НЕЗАВИСИМЫЙ МОСКОВСКИЙ УНИВЕРСИТЕТ

globus ГЛОБУС

Общематематический семинар. Выпуск 3

Под редакцией М. А. Цфасмана и В. В. Прасолова

Москва
Издательство МЦНМО
2006

УДК 51(06)
ББК 22.1я5
Г54

Издание осуществлено при поддержке РФФИ (издательский проект № 03-01-14113).

РФФИ

Глобус. Общематематический семинар / Под ред. М. А. Цфасмана и В. В. Прасолова. — М.: МЦНМО, 2004— . — ISBN 5-94057-064-X.

Г54

Вып. 3. — 2006. — 164 с. — ISBN 5-94057-259-6.

Цель семинара «Глобус» — по возможности восстановить единство математики. Семинар рассчитан на математиков всех специальностей, аспирантов и студентов.

Третий выпуск включает доклады С. Алескера, В. М. Бухштабера, П. Делиня, С. Б. Каток, А. Н. Паршина, А. Б. Сосинского, А. Г. Хованского, М. А. Цфасмана, С. Б. Шлосмана.

УДК 51(06), ББК 22.1я5

ГЛОБУС
Общематематический семинар. Выпуск 3

Научный редактор *М. А. Цфасман*
Редактор *В. В. Прасолов*
Тех. редактор *А. С. Протопопов*

Лицензия ИД №01335 от 24.03.2000 г. Подписано в печать 12.10.2006 г. Формат $70 \times 100\ {}^{1}/_{16}$. Бумага офсетная №1. Печать офсетная. Печ. л. 10,25. Тираж 800 экз.
Заказ №

Издательство Московского центра непрерывного математического образования.
121002, Москва, Большой Власьевский пер., 11. Тел. (095) 241–74–83.

Отпечатано с готовых диапозитивов в ППП «Типография „Наука“».
119009, Москва, Шубинский пер., 6.

Книги издательства МЦНМО можно приобрести в магазине «Математическая книга», Большой Власьевский пер., д. 11. Тел. (095) 241–72–85. E-mail: biblio@mccme.ru

ISBN 5-94057-064-X
ISBN 5-94057-259-6 (Вып. 3)

Предисловие

Предлагаем Вам третий сборник докладов на семинаре «Глобус» — общематематическом семинаре Независимого Московского университета. Как и ранее, авторы рассказывают математикам других специальностей, как они видят свою область и что в ней нового. Начинается сборник с двух докладов А. Г. Хованского. Первый — по алгебре: описывается роль многогранников Ньютона при изучении систем алгебраических уравнений. Второй — по геометрии: изучается задача описания всюду гиперболических поверхностей. В докладе С. Б. Шлосмана рассказывается, как в комбинаторике и статфизике возникают похожие геометрические вариационные задачи. А. Н. Паршин рассказывает об n-мерных локальных полях и законах взаимности. Доклад А. Б. Сосинского посвящен связи между топологией и математической логикой: можно ли свести проблему Пуанкаре к алгоритмической разрешимости некоторых задач о представлениях групп? С. Алескер излагает теорию валюаций на выпуклых множествах — функционалов аддитивных относительно объединений. В моем докладе делается попытка объяснить математикам далёким от алгебраической геометрии и теории чисел, какая геометрия может быть над конечным полем. В. М. Бухштабер рассказывает красивую задачу из теории инвариантов. П. Делинь говорит о значениях поли-дзет и их связи со связностями и алгебрами Хопфа. Завершает сборник статья С. Б. Каток о геодезическом потоке на модулярной группе.

Как и в предыдущих сборниках, набор сюжетов весьма разнообразен. Несмотря на «популярность» этих лекций, понять все — не просто.

Доклады воссозданы по записям В. В. Прасолова с помощью авторов и издаются трудами А. С. Протопопова, В. Ю. Радионова, Ю. Н. Торхова и других сотрудников издательства МЦНМО. Всем им — большая благодарность.

Семинар продолжает работать; можно надеяться, что и издание его трудов будет продолжаться. До новых встреч.

М. А. Цфасман

А. Г. Хованский

СИСТЕМЫ УРАВНЕНИЙ С МНОГОГРАННИКАМИ НЬЮТОНА ОБЩЕГО ПОЛОЖЕНИЯ

Я буду рассказывать об одной довольно необычной ситуации с многогранниками Ньютона. Сначала я расскажу вообще о многогранниках Ньютона, потом расскажу про эту ситуацию, а потом немножко расскажу при теорию Паршина—Като, которая с этим связана. Программа лекции:

1) Многогранники Ньютона (вообще — что это такое, какие есть варианты, какие есть решённые задачи и какие нерешённые).

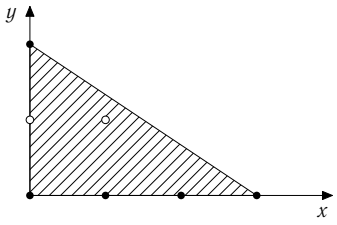
2) Системы n уравнений от n неизвестных, многогранники Ньютона которых находятся в общем положении.

В пункте 2) — две части. Одна более старая, которую мы получили с Ольгой Гельфонд [1, 2]. Это — явная формула для суммы любой функции по корням системы. Умея вычислять такую сумму, можно вычислить всё, что угодно. Можно исключать неизвестные, можно находить число вещественных корней, число вещественных корней в области, ограниченной заданными полиномиальными неравенствами и т. д. У меня есть значительно более новый результат [3]. Я нашёл произведение всех корней системы уравнений. Дело в том, что корни лежат в группе $(\mathbb{C}^*)^n$. Все корни такой системы можно перемножить, и для произведения получается совершенно явная формула, аналогичная формуле Виета. У меня получились две формулы для произведения корней, абсолютно непохожие друг на друга. Одна формула в духе многогранников Ньютона. Там фигурируют смешанные объёмы, производные. А во второй формуле фигурирует необычный объект, который называется *символом Паршина—Като*. Когда эта формула написалась, возникло удивительно симметричное выражение, которое напоминало те выражения, которые встречаются в одномерном законе взаимности Вейля. Я спросил Сашу Бейлинсона, знает ли он какие-либо многомерные обобщения теоремы Вейля. Он сказал: «А как же!» и сослался на теорию Паршина—Като. Я немножко расскажу про эту теорию. У меня была надежда, что при помощи этой теории можно будет наши результаты упростить. Но, честно говоря, вышло всё наоборот. Я, скорее, сильно упростил теорию Паршина—Като. Точнее, упростил законы взаимности из этой теории в случае,

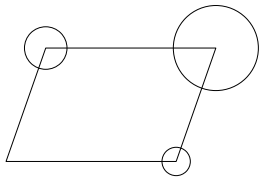
когда основное поле является полем комплексных чисел. Я, правда, про это ещё не готов рассказывать. Но надеюсь, что я это скоро закончу, и что эта вещь будет совсем элементарной.

Многогранники Ньютона

Я начну с простейшего примера. Рассмотрим уравнение $P(x, y) = y^2 + a_0 + a_1 x + a_2 x^2 + a_3 x^3 = 0$. Какая у этого уравнения степень? По определению степень этого уравнения 3, но в нём присутствуют далеко не все члены степени $\leqslant 3$. Давайте рассмотрим вещественную плоскость и отметим все целые точки, которые соответствуют мономам, входящим в уравнение с ненулевыми коэффициентами. Рассмотрим выпуклую оболочку всех этих точек (рис. 1). Эта выпуклая оболочка называется *многогранником Ньютона* полинома P и обозначается $\Delta(P)$. Она играет такую же роль, как степень. Только степень — это число, а многогранник Ньютона — это геометрическая фигура. Поэтому всё получается значительно интересней.



Р и с. 1. Многогранник Ньютона



Р и с. 2. Сумма выпуклых фигур по Минковскому

Что получается интересного в нашем простейшем примере? Рассматриваемая кривая эллиптическая; она имеет род 1. А внутри многоугольника Ньютона есть ровно одна целая точка. И это не случайно; так бывает всегда. Потом я расскажу про это подробнее.

Что мы вообще знаем про степень? Мы знаем, что $\deg PQ = \deg P + \deg Q$. Оказывается, что для многогранников Ньютона тоже $\Delta(PQ) = \Delta(P) + \Delta(Q)$, только эту сумму нужно понимать в смысле Минковского. Как складываются выпуклые фигуры по Минковскому? Пусть в линейном пространстве есть две выпуклые фигуры. Рассмотрим суммы всех пар векторов, у которых один конец лежит в одной фигуре, а другой конец лежит в другой фигуре (рис. 2). Получается фигура, которая тоже будет выпуклой. Она называется суммой Минковского рассматриваемых выпуклых фигур.

Легко доказывается, что если мы перемножим два полинома, то их многогранники Ньютона сложатся. Это происходит из-за того, что при перемножении мономов их степени складываются.

Многогранники Ньютона ведут себя похоже на степени многочленов. Какие есть классические результаты про системы уравнений фиксированных степеней? Если есть система уравнений, причём все уравнения имеют заданные степени и достаточно общие коэффициенты, то оказывается, что дискретные комплексные геометрические инварианты полученного многообразия (скажем, в проективном пространстве) зависят только от степеней и совершенно явно вычисляются. Рассмотрим, например, кривую степени n, заданную уравнением $P_n(x, y) = 0$. Если коэффициенты уравнения достаточно общие, то род g этой кривой равен $\frac{(n-1)(n-2)}{2}$ (формула Римана). Как всё это обобщается на системы уравнений с фиксированными многогранниками Ньютона и с достаточно общими коэффициентами? Прежде всего, все классические вычисления, какие только есть, переносятся на случай систем с фиксированными многогранниками Ньютона. Более того, со времён, когда классики это считали, было открыто много новых дискретных инвариантов. Я имею в виду, например, числа Ходжа смешанных структур Ходжа. Все эти новые инварианты, так же, как и старые, тоже вычисляются в терминах многогранников Ньютона, если коэффициенты уравнений достаточно общие.

Пример 1. Формула для рода кривой обобщается следующим образом. Я расскажу обобщение для случая гиперповерхности. Пусть есть одно уравнение $P(x_1, \ldots, x_n) = 0$ от n переменных, с многогранником Ньютона Δ. Спрашивается, каков род поверхности, заданной этим уравнением. Род поверхности — это число независимых голоморфных форм старшей степени на произвольной гладкой компактификации этой поверхности. Оказывается, что род гиперповерхности Γ задается формулой $g(\Gamma) = \#\big(\mathbb{Z}^n \cap (\Delta \setminus \partial\Delta)\big)$, т. е. род равен числу точек целочисленной решётки, лежащих строго внутри многогранника Ньютона. Например, формула для рода кривой получается таким образом. Многоугольник Ньютона общего уравнения степени n — это стандартный треугольник (рис. 3). Посчитаем число целых точек внутри этого стандартного треугольника. В этом треугольнике в верхнем ряду одна целая точка, в следующем ряду две целые точки, затем три и т. д. Если вы просуммируете арифметическую прогрессию, то как раз и получите число $\frac{(n-1)(n-2)}{2}$. Так что формула для рода гиперповерхности — прямое обобщение формулы Римана. Только ответ получается геометрический — число целых точек, лежащих внутри многогранника Ньютона.

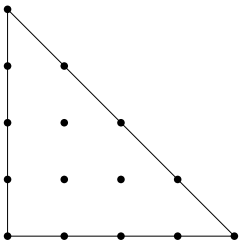


Рис. 3. Многоугольник Ньютона общего уравнения степени n

Здесь важно сказать, что все уравнения значительно лучше рассматривать не в $\mathbb{C}^n$, а в $(\mathbb{C}^*)^n$. Это означает, что мы должны выбросить все координатные плоскости из пространства $\mathbb{C}^n$. Почему лучше? Потому, что все ответы получаются значительно более симметричными. А если мы знаем ответы в пространстве $(\mathbb{C}^*)^n$, то ответы в $\mathbb{C}^n$ по ним восстанавливаются. Многие инварианты оказываются аддитивными, нужно их просто сложить по координатным плоскостям. А неаддитивные инварианты обычно выражаются через аддитивные. Просто формулы в $\mathbb{C}^n$ оказываются сложнее, а в $(\mathbb{C}^*)^n$ они удивительно симметричны.

Раз уж мы об этом заговорили, то я объясню, почему это происходит, и какое есть разумное обобщение всех вопросов про многогранники Ньютона. Формулы симметричны потому, что $(\mathbb{C}^*)^n$ — группа по умножению: наборы n ненулевых комплексных чисел можно покоординатно перемножать. А мономы — это не что иное, как характеры этой группы, т. е. гомоморфизмы $\chi\colon (\mathbb{C}^*)^n \to \mathbb{C}^*$. Каждый моном имеет вид $x_1^{a_1}\dots x_n^{a_n}$. Среди характеров встречаются и такие, в которых некоторые степени $a_1, \dots, a_n$ отрицательные. И для теории многогранников Ньютона это тоже совершенно неважно — можно допускать мономы с отрицательными степенями. Линейные комбинации таких обобщённых мономов называются полиномами Лорана. Переход от полиномов к полиномам Лорана доставляет дополнительные удобства и никак не усложняет задачу вычисления дискретных инвариантов.

Поэтому сам вопрос относится к группе. У нас есть группа $\mathcal{G}$, на ней есть характеры и их линейные комбинации. Мы спрашиваем, что можно сказать про нули такой линейной комбинации. Вот более общий вопрос. (Я действительно верю, что он должен разрешиться. Он фактически не двигался с места только потому, что никто этого не делал.) Давайте вместо $(\mathbb{C}^*)^n$ возьмём любую редуктивную группу $\mathcal{G}$, а вместо многогранника Ньютона её представление $\mathcal{G} \to \mathrm{GL}(n)$. С представлением связано линейное пространство функций на группе — линейные комбинации матричных элементов. Можно говорить об общих функциях этого линейного пространства. Возьмём такую функцию и приравняем её нулю. У нас получится гиперповерхность в группе. Разумеется, её дискретные свойства должны зависеть только от группы и от представления. Когда есть несколько характеров группы $(\mathbb{C}^*)^n$, о них можно думать как о диагональном представлении этой группы. Общая матричная функция такого представления будет как раз полиномом Лорана. Здесь получится много вопросов, но ни для каких других групп почти ничего не сделано. Есть редкие счастливые исключения. Одно счастливое исключение — это формула Бори Казарновского. Он посчитал

в этой общей ситуации число решений системы n уравнений на n-мерной группе.

Итак, эта область почти полностью открыта. Ясно, что тут должно быть много интересного.

Пример 2. Рассмотрим n уравнений от n неизвестных: $P_1 = \ldots = P_n = 0$. Предположим, что у всех у них многогранник Ньютона одинаков: $\Delta(P_1) = \ldots = \Delta(P_n) = \Delta$. Тогда оказывается, что число решений этой системы (если уравнения достаточно общие и если решения рассматривать в $(\mathbb{C}^*)^n$) равно $n!V(\Delta)$, где V — объём. Это теорема Кушниренко 1975 года. В ответах, как вы видите, встречаются целые точки, встречаются объёмы.

В ответах на более сложные вопросы начинает встречаться вся комбинаторика многогранника: число граней разных размерностей, число флагов разных граней, число точек в этих флагах, объёмы разных измерений на разных целочисленных гранях. Всё это постепенно завязывается в ответы. Мы получаем связь между обычной геометрией многогранников (с числами целых точек на них, с их комбинаторикой и т. д.) и алгебраической геометрией. И эта связь работает в обе стороны. Иногда алгебраическая геометрия подсказывает совершенно неожиданные формулы про многогранники. Иногда их удаётся доказать отдельно, не используя алгебраической геометрии. Некоторые вещи до сих пор не доказаны отдельно. И наоборот, геометрия многогранников помогает продвинуться в алгебраической геометрии. Я надеюсь, что мне удалось упростить теорию Паршина—Като. И, конечно же, для меня это упрощение пришло из геометрии многогранников.

Итак, у нас есть связь алгебраической геометрии с обычной геометрией многогранников. Мы обсудим примеры ответов для рода и числа решений. Кстати, формула для числа решений была обобщена Д. Бернштейном на тот случай, когда многогранники Ньютона $\Delta_1, \ldots, \Delta_n$ уравнений $P_1 = \ldots = P_n$ системы могут быть разными. В этом случае число решений равно $n!V(\Delta_1, \ldots, \Delta_n)$, где $V(\Delta_1, \ldots, \Delta_n)$ — смешанный объём многогранников. Смешанный объём — это замечательная функция, которая линейна (относительно сложения по Минковскому) по каждому аргументу, симметрична, а на диагонали совпадает с объёмом. Такая функция только одна. Она была открыта Минковским, и активно им использовалась.

Пример 3. Теперь я приведу формулу для эйлеровой характеристики. Пусть есть гиперповерхность Γ, заданная уравнением $P = 0$ в торе $(\mathbb{C}^*)^n$. Тогда эйлерова характеристика гиперповерхности Γ равна $(-1)^{n-1}n!V(\Delta)$, где $\Delta = \Delta(P)$. Если гиперповерхность задана в $\mathbb{C}^n$, то надо воспользоваться аддитивностью эйлеровой характеристики и сложить

эйлеровы характеристики пересечений гиперповерхности с координатными плоскостями. Получится аналогичная формула, гораздо менее красивая и более громоздкая.

Кстати сказать, совпадение (с точностью до знака) эйлеровой характеристики и числа Кушниренко в действительности выглядит загадочно. Не видно никаких причин, чтобы эти два числа совпадали. В формуле для эйлеровой характеристики гиперповерхности в компактном многообразии фигурируют классы Черна. Формула сложная, и первый её член действительно является аналогом числа $(-1)^{n-1}n!V(\Delta)$. Но там есть ещё много других членов. Они все фантастическим образом сокращаются. Я когда-то просчитал эйлерову характеристику гиперповерхности в торе $(\mathbb{C}^*)^n$ при помощи классов Черна. Оказалось, что всё действительно сокращается. Если вместо тора взять другую группу, то такого сокращения не происходит.

Пример 4. Я приведу ещё один пример, чтобы показать, как из алгебры иногда вытекают содержательные геометрические утверждения. Правда, то утверждение, которое я сейчас приведу, очень простое. Его, конечно, можно доказать другим способом.

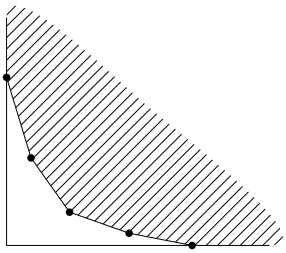
Рассмотрим уравнение $P(x, y) = 0$ на плоскости (точнее, в торе $(\mathbb{C}^*)^2$). Та кривая, которую мы получим, будет некомпактна. Компактную кривую вообще нельзя аналитически вложить в аффинное пространство. Такое вложение противоречило бы принципу максимума. Значит, эта кривая является сферой с ручками, имеющей проколы. Число ручек, как мы знаем, это число целых точек внутри многоугольника $\Delta = \Delta(P)$. Оказывается, что число проколов (число дырок) равно числу целых точек на границе многоугольника $\Delta = \Delta(P)$. А эйлерова характеристика — это удвоенный объём (в данном случае — удвоенная площадь $2S(\Delta)$). Но если мы знаем, что наше пространство есть сфера с ручками и с проколами, и знаем число ручек и число проколов, мы можем вычислить эйлерову характеристику. Поэтому у нас получается соотношение: эти числа не независимые. Значит, есть соотношение между площадью многоугольника, числом целых точек внутри многоугольника и числом целых точек на границе многоугольника. Если это соотношение написать, то получится так называемая *формула Пика*:

$$S(\Delta) = \#\mathbb{Z}^2 \cap (\Delta \setminus \partial\Delta) + \frac{1}{2}\#\mathbb{Z}^2 \cap \partial\Delta - 1.$$

Эта формула есть прямое применение алгебраических вычислений в геометрии многоугольников. Правда, это применение ерундовое. К тому же мы таким способом доказали формулу Пика только для выпуклых многоугольников, а она верна и для невыпуклых многоугольников тоже.

