов.

Ни вычислители с их компьютерами, ни алгебраические геометры со своими аксиоматиками не внесли почти никакого вклада в решение этих настоящих (real, $\mathbb{R}$) задач. Наибольшие достижения в этой области, начатой Декартом и Ньютоном, Гурвицем и Клейном, Харнаком и Гильбертом, принадлежат российской математической школе: И. Г. Петровско-

му и В. А. Рохлину, О. Я. Виро и В. М. Харламову, Г. М. Полотовскому и Е. И. Шустину.

Но, к сожалению, самые естественные вопросы о топологической структуре обычных и тригонометрических многочленов остались, кажется, открытыми во всех этих исследованиях.

Рассмотрим, например, топологическую классификацию вещественных многочленов от одной переменной степени $n+1$ с n вещественными критическими точками с разными критическими значениями, со старшим членом x^{n+1}.

Числа N топологических типов при небольших n нетрудно сосчитать:

n	0	1	2	3	4	5	6	7	8
N	1	1	1	2	5	16	61	272	1385

(мы рассматриваем здесь «топологические типы» с точностью до топологических преобразований, сохраняющих ориентации осей координат). Эта замечательная последовательность чисел $N(n)$, легко узнаваемая по числу Эйлера 61, доставляет разложение в ряд Тейлора для тангенса:

$$\sum_{n=0}^{\infty} \frac{N(n)t^n}{n!} = \sec t + \operatorname{tg} t, \quad \operatorname{tg} t = \frac{1}{1!}t + \frac{2}{3!}t^3 + \frac{16}{5!}t^5 + \dots$$

Теоремы настоящей лекции получены при попытке перенести эти результаты на функции нескольких переменных.

Вопрос о топологической классификации вещественных многочленов не был включен Гильбертом в его классическую формулировку проблемы, а эти догматические формулировки гипнотизируют исследователей, стремящихся скорее преуспеть в решении классической задачи, чем разобраться в сути дела.

На международном математическом конгрессе в Стокгольме в 1962 году я рассказывал о своем решении проблемы устойчивости Биркгофа для нерезонансных положений равновесия систем Гамильтона. Решив эту классическую проблему, я не заметил, что доказал большее, а именно устойчивость в большинстве резонансных случаев (для резонансов порядка 5 и выше), потому что в классической проблеме исключались все резонансы. Американский математик Ю. Мозер тут же отметил это доказанное мной обстоятельство (устойчивость для «слабых резонансов»), но классическая формулировка классической проблемы помешала мне самому включить это свой результат в доклад на Конгрессе.

К счастью, ни доклады на конгрессах, ни выборы в академии наук, ни включение проблем в список Гильберта, ни награды вроде филдсовской и нобелевской не оказывали большого влияния на развитие математики, да и

других наук, так что я надеюсь, что влияние школы в Дубне будет бóльшим (и приведет к решения слушателями хотя бы части задач, обсуждавшихся в настоящей лекции).

Например, Л. Николаеску, продолжая исследования дубнинской лекции 2005 года, доказал в 2006 году содержащуюся в ней гипотезу о скорости роста порядка T^{2T} числа типов функций Морса с T седлами на S^2 (см. его статью «Morse function statistics» в журнале «Функциональный анализ и другая математика». 2006. V. 1, № 1. С. 97—103). Как растет число типов многочленов степени n, неясно (возможно, как степень n, а не как экспонента числа n).

Работа Николаеску устанавливает удивительную связь между задачей классификации топологических типов функций Морса на S^2 и теорией зеркальной симметрии квантовой теории поля (построенной А. Б. Гивенталем), хотя в статье Николаеску об этой связи явно не сказано.

Дело в том, что он нашел рекуррентное соотношение между числами $\varphi(T)$ типов функций Морса на сфере с T седлами, доказывающее, что эти числа топологический типов (17746 и т. д.) появляются при исследовании разложений некоторых эллиптических интегралов в ряды по степеням параметра, от которого (рационально) зависят подынтегральные функции $(x^3 + ax + b)^{-1/2}$.

Доказательства теорем Гивенталя о зеркальной симметрии также основаны на выражении целочисленных характеристик чисел Ходжа (комплексных алгебраических многообразий комплексной размерности 3) через коэффициенты разложения в ряд некоторых специальных абелевых интегралов (вдоль циклов «зеркального образа» многообразия).

Причем такая же топологическая характеристика многообразия-образа выражается таким же образом через аналогичные интегралы вдоль циклов исходного трехмерного многообразия. Интересно, что двумерным аналогом зеркальной симметрии физиков оказалась ранее открытая «странная двойственность» треугольников на плоскости Лобачевского (Арнольд, 1974).

Что окажется зеркальным образом задачи о топологической классификации гладких функций Морса по сфере S^2 остается, к сожалению, неясным.

Топологическая классификация тригонометрических многочленов и функций на торе, начатая выше в § 3, привела меня к неожиданным результатам, которые будут описаны на лекции в Дубне в 2006 году: классов в некоторых классификационных задачах конечное число, а в некоторых — бесконечное.

Ответы в задаче о классификации тригонометрических многочленов (1) и гладких функций с 6 фиксированными критическими значениями на торе получились такие:

Анализ C^∞: гладкие функции Морса с 6 критическими точками на торе T^2	16 классов	∞ классов
Алгебра: тригонометрические многочлены (1)	2 класса многочленов	6 классов многочленов
	$\mathrm{Diff}(T^2)$ классификация функций с точностью до диффеоморфизмов тора	$\mathrm{Diff}_o(T^2)$ классификация функций (с точностью до гомотопных тождественному диффеоморфизмов тора)

См. препринты

— Arnold V. Topological Classification of Trigonometric Polynomials, Related to Affine Coxeter Group $\tilde{A}_2$ // Abdus Salam International Centre for Theoretical Physics, ICTP. 2006. 15pp. IC/2006/039. 15pp.

`http://www.ictp.it/~pub_off`

— Arnold V. Smooth functions statistics // Abdus Salam International Centre for Theoretical Physics, ICTP. 2006. IC/2006/012. 9pp.

`http://www.ictp.it/~pub_off`

— Арнольд. В. Статистика и классификация топологий периодических функций и тригонометрических многочленов // Труды Института Математики и Механики УрО РАН. 2006. том 12, № 1, 2006, 10 стр.

ЛЕКЦИЯ 2

КОМБИНАТОРНАЯ СЛОЖНОСТЬ И СЛУЧАЙНОСТЬ

Человеку свойственно ошибаться, но
по-настоящему запутывает все
только компьютер

из «законов Мэрфи»

Вера и знание — две чаши весов:
чем выше одна, тем ниже другая

А. Шопенгауэр

Сегодняшняя лекция плохо укладывается в традиционное деление математики на части (вроде «теории чисел» и «теоретико-множественной топологии»).

Вопрос о том, *насколько случайна данная функция или последовательность*, не поставлен математически, хотя каждый из нас понимает, что, например, последовательность

$$001\ 001\ 001\ 001$$

менее случайна, чем последовательность

$$010\ 010\ 111\ 001.$$

Я попытаюсь теперь придать этим словам некоторый математический смысл. А именно, мы свяжем с такими последовательностями некоторые геометрические объекты, и их естественные геометрические характеристики будут служить мерой сложности исходной последовательности.

Этот анализ можно отнести и к геометрии конечных функциональных пространств, и к комбинаторике, и к арифметике конечных полей, и к статистике случайных процессов, и к линейной алгебре матриц, и к теории графов, и к арифметике жордановых нормальных форм операторов, и к теории конечных разностей и численного интегрирования, но я предпочитаю не включать его ни в один из этих узких разделов математики.

§1. Геометрия бинарных последовательностей

Пусть x — последовательность из n нулей и единиц,

$$x = (x_1, x_2, \ldots, x_n), \quad x_j \in \mathbb{Z}_2 (= \mathbb{Z}/2\mathbb{Z}).$$

Множество M всех таких последовательностей конечно, оно состоит из 2^n элементов. Эти элементы можно считать вершинами n-мерного куба

(рис. 1), а можно считать множество $\mathbb{Z}_2^n$ векторным пространством размерности n (над полем $\mathbb{Z}_2$ из двух элементов 0 и 1 с операциями Митрофанушки,

$$1+1=0+0=0, \quad 0+1=1+0=1,$$
$$0\cdot 0=0\cdot 1=1\cdot 0=0, \quad 1\cdot 1=1).$$

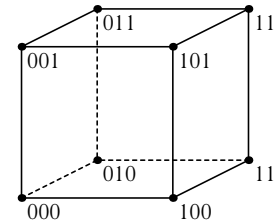


Рис. 1. Вершины n-мерного куба при $n=3$

Для исследования закономерности последовательности x мы будем рассматривать ее как *функцию со значениями* 0 *и* 1, *определенную на множестве из* n *элементов* j:

$$x\colon \{1, 2, \ldots, n\} \to \mathbb{Z}_2$$

(так что $x(j)=x_j$).

Чтобы исследовать такую функцию, мы последуем рецепту Ньютона и составим *последовательность ее разностей*,

$$y_j = x_{j+1} - x_j \in \mathbb{Z}_2.$$

Чтобы последовательность разностей состояла по-прежнему из n элементов, мы определим x_{n+1} как x_1, замкнув нашу последовательность длины n до цикла (или считая функцию x натурального элемента j периодической, с периодом n). В этом смысле мы можем считать областью определения функции x конечную окружность из n точек,

$$\mathbb{Z}_n = \mathbb{Z}/(n\mathbb{Z}).$$

Эта функция

$$x\colon \mathbb{Z}_n \to \mathbb{Z}_2$$

вовсе не предполагается здесь линейной.

Среди таких функций самые простые — постоянные, $x=0$ и $x=1$. Мы будем считать многочлены меньшей степени более простыми функциями, чем многочлены большей степени, и приведенные далее определения придадут этому замыслу точный смысл.

Согласно Ньютону, для многочленов степени меньше m (и только для них) операция взятия разностей приведет к нулю после m-кратного применения:

$$A^m x = 0.$$

Оператор взятия разностей $A\colon \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ линеен, так что A^m — это m-я степень линейного оператора A.

С другой стороны, операция $A\colon M \to M$ отображает конечное множество M последовательностей в себя. Поэтому мы можем связать с ней

следующий замечательный *граф операции* с $|M| = 2^n$ вершинами x: *из каждой вершины x выходит ровно одна стрелка, и она ведет из вершины x в вершину Ax.*

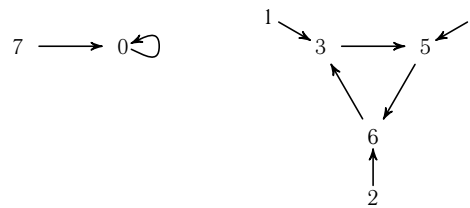
Сосчитаем, например, этот граф для случая $n = 3$, когда вершин 8. По определению, мы находим

$$A(0,\,0,\,0) = (0,\,0,\,0); \quad A(0,\,0,\,1) = (0,\,1,\,1), \quad A(0,\,1,\,0) = (1,\,1,\,0),$$
$$A(0,\,1,\,1) = (1,\,0,\,1), \quad A(1,\,0,\,0) = (1,\,0,\,1), \quad A(1,\,0,\,1) = (1,\,1,\,0),$$
$$A(1,\,1,\,0) = (0,\,1,\,1), \quad A(1,\,1,\,1) = (0,\,0,\,0).$$

Чтобы не писать таких длинных формул, мы используем в качестве обозначения для $x = (x_1, \ldots, x_n)$ целое число (или вычет по модулю 2^n) с такими бинарными цифрами (в двоичной системе счисления, как в компьютере):

$$X = x_1 2^{n-1} + x_2 2^{n-2} + \ldots + x_n 2^0.$$

В этих обозначениях вершина куба $x = (1,\,1,\,1)$ превращается в число $X = 7$. Вся предыдущая таблица действия операции взятия разностей A бинарных последовательностей периода $n = 3$ принимает вид ориентированного графа с восемью вершинами:



В графе любого отображения конечного множества в себя из каждой вершины выходит ровно одно ребро. Легко доказывается

Теорема 1. *Каждая компонента связности графа любого отображения конечного множества в себя содержит один и только один цикл.*

Вся компонента получается из этого притягивающего цикла-аттрактора добавлением к каждой вершине притягиваемого к ней дерева.

Например, для операции взятия разностей в $\mathbb{Z}_2^3$ мы нашли две компоненты, с циклами длин 1 и 3, оснащенными простейшими деревьями с двумя вершинами:

Доказательство теоремы 1 таково. Рассмотрим орбиту какой-либо точки x при применении отображения A (она состоит из точек x, Ax, A^2x, A^3x, ...).

Поскольку все отображаемое множество конечно, последовательные точки орбиты не могут все быть различными: $A^p x = A^q x$ для некоторых (неравных) p и q. Пусть, скажем, $p > q$. Тогда $A^r y = y$ для $r = p - q$, $y = A^q x$, т. е. мы нашли цикл периода r (в компоненте любой точки x, т. е. в любой компоненте графа).

Если бы в одной компоненте было два цикла, соединенных конечной цепочкой ребер графа, $(a, b, \ldots, z)$, то ребро a было бы направлено к первому циклу, а ребро z ко второму. Поэтому в цепочке нашлась бы вершина, из которой выходили бы ребра к обоим циклам, что невозможно, так как из каждой вершины выходит только одно ребро (а ребро, вышедшее из вершины цикла, все принадлежит этому циклу).

Итак, цикл в компоненте один, а все остальные ребра (его компоненты) ведут к нему, и теорема 1 доказана. □

Мы введем следующие обозначения: цикл из m ребер обозначается знаком O_m. Если T — некоторое корневое дерево, то $O_m * T$ будет обозначать граф, составленный из m копий корневого дерева T, где ребра каждой копии направлены к корню дерева, а эти корни всех m деревьев составляют цикл O_m в графе.

В этих обозначениях предыдущее описание операции взятия разностей $A\colon \mathbb{Z}_2^3 \to \mathbb{Z}_2^3$ принимает вид графа из двух компонент, $(O_1 * T_2)$ и $(O_3 * T_2)$.

Знаком T_{2^n} мы будем обозначать бинарное корневое дерево (с n этажами выше корня и 2^n вершинами), ребра которого направлены к корню (рис. 2).

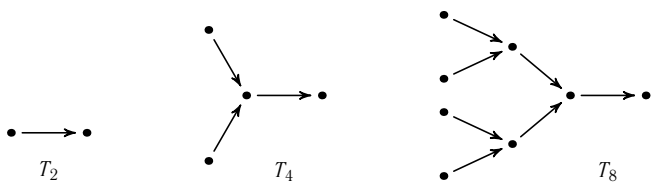


Рис. 2. Бинарные корневые деревья

Бинарность дерева означает, что в каждую вершину любого этажа, кроме первого и самого верхнего, приходит ровно 2 ребра (вышедшие из вершин следующего этажа, а для находящегося на первом этаже корня — из (единственной) вершины второго этажа).

Например, знак $(O_2 * T_4)$ означает направленный граф с восемью вершинами, изображенный на рис. 3.

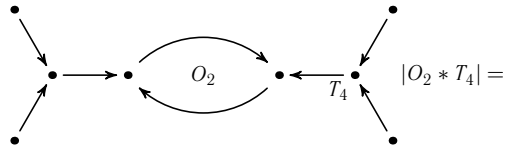


Рис. 3. Лес деревьев T_4 на цикле O_2

Наш план исследования сложности последовательности $x \in \mathbb{Z}_2^n$ состоит в том, чтобы рассматривать точку x как вершину графа операции Ньютона (взятие разностей) $A\colon \mathbb{Z}_2^n \to \mathbb{Z}_2^n$.

Последовательность x будет считаться тем более сложной, чем большей компоненте графа операции A принадлежит эта точка x (т. е. чем больше длина цикла этой компоненты и чем больше лес оснащающих этот цикл в компоненте деревьев), а также при фиксированной длине цикла для сложности важно, насколько далеко от цикла расположена точка x, на сколь высокой ветке дерева она находится.

Для этого мы, прежде всего, перечислим все компоненты графа, их циклы и леса. Это — уже не легкая задача и я полностью знаю ответы только при $n \leqslant 12$.

§2. Графы операций взятия разностей

Подобно тому, как это сделано выше для периодических последовательностей длины $n = 3$, я сосчитал графы операций взятия разностей периодических бинарных последовательностей периода n, $A\colon \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ для всех $n \leqslant 12$.

При бóльших n я этого не сделал потому, что число вершин графа, 2^n, равное при $n = 12$ всего 4096, еще позволяет в этом случае нарисовать весь этот граф на одной странице, тогда как при бóльших значениях n он занимает несколько страниц и я начинаю путаться в своих (бескомпьютерных) вычислениях.

Между тем, продолжить эти мои вычисления на бóльшие значения n было бы очень интересно, в особенности, потому, что уже анализ случаев $n \leqslant 12$ приводит к целому ряду увлекательных гипотез, которые хотелось бы проверить экспериментально, прежде чем пытаться их доказывать.

Именно эта возможность быстро (и ничего заранее не зная) продолжить мои исследования побудила меня включить эту тему в настоящий курс

лекций для школьников в Дубне: я надеюсь, что слушатели добьются новых успехов.

Теорема 1. *Графы операций A взятия разностей бинарных периодических последовательностей длины $n \leqslant 12$ имеют такие структуры, как указано в следующей таблице*:

n	b	компоненты графа операции A	соотношение
2	1	$(O_1 * T_4)$	$A^2 = 0$
3	2	$(O_3 * T_2) + (O_1 * T_2)$	$A^4 = A$
4	1	$(O_1 * T_{16})$	$A^4 = 0$
5	2	$(O_{15} * T_2) + (O_1 * T_2)$	$A^{16} = A$
6	4	$2(O_6 * T_4) + (O_3 * T_4) + (O_1 * T_4)$	$A^8 = A^2$
7	10	$9(O_7 * T_2) + (O_1 * T_2)$	$A^8 = A$
8	1	$(O_1 * T_{256})$	$A^8 = 0$
9	6	$4(O_{63} * T_2) + (O_3 * T_2) + (O_1 * T_2)$	$A^{64} = A$
10	10	$8(O_{30} * T_4) + (O_{15} * T_4) + (O_1 * T_4)$	$A^{32} = A^2$
11	4	$3(O_{341} * T_2) + (O_1 * T_2)$	$A^{342} = A$
12	24	$20(O_{12} * T_{16}) + 2(O_6 * T_{16}) + (O_3 * T_{16}) + (O_1 * T_{16})$	$A^{16} = A^4$

В столбце b («число Бетти») указано число компонент связности графа, перечисленных в следующем столбце. Обозначение $20(O_{12} * T_{16})$ в строке $n = 12$ означает, что имеется 20 (изоморфных друг другу) компонент с циклами периода 12, оснащенными в каждой точке цикла корневым (четырехэтажным) бинарным деревом с 16 вершинами.

Таким образом, при $n = 12$ все 24 компоненты графа доставляют следующее число вершин:

$$20 \cdot 12 \cdot 16 + 2 \cdot 6 \cdot 16 + 1 \cdot 3 \cdot 16 + 1 \cdot 1 \cdot 16 = (240 + 12 + 3 + 1)16 = 256 \cdot 16 = 2^{12},$$

(как и следовало для $|\mathbb{Z}_2^{12}| = 2^{12}$). Число всех вершин всех их циклов составляет при этом $20 \cdot 12 + 2 \cdot 6 + 1 \cdot 3 + 1 \cdot 1 = 256$.

В последнем столбце таблицы указано тождество, которому удовлетворяет линейный оператор A. Например, для компоненты $(O_3 * T_2)$ при $n = 3$ мы замечаем, что точка $y = Ax$ всегда принадлежит циклу, поэтому $A^3 y = y$, так что $A^4 = A$.

Это тождество, однако, можно было бы легко предвидеть заранее алгебраически, до вычисления графа (построить который тождество помогает).

Обозначим через $\delta\colon \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ линейный оператор циклического сдвига последовательности длины n (так что $(\delta x)_j = x_{j-1}$ для $j \in \mathbb{Z}_n$).

Очевидно, $A = 1 + \delta$ (так как для вычетов по модулю 2 разность совпадает с суммой), и $\delta^n = 1$ (так как δ — циклический сдвиг правильного n-угольника на угол $2\pi/n$).

Теперь, для $n = 3$, мы последовательно вычисляем степени оператора взятия разностей A так:

$$A = 1 + \delta, \quad A^2 = 1 + 2\delta + \delta^2 = 1 + \delta^2,$$
$$A^3 = (1 + \delta)(1 + \delta^2) = 1 + \delta + \delta^2 + \delta^3 = \delta + \delta^2$$

(поскольку $\delta^3 = 1$, $1 + 1 = 0$). Наконец,

$$A^4 = (1 + \delta)(\delta + \delta^2) = \delta + \delta^2 + \delta^2 + \delta^3 = \delta + 1 = A.$$

Такими же выкладками доказываются и остальные соотношения последнего столбца таблицы.

Замечание. Разглядывая таблицу, можно сделать ряд интересных наблюдений, некоторые из которых уже превращены сегодня в доказанные теоремы.

Например, если $n = 2^k$, то компонента в графе всего одна, и период равен 1, так что весь граф сводится к оснащающему корень $x = 0$ дереву T_{2^n} с 2^{2^k} вершинами.

В этом случае все функциональное пространство n-периодических функций со значениями в $\mathbb{Z}_2$ совпадает с кольцом «многочленов» (в общем случае подкольцо «многочленов» составляет выписанную последней компоненту связности $O_1 * T_r$, где $r = 2^{2^k}$ для $n = 2^k(2l + 1)$).

Под «многочленом» я понимаю здесь многочлен с рациональными коэффициентами и целыми в целых точках j значениями

$$x(j) = a_0 j^m + a_1 j^{m-1} + \ldots + a_m,$$

приведенный по модулю 2, чтобы значения попали в $\mathbb{Z}_2$.

Примерами таких «многочленов» являются удивительным образом целочисленные числа сочетаний

$$C_j^2 = \frac{j(j-1)}{2}, \quad C_j^3 = \frac{j(j-1)(j-2)}{6} \quad \text{и т. д.}$$

«Многочлены» периода n образуют подкольцо кольца n-периодических функций (со значениями в $\mathbb{Z}_2$).

Для дальнейшего интересен вопрос, какой период имеет «многочлен» $x(j) = C_j^k \pmod 2$ при фиксированном k, а также какова размерность векторного пространства «многочленов» периода n со значениями в $\mathbb{Z}_2$ (над полем $\mathbb{Z}_2$), то есть сколько из «многочленов» линейно независисмы.