Но так как у алгебраических многообразий дискретных инвариантов много, и соотношений между ними тоже много, то можно себе представить, что из них получаются совершенно странные, с геометрической точки зрения удивительные, утверждения про целочисленные многогранники.

У теории многогранников Ньютона есть много вариантов. Один — это обобщение теории на другие группы. Есть и другие варианты, которые продвинуты не хуже, чем основной. Основной вариант — это когда мы рассматриваем общие полиномиальные уравнения в $(\mathbb{C}^*)^n$. Другие варианты здесь такие. Один — глобальный, когда мы рассматриваем общие полиномиальные уравнения в $\mathbb{C}^n$. Я уже говорил, что в этом варианте вопросы естественней, а ответы сложнее. Есть ещё локальный вариант. Представьте себе, что у нас есть аналитическая функция $f(z_1, \ldots, z_n)$, у которой в нуле есть особая точка. Возьмём ряд Тэйлора этой функции в нуле. Если в нуле особая точка, то можно считать, что первых членов ряда Тэйлора у этой функции нет. Отметим точки решётки, соответствующие ненулевым мономам, и возьмём их выпуклую оболочку (рис. 4). Получится бесконечный многогранник. У этого многогранника есть набор компактных граней. У особенности есть много дискретных инвариантов. Оказывается, что если функция f достаточно общая, то все дискретные инварианты выражаются через многогранник Ньютона.



Р и с. 4. Многогранник Ньютона особой точки

Теория особенностей занимается тем, что она рассматривает всё более и более особые точки. И чем более точка особая, тем больше в эту точку засасывается алгебраической геометрии. В данном случае, та алгебраическая геометрия, которая попадает сюда, — это теория многогранников Ньютона.

Ещё есть вещественный вариант. Многогранники Ньютона помогают строить нетривиальные примеры вещественных алгебраических многообразий с предписанными свойствами. Этот метод придумал Олег Виро. Ситуация здесь такая. До появления работы Виро существовали единичные, отдельные примеры хитрых вещественных алгебраических многообразий. Но никто не знал никакого большого списка примеров (были две-три серии и несколько отдельных примеров). После появления метода Виро возникла масса примеров. Иногда они окончательные, иногда почти окончательные. Этот метод настолько силён, что есть даже гипотеза (я, правда, сомневаюсь, что она верна), что основная масса всех возможных примеров может быть построена этим методом.

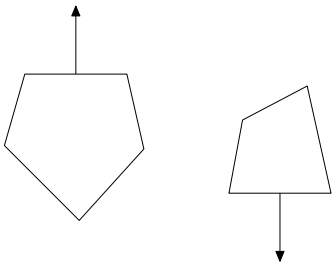
Ещё один вариант очень похож на замену группы: вместо группы $(\mathbb{C}^*)^n$ берётся аддитивная группа $\mathbb{C}^n$. Этот вариант такой. Вместо полиномиальных уравнений можно рассматривать уравнения вида $\sum c_\alpha \exp(\alpha_1 z_1 + \ldots + \alpha_n z_n) = 0$, где $\alpha = (\alpha_1, \ldots, \alpha_n)$. Выражение $\exp(\alpha_1 z_1 + \ldots + \alpha_n z_n)$ очень похоже на моном $z_1^{\alpha_1} \cdot \ldots \cdot z_n^{\alpha_n}$. Более того, если все числа α_k целые, то логарифмическим преобразованием $z_1 = \ln x_1, \ldots, z_n = \ln x_n$ функция $\sum c_\alpha \exp(\alpha_1 z_1 + \ldots + \alpha_n z_n)$ сводится к обычному полиному Лорана. Но эти числа могут быть и не целыми. Они могут быть вещественными и даже комплексными.

Если есть такая конечная экспоненциальная суммы, то у неё есть многогранник Ньютона. Он строится следующим образом. Отметим все точки α (в $\mathbb{R}^n$ или в $\mathbb{C}^n$) и возьмём их выпуклую оболочку. Оказывается, что выпуклая оболочка в очень многом ответственна за свойства такой суммы. Например, она влияет на асимптотику роста числа нулей (в одномерном случае), на асимптотику роста дискретных инвариантов. Многообразия, которые здесь получаются, не алгебраические, а скорее, квазипериодические. Можно, например, рассматривать шар большого радиуса R, брать порцию многообразия в этом шаре и вычислять какую-нибудь его характеристику (например, эйлерову характеристику), а затем делить её на подходящую степень радиуса R. Оказывается, что такие отношения часто имеют пределы. Иногда эти пределы — те, которые можно ожидать. Иногда они совершенно неожиданные. Обычно ситуация такая. Если числа α_k вещественные, то получается вещественный многогранник — вроде многогранника Ньютона. Только его вершины не целые точки, а произвольные. В этом случае ответы получаются очень похожие на соответствующие ответы в теории многогранников Ньютона. Однако если эти числа комплексные, ответы получаются совсем другие. Здесь сделано не так уж много, и почти всё, что сделано, сделано Борей Казарновским. Некоторые его ответы очень красивые.

Общие системы из n уравнений от n неизвестных

Теперь я хочу перейти собственно к тому, о чём я хотел рассказать. Это будет состоять из двух частей. Одна старая, а другая более новая. Я начну с более старой, с нашей совместной работы с Ольгой Гельфонд. Пусть есть система уравнений $P_1 = \ldots = P_n = 0$, у которых многогранники Ньютона $\Delta_1, \ldots, \Delta_n$ расположены достаточно общим образом друг относительно друга.

Для начала я определю, что означает общность взаимного расположения многоугольников на плоскости. В n-мерном пространстве определение



Р и с. 5. Параллельные стороны с противоположно направленными нормалями

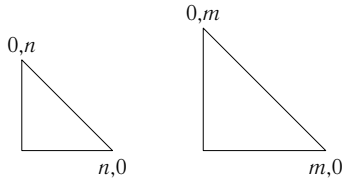
более сложное; я дам его позже. Скажем, что два многоугольника на плоскости *расположены общим образом относительно друг друга*, если у них нет параллельных сторон с одинаковым направлением внешних нормалей. Параллельные стороны с противоположно направленными внешними нормалями у них могут быть (рис. 5).

Пусть, например, есть два уравнения с двумя неизвестными, у которых многоугольники Ньютона расположены достаточно общим образом относительно друг друга. Что можно сказать про такую систему уравнений? По теореме Бернштейна мы знаем число её решений: оно равно $2\,\mathrm{vol}(\Delta_1, \Delta_2)$. Удивительным образом оказывается, что для такой системы можно сказать гораздо больше. В каком-то смысле, эту систему можно решить. «Решить» здесь означает, что можно исключить все неизвестные, кроме одного, т. е. свести к одному уравнению с одним неизвестным. Это можно сделать совершенно явно. Можно также явно вычислить число вещественных корней (если коэффициенты вещественные); можно вычислить число вещественных корней в какой-либо полуалгебраической области. В общем, с этой системой можно сделать всё, что угодно, и это можно сделать достаточно явно. Я бы сказал, что это поразительно. Это выпадает из общей идеологии многогранников Ньютона. В многогранниках Ньютона главная идеология заключается в следующем. Пусть уравнения достаточно общие. Тогда от их коэффициентов не зависит ничего — все дискретные инварианты вычисляются по многогранникам. Например, для одного уравнения от одного неизвестного число корней равно степени уравнения. Но это рассуждение ничего не может сказать о расположении корней. Конечно же, нельзя найти корни, не зная коэффициентов уравнения. Все коэффициенты будут играть роль в тех формулах, которые я напишу. В каком-то плане эти формулы не из нашей оперы про многогранники Ньютона. Но это не совсем так. Оказывается, что эта формула распадается на части. Одни части этой формулы абсолютно из нашей оперы — они зависят только от многогранников Ньютона. Другие части связаны с вычислением с коэффициентами, и это вычисление абсолютно явное. Никаких систем решать не надо, а надо только делить один многочлен на другой — что-то вроде того. Надо делать совершенно явные операции.

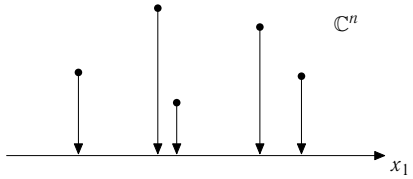
Такая же вещь есть не только на плоскости, но и в любой размерности. Оказывается, что если многогранники Ньютона расположены достаточно

общим образом друг относительно друга (позже я дам точное определение), то можно сделать всё то же самое. Оказывается, что такие системы бесконечно проще, чем рассматриваемые обычно системы.

На плоскости обычно рассматривают систему из уравнения степени n и уравнения степени m, многоугольники Ньютона у которых выглядят так, как показано на рис. 6. У этих многоугольников есть параллельные стороны. Поэтому если вы берете обычные системы, то для них наша формула неприменима. И правда, для них всё достаточно сложно.



Р и с. 6. Два многоугольника Ньютона



Р и с. 7. Проекция на ось x_1

Что конкретно будет выражать формула? Пусть есть n уравнений $P_1 = 0, \ldots, P_n = 0$ от n неизвестных, и пусть есть ещё какой-то многочлен Q от n переменных. Формула будет явно выражать сумму $\sum Q(a)$, где суммирование ведётся по всем $a \in (\mathbb{C}^n)^*$, для которых $P_1(a) = \ldots = P_n(a) = 0$.

Прежде чем выписывать эту формулу, зададимся вопросом, зачем вообще она нужна. Представьте себе, что мы умеем вычислять такую сумму для любого многочлена Q. Если мы это умеем, то мы можем сделать всё остальное: можно и исключить неизвестные, и найти вещественные корни и т. д. Давайте, например, исключим неизвестные. Пусть есть конечное множество точек $A \subset \mathbb{C}^n$. Предположим, что мы знаем $\sum\limits_{a \in A} Q(a)$. Как получить уравнение на x_1? Спроектируем все точки множества A на ось x_1 (рис. 7). Мы хотим написать уравнение, корнями которого являются все получившиеся проекции. Сначала научимся суммировать полином одной переменной по проекции множества A. Возьмём функцию Q, которая зависит только от x_1. Если Q зависит только от x_1, то её значения в точке и в проекции одинаковы. Поэтому суммирование функции Q по проекциям и суммирование по точкам множества A — это одно и то же. По точкам множества A мы умеем суммировать. Давайте просуммируем такие функции. Сначала просуммируем функцию 1. Тогда мы узнаем число точек N. Затем просуммируем функции x_1, x_1^2, …, x_1^{N-1}. Так мы получим основные симметрические функции Ньютона от точек проекции множества A. По ним выписываются коэффициенты многочлена, корнями которого являются точки проекции множества A. Мы исключили все неизвестные, кроме одного.

Примерно таким же способом (только надо рассматривать квадратичные формы, их сигнатуры и т. д.) можно узнать число вещественных точек в множестве A, если A инвариантно относительно комплексного сопряжения. Можно найти число точек в любой области. Если вы знаете сумму значений любого полинома Q по всем точкам конечного множества A, вы знаете всё про множество A.

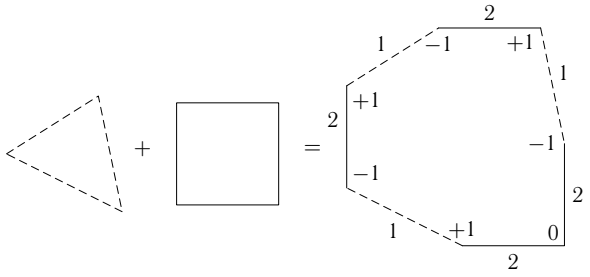
Комбинаторный коэффициент

Чтобы написать формулу для суммы значений полинома, мне понадобятся две величины. Одна из них — геометрическая величина (точнее, много однотипных геометрических величин, которые характеризуют взаимное расположение многогранников). Она самая забавная, и о ней я буду специально говорить. Она называется *комбинаторный коэффициент*. Другая величина алгебраическая, та самая, которая зависит от всех коэффициентов и получается делением. Формула, как я и обещал, будет состоять из двух частей. Одна часть геометрическая, другая алгебраическая. Формула имеет следующий вид:

$$\sum_{P_1(a)=\ldots=P_n(a)=0} Q(a) = \sum_{A\in \mathrm{vert}\ \Delta} C_A\ \mathrm{res}_A\ \omega.$$

Суммирование в правой части формулы ведётся по вершинам некоторого многогранника Δ. Многогранник Δ нужно взять равным $\Delta_1 + \ldots + \Delta_n$, т. е. многогранник Δ — это сумма по Минковскому многогранников Ньютона всех уравнений системы.

Нарисуем, например, на плоскости сумму по Минковскому треугольника и квадрата (рис. 8). Получается многоугольник, стороны которого — это все стороны треугольника и квадрата, расположенные в таком порядке, чтобы получился выпуклый многоугольник. С точностью до параллельного переноса сумма по Минковскому двух многоугольников на

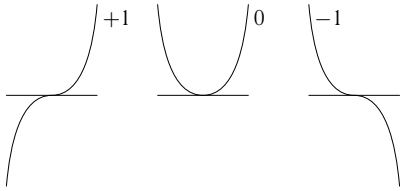


Р и с. 8. Сумма по Минковскому

плоскости — это многоугольник, стороны которого — это все стороны первого многоугольника и все стороны второго.

Во-первых, в каждой вершине A многогранника Δ возникнет своё целое число C_A — геометрическая часть формулы. Во-вторых, возникнет арифметическое число $\operatorname{res}_A \omega$. Нужно вычислить целое число C_A, которое характеризует расположение многогранников. Нужно провести некую операцию деления и вычислить число $\operatorname{res}_A \omega$. В этой операции деления фигурируют коэффициенты всех уравнений. Потом нужно сложить числа $\operatorname{res}_A \omega$, взятые с коэффициентами C_A, по всем вершинам A многогранника Δ, и получится ответ. Так выглядит эта формула.

Теперь мне нужно определить комбинаторный коэффициент C_A. Я начну с определения в плоском случае, потому что плоский случай проще всех остальных. Оказывается, что в плоском случае комбинаторный коэффициент принимает всего три значения: -1, 0 и $+1$. Я замечу, что в одномерном случае локальный индекс пересечения, скажем, прямой и графика функции тоже может принимать те же самые значения -1, 0 и $+1$ (рис. 9). А в многомерном случае индекс может быть принимать любое целое значение. В многомерном случае комбинаторный коэффициент тоже может быть любым целым числом.



Р и с. 9. Комбинаторный коэффициент в плоском случае

В плоском случае определение комбинаторного коэффициента C_A таково. Возьмём сумму многоугольников Ньютона, и будем её обходить против часовой стрелки. Если в вершине A сторона, пришедшая из первого многоугольника, меняется на сторону, пришедшую из второго многоугольника, то комбинаторный коэффициент C_A равен $+1$; если сторона, пришедшая из второго многоугольника, меняется на сторону, пришедшую из первого, то комбинаторный коэффициент в этой вершине равен -1; если обе стороны, прилегающие к вершине A, пришли из одного и того же многоугольника, то комбинаторный коэффициент C_A вершине равен 0 (рис. 8).

В многомерном случае прежде всего надо определить понятие общего набора многогранников. Давайте проанализируем определение в двумерном случае. Пусть у двух многоугольников нет параллельных сторон

с одинаковым направлением внешних нормалей. Это означает следующее. Возьмём произвольный ненулевой ковектор: $\xi \in (\mathbb{R}^2)^*$, $\xi \neq 0$. Ковектор — это линейная функция. Будем искать её максимум на наших многоугольниках. Я утверждаю, что хотя бы в одном из многоугольников максимум будет в вершине. Он случайно может достигаться на стороне в одном многоугольнике. Но тогда во втором многоугольнике на стороне он достигаться не может, иначе у этих многоугольников были бы параллельные стороны с одинаковым направлением внешних нормалей. Итак, этот набор многоугольников обладает следующим свойством: для любого ковектора $\xi \neq 0$ существует номер i, для которого максимум $\max\limits_{x \in \Delta_i}(\xi, x)$ на i-м многоугольнике достигается лишь в вершине. Скажем, что n многогранников в n-мерном пространстве расположены общим образом относительно друг друга, если любая ненулевая линейная функция хотя бы в одном из них достигает максимума в вершине. Это свойство выполняется в случае общего положения. Если n многогранников не такие, то, чуть-чуть повернув их, вы добьетесь того, что это свойство будет выполнено.

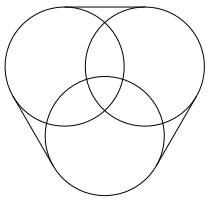
Теперь я определяю комбинаторный коэффициент в вершине. Это будет локальная степень некоторого отображения. Построим отображение границы многогранника $\Delta = \Delta_1 + \ldots + \Delta_n$ в границу положительного ортанта $F\colon \partial\Delta \to \partial\mathbb{R}^n_+ \subset \mathbb{R}^n$. Граница многогранника — это, в сущности, многообразие. Негладкое, конечно, но многообразие. Граница положительного ортанта — это тоже многообразие; тоже негладкое. Я построю такое отображение F, что в нуль перейдут только вершины многогранника. Другими словами, прообраз нуля при этом отображении состоит в точности из вершин. Тогда в каждом прообразе есть понятие локальной степени отображения. Эта локальная степень отображения и будет комбинаторным коэффициентом, соответствующим вершине. Причём отображение я построю не одно; отображений я построю много. Но все они будут гомотопны в классе отображений, у которых прообраз нуля состоит из вершин, и локальная степень у них у всех будет одинаковая. Неважно, какое из них взять.

Положительный ортант $\mathbb{R}^n_+$ лежит в $\mathbb{R}^n$. Поэтому отображение в $\mathbb{R}^n_+$ можно задать, задав n функций $F_1, \ldots, F_n$. Функция F_i определяется так. Вспомним, что $\Delta = \Delta_1 + \ldots + \Delta_n$. Это означает, что каждая грань Γ многогранника Δ есть сумма граней: $\Gamma = \Gamma_1 + \ldots + \Gamma_n$, где Γ_i — грань многогранника Δ_i. Я утверждаю, что среди граней Γ_i обязательно есть вершины. Действительно, возьмём линейную функцию ξ, которая достигает максимума в точности на грани Γ. Тогда на первом многограннике она достигает максимума в точности на грани Γ_1, на втором — на грани Γ_2,

на последнем — на грани Γ_n. Но, как мы знаем, для каждой ненулевой линейной функции ξ найдётся многогранник, в котором максимум достигается в вершине. Итак, некоторые грани Γ_i — вершины. Положим $F_i(\Gamma) > 0$, если Γ_i — не вершина, и положим $F_i(\Gamma) = 0$, если Γ_i — вершина. Такие функции существуют. Действительно, если мы зафиксируем номер i, то набор тех граней Γ, в которые Γ_i входит вершиной, будет замкнутым множеством. Мы всегда можем построить непрерывную функцию, которая на замкнутом множестве равна нулю и положительна на дополнении.

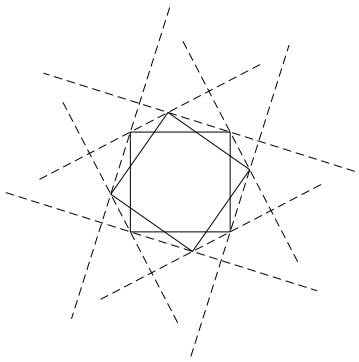
Возьмём любой набор функций $F_1, \ldots, F_n$, удовлетворяющих указанному свойству. Покажем, что отображение $F = (F_1, \ldots, F_n)$ переводит границу многогранника Δ в границу ортанта. Действительно, во-первых, все функции неотрицательные, во-вторых, в каждой точке одна из них обязательно обращается в нуль. Итак, у нас возникает отображение границы многогранника в границу ортанта. В каждой вершине A есть локальная степень отображения C_A. Это и есть комбинаторный коэффициент.

Оказывается, у чисел C_A есть совершенно другое определение. Возьмём вершину A многогранника Δ. Около одной вершины любой многогранник выглядит как конус. В этот конус после параллельного переноса попадают конусы многогранников Δ_i около вершин A_i, где $A = A_1 + \ldots + A_n$. Эту картинку нужно проективизировать. Например, если речь идёт о трёхмерном многограннике, то проективизированная картинка плоская. На этой плоскости есть три выпуклые фигуры — это то, как выглядят три выпуклых многогранника Δ_i в окрестностях вершин A_i, параллельно перенесённых в точку A (рис. 10). Проективизация около вершины A многогранника $\Delta = \Delta_1 + \Delta_2 + \Delta_3$ будет выпуклой оболочкой этих трёх фигур.



Р и с. 10. Проективизированная картинка

В $(n+1)$-мерном случае получаем $n+1$ выпуклое тело в n-мерном пространстве. Предположим, что эти тела достаточно общие. Я хочу сопоставить каждому такому набору выпуклых тел целое число, комбинаторный коэффициент C_A. Выпуклые тела занумерованы (упорядочены). Ограничение по общности будет такое. Например, на плоскости у двух выпуклых областей может быть общая опорная прямая (общая касательная), но у трёх областей сразу общая опорная прямая может быть только случайно; если они находятся в общем положении, то общей опорной прямой у них нет. В n-мерном случае мы будем говорить, что $n+1$ выпуклых тел общим образом расположены в n-мерном пространстве, если у них нет общей опорной гиперплоскости. Для таких наборов я определю целое число. Определение будет немножко странное, но вполне в духе



Р и с. 11. Опорные прямые

степени. Оно будет несимметрично. Тела будут играть разную роль. Пусть есть выпуклые тела $\Delta_1, \ldots, \Delta_n, \Delta_{n+1}$. Последнее тело Δ_{n+1} отложим в сторону. Тогда останется n тел. У n тел может быть общая опорная гиперплоскость; их даже может быть очень много, не обязательно одна. Например, если вы возьмёте квадрат, повернете его относительно центра, то у получившейся пары квадратов общих опорных прямых будет много (рис 11).

Итак, берем n выпуклых тел в n-мерном пространстве и берем у них какую-то опорную гиперплоскость. Точнее говоря, нас интересуют только такие опорные гиперплоскости, что одно из полупространств, на которые эта гиперплоскость делит $\mathbb{R}^n$, содержит все $n+1$ тел. Давайте будем считать, что наши тела удовлетворяют немного более сильному свойству общности, чем я предположил с самого начала. Предположим, что общая опорная гиперплоскость пересекает каждое из n тел по одной точке и будем считать, что эти n точек образуют вершины $(n-1)$-мерного симплекса. Наша гиперплоскость ориентирована: она была границей полупространства, содержащего $n+1$ тело, а граница ориентированного многообразия ориентирована. Итак, наша гиперплоскость ориентирована, и в ней есть симплекс, вершины которого занумерованы. Порядок вершин симплекса тоже задаёт ориентацию. Если эти две ориентации совпадают, то будем считать вклад этого опорного полупространства со знаком плюс, а если не совпадают — со знаком минус. Посчитаем все опорные полуплоскости описанного вида, с учётом приписанных им знаков. Получится целое число. Это целое число и есть комбинаторный коэффициент для $n+1$ тела, удовлетворяющего немного более сильному свойству общности.

Пусть есть $n+1$ тело, общие в слабом смысле, т. е. у них нет общей опорной гиперплоскости. Давайте их слегка пошевелим, чтобы у них попрежнему не было общей опорной гиперплоскости и чтобы полученные после шевеления тела были в общем положении в сильном смысле слова. Для полученных тел определён комбинаторный коэффициент. Разные шевеления могут приводить к очень разным картинкам. Но получающийся комбинаторный коэффициент не зависит от способа шевеления. Он и называется комбинаторным коэффициентом исходного набора тел. Как при определении степени отображения: вы можете взять непрерывное отображение, приблизить его гладким, взять точку, посчитать число прообразов

с учётом знака якобиана. При разных непрерывных аппроксимациях могут получаться очень разные картинки, но число прообразов, посчитанных со знаком, не меняется.

Я думаю, что я рассказал, что такое комбинаторный коэффициент. Видно, что это действительно геометрический инвариант, который никакого отношения к коэффициентам уравнений не имеет, а зависит именно от того, как многогранники расположены. И видно, что этот инвариант довольно грубый и устойчивый относительно малых шевелений.

Число $\mathrm{res}_A\ \omega$

Давайте теперь разбираться с числом $\mathrm{res}_A\ \omega$. Это число, как можно догадаться из формулы, будет вычетом некоторой дифференциальной формы ω. Сейчас я напишу эту дифференциальную форму и расскажу, о каком вычете идёт речь. Этот вычет будет комплексным числом, которое будет получаться явным делением; в его вычислении будут участвовать все коэффициенты всех уравнений.

Прежде чем переходить к n-мерному случаю, разберём совсем понятный 1-мерный случай. Пусть есть рациональная дифференциальная форма $\frac{Q}{P}\frac{dz}{z}$, где P и Q — многочлены от одной переменной z. Я хочу раскладывать рациональную функцию $\frac{Q}{P}$ в ряд Лорана по переменной z; не по переменной $z-a$, а именно по переменной z. Давайте посмотрим, сколько таких рядов. У нас есть знаменатель P, у знаменателя есть нули. Рассмотрим на комплексной плоскости окружности с центром в нуле, проходящие через нули знаменателя. Функция $\frac{Q}{P}$ раскладывается в ряд Лорана по переменной z в любом кольце между соседними окружностями. (Любая аналитическая функция, регулярная в кольце, раскладывается в этом кольце в ряд Лорана). Число таких колец на единицу меньше числа нулей у полинома P. Вообще говоря, ряд Лорана бесконечен в обе стороны. Но есть два ряда, конечных в одну из сторон. Это ряды Лорана в точке 0 и в точке ∞. Эти ряды особенно приятные. Они-то нам и нужны.

Любой коэффициент в этих рядах находится чисто алгебраически, делением. Я сразу расскажу об алгоритме нахождения коэффициентов ряда в многомерной ситуации. В одномерной ситуации многогранник Ньютона полинома P от одной переменной — отрезок. У этого отрезка есть два конца. Замечательных областей для формы $\frac{Q}{P}\frac{dz}{z}$ тоже две. Они связаны с вершинами многогранника Ньютона знаменателя: ясно, что младшая степень полинома P отвечает за точку 0, а старшая степень отвечает за

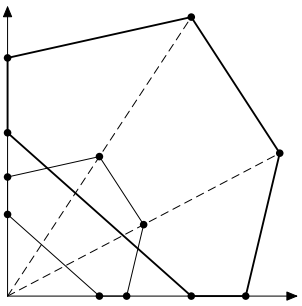
точку ∞. Как это будет выглядеть в многомерном случае? В многомерном случае полином Лорана P будет иметь свой многогранник Ньютона. Оказывается, что если в многомерном пространстве мы возьмём форму

$$\frac{Q}{P}\frac{dz_1}{z_1}\wedge\ldots\wedge\frac{dz_n}{z_n},$$

то с каждой вершиной многогранника Ньютона полинома Лорана P будет связан свой ряд Лорана по переменным $z_1, \ldots, z_n$. Мономы, входящие в этот ряд, принадлежат некоторому выпуклому конусу. Суммирование будет идти не по всем целым точкам в пространстве, а только по целым точкам, лежащим в конусе. Этот конус будет, по существу, тем конусом, который связан с вершиной многогранника (около каждой вершины многогранник совпадает с конусом).

Есть ещё и другие ряды Лорана по степеням $z_1, \ldots, z_n$. Они тоже очень забавные геометрически. Правда, с нашей формулой они никак не связаны. Зато они тесно связаны с тем, что называется *амёба*. Амёбу придумали Виро, Гельфанд, Зелевинский, Капранов, Михалкин.

Я напишу ряд Лорана, связанный с вершиной. Вы увидите, что он получается чистым делением. Пусть у нас есть рациональная функция $\frac{Q}{P}$, и мы хотим разложить её в ряд Лорана около вершины многогранника Ньютона полинома Лорана P. Давайте сначала рассмотрим самый простой случай, к которому всё потом сведётся. Пусть точка O является вершиной многогранника $\Delta(P)$, пусть свободный член полинома P равен единице и мы хотим написать ряд Лорана функции $\frac{1}{P}$, соответствующий нулевой вершине. Положим $P = 1 - \widetilde{P}$, где $\widetilde{P}$ — полином Лорана, не имеющий свободного члена. Так и хочется написать $\frac{1}{1-\widetilde{P}} = 1 + \widetilde{P} + \widetilde{P}^2 + \ldots$ Давайте так и сделаем. Что получится? Многогранник Ньютона для P содержит точку 0. Когда мы определяем $\widetilde{P}$, единичку у P мы забрали, поэтому многогранник Ньютона для $\widetilde{P}$ не содержит точку 0. Многогранник Ньютона для $\widetilde{P}^2$ такой же, как и для $\widetilde{P}$, но в два раза больше (рис. 12); для $\widetilde{P}^3$ — в три раза больше и т. д. Как вы видите, многогранник Ньютона для $\widetilde{P}^k$ всё дальше и дальше отходит от точки 0. Чтобы узнать коэффициент при любом конкретном мономе, нам придётся считать лишь конечное число степеней $\widetilde{P}^k$ полинома Лорана $\widetilde{P}$. Это вполне конечная процедура.



Р и с. 12. Многогранники Ньютона для $\widetilde{P}$ и для $\widetilde{P}^2$

Пусть теперь $P = c_a z^a + \ldots$, точка a является вершиной многогранника $\Delta(P)$, и мы хотим написать ряд Лорана функции $\frac{1}{P}$, соответствующий вершине $A = a$. Этот ряд по определению равен умноженному на $c_a z^a$ ряду Лорана функции $\frac{1}{c_a^{-1} z^{-a} P}$, соответствующему нулевой вершине многогранника $\Delta(c_a^{-1} z^{-a} P)$.

Если ряд для P^{-1} написан, то умножить его на полином Лорана Q уже несложно. Итак, если у вас есть рациональная функция $\frac{Q}{P}$, то в каждой вершине многогранника $\Delta(P)$ она раскладывается в ряд Лорана. Это вычисление явное. В каждой вершине нам нужен будет только свободный член. Число $\operatorname{res}_A \omega$ для формы $\omega = \frac{Q}{P} \frac{dz_1}{z_1} \wedge \ldots \wedge \frac{dz_n}{z_n}$ по определению равно свободному члену ряда Лорана функции $\frac{Q}{P}$ в вершине A многогранника $\Delta(P)$.

Давайте немножко обобщим нашу задачу. Будем считать не сумму значений полиномов по корням, а сумму значений так называемых *вычетов Гротендика*. Эти вычеты Гротендика в частном случае будут тем, что нам надо (значениями функций), а вообще-то это немножко более общие вещи. Сейчас я определю, что такое вычет Гротендика, и сформулирую нашу с Ольгой Гельфонд теорему в полном объёме. Пусть есть гиперповерхность, которая представлена как объединение n гиперповерхностей: $P = P_1 \cdot \ldots \cdot P_n = 0$. Тогда около каждого решения a системы уравнений $P_1 = \ldots = P_n = 0$ определён цикл, который лежит в дополнении к гиперповерхности. Он определяется так: это лежащая в малой окрестности решения a компонента связности многообразия, определённого уравнениями $\|P_1\| = \varepsilon_1, \ldots, \|P_n\| = \varepsilon_n$, где числа ε_i достаточно малы и достаточно общи. Все такие многообразия между собой гомологичны в дополнении к гиперповерхности. Определение вычета Гротендика таково. Пусть есть n-форма ω, особенности которой лежат на гиперповерхностях $P_1 = 0, \ldots, P_n = 0$. Интеграл $\left(\frac{1}{2\pi i}\right)^n \int \omega$ по такому циклу называется вычетом Гротендика. Этот вычет — прямое обобщение вычета Коши: если берётся многочлен P от одной переменной, то $P = 0$ — это точка, компонента связности множества $P = \varepsilon$, лежащая около точки — это кривая, обходящая вокруг этой точки. В 1-мерном случае особая точка обходится по кривой, в n-мерном пространстве, если уравнения трансверсальны, то по тору, а если не трансверсальны — по подмногообразию, которое задаётся уравнениями $\|P_1\| = \varepsilon_1, \ldots, \|P_n\| = \varepsilon_n$.

Теперь я сформулирую нашу теорему.

Теорема 1. *Пусть в* $(\mathbb{C}^*)^n$ *есть* n *уравнений* $P_1 = 0, \ldots, P_n = 0$, *многогранники Ньютона* $\Delta_1, \ldots, \Delta_n$ *которых расположены*

достаточно общим образом. Возьмём произвольный многочлен Лорана Q. Положим $P = P_1 \cdot \ldots \cdot P_n$. Тогда

$$\sum_{\text{по всем корням системы}} \operatorname{res} \frac{Q}{P} \frac{dz_1}{z_1} \wedge \ldots \wedge \frac{dz_n}{z_n} = \sum_{A \in \operatorname{vert}\Delta(P)} (-1)^n C_A \operatorname{res}_A \omega,$$

где слева суммируются вычеты Гротендика, а справа — числа $\operatorname{res}_A \omega$ *— свободные члены рядов Лорана рациональной функции* $\frac{Q}{P}$, *вычисленные в вершине A многогранника* $\Delta(P)$.

Кстати сказать, ряды Лорана в разных вершинах связаны, потому что сумма каждого ряда — данная рациональная функция. Но связаны они не сильнее, чем, скажем, ряды Лорана рациональной функции одной переменной в нуле и в бесконечности. Это абсолютно разные ряды. Аналитические продолжения их сумм, конечно, одинаковые, но они абсолютно не похожи друг на друга. С каждой вершиной у нас связан свой ряд; берём его свободный коэффициент, и вычисляем сумму.

В частном случае, если форму $\frac{Q}{P} \frac{dz_1}{z_1} \wedge \ldots \wedge \frac{dz_n}{z_n}$ подобрать правильно, это будет сумма значений многочлена. Как же её подобрать правильно? Если есть форма

$$L \frac{dP_1}{P_1} \wedge \ldots \wedge \frac{dP_n}{P_n},$$

то вычет такой формы в корне a системы $P_1 = \ldots = P_n = 0$ равен $\mu(a)L(a)$, где $\mu(a)$ — кратность корня a. Представьте себе, что я хочу суммировать многочлен L по корням. Это то же самое, что суммировать вычеты формы

$$\frac{dP_1}{P_1} \wedge \ldots \wedge \frac{dP_n}{P_n} = \frac{L \det\left(\frac{\partial P}{\partial z}\right) z_1 \cdot \ldots \cdot z_n}{P_1 \cdot \ldots \cdot P_n} \frac{dz_1}{z_1} \wedge \ldots \wedge \frac{dz_n}{z_n}.$$

Задачу суммирования значений функций мы свели к задаче суммирования вычетов Гротендика, которую мы умеем решать.

Частные случаи

Первый частный случай — суммирование единицы. С одной стороны, мы должны получить число корней. С другой стороны, получается довольно странное выражение $\sum (-1)^n C_A \det(A_1, \ldots, A_n)$, которое никак не похоже на смешанный объём. Это выражение устроено следующим образом. В сумме Минковского $\Delta = \Delta_1 + \ldots + \Delta_n$ каждая вершина A помнит, суммой каких вершин она является. $A = A_1 + \ldots + A_n$. Поэтому с каждой вершиной A суммы Минковского в n-мерном пространстве связано n векторов $A_1, \ldots, A_n$. Для них можно рассмотреть детерминант $\det(A_1, \ldots, A_n)$,

где A_1 — вершина в первом многограннике, …, A_n — вершина в n-м многограннике (имеются в виду те вершины, суммой которых является наша вершина A).

По теореме Бернштейна число корней равно $n!\,\mathrm{vol}(\Delta_1, \ldots, \Delta_n)$, поэтому

$$\sum_{A\in \mathrm{vert}\Delta(P)} (-1)^n C_A \det(A_1, \ldots, A_n) = n!\,\mathrm{vol}(\Delta_1, \ldots, \Delta_n).$$

Почему должно выполняться такое равенство, непонятно. Мы суммируем самые простые определители, правда, с хитрыми коэффициентами. В результате получается смешанный объём!

Мы доказали эту странную геометрическую теорему в том случае, когда наши многогранники целочисленные и достаточно общим образом расположены друг относительно друга. Я уже говорил, что теория многогранников Ньютона позволяет из алгебры получать геометрические следствия. После того как такая формула получилась для целочисленных многогранников, не нужно иметь слишком богатое воображение, чтобы догадаться, что такая же формула верна и для нецелочисленных многогранников. Но при помощи алгебры её никак не докажешь. Никакого отношения системы алгебраических уравнений к не целочисленным многогранникам не имеют; разве что с экспоненциальными суммами есть связь, но это уже не алгебра. Алгебра просто диктует справедливость следующего геометрического факта. Пусть в n-мерном пространстве задано n выпуклых многогранников, которые находятся в общем положении, т. е. для каждого ненулевого ковектора хотя бы в одном многограннике максимум достигается строго в вершине. Тогда для таких наборов многогранников должна быть верна формула

$$\sum (-1)^n C_A \det(A_1, \ldots, A_n) = n!\,\mathrm{vol}(\Delta_1, \ldots, \Delta_n).$$

Мы с Ольгой гипотетически считали, что эта формула верна. Ольга её даже в отдельных случаях доказала, но не в общем случае. Мне стоило большого труда доказать её геометрически. Я придумал доказательство, которое, с одной стороны, доказывает эту теорему для любых многогранников, а с другой стороны, очень близкие рассуждения (прямой перевод, как подстрочником, с русского на английский) в алгебраической геометрии дают доказательство теоремы Бернштейна.

Произведения корней

Теперь я расскажу о формуле для произведения корней. Я хочу перемножить корни n уравнений от n неизвестных в группе $(\mathbb{C}^*)^n$. Я снова буду

предполагать, что многогранники Ньютона $\Delta_1, \ldots, \Delta_n$ достаточно общи. Сейчас я напишу ответ, который очень похож на предыдущую формулу для суммы значений функции. (Я знаю два ответа. Один похож на эту формулу, а второй похож скорее на формулу Бернштейна.)

Произведение корней я просто посчитал не используя никакой техники. Правда, считал я долго. Делать что-либо без всякой техники, просто руками, всегда трудно.

Когда мы перемножим корни, мы получим точку в торе $(\mathbb{C}^*)^n$. Найти точку в группе $(\mathbb{C}^*)^n$ — это по существу то же самое, что вычислить на ней значение любого характера $\chi\colon (\mathbb{C}^*)^n \to \mathbb{C}$. Координаты $z_1, \ldots, z_n$ — это характеры. Если вычислить по формуле, которую я сейчас напишу, значения этих характеров, то мы получим координаты точки.