Эти вопросы арифметики биномиальных коэффициентов по модулю 2 легко решает геометрия треугольника Паскаля, и я оставляю их слушателям в надежде, что они где-нибудь видели этот треугольник (хотя его и нет в современных учебниках теории вероятностей для школьников).

Оставляю пока слушателям и разгадку загадочного поведения при изменении n наибольших (и других) периодов T циклов компонент графа оператора взятия разностей в $\mathbb{Z}_2^n$:

n	2	3	4	5	6	7	8	9	10	11	12
T	1	3	1	15	6	7	1	63	30	341	12

Если период T наибольшего цикла отличен от 1, то он делится на n, и вдобавок частное почему-то имеет вид $2^s - 1$.

Например, $341 = 11 \cdot 31$.

Число 341 доставляет первый пример, опровергнувший продержавшееся несколько тысячелетий китайское обращение «малой теоремы Ферма». Древние китайцы думали, что если $2^a \equiv 2 \pmod a$, то число a простое. Но для $a = 341$ это сравнение выполняется, поскольку

$$2^{11} \equiv 2 \pmod{11}, \quad 2^{11} \equiv 2 \pmod{31},$$
$$2^{31} \equiv 2 \pmod{11}, \quad 2^{31} \equiv 2 \pmod{31}.$$

Удивительно в таблице и то, что каждая компонента связности графа имеет вид $O_m * T_{2^k}$, где бинарное дерево T_{2^k} — такое же, как для кольца «многочленов» соответствующего периода.

Этот факт объясняется элементарной геометрией линейного оператора

$$A\colon \mathbb{Z}_2^n \to \mathbb{Z}_2^n.$$

Эта геометрия утверждает, что множество $\mathrm{Ker}(A^s)$ решений x-линейного однородного уравнения

$$A^s x = 0,$$

является векторным подпространством векторного пространства $\mathbb{Z}_2^n$, а множество решений неоднородного уравнения (с правой частью y) — параллельное ему аффинное подпространство.

Подпространства решений однородных уравнений с разными s образуют растущую последовательность

$$\mathrm{Ker}(A) \subseteq \mathrm{Ker}(A^2) \subseteq \mathrm{Ker}(A^3) \subseteq \ldots \subseteq \mathrm{Ker}(A^\infty),$$

(где $\mathrm{Ker}(A^\infty)$ означает просто наибольшее из пространств последовательности). Знак «Ker» происходит от слова «Kernel» = «ядро», каким называют пространство решений однородного уравнения.

Это пространство $\mathrm{Ker}(A^\infty)$ и есть кольцо «многочленов», так как, по теореме Ньютона, «многочлены» образуют объединение всей последовательности (пространств многочленов разных степеней). В графе операции A это компонента, притягиваемая аттрактором 0 периода 1.

Другую последовательность подпространств образуют образы степеней оператора A:

$$\mathbb{Z}_2^n \supseteq A(\mathbb{Z}_2^n) \supseteq A^2(\mathbb{Z}_2^n) \supseteq A^3(\mathbb{Z}_2^n) \supseteq \dots.$$

Эта последовательность векторных подпространств убывает. Пересечение всех этих подпространств я обозначу через $A^\infty(\mathbb{Z}_2^n)$ (не заботясь, как и в случае пространства решений однородного уравнения, о смысле оператора A^∞).

Пространство $A^\infty(\mathbb{Z}_2^n)$ имеет в терминах графа операции A простое описание: ему принадлежат в точности все точки всех циклов всех компонент (и только они), так как любая другая вершина графа, леса которого имеют высоту h, не имеет вершин выше себя больше чем на h, а потому не принадлежит образу оператора A^{h+1}.

Например, из всего этого следует, что сумма $\sum$ длин всех циклов графа является степенью двойки (а именно, 2^q, если векторное пространство $A^\infty(\mathbb{Z}_2^n)$ имеет размерность q над $\mathbb{Z}_2$), что и наблюдается в нашей таблице:

n	2	3	4	5	6	7	8	9	10	11	12
$\sum$	1	4	1	16	16	64	1	128	256	1024	256

Все утверждения теоремы 2 проверяются непосредственными вычислениями, и я не привожу этих вычислений в надежде, что слушатели (вооруженные, возможно, компьютерами) проведут их сами не только при $n \leqslant 12$, но и для гораздо бóльших периодов n периодических бинарных последовательностей.

§3. Логарифмическая функция и ее сложность

Таблица показывает, что, если n — не степень двойки, то не всякая n-периодическая функция со значениями из $\mathbb{Z}_2$ является «многочленом». В этому случае мы будем считать эти неполиномиальные функции x «более сложными, чем все многочлены». При желании можно было бы продолжать иерархию функций нашего конечного функционального пространства (в котором всего 2^n элементов) — рассмотреть экспоненты, синусы, квазимногочлены и т. д., и сравнивать их «сложности» (в зависимости от того, какому простейшему дифференциальному уравнению они удовлетворяют, например).

Но я не буду сейчас этого делать, а рассмотрю одну специальную функцию — дискретный «теоретико-числовой логарифм» со значениями 0 и 1, который почему-то оказывается «очень сложной» в нашем смысле функцией.

Для определения этого арифметического логарифма я предположу, что $n + 1 = p$ — нечетное простое число.

Пусть a — первообразный остаток от деления на p, так что $a^{p-1} \equiv 1(p)$ и n остатков

$$\{a,\ a^2,\ \ldots,\ a^{p-1}\}$$

доставляют все ненулевые остатки k от деления на p.

В случае, когда

$$a^l \equiv k \pmod{p}$$

мы будем называть число l (или лучше остаток от деления l на $n = p - 1$) *арифметическим логарифмом* l числа (или остатка) k:

$$l = \log_a k.$$

Мы определили функцию l аргумента k. Нужная нам функция L со значениями в $\mathbb{Z}_2$ — это остаток от деления числа l на 2:

$$L(k) \equiv l(k) \pmod{2}.$$

Заметим, что, поскольку число $p - 1$ четно, то значение $L(k)$ однозначно определено (несмотря на то, что l определено лишь как остаток от деления на $p - 1$).

Мы будем рассматривать эти значения как бинарную последовательность длины $p - 1 = n$:

$$L(1),\ L(2),\ \ldots,\ L(n).$$

Значение в точке k равно нулю если и только если $k \neq 0$ является квадратным вычетом по модулю p, и равно 1, если $k \neq 0$ является квадратичным невычетом.

Из этого видно, что наш «арифметический бинарный логарифм» $L\colon \mathbb{Z}_n \to \mathbb{Z}_2$ не зависит от выбора того первообразного остатка a, при помощи которого мы его определили, а зависит только от числа $n = p - 1$.

Прямые вычисления расположения этой «логарифмической» функции в графах теоремы 2 показывают, что *эта специальная функция оказывается либо максимально сложной* (среди всех бинарных функций периода n), *либо почти максимально сложной* (в смысле сложности определенной нами геометрией графа операции взятия графа).

Эти вычисления, сами по себе не сложные, довольно длинны, и для описания их результатов мы будем задавать функцию $L \in \mathbb{Z}_2^n$ двоичным числом $X < n$ цифрами $L(k)$:

$$X = L(1)2^{n-1} + L(2)2^{n-2} + \ldots + L(n)2^0 \pmod{2^n}.$$

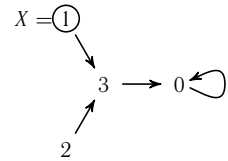
В простейшем случае $p = 3$, $n = 2$, логарифмическая функция L вычисляется при помощи геометрической прогрессии n остатков

$$\{2^l \pmod{3}\} = (2,\ 1).$$

Стало быть, $l(2) = 1$, $l(1) = 2$, так что

$$L(1) = 0, \quad L(2) = 1, \quad X = 1 \pmod 4.$$

Граф операции A имеет в случае $n = 2$ вид



Точка $X = 1$ — наиболее сложная бинарная функция периода 2, так как эта вершина дерева T_4 удалена от цикла (корня 0) наиболее далеко.

При следующих простых значениях ($p = 5, 6, 11, 13$) вычисления совершенно аналогичны, но более длинны, и я приведу только их результаты.

Теорема 1. *Арифметические логарифмы L доставляют в графах операции взятия разностей бинарных последовательностей периода $n = p - 1$ вершины X, приведенные для $p = 5, 7, 11$ и 13 в следующих таблицах.*

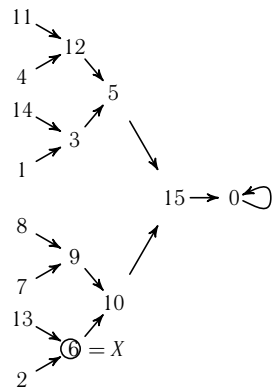
При $p = 7, 11$ это — самые сложные бинарные функции периода $n = p - 1$, а при $p = 5, 13$ — почти самые сложные (вершина X расположена в первом случае на наибольшем расстоянии от цикла, а во втором — на 1 ближе).

Случай $p = 5$, $n = 4$. Выбирая $a = 2$, находим

$$l(1) = 0, \ l(2) = 1, \ l(3) = 3, \ l(4) = 2,$$

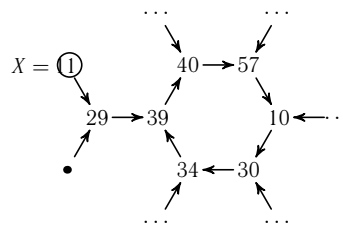
откуда получается для «арифметического бинарного логарифма» $X = 6$.

Нужная компонента графа имеет вид $O_1 * T_{16}$, а именно



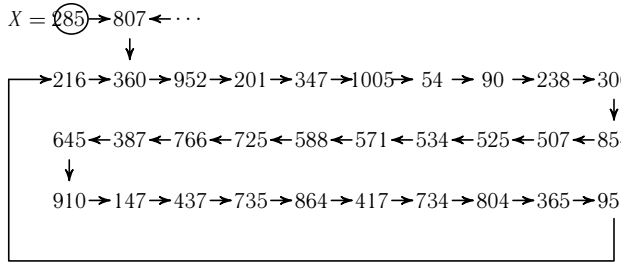
Случай $p = 7$, $n = 6$. Выбирая $a = 3$, приходим к «арифметическому бинарному логарифму» $X = 11$.

Нужная компонента графа имеет вид $O_6 * T_4$, и я выпишу только орбиту точки X в этом графе:



Случай $p = 11$, $n = 10$. Выбирая опять первообразный остаток $a = 2$, мы получаем геометрическую прогрессию (2, 4, 8, 5, 10, 9, 7, 3, 6, 1) (mod 11), откуда $L = (0, 1, 0, 0, 0, 1, 1, 1, 0, 1)$, $X = 285$.

Соответствующая компонента имеет вид $O_{30} * T_4$, и я выпишу из ее 120 точек только 32 вершины орбиты бинарного арифметического логарифма X:



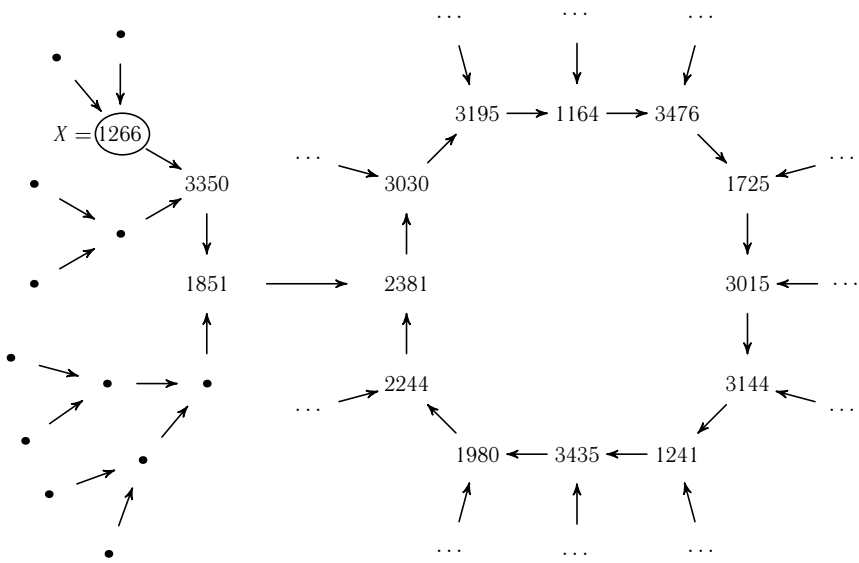
Случай $p = 13$, $n = 12$. Примитивный остаток $p = 2$ доставляет при $k = (1, 2, \ldots, 12)$ значения

$$\log_2 k = (12, 1, 4, 2, 5, 11, 3, 8, 10, 7, 6),$$

откуда для бинарного арифметического логарифма получается $X = 1266$.

Эта вершина принадлежит самой большой компоненте графа, $O_{12} * T_{16}$. Из 192 вершин этой компоненты в орбиту бинарного арифметического логарифма X попадает только 15 вершин, которые выписаны ниже:

Из этого вида компоненты следует, что бинарный арифметический логарифм периода 12 — почти самая сложная из бинарных функций периода 12. Эта вершина X принадлежит компоненте с наибольшим периодом цикла, и в ней удалена от цикла на почти наибольшее расстояние (равное 3, в



то время как самые верхушки оснащающего цикл леса удалены от цикла на расстояние 4).

Вся описанная выше теория остается пока эмпирической: ни ее перенесение на случай $n > 12$, ни доказательства высказанных многочисленных гипотез, (например, гипотезы о большой сложности бинарного арифметического логарифма при бóльших значениях n), еще не известны.

Кроме того, развитая здесь теория бинарных функций (со значениями в $\mathbb{Z}_2$), то есть последовательностей нулей и единиц, должна бы быть обобщена на случай функций с бóльшим числом значений (например, для функций со значениями в $\mathbb{Z}_p$ или в $\mathbb{Z}_n$).

Это относится, например, к вопросу о сложности определенной выше функции $l(k) = \log_a k$ аргумента $k \in \mathbb{Z}_p \setminus 0$ со значениями в $\mathbb{Z}_{p-1}$, возникающими в ситуации малой теоремы Ферма: насколько беспорядочно (случайно) распределены логарифмы последовательных чисел (или, обратным образом, элементы геометрических прогрессий вычетов)?

§4. Сложность и случайность таблиц полей Галуа

Аналогичные вопросы о случайности возникают и в теории полей Галуа: поле $\mathbb{Z}_p$ — один из примеров, общие поля Галуа имеют p^k элементов, например, поля из p^2 элементов уже доставляют много случайных на вид таблиц, но их сложность не измерена пока никакой точно определенной мерой, а остается лишь эмпирически наблюденным фактом.

Вот пример такой «случайной» таблицы из p^2 клеточек, заполненных числами от 1 до p^2.

Пусть $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ — матрица второго порядка, элементы которой — остатки от деления на простое число p, обладающая тем свойством, что все $p^2 - 1$ матрицы $A^k (1 \leqslant k \leqslant p^2 - 1)$ различны и $A^{p^2-1} = 1$.

Такое случается: для любого простого числа p такая матрица существует. Например, при $p = 5$ годится матрица

$$A = \begin{pmatrix} 0 & 2 \\ 1 & 2 \end{pmatrix}.$$

В этом случае $A^2 = \alpha A + \beta$ (в примере $\alpha = 2$, $\beta = 2$), что позволяет легко вычислить

$$A^3 = \alpha A^2 + \beta A = (\alpha^2 + \beta)A + \alpha\beta$$

и так далее, $A^k = u_k A + v_k$ (где $0 \leqslant u < p$, $o \leqslant v < p$).

Расположим число k в клеточке с координатами (v_k, u_k) таблицы размера $p \times p$. Полученные матрицы A^k с добавлением матрицы 0, для $k = \infty$, $u_\infty = v_\infty = 0$, образуют поле Галуа из p^2 элементов. Операции в этом поле такие:

$$A^k \cdot A^l = A^{k+l}, \quad A^k + A^l = A^{k*l},$$

где $*$ — операция сложения мест в таблице,

$$v_{k*l} = v_k + v_l, \quad u_{k*l} = u_k + u_l.$$

Символ A^∞ означает здесь нулевую матрицу. Таблица в случае $p = 5$, построенная по описанному правилу, имеет вид, изображенный на рис. 4.

13	15	5	16	20
7	10	9	14	23
19	11	2	21	22
1	8	4	17	3
∞	24	18	6	12

Рис. 4. Таблица поля Галуа из 25 элементов

Глядя на эти числа, можно убедиться, что они заполняют таблицу «случайным образом». При попытке проверить «частоты» различных событий,

которые предсказывает «случайное» заполнение, получаются тем более близкие к ожидаемым результаты, чем больше p.

Гипотетически это асимптотически так для попадания первых N чисел таблицы в любую (не слишком сложно определяемую) область: ожидаемое число этих попаданий равно ϑN, если отношение площади области к площади таблицы составляет ϑ.

Возьмем, например, первые $N=10$ из 25 значений ($k=1, \ldots, 10$). Первые два столбца таблицы из 5 столбцов составляют $\vartheta = 2/5$ ее площади. Случайное заполнение заставляет ожидать $(2/5)N$ попаданий. В таблице это значения $k=1, 8, 7, 10$ — их как раз 4. Столь точное совпадение числа попаданий с ϑN имеет место не всегда, но точность повышается с ростом таблицы, если область определена не слишком сложным алгоритмом.

Слушатели могут сами выбирать критерии «случайности» и экспериментально проверять их. Некоторое количество примеров имеется в моей статье «Геометрия и динамика полей Галуа» (Успехи матем. наук. 2004. Том 59, №6. С. 23—40) и в книжке «Динамика, статистика и проективная геометрия полей Галуа», МЦНМО, 2006.

Утверждение, что критерии стохастичности выполняются при $p \to \infty$ в пределе точно, хорошо подтверждается экспериментами, но доказано далеко не всегда, и я рассказываю здесь о нем именно в надежде на помощь читателей в (хотя бы экспериментальном) исследовании этого вопроса.

Следующий эксперимент относится к теории тасования карт. Мы сопоставим таблице поля Галуа из p^2 элементов следующую перестановку p^2 элементов.

Каждому номеру $k (1 \leqslant k < p^2)$ сопоставим адрес $K = pu_k + v_k$ той клетки таблицы, где этот номер расположен. Например, в таблице рис. 4 получаемая $K(1)=5$, $K(2)=12, \ldots, K(20)=24$.

Гипотеза состоит в том, что эта перестановка столь же «случайна», как хорошая перетасовка колоды из p^2-1 карты (качество тасовки улучшается при $p \to \infty$).

При желании можно добавить еще один элемент, переставляя все p^2 элементов поля Галуа. Для этого надо заменить символ A^∞ на 0 (полагая $k = p^2$ вместо ∞), добавив адрес $K=0$ места, где стоит нулевая матрица.

Насколько «случайна» получающаяся перестановка $p^2=25$ элементов поля Галуа, мы обсудим в следующей лекции.

Для полей Галуа из p^n элементов таблица поля заполняет n-мерный куб со стороной p. Она основана на выражении $n \times n$ матриц A^k, элементы которых — остатки от деления на p, в виде линейных комбинаций n матриц $1, A, A^2, \ldots, A^{n-1}$:

$$A^k = u_k A^{n-1} + v_k A^{n-2} + \ldots + w_k A^{n-n}.$$

«Случайность» таблиц таких полей растет при $p \to \infty$ подобно тому. как это выше объяснено для $n = 2$. Слушатели могут сами проверить это экспериментально, хотя доказано это далеко не всегда.

Скажу еще о соотношении «сложности» и «случайности», рассматривавшихся в настоящей лекции, со сложностью и случайностью, определяемыми совершенно иначе в теории алгоритмов и в теории вероятностей.

Статистическая точка зрения на эти понятия состоит в том, что общие статистические законы (вроде «закона больших чисел» о стремлении частот успеха в повторяющихся испытаниях к вероятности успеха в одном испытании) выполняются лишь для *большинства* последовательных испытаний.

Большинство изучаемых нами последовательностей тоже «сложны» в нашем смысле, потому что бóльшая часть вершин соответствующего графа расположена на компонентах с длинными циклами, и притом на верхних ветвях оснащающих эти циклы деревьев.

Гипотеза состоит, однако, в том, что критерии стохастичности выполняются (с тем большей точностью, чем больше длина n рассматриваемых последовательностей) не только для типичных последовательностей (т. е. для большинства их), но и для нетипичных последовательностей, которые «сложны» в нашем смысле.

Некоторые статистически «типичные» объекты могут оказаться не сложными в нашем смысле, но они, предположительно, составляют малую долю всех (как и статистически нетипичные «сложные» в нашем смысле последовательности, если они существуют).

Можно также предполагать, что конечные модели алгоритмически невычислимых последовательностей окажутся, как правило, более сложными в нашем смысле, чем аналогичные конечные модели алгоритмически вычислимых последовательностей.

Но эта гипотеза не только не доказана, но и не сформулирована пока достаточно точно, чтобы ее можно было пытаться доказывать (и формулировки, и доказательства я ожидаю от слушателей, которым для того о них и рассказывал).

Многие математики считают, что понять теорему можно только обобщив ее, чтобы найденные закономерности оказались распространенными на более широкий круг явлений.

Поэтому я не остановился на описанной выше теории сложности последовательностей двоичных цифр, а провел аналогичные эксперименты для последовательностей, состоящих из иных объектов, например, для тренарных последовательностей, состоящих из остатков от деления на 3 (т. е. заменяя остатки {0, 1} от деления на 2 остатками {0, 1, 2} от деления на 3).

Вот простейшие результаты этих экспериментов (относящиеся к последовательностям из $n \leqslant 7$ знаков, составляющим множество из 3^n элементов).

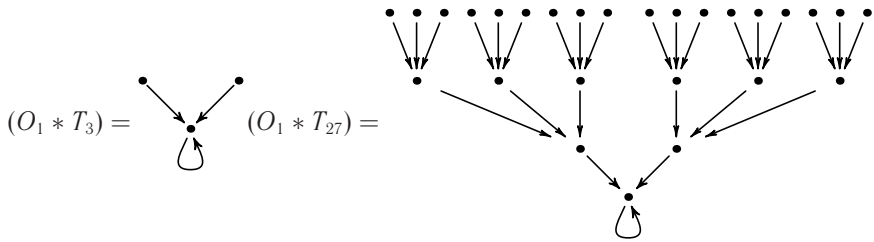
Оператор $A = \delta - 1\colon \mathbb{Z}_3^n \to \mathbb{Z}_3^n$ действует по прежней формуле

$$(A_x)_k = x_{k+1} - x_k \quad (\text{где } x_{n+1} = x_1).$$

Прямое вычисление графов этих операторов приводит к следующей таблице ответов:

n	b	компоненты графа	соотношение
2	1	$(O_1 * T_3)$	$A^2 = A$
3	1	$(O_1 * T_{27})$	$A^3 = 0$
4	6	$3(O_8 * T_3) + 3(O_1 * T_3)$	$A^9 = A$
5	2	$(O_{80} * T_3) + (O_1 * T_3)$	$A^{81} = A$
6	11	$8(O_3 * T_{27}) + 3(O_1 * T_{27})$	$A^6 = A$
7	3	$2(O_{364} * T_3) + (O_1 * T_3)$	$A^{365} = A$

В тренарных деревьях (T_3 и T_{27}) каждая вершина, кроме самых верхних, имеет 3 прообраза:



Я предоставляю читателю перенести на случай остатков от деления на p предыдущие теоремы о бинарных последовательностях, например, исследовав «дерево многочленов» $(O_1 * T_{3^k})$ в каждой компоненте предыдущей таблицы.

Период $T = 3$ самого длинного цикла графа операции $A\colon \mathbb{Z}_3^6 \to \mathbb{Z}_3^6$ (в случае $n = 6$) не делится на n, так что обобщение соответствуюего свойства бинарных последовательностей не тривиально.

Случай $n = 7$ нашей таблицы явно связан с астрономией года из $364 = (2^6 - 1)/2$ дней и $364/28 = 13$ месяцев.

Дубнинская лекция 2005 года породила продолжающие ее исследования О. Карпенкова и А. Гарбера. Последний, в частности, обнаружил, что если $n = p - 1$ не делится на 8, то «логарифм» сложен (например, для

$p = 4k + 3$), но для $p = 73$ «логарифм» принадлежит циклу. О. Карпенков обнаружил, что дробь T/n — не всегда уменьшенная на 1 степень двойки. Например, для $n = 23$: $T = 2047$, $T/n = 89$. Я надеюсь, что все эти результаты будут вскоре опубликованы. См также: Arnold V. Complexity of finite sequences of zeros and ones and geometry of finite spaces of functions // Functional Analysis and other mathematics. 2006. V. 1, № 1. 2006. pp. 1–18.

ЛЕКЦИЯ 3

СЛУЧАЙНЫЕ ПЕРЕСТАНОВКИ И ДИАГРАММЫ ЮНГА ИХ ЦИКЛОВ

Каждый считает себя экспертом в том, что знает хуже всего.

О. Уайльд

Опубликованный манускрипт подобен публичной женщине.

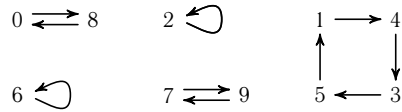
Дж. Свифт (*согласно В. Набокову*)

Каждая перестановка n элементов разбивается на циклы. Например, для следующей перестановки, переводящей цифру x в цифру y:

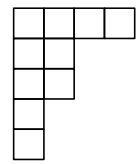
x	0	1	2	3	4	5	6	7	8	9
y	8	4	2	5	3	1	6	9	0	7

,

циклов 5, а именно



Таким образом, получается разбиение числа n на длины циклов: $10 = 4 + 2 + 2 + 1 + 1$. Разбиение натурального числа на натуральные слагаемые обычно изображается своей «диаграммой Юнга», имеющей в нашем примере вид



Для разбиения $n = x_1 + \ldots + x_y$, $x_1 \geqslant x_2 \geqslant \ldots \geqslant x_y$ в первой строчке диаграммы Юнга стоят x_1 единичных квадратов, во второй x_2, и т. д. до самой короткой из y строк.

Числа x_1 и y называются *длиной* и *высотой* диаграммы. Диаграмма Юнга разбиения десяти цифр на циклы нашей перестановки имеет, таким образом, длину $x = 4$ и высоту $y = 5$. Она заполняет в описанном вокруг нее прямоугольнике (со сторонами x и y) площадь n и долю $\lambda = n/(xy)$

($\lambda = 10/20 = 1/2$ в нашем примере). Значение λ я буду называть *полнотой* диаграммы.

В этой лекции мы исследуем диаграммы Юнга некоторых специальных перестановок.

Всего перестановок n элементов имеется $n!$ штук, и интересно, какие диаграммы Юнга разбиения n на циклы встречаются среди этих $n!$ «случайных» перестановок чаще, а какие реже: каковы в среднем длина, ширина, полнота диаграммы Юнга, что обычно больше, длина или ширина?

При небольших n эти вопросы можно решить, перечислив все диаграммы Юнга площади n всех $n!$ перестановок, но их число p_n с ростом n становится большим:

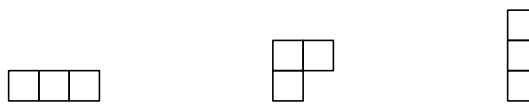
n	1	2	3	4	5	6	7	8	9	10
p_n	1	2	3	5	7	11	15	22	30	42

При больших значениях n проще пойти путем эвристического эксперимента: создать искусственно одну «случайную» перестановку n элементов (скажем, с $n = 100$) и сосчитать ее диаграмму Юнга (предполагаю, что диаграммы Юнга большинства из 100! перестановок 100 элементов похожи между собой, что, конечно, вовсе не очевидно, как не очевидна и типичность искусственно созданной перестановки). Для проверки типичности стоит повторить эксперимент несколько раз и сравнить, похожи ли ответы.

Кроме того, мы сравним «случайные» перестановки с доставляемыми алгеброй и теорией чисел перестановками точек в динамических системах с конечным фазовым пространством (что моделирует также конечные циклы периодических точек в динамических системах с дискретным временем).

§ 1. Статистика диаграмм Юнга перестановок небольшого числа элементов

При небольших n не составляет труда перечислить все разбиение n на натуральные слагаемые. Например, при $n = 3$ их всего 3, с диаграммами Юнга



а именно $3 = 2 + 1 = 1 + 1 + 1$.

Для сокращения обозначений я буду записывать разбиение $n = x_1 + \ldots + x_y$ в виде «одночлена»

$$a^{\alpha} b^{\beta} \ldots, \quad a > b > \ldots,$$

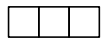
если наибольшее слагаемое $x_1 = a$ повторено в разбиении α раз, следующее по величине — β раз и так далее.

Три перечисленных разбиения числа $n = 3$ соответствуют одночленам $D = 3$, 2.1 и 1^3 (показатель 1 не указывается).

Группа $S(3)$ из шести перестановок трех элементов может рассматриваться как группа симметрий правильного треугольника (переставляющая его три вершины).

Она состоит из тождественного преобразования, двух поворотов (на 120° и 240°) третьего порядка и трех симметрий второго порядка (отражений относительно высот треугольника).

Диаграммы Юнга этих преобразований имеют, соответственно, вид


для поворотов;


для отражений;


для тождественного преобразования.

Мы запишем эти выводы в виде таблицы.

Случай $n = 3$. *Классы сопряженных элементов симметрической группы $S(3)$ (симметрий правильного треугольника).*

D	3	2.1	1^3
x	3	2	1
y	1	2	3
N	2	3	1

.

В строке N указано число перестановок с данной диаграммой Юнга D. Все эти перестановки, при данной диаграмме Юнга, «одинаковы» (подобны в группе $S(n)$ перестановок n элементов) и различаются только обозначениями переставляемых элементов: они «неразличимы с релятивистской точки зрения».

Общее число всех перестановок есть сумма по всем диаграммам площади n,

$$n! = \sum_D N(D),$$

так как каждая перестановка единственным образом разбивается на циклы.

Подсчитать число $N(D)$ представлений перестановками для данной диаграммы Юнга D — элементарная комбинаторная задача, нужно только указать число способов расставить символы $(1, 2, \ldots, n)$ в клетках диаграммы Юнга, приводящих к разным перестановкам.

Для диаграммы из одного цикла длины n (где $D = n$) число разных перестановок составляет $N(n) = (n-1)!$, например, $N(4) = 6$.

Действительно, в цикле после обязательно входящего в него элемента 1 следует один из остальных $(n-1)$ элементов, затем — один из оставшихся $(n-2)$ элементов, и т. д., так что число всех разных циклических перестановок множества из n элементов равно $(n-1)(n-2)\ldots = (n-1)!$.

Слушатели легко смогут сами доказать результаты о перестановках $n \leqslant 7$ элементов аналогичные приведенной выше для $n = 3$ таблице.

Теорема 1. *Разбиения групп $S(n)$ перестановок n элементов на классы сопряженных элементов (с диаграммами Юнга длины x и высоты y для $N(D)$ сопряженных друг другу перестановок) доставляются при $n \leqslant 7$ следующими таблицами.*

Случай $n = 4$. *Классы сопряженных элементов симметрической группы $S(4)$ (симметрий тетраэдра).*

D	4	3.1	2^2	2.1^2	1^4
x	4	3	2	2	1
y	1	2	2	3	4
N	6	8	3	6	1

.

Столбец $N(2^2) = 3$ соответствует трем осям симметрии тетраэдра, соединяющим середины противоположных (скрещивающихся) ребер. Эти три симметрии тетраэдра, вместе с его тождественным преобразованием, образуют замечательную коммутативную подгруппу $\mathbb{Z}_2 \times \mathbb{Z}_2 \subset S(4)$, благодаря которой алгебраические уравнения степени 4 решаются в радикалах.

Случай $n = 5$. *Классы сопряженных элементов симметрической группы $S(5)$.*

D	5	4.1	3.2	3.1^2	$2^2.1$	2.1^3	1^5
x	5	4	3	3	2	2	1
y	1	2	2	3	3	4	5
N	24	30	20	20	15	10	1

.

Эта таблица доставляет всю геометрию группы симметрий додекаэдра, с его 12 гранями, 20 вершинами и 30 ребрами.

Каждая грань порождает два циклических вращения пятого порядка, каждая вершина — сохраняющие ее вращения порядка 3, каждое ребро — сохраняющее пару противоположных ребер вращение порядка 2.

Чтобы интерпретировать симметрии додекаэдра как перестановки 5 элементов, Кеплер вписал в додекаэдр пять кубов, ребра которых являются диагоналями граней додекаэдра. Вот как он их строил (рис. 1).

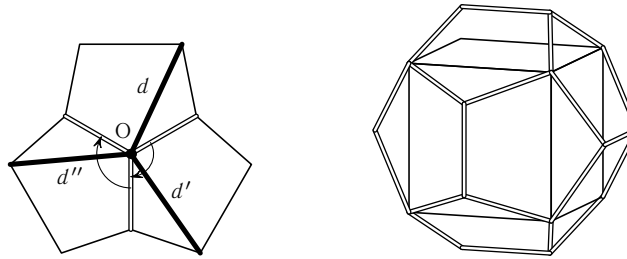


Рис. 1. Вписанный в додекаэдр куб Кеплера

Выбрав диагональ d одной из граней, можно провести через ее концевую вершину O две диагонали d', d'' двух соседних граней, получающиеся из выбранной диагонали сохраняющими упомянутую вершину вращениями додекаэдра. Полученные 3 диагонали граней ортогональны друг другу.

Повторяя эту конструкцию в свободных концах уже полученных трех диагоналей граней додекаэдра, мы будем строить новые диагонали граней, пока построенные 12 диагоналей граней не составят куб (имеющий по ребру на каждой из 12 граней додекаэдра).

В зависимости от того, с какой из пяти диагоналей исходной грани началась эта конструкция, получатся пять разных вписанных в додекаэдр кубов Кеплера, которые и переставляются группой симметрий додекаэдра.

Случай $n = 6$. *Классы сопряженных элементов симметрической группы* $S(6)$.

D	6	5.1	4.2	4.1^2	3^2	3.2.1	3.1^3	2^3	$2^2.1^2$	2.1^4	1^6
x	6	5	4	4	3	3	3	2	2	2	1
y	1	2	2	3	2	3	4	3	4	5	6
N	120	144	90	90	40	120	40	15	45	15	1

.

Случай $n = 7$. *Классы сопряженных элементов симметрической группы* $S(7)$.

D	7	6.1	5.2	5.1^2	4.3	4.2.1	4.1^3	
N	720	840	504	504	420	630	210	

D	$3^2.1$	3.2^3	$3.2.1^2$	3.1^4	$2^3.1$	$2^2.1^3$	2.1^5	1^7
N	280	210	420	70	105	105	21	1

.

Докажем, например, что $N(3^2.1) = 280$. Цикл длины 1 может составляться любым из 7 элементов и определен перестановкой с диаграммой $D = 3^2.1$ однозначно. Остающиеся 6 элементов нужно разбить на два цикла по 3 элемента. Одна тройка выбирается $C_6^3 = 20$ способами, но разных разбиений получается 10, так как неизвестно, которая из двух троек первая.

Если разбиение фиксировано, то остается выбрать циклический порядок в каждой из троек (независимо от другой). Циклических порядков $m = 3$ элементов имеется $(m-1)! = 2$. Поэтому разных выборов циклических порядков в обеих тройках $2 \cdot 2 = 4$, а всего нужных перестановок данных 6 элементов получается $4 \cdot 10 = 40$.

Учитывая произвольность седьмого неподвижного при перестановке элемента, получаем общее число перестановок с диаграммой $D = 3^2.1$ равным произведению $40 \cdot 7 = 280$.

Остальные элементы таблицы вычисляются аналогично.

Приведенные таблицы позволяют вычислить (при $n \leqslant 7$) для любой функции f от диаграммы Юнга D площади n ее среднее значение $\hat{f}$ по всем $n!$ перестановкам n элементов:

$$\hat{f}(n) = \sum_D (N(D) f(D))/n!.$$

Теорема 2. *Средние значения характеристик диаграмм* (*длины* x, *высоты* y, *полноты* $\lambda = n/(xy)$, *асимметрии* $\mu = y/x$) *имеют при* $n \leqslant 7$ *следующие значения*:

n	$\hat{x}$	$\hat{y}$	$\hat{\lambda}$	$\hat{\mu}$
2	$3/2 = 1{,}50$	$3/2 = 1{,}50$	$1{,}00$	$5/4 = 1{,}25$
3	$2\frac{1}{6} \approx 2{,}17$	$1\frac{5}{6} \approx 1{,}83$	$7/8 \approx 0{,}88$	$10/9 \approx 1{,}11$
4	$2\frac{19}{24} \approx 2{,}79$	$2\frac{1}{12} \approx 2{,}08$	$\frac{29}{56} \approx 0{,}80$	$\frac{137}{144} \approx 0{,}95$
5	$3\frac{17}{40} \approx 3{,}42$	$2\frac{17}{60} \approx 2{,}28$	$\frac{325}{432} \approx 0{,}75$	$\frac{184}{225} \approx 0{,}82$
6	$4\frac{31}{720} \approx 4{,}04$	$2\frac{408}{720} \approx 2{,}57$	$\approx 0{,}72$	$\approx 0{,}76$
7	$\approx 4{,}68$	$\approx 2{,}71$	$\approx 0{,}69$	$\approx 0{,}70$

Хотя эта статистика получена лишь при довольно небольших значениях n, она подсказывает целый ряд предположений.

Например, можно предполагать, что средняя длина диаграммы растет с ростом ее площади n приблизительно линейно, $\hat{x} \sim c_1 n$ (таблица подсказывает значение коэффициента c_1, близкое к $2/3$).

Напротив того, средняя высота диаграммы растет с ее площадью гораздо медленнее, предположительно даже

$$\hat{y} \sim x_2 \ln n$$

(высота примерно удваивается при возведении n в квадрат).

Средний коэффициент заполнения прямоугольника диаграммой медленно убывает с ростом площади n, таблица подсказывает его поведение типа

$$\hat{\lambda} \sim c_3/(\ln n),$$

(с уменьшением коэффициента заполнения примерно вдвое при возведении площади n в квадрат).

Величина средней асимметрии μ убывает в том смысле, что, по мере роста площади n, типичные диаграммы уплощаются и становятся все более низкими и длинными.

Однако, квадратичное среднее значении логарифма асимметрии μ составляет $\sigma \approx 0{,}30$ при $n = 2$ и вырастает до $\sigma \approx 0{,}42$ при $n = 7$. Этот рост величины σ показывает, что довольно значительная асимметрия сохраняется у значительной доли диаграмм и при большой площади n, то есть что диаграммы не все становятся при росте площади более симметричными, а, напротив, достигают бóльшего разнообразия форм, включающего и длинные диаграммы с $\mu < 1$, и высокие диаграммы с $\mu > 1$ (какая из этих двух асимметрий имеет место, средняя квадратичная величина σ не различает).

§2. Экспериментирование со случайными перестановками большего числа элементов

Хотя компьютерные вычисления могли бы продолжить вычисление средних характеристик диаграмм Юнга перестановок при бóльших, чем в § 1, значениях n, эти вычисления, требующие суммирования по всем $n!$ перестановкам n элементов, останутся нереальными (даже с использованием компьютеров) при таких, например, значениях n, как 100: $100!$ слагаемых сложить не удастся.

Поэтому я изобрел другой (естественно-научный скорее, чем математический) подход к этой задаче: вместо суммирования по всем $100!$ перестановкам, я реализовал одну перестановку сотни элементов, выбрав ее действительно случайным образом и рассматривая характеристики ее диаграммы Юнга разбиения числа $n = 100$ на длины циклов как типичные характеристики диаграммы площади n «случайной» перестановки.

Мой способ построения «случайной» перестановки n элементов был таким (я опишу его ниже для $n = 100$, чтобы упростить обозначения).

Начнем с какой-либо последовательности случайных цифр. Я опишу ниже несколько источников таких последовательностей: использовались телефонные номера членов Национальной Академии наук США в одном случае и номера автомобилей, проезжавших на улице Вавилова мимо математического института Российской академии наук в другом, и результаты получились сходные.

Если в случайной последовательности цифр $\alpha\beta\gamma\delta\ldots$ рассматривать $(\alpha\beta)$ как двузначное число, то мы начинаем перестановку (то есть упорядочение) всех 100 двузначных чисел (00, 01, ..., 99) с числа $(\alpha\beta)$. Если число $(\gamma\delta)$ отлично от $(\alpha\beta)$, то мы поставим его вторым членом переставленной последовательности двузначных чисел, а если нет, то пропустим его и перейдем к следующему числу, $(\varepsilon\zeta)$. И так далее: если несколько первых членов переставленной последовательности уже выбраны, то в качестве следующего ее элемента будет браться первая из последующих пар членов исходной последовательности случайных цифр, которая отлична от всех уже выбранных пар.

Этот алгоритм создает в конце концов случайную перестановку всех 100 двузначных чисел. Но, по мере приближения к концу, он работает все медленнее и медленнее, так как до встречи нового, (еще не выбиравшегося) двузначного числа приходится долго двигаться среди уже повторяющихся кандидатов, испытывая их на новизну одного за другим.

А именно, если мы переставляли бы n элементов, то число кандидатов, которых пришлось бы рассматривать, было бы, как это объяснено ниже, порядка $n \ln n$. Для $n = 100$ получаем $\ln 100 \approx 4{,}6$, то есть нужно располагать таблицей примерно 500 случайных двузначных чисел.

Список академиков для этого достаточно велик, и я выбирал в качестве «случайных» двузначных чисел пары цифр, образованные четвертой и пятой цифрой семизначного телефонного номера академиков (упорядоченных в справочнике по алфавиту).

В следующем протоколе такого эксперимента подряд выписаны выбранные двузначные числа, а отвергнутые кандидаты заключены в скобки. В этом эксперименте в 100 попытках выбрано 64 двузначных числа, и частота повторяющихся кандидатов в конце работы сильно ее тормозит.

47	99	07	32	02	91	52	66	21	81	27	82	70
43	17	65	76	28	63	08	94	11	01	95	(52)	(76)
87	(65)	29	16	20	80	10	25	37	(65)	(32)	35	(21)
74	05	36	48	(24)	73	(48)	90	18	75	12	(02)	15
41	72	38	61	(73)	(73)	(63)	(11)	24	83	56	(32)	(74)

06 84 (56) (81) 67 14 03 (83) (56) 96 (48) (27) (37)
97 (08) (37) 89 (02) (97) (38) (52) 44 19 (24) (28) (12)
(01) 13 69 (20) (17) (84) 88 53 (61).

Чтобы бороться с этим замедлением в конце выбора, я изобрел еще несколько усовершенствований. Во-первых, можно вместо n выбрать только $n/2$ элементов, а потом из оставшихся $n/2$ кандидатов выбрать $n/2$ элементов тем же приемом. Например, можно как-либо биективно отобразить оставшееся множество на уже готовые $n/2$ переставленных элементов, а затем упорядочить их при помощи уже готового упорядочения $n/2$ выбранных элементов.

Другой способ состоит в том, чтобы использовать в качестве новой таблицы случайных двузначных чисел (для второй половины упорядочения) не исходную последовательность $(\alpha\beta)(\gamma\delta)\ldots$, а последовательность $(\beta\gamma)(\delta\varepsilon)\ldots$.

Третий способ состоит в использовании для упорядочения $m = 100/k$ элементов последовательность остатков от деления исходных случайных двузначных чисел $(\alpha\beta)$, $(\gamma\delta)$, ... на делитель m числа 100.

Каждый из этих способов позволяет быстро «случайно» упорядочить все $n = 100$ двузначных чисел, что и задает нужную «случайную» перестановку множества из n элементов.

При $n = 16$ я получил таким путем из случайной последовательности остатков от деления на 16 следующую перестановку элементов $\{0, 1, \ldots, 15\}$:

$$0\ 4\ 3\ 12\ 9\ 8\ 7\ 14\ 5\ 1\ 2\ 11\ 6\ 15\ 10\ 13.$$

Соответствующие перестановке циклы легко вычислить:

$0 \to (0)$ длины 1;

$1 \to 4 \to 9 \to (1)$ длины 3;

$2 \to 3 \to 12 \to 6 \to 7 \to 14 \to 10 \to (2)$ длины 7:

$5 \to 8 \to (5)$ длины 2;

$11 \to (11)$ длины 1;

$13 \to 15 \to (13)$ длины 2.

Итак, диаграмма Юнга нашей «случайной» перестановки 16 элементов есть $D = 7.3.2^2.1^2$.

Параметры этой диаграммы имеют значения

$$x = 7,\ y = 6, \quad \lambda = 16/42 \approx 0{,}38, \quad \mu = 6/7 \approx 0{,}86.$$

Они удовлетворительно укладываются в гипотетическое поведение средних характеристик, предсказанное на с. 72 на основании таблиц, где число элементов было $n \leqslant 7$.

Замечание. Выражение $n \ln n$ для числа наших попыток случайно упорядочить n элементов имеет следующее эвристическое объяснение.