Итак пусть есть группа $(\mathbb{C}^*)^n$. Фиксируем характер, или, если хотите, фиксируем моном $z^k = z_1^{k_1} \ldots z_n^{k_n}$. Я хочу перемножить корни и на полученной точке вычислить этот моном. Получится число. Для числа легче писать формулу, чем для точки.

Ответ получается как произведение по вершинам многогранника $\Delta = \Delta_1 + \ldots + \Delta_n$ некоторых странных выражений. Вспомним, что каждая вершина A многогранника Δ есть сумма вершин $A_1, \ldots, A_n$ многогранников Ньютона $\Delta_1, \ldots, \Delta_n$, т. е. $A = A_1 + \ldots + A_n$. Ещё давайте вспомним, что в каждом многограннике Δ_i у нас был написан свой полином Лорана P_i; у этого полинома есть коэффициент, который отвечает соответствующей вершине A_i. Обозначим его $P_i(A_i)$. Здесь имеется в виду следующее. Первый многочлен Лорана есть сумма $P_1 = \sum P_1(a_k) z^k$ по целочисленным точкам $a_k \in \Delta_1$. Обозначим через $P_1(A_1)$ коэффициент $P_1(a_k)$ в соответствующей вершине A_1, т. е. $a_k = A_1$. Аналогичный смысл имеют $P_2(A_2), \ldots, P_n(A_n)$ — это коэффициенты полиномов Лорана P_i в соответствующей вершине A_i.

Число $[P_1(A_1), \ldots, P_n(A_n), k]_A$, которое я хочу сейчас определить, будет зависеть лишь от вершин $A_1, \ldots, A_n$, от коэффициентов $P_i(A_i)$ и от характера $z^k = z_1^{k_1} \ldots z_n^{k_n}$.

Данные, определяющие это число, можно записать в виде $(n+1) \times$ $\times (n+1)$ матрицы $M(A)$, где

$$M(A) = \begin{array}{|c|c|c|c|c|} \hline P_1(A_1) & P_2(A_2) & \ldots & P_n(A_n) & 1 \\ \hline A_1 & A_2 & \ldots & A_n & k \\ \hline \end{array}$$

В первой строке этой матрицы стоит набор чисел $P_1(A_1), \ldots, P_n(A_n), 1$ (единица в этой строке появилась из-за того, что характер z^k — это моном, взятый с единичным коэффициентом). Матрица $\widetilde{M}(A)$, полученная из матрицы $M(A)$ вычёркиванием первой строки, определяется так: столбец этой

матрицы с номером m, где $1 \leqslant m \leqslant n$, — вектор A_m. Столбец с номером $(n+1)$ — вектор k. По определению

$$[P_1(A_1), \ldots, P_n(A_n), k]_A = \\ = (-1)^{D(\widetilde{M}(A))} P_1(A_1)^{\det[A_2,\ldots,A_n,k]} \ldots P_n(A_n)^{(-1)^{n+1} \det[A_1,\ldots,A_{n-1}]}.$$

Здесь $D(\widetilde{M}(A))$ — своеобразный аналог определителя целочисленной $(n+1) \times n$ матрицы $\widetilde{M}(A)$, определённый по mod 2. Это единственная не равная тождественно нулю полилинейная кососимметричная функция от столбцов матрицы, принимающая значение в поле $\mathbb{Z}/2\mathbb{Z}$ и обладающая следующим свойством. Если ранг матрицы над полем $\mathbb{Z}/2\mathbb{Z}$, полученной приведением матрицы $\widetilde{M}(A)$ по модулю 2, меньше чем n, то $D(\widetilde{M}(A)) = 0$. Конечно, этот замечательный аналог определителя заслуживает более подробного обсуждения, но у меня сейчас нет на это времени.

Число $[P_1(A_1), \ldots, P_n(A_n), k]_A$ называется *символом Паршина*. Именно такие числа фигурируют в теории Паршина—Като.

Т е о р е м а 2. *Пусть в* $(\mathbb{C}^*)^n$ *есть система уравнений* $P_1 = \ldots = P_n = 0$, *многогранники Ньютона* $\Delta_1, \ldots, \Delta_n$ *которых расположены достаточно общим образом. Тогда*

$$\prod_{\text{по корням } a \text{ системы}} a^k = \prod_{A \in \operatorname{vert}\Delta} [P_1(A_1), \ldots, P_n(A_n), k]_A^{(-1)^n C_A}.$$

Другими словами, чтобы вычислить значение характера $z^k = z_1^{k_1} \ldots z_n^{k_n}$ на произведении корней системы уравнений $P_1 = \ldots = P_n = 0$ с многогранниками Ньютона $\Delta_1, \ldots, \Delta_n$, надо по всем вершинам A многогранника $\Delta = \Delta_1 + \ldots + \Delta_n$ перемножить символ $[P_1(A_1), \ldots, P_n(A_n), k]_A$, возведённый в степень $(-1)^n C_A$, где C_A — комбинаторный коэффициент в вершине A.

Формула для произведения корней очень похожа на формулу

$$\sum_{A \in \operatorname{vert}\Delta} (-1)^n C_A \det(A_1, \ldots, A_n)$$

для числа корней такой системы. И, вообще, формула для произведения корней очень похожа на формулу для суммы вычетов Гротендика (см. теорему 1). Почему? Существует ли единое доказательство этих формул? Оказалось, что ответ на этот вопрос положителен.

Пришла пора сказать о законах взаимности Паршина. Я начну с классического одномерного случая. Для алгебраических кривых известны две очень похожие формулы. Согласно первой формуле для всякой рациональной дифференциальной формы ω на любой алгебраической кривой Γ

выполняется равенство

$$\sum_{a\in\Gamma} \operatorname{res}_a \omega = 0.$$

Согласно второй формуле (называемой законом взаимности Вейля) для всякой пары рациональных функций f, g на любой алгебраической кривой Γ выполняется равенство

$$\prod_{a\in\Gamma} \{f,\ g\}_a = 1.$$

Здесь $\{f,\ g\}_a$ — так называемый символ пары функций f, g в точке a. Напомним определение этого символа. Пусть u — локальная координата около точки a, на кривой Γ и $u(a)=0$. Пусть $f = au^k + \dots$ и $g = bu^m + \dots$ — старшие члены разложения функций f и g в ряды Лорана по координате u. Тогда символ $\{f,\ g\}_a$ по определению равен $(-1)^{km}a^m b^{-k}$. Если в точке a ни f, ни g не обращаются в нуль, ни в бесконечность, то $\{f,\ g\}_a = 1$. Поэтому в произведении, фигурирующем в законе взаимности, лишь конечное число сомножителей отлично от 1 и это бесконечное произведение имеет смысл. Определение символа $\{f,\ g\}_a$ можно переписать в следующем виде. Рассмотрим 2×2 матрицу M, первая строка которой равна $(a,\ b)$, а вторая строка равна $(k,\ m)$.

Рассмотрим еще 2×1 матрицу $\widetilde{M}$, состоящую из строки $(k,\ m)$. Тогда $\{f,\ g\}_a = (-1)^{D(\widetilde{M})}a^m b^{-k}$, что полностью согласуется с определением символа Паршина.

Андре Вейль пришел к своему закону взаимности из теории чисел, рассматривая его как функциональный аналог квадратичного закона взаимности Гаусса.

Для Паршина эти классические теоремы были исходным пунктом. Он обобщил их на многомерный случай. В его теории вместо алгебраической кривой фигурирует произвольное n-мерное алгебраическое многообразие (над произвольным алгебраически замкнутым полем). Для произвольной рациональной формы ω старшей степени на алгебраическом многообразии он определяет понятие вычета, и доказывает, что определенные суммы вычетов любой формы равны нулю. Для $n+1$ мероморфной функции на n-мерном алгебраическом многообразии Паршин определяет понятие символа и доказывает, что определенные произведения символов любых наборов из $n+1$ функции равны единице.

Итак, в теории Паршина есть результаты о суммах и о произведениях, очень похожие друг на друга. Разумеется, я надеялся, что наши теоремы о суммах по корням систем уравнений и о произведении корней систем уравнений должны вытекать из теории Паршина. Мой аспирант Ваня

Сопрунов доказал, что так оно и есть. Но мои надежды, пожалуй, не вполне оправдались — доказательство наших теорем от этого не упростились и ситуация с ними не стала более понятной. Скорее наоборот, думая об этом предмете, я очень упростил законы взаимности Паршина для поля комплексных чисел (отмечу, что Паршина, как и Вейля, больше интересовали числовые поля, а совсем не поле комплексных чисел).

Литература

[1] Гельфонд О. А., Хованский А. Г. Многогранники Ньютона и вычеты Гротендика // Доклады Академии Наук. 1996. № 3(350). С. 298—300.

[2] Gelfond O. A., Khovanskii A. G. Toric geometry and Grothendieck residues // Moscow Mathematical Journal. 2002. V. 2, № 1. P. 99—112.

[3] Khovanskii A. G. Newton polyhedrons, a new formula for mixed volume, product of roots of a system of equations // The Arnoldfest, Proceedings of a Conference in Honour of V. I. Arnold for his Sixtieth Birthday. Fields Institute Communications, V. 24. Amer. Math. Soc., 1999. P. 325—364.

14 декабря 2000 г.

А. Г. Хованский

ПРОБЛЕМА АРНОЛЬДА О ГИПЕРБОЛИЧЕСКИХ ПОВЕРХНОСТЯХ В ПРОЕКТИВНЫХ ПРОСТРАНСТВАХ

Я хочу рассказать о нашей совместной работе с Митей Новиковым [1]—[3], посвященной проблеме Арнольда. Проблему мы не решили. Я больше не в силах думать про эту задачу. Кое-что нам удалось сделать, особенно в аффинном случае. В проективном случае мы тоже доказали одну теорему, но наш результат никак не соответствует затраченным усилиям. Мне хочется рассказать о наших результатах и устроить рекламу проблеме Арнольда.

В $\mathbb{R}^n$ выделяются своей простотой выпуклые гиперповерхности. Они обладают тем свойством, что в каждой точке их вторая квадратичная форма невырожденная; около каждой точки гиперповерхности можно ввести систему координат, для которой одна из координатных гиперплоскостей касается гиперповерхности и одна из осей координат — нормаль к гиперповерхности, глядящая из области, ограниченной гиперповерхностью. В такой системе координат гиперповерхность локально выглядит как график функции, которая с точностью до членов меньшего порядка является отрицательно определённой квадратичной формой. Никаких других компактных гиперповерхностей с невырожденной второй квадратичной формой, кроме выпуклых, нет. Действительно, по условию гиперповерхность компактна. Поэтому она помещается внутри сферы достаточно большого радиуса. Возьмём минимальную сферу, которая содержит рассматриваемую гиперповерхность. Тогда в самой удалённой от центра сферы точке вся гиперповерхность лежит по одну сторону от сферы. В этой точке гиперповерхность выпукла. Поэтому не может быть компактной гиперповерхности в $\mathbb{R}^n$, которая была бы всюду гиперболична. Но в проективном пространстве такие гиперповерхности есть. Простейший пример — гиперболоид в $\mathbb{R}P^3$. Эта поверхность гиперболична в следующем смысле: в окрестности каждой точки поверхность не лежит по одну сторону от плоскости, касающейся поверхности в этой точке. Такая поверхность по одним направлениям выпукла, а по другим вогнута.

Про выпуклые гиперповерхности есть масса замечательных результатов; некоторые из них, например, теорема Хелли и теорема Браудера, нам сегодня понадобятся. Теория выпуклых гиперповерхностей встречается практически всюду. Она простая, многое в ней до конца сделано.

Как описать гиперповерхности в проективном пространстве $\mathbb{R}P^n$, у которых вторая квадратичная форма нигде не вырождена? В этом вопросе и заключается проблема Арнольда. Эту же проблему независимо от Арнольда поставил Громов [5]. Это совершенно открытая задача, про которую не известно практически ничего.

Арнольд конкретизировал свой вопрос. Перейдём к этому более конкретному вопросу Арнольда в простейшем случае. Пусть в $\mathbb{R}P^3$ есть связная поверхность, которая является границей области. Пусть известно, что вторая квадратичная форма этой поверхности в каждой точке невырождена и имеет один плюс и один минус. Примером такой поверхности может служить гиперболоид. Гиперболоид ограничивает область. Про эту область верно следующее: внутри неё находится прямая, и вне неё находится прямая. Более того, если мы возьмём одну точку на внутренней прямой и одну точку на внешней прямой и соединим эти точки отрезком, то этот отрезок пересекает гиперболоид ровно в одной точке. Очень похоже дело обстоит и для выпуклых гиперповерхностей. Для выпуклой гиперповерхности вторая квадратичная форма всюду отрицательна (если её рассматривать как график функции в направлении нормального вектора, глядящего из области, ограниченной гиперповерхностью); у неё есть $n-1$ отрицательных квадратов. И вне нашей выпуклой гиперповерхности существует $(n-1)$-мерное пространство. У квадратичной формы есть 0 положительных квадратов, и внутри поверхности находится 0-мерное пространство, т. е. точка. Если мы возьмём любой отрезок, соединяющий точку, лежащую внутри поверхности, с точкой $(n-1)$-мерного пространства, расположенного вне поверхности, то этот отрезок пересекает выпуклую гиперповерхность ровно один раз.

Пусть в $\mathbb{R}P^n$ задана гладкая связная гиперповерхность Γ, которая является границей области U. Пусть вторая квадратичная форма гиперповерхности Γ в каждой точке невырождена. Тогда сигнатура второй квадратичной формы от выбора точки не зависит. Пусть эта форма имеет k минусов и, соответственно, $n-1-k$ плюсов (размерность Γ равна $n-1$). Тогда согласно гипотезе Арнольда вне области U лежит проективное подпространство $\mathbb{R}P^k$, а внутри области U лежит проективное подпространство $\mathbb{R}P^{n-1-k}$. Более того, согласно гипотезе Арнольда эти подпространства можно выбрать так, что любой отрезок, их соединяющий, пересекает гиперповерхность Γ ровно в одной точке.

Гипотеза Арнольда верна для выпуклых гиперповерхностей в проективном пространстве. Пусть в $\mathbb{R}P^n$ есть гиперповерхность, у которой вторая квадратичная форма всюду отрицательна. Тогда есть классическая теорема о том, что эта гиперповерхность — выпуклая гиперповерхность в некотором аффинном пространстве. Точнее: существует гиперплоскость, не пересекающая эту гиперповерхность. Дополнение проективного пространства до гиперповерхности имеет структуру аффинного пространства. В этом аффинном пространстве рассматриваемая гиперповерхность является выпуклой в обычном смысле слова.

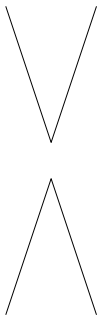
Арнольд кроме исходной проблемы сформулировал близкие задачи. Пусть гиперповерхность задана не в $\mathbb{R}P^n$, а в $\mathbb{R}^n$. Но зато известна её асимптотика ухода на бесконечность (я приведу примеры таких асимптотик). Вопрос остаётся тем же самым: верно ли, что внутри гиперповерхности находится аффинное подпространство нужной размерности и вне гиперповерхности находится аффинное подпространство нужной размерности?

Мы доказали (даже в многомерном случае), что для той конкретной асимптотики, про которую спрашивал Арнольд, ответ положительный. Внутри гиперповерхности с такой асимптотикой находится пространство нужной размерности и вне неё находится пространство дополнительной размерности. Про отрезок, соединяющий подпространства, мы получили лишь частичный результат. Кроме того, Арнольд спрашивал и про другие асимптотики. И мы с большим удивлением увидели, что для некоторых других асимптотик ответ отрицательный. После того как получилось очень простое доказательство задачи Арнольда для исходной асимптотики, мы были уверены, что гипотеза верна и для других асимптотик. Мы долго это доказывали, но в конце концов построили контрпримеры. Сначала мы эти контрпримеры даже клеили из бумаги, чтобы себя убедить в том, что они верны. Контрпример строить довольно трудно. Непонятно, как про данную гиперповерхность удостовериться, что она всюду гиперболическая, и почему внутри ограниченной ею области нет прямой. У нас были сомнения. Однако потом их удалось ликвидировать. Для аффинной проблемы все доказательства получились простыми.

Когда мы построили аффинные контрпримеры, мы попытались в этом же классе поверхностей построить проективный контрпример. Но в конце концов мы доказали, что в рассматриваемом классе проективных поверхностей контрпримеров нет. Эта задача свелась к массе частных случаев, многомерных, но всё же конечномерных. Мы постепенно разбирали случай за случаем. На первый случай мы потратили месяц. Потом дело пошло быстрее. В конце концов мы убедились, что во всех случаях

внутри поверхности есть прямая. Потом нам удалось резко сократить число случаев, подлежащих рассмотрению. В окончательном варианте их оказалось 6. Это удивительно трудное доказательство; оно никак не соответствует общности теоремы. Мы доказали лишь теорему о том, что внутри «выпукло-вогнутой» поверхности в $\mathbb{R}P^3$ есть прямая.

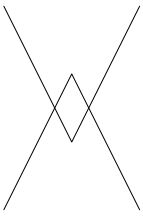
Теперь я более подробно расскажу про аффинную задачу. Я ограничусь пространством $\mathbb{R}^3$, хотя очень многое, из того, что я расскажу, дословно переносится в n-мерное пространство. Рассмотрим простейший конус $z^2 = x^2 + y^2$. Арнольд предложил в качестве начального шага в его проблеме рассмотреть следующую ситуацию. Пусть есть поверхность в $\mathbb{R}^3$, которая асимптотически на бесконечности совпадает с этим конусом (т. е. пусть замыкание поверхности в $\mathbb{R}P^3$ пересекается с замыканием конуса в $\mathbb{R}P^3$ по окружности, лежащей в бесконечно удалённой плоскости) и пусть эта поверхность гиперболична, т. е. имеет отрицательную гауссову кривизну. Верно ли, что внутри этой поверхности найдётся прямая? (То, что вне поверхности найдётся прямая, сомнений не вызывает.) Ответ: да, верно.



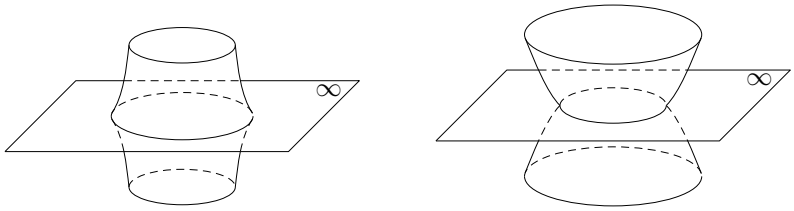
Теперь чуть-чуть изменим ситуацию. Рассмотрим такую же задачу, но полы конуса немного раздвинем (рис. 1). Пусть есть поверхность, которая всюду гиперболична и выходит на бесконечности на эти конусы. Верно ли, что внутри поверхности есть прямая? Ответ: нет, не всегда верно.

Р и с. 1. Раздвинутый конус

Справедлив следующий «принцип максимума»: ни в одном из этих случаев поверхность внутрь конуса не входит. Собственно говоря, всё дело в этом принципе максимума. Ось конуса и любая другая прямая, проходящая через вершину конуса и лежащая внутри конуса, по принципу максимума лежит и внутри поверхности. Для раздвинутого конуса принцип максимума не доказывает существования прямой внутри поверхности, и такой прямой и в самом деле может не быть. Если же полы конуса, наоборот, сдвинуть (рис. 2), то тот же самый принцип максимума показывает, что внутри прямая есть. Обе поверхности (раздвинутый конус и сдвинутый конус) на бесконечности имеют особенности, но особенности у них разные. Для раздвинутого конуса поверхность в каждой точке бесконечно удалённой окружности будет эллиптической (т. е. поверхность локально расположена по одну сторону от некоторой опорной плоскости), а для сдвинутого — гиперболической (т. е. она локально рассекается каждой плоскостью, проходящей через точку);



Р и с. 2. Сдвинутые конусы



Р и с. 3. Особенности на бесконечности

оба случая изображены на рис. 3. Поверхность становится особой, но гиперболичность в одном случае не нарушается, а в другом нарушается.

Теорема о существовании прямой в аффинном случае получилась достаточно общая, и доказывается она достаточно просто. Доказательство основано на следующей простой топологической лемме.

Л е м м а 1. *Пусть M_1^n — компактное ориентированное n-мерное многообразие, причём $\pi_1(M_1^n)=0$. Пусть есть другое компактное ориентированное n-мерное многообразие M_2^n, у которого, возможно, есть непустая граница ∂M_2^n. Рассмотрим отображение $\pi\colon M_2^n \to M_1^n$, которое обладает следующими свойствами*: 1) *π — локальный гомеоморфизм* (*это означает, что в окрестности каждой точки π является гомеоморфизмом на образ этой окрестности*); 2) *ограничение отображения π на каждую компоненту связности границы ∂M_2^n является вложением. Тогда π — вложение.*

В частности, в условиях леммы 1 образы разных компонент границы ∂M_2^n не могут пересекаться.

Самый простой случай — когда у многообразия M_2^n границы нет. Тогда π — локальный гомеоморфизм одного компактного многообразия на другое. Локальный гомеоморфизм компактных многообразий является накрытием. Но $\pi_1(M_1^n)=0$, значит, это накрытие однолистное, т. е. оно — гомеоморфизм. Так что, если у M_2^n границы нет, то и доказывать нечего.

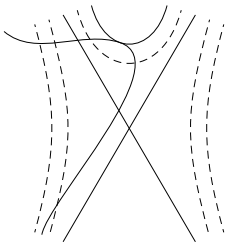
Предположим теперь, что у многообразия M_2^n граница есть. Мы сведём этот случай к случаю, когда границы нет. Возьмём компоненту границы ∂M_2^n и рассмотрим её образ. По условию ограничение отображения π на эту компоненту — вложение. В M_1^n возникает вложенная гиперповерхность. Она к тому же кориентирована: на компоненте границы выделена сторона, с которой локально лежит образ M_2^n. У нас возникла кориентированная гиперповерхность $\pi(\partial M_2^n)$ в M_1^n. Но кориентированная гиперповерхность задаёт 1-мерный класс когомологий, сопоставляющий каждой кривой число точек пересечения с этой гиперповерхностью, посчитанное с учётом знаков. Но ненулевого 1-мерного класса когомологий быть не может, потому что если нет π_1, то нет и H^1. Гиперповерхность $\pi(\partial M_2^n)$

задаёт нулевой класс когомологий. Это означает, что она затягивается плёнкой. Возьмём эту плёнку и сделаем с M_2^n следующую хирургию: приклеим к компоненте границы эту плёнку (дальше мы называем её шапкой) по тому отображению, которое здесь возникает; и так сделаем с каждой компонентой границы. Все компоненты границы M_2^n мы заклеили; получилось многообразие без границы. У этого многообразия есть отображение $\tilde{\pi}$ в M_1^n. Оно устроено следующим образом. Если мы находимся в старой части многообразия M_2^n, то отображение $\tilde{\pi}$ — это старое отображение π. Новая часть многообразия M_2^n состоит из шапок. Но шапки сами собой отображены в M_1^n тождественным отображением. Положим $\tilde{\pi}$ на шапках равным этому тождественному отображению. Отображение $\tilde{\pi}$ является локальным гомеоморфизмом. Итак, для построенного многообразия без границы есть отображение $\tilde{\pi}$, которое является локальным гомеоморфизмом. Такое отображение является взаимно однозначным. Значит, исходное отображение тоже взаимно однозначно.

Вернемся к нашей задаче. Возьмём нашу поверхность и рассмотрим отображение Гаусса поверхности в единичную сферу, которое сопоставляет точке x поверхности единичную нормаль $n(x)$ в точке x к данной поверхности. Это отображение удовлетворяет условиям леммы 1. Действительно, в обычной части поверхности это отображение является локальным гомеоморфизмом. Дело здесь вот в чём. Из-за того, что кривизна поверхности всё время отрицательна, мы знаем, что якобиан гауссова сферического отображения всё время отрицателен: согласно теореме Гаусса якобиан этого отображения совпадает с кривизной, а кривизна по условию отрицательна, потому что вторая квадратичная форма в каждой точке имеет один плюс и один минус. В частности, якобиан в каждой точке отличен от нуля. Рассматриваемая поверхность в $\mathbb{R}^3$ не является многообразием с краем. Но на бесконечности край у него есть: на бесконечности мы можем к ней приделать окружности, которые происходят из конуса. Ведь поверхность на бесконечности постепенно сливается с конусом. Окружность на проективной плоскости нужно раздвоить: одну окружность приклеить с одной стороны, а другую с другой. Мы компактифицировали исходное некомпактное многообразие без края. У нас получилось многообразие с краем, которое отображено на сферу.

Как устроено отображение на бесконечных окружностях, мы знаем, потому что на окружностях будет та же самая нормаль, что и у конуса; поверхность постепенно совпадает с конусом, и нормаль постепенно совпадает с нормалью к конусу. Поэтому на граничных окружностях отображение является гауссовым отображением для конуса; значит, это отображение является вложением.

В итоге получаем, что отображение Гаусса взаимно однозначно отображает нашу поверхность на полоску на сфере, поскольку выполнены все условия леммы 1. Дополнение сферы к полоске состоит из двух компонент, которые мы будем называть выброшенными шапками. Прообраз при отображении Гаусса выброшенных шапок по определению пуст. В частности, мы доказали, что у поверхности никаких ручек нет; эта поверхность — цилиндр. Между прочим, этот факт справедлив для раздвинутого конуса и для сдвинутого конуса. Эта часть рассуждений проходит без изменений для любой из этих асимптотик.



Р и с. 4. Поверхность внутри конуса

Я утверждаю, что поверхность не может войти внутрь конуса. Предположим, что поверхность вошла внутрь конуса (рис. 4). Рассмотрим наряду с конусом $z^2 = x^2 + y^2$ семейство слоёв $z^2 = x^2 + y^2 + c$, зависящих от параметра c. Выберем из этих слоёв последний, который пересекает наша поверхность. Последний слой можно выбрать потому, что наша поверхность на бесконечности приближается к конусу. Последний слой и наша поверхность касаются; в точке касания у них одинаковые отображения Гаусса (нормали одинаковые). Но отображение Гаусса переводит слой, зашедший внутрь конуса, как раз в выброшенную шапку, что противоречит тому, что мы знаем про отображение Гаусса данной поверхности. Значит, внутрь конуса наша поверхность зайти не может. Для раздвинутого и сдвинутого конуса поверхность тоже не заходит внутрь. Этот факт тоже справедлив для любой из этих асимптотик.

Если поверхность не заходит внутрь конуса, то внутри её существует прямая. Это утверждение справедливо и для сдвинутого конуса. Но раздвинутый конус состоит из двух компонент связности и не содержит внутри себя прямой линии.

Теперь я докажу, что каждый луч, проведённый из вершины конуса, пересекает нашу поверхность только один раз. Это примерно соответствует гипотезе Арнольда про отрезок, соединяющий точки внешнего и внутреннего подпространства. Но мы умеем это доказывать только для лучей, выходящих из вершины конуса.

Любая прямая пересекает выпуклую поверхность не более чем два раза. Про гиперболическую поверхность в проективном пространстве это заведомо неверно. Давайте для примера рассмотрим гиперболоид в проективном пространстве. Эта поверхность строго гиперболична и на ней лежит много прямых. Возьмём одну из этих прямых и немножко продеформируем нашу поверхность. Поверхность гиперболична, поэтому при

любой малой деформации она останется гиперболичной. Но чуть продеформировав гиперболоид, его можно превратить в «гармошку» около нашей прямой и сделать столько пересечений с нашей прямой, сколько мы хотим. Поэтому не приходится надеяться на то, что любая прямая пересекает гиперболическую поверхность не более чем в двух точках. Чтобы доказывать вторую часть гипотезы Арнольда, отрезки нужно выбирать удачно. Например, отрезок, один из концов которого является вершиной конуса, пересекает поверхность не более одного раза. Для доказательства этого я воспользуюсь теоремой Арнольда, а для доказательства теоремы Арнольда мне понадобится интеграл по эйлеровой характеристике. Интеграл по эйлеровой характеристике — красивая и полезная вещь, которая недостаточно популярна.

Теорему Арнольда Владимир Игоревич придумал, иногда размышляя над своей проблемой. Теорема Арнольда — это общий факт о пересечении прямой с поверхностью в проективном пространстве. Пусть в проективном пространстве дана поверхность Γ. Сопоставим каждой прямой l два числа: первое число — это число геометрически различных точек пересечения прямой с поверхностью $\#(l \cap \Gamma)$ (посчитанных без учёта кратностей); второе число — количество плоскостей (со знаком), проходящих через прямую l и касающихся поверхности Γ. Вообще говоря, во втором числе фигурирует не только знак, там фигурируют целочисленные кратности. Но в случае общего положения эти кратности равны ± 1. В случае общего положения поверхность либо локально лежит по одну сторону от касательной плоскости, как в случае сферы (это эллиптический случай), либо локально лежит по обе стороны от касательной плоскости, как в случае седла (это гиперболический случай). В эллиптическом случае касательная плоскость считается со знаком плюс, а в гиперболическом случае — со знаком минус. Теорема Арнольда утверждает, что для общей прямой l справедлива следующая формула:

$$\#(l \cap \Gamma) + (\text{посчитанное с учётом знаков число касательных плоскостей, содержащих прямую } l) = E(\Gamma),$$

где $E(\Gamma)$ — эйлерова характеристика поверхности Γ.

Рассмотрим, например, сферу. Если прямая пересекает сферу, то $\#(l \cap \Gamma) = 2$, а число касательных плоскостей, проходящих через прямую, равно нулю. Сумма равна 2, что как раз равно эйлеровой характеристике сферы. Если же прямая не пересекает сферу, то $\#(l \cap \Gamma) = 0$, а число касательных плоскостей равно 2 (оба касания эллиптические). Сумма снова равна эйлеровой характеристике сферы. Утверждается, что так будет и для любой поверхности и любой прямой в случае общего положения.

Совершенно замечательное доказательство этой теоремы придумал Олег Виро. Он построил исчисление (своеобразную теорию интегрирования), которое очень просто, но содержательно. Например, теорему Арнольда это исчисление сразу доказывает. Это так называемый *интеграл по эйлеровой характеристике*. Для построения теории меры нужно пользоваться какой-либо булевой алгеброй. Удобно пользоваться булевой алгеброй вещественных полуалгебраических множеств. Давайте для простоты считать, что все встречающиеся множества полуалгебраические. (На самом деле, во многих конкретных случаях этого предположения можно избежать.) Итак, мы берём $\mathbb{R}^n$ или $\mathbb{R}P^n$ и рассматриваем полуалгебраические множества, т. е. множества, заданные алгебраическими уравнениями и неравенствами и их конечные объединения. Для каждого замкнутого полуалгебраического множества X определена эйлерова характеристика $E(X)$. Для замкнутых полуалгебраических множеств X и Y справедливо равенство

$$E(X \cup Y) = E(X) + E(Y) - E(X \cap Y). \tag{1}$$

Этот факт можно доказывать либо «научно», с помощью теоремы Майера—Вьеториса, либо с помощью здравого смысла. Эйлерова характеристика — это альтернированная сумма количеств симплексов разных размерностей. Возьмём триангуляцию множества $X \cup Y$, которая уважает пересечение $X \cap Y$; для полуалгебраических множеств такая триангуляция всегда существует. Для такой триангуляции формула, аналогичная формуле (1), верна не только для эйлеровой характеристики, но и для числа симплексов любой размерности. Например, число вершин в объединении равно числу вершин в первом множестве плюс число вершин во втором множестве минус число вершин в пересечении первого и второго множества (они были посчитаны дважды). То же самое верно и для симплексов других размерностей. Поэтому формула (1) верна. Формула (1) показывает, что эйлерова характеристика ведёт себя так же, как мера: $\mu(X \cup Y) = \mu(X) + \mu(Y) - \mu(X \cap Y)$. Оказывается, что эйлерова характеристика единственным образом продолжается на все полуалгебраические множества с выполнением свойств меры. Возникает конечно-аддитивная мера на пространстве полуалгебраических множеств. Нужно сказать, что для произвольного полуалгебраического множества мера, которую мы получим, не обязательно будет равна эйлеровой характеристике этого множества.

Эта мера — очень простой объект. Например, «эйлерова характеристика» открытого круга равна 1: в открытом круге есть одна 2-мерная клетка и нет ни 1-мерных клеток, ни 0-мерных. Если полуалгебраическое

множество разрезано на открытые полуалгебраические симплексы, то для вычисления меры этого полуалгебраического множества нужно просто посчитать число симплексов и взять их альтернированную сумму. Эта мера — самое наивное определение эйлеровой характеристики.

Итак, в пространстве полуалгебраических множеств есть конечно-аддитивная мера. По этой мере можно интегрировать. Интегрируемая функция — это такая функция, которая принимает конечное число значений и каждая линия уровня которой есть полуалгебраическое множество. Для интегрируемой функции f можно определить интеграл по эйлеровой характеристике: $\int f\,dE = \sum_{\lambda} E(f^{-1}(\lambda))$. Мы получаем интегральное исчисление. Оно проще обычного интегрального исчисления, но в нём верны некоторые ключевые теоремы обычного интегрального исчисления. Например, в нём есть теорема Фубини, согласно которой функцию можно интегрировать послойно. Вот её полная формулировка: Пусть $\pi\colon X \to Y$ — полуалгебраическое отображение полуалгебраического множества X в полуалгебраическое множество Y и $f\colon X \to \mathbb{R}$ — интегрируемая функция. Тогда $\int\limits_X f dE = \int\limits_Y \left(\int\limits_{\pi^{-1}(Y)} f\,dE \right) dE$. По определению интеграла по эйлеровой характеристике, интеграл характеристической функции замкнутого множества равен эйлеровой характеристике этого множества.

Это интегральное исчисление очень геометрично. Олег Виро для этого исчисления нашел аналог преобразования Радона. Оказывается, что теорема Арнольда есть прямое следствие формулы обращения преобразования Радона для интеграла по эйлеровой характеристике.

Пусть есть проективное пространство $\mathbb{R}P^n$ и двойственное проективное пространство $(\mathbb{R}P^n)^*$. Точка двойственного пространства — это гиперплоскость исходного пространства. Пусть на $\mathbb{R}P^n$ задана интегрируемая функция f. Тогда можно построить функцию f^* на двойственном пространстве $(\mathbb{R}P^n)^*$, которая определена следующим образом: $f^*(h) = \int\limits_h f\,dE$. Если бы здесь стоял обычный интеграл, то это было бы обычное преобразование Радона. Теорема Радона даёт выражение для f^{**}. Как и в классическом случае, ответ зависит от чётности размерности.

Т е о р е м а (Радона для интеграла по эйлеровой характеристике).

$$f^{**} = \begin{cases} f, & \text{если } n \text{ нечётно;} \\ -f + \int f\,dE, & \text{если } n \text{ чётно.} \end{cases}$$

Классическая формула Радона далеко не очевидна. Формула Радона, найденная Виро, доказывается просто сложением и вычитанием. Однако из этой теоремы сразу вытекает теорема Арнольда.

Теорема Арнольда соответствует случаю $n=2$. Представьте себе, что мы проектируем поверхность Γ из некоторой точки $a \notin \Gamma$ на какую-нибудь плоскость L. Рассмотрим на этой плоскости L функцию $f(x)=\#(l_x \cap \Gamma)$, где l_x — прямая, проходящая через точку $x \in L$. Во-первых, $\int f(x)\,dE = E(\Gamma)$. Действительно, рассмотрим на нашей поверхности функцию, тождественно равную 1. По определению интеграл по эйлеровой характеристике от этой функции равен $E(\Gamma)$. Вычислим теперь этот интеграл по теореме Фубини. Мы должны взять прямую l_x и проинтегрировать нашу функцию по этой прямой; в результате получим число точек пересечения $\#(l_x \cap \Gamma)$. Если мы проинтегрируем это число по базе, то по теореме Фубини получим исходный интеграл. Соотношение

$$\#(l \cap \Gamma) + f^{**} = E(\Gamma)$$

следует из формулы Радона $f + f^{**} = \int f(x)\,dE$. Нам осталось посчитать f^{**}. Чтобы посчитать f^{**} в точке x, мы должны взять точку x, провести через неё гиперплоскость (в данном случае — прямую) h и посмотреть, чему равно $f^*(h)$. После этого нужно взять пучок прямых, проходящих через x, и проинтегрировать f^* по этому пучку. То, что получится, это и будет по определению $f^{**}(x)$. Посмотрим теперь, что такое $f^*(h)$. Мы берём прямую h, и по ней мы должны интегрировать нашу функцию. Это означает, что из каждой точки прямой h мы должны провести прямую, проходящую через точку a. При этом получается плоскость L_h, содержащая точку a и прямую h. Мы должны вычислить эйлерову характеристику пересечения нашей поверхности с плоскостью L_h. В общем положении пересечение плоскости и поверхности состоит из окружностей, а эйлерова характеристика окружности равна 0. Значит, функция f^* всегда равна нулю, за исключением прямых h, соответствующих плоскостям L_h, касающимся поверхности. Эллиптический случай касания соответствует тому, что одна из окружностей пересечения стала точкой. Её эйлерова характеристика равна 1. В гиперболическом случае у нас возникает восьмёрка. Эйлерова характеристика восьмёрки равна -1. Поэтому в эллиптическом случае мы берём $+1$, а в гиперболическом случае берём -1. Значит, интеграл функции f^* равен сумме по тем плоскостям, которые касаются поверхности Γ, причём в случае общего положения одни из них входят в сумму со знаком $+1$, а другие со знаком -1. Это и доказывает теорему Арнольда.

В формуле Радона—Виро спрятано много геометрии. Например, используя эту формулу для различных функций на комплексной алгебраической кривой, можно получить формулы Плюккера. Рассматривая различные функции на алгебраической поверхности, можно получить двумерные аналоги формул Плюккера и т. д.

Итак, мы доказали теорему Арнольда. Будем теперь её использовать. Наша поверхность гиперболическая, поэтому все точки касания будут входить со знаком минус. Согласно теореме Арнольда, число точек пересечения минус число точек касания равно эйлеровой характеристике поверхности. Для гиперболической поверхности эйлерова характеристика равна нулю. Если прямая не пересекает нашу поверхность, то через неё не проходит ни одна касательная плоскость. Действительно, если число точек пересечения равно 0, то по теореме Арнольда и число точек касания тоже равно 0. Возьмём прямую, лежащую внутри конуса. Мы раньше доказали, что такая прямая не пересекает нашу поверхность. Поэтому через эту прямую не проходит ни одна касательная плоскость.