Последний из n выбираемых элементов встречается при продолжении нашей случайной последовательности с вероятностью $1/n$, так что его появление ожидается в среднем после n попыток.

Для выбора предпоследнего элемента удачей будет встреча любого из двух еще не выбранных элементов из n возможных. Ожидаемое число попыток, стало быть, есть $n/2$.

Выбор предыдущего элемента потребует в среднем $n/3$ попыток, и так далее. Итого общее число попыток ожидается в среднем такое:

$$n\left(\sum_{k=1}^{n}\frac{1}{k}\right) \sim n \ln x|_1^n \sim n \ln n$$

(мы использовали соотношение $\int \frac{dx}{x} = \ln x$).

Сходное эвристическое рассуждение показывает, почему число циклов у перестановки n элементов ожидается порядка $c_2 \ln n$.

Естественно думать, что в случайной последовательности независимых выборов из n элементов первое повторение наступит после cn попыток, где c — некоторая константа. Это дает ожидаемую длину cn первого цикла.

С этого момента выбор производится из $n_1 \equiv n - cn = (1-c)n$ кандидатов. Это подсказывает для второго цикла ожидаемую длину cn_1, и после создания останется $n_2 = (1-c)n_1 = (1-c)^2 n$ кандидатов после создания второго цикла.

Таким же образом, после создания y циклов число оставшихся кандидатов будет $n_y = (1-c)^y n$.

Но вся процедура заканчивается при n_y порядка 1. Это соотношение доставляет для числа циклов y следующее выражение:

$$y \ln(1-c) + \ln n \approx 0, \quad y \approx \frac{\ln n}{\ln(1-c)^{-1}} \sim c_2 \ln n.$$

Проделав довольно много подобных экспериментов, я получил следующие диаграммы Юнга «случайных» перестановок n элементов:

n	D	x	y	λ	μ
16	$7.3.2^2.1^2$	7	6	0,38	0,86
25	9.7.5.3.1	9	5	0,56	0,56
64	$35.15.7.3.2.1^2$	35	7	0,26	0,20
100	$42.36.18.2.1^2$	42	6	0,40	0,14
100	90.4.3.2.1	90	5	0,22	0,06
100	$88.9.1^3$	88	5	0,23	0,06
169	147.13.8.1	147	4	0,29	0,03

Все эти характеристики эмпирически «случайных» перестановок удовлетворительно согласуются с гипотезами с. 72. Но, конечно, дальнейшая экспериментальная проверка (особенно с бóльшими значениями n) была бы очень желательна, с одной стороны, и не требует практически никаких предварительных знаний и доступна школьникам — с другой.

Все же я провел еще одну независимую проверку, использовав как источник случайных перестановок таблицу полей Галуа из p^2 элементов.

§3. Случайные перестановки p^2 элементов, порожденные полями Галуа

Построение этих перестановок объяснено в § 4 лекции 2. Напомню, что поле Галуа из p^2 элементов состоит из матриц

$$0 \text{ и } A^k,\ 1 \leqslant k < p^2, \quad \text{где } A^k = u_k A + v_k;$$

$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ — матрица, элементы которой являются остатками от деления на простое число p и такая, что

$$A^{p^2-1} = 1, \quad A^{k<p^2-1} \neq 1.$$

Перестановка множества из p^2 элементов $1 \leqslant k \leqslant p^2$ определяется этим полем так: элементу $k < p^2$ сопоставляется адрес $K = u_k p + v_k$ матрицы $A^k = u_k A + v_k$ в таблице поля, а для $k = p^2$ полагаем $K = 0$ (так что $u_{p^2} = v_{p^2} = 0$).

Заметим, что адреса элементов таблицы поля пробегают множество $0 \leqslant K \leqslant (p-1)p + (p-1) = p^2 - 1$ из p^2 точек (так что мы можем истолковывать и k, и K как остатки от деления на p^2).

Отображение $k \to K$ является перестановкой этого множества из p^2 элементов, и мы можем применить к этой перестановке предыдущую теорию: разложить ее на циклы, построить диаграмму Юнга, найти ее характеристики — длину x, высоту y, полноту λ, асимметрию μ.

Сделав эти вычисления при $p \leqslant 13$, я получил удивительно похожие на диаграммы Юнга случайных перестановок ответы, приведенные в следующей таблице.

Теорема 1. *Диаграммы Юнга перестановок таблиц полей Галуа из p^2 элементов таковы*:

p	n	D	x	y	λ	μ
3	9	$6.2.1$	6	3	0,50	0,50
5	25	$14.5.4.1^2$	14	5	0,36	0,36
7	49	$16.11.7.6.4.3.1^2$	16	8	0,38	0,50
11	121	$65.39.5.3^3.2.1$	65	8	0,25	0,12
13	169	$98.55.12.2.1^2$	98	6	0,29	0,06

Сравнение этой таблицы с предыдущей таблицей можно считать подтверждением гипотезы о том, что таблицы полей Галуа обладают, особенно при больших p, статистическими свойствами таблиц случайных чисел (например, что задаваемые этими таблицами перестановки элементов поля обладают свойствами типичных случайных перестановок).

Для этого вывода нет, кроме приведенных эмпирических данных, никаких теоретических оснований, и это явилось одной из причин, почему я включил описание этих экспериментов в настоящие лекции.

С другой стороны, приведенная таблица доставляет новый довод в пользу гипотез с. 72 для случайных перестановок (если мы согласимся считать перестановку, заданную таблицей поля Галуа, случайной).

§4. Статистика циклов автоморфизмов Фибоначчи

В качестве еще одного алгебраического примера перестановок конечных множеств мы рассмотрим периодические точки специальных динамических систем на двумерном торе — так называемых гиперболических автоморфизмов.

Автоморфизм тора $T^2 = \mathbb{R}^2/\mathbb{Z}^2$ задается целочисленной матрицей $A \in \in \mathrm{SL}(2, \mathbb{Z})$ определителя 1. Линейное преобразование $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ переводит решетку целочисленных векторов $\mathbb{Z}^2$ в себя, а потому диффеоморфизм переводит в себя и тор.

Периодические точки этого автоморфизма (который мы будем попрежнему обозначать A) — это решения уравнений

$$A^k x = x \ (\in T^2).$$

Мы будем предполагать, что они изолированы, т. е. что $\det(A^k - 1) \neq 0$. В этом случае периодические точки принадлежат конечным торам

$$M \simeq \mathbb{Z}_n^2,$$

состоящим из точек с рациональными координатами

$$z = (z_1, z_2), \quad z_j = u_j/n, \quad u_j \in \mathbb{Z}_n = \mathbb{Z}/(n\mathbb{Z}).$$

Число таких точек (с данным знаменателем n) конечно:

$$|\mathbb{Z}_n^2| = n^2.$$

Отображение A переставляет точки конечного множества M, и эта перестановка разбивается на циклы.

Мы собираемся теперь исследовать диаграммы Юнга этих разбиений.

Рассмотрим специальный *автоморфизм Фибоначчи* (связанный также с золотым сечением), соответствующий матрице $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$:

$$A\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 2z_1 + z_2 \\ z_1 + z_2 \end{pmatrix}. \tag{1}$$

Чтобы упростить обозначения, мы можем умножить (дробные) координаты на общий знаменатель n, то есть считать компоненты z_1 и z_2 остатками от деления на n:

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \in M = \mathbb{Z}_n^2.$$

Диаграмма Юнга автоморфизма $A\colon M \to M$ конечного тора M в себя имеет площадь n^2, и мы собираемся исследовать ее форму для автоморфизма Фибоначчи при разных M.

Замечание. Автоморфизм A я называю автоморфизмом Фибоначчи потому, что он действует на базисные векторы следующим образом:

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} 3 \\ 2 \end{pmatrix} \mapsto \begin{pmatrix} 8 \\ 5 \end{pmatrix} \mapsto \begin{pmatrix} 21 \\ 13 \end{pmatrix} \mapsto \begin{pmatrix} 55 \\ 34 \end{pmatrix} \mapsto \dots$$

Компоненты этих векторов образуют последовательность Фибоначчи 1, 1, 2, 3, 5, 8, 13, ..., а их отношения стремятся к золотому сечению, $\dfrac{\sqrt{5}-1}{2} \approx 0,6$, (так как собственные числа матрицы A равны $\dfrac{3 \pm \sqrt{5}}{2}$).

Говорят также, что диффеоморфизм $A\colon T^2 \to T^2$ готовит из кошки окрошку, так как образ вложенной в T^2 кошки C становится после применения нескольких итераций сохраняющего площади отображения A «размазанным по тору» очень равномерно (рис. 2).

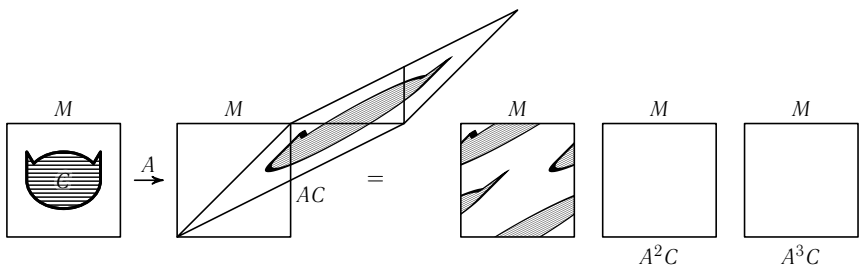


Рис. 2. Приготовление окрошки из кошки

Прямым подсчетом итераций отображения A доказывается

Теорема 1. *Диаграмма Юнга D действия автоморфизма Фибоначчи* (1) *на конечные торы $\mathbb{Z}_n^2$ при $n \leqslant 20$ имеют следующие пара-*

метры:

n	D	x	y	$\lambda \approx$	$\mu \approx$
2	3.1	3	2	0,67	0,67
3	$4^2.1$	4	3	0,75	0,75
4	$3^5.1$	3	6	0,89	2,00
5	$10^2.2^3.1$	10	5	0,50	0,50
6	$12^2.4^2.3.1$	12	6	0,50	0,50
7	$8^6.1$	8	7	0,88	0,88
8	$6^8.3^5.1$	6	14	0,76	2,33
9	$12^6.4^2.1$	12	9	0,75	0,75
10	$30^2.10^2.6^2.3.2^2.1$	30	10	0,33	0,33
11	$5^{24}.1$	5	25	0,97	5,00
12	$12^{10}.4^2.3^5.1$	12	18	0,67	1,5
13	$14^{12}.1$	14	13	0,93	0,93
14	$24^6.8^6.3.1$	24	14	0,58	0,58
15	$20^8.10^2.4^{10}.2^2.1$	20	23	0,49	1,15
16	$12^{16}.6^8.3^5.1$	12	30	0,71	2,50
17	$18^{16}.1$	18	17	0,94	0,94
18	$12^{26}.4^2.3.1$	12	30	90,90	2,50
19	$9^{40}.1$	9	41	0,98	4,8
20	$30^{10}.10^2.6^{10}.3^5.2^2.1$	30	30	0,44	1,00
41	$20^{84}.1$	20	85	0,99	4,25
97	$98^{96}.1$	98	97	0,99	0,99

Замечание. Эта теорема подсказывает ряд гипотез, например, для простых значений $n = p > 5$ диаграммы имеют стандартный простой вид с символом $D = x^z.1$, $z = y - 1$. Видна также связь мультипликативного характера, $D(pq)$ связано с $D(p)$ и $D(q)$.

Точные формулировки ряда таких гипотез я оставляю слушателям (надеясь, что они даже докажут со временем некоторые из них).

Некоторые специальные диаграммы для некоторых специальных автоморфизмов были вычислены раньше Персивалем, Вивальди, Дайсоном и другими. Но их специальные примеры составляют малую долю всех случаев, и я не решался бы делать из них какие-либо выводы о типичных диаграммах длин циклов (даже для случая автоморфизма Фибоначчи при типичных значениях n).

Вместе с конечным тором $M = \mathbb{Z}_p^2$ можно рассмотреть конечную проективную прямую

$$P = P^1(\mathbb{Z}_p) = (\mathbb{Z}_p^2 \setminus 0)/(\mathbb{Z}_p \setminus 0),$$

она состоит из $p+1$ точки.

Линейный оператор $A\colon M \to M$ действует на P как специфическая (проективная) перестановка этих точек,

$$A_p \in \mathrm{GP}(\mathbb{Z}_p) \subset S(p+1).$$

Разбиение этой проективной прямой на циклы проективной перестановки A_p определяет диаграмму Юнга (площади $p+1$).

Теорема 2. *Диаграммы Юнга циклов проективных перестановок A_p, порожденных автоморфизмами Фибоначчи $A(z_1, z_2) = (2z_1 + z_2,\ z_1 + z_2) \pmod p$ имеют при $p < 20$ следующие значения параметров*:

p	D	x	y	$\lambda \approx$	$\mu \approx$
2	3	3	1	1,00	0,33
3	2^2	2	2	1,00	1,00
5	5.1	5	2	0,60	0,40
7	4^2	4	2	1,00	0,50
11	$5^2.1^2$	5	4	0,60	0,80
13	7^2	7	2	1,00	0,29
17	9^2	9	2	1,00	0,22
19	$9^2.1^2$	9	4	0,53	0,44
41	$10^4.1^3$	10	6	0,70	0,60
97	49^2	49	2	1,00	0,04

Теорема 2 доказывается прямыми вычислениями вместе с теоремой 1, но ее можно и получить из теоремы 1, факторизуя действие перестановки A по подгруппе скаляров.

Впрочем, можно и обратно, начать с более простых вычислений теоремы 2, а потом надстраивать проективные перестановки линейными операторами.

Персиваль и Вивальди шли именно этим путем, но скрыли всю простую проективную геометрию ситуации за сложной алгеброй теории расширения полей.

Сравнивая теоремы 1 и 2 с характеристиками случайных перестановок из §2 и §3, мы замечаем, что диаграммы Юнга автоморфизмов конечных торов сильно отличаются и от диаграмм типичных случайных перестановок того же числа элементов, и от диаграмм Юнга перестановок, определенных таблицами полей Галуа.

А именно, параметр полноты λ в случае автоморфизмов принимает значительно бóльшие значения, причем наблюдающееся для случайных перестановок уменьшение полноты по мере роста площади диаграммы в случае автоморфизма Фибоначчи, видимо, отсутствует.

Асимметрия μ диаграмм автоморфизмов тоже заметно выше, чем для случайных перестановок и для перестановок, порожденных таблицами полей Галуа. При этом диаграммы для автоморфизмов чаще бывают высокими ($\mu > 1$ в теореме 1), в то время как диаграммы случайных перестановок большого числа элементов обычно низкие ($\mu < 1$), а при увеличении площади диаграммы среднее отношение $\mu = y/x$ высоты диаграммы к длине стремится, кажется, к нулю.

Значительная асимметрия диаграмм Юнга автоморфизмов делает их статистику сильно отличающейся не только от равномерно-усредненной по всем $n!$ перестановкам статистики, которую мы вычисляем в §1, но и от статистики Вершика—Планшереля, в которой среднее значение отношения y/x равно 1.

На основании всего этого мы приходим к выводу, что распределение периодических точек автоморфизмов конечного тора по периодам приводит к новым асимптотикам диаграмм Юнга (по сравнению с двумя изученными — равномерной и планшерелевской) уже для автоморфизма Фибоначчи с матрицей $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$.

Можно думать, что сходные явления (может быть, даже с такой же универсальной асимптотикой при $n \to \infty$, а может быть и с зависимостью этой асимптотики от соответствующей матрице цепной дроби) имеют место и для других автоморфизмов торов.

Было бы интересно изучить поведение при $n \to \infty$ средних значений параметров x, y, λ, μ диаграмм Юнга циклов автоморфизмов A n^2-точечного тора M вдоль всей группы автоморфизмов (или хотя бы поведение их чезаровских средних при $n \to \infty$).

Было бы также интересно сравнить средние по всей симметрической группе $S(n)$ со средними по подгруппе проективных перестановок $n = p + 1$ точки конечной проективной прямой $P^1(\mathbb{Z}_p)$.

В случае автоморфизмов конечного m-мерного тора придется рассматривать «проективные перестановки» множества из $n = p^{m-1} + \ldots + p + 1$ точек конечного $m-1$-мерного проективного пространства $P^{m-1}(\mathbb{Z}_p)$. В этом случае следует, возможно, различать в этом конечном проективном пространстве его лобачевскую часть Λ (модели Клейна) и дополнительный конечный релятивистский мир де-Ситтера, $P^{m-1} \setminus \Lambda$.

Наконец, поведение всех этих объектов при стремлении к бесконечности размерности m может привести к новым интересным асимптотикам «больших диаграмм Юнга». Дело в том, что для автоморфизмов m-мерного тора

$$A\colon \mathbb{Z}_n^m \to \mathbb{Z}_n^m$$

эти асимптотические характеристики могут зависеть от многомерной цепной дроби оператора $A \in \mathrm{GL}(n, \mathbb{Z})$. Например, ответы могут зависеть от «триангуляции» выпуклыми целочисленными многогранниками непрерывного тора $(S^1)^{m-1}$, определяющей геометрию «периодов» многомерной цепной дроби оператора A.

В качестве «диаграмм Юнга циклов» в этом случае нужно использовать, вероятно, не только длины циклов оператора A, но и описание действия всей коммутативной группы $\mathbb{Z}^{m-1}$ симметрий периодической многомерной цепной дроби.

Даже средне значения параметров диаграмм Юнга циклов операторов A, усредненные по всей группе автоморфизмов (и уже не зависящие от цепной дроби) заслуживают внимательного изучения (хотя бы эмпирического, с помощью экспериментов, подобных описанным выше).

Замечание. Перестановка n^m точек конечного тора $\mathbb{Z}_n^m$, заданная матрицей $A \in \mathrm{SL}(m, \mathbb{Z})$, имеет определенный период $T(A, n)$, для которого

$$A^{T(A,n)} = 1 \pmod{n}.$$

Все длины циклов этой перестановки n^m элементов делят, очевидно, целое число $T(A, n)$.

Поэтому можно было бы сравнивать статистику диаграмм Юнга циклов операторов A (даже оператора Фибоначчи $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$) при $n \to \infty$ не со статистикой всех диаграмм Юнга всех перестановок n^m элементов, а со статистикой лишь тех из них, все длины циклов которых делят данное целое число $T(A, n)$.

Неясно, насколько эта статистика отличается от общей статистики всех $n^m!$ перестановок множества из $n^m = |\mathbb{Z}_n^m|$ элементов. Неясен и (несомненно, более легкий) вопрос о поведении периода $T(A, n)$ при $n \to \infty$ (ни для матрицы Фибоначчи $= \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$), ни для случайной матрицы A из $\mathrm{SL}(2, \mathbb{Z}_n)$, и для случайной матрицы из $\mathrm{SL}(m, \mathbb{Z}_n)$, ни для их (рассматривавшихся выше, на с. 80) проективных версий.

Все эти случаи интересны и доступны слушателям (по меньшей мере экспериментально). Теорема 1 доставляет для $n = (2, 3, \ldots, 20)$ следующие периоды $T(A, n)$ для матрицы Фибоначчи $= \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$:

$$T(A, n) = (3, 4, 3, 10, 12, 8, 6, 12, 60, 5, 12, 14, 24, 20, 12, 18, 12, 9, 60).$$

Я вычислил при $n = 5$ средние значения характеристик перестановок по всем перестановкам $n^2 = 25$ элементов, длины всех циклов которых делят

число $T\begin{pmatrix}2&1\\1&1\end{pmatrix}=10$, имеющим, как и перестановка Фибоначчи, ровно один цикл длины 1.

Число таких перестановок — примерно $25!!\cdot 2^2\cdot 3^5\cdot 7\cdot 11/5$, где $(25)!!=$ $=25\cdot 23\cdot 21\cdot\ldots\cdot 3\cdot 1$.

Средние значения параметров этих перестановок получились такие:

$$x_T\approx 10{,}00,\quad y_T\approx 6{,}96,\quad \lambda_T\approx 0{,}56,\quad \mu_T\approx 0{,}70.$$

Наблюденные для перестановок Фибоначчи при $n=5$ значения

$$x=10,\quad y=5,\quad \lambda=0{,}50,\quad \mu=0{,}50$$

ближе к приведенным выше средним по перестановкам дины циклов которых делят число T, чем к средним по всем 25! перестановкам 25 элементов, приведенным на стр. 75 и равным

$$\hat{x}\approx 9,\quad \hat{y}\approx 5,\quad \hat{\lambda}\approx 0,\,56,\quad \hat{\mu}\approx 0,\,56.$$

Но все же наблюденные значения сильно отличаются и от указанной выше статистики по перестановкам, длины циклов которых делят число T.

Поведение этих статистик при $n\to\infty$ остается неизвестным. Я не знаю даже, как ведет себя при больших n странная функция $T\left(\begin{pmatrix}2&1\\1&1\end{pmatrix},n\right)$ от числа n, определяющего переставлявшийся выше конечный тор из n^2 элементов. Приведенные выше первые два десятка значений этой функции не позволили мне предсказать какую-либо гипотезу о ее поведении, но слушатели могли бы здесь быстро продвинуться, хотя бы вычислив следующие значения периода T, при бóльших n.

Читатели лекций 2005 года уже продолжили мои вычисления (при помощи своих компьютеров). Они подтверждают справедливость моих гипотез о характеристиках случайных подстановок (с. 71), а для автоморфизмов Фибоначчи конечных торов приводят к росту длин и высот диаграмм порядка квадратного корня из числа точек конечного тора, при средней полноте примерно 0,8 и при средней асимметрии порядка того же квадратного корня.

Для перестановок Фибоначчи конечной проективной прямой их диаграммы Юнга оказались почти прямоугольниками (вида k^a или $k^a.1^2$ с четным a, которое равно 2 в 60% случаев, согласно М. Казаряну и В. Клепцыну).

Я надеюсь, что читатели докажут и эти новые гипотезы.

См. также статью: Arnold V. Statistics of Young Diagrams of Cycles of Dynamical Systems for Finite Tori Automorphisms // Moscow Mathematical Journal. 2006. V. 6, № 1. P. 43—56.

ЛЕКЦИЯ 4

ГЕОМЕТРИЯ ЧИСЕЛ ФРОБЕНИУСА ДЛЯ АДДИТИВНЫХ ПОЛУГРУПП

Математика подобна любви: в обоих случаях наибольшее удовольствие доставляет познание

«Темнота Ньютона»
(«Newton's Darkeness»,
C. Djerassi, Imperial College Press, 2003).