Рассмотрим проекцию нашей поверхности из вершины конуса на сферу с центром в вершине конуса. Докажем, что это отображение тоже взаимно однозначное, как и отображение Гаусса. Другими словами, каждый луч, выходящий из вершины конуса, пересекает поверхность не более одного раза. Допустим, что наше отображение не взаимно однозначное. Тогда у него есть складка. Это означает, что существует касательная плоскость к нашей поверхности, проходящая через вершину конуса. Докажем, что это невозможно. Допустим, что касательная плоскость пересекает конус лишь в его вершине. Тогда её нормаль попадает в выброшенную шапочку. Но мы доказали, что этого не может быть. Теперь допустим, что касательная плоскость рассекает конус. Тогда внутри конуса найдётся прямая, лежащая в плоскости. Эта прямая не пересекает нашу поверхность в силу принципа максимума, но через неё проходит касательная плоскость к поверхности. Это противоречит формуле Арнольда. Поэтому проекция поверхности на сферу взаимно однозначна.

Я перечислил все положительные результаты, которые нам удалось доказать. Второй результат, о проекции из центра на сферу, основан на теореме Арнольда; он доказан только в трёхмерном пространстве. А первый результат, о том, что поверхность не может заходить внутрь конуса, доказан и в многомерном случае (доказательство в многомерном случае в точности то же самое, как в $\mathbb{R}^3$).

Теперь я расскажу об отрицательном аффинном результате, который состоит в построении гиперболической поверхности, которая асимптотически совпадает с раздвинутым конусом и не содержит внутри себя никакой прямой. Чтобы строить такие примеры, мы ввели специальный класс поверхностей. Он нужен и для дальнейшего: именно для такого класса поверхностей в проективном случае мы доказали, что внутри ограниченных ими областей есть прямая. Поверхность должна быть всюду гладкой и всюду гиперболической. Свойство гладкости не очень важно. Выпуклые

поверхности бывают негладкими, например, многогранники. От этого с ними ничуть не сложнее работать. Я сейчас хочу определить нечто в этом духе, но только гиперболическое. Рассмотрим область в пространстве $\mathbb{R}^3$, которая обладает следующими свойствами: 1) каждое горизонтальное сечение выпукло; 2) горизонтальное сечение вогнутым образом зависит от высоты, т. е. если вы возьмёте три горизонтальных сечения, то выпуклая оболочка двух крайних сечений содержит среднее сечение. Если бы эта поверхность была гладкой, то она не только была бы седловая в каждой точке, но по горизонтальному направлению она была бы всегда выпукла, а по вертикальному вогнута. С такими поверхностями легче разобраться. Они в каком-то смысле напоминают выпуклые поверхности.

Мы построим негладкую поверхность такого рода, внутри которой нет прямой. Когда такая поверхность построена, её легко подправить: сделать гладкой и гиперболической. Вот аналогичная ситуация. Если есть выпуклый многогранник, то его можно чуть-чуть раздуть, и он станет строго выпуклой гладкой поверхностью. Ясно, что этот процесс нуждается в точном описании, но совершенно очевидно, что это сделать можно. Это дело техники; здесь нет ничего ни удивительного, ни полезного.

Сделаю небольшое отступление и определю аналогичный класс гиперповерхностей в проективном пространстве. В аффинном пространстве есть семейство горизонтальных плоскостей. В проективном пространстве этому соответствует семейство плоскостей, проходящих через некоторую прямую l. Пусть в проективном пространстве есть поверхность, ограничивающая тело, которое обладает следующими свойствами: 1) каждое сечение тела, проходящее через фиксированную прямую l, выпукло; 2) если мы возьмём любую точку прямой l и спроектируем из неё тело, то получится дополнение до выпуклого множества. Тогда такую поверхность мы будем называть *l-выпукло-вогнутой*. Аналогичные гиперповерхности можно рассмотреть в многомерном проективном пространстве. Представьте себе, что в многомерном проективном пространстве есть плоскость L размерности k. Рассмотрим гиперповерхности, ограничивающие тела, которые обладают следующими свойствами: 1) сечение тела каждой плоскостью, проходящей через L и имеющей размерность $k+1$, выпукло; 2) если спроектировать тело из любого центра в L размерности $k-1$, то получится вогнутое множество, т. е. проекция является дополнением до выпуклого множества. Такие поверхности можно назвать *L-выпукло-вогнутыми*. Они выпуклы в одном направлении и вогнуты в другом. Никакой гладкости здесь не нужно. Гипотеза Арнольда о том, что внутри находится проективное пространство нужной размерности, может быть сформулирована и для таких L-выпукло-вогнутых поверхностей. Кстати

сказать, тела, ограниченные такими поверхностями, гораздо больше похожи на выпуклые. Для выпуклых тел неважно, гладкие у них границы или негладкие.

Проективное определение можно дать ещё так. В аффинном случае у нас было условие на три сечения: то сечение, которое лежит между двумя другими, должно содержаться в их выпуклой оболочке. В проективном пространстве бесконечно удалённое сечение тоже может лежать между двумя сечениями; нам нужно, чтобы только что сформулированное свойство выполнялось для любой тройки сечений в аффинном пространстве, полученном из проективного пространства вычёркиванием произвольной проективной плоскости.

Для аффинных поверхностей мы строим контрпример, а для проективных поверхностей такого рода нам с большим трудом удалось доказать, что в размерности 3 внутри ограниченной ими области всегда есть прямая.

Перейдем к конструкции контрпримера. В чём здесь трудность? По самой формулировке задачи картинку контрпримера нарисовать непросто. С какой стороны мы бы ни посмотрели на такую поверхность, она выглядит так, как будто внутри неё есть прямая: любая проекция поверхности содержит прямую.

Прежде чем строить пример, опишем некоторые хирургии выпукло-вогнутых множеств. Первая хирургия такая. Возьмём два горизонтальных сечения, удалим заключённую между ними часть области и заменим её на выпуклую оболочку этих двух сечений. Легко проверить что после такой хирургии выпукло-вогнутое тело останется выпукло-вогнутым. Вторая операция такая. Представьте себе, что у нас есть поверхность, её проекция выглядит так, как для выпукло-вогнутой поверхности, т. е. дополнение до проекции является объединением двух неограниченных выпуклых областей. Тогда если каждое горизонтальное сечение заменить его выпуклой оболочкой, то в результате получится выпукло-вогнутое множество.

Конструкция примера основана на построении удивительного выпукло-вогнутого множества, которое вовсе не тело, а полоска. Среди выпукло-вогнутых тел есть полоски, для которых любое горизонтальное сечение — отрезок с центром на фиксированной вертикальной прямой. Я сначала построю такую полоску, внутри которой есть эта фиксированная вертикальная прямая. А потом эту полоску пошевелю так, что никакой прямой уже в ней не останется.

Оказывается, что можно построить полоску так, что с какой бы стороны на неё ни посмотреть, она будет гиперболической. Если немножко подумать, то это довольно удивительно. И это тесно связано с линейными дифференциальными уравнениями второго порядка. Такие поверхности

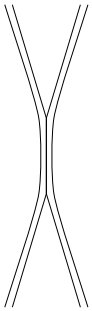
очень жёсткие. Я сейчас опишу все гладкие полоски. Конечно, существуют и негладкие полоски такого рода. Отметим, что на полоску можно посмотреть сбоку так, чтобы один из составляющих её горизонтальных отрезков выглядел, как точка.

Пусть z — высота, $x(z)$ и $y(z)$ — координаты одного из концов отрезка на высоте z. Вектор-функция $\boldsymbol{u}(z) = (x(z), y(z))$ полностью определяет полоску. Оказывается, что нужные нам функции $\boldsymbol{u}$ устроены следующим образом. Рассмотрим линейное дифференциальное уравнение $\boldsymbol{u}'' = q(z)\boldsymbol{u}$, где $q(z) > 0$. Возьмём два его независимых решения $x(z)$ и $y(z)$. Вектор-функции $\boldsymbol{u}(z) = (x(z), y(z))$ соответствует выпукло-вогнутая полоска. Более того, всякую гладкую выпукло-вогнутую полоску можно получить таким способом. Действительно, если мы смотрим на полоску сбоку, то полоска выглядит, как область, симметричная относительно вертикальной прямой, негоризонтальные компоненты границы которой являются линейными комбинациями решений $x(z)$ и $y(z)$ нашего линейного дифференциального уравнения. Но линейная комбинация двух решений тоже является решением. У решения дифференциального уравнения $\boldsymbol{u}'' = q(z)\boldsymbol{u}$, где $q > 0$, вторая производная имеет тот же знак, что и решение. Это значит, что граница области по одну сторону от вертикальной прямой выпукла в одну сторону, а по другую — в другую, с какого бы бока мы на полоску ни смотрели.

Полученная выпукло-вогнутая полоска содержит фиксированную вертикальную прямую; легко проверить, что для любой функции $q > 0$ эта прямая ровно одна — никаких других прямых эта полоска не содержит. Теперь я хочу сделать из этой полоски выпукло-вогнутую область, не содержащую прямой. Сделаем хирургию первого типа. Возьмём два горизонтальных сечения. Эти сечения — отрезки. Их выпуклая оболочка является тетраэдром. Разбивая полоску серией горизонтальных сечений и делая хирургию первого типа, превратим полоску в область, состоящую из серии тетраэдров. Легко доказать, что другой прямой при этом не появится. Область содержит одну вертикальную прямую. У неё есть много горизонтальных сечений, являющихся отрезками с центрами на этой прямой. Гиперболичность сохраняется при малом шевелении. Я возьму один отрезок и немножко его подвину. Тогда прямая исчезнет, а гиперболичность останется. Итак, мы построили выпукло-вогнутую область, внутри которой нет прямой. Покажем, как построить такую область с заданной асимптотикой на бесконечности.

Пусть есть два раздвинутых конуса, и есть прямая, проходящая через вершины этих конусов (рис. 5). Вместо этой прямой вставим нашу полоску, настолько узкую, чтобы сбоку это выглядело, как на рисунке 5. То, что

получится, не будет выпукло-вогнутым, но с любого бока это выглядит гиперболично. Поэтому мы можем применить хирургию второго типа — заменить каждое сечение его выпуклой оболочкой. На этом построение примера закончено.



Р и с. 5. Раздвинутые конусы

Мы с Митей Новиковым были абсолютно уверены, что, немножко поработав, мы подправим нашу конструкцию на бесконечности и получим проективный контрпример. Мы долго мучились и в конце концов доказали, что каждое проективное l-выпукло-вогнутое тело содержит внутри себя проективную прямую. Доказательство получилось довольно тяжёлое и морально неправильное. В трёхмерном пространстве, разобрав море случаев, которые мы потом свели к шести основным случаям, мы что-то сделали. Но как быть даже с 4-мерным пространством, мы совершенно не представляем.

Несколько слов о доказательстве этой теоремы. Она формулируется следующим образом. Всякое l-выпукло-вогнутое множество в трёхмерном проективном пространстве содержит проективную прямую.

Рассмотрим l-выпукло-вогнутое множество как объединение выпуклых сечений, проходящих через прямую l. Нам нужно провести прямую через все эти сечения. Для этого согласно теореме Хелли *) достаточно доказать, что через любые 5 сечений проходит прямая. Действительно, рассмотрим все прямые в проективном пространстве, которые не пересекают прямую l. Они образуют 4-мерное аффинное пространство. (В этом аффинном пространстве прямую можно задать так: возьмём две плоскости, проходящие через прямую l, на каждой отметим по точке, не лежащей на прямой l, и через эти точки проведём прямую; так мы параметризуем все прямые, не пересекающие прямой l. Каждая из двух отмеченных точек лежит в аффинной плоскости, поэтому наше пространство прямых имеет естественную структуру 4-мерного аффинного пространства.) Рассмотрим теперь множество прямых из нашего пространства, пересекающих одно сечение. Разумеется, это будет множество будет выпуклым в нашем пространстве прямых: если две прямые пересекают выпуклое сечение, то их линейная комбинация тоже пересекает сечение. У нас есть семейство выпуклых множеств в 4-мерном пространстве прямых (каждому сечению соответствует одно множество семейства). Если мы хотим доказать, что все эти множества пересекаются, то согласно теореме Хелли достаточно

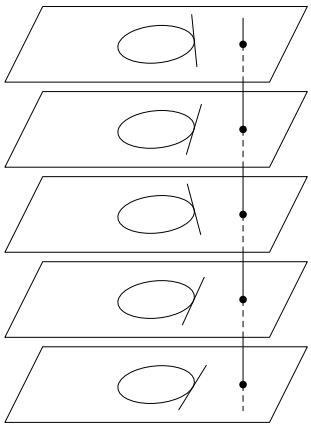
*) Теорема Хелли утверждает, что если в n-мерном аффинном пространстве имеется семейство компактных выпуклых множеств такое, что любые $n+1$ множеств из этого семейства имеют общую точку, то пересечение всех выпуклых множеств из этого семейства непусто.

доказать, что любые 5 из них пересекаются. Остаётся доказать, что через любые 5 сечений проходит прямая.

Мы умеем очень просто доказывать, что через любые 4 сечения проходит прямая. Чтобы это сделать, мне будет нужна ещё одна замечательная теорема из выпуклой геометрии, которая является обобщением теоремы Брауэра. Придумал её человек с очень похожей фамилией — Браудер. Теорема Брауэра такова. Пусть есть непрерывное отображение $f\colon B^n \to B^n$, где B^n — замкнутый шар. Тогда существует точка x, для которой $f(x) = x$. В теореме Браудера рассматриваются многозначные отображения шара в себя, которые каждой точке сопоставляют выпуклое множество, т. е. $f(x)$ — выпуклое множество в B^n. Предположим, что это отображение полунепрерывное сверху, т. е. если точке соответствует выпуклое множество V, то соседним точкам соответствуют выпуклые множества, лежащие в малой окрестности множества V; при изменении точки x множество $f(x)$ не может резко увеличиваться, но может резко уменьшиться. Если отображение однозначно, то полунепрерывность сверху означает непрерывность этого отображения. Теорема Браудера утверждает, что существует точка x, для которой $x \in f(x)$. Например, если f — однозначное отображение, то это в точности теорема Брауэра. Давайте я расскажу неправильное доказательство теоремы Браудера. Сопоставим каждому выпуклому телу $f(x)$ его центр тяжести. У нас возникает отображение шара в себя. По теореме Брауэра у него есть неподвижная точка. Из этого следует, что существует точка x, являющаяся центром тяжести тела $f(x)$. Это неправильное рассуждение. Неправильное оно потому, что центр тяжести выпуклого тела разрывно зависит от тела. Например, центр тяжести треугольника — точка пересечения медиан, а центр тяжести отрезка — его середина. Поэтому если треугольник схлопывается в отрезок, то центр тяжести скачет. Теорема Браудера нуждается в отдельном доказательстве. Но доказывается это примерно так же, как и теорема Брауэра. Эта теорема ещё называется теоремой Какутани (Sh. Kakutani. A generalization of Brouwer's fixed point theorem // Duke Math. J. 1941. V. 8. P. 457—458).

Докажем, что через любые 4 сечения проходит прямая. Представьте себе, что есть три последовательных горизонтальных сечения. Для выпукло-вогнутого тела среднее сечение находится в выпуклой оболочке двух других. Это означает, что через каждую точку среднего сечения можно провести прямую, пересекающую два других сечения. Я буду пользоваться этим свойством. Пусть теперь есть 4 сечения A, B, C, D. Построим многозначное отображения сечения B в себя следующим образом. Возьмём произвольную точку $x \in B$ и проведём через нее прямую, протыкающую сечения A и C. Затем через полученную точку пересечения

$y \in C$ проведём прямую, протыкающую сечения B и D. Получилась ломаная. Отобразим точку x в множество $f(x)$ пересечений всевозможных прямых такого вида с сечением в B. Для каждого x множество $f(x)$ выпукло. Согласно теореме Какутани—Браудера есть точка x, которая переходит в выпуклое множество $f(x)$. Для такой точки две прямые сливаются, и мы получаем прямую, пересекающую 4 сечения. Это доказательство годится для аффинных выпукло-вогнутых тел. Однако как мы знаем, в аффинном случае есть выпукло-вогнутая область, внутри которой нет прямой.



Р и с. 6. Пять сечений

Поэтому утверждение о существовании прямой, пересекающей 5 заданных сечений, принципиально более трудное. Несколько слов о нашем доказательстве этого утверждения. Пусть есть 5 сечений (рис. 6). Рассмотрим всевозможные невертикальные прямые. Каждой прямой сопоставим максимум из расстояний до данных сечений. Выберем ту прямую, для которой максимум самый маленький (чебышевскую прямую). Возникают 5 опорных полуплоскостей (см. рис. 6). Допустим, что существует прямая, проходящая через эти 5 полуплоскостей. Тогда я утверждаю, что исходная прямая была не чебышевская. Действительно, нашу прямую можно подвинуть к прямой, пересекающей 5 опорных плоскостей, сократив расстояния до всех сечений. Негоризонтальные прямые образуют аффинное пространство; поэтому прямую можно подвинуть. Итак, нужно доказать, что для 5 опорных полуплоскостей можно провести прямую, которая их протыкает. Это — отдельная задача, которую и надо решать. Эта задача зависит от большого, но конечного числа параметров: нужно задать высоты, на которых находятся горизонтальные плоскости, направления граничных прямых полуплоскостей и их расстояния до вертикальной прямой. Вырожденный случай в рассматриваемой задаче, когда граничные прямые каких-либо двух полуплоскостей параллельны, при помощи специальной двойственности (см. [2]) сводится к задаче о существовании прямой, пересекающей 4 сечения. Невырожденные случаи распадаются на 6 принципиально различных случаев. Они разбираются отдельно (см. [1]). В этом рассмотрении задача о существовании прямой, пересекающей 4 сечения, тоже играет ключевую роль.

Литература

[1] Khovanskii A., Novikov D. L-convex-concave sets in real projective space and L-duality // Moscow Mathematical Journal, 2003. V. 3, № 3. P. 1013—1037.

[2] Khovanskii A., Novikov D. L-convex-concave body in $\mathbb{RP}^3$ contains a line // Geometric and Functional Analysis (GAFA), 2003. V. 13. P. 1082—1118.

[3] Khovanskii A., Novikov D. On affine hypersurfaces with everywhere nondegenerate second quadratic form // Moscow Mathematical Journal, 2006. V. 3. № 1. P.135—152.

[4] Arnold V. I. Problem 1987-4 // Arnold Problems. — Springer—PHASIS, 2004.

[5] Громов М. Знак и геометрический смысл кривизны. Ижевск: РХД, 1999.

26 апреля 2001 г.

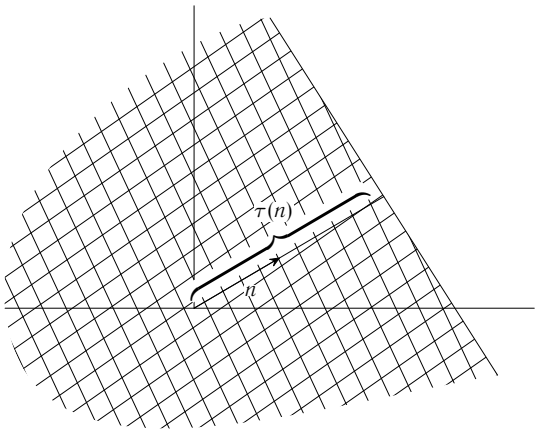
С. Б. Шлосман

ГЕОМЕТРИЧЕСКИЕ ВАРИАЦИОННЫЕ ЗАДАЧИ КОМБИНАТОРИКИ И СТАТФИЗИКИ

Я начну с очень простого вопроса, ответ на который очевиден. Он состоит в следующем. Представьте себе, что в воде плавает капля масла. Когда эта капля успокоится, она примет некую форму, и нам бы хотелось знать, какую. Вопрос не такой простой, если учитывать все существенные факторы. Но давайте поместим это в невесомость. Тогда понятно, что капля должна стать шариком, потому что сфера — это та поверхность, которая при заданном объёме имеет минимальную площадь. Этому учат в школьном курсе физики, говоря такие слова, что есть такая штука, как поверхностное натяжение, и всё происходит так, чтобы поверхность, которая получается в результате, обладала тем свойством, что она минимизирует поверхностную энергию, а поверхностное натяжение — это то, что надо проинтегрировать по поверхности, чтобы получить поверхностную энергию. Так что в простейшей ситуации ответ на наш вопрос — это просто сфера (или окружность, если ситуация плоская).