А ошибусь — мне это трын-трава:
Я все равно с ошибкой не расстанусь

Б. Л. Пастернак — А. А. Ахматовой

Предмет этой лекции относится к самым простым вопросам арифметики: какие числа можно получить из данных слагаемых, складывая их (в любом количестве)?

Пусть, скажем, имеются монеты достоинством 3 копейки (алтын) и 5 копеек (пятак).

Какие суммы можно составить из алтынов и пятаков?

Очевидно 1, 2 и 4 копейки составить из этих монет нельзя, 3, 5, 6 копеек можно, 7 нельзя. Далее следуют

$$8 = 3 + 5, \quad 9 = 3 + 3 + 3, \quad 10 = 5 + 5.$$

А из этого видно, что можно получить любое большее целое число (достаточно добавлять трехкопеечные монеты к 8, 9, 10).

Интересно нарисовать множество всех допустимых сумм: оно образует аддитивную полугруппу, т. е. с любыми двумя своими элементами x и y содержит и их сумму, $x + y$. Ноль мы тоже включили в число сумм.

Полугруппа с образующими 3 и 5 изображена на рис. 1 (квадратиками).

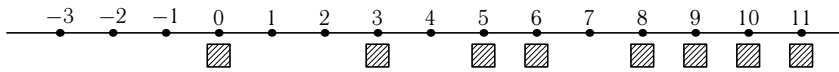


Рис. 1. Полугруппа, порожденная алтыном и пятаком

Интересно отметить, что дополнительное множество целых чисел расположено симметрично этой полугруппе (см. квадратики на рис. 2).

А именно, если x входит в полугруппу, то $7 - x$ входит в дополнение. Например, в полугруппу входят все целые числа, большие или равные 8, а

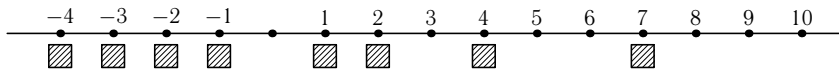


Рис. 2. Дополнение к полугруппе, порожденной алтыном и пятаком

в дополнение — все целые числа, меньшие дополнительного к 8 числа -1 (или равные ему).

§ 1. Теорема Сильвестра и числа Фробениуса

Первый американский математик Дж. Сильвестр доказал, что положение будет аналогичным для полугруппы, порожденной любыми натуральными числами a и b, у которых наибольший общий делитель равен 1. Эта полугруппа состоит из всевозможных комбинаций $xa + yb$ (с целыми неотрицательными коэффициентами x и y).

Теорема 1 (Сильвестр). *Порожденная взаимно простыми числами a и b полугруппа содержит все целые числа, начиная с $N(a, b) = (a-1)(b-1)$.*

Мы докажем теорему 1 ниже, в § 6.

Симметрия (с центром $(N-1)/2$) тоже всегда имеет место: *x принадлежит полугруппе, если и только если $y = N - 1 - x$ не принадлежит ей.*

Задача Фробениуса состоит в том, чтобы понять, как обстоит дело в случае больше двух образующих. Пусть например выбраны n образующих (натуральных чисел)

$$a_1, a_2, \ldots, a_n.$$

Если наибольший делитель всех этих чисел равен 1, то *аддитивная полугруппа их целочисленных комбинаций*

$$\{a = x_1a_1 + \ldots + x_na_n\}, \quad x_s \geqslant 0,$$

содержит все целые числа, больше или равные некоторого числа Фробениуса $N(a_1, \ldots, a_n)$.

Это легко вывести из теоремы Сильвестра, добавляя образующие.

Задача Фробениуса состоит в том, чтобы вычислить число Фробениуса $N(a_1, \ldots, a_n)$ или хотя бы исследовать его поведение при изменении образующих a_s (например, при $a_s \to \infty$).

Формула Сильвестра указывает при $n = 2$ на рост порядка произведения образующих, $N(a, b) \sim ab$.

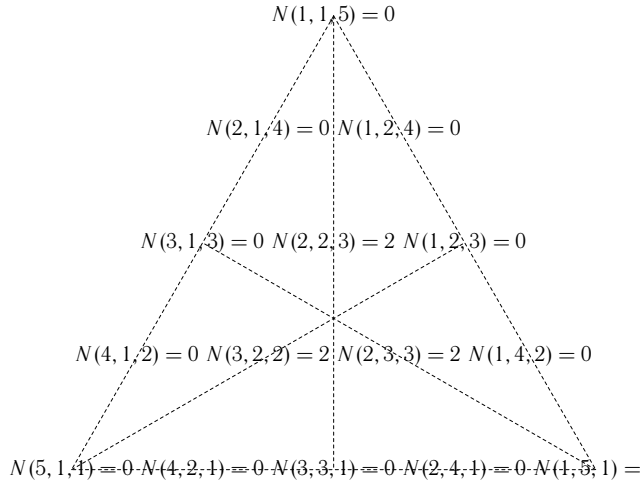
При n образующих, как мы увидим ниже, роль произведения ab начнет играть величина

$$N_0(a_1, \ldots, a_n) = \sqrt[n-1]{(n-1)!a_1 \ldots a_n}.$$

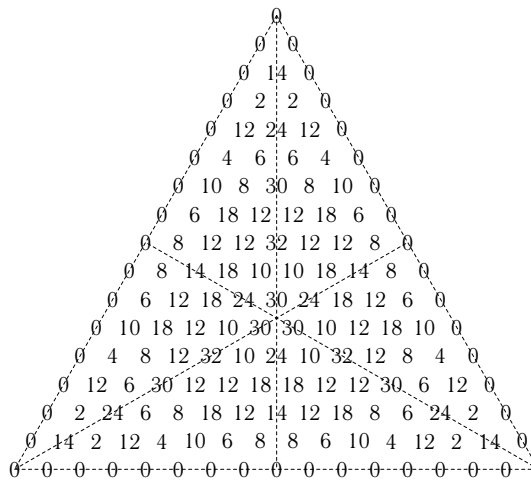
Например, для $n = 3$ формула такая:

$$N_0(a,\, b,\, c) = \sqrt{2abc}.$$

Пример 1. Значения $N(a,\, b,\, c)$ для $a + b + c = 7$ образуют следующий равносторонний треугольник (в котором я отметил его 3 оси симметрии).



Пример 2. Для $a + b + c = 19$ числа Фробениуса $N(a,\, b,\, c)$ образуют следующий равносторонний треугольник.



Как мы увидим ниже (следуя работе Арнольд В. И. Слабые асимптотики чисел решений диофантовых задач // Функ. анализ и его прил. 1999. Т. 33, № 4. С. 65—66), при росте суммы $\sigma = a_1 + \ldots + a_n$ эти заполнения (правильных n-мерных симплексов) числами Фробениуса $N(a_1, \ldots, a_n)$ порождают своеобразные асимптотические закономерности, некоторые из которых мы далее докажем, другие же останутся естественно-научными закономерностями, ожидающими строгой математической теории (возможно, от слушателей этих лекций).

§2. Загораживающие деревья леса

Предположим, что велосипедист едет по прямому шоссе (рис. 3) и подъезжает к углу леса. Глядя на этот угол, он сначала видит крайнее дерево, но потом деревьев становится все больше, и, начиная с некоторого места N, лес становится сплошным.

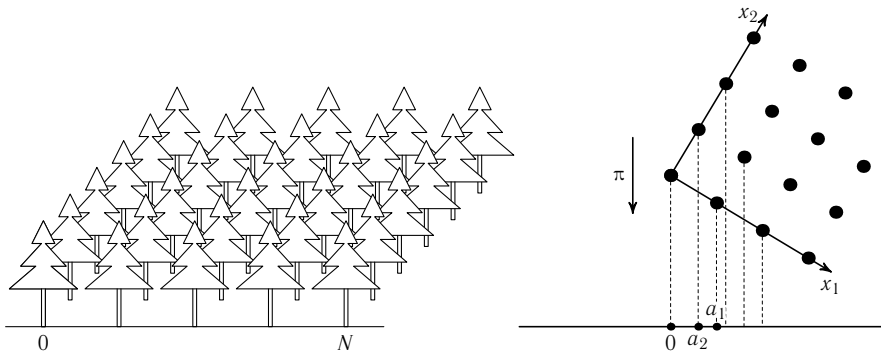


Рис. 3. Проекция π квадратно-гнездового леса на шоссе

Если лес посажен квадратно-гнездовым способом, то те места шоссе, где велосипедист, смотря точно перпендикулярно ему, увидит дерево, образуют аддитивную полугруппу (вдоль шоссе, где начало координат выбрано в ближайшей к углу леса точке).

Точно так же и общую задачу об аддитивных полугруппах можно рассматривать как задачу о линейной проекции

$$\pi\colon \mathbb{Z}_+^n \to \mathbb{R}$$

n-мерного квадранта $\{x_1, \ldots, x_n\}$, $x_s \geqslant 0$, $x_s \in \mathbb{Z}$. Проекция π сопоставляет «дереву» $x \in \mathbb{Z}_+^n$ элемент полугруппы

$$a = x_1 a_1 + \ldots + x_n a_n.$$

Не заботясь вначале о математической строгости, постараемся оценить то место N, начиная с которого лес кажется сплошным.

В нашей модели мы будем предлагать образующие полугруппы $(a_1, \ldots \ldots, a_n)$ натуральными числами, так что вся полугруппа $\pi(\mathbb{Z}_+^n)$ состоит из целых чисел.

Постараемся понять, сколько из них меньше фиксированного числа l или равно ему (рис. 4). Обозначим это число через $M(l)$.

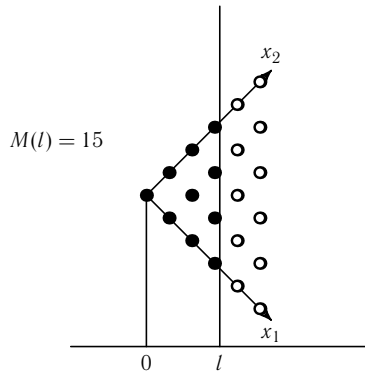


Рис. 4. Симплекс первых деревьев

Будем считать, что площадь (n-мерный объем при $n > 2$) одного квадрата (n-мерного куба) квадратно-гнездового леса равна 1. Тогда число $M(l)$ будет примерно равно площади (n-мерному объему) треугольника (симплекса размерности n) $S(l)$ в $\mathbb{R}^n$, заданного неравенствами

$$S(l) = \left\{ x \in \mathbb{R}^n : x_s \geqslant 0, \sum_{s=1}^{n} x_s \leqslant l \right\}.$$

Катеты этого прямоугольного треугольника (симплекса) имеют длины l/a_s.

Поэтому эта площадь есть $V(l) = \dfrac{1}{2} \dfrac{l}{a_1} \dfrac{l}{a_2}$. В n-мерном случае n-мерный объем симплекса $S(l)$ есть

$$V(l) = \frac{1}{n!} \prod_{s=1}^{n} \frac{l}{a_s}.$$

Итак мы приходим к гипотетической приближенной формуле для числа деревьев в симплексе $S(l)$:

$$M(l) \approx \frac{l^n}{n!\,\Pi}, \quad \text{где} \quad \Pi = \prod_{s=1}^{n} a_s.$$

Чтобы проекция π покрывала точку l, нужно, чтобы выполнялось неравенство $M(l) - M(l-1) \geqslant 1$.

Считая l большим и заменяя разности производными, мы получаем условие $dM/dl \geqslant 1$, т. е.

$$\frac{l^{n-1}}{(n-1)!\,\Pi} \geqslant 1,$$

то есть

$$l \geqslant N_0 = \sqrt[n-1]{(n-1)!\,\Pi}. \tag{1}$$

При $n = 2$ получается приближенное условие $l \geqslant a_1 a_2$, (близкое асимптотически к точному условию Сильвестра $l \geqslant (a_1 - 1)(a_2 - 1)$).

Вопрос о строгом математическом обосновании формулы (1) очень не прост, и мы обсудим его ниже, в §3 и §4.

§3. Геометрия чисел

Попытку обоснования формулы (1) мы начнем с очень простых соображений (Минковского), связывающих число целых точек $M(l)$ с объемом $V(l)$: в каком именно смысле $M(l) \approx V(l)$? Обозначаем через σ сумму $a_1 + \ldots + a_n$.

Теорема 1. *Имеют место неравенства*

$$V(l) \leqslant M(l) \leqslant V(l + \sigma).$$

Доказательство. Каждой целой точке x замкнутой области $S(l)$ сопоставим единичный n-мерный куб (рис. 5)

$$\{y \in \mathbb{R}^n \colon x_s \leqslant y_s \leqslant x_x + 1,\ s = 1, \ldots, n\}$$

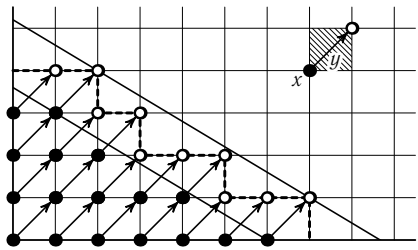
Объединение таких кубов, построенных для всех целых точек x замкнутого симплекса $S(l)$, образует многогранник P. Этот многогранник содержится в замкнутом симплексе $S(p + \sigma)$.

Действительно, для точки y куба, построенного по точке x, мы находим

$$(a,\, y) \leqslant (a,\, x) + (a,\, 1) = l + \sigma.$$

Многогранник P содержит замкнутый симплекс $S(l)$.

Действительно, заменим каждую координату x_s точки x из $S(l)$ ее целой частью z_s. Полученная точка z также лежит в замкнутом симплексе $S(l)$,



$$a_1 = 3,\ a_2 = 5,\ l = 18.$$
$$M(l) = 17$$
$$V(l) = 10{,}8 \quad V(l+1) = 12\frac{1}{30}.$$
$$V(l+\sigma) = 22\frac{8}{15}.$$
$$V\left(l + \frac{\sigma}{2}\right) = 16\frac{2}{15} < M(l),$$
$$V\left(l + \frac{\sigma+2}{2}\right) = 17\frac{19}{30} > M(l)$$

Рис. 5. Описанный многогранник P симплекса S

поэтому исходная точка x принадлежит кубу, построенному по целой точке z, а значит x принадлежит многограннику P.

Из доказанных включений многогранников

$$S(l) \subseteq P \subseteq S(l+\sigma)$$

вытекает неравенство теоремы 1 между объемами этих многогранников. □

Замечание. Примеры показывают, что имеют место гораздо более интересные неравенства для числа $M(l)$ целых точек в замкнутом симплексе $S(l)$:

$$V\left(l + \frac{\sigma-2}{2}\right) \leqslant M(l) \leqslant V\left(l + \frac{\sigma+2}{2}\right).$$

Среднее арифметическое левой и правой оценивающих величин дает особенно хорошее приближение к числу целых точек $M(l)$.

Пример. Для тройки $\{a_s\} = \{3, 5, 8\}$ легко вычислить значения. В этом случае $\sigma/2 = 8$.

l	0	1	2	3	4	5	6	7	8	9
$M(l)$	1	1	1	2	2	3	4	4	6	7
$V(l)$	0	$\frac{1}{720}$	$\frac{1}{90}$	$\frac{3}{80}$	$\frac{4}{45}$	$\frac{25}{144}$	$\frac{3}{10}$	$\frac{343}{720}$	$\frac{32}{45}$	$\frac{81}{80}$

l	10	11	12	13	14	15	16	17	18	19	20	21
$V(l)$	8	10	11	13	15	17	20	22	25	28	31	35
$M(l)$	$\frac{25}{18}$	1,84	2,4	3,05	3,81	4,69	5,69	7,60	8,10	9,52	11,11	12,86

Случаи четной и нечетной сумм σ немного различаются, поэтому приведу еще таблицу для тройки $\{a_s\} = \{3, 5, 7\}$, $\sigma = 15$, $(\sigma+1)/2 = 8$.

l	0	1	2	3	4	5	6	7	8	9
$M(l)$	1	1	1	2	2	3	4	5	6	7
$V(l)$	0	$\frac{1}{630}$	$\frac{8}{630}$	$\frac{3}{70}$	0,102	0,198	0,343	0,544	0,812	1,157

l	10	11	12	13	14	15	16	17	18	19	20
$M(l)$	9	10	12	14	16	19	21	24	27	30	34
$V(l)$	1,59	2,11	2,74	3,49	4,35	5,36	6,50	7,79	9,25	10,88	12,69

Здесь неравенства

$$V\left(l+\frac{\sigma-1}{2}\right)=V(l+7)\leqslant M(l)\leqslant V(l+8)=V\left(l+\frac{\sigma+1}{2}\right)$$

выполняются при $l\geqslant 3$, но при $l=2$ положение иное:

$$\left(V\left(7\,\frac{1}{2}\right)\approx 0{,}984\right)<(M(2)=1)<V(8).$$

Я не знаю общего доказательств отмеченных здесь неравенств, даже для случая $n=2$ плоских многоугольников, (где вселяют надежду соображения симметрии по отношению к точке прямой $(a,\,x)=l+\frac{\sigma}{2}$).

Теперь мы используем доказанную выше теорему 2 для оценки числа Фробениуса снизу.

Рассмотрим интервал $N\leqslant l<N+r$, где N — число Фробениуса. Для каждой точки l этого интервала существует целая точка x $(x_s\geqslant 0)$, в которой $(a,\,x)=l$. Поэтому в симплексе $S(N+r-1)$ лежит не меньше, чем r целых точек:

$$M(N+r-1)\geqslant r.$$

Подставляя оценку числа целых точек M сверху из теоремы 2, мы получаем оценку снизу для объема соответствующего симплекса, а значит оцениваем снизу и число Фробениуса N.

Из теоремы 2 мы заключаем, что

$$r\leqslant M(N+r-1)\leqslant V(N+\sigma+r-1)=\frac{(N+\sigma+r-1)^n}{n!\,\Pi}.$$

Для значения $r=\lambda(N+\sigma-1)$ мы находим

$$\lambda(N+\sigma-1)\leqslant\frac{(N+\sigma-1)^n(1+\lambda)^n}{n!\,\Pi},$$

откуда

$$N+\sigma-1\geqslant\left(\frac{\lambda n!\,\Pi}{(1+\lambda)^n}\right)^{1/(n-1)}.$$

Итак,

$$N \geqslant \omega \Pi^{\frac{1}{n-1}} - \sigma + 1,$$

где коэффициент ω имеет следующий вид:

$$\omega(\lambda) = \left(\frac{\lambda n!}{(1+\lambda)^n}\right)^{\frac{1}{n-1}}.$$

При $\lambda = \dfrac{1}{n-1}$,

$$\left(\frac{\lambda}{(1+\lambda)^n}\right)^{\frac{1}{n-1}} = \frac{n-1}{n^{1+\frac{1}{n-1}}} \geqslant \frac{1}{4},$$

так что наша оценка числа Фробениуса имеет вид

$$N \geqslant \frac{1}{4}(n!\,\Pi)^{\frac{1}{n-1}} - \sigma + 1.$$

Это неравенство указывает на именно такой рост порядка $\Pi^{\frac{1}{n-1}}$, то есть $\sigma^{n/(n-1)}$, как дали наши эвристические рассуждения в § 4.2, хотя коэффициент в теперь доказанной оценке при $n = 2$ меньше, чем в формуле Сильвестра, $N_0 = \Pi - \sigma + 1$.

§4. Оценка числа Фробениуса сверху

Наша оценка будет использовать простые общие факты геометрии чисел.

Рассмотрим в Евклидовом пространстве $\mathbb{R}^n$ базис $(P_1, P_2, \ldots, P_n)$, и породим этими векторами решетку $\mathbb{Z}^n \subset \mathbb{R}^n$. Мы оценим теперь сверху наибольший возможный радиус R шара в этом пространстве, не содержащего ни одной точки решетки:

Теорема 1. *Имеет место неравенство*

$$R \leqslant \frac{\sqrt{R_1^2 + R_2^2 + \ldots + R_n^2}}{2},$$

где R_s означает расстояние от точки P_s до пространства, порожденного векторами $(P_1, \ldots, P_{s-1})$.

Доказательство. При $n = 1$ это очевидно: $R \leqslant |P_1|/2$. Предположим, что при $n = k$ это уже доказано.

Рассмотрим в пространстве $\mathbb{R}^{k+1}$, порожденном векторами $(P_1, \ldots, P_{k+1})$, гиперплоскость Q, порожденную векторами $(P_1, \ldots, P_k)$, и параллельные ей гиперплоскости Q_j, проходящие через точки jP_{k+1} соответственно (рис. 6).

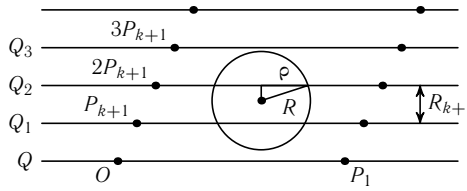


Рис. 6. Пустой шар в пространстве с решеткой размерности $k+1$

Центр пустого шара в пространстве $\mathbb{R}^{k+1}$ попадет в один из слоев толщины R_{k+1} между гиперплоскостями Q_j, а потому удален от одной из них не более, чем на расстояние $R_{k+1}/2$.

Пересечение нашего пустого шара с этой гиперплоскостью Q_j пусто в этой гиперплоскости, а потому его радиус ρ уже оценен предполагающейся известной при $n = k$ теоремой 3:

$$\rho \leqslant \frac{\sqrt{R_1^2 + \ldots + R_k^2}}{2}.$$

По теореме Пифагора мы получаем

$$R^2 \leqslant \rho^2 + (R_{k+1}/2)^2 = (R_1^2 + \ldots + R_k^2 + R_{k+1}^2)/4,$$

чем и доказана теорема 1. □

Теперь мы применим эту теорему к решетке целых точек в плоскости

$$\{x \in \mathbb{R}^n \colon (a,\, x) = l\},$$

в которой нет целых точек в ортанте $x_s \geqslant 0$. Симплекс размерности $n-1$, по которому гиперплоскость пересекает ортант, мы обозначим через $S(l)$. Разделим зависимость функций от вектора $a = (a_1, \ldots, a_s)$ на зависимость от его направления $\alpha(a) = a/\sigma(a)$ и от его «размера» $\sigma(a)$.

Теорема 2. *Радиус шара, вписанного в симплекс $S(l)$, равен $\beta(\alpha)l/\sigma(a)$, где безразмерный коэффициент β, зависящий лишь от направления вектора a, есть*

$$\beta(\alpha) = \frac{|\alpha|}{\sum\limits_{s=1}^{n} (\alpha_s \sqrt{|\alpha|^2 - \alpha_s^2})}.$$

Пример. При $n = 2$ эта формула имеет вид

$$\beta(\alpha) = \frac{|\alpha|}{2\Pi} = \frac{\sqrt{\alpha_1^2 + \alpha_2^2}}{2\alpha_1\alpha_2} \geqslant \sqrt{2}.$$

При любом n $\beta(\alpha) \geqslant \sqrt{n/(n-1)} \geqslant 1$.

Доказательство теоремы 2. Пирамида с основанием $S(l)$ имеет объем $V(l) = \frac{1}{n!}\prod_{s=1}^{n}\left(\frac{l}{a_s}\right) = \frac{l^n}{n!\,\Pi}$. В то же время этот объем выражается через объем $|S(l)|$ основания и длину h опущенной на него из O высоты:

$$V(l) = \frac{1}{n}h|S(l)|, \quad h = l/|a|.$$

Поэтому

$$|S(l)| = \frac{nV(l)}{h} = \frac{l^{n-1}|a|}{(n-1)!\,\Pi} = \frac{l^{n-1}\sigma(a)|\alpha|}{(n-1)!\,\Pi}. \tag{1}$$

С другой стороны, объем симплекса $S(l)$ равен $\frac{1}{n-1}$ от произведения радиуса r вписанного в него шара на сумму площадей $|S_s|$ его граней $\{x_s = 0\}$, поэтому

$$r = \frac{n-1}{\sum_{s=1}^{n} S_s}|S(l)|. \tag{2}$$

Вычисляя площади граней по формуле (1), мы получаем

$$S_s = \frac{l^{n-3}}{(n-2)!}\frac{a_s\sqrt{|a^2| - a_s^2}}{\Pi} = \frac{\sigma^2(a)l^{n-2}\alpha_s\sqrt{|\alpha^2| - \alpha_s^2}}{(n-2)!\,\Pi}. \tag{3}$$

Подставляя в формулу (2) выражения (1) и (3), мы находим

$$r = \frac{(n-1)l^{n-1}\sigma(a)|\alpha|}{(n-1)!\,\Pi}\frac{(n-2)!\,\Pi}{\sigma^2(a)l^{n-2}\sum_{s=1}^{n}(\alpha_s\sqrt{|\alpha|^2 - \alpha_s^2})} = \frac{l}{\sigma(a)}\frac{|\alpha|}{\sum_{s=1}^{n}(\alpha_s\sqrt{|\alpha|^2 - \alpha_s^2})},$$

что и доказывает теорему 2. □

Заметим теперь, что, если в симплексе $S(l)$ нет целых точек, то их нет и во вписанном в него шаре. Поэтому радиус этого вписанного шара не может превосходить границы, доставляемой теоремой 1 (примененной к $n-1$-мерной гиперплоскости $\{x \in \mathbb{R}^n\colon (a, x) = l\}$). Эта теорема доставляет неравенство

$$\frac{\beta(\alpha)l}{\sigma(a)} \leqslant \frac{\sqrt{R_1^2 + \ldots + R_{n-1}^2}}{2},$$

$$l \leqslant \frac{\sigma(a)}{2\beta(a)}\sqrt{R_1^2 + \ldots + R_{n-1}^2}. \tag{4}$$

Теорема 3. *Для вектора $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ любого направления $\alpha = a/\sigma(a)$ имеет место оценка числа Фробениуса сверху*

$$N(a) \leqslant 1 + \gamma(\alpha)\sigma^2(a)$$

(*с указанной в доказательстве постоянной* γ).

Доказательство теоремы 3. Рассмотрим флаг из подпространств $\mathbb{R}^1 \subset \mathbb{R}^2 \subset \ldots \subset \mathbb{R}^n$, где $\mathbb{R}^s$ натянуто на первые s координатных осей (рис. 7).

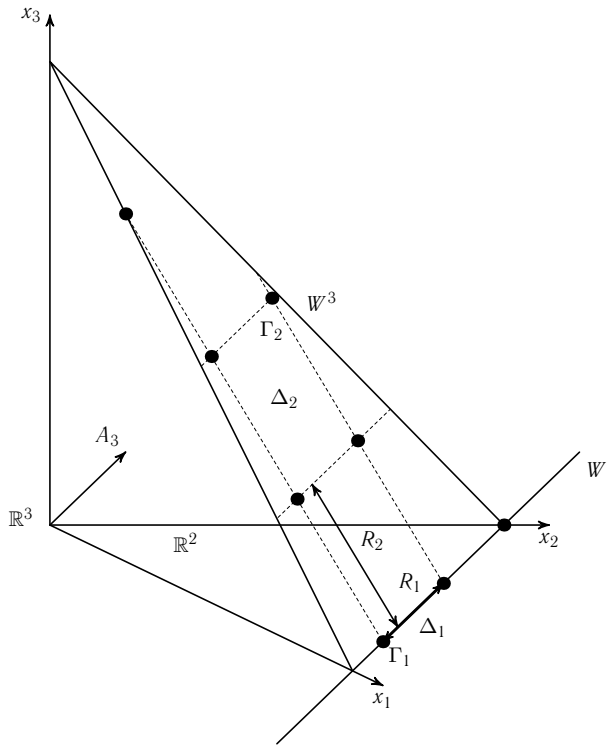


Рис. 7. Последовательные высоты R_s и параллелепипеды Δ_s при $n = 3$. Вместо плоскости $(A_3, x) = 0$ нарисована плоскость $(A_3, x) = \text{const}$, где решетка Γ_2 такая же.

Зададим в $\mathbb{R}^n$ гиперплоскость W^{n-1} уравнением $a_1x_1 + \ldots + a_nx_n = 0$. Мы введем в каждом пространстве $\mathbb{R}^s$ вектор A_s компонентами $a_1, \ldots, a_s$, и определим в $\mathbb{R}^s$ ортогональную этому вектору гиперплоскость W^{s-1} уравнением

$$(A_s, X_s) = 0 \quad (\text{где } X_s = (x_1, \ldots, x_s) \in \mathbb{R}^s).$$

Мы определили, таким образом, флаг векторных евклидовых подпространств в гиперплоскости W^{n-1}:

$$0 \subset W^1 \subset \ldots \subset W^{n-1}$$

(например, W^1 — это прямая на плоскости $\mathbb{R}^2$, заданная уравнением

$$(A_2,\, X_2) = 0, \quad \text{то есть уравнением } a_1x_1 + a_2x_2 = 0.$$

Пересечение гиперплоскости W^s с целочисленной решеткой $\mathbb{Z}^{s+1}$ пространства $\mathbb{R}^{s+1}$ определяет в этой s-мерной гиперплоскости s-мерную решетку Γ_s.

Лемма. *Объем фундаментального s-мерного параллелепипеда Δ_s решетки Γ_s в евклидовом пространстве W^s равен $|\Delta_s| = |A_{s+1}|/d_s$, где d_s — наибольший общий делитель чисел $(a_1, \ldots, a_{s+1})$ и $|A_s|^2 = a_1^2 + \ldots + a_s^2$, так что $|A_s|$ — евклидова длина вектора A_s.*

Доказательство леммы. Фундаментальный параллелепипед решетки $\mathbb{Z}^{s+1}$ в пространстве $\mathbb{R}^{s+1}$ можно получить из параллелепипеда Δ_s решетки Γ_s в W^s, лежащего в гиперплоскости W^s добавлением ближайшего к этой гиперплоскости целочисленного вектора $x \in \mathbb{Z}^{s+1}$ (рис. 8).

Скалярное произведение этого вектора с нормальным к гиперплоскости W^s вектором A_{s+1} имеет наименьшее положительное значение, возможное для целочисленных линейных комбинаций

$$a_1x_1 + \ldots + a_{s+1}x_{s+1} = (A_{s+1},\, x).$$

Значение этой комбинации делится на общее кратное d_s ее коэффициентов, и ее наименьшее значение есть d_s. Итак, для ближайшего к гиперплоскости W^s не лежащего в ней целочисленного вектора x выполняется соотношение $(A_{s+1},\, x) = d_s$ (рис. 8).

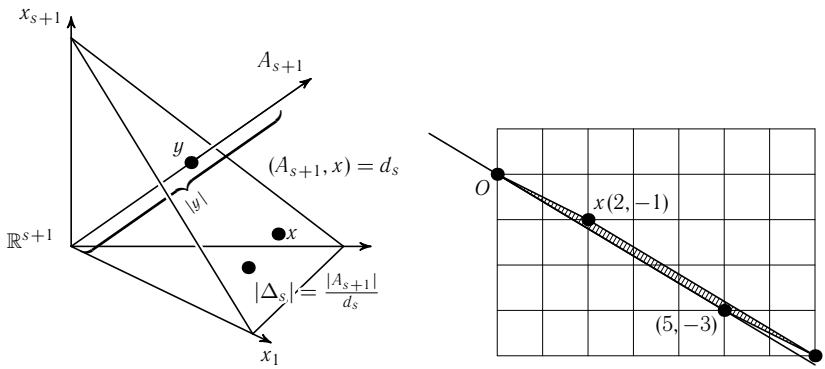


Рис. 8. Вычисление объема $|\Delta_s|$ фундаментального параллелепипеда Δ_s решетки Γ_s в W^s

Найдем теперь расстояние от этого ближайшего вектора до гиперплоскости W^s (рис. 8).

Для этого проведем нормаль к гиперплоскости, $\{y = |y|A_{s+1}/|A_{s+1}|\}$. Точка y на этой нормали лежит от гиперплоскости W^s на таком же расстоянии $\rho = |y|$, как и ближайшая целая точка x, если $(A_{s+1}, y) = (A_{s+1}, x) = d_s$.

Умножая y скалярно на вектор A_{s+1}, мы находим

$$(A_{s+1}, y) = |y|(A_{s+1}, A_{s+1})/|A_{s+1}| = |y||A_{s+1}| = d_s,$$

откуда $\rho = |y| = d_s/|A_{s+1}|$.

Итак, объем фундаментального параллелепипеда решетки $\mathbb{Z}^{s+1}$ в евклидовом пространстве $\mathbb{R}^{s+1}$ есть $|\Delta_s||y| = \rho|\Delta_s|$. Но этот объем равен 1, так как $\mathbb{Z}^{s+1}$ — стандартная целочисленная решетка.

Значит $|\Delta_s| = 1/\rho = |A_{s+1}|/d_s$, и лемма доказана. □

Следствие. *Длина R_s высоты фундаментального параллелепипеда Δ_s решетки Γ_s в евклидовом пространстве W^s, основанием которого является фундаментальный параллелепипед Δ_{s-1} подрешетки Γ_{s-1}, равна*

$$R_s = |\Delta_s|/|\Delta_{s-1}|.$$

Доказательство следствия. s-мерный объем $|\Delta_s|$ параллелепипеда Δ_s равен произведению $(s-1)$-мерного объема $|\Delta_{s-1}|$ параллелепипеда-основания на длину R_s высоты (рис. 7). □

Нам будет удобнее записывать эту формулу для длины R_s очередной высоты в виде

$$R_s^2 = \frac{(a_1^2 + \ldots + a_{s+1}^2)d_{s-1}^2}{(a_1^2 + \ldots + a_s^2)d_s^2}. \tag{1}$$

Заметим, что число d_{s-1} делится на d_s нацело (так как наибольший общий делитель набора чисел является делителем любого его поднабора).

Таким образом, последовательность целых чисел $q_s = d_{s-1}/d_s$ имеет произведение

$$q_1 q_2 \ldots q_{n-1} = d_0 = a_1,$$

в то время как произведение длин высот имеет вид

$$R_1^2 \ldots R_{n-1}^2 = \frac{|\Delta_1|^2}{|\Delta_0|^2}\frac{|\Delta_2|^2}{|\Delta_1|^2} \ldots \frac{|\Delta_{n-1}|^2}{|\Delta_{n-2}|^2} = \frac{|\Delta_{n-1}|^2}{|\Delta_0|^2}.$$

Однако нам известны граничные условия $|\Delta_0| = |A_1|/d_1 = a_1/a_1 = 1$, $|\Delta_{n-1}|^2 = |A_n|^2/d_n^2 = |A_n|^2/1$. Из них мы заключаем, что произведение длин всех $n-1$ высот дается удивительной формулой

$$R_1^2 \ldots R_{n-1}^2 = a_1^2 + \ldots + a_n^2. \tag{2}$$

Это равенство и даст нам оценки длин высот сверху. Для этого сначала оценим их снизу.

Из формулы (1) для длин высот следует, что $R_s^2 > 1$ (поскольку $|A_{s+1}|^2 > |A_s|^2$ и $q_s \geqslant 1$).

Поэтому из формулы (2) для произведения длин высот вытекают оценки длин высот сверху:

$$R_s^2 = \frac{|A_n|^2}{R_1^2 \dots R_{s-1}^2 R_{s+1}^2 \dots R_{n-1}^2} < |A_n|^2.$$

Этим доказано неравенство

$$\sum_{s=1}^{n-1} R_s^2 < (n-1)(a_1^2 + \dots + a_n^2).$$

Эта оценка квадратов длин высот сверху порождает, согласно теореме 1 (§4), оценку сверху радиуса r пустого шара в гиперплоскости W^{n-1} (не имеющего общих точек с решеткой $\Gamma_{n-1} = \mathbb{Z}^n \cap W^{n-1}$):

$$r^2 \leqslant \frac{n-1}{4}|A_n|^2. \tag{3}$$

По теореме 2 в гиперплоскости W^{n-1} существует пустой шар радиуса $\frac{l}{\sigma(a)}\beta(\alpha)$, (с зависящей лишь от направления α вектора a постоянной β), если $l = N(a) - 1$ (так как в симплексе $A_n(x) = l$, $x_j \geqslant 0$, нет целых точек, то их нет и в вписанном в него шаре).

Из неравенства (3) мы находим для l оценку сверху

$$\frac{l^2}{\sigma^2(a)}\beta^2(\alpha) \leqslant \frac{n-1}{4}|A_n|^2,$$

откуда

$$N(a) = l + 1 \leqslant 1 + \frac{\sqrt{n-1}}{2\beta}|A_n|\sigma(a).$$

Но $|A_n|^2 = a_1^2 + \dots + a_n = \sigma^2(a)(\alpha_1^2 + \dots + \alpha_n^2)$, так что мы получаем оценку числа Фробениуса сверху,

$$N(a) \leqslant 1 + \frac{\sqrt{n-1}}{2\beta(\alpha)}\sigma^2(a)\sqrt{\alpha_1^2 + \dots + \alpha_n^2},$$

что и доказывает теорему 3 страницы 95 (с постоянной $\gamma(\alpha) = \frac{\sqrt{n-1}|\alpha|}{2\beta(\alpha)}$). □

Впрочем, иногда полезнее доставляющая меньшую оценку числа Фробениуса сверху оценка

$$N \leqslant 1 + \frac{\sigma(a)}{\beta(\alpha)}\sqrt{\left(\sum_{s=1}^{n-1} R_s^2\right)} \Big/ 2 \tag{4}$$

с выражениями (1) для R_s^2. Например, когда все $d_s = 1$, все длины высот R_s зависят только от направления α вектора a, исключая лишь $R_1^2 = a_1^2 + a_2^2$.

Замечание. Оценка числа Фробениуса N снизу (§ 3) имела вид $N \geqslant \mathrm{const}(\alpha)\Pi^{\frac{1}{n}-1}(\geqslant \mathrm{const}(\alpha)\sigma^{1+\frac{1}{n-1}})$. Эта величина растет с σ медленнее, чем σ^2, при $n > 2$. Например, при $n = 3$ оценка снизу получается $N \geqslant \mathrm{const}\,\sigma^{3/2}$, а сверху $N \leqslant \mathrm{const}\,\sigma^2$, что гораздо больше.

Следующий пример показывает неизбежность этого явления: мы приведем примеры, где, действительно, $N(a, b, c) \geqslant \mathrm{const}\ \sigma^2$ (что при больших σ во сколько угодно раз больше, чем $\sigma^{3/2}$), так что оценка сверху величиной роста $\sigma^{3/2}$ при $n = 3$ невозможна.

Пример. Рассмотрим три (взаимно простые) числа

$$a,\ b,\ c = a + b.$$

В этом случае число Фробениуса легко вычислить:

$$N(a, b, c) = N(a, b),$$

(поскольку суммы копий чисел (a, b, c) являются суммами копий чисел a и b).

Итак, $N(a, b, c) = (a - 1)(b - 1)$ (по формуле Сильвестра). Нам нужен будет лишь квадратичный рост по σ, доказанный выше и без формулы Сильвестра (впрочем, доказанной ниже, в § 6).

Предположим теперь, что $1/9 < a/b < 9$, $a > 2$, $b > 2$ (рис. 9). Тогда $a - 1 > a/2$, $b - 1 > b/2$, поэтому выполняется неравенство

$$N(a, b, c) = (a - 1)(b - 1) > ab/4.$$

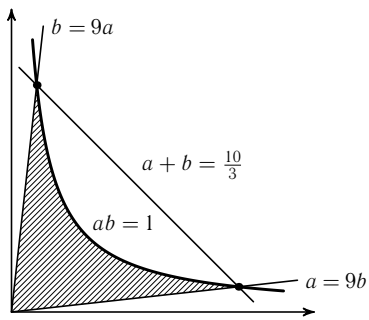


Рис. 9. Сектор, в котором $(a + b)^2 < \left(\frac{10}{3}\right)^2 ab$

Но $ab > (9/100)(a + b)^2$ в секторе $1/9 < a/b < 9$. Действительно, функция $a + b$ достигает на отрезке гиперболы $ab = 1$ в нашем секторе макси-

мума в концевых точках отрезка, где $a=3$, $b=\frac{1}{3}$ или $a=\frac{1}{3}$, $b=3$, так что всюду в этом секторе

$$a+b\leqslant 3+\frac{1}{3}=\frac{10}{3}.$$

Значит по однородности и при любом значении ab в этом секторе выполняется неравенство $(a+b)^2<(10/3)^2ab$. Итак, всюду в указанном секторе выполнено неравенство

$$\frac{ab}{4}>\frac{9(a+b)^2}{400},$$

т. е.

$$N(a,\,b,\,c)>\frac{9}{400}\sigma^2.$$

В частности, отношение

$$\frac{N(a,\,b,\,c)}{\sigma^{3/2}}>\frac{9}{400}\sqrt{\sigma}$$

принимает в точках указанной области (со взаимно простыми a и b) сколь угодно большие значения, делая оценку сверху вида

$$N(a,\,b,\,c)<\mathrm{const}\,(\alpha)\sigma^{3/2}$$

невозможной (для многих троек в целом секторе направлений, α).

Таким образом, примеры квадратичного по $\sigma(a)$ роста чисел Фробениуса имеются. Тем не менее, я не сумел выяснить *насколько многочисленны* тройки, обладающие описанным выше свойством квадратичного роста числа Фробениуса N с σ: типичны они или исключительны?

Дело в том, что рассуждения приведенного выше доказательства неравенства $N\leqslant\mathrm{const}\,\sigma^2$ доказывают в действительности больше (см. выше неравенство (4). Величина σ^2 вместо $\sigma^{1+\frac{1}{n-1}}$ получилась из-за того, что мы оценили рост всех длин высот R_s величиной $|A_n|^2$, в то время как мы знаем, что с этой скоростью растет только их произведение

$$R_1^2\ldots R_{n-1}^2=|A_n|^2,$$

так что сами длины высот растут медленнее.

В случае, когда все *отношения* длин высот R_s/R_t ограничены сверху некоторой (не зависящей от $|A_n|$) постоянной, из приведенного выражения (2) для произведения получались бы оценки длин высот гораздо меньшей величиной, $R_s^2\leqslant\mathrm{const}\,|A_n|^{\frac{2}{n-1}}$, чем использованное у нас $|A_n|^2$.

Рассмотрим, например, случай, когда коэффициенты $(a_1,\,a_2,\,\ldots,\,a_n)$ можно расположить в таком порядке, чтобы все наибольшие общие делители первых из них,

$$d_1=(a_1,\,a_2),\;\;d_2=(a_1,\,a_2,\,a_3),\;\;\ldots,\;\;d_{s-2}=(a_1,\,a_2,\,\ldots,\,a_{n-1}),$$

были равны 1, так что формула (1) принимает не зависящий от $\sigma(a)$ вид

$$R_s^2 = \frac{a_1^2 + \ldots + a_{s+1}^2}{a_1^2 + \ldots + a_s^2} = \frac{\alpha_1^2 + \ldots + \alpha_{s+1}^2}{\alpha_1^2 + \ldots + \alpha_s^2} = \Psi_s(\alpha).$$

В этом случае условие ограниченности отношений длин высот R_s/R_t выделяет сектор в пространстве направлений, в пределах которого действует оценка $N \leqslant 1 + \mathrm{const}\,\sigma^{1+\frac{1}{n-1}}$, так что асимптотика числа Фробениуса N порядка $\sigma^{\frac{n}{n-1}}$ (то есть $\Pi^{\frac{1}{n-1}}$) имеет место для одних направлений, а порядка σ^2 (то есть $\Pi^{2/n}$) — для других, и вопрос о том, каких направлений больше, не прост.

Даже вопрос о поведении в зависимости от σ средних значений $N(\sigma)$ чисел $N(a)$ по всем направлениям векторов a с данной суммой координат $\sigma(a)$ не прост и заслуживает экспериментального исследования.

Я сделал в этом направлении только первые шаги, но при привлечении надлежащей компьютерной техники можно быстро продвинуться вперед в этих эмпирических исследованиях.

Приведенные ниже численные данные подсказывают достижение средними числами Фробениуса $\hat{N}$ роста со скоростью меньше σ^2, но он может замедлиться при бóльших значениях суммы $\sigma = a + b + c$ (я дошел только до $\sigma = 41$, но см. описание последующих результатов на стр. 104).

§5. Средние значения чисел Фробениуса

Для сравнения доказанных оценок с реальностью я сосчитал явно все значения чисел Фробениуса $N(a, b, c)$ для $\sigma = 41$, т. е. для всех 780 троек натуральных чисел с суммой $a + b + c = 41$.