Вопрос становится более интересным, если поверхностное натяжение не постоянно, т. е. зависит от плоскости, вдоль которой оно считается. Поверхностное натяжение можно рассматривать как функцию на единичной сфере. Задача, которая имеет к нам непосредственное отношение, была решена 100 с небольшим лет назад в работе кристаллографа Вульфа, который работал в то время в Казани, в том же месте, где и Лобачевский. Работа, о которой я говорю, была опубликована в 1900 г. Эта задача формулируется следующим образом. Пусть задана функция поверхностного натяжения, которая зависит от направления. Как по ней узнать форму соответствующей капли или, может быть, кристалла, который образуется в соответствующей физической задаче? Математическая постановка вопроса следующая. Пусть задана функция поверхностного натяжения $\tau(\bar{n})$, где $\bar{n} \in S^k \subset \mathbb{R}^{k+1}$. Предполагается, что функция τ непрерывна, положительна и $\tau(\bar{n}) = \tau(-\bar{n})$. С каждой гиперповерхностью $M^k \subset \mathbb{R}^{k+1}$ можно связать число, про которое естественно думать, что это — поверхностная энергия:

$$w(M^k) = \int\limits_{M^k} \tau(\bar{n}_s)\, ds.$$



Р и с. 1. Конструкция Вульфа

Если мы хотим говорить не о произвольных поверхностях, а о каплях, то тогда мы должны сказать, что M^k — многообразие без края, для которого объём внутри M^k фиксирован, например, $\mathrm{vol}(M^k) = 1$. Пусть D_1 — семейство всех таких гиперповерхностей M^k. Мы хотим найти те гиперповерхности M^k из этого семейства, для которых минимум $\min\limits_{M \in D_1} w(M)$ равен $w(M^k)$. Я уже сказал, что если у нас случай изотропный, т. е. функция поверхностного натяжения не зависит от направления, то получается задача о теле с минимальной поверхностью и заданным объёмом. Ответ всем известен: это — шар. В общем случае геометрическая конструкция, которая даёт ответ на этот вопрос, была найдена Вульфом в работе 1900 г. Она состоит в следующем. У рассматриваемого функционала есть единственный минимум с точностью до сдвигов. Этот единственный минимум строится так: нужно рассмотреть множество

$$K = \{x \in \mathbb{R}^{k+1} \colon (\bar{n},\, x) \leqslant \tau(\bar{n})\ \forall \bar{n}\}.$$

Другими словами, нужно взять все лучи, выходящие из начала координат, на каждом луче отложить отрезок длины $\tau(\bar{n})$, через конец этого отрезка провести ортогональную гиперплоскость, взять соответствующее полупространство и рассмотреть пересечение всех этих полупространств (рис. 1). Искомая гиперповерхность — граница множества K с точностью до того, что у множества K объём может быть отличным от 1; тогда его нужно сжать или растянуть так, чтобы объём стал равным 1.

С помощью этой конструкции можно находить форму кристалла, если мы знаем поверхностное натяжение в среде как функцию направления. И если ситуация такая, что в веществе есть две фазы, между которыми

есть поверхностное натяжение, отличное от нуля, то там может происходить кристаллизация. И если она происходит, то форма кристалла, который возникает, получается с помощью этой конструкции. Когда у нас есть вещество, и оно находится в двух фазах, тогда это можно применять. Для таких фаз, как лёд и вода, на таком языке ответ получить трудно. Лучше говорить, например, о явлении спонтанной намагниченности, когда есть домены, в которых одна намагниченность, а окружены они областью, где намагниченность локально направлена в противоположную сторону. Две фазы должны быть одной физической природы, т. е. между ними должна быть некая симметрия. Тогда остаётся только поверхностное натяжение, никаких других эффектов нет. Только тогда конструкция Вульфа и применима.

Ответом всегда является выпуклая область. По поводу отдельных доменов нужно либо бесконечно долго ждать, пока они не объединятся в один домен, либо считать, что то, что вы видите, уже близко к равновесию, и тогда остальным вы уже не интересуетесь, т. е. домен растет настолько медленно, что можно забыть про всё остальное.

Конструкция Вульфа известна уже больше века. Статистическая механика, которая рассматривала строение вещества с микроскопической точки зрения (т. е. различая отдельные частицы), существовала сама по себе. И вопрос о том, можно ли увидеть в микроскопической статистической механике явление, которое соответствует той ситуации, которая здесь описана, по какой-то причине не рассматривался довольно долго. Соответствующие математические результаты, которые тем не менее имеются в строгой статистической механике, были получены сравнительно недавно. То, что я сейчас расскажу, составляет содержание книги, которая была опубликована 10 лет назад и была написана Добрушиным, Котецким и мной. Там мы описали ситуацию, которую я сейчас объясню. В этой ситуации конструкция Вульфа даёт ответ, хотя априори не видно, какое отношение могла бы иметь такая задача к той ситуации, про которую я сейчас расскажу.

Я хочу рассказать про задачу микроскопической статистической физики, в которой, на мой взгляд довольно удивительным образом, в конце концов возникает объект, имеющий непосредственное отношение к геометрической вариационной задаче, которую я объяснил. Самая хорошо изученная модель в статистической физике называется *моделью Изинга*. Я для простоты буду рассматривать случай размерности 2. Берётся двумерная целочисленная решётка $\mathbb{Z}^2$. В каждом узле $t \in \mathbb{Z}^2$ сажается $+1$ или -1. Тогда $\sigma = \{\sigma_t = \pm 1\}$ называется конфигурацией модели Изинга. Нужно представлять себе элементарные атомы, у которых есть магнитный

момент, и этот момент принимает всего два значения: вверх или вниз. Когда такая конфигурация задана, ей сопоставляется энергия

$$H(\sigma) = -\sum_{s,t\,:\,|s-t|=1} \sigma_s\sigma_t - h\sum_s \sigma_s.$$

Первый член соответствует ферромагнитному взаимодействию. Если ферромагнетик помещён в магнитное поле h, то возникает второй член.

Суммирование здесь происходит по всей решётке, поэтому эта сумма ни в каком смысле не сходится. Чтобы получилось нечто осмысленное, нужно сделать следующее. Нужно написать аналог этого выражения для конечной области V. Потом физически можно сказать просто, что нужно взять V очень большое, и на этом ограничиться. Математически это означает, что нужно сделать предельный переход, когда $V \to \infty$. Этот предельный переход называется *термодинамическим предельным переходом*. Только тогда возникают разные интересные эффекты в этой модели. А до этого мы имеем какую-то конечную систему, а для конечной системы трудно рассчитывать на то, что получится нечто обозримое или интересное.

На какой объект нужно смотреть? Нужно зафиксировать то, что находится вне коробки V. Пусть $\sigma_{\bar V}$ — фиксированная конфигурация вне коробки; внутри коробки конфигурация принимает все разрешённые значения. Запишем теперь энергию конфигурации в коробке V при условии, что конфигурация вне V фиксирована:

$$H(\sigma_V|\sigma_{\bar V}) = -\sum_{s\in V,\,t\in\mathbb{Z}^2\,:\,|s-t|=1} \sigma_s\sigma_t - h\sum_{s\in V} \sigma_s.$$

В первом члене суммирование происходит по тем рёбрам, которые либо целиком лежат внутри коробки V, либо один конец лежит в V, а другой снаружи.

Дальше нужно рассмотреть то, что называется *гиббсовским распределением*. Это — основной объект изучения статфизики. Гиббсовское распределение — это следующее распределение вероятностей на конфигурациях внутри коробки:

$$q^h_\beta(\sigma_V|\sigma_{\bar V}) = \exp\{-\beta H(\sigma|\sigma_{\bar V})\}/Z(\sigma_{\bar V}),$$

где

$$Z(\sigma_{\bar V}) = \sum_{\sigma_V} \exp\{-\beta H(\sigma|\sigma_{\bar V})\},$$

$\beta = 1/T$ называется *обратной температурой*.

Про это нужно думать так. Вы берете ферромагнетик. Если вы имеете возможность наблюдать за индивидуальными атомами, то неизвестно, что вы там увидите. Нужно мысленно представить себе много разных экземпляров одного и того же ферромагнетика. Если вы случайно выберете один из них и посмóтрите, какова же там конфигурация спинов частиц, то какие-то конфигурации вы будете видеть чаще, а какие-то реже. Вероятность, с которой вы будете видеть конфигурацию σ, есть энергия соответствующей системы взаимодействующих спинов или магнитных моментов, определенная выше.

Почему правильно смотреть именно на эту формулу, а не на какую-нибудь другую, я сейчас, к сожалению, объяснить не могу. Это — то, что называется распределением Гиббса. Об этом распределении вероятностей и будет некоторое время идти сегодня речь. Я хочу объяснить, какое отношение имеет это распределение вероятностей к геометрической конструкции, о которой я рассказал.

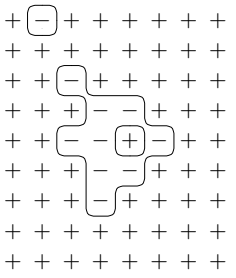
Но прежде я должен сказать ещё несколько слов о свойствах того объекта, который здесь определён. Несмотря на кажущуюся простоту этого распределения вероятностей и тех объектов, на которых оно сидит, у этой модели есть богатая внутренняя структура. Она состоит в следующем. Описываемая теория была задумана для того, чтобы изучать явление фазового перехода. Фазовый переход в этой модели происходит, и он происходит достаточным образом похоже на то, что происходит в реальных ферромагнетиках. Явление фазового перехода состоит в следующем. Нас интересует предел $\lim\limits_{V\to\infty} q^h_\beta(\quad|\sigma_{\bar V}) = \mu^h_{\beta,\bar\sigma}$. Здесь $\sigma_{\bar V} = \bar\sigma|_{\partial V}$. Оказывается, что семейство мер $\mu^h_{\beta,\bar\sigma}$ обладает следующими свойствами:

1. Если $h \neq 0$, то мера $\mu^h_{\beta,\bar\sigma}$ не зависит от $\bar\sigma$, т. е. если вы помещаете вещество в магнитное поле, то оно не реагирует на то, как устроена система далеко от того места, где вы производите наблюдения. Всё определяется тем, какое магнитное поле вы наложите.

2. Если $h = 0$, а $T > T_{\text{cr}}$, т. е. температура больше некоторого критического значения, то $\mu_{\beta,\bar\sigma}$ снова не зависит от $\bar\sigma$. Это ситуация высокотемпературной фазы. Когда температура высокая, тепловые флуктуации такие сильные, что вещество ни на что другое внимания не обращает, и взаимодействие между соседями слабое.

3. Если $h = 0$, а $T < T_{\text{cr}}$ (т. е. $\beta > \beta_{\text{cr}}$), то тогда у вещества возникают фазы. Предельная мера зависит от того, какое граничное условие мы возьмём. Например, если мы возьмём граничное условие $\bar\sigma \equiv +1$ и обозначим его $\oplus$, и возьмём граничное условие $\bar\sigma \equiv -1$ и обозначим его $\ominus$, то меры $\mu_{\beta,\oplus}$ и $\mu_{\beta,\ominus}$ различны. Это и означает, что в веществе происходит фазовый переход, или что в веществе имеется дальний порядок.

Математически различие мер $\mu_{\beta,\oplus}$ и $\mu_{\beta,\ominus}$ означает следующее. Если взять коробку, на границе написать все плюсы, рассмотреть такое распределение вероятностей (считая, что параметр β принимает достаточно большое значение) и перейти к пределу, то получится некая мера — предельное распределение вероятностей. Эта мера будет иной, чем если вместо всех плюсов поставить минусы. Для других выборов σ может так случиться, что предела не будет. Фазовый переход — это та ситуация, когда предельный объект (предельная мера) зависит от того, что стоит на границе области. Другими словами можно сказать, что в системе существует дальний порядок, а это — то же самое, что поведение внутри коробки зависит от того, что стоит на границе, как бы далеко сама граница при этом ни находилась. Это называется дальним порядком; система помнит граничные условия, из которых она возникла.

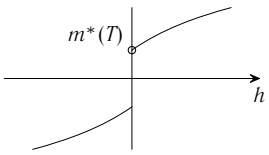


Р и с. 2. Плюс-фаза

На двумерной решётке у нас есть два распределения вероятностей: $\mu_\oplus$ и $\mu_\ominus$. Они называются плюс-фазой и минус-фазой. Если мы посмотрим на типичную конфигурацию плюс-фазы, то мы увидим много плюсов и кое-где будут встречаться островки минусов (рис. 2). Другая мера получается из этой меры заменой плюсов на минусы и минусов — на плюсы. Минус-фаза сосредоточена в основном на отрицательных конфигурациях, а плюс-фаза — на положительных. Это ровно та ситуация, в которой можно надеяться применить теорию, с которой я начал. Есть вещество, существующее в двух разных фазах. Эти фазы равноправны, так как одна переходит в другую под действием группы симметрии $\mathbb{Z}_2$. Теперь можно спросить, как выглядит капля одной фазы, если её заставить плавать внутри другой фазы.

Чтобы ответить на этот вопрос, сначала нужно его точно сформулировать. Что математически означает, что капля одной фазы плавает внутри другой фазы? Что нужно сделать, чтобы такая ситуация возникла?

Можно посчитать математическое ожидание какой-либо случайной величины по одной из этих мер. Например, математическое ожидание того, что происходит в начале координат. Случайная величина σ_0 принимает два значения. С какой-то вероятностью она принимает значение $+1$, с какой-то вероятностью она принимает значение -1. У неё есть среднее значение $m(h) = \langle\sigma_0\rangle_h$ при данном магнитном поле h. График функции $m(h)$ изображён на рис. 3. При $h = 0$ есть две разных меры. Если мы посчитаем средние значения для



Р и с. 3. Среднее значение σ_0

этих двух мер, то у нас получатся два разных числа. То, что эти меры разные, видно уже из того, что среднее по плюс-фазе равно $m^*(T) > 0$; эта величина называется *спонтанной намагниченностью*. Среднее по минус-фазе равно, соответственно, $-m^*(T)$.

В размерности 3 кроме этих двух предельных мер, есть ещё и другие. В размерности 2 тоже есть некоторые промежуточные меры, но все они являются линейными комбинациями мер $\mu_\oplus$ и $\mu_\ominus$. Грубо говоря, если взять граничное условие, которое содержит половину плюсов и половину минусов (например, в верхней полуплоскости взять плюсы, а в нижней — минусы) и перейти к термодинамическому пределу, то этот предел существует, и та мера, которая получится, есть полусумма мер $\mu_\oplus$ и $\mu_\ominus$. В размерности 3 ситуация более интересная. Если то же самое проделать для размерности 3, получится мера, которая через эти две не выражается. Она является ещё одной крайней точкой в множестве всех возможных значений этой меры. Но уже наличие двух разных пределов говорит о том, что мы имеем дело с ситуацией, где имеет место фазовый переход.

Плюс-фаза характеризуется тем, что доля плюс-значений составляет примерно $\frac{1+m^*(T)}{2}$. В противоположной минус-фазе доля плюс-значений составляет $\frac{1-m^*(T)}{2}$. Зная это, мы можем рассмотреть такой вопрос: «Что будет, если мы зафиксируем в нашей системе долю плюс-частиц?» Повторим всё то, что было сказано, но только теперь будем рассматривать те конфигурации σ_V, для которых $\sum\limits_{t\in V} \sigma_t / |V| = \rho$, где $-m^*(T) < \rho < +m^*(T)$.

Я повторю, что есть две фазы $\mu_\oplus$ и $\mu_\ominus$. Первая из них сосредоточена в основном на положительных конфигурациях, а вторая в основном на отрицательных. А я хочу наблюдать ситуацию, когда эти две фазы сосуществуют. Оказывается, что это происходит, если я дополнительно скажу, что я разрешаю не все конфигурации, а только те, у которых доля плюсов и доля минусов фиксированы, причём доля плюсов находится между $\frac{1+m^*(T)}{2}$ и $\frac{1-m^*(T)}{2}$. Этим ограничением я заставляю систему в какой-то части пространства быть в одной фазе, а в какой-то части быть в другой фазе. Во всяком случае, апостериори оказывается, что она ведёт себя именно таким образом.

Ситуация, когда мы рассматриваем формализм гиббсовских состояний и дополнительно фиксируем долю одного (и, соответственно, другого) сорта частиц, называется *каноническим ансамблем*. Раньше я говорил о ситуации, когда этого дополнительного ограничения нет. Такая ситуация называется *большим каноническим ансамблем*.

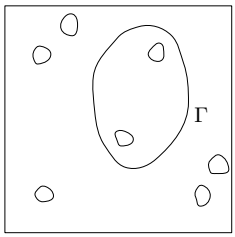
Вместо того чтобы описывать конфигурацию нашей системы, задавая в каждом узле, какое именно значение там принимается (плюс или минус), эквивалентным образом можно описывать нашу систему, задав так называемые контуры. А именно, возьмём какое-нибудь ребро. Если на одном его конце стоит плюс, а на другом минус, то мы проведём отрезок, который тоже имеет единичную длину, перпендикулярен этому ребру и делит его пополам. Если же на обоих концах ребра стоят одинаковые значения, то делать ничего не будем. Контуры — это линии, составляющие из этих отрезков. Если стереть конфигурацию, то по контурам её можно восстановить, если мы знаем, что снаружи (на границе области) стоят плюсы. А именно, мы должны рисовать плюсы, пока не дойдём до контура. После этого мы должны ставить минусы до следующего контура, потом опять плюсы и т. д. Так что есть взаимно однозначное соответствие между конфигурациями и контурами.

Если мы хотим стать на такую точку зрения, то мы должны выразить распределение вероятностей в терминах контуров. Оказывается, что это очень легко сделать в наиболее интересном для нас случае, когда $h=0$. А именно,

$$q_{V,\beta}(\{\gamma_i\}) = \frac{\exp\{-2\beta \sum\limits_i |\gamma_i|\}}{Z(V,\beta)}.$$

Здесь $\{\gamma_i\}$ — набор контуров, $|\gamma_i|$ — длина контура γ_i, $Z(V,\beta)$ — нормирующий множитель. Мы видим, что это распределение вероятностей в основном сосредоточено на конфигурациях, где мало контуров, потому что они экспоненциально сильно подавляются. При низкой температуре должно быть редкое поле маленьких контуров.

Это в самом деле верно, но только если мы говорим о случае большого канонического ансамбля. Если же мы говорим о случае, когда введено такое ограничение, что есть много плюс-частиц и много минус-частиц, то так быть не может. Оказывается, что если на эту систему глядеть с точки зрения контуров, то происходит в точности то явление, которое нам хочется наблюдать. Оно состоит в следующем. Есть распределение вероятностей; β — большое число (что соответствует тому, что температура низкая); кроме того, фиксирована доля плюсов и доля минусов, причём так, что выполнено неравенство $-m^*(T) < \rho < +m^*(T)$. Оказывается, что в этом случае типичная конфигурация системы выглядит так: есть один большой контур Γ_0, а кроме того, есть много маленьких контуров (рис. 4). Такая



Р и с. 4. Типичная конфигурация

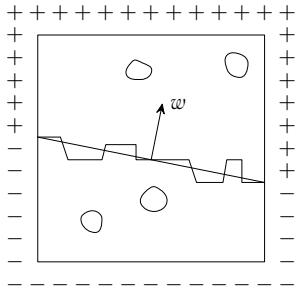
конфигурация наблюдается с вероятность, которая стремится к 1, когда мы делаем термодинамический предельный переход. Я повторяю, что это — типичная конфигурация для данного распределения вероятностей на контурах, когда значение параметра β велико.