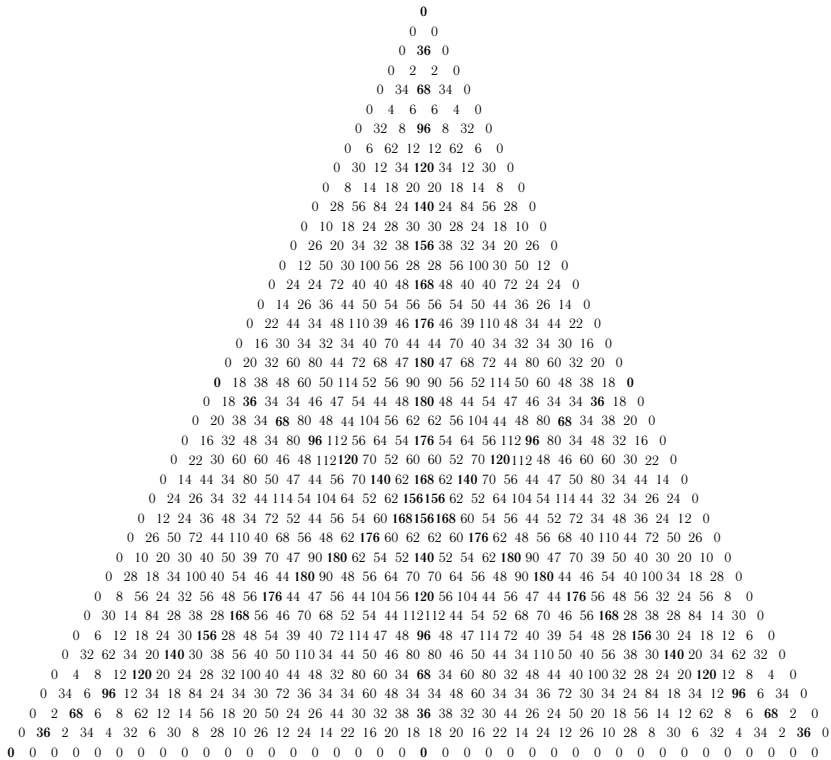
В качестве суммы σ я выбрал простое число для того, чтобы исключить случай «соизмеримости», когда все числа a_s имеют нетривиальный общий делитель, так что порожденная ими аддитивная полугруппа не содержит всех целых чисел, начиная с N, ни при каком N.

Тройки чисел $N(a, b, c)$ с данной суммой σ естественно образуют равносторонний треугольник. Для $\sigma = 7$ и 19 такие треугольники приведены выше, на стр. 87. При $\sigma = 41$ достаточно рисовать только часть всего этого треугольника (6 симметрий которого легко позволяют восстановить недостающие части), см. с. 103.

Из всех приведенных выше таблиц значений $N(a, b, c)$ видно, что отношение

$$\vartheta = \frac{N(a, b, c)}{(2abc)^{1/2}}$$

может сильно меняться даже при небольших изменениях аргументов a, b, c.



Пример (при $a+b+c=41$).

$$N(7,\,14,\,20)=114,\quad \text{гораздо больше, чем } \sqrt{2\cdot 7\cdot 14\cdot 20}\approx 62{,}4.$$
$$N(7,\,15,\,19)=47,\quad \text{гораздо меньше, чем } \sqrt{2\cdot 7\cdot 15\cdot 19}\approx 63{,}2.$$

В первом случае $\vartheta\approx 1{,}82$, а во втором $\vartheta\approx 0{,}74$ вдвое меньше, хотя тройки — соседи.

Я давно уже высказал гипотезу, что эмпирически естественная «асимптотика» $N\sim\left((n-1)!\prod_{s=1}^{n}a_s\right)^{1/n-1}$ должна считаться слабой, т. е. что приближение к ней возникает (по мере роста величины $\sigma(a)$) лишь для средних значений N по векторам a, направления которых, α, пробегают некоторую область в пространстве направлений (Слабые асимптотики чисел решения диофантовых задач // Функц. анализ и его приложения. 1999. Т **33**, вып. 4. С. 65 — 66).

Эта гипотеза остается недоказанной, но я решил поверить хотя бы поведение средних значений функций N и $I=\sqrt{abc}$ по всему симплексу $\{a+b+c=\sigma,\ a\geqslant 1,\ b\geqslant 1,\ c\geqslant 1\}$.

Эти вычисления, для приведенных выше таблиц с $\sigma=7$, 19 и 41, привели к следующим средним значениям, $\hat{N}$ и $\hat{I}$.

σ	$\sum N$	$\sum 1$	$\hat{N}=(\sum N)/(\sum 1)$	$\sum I$	$\hat{I}=(\sum I)/(\sum 1)$
7	6	15	0,4	43,04	2,87
19	1332	153	8,7	1880	12,29
41	33126	780	42,47	31068	39,83
97	909930	4560	199,546		
199	12975216	19503	665,293		

Из этих данных можно логарифмированием извлечь порядок предположительной степенной асимптотики,

$$\hat{N}\sim C\sigma^{u}.$$

Действительно, если $\ln\hat{N}=\ln C+u\ln\sigma$, то коэффициент u находится как наклон графика зависимости $\hat{N}(\sigma)$, нарисованного на двойной логарифмической бумаге:

$$u\approx\frac{\ln\hat{N}(\sigma_2)-\ln\hat{N}(\sigma_1)}{\ln\sigma_2-\ln\sigma_1}.$$

Для $\sigma_1=7$, $\sigma_2=41$ и для $\sigma_1=19$, $\sigma_2=41$ получаются

$$u\approx\frac{3{,}83+0{,}92}{3{,}70-1{,}95}\approx 2{,}5\quad\text{и}\quad u\approx\frac{3{,}83-2{,}20}{3{,}70-2{,}94}\approx 2,1.$$

выбор ($\sigma_1=41$, $\sigma_2=97$) приводит к наклону прямых $u=1{,}8$, а выбор ($\sigma_1=97$, $\sigma_2=199$) — $ku\approx 1{,}6$. Эти приближающиеся к 1,5 числа отчасти поддерживают мою гипотезу 1999 года, что при больших векторах a числа Фробениуса $N(A)$ растут в среднем как σ в степени $1+1/(n-1)$, то есть как σ в степени $3/2$ при $n=3$. Вычисления для $\sigma=97$ и 199 — компьютерные, их провел по моей просьбе А. Годер.

Аналогичные вычисления для $\hat{I}$ приводят скорее к $u\approx 1{,}5$, что дают и доказанные выше асимптотики (а также соображения подобия, примененные к интегралу от функции $I=\sqrt{abc}$ по симплексу $a+b+c=\sigma$).

§6. Доказательство теоремы Сильвестра

Речь идет о целых точках ($x\geqslant 0$, $y\geqslant 0$) на прямой $ax+by=l$ в плоскости $\mathbb{R}^2$ (где a и b — натуральные взаимно простые числа, l — целое число).

Теорема 1. *Если $l=N-1=ab-a-b$, то на указанном отрезке прямой нет целых точек, а если $l\geqslant N=(a-1)(b-1)$, то есть.*

Доказательство. Рассмотрим сначала все целые точки на прямой $ax+by=0$. Поскольку x и y взаимно просты, число x делится на b, а число y на a. Поэтому ближайшая к 0 ненулевая целая точка P нашей прямой имеет координаты $x=b$, $y=-a$ (рис. 10).

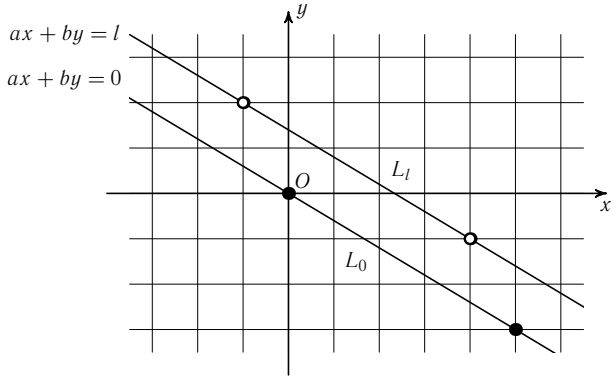


Рис. 10. Целые точки на прямых $ax+by=l$ ($a=3$, $b=5$)

Следовательно, расстояние между соседними целыми точками на нашей прямой есть $L=\sqrt{a^2+b^2}$.

На всякой параллельной ей прямой $ax+by=l$ (с целым l) целые точки образуют такую же решетку с шагом L (рис. 10), так как эти прямые переводятся друг в друга сохраняющим решетку целых точек сдвигом плоскости (поскольку уравнение $ax+by=1$ разрешимо, что видно, например, из алгоритма Евклида для нахождения наибольшего общего делителя 1 чисел a и b).

Из всего этого следует, что при $l \geqslant ab$ на отрезке прямой $ax+by=l$, где $x \geqslant 0$, $y \geqslant 0$, целые точки есть (так как длина этого отрезка есть $\sqrt{\left(\frac{l}{a}\right)^2+\left(\frac{l}{b}\right)^2}=\frac{l}{ab}\sqrt{a^2+b^2} \geqslant L$).

Тем самым второе утверждение теоремы доказано для $l \geqslant ab$.

Рассмотрим теперь прямую, на которой $ax+by=ab$ (рис. 11).

На ней лежат точки $A(x=b,\ y=0)$ и $B(x=0,\ y=a)$.

Точки $A'(x=b-1,\ y=-1)$ и $B'(x=-1,\ y=a-1)$ лежат на прямой $ax+by=l$, где $l=ab-a-b=N-1$.

Расстояния $|AB|$ и $|A'B'|$ равны L, поэтому на отрезке $A'B'$ прямой $ax+by=l'$ нет целых точек, кроме концевых точек A' и B'.

Тем самым первое утверждение теоремы (о пустоте отрезка прямой с $l=N-1$) доказано.

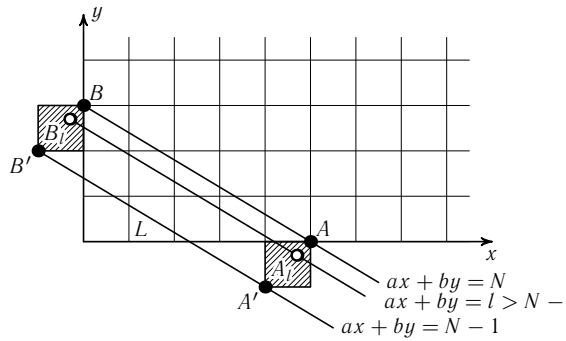


Рис. 11. Прямые Сильвестра $ax + by = N(ab) - 1$ и $ax + by = l > N(ab) - 1$

Для полного доказательства второго утверждения теоремы остается проверить его лишь при $N \leqslant l < ab$. В этих условиях прямая $ax + by = l$ пересекает диагонали (AA') и (BB') заштрихованных у точек A и B квадратов в точках A_l и B_l соответственно, причем в A_l отрицательная координата y, а в B_l отрицательная координата x.

Расстояние $|A_lB_l|$ равно L, подобно расстоянию $|AB|$. Поэтому на отрезке A_lB_l обязательно есть целая точка (x_ly_l). В этой целой точке обе координаты неотрицательны, так как пересечения отрезка A_lB_l прямой $ax + by = l$ с областями $x < 0$ и $y < 0$ лежат внутри заштрихованных квадратов (рис. 11), где целых точек нет. Наличие целой точки $(x_k,\ y_l)$, где $x_l \geqslant 0$, $y_l \geqslant 0$, $ax_l + by_l = l \geqslant N$ доказывает теорему Сильвестра до конца. □

§7. Геометрия цепных дробей чисел Фробениуса

Приведенные ниже геометрические теоремы о числах Фробениуса я придумал в 1999 году, когда писал работу о слабых асимптотиках чисел решений диофантовых задач и вычислял тысячи чисел Фробениуса $N(a, b, c)$. Но я не опубликовал этих результатов, считая их тогда слишком очевидными. Эта их очевидность ниже и доказывается. В отличие от доказательств, открытие этих геометрических фактов вовсе не просто.

Пусть (a, b, c) — положительные целые числа, не имеющие большего 1 общего делителя. Рассмотрим функцию со значениями $l(y, z) = by + cz$ на замкнутом положительном квадранте $\{(y \geqslant 0, z \geqslant 0)\}$.

Определение 1. *Реализатором* остатка k от деления на a называется такая целая точка r квадранта, где значение $l(r)$ дает при делении на a остаток k, причем значение $l(r)$ — минимальное (среди всех дающих в остатке k при делении на a значений в точках замкнутого положительного квадранта $\{(y \geqslant 0, z \geqslant 0)\}$).

Определение 2. *Областью* $D(a, b, c)$ тройки (a, b, c) называется множество всех реализаторов всех a остатков ($k = 0, 1, \ldots, a-1$). Рисовать удобнее не конечное множество реализаторов, а соответствующую вещественную область, где они лежат (см. теорему 1).

Все остатки реализуются, если $(a, b, c) = 1$ — это следует из алгоритма Евклида. Как правило, у каждого остатка только один реализатор, в этом случае число всех реализаторов равно a. Но если реализаторов больше, то это нам не помешает.

Теорема 1. *Ограничивающая область D ломаная является всегда ступенчатой лестницей* (*диаграммой Юнга*): *если точка q лежит вне области D, то любая точка* $Q \geqslant q$ (*с координатами* $\{(y(Q) \geqslant y(q), z(Q) \geqslant z(q))\}$) *тоже лежит вне области D.*

Пример 1. Для $a = 21$, $b = 31$, $c = 45$, выписывая у каждой целой точки $q = (y, z)$ остаток от деления числа $l(q)$ на a, мы получаем набор остатков, доставляющий выделенные жирным шрифтом реализаторы, образующие обведенную лестничной ломаной область D. Вне этой области те же остатки достигаются (в целых точках положительного квадранта $\{(y \geqslant 0, z \geqslant 0)\}$) при бóльших значениях функции.

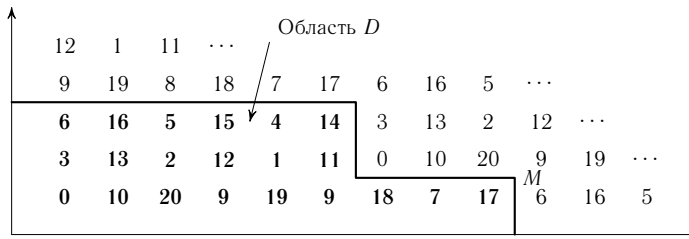


Рис. 12. Лестничная ломаная тройки (21, 31, 45)

Рассмотрим вершину M ограничивающей область D лестничной ломаной, где значение L линейной функции l максимально (таких «максимальных вершин» может быть несколько).

Теорема 2. *Число Фробениуса* $N(a, b, c)$ *дается формулой*

$$N = L - a + 1.$$

Пример. В предыдущем примере (единственная) максимальная вершина M имеет координаты $y = 8$, $z = 0$, так что

$$L = l(8, 0) = 8 \cdot 31 + 0 \cdot 45 = 248 \ (\equiv 17 \pmod{21}).$$

Теорема 2 утверждает в этом случае, что $N(21, 31, 45) = 248 - 21 + 1 = 228$, что верно, но не так уж легко доказывается.

Доказательство теоремы 1. Если точка q лежит вне области D, то в D есть реализатор r со сравнимым с $l(q)$ по модулю a значением $l(r) < l(q)$. Пусть $Q \geqslant q$, докажем, что точка Q тоже лежит вне области D.

Рассмотрим вектор $R = r + (Q - q)$. Имеем $l(R) = l(Q) \pmod a$, $l(R) < l(Q)$, поэтому точка Q не реализатор и лежит вне области D. □

Доказательство теоремы 2.

Лемма 1. *Число $L - a$ не входит в аддитивную полугруппу $P = \{ax + by + cz,\ x \geqslant 0,\ y \geqslant 0,\ z \geqslant 0\}$ комбинаций с целыми коэффициентами (x, y, z).*

Доказательство леммы 1. Если бы имелось представление

$$L - a = aX + bY + cZ, \quad X \geqslant 0,\ Y \geqslant 0,\ Z \geqslant 0, \tag{1}$$

то точка Q с координатами (Y, Z) удовлетворяла бы условиям

$$(l(Q) = bY + cZ) \leqslant (aX + bY + cZ = L - a) < (L = l(M)).$$

Значит точка M не была бы реализатором, вопреки своему определению. Полученное противоречие доказывает невозможность представления (1). Лемма 1 доказана. □

Лемма 2. *Любое целое число $K > L - a$ входит в аддитивную полугруппу P.*

Доказательство леммы 2. Реализатор r остатка от деления целого числа $K > L - a$ на a принадлежит области D тройки (a, b, c). Поэтому выполняются неравенства $l(r) \leqslant L < K + a$ (ведь число L — максимум функции l на реализаторах).

Остатки от деления обоих чисел K и $l(r)$ на a одинаковы. Значит из неравенства $l(r) < K + a$ вытекает неравенство $l(r) \leqslant K$. Поэтому $K = l(r) + xa$, где целое число x неотрицательно. Обозначая координаты точки r через (y, z), мы получаем равенство

$$K = ax + by + cz,$$

доказывающее лемму 2. □

Теорема 2 вытекает из лемм 1 и 2: число Фробениуса N больше $L - a$ по лемме 1 и не превосходит $L - a + 1$ по лемме 2, а потому равно $L - a + 1$. □

Пример. В случае $a = b$ мы находим в качестве максимального реализатора точку M с координатами $(y = 0, z = a - 1)$, поэтому $L = c(a - 1)$ и теорема 2 доставляет ответ $N(a, a, c) = c(a - 1) - a + 1 = (a - 1)(c - 1)$, так что мы по-новому доказали теорему Сильвестра: $N(a, c) = (a - 1)(c - 1)$.

Для быстрого вычисления области D (а значит, по теореме 2, и числа Фробениуса) полезно следующее описание лестничной границы области D.

Теорема 3. *Если точка Q лежит вне области D на 1 выше горизонтального участка границы* (*где $z = \text{const}$*), *то число $l(Q)$ сравнимо*

по модулю a со значением $l(R)$ в одной из точек R нижней границы области D (где $z = 0$).

Более того, любой реализатор R остатка от деления числа $l(Q)$ на a лежит на нижнем краю области D.

Доказательство теоремы 3. Если бы в реализаторе R выполнялось неравенство $z(R) > 0$, то области D принадлежала бы также точка $r = (y(R),\ z(R) - 1)$. В этом случае точка $q = (y(Q),\ z(Q) - 1)$ давала бы при делении значения $l(q)$ на a такой же остаток, как остаток от деления на a меньшего значения $l(r)$.

Поэтому точка q лежала бы вне области D (не была бы реализатором по причине существования соперника r). Иными словами, если бы $z(R)$ было положительным, то исходная точка Q не могла бы лежать сразу над горизонтальным участком границы области D, вопреки условию. Значит, $z(R) = 0$, и теорема 3 доказана. □

Замечание. Меняя местами y и z, мы получаем аналогичное теореме 3 описание *вертикальных* участков ($y = \text{const}$) границы области D: $l(Q) = l(R = (0,\ z(R)) \pmod a$ на следующей вертикали за граничной.

Следствие. *За любой входящей вершиной граничной лестничной ломаной (направленной здесь острием к началу координат) значение функции l в отстоящей на 1 и от горизонтального, и от вертикального отрезков ломаной точке V, делится на a.*

Доказательство следствия. Реализатор r остатка от деления числа $l(V)$ на a должен принадлежать и оси $y = 0$, и оси $z = 0$ (по теореме 3). Значит $r = 0$, $l(r) = 0$, а потому $l(V)$ делится на a. □

Пример. На рис. 13 входящая вершина V имеет координаты $y(V) = 6$, $z(V) = 2$. В ней указан остаток нуль, так как значение $l(V) = 186 + 45 = 231 = 21 \cdot 11$ делится на $a = 21$.

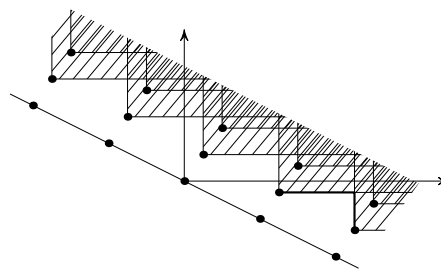


Рис. 13. Построение пилы П с зубцами K_q

Из теорем 1—3 следует быстрый способ построения области D (а стало быть и вычисления числа Фробениуса).

А именно, рассмотрим решетку Γ тех целых точек q плоскости с координатами (y, z), в которых значение $l(q)$ делится на a. Например, в их число входят точки $(c, -b)$ и $(a, 0)$ (не обязательно образующие базис решетки Γ).

Рассмотрим ту часть Γ_+ решетки Γ («полурешетку»), где функция $l = by + cz$ принимает положительные значения. Каждой точке q из полурешетки Γ_+ сопоставим замкнутый «зубец»

$$K_q = \{Q \geqslant q \colon y(Q) \geqslant y(q),\ z(Q) \geqslant z(q)\}$$

с вершиной q. Рассмотрим (рис. 13) пилу из всех таких зубцов

$$\Pi = \bigcup_{l(q)>0} K_q.$$

Теорема 4. *Область D представляет собой дополнение к пиле Π в положительном квадранте $\mathbb{R}^2_+ = \{y \geqslant 0,\ z \geqslant 0\}$ (плоскости с координатами y и z):*

$$D = \mathbb{R}^2_+ \setminus \Pi.$$

Доказательство теоремы 4.

Лемма 1. *Никакой квадрант K_q не пересекается с областью D.*

Доказательство леммы 1. Действительно, если точка q лежит в положительном квадранте плоскости, то (единственным) реализатором остатка от деления числа $l(q)$ на a является точка $0 \in D$, поэтому $q \neq 0$ не принадлежит области D

Если $y(q) > 0$, $z(q) < 0$, то мы рассмотрим разложение

$$q = q' - q'', \quad q' = (y(q), 0), \quad q'' = (0, -z(q)).$$

В этих обозначениях

$$l(q') = l(q) + l(q'') > l(q''),$$

поэтому точка q' не принадлежит области D (ибо в точке q'' остаток от деления l на a такой же, а значение меньше).

По теореме 1 весь квадрант бóльших q' точек не пересекается с областью D, поэтому и весь квадрант K_q больших q точек с областью D не пересекается (ведь в D всюду $z \geqslant 0$).

Итак, $D \cap K_q = \varnothing$ и в рассматриваемом случае, так что лемма 1 доказана. □

Лемма 2. *Если точка $Q \in \mathbb{R}^2_+$ не принадлежит ни одному из квадрантов K_q (где $l(q) > 0$), то точка Q принадлежит области D.*

Доказательство леммы 2. Рассмотрим (рис. 14) реализатор $R \in D$ остатка от деления числа $l(Q)$ на a. Если $l(R) < l(Q)$, то в точке

$q = Q - R$ выполнены такие условия (рис. 14): разность $l(q) = l(Q) - l(R) > 0$ делится на a,

$Q \geqslant q$ (т. е. $y(Q) \geqslant y(q)$, $z(Q) \geqslant z(q)$).

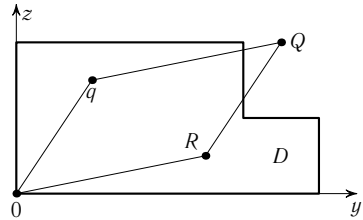


Рис. 14. К доказательству леммы 2

Значит $Q \in K_q$ для точки q полурешетки, где $l(q) > 0$ (точка q не обязана лежать в D).

Следовательно, если точка Q не лежит ни в одном из квадрантов K_q точек полурешетки, то $l(R) = l(Q)$, т. е. точка Q сама является реализатором и принадлежит области D.

Лемма 2 доказана. □

Теорема 4 является прямым объединением утверждений лемм 1 и 2, так что она теперь доказана. □

Доказанные выше результаты о структуре области реализаторов $D(a, b, c)$ можно переформулировать в терминах теории цепных дробей следующим образом.

Рассмотрим для тройки целых чисел (a, b, c) плоскость $\Pi = \{(x, y, z): ax + by + cz = 0\}$ в $\mathbb{R}^3$.

Три прямые (X, где $x = 0$; Y, где $y = 0$; Z, где $z = 0$) делят плоскость Π на шесть «камер Вейля».

Целые точки плоскости Π образуют двумерную решетку Γ. Например, она всегда содержит «точки Косуля»

$$\pm(x = 0,\ y = c,\ z = -b), \quad \pm(x = -c,\ y = 0,\ z = a),$$
$$\pm(x = b,\ y = -a,\ z = 0),$$

но базисные векторы решетки Γ могут быть и иными.

Отличные от O точки решетки Γ в каждой (замкнутой) камере Вейля K образуют там аддитивную полугруппу. Выпуклая оболочка этой полугруппы ограничена ломаной (обращенной выпуклостью к нулю), которая называется *цепной дробью* (тройки (a, b, c) в камере K).

Все эти 6 цепных дробей образуют звездчатый шестиугольник на плоскости Π с вершинами в точках решетки Γ, внутри которого лежит только одна точка O решетки Γ (рис. 15).

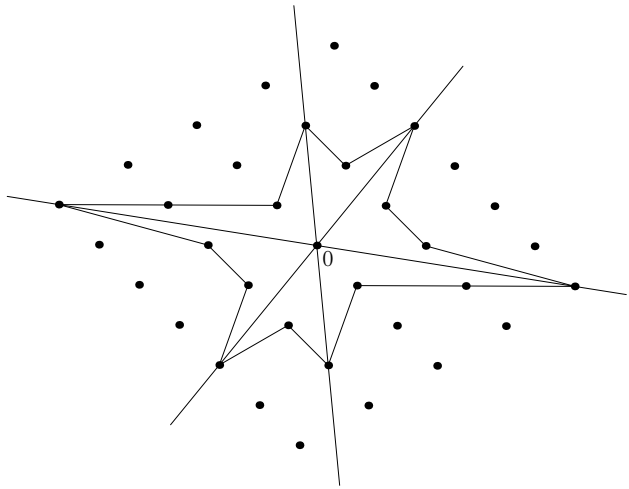


Рис. 15. Шестиугольник цепных дробей решетки с тремя прямыми

Теперь мы опишем область реализаторов $D(a, b, c)$ в терминах геометрии этих цепных дробей (так что и число Фробениуса N тройки (a, b, c) получит описание в терминах цепных дробей).

Замечание. Можно надеяться, что эти геометрические конструкции позволят в дальнейшем обобщить числа Фробениуса на случай несоизмеримых аргументов (a, b, c) (подобно тому, как обычная теория цепных дробей позволяет обобщить алгоритм Евклида нахождения наибольшего общего кратного двух целых чисел, переходя от конечных выпуклых ломаных к бесконечным).

Итак, начнем с трех прямых, проходящих через точку O плоскости $\mathbb{R}^3$ с какой-либо решеткой Γ (с началом O). Выберем один из 6 углов, на которые эти прямые делят плоскость.

Обозначим этот угол через K, составляющие его стороны — через (Y, Z), а третью прямую — через X (рис. 16). Обозначим через Γ_+ ту полурешетку решетки Γ, которая состоит из точек решетки Γ, лежащих от прямой X строго по ту же сторону, что и угол K. Перенесем угол K параллельно в каждую точку q положительной полурешетки Γ_+, так что получится угол K_q с вершиной $q \in \Gamma_+$.

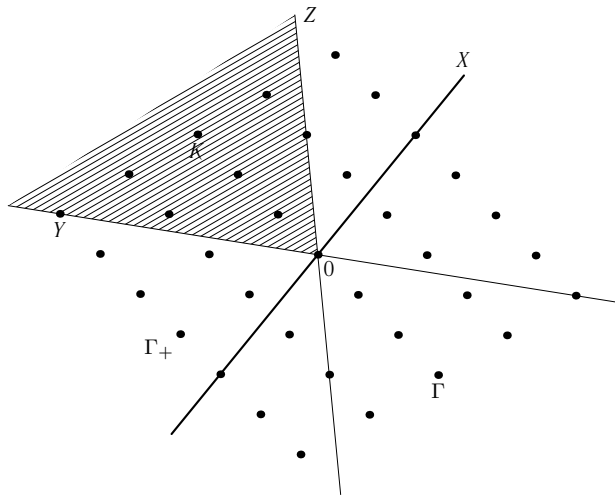


Рис. 16. Исходные данные для построения области D: угол K, решетка Γ и три прямые (X, Y и Z).

Определение. *Областью D угла K* (для решетки Γ и тройки прямых X, Y, Z) называется дополнение в угле K к объединению всех перенесенных углов K_q с вершинами $q \in \Gamma_+$:

$$D = K \setminus \bigcup K_q, \quad q \in \Gamma_+.$$

Пусть три положительных целых числа a, b, c не имеют большего 1 общего делителя.

Теорема 5. *При проектировании трехмерного пространства с координатами (x, y, z) на плоскость с координатами (y, z) (вдоль оси x) область D угла K (где $y \geqslant 0, z \geqslant 0$) плоскости $\Pi = \{(x, y, z)\colon ax + by + cz = 0\}$ с решеткой Γ ее целых точек и тройкой прямых*

$$(X\colon x = 0,\ Y\colon y = 0,\ Z\colon z = 0)$$

проектируется на область $D(a, b, c)$ реализаторов $q = (y, z)$ остатков от деления на a значений линейной функции $l(y, z) = by + cz$.

Замечание. Из этой теоремы следует, в частности, совпадение чисел Фробениуса всех шести цепных дробей камер Вейля на плоскости Π, само по себе геометрически вовсе не очевидное: ведь $N(a, b, c) = N(b, c, a) = \ldots$

Доказательство теоремы 5. Запишем уравнение $ax + by + cz = 0$ плоскости Π в виде

$$z = -(b/a)y - (c/a)z,$$

т. е. в виде $z = -l(y, z)/a$.

Из этой формулы следует, что точки решетки Γ (целых точек плоскости Π) проектируются в точности в те целые точки q плоскости с координатами (y, z), где значение $l(q)$ делится на a.

Описание лестничной границы области $D(a, b, c)$ на плоскости с координатами (y, z), данное в теоремах 1—4, доставляет в терминах проекции (из трехмерного пространства на плоскость с координатами (y, z) вдоль оси x) в точности приведенное выше геометрическое описание области D для угла K (где $y \geqslant 0$, $z \geqslant 0$) на плоскости Π (снабженной решеткой целых точек Γ и тройкой прямых (X, Y, Z), что и доказывает теорему 5. □

Замечание. Кроме доказанных выше теорем мои давние вычисления тысяч чисел Фробениуса привели и к сотням других наблюдений, не получивших пока научных объяснений и общих формулировок. Вот некоторые из этих странных экспериментальных наблюдений:

$$\frac{N(13, 32, 52) = 372}{N(13, 33, 51) = 186} = 2, \quad \frac{N(9, 43, 45) = 336}{N(9, 42, 46) = 168} = 2,$$
$$\frac{N(5, 35, 57) = 224}{N(5, 34, 58) = 112} = 2, \quad \frac{N(4, 20, 73) = 216}{N(4, 19, 74) = 54} = 4,$$
$$N(4, 6 + 4k, 87 - 4k) = 90 \quad (k = 0, 1, 2, \ldots, 14, k \neq 8),$$
$$N(9, 9k \pm 3, 88 - (9k \pm 3)) = 168 \quad (k = 1, 2, \ldots, 7).$$

Интересно было бы понять, как связаны между собой аддитивные полугруппы и цепные дроби разных членов этих серий троек чисел, для троек, имеющих связанные между собой приведенными выше формулами числа Фробениуса.

Перенесение описанной выше теории на случай чисел Фробениуса $N(a_1, a_2, \ldots, a_n)$, где $n > 3$, мало что меняет в ней, только цепные дроби — многомерные.

Продолжение исследований асимптотического поведения чисел Фробениуса больших векторов трехмерного пространства, содержащихся в лекции 2005 года в Дубне, описано в статье Arnold V. I. Geometry and growth rate of Frobenius numbers of additive semigroups // Mathematical Physics, Analysis and Geometry, 2006. 13 p.

Эти результаты, подтверждающие автомодельность усредненного распределения чисел Фробениуса по n-мерному пространству векторов, по-

казывает, что рост чисел Фробениуса порядка σ в степени $1+\frac{1}{n-1}$ встречается чаще, чем рост порядка σ^2 (встречающийся в резонансных местах).

В частности, средние значения чисел Фробениуса $N(a, b, c)$ по треугольникам $a+b+c=\sigma$ растут, по-видимому, как $\sigma^{3/2}$ при больших значениях σ.

Подробности этих исследований будут опубликованы в большой статье Arnold V. I. Arithmetical Turbulence of Selfsimilar Fluctuations Statistics of Large Frobenius Numbers of Additive Semigroups of Integers // Moscow Mathematical Journal. 2007. V. 7, № 2, P. 173—193.

§8. Рапределение точек аддитивной полугруппы на отрезке до числа Фробениуса

Сильвестр доказал, что *точки аддитивной полугруппы с двумя взаимно простыми образующими* ($P=\{xa+yb\}$, $x\geqslant 0$, $y\geqslant 0$) *заполняют ровно половину*[1] *целочисленного отрезка* $\{0, N-1\}$ *до числа Фробениуса* $N(a, b)$ (*а именно, их число равно* $N/2$, *причем точка* p *лежит в полугруппе* P *если и только если сопряженная точка* $q=N-1-p$ *не принадлежит подгруппе* P).

Если число образующих больше 2, то *полугруппа* P *занимает не более половины отрезка* $\{0, N-1\}$. Действительно, если точка p лежит в P, то сопряженная точка $q=N-1-p$ лежать в P не может (иначе сумма $p+q=N-1$ входила бы в полугруппу P, что противоречит минимальности из определения числа Фробениуса N).

В некоторых случаях полугруппа с тремя образующими занимает ровно половину целочисленного отрезка $\{0, N-1\}$.

Пример. $N(3, 4, 7)=N(3, 4)=6$, так как третья образующая 7 не добавляет ничего в полугруппу: $P(3, 4, 7)=P(3, 4)$. Из 6 точек $\{0, \ldots, 5\}$ в эту полугруппу входят 3: $\{0, 3, 4\}$.

В других случаях занятая полугруппой часть целочисленного отрезка до числа Фробениуса меньше его половины.

Пример. $N(4, 5, 7)=7$, а из 7 точек целочисленного отрезка $\{0, 1, \ldots, 6\}$ в полугруппу $P\{4, 5, 7\}$ входят только три точки $\{0, 4, 5\}$, и $3/7<1/2$.

По-видимому, полугруппа $P(a, b, c)$ всегда покрывает не менее трети точек целочисленного отрезка $\{0, 1, \ldots, N(a, b, c)-1\}$ (может быть, нижняя грань занимаемой ею доли даже больше $1/3$). Но это не доказано.

[1] Число N точек этого целочисленного отрезка четно, так как $N=(a-1)(b-1)$ было бы нечетным только если обе образующие были бы четными, что невозможно из-за взаимной простоты.

Попытки понять, почему полугруппа не может занимать слишком малую часть отрезка $\{0, 1, \ldots, N-1\}$ привели к удивительным экспериментальным наблюдениям. Я расскажу здесь о них потому, что надеюсь на участие школьников Дубнинской школы в доказательстве (или опровержении) удивительных гипотез, сформулированных ниже.

Пусть $p \in \{0, 1, \ldots, N-1\}$, где $N = N(a, b, c)$ — число Фробениуса трех образующих (a, b, c) аддитивной полугруппы $P = \{xa + yb + zc : x \geqslant 0, y \geqslant 0, z \geqslant 0\}$ $(x, y, z \in \mathbb{Z}_+)$, не имеющих большего 1 общего делителя.

Определение 1. Число p называется $(+)$-*числом*, если оно входит в полугруппу P.

Пример. 0 является $(+)$-числом, а $N-1$ нет.

Определение 2. Число p называется $(-)$-*числом*, если оно не входит в полугруппу P.

Пример. Число $N-1$ является $(-)$-числом, а число a — нет.

Определение 3. Число q называется *сопряженным* к числу p, если $p + q = N - 1$. Обозначение $q = \bar{p}$.

Очевидно, сопряженным числом к сопряженному к числу p является само число p.

Если p является $(+)$-числом, то сопряженное к нему число $\bar{p}$ является $(-)$-числом.

Ибо иначе число $p + \bar{p} = N - 1$ входило бы в полугруппу P, вопреки определению числа Фробениуса N.

Определение 4. Число p называется $(-, -)$-*числом*, если и оно, и сопряженное ему число являются $(-)$-числами:

$$p \notin P, \quad \bar{p} \notin P.$$

Нашей ближайшей целью будет исследование множества всех $(-, -)$-чисел и числа $\#(\{-, -\})$ его элементов.

Замечание. Количества $\#(\{+\})$ $(+)$-, $\#(\{-\})$ $(-)$-, и $\#(\{-, -\})$ $(-, -)$-чисел связаны следующими очевидными соотношениями:

$$\#(\{+\}) + \#(\{-\}) = N,\ \#(-) - \#(\{+\}) = \#(\{-, -\})$$

Это следует из того, что

$$\#(\{+\}) = \#(\{+, -\}) = \#(\{-, +\}),$$
$$\#(\{-\}) = \#(\{-, -\}) + \#\{(-, +)\}$$

(где $\#(\{\alpha, \beta\})$ есть число точек p класса α, для которых сопряженная точка $\bar{p}$ принадлежит классу β).

В частности,

$$N = 2\ \#(\{+\}) + \#(\{-, -\}),$$

поэтому *число Фробениуса N и число точек* $(-, -)$ *одинаковой четности*.

В частности, *если число Фробениуса* $N(a, b, c)$ *нечетно, то* $\#(\{-, -\}) > 0$, *т. е. точки типа* $(-, -)$ *существуют* (чего не бывает для полгрупп с двумя образующими).

Пример. $N(5, 17, 19) = 34$. Числа типа $(+)$: {0, 5, 10, 15, 17, 19, 20, 22, 24, 25, 27, 29, 30, 32}. Их 14.

Числа типа $(-)$: {1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 16, 18, 21, 23, 26, 28, 31, 33}. Этих чисел 20.

Числа типа $(-, -)$: {2, 7, 12, 21, 26, 31}. Этих чисел 6.

6 чисел типа $(-, -)$ образуют замечательную диаграмму

$$\begin{array}{ccccc} 2 & \longrightarrow & 7 & \longrightarrow & 12 \\ \downarrow & & \downarrow & & \downarrow \\ 21 & \longrightarrow & 26 & \longrightarrow & 31, \end{array}$$

где «$\rightarrow$» означает «прибавить $a = 5$», «$\downarrow$» означает «прибавить $c = 19$».

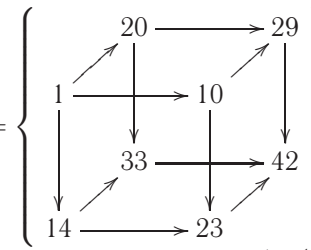
Обозначим через m наименьшее из числе типа $(-, -)$, а через M — наибольшее из чисел типа $(-, -)$.

Гипотеза 1. *Разность M и m — всегда число типа* $(+)$.

Пример. В предыдущем примере $M = 31$, $m = 2$, $M - m = 29 = 2a + c$ (потому что приведенная выше прямоугольная диаграмма имеет длину 2 направления a и высоту 1 наравления c).

Основание гипотезы состоит в том, что числа типа $(-, -)$ в многочисленных примерах образуют прмолинейные, прямоугольные или параллелипипедальные диаграммы, подобные приведенному прямоугольнику из 6 точек типа $(-, -)$.

Пример. $N(10, 13, 48) = 56$, $\{(-, -)\} = \{8 \rightarrow 21 \rightarrow 34 \rightarrow 47\}$ — отрезок $b = 13$ — направления из 4 вершин

Пример. $N(9, 13, 19) = 44$, $\{(-, -)\} =$  — параллелипипед из 8 вершин с ребрами «$\rightarrow$» a-направления, «$\downarrow$» b-направления, «$\nearrow$» c-направления длины 1.

Гипотеза 2. *Множество точек типа* $\{(-, -)\}$ *всегда представляет собой параллелепипед* (*размерности* 1, 2 *или* 3), *ребра которого*

соединяют вершины m и M направленными отрезками, изображающими увеличение вершины на a, на b или на c. А именно, если $M = m + (ua + vb + wc)$, то ребра a-направления имеют длину u, b-направления — длину v, c-направления — длину w.

Я сформулировал здесь эти гипотезы не потому, что умею их доказывать, а потому, что надеюсь, что эксперименты учащихся Дубнинской школы (возможно, компьютеризированные) помогут либо опровергнуть их контрпримером, либо пополнить имеющиеся у меня списки подтверждающих примеров более убедительными доводами в пользу этих гипотез.

Надежда использовать описанную выше предполагаемую структуру множеств $\{(-,\,-)\}$ для оценки количества $(+)$-точек снизу основана на надежде извлечь из этой структуры оценку количества $(-,\,-)$-точек сверху.

Например, *чтобы доказать оценку* $\{(+)\} \geqslant \varepsilon N$ *достаточно было бы проверить, что* $\{(-,\,-)\} \leqslant (1 - 2\varepsilon N)$.

В известных мне примерах неравенство $\{(-,\,-)\} \leqslant N/3$ всегда выполнено (с запасом) так, что выполнено и неравенство $\{(+)\} \geqslant N/3$ (и даже более сильное неравенство), но в общей ситуации эти оценки не доказаны.

Некоторый аналог этих гипотетических оценок доставляет следующая

Задача. *Какую часть объема тетраэдра может покрывать целиком содержащийся в нем параллелипипед*?

Гипотеза 3. $\dfrac{\text{(объем параллелепипеда)}}{\text{(объем симплекса в } \mathbb{R}^n)} \leqslant \dfrac{n!}{n^n}$.

Пример. При $n = 2$ правая часть равна $1/2$, и соедержащийся в треугольнике параллелограмм может покрыть половину его площади и не может покрыть больше, что легко доказать.

При $n = 3$ правая часть $n!/n^n$ равна $2/89$. Такую долю объема пирамиды нетрудно накрыть подходящим лежащим в ней параллелипипедом. Например, для пирамиды $\{x \geqslant 0,\ y \geqslant 0,\ z \geqslant 0,\ x + y + z = 1\}$ две девятых ее объема покрывает куб $\{0 \leqslant x \leqslant 1/3,\ 0 \leqslant y \leqslant 1/3,\ 0 \leqslant z \leqslant 1/3\}$. Доказать, что нельзя покрыть бóльшую $2/9$ часть объема не так просто (но, думаю, не слишком и трудно для дубнинских школьников).

Надежда использовать эту задачу для оценки сверху числа целых точек предполагаемого параллелипипеда $\{(-,\,-)\}$ основана на том, что результаты выпуклой геометрии многогранников обычно имеют в геометрии Минковского целочисленные аналоги (где роль объема многогранника играет число его целых точек)

Замечание. Число Фробениуса N — не число целых точек в пирамиде $\{x \geqslant 0,\ y \geqslant 0,\ z \geqslant 0,\ ax + by + cz = N - 1\}$, куда предполагается вкладывать параллелипипед $\{-,\,-\}$, а своеобразный численный аналог высоты этой пирамиды (опущенной на «гипотенузу» из вершины O). Такой же «высотой» является и оцениваемое сверху число точек типа $(-,\,-)$.

Поэтому для доказательства оценки числа точек типа $(-, -)$ сверху через число Фробениуса $N(a, b, c)$ гипотеза 3 недостаточна (хотя, она, вероятно, практически необходима для изобретения этого доказательства).

Все же объем пирамиды оцене сверху величиной порядка N (при доказательстве в §4.3 оценки числа Фробениуса снизу $N \geqslant \mathrm{const} \cdot \sigma^{3/2}$). Высота $\#\{(-, -)\}$ф предполагаемого параллелипипеда не превосходит числа его целых точек, которое напоминает его объем, оцениваемый гипотезой 3. Поэтому гипотеза 3 позволяет надеяться на оценку числа точек типа $(-, -)$ сверху величиной определенной доли числа Фробениуса, $\#(\{-, -\}) \leqslant (1 - 2\varepsilon)N$.

Было бы интересно исследовать не только полную массу точек полугруппы на отрезке до числа Фробениуса, но и характер распределения точек полугруппы на этом отрезке.

Эмпирическая плотность этого распределения часто оказывается растущей примерно степенным образом (с предполагаемым показателем 2, превращающимся в $n - 1$ для полугрупп с n образующими).

Показатель $n - 1$ объясняется соотношением $dl^n = nl^{n-1}dl$ (рис. 4 на с. 89). Но резонансы (вроде соотношения $P(a, b, c) = P(a, b)$ при $c = a + b$) нарушают связь между числом точек проекции целых точек n-мерной пирамиды и l^n, поэтому превратить указанное наблюдение роста плотности в теорему не так уж легко.

Я сформулировал выше программу оценки числа точек типа $(+)$ снизу для случая трех образующих полугруппы. Но аналогичная программа пригодна и для $n > 3$ образующих (с предположенной еще в моей цитированной выше работе 1999 года гипотетической оценкой $\#\{(+)\} \geqslant N/n$ и с гипотезой о росте плотности)

Владимир Игоревич Арнольд

Экспериментальное наблюдение математических фактов

Подписано в печать 15.12.2006 г. Формат $60 \times 90\ {}^{1}/_{16}$. Бумага офсетная № 1.
Печать офсетная. Печ. л. 7,5. Тираж 2000 экз. Заказ №

Издательство Московского центра
непрерывного математического образования

119002, Москва, Большой Власьевский пер., 11.

Отпечатано с готовых диапозитивов в ФГУП «Полиграфические ресурсы».

Книги издательства МЦНМО можно приобрести в магазине «Математическая книга»,
Большой Власьевский пер., д. 11. Тел. 241 72 85. E-mail: biblio@mccme.ru