Большой контур Γ_0 — это случайный полигон. Теорема состоит в том, что с вероятностью, стремящейся к 1, большой контур только один.

Т е о р е м а 1. *Случайный контур* Γ_0, *в пределе, когда* $V \to \infty$, *имеет асимптотически неслучайную предельную форму. Эта предельная форма — универсальная кривая, которая зависит от* β *и* ρ. *Эта кривая получается вышеописанной конструкцией Вульфа.*

Как вы помните, у конструкции Вульфа есть вход: нужно задать некую функцию τ. Я не буду давать точного определения поверхностного натяжения τ для модели Изинга. Скажу просто, что если есть модель Изинга и у нас фиксирована температура T, то определена функция $\tau^I_\beta(\bar{n})$, где $\beta = \frac{1}{T}$. (Верхний индекс I означает, что мы рассматриваем модель Изинга.)

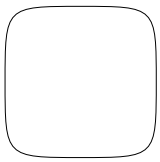
Для её построения нужно взять вектор $\bar{n}$, провести к нему нормальную гиперплоскость (в 2-мерном случае — прямую) и поставить на границе V плюсы сверху неё, а минусы снизу (рис. 5). В этом случае возникнет один длинный контур и много маленьких. Грубо говоря, статистика этого длинного контура и то направление, в котором мы заставляем его идти, наложив эти специальные граничные условия, приводит к определению функции поверхностного натяжения $\tau^I_\beta(\bar{n})$.



Р и с. 5. Поверхностное натяжение в модели Изинга

Функция поверхностного натяжения не зависит от ρ. Мы определяем функцию τ, решаем вариационную задачу и находим оптимальную форму. После этого нужно её масштабировать так, чтобы площадь внутри составляла бы $\frac{m-p}{2m}|V|$.

Теорема, которую можно доказать, состоит в следующем. Пусть $W^{I,\beta}$ — кривая, полученная для модели Изинга конструкцией Вульфа при заданной температуре T (рис. 6). Если мы возьмём большую коробку размером $N \times N$, то там, как я уже сказал, почти наверное окажется ровно один большой контур. Его размер порядка cN, где $c = c(\rho)$. Мы должны взять фигуру $W^{I,\beta}$, нарисовать её в подходящем масштабе, и сдвинуть её так, чтобы она была ближе всего к контуру Γ_0. Потом мы

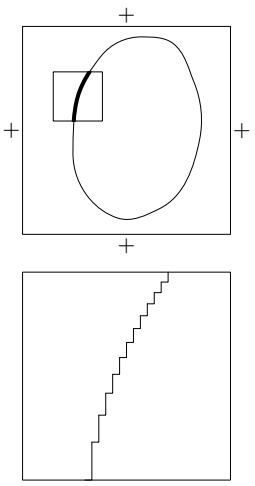


Р и с. 6. Кривая Вульфа

находим расстояние по Хаусдорфу между случайным контуром и неслучайной кривой, т. е. мы находим минимальный радиус окрестности кривой, содержащей контур. Пусть это расстояние равно Δ. Оказывается, что с вероятностью, стремящейся к 1, в термодинамическом пределе верно, что $\Delta \leqslant N^{2/3}$. Если же мы сделаем размер коробки единичным, то после предельного перехода мы даже не будем видеть, что контур Γ_0 случаен, потому что он отличается от неслучайной кривой на величину, которая будет стремиться к нулю, при $N \to \infty$.

Таким образом в модели Изинга возникает такой замечательный объект — неслучайная кривая, которая является асимптотической формой капли. Её хотелось бы назвать «кристаллом», но в двумерном случае это — гладкая (даже аналитическая) кривая; у неё нет плоских участков. Так что правильнее всё-таки сказать, что эта конструкция позволяет определить форму капли (а не кристалла) одной фазы внутри другой фазы, когда фазы сосуществуют.

Доказательство этой теоремы довольно замысловатое и длинное. В книге, про которую я говорил, доказывается эта теорема и только она; там ничего больше нет. Мы знаем, что образуется случайная капля; как угадать, какой она будет формы? Давайте возьмём кусочек нашей системы. Что мы увидим, когда поглядим в маленькое окошко (рис. 7)? Я утверждаю, что то, что мы увидим, будет выглядеть примерно так же, как на рис. 5. Если мы посмотрим под увеличением, то вместо непрерывной кривой мы увидим ступенчатую кривую. Дальше мы должны посчитать, какова стоимость того, чтобы у нашей системы под таким углом шла лестница. Эта стоимость и есть та величина, из которой предельным переходом получается поверхностное натяжение. Дальше мы смотрим на распределение вероятностей, которое у нас имеется. Я напомню, что это распределение имеет вид $e^{-H(\sigma_V)}$. А раз так, то типичной будет та конфигурация, для которой энергия близка к минимуму. Стало быть, энергия конфигурации должна быть такой, что если мы проинтегрируем локальные вклады всех кусочков и сложим их, то должно получиться нечто, имеющее отношение к минимуму уже глобального непрерывного функционала. Эту идею нужно как-то реализовать. Это весьма замысловатая и длительная процедура. Но по пути действительно решается буквально такая задача. Вы берете произвольную каплю,

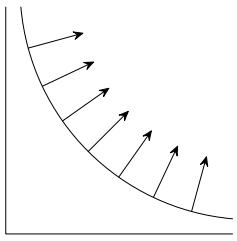


Р и с. 7. Маленький участок капли

и спрашиваете себя, сколько есть конфигураций, у которых большие контуры имеют примерно такую форму? Этот вопрос состоит в том, что вы рисуете ломаные; в каждом квадратике выбираете своё направление; считаете, какие же конфигурации дают вклад в множество всех разрешённых конфигураций, где плотность заданная. После этого вы спрашиваете, какова же та кривая, которая даёт максимальный вклад в множество всех конфигураций. Это — та конфигурация, у которой поверхностная энергия минимальна. Действительно получается, что можно доказать то, что физически хотелось бы доказать. Технически это занятие замысловатое и длительное. Частичным ответом на это будет то, что я хочу рассказать дальше. А именно, я хочу рассказать, каким образом похожие вопросы и похожие ответы возникают в комбинаторике. Там ситуация более обозрима.

Асимптотическая комбинаторика

Я хочу рассказать о том, о чём здесь уже рассказывал Анатолий Моисеевич Вершик. Я расскажу главным образом его результаты, которые я воспринимаю с несколько другой точки зрения, чем он. Начну я с другой вариационной задачи. Апостериори оказывается, что именно она возникает в некоторых задачах асимптотической комбинаторики. Задача, которую я хочу объяснить, похожа на задачу, с которой я начал. Эта задача состоит вот в чём. Раньше у нас была функция $\tau(\bar{n})$, $\bar{n} \in S^k$. Теперь рассмотрим функцию $\eta(\bar{n})$, $\bar{n} \in S^k \cap (\mathbb{R}^{k+1})_+ = \Delta^k$ (здесь $\mathbb{R}^{k+1}_+$ — положительный ортант). Функция η тоже положительная и непрерывная. Тогда тоже можно определить некий интегральный функционал на поверхностях. Но теперь функция задана не везде, поэтому и поверхности можно рассматривать не все, а только те, у которых нормальные вектора удовлетворяют указанному ограничению (рис. 8). Тогда поверхности G можно сопоставить число $v(G) = \int\limits_G \eta(\bar{n}_s)ds$. Чтобы такой интеграл был конечным, нужно дополнительно потребовать, чтобы $\eta(\bar{n}) \to 0$, когда $\bar{n} \to \partial\Delta^k$. Без этого условия интеграл будет расходиться. Вариационная задача, которую я хочу рассматривать, следующая. Я определю число $\mathrm{vol}(G)$ так: когда такая поверхность разделяет положительный ортант на две части, объёмом называется объём той части, которая прилегает к координатным плоскостям. Объём под кривой я снова фиксирую: $\mathrm{vol}(G) = 1$. Пусть $\mathcal{D}^1$ — множество всех поверхностей G, для которых $\mathrm{vol}(G) = 1$, причём для



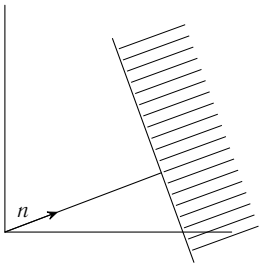
Р и с. 8.
Поверхность G

любой точки $s \in G$ вектор $\bar{n}_s$ принадлежит Δ^k. Меня интересует максимум $\max\limits_{G \in \mathcal{D}^1} v(G)$.

В некотором смысле эта задача отличается от той задачи Вульфа, потому что если рассматривать максимум функционала без такого рода ограничений, то он обязательно будет равен бесконечности. Поверхность может быть сильно изрезанной. Тогда объём под ней может быть конечным, а площадь поверхности может быть бесконечной. Но из-за того, что требуется, чтобы поверхность была монотонной, задача делается осмысленной. И оказывается, что максимум интеграла в этом семействе часто бывает конечным. Более того, снова есть геометрическая конструкция, которая строит максимизирующую поверхность. Она очень похожа на конструкцию Вульфа, с которой я начал. А именно, нужно рассмотреть множество

$$K = \{x : (x, \bar{n}) \geqslant \eta(\bar{n}) \ \forall \bar{n} \in \Delta^k\}.$$

Здесь знак неравенства обратный по сравнению с конструкцией Вульфа. Другими словами, снова нужно нарисовать луч, который идёт в направлении вектора $\bar{n}$, и снова нужно нарисовать нормальную гиперплоскость (рис. 9). Но теперь нужно взять внешнюю часть. Все эти внешние части нужно пересечь. Утверждение состоит в том, что поверхность, которая так получается, является максимизирующей для данной вариационной задачи. Правда, лишь в том случае, если площадь под ней конечна. Тогда её можно нормировать. А если площадь бесконечна, то тогда максимизирующей поверхности нет: максимум функционала на этом множестве поверхностей равен бесконечности.



Р и с. 9. Построение максимизирующей поверхности

Оказывается, что эта вариационная задача отвечает на некоторые вопросы асимптотической комбинаторики. Один из этих вопросов, который был поставлен и решён в работе Вершика и Керова, хорошо известен. Он состоит в следующем. Рассмотрим все диаграммы Юнга с N клетками, и рассмотрим на этом конечном множестве равномерное распределение вероятностей: вероятность каждой диаграммы обратна числу диаграмм Юнга с N клетками. Раз есть распределение вероятностей, то есть диаграммы, типичные с точки зрения этого распределения вероятностей. Можно ли утверждать, что у случайной (в смысле этого распределения вероятностей) диаграммы Юнга есть неслучайная предельная форма? В том же самом смысле, в каком неслучайная форма была у случайного контура в модели Изинга. Ответ на этот вопрос положителен. Он состоит

в том, что существует универсальная кривая

$$e^{-x\frac{\pi}{\sqrt{6}}} + e^{-y\frac{\pi}{\sqrt{6}}} = 1, \tag{1}$$

которая определяет асимптотическую форму большой случайной (в смысле этого распределения) диаграммы Юнга. А именно, если взять эту кривую, растянуть её в обоих направлениях на $\sqrt{N}$, нарисовать её вместе с диаграммой Юнга из N клеток и посчитать расстояние между этой неслучайной кривой и случайной лестницей, которая ограничивает диаграмму Юнга, то вероятность того, что расстояние не превосходит $\varepsilon N^{1/2}$, при любом $\varepsilon > 0$ стремится к 1 при $N \to \infty$.

Кривая (1) является решением вариационной задачи, о которой я только что говорил, для подходящей функции η. Можно догадаться какова должна быть функция η, и почему так получается, что типичная диаграмма имеет неслучайную форму, если задать себе следующий вопрос. Давайте возьмём какую-нибудь монотонную кривую $\gamma(x)$ (чтобы вообще были диаграммы, которые на неё похожи) и посчитаем, сколько же есть диаграмм, примерно такой формы.

Растянем γ в $\sqrt{N}$ раз по горизонтали и вертикали, и оценим, сколько есть диаграмм из N клеток, «похожих на γ». Для этого расставим точки $A_1, A_2, \ldots, A_k, \ldots$ на кривой $\sqrt{N}\gamma$, с расстоянием $\sim \sqrt[4]{N}$ между соседними. Ломаная Γ_N, составленная из отрезков $[A_k, A_{k+1}]$, приближает кривую $\sqrt{N}\gamma$, и представляет собой график убывающей функции. Пусть L_k — число нисходящих лестниц на клетчатой бумаге, соединяющих точки A_k и A_{k+1}. Тогда произведение $\prod\limits_k L_k$ и есть примерное число диаграмм, идущих вдоль $\sqrt{N}\gamma$. Числа L_k нетрудно вычислить, используя формулу Стирлинга. А именно,

$$L_k \approx \exp\left(|A_{k+1} - A_k|\eta(n_k)\right).$$

Здесь n_k — единичная нормаль к отрезку $[A_k, A_{k+1}]$, а

$$\eta(n) = -\left(n^{(1)} \ln \frac{n^{(1)}}{n^{(1)} + n^{(2)}} + n^{(2)} \ln \frac{n^{(2)}}{n^{(1)} + n^{(2)}}\right).$$

Тем самым мы получаем, что число диаграмм, идущих вдоль кривой $\sqrt{N}\gamma$ примерно равно

$$\exp\left\{\sqrt{N} \int\limits_0^\infty \eta\left(-\frac{\gamma'(x)}{\sqrt{(1+\gamma'(x))^2}}, \frac{1}{\sqrt{(1+\gamma'(x))^2}}\right)\right\} dx.$$

А кривая (1) как раз и максимизирует интеграл в последнем выражении среди всех кривых $\gamma(x) > 0$, монотонно убывающих по x и таких, что $\int \gamma(x)dx = 1$.

15 ноября 2001 г.

А. Н. Паршин

ЛОКАЛЬНЫЕ КОНСТРУКЦИИ В АЛГЕБРАИЧЕСКОЙ ГЕОМЕТРИИ

Мой доклад состоит из двух частей. Сначала общее введение про n-мерные локальные поля, это — старые вещи, но короткий очерк необходим. А во второй половине я расскажу совсем свежие вещи, посвящённые вполне конкретному вопросу.

Что такое n-мерные локальные поля и откуда они возникают? Если есть многообразие X размерности 1, то, что такое «локально» хорошо известно и никаких вопросов не возникает. Имеется область на комплексной плоскости, точка P и любую мероморфную функцию можно разложить в ряд Тейлора в окрестности P. В алгебраической геометрии, когда мы рассматриваем алгебраическую кривую C над произвольным полем k, понятия сходимости в окрестности точки нет и поэтому единственное, что можно и нужно сделать, это рассматривать формальные разложения. Иными словами, на кривой имеется такая диаграмма полей:

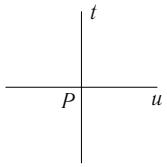
$$\begin{array}{c} K = k(C) \\ \big\downarrow \\ K_P \cong k((t_P)) \end{array}$$

Здесь $k(C)$ — поле рациональных функций на кривой C. Кривая может быть компактной, т. е. проективной, а может быть и аффинной. Если мы выберем какую-то точку P, то поле K вкладывается в пополнение K_P. Точке P отвечает нормирование $v_P\colon K^* \to \mathbb{Z}$ поля K (v_P — это порядок нуля или полюса в точке P). Поле K можно пополнить по этому нормированию и получить новое поле K_P, содержащее исходное поле K. Оно и называется *локальным* полем в точке P.

В данном случае для нас важно, что K_P — поле степенных рядов от некоторого параметра t_P, который мы можем выбирать разными способами, но который всегда связан с точкой P (для простоты мы считаем, что поле k алгебраически замкнуто). Параметр t_P чисто локальный; он ничего не знает о глобальном поведении кривой. Мы имеем конструкцию, которая существует, как говорят, в формальной окрестности точки на кривой.

Мы получили то, что является частным случаем локального поля, а точнее, *одномерным* локальным полем. Теперь я дам общее определение, а потом буду приводить результаты, конструкции и примеры, которые объясняют, почему это интересно.

Но сначала все-таки еще один пример. Я уже сказал, что локальные поля естественно возникают в случае размерности 1, для кривых. Когда имеется n-мерная ситуация, то, с обычной точки зрения, самое локальное, что может быть на многообразии, это какая-то окрестность точки. Пусть у нас даны комплексная плоскость $\mathbb{C}^2$ и точка $P \in \mathbb{C}^2$. Рассмотрим координатный крест с координатами u и t (рис. 1). Тогда функции на плоскости можно разлагать в степенные ряды от двух переменных. Точка зрения, которую я хочу здесь изложить, и привести аргументы в ее пользу, состоит в том, что уже эта конструкция не локальна. Она занимает промежуточное положение между локальным и глобальным подходом, иначе говоря, существует иерархия локальностей, в которой она занимает не самое низшее место.



Р и с. 1. Координатный крест

Для случая размерности 1 имеются два сорта объектов: глобальные поля, которые связаны со всей кривой, и локальные поля, которые связаны с точкой, и больше ничего нет. В случае же 2-мерном (и n-мерном), как мы увидим, имеются гораздо более разнообразные конструкции. Но посмотрим еще на привычную конструкцию — пополненное локальное кольцо поверхности X в точке P; если мы перейдем к формальной точке зрения, как и в предыдущем случае, то это будет не что иное, как кольцо $\hat{\mathcal{O}}_{X,P} = k[[u, t]]$ формальных степенных рядов Тейлора от двух переменных. Это обычное локальное понятие, которое используется и в алгебраической, и в аналитической геометрии.

Сейчас я введу определение *n-мерного локального поля*. Это прежде всего поле K и на нем имеется дополнительная структура, зависящая от числа n. Именно, в поле K выделено подкольцо $\mathcal{O}_K$, которое является полным кольцом дискретного нормирования. Раз таковое есть, то имеется максимальный идеал $\wp$ и поле вычетов $\mathcal{O}_K/\wp = \bar{K}$. Мы предполагаем, что это поле вычетов является $(n-1)$-мерным локальным полем. Кроме того, поле K должно быть полем отношений кольца $\mathcal{O}$, т. е. $K = \mathrm{Frac}(\mathcal{O})$.

В частности, 0-мерное локальное поле — это просто поле, без всякой дополнительной структуры.

Мы имеем здесь индуктивное обобщение ситуации, хорошо известной в одномерной коммутативной алгебре или алгебраической геометрии. В частности, если у нас имеется поверхность и мы хотим построить на

этой поверхности набор локальных объектов с ней связанных (локальных полей), то естественно возникает следующая диаграмма:

$$\begin{array}{ccc} K = k(X) & \longrightarrow & K_P \\ \downarrow & & \downarrow \\ K_C & \longrightarrow & K_{P,C} \end{array}$$

Самый глобальный объект — это поле $k(X)$ (поле рациональных функций на поверхности). Затем есть промежуточные «поля» (не все из них являются полями; некоторые будут кольцами). Мы рассматриваем флаг $X \supset C \ni P$, где C — кривая на поверхности X (дивизор), и на этой кривой выбрана точка $P \in C$. Поля K_C, отвечающие кривым C, действительно является полями. «Поля» K_P, отвечающие точкам, на самом деле являются только кольцами. Самое большое поле, которое все их содержит, — это поле $K_{P,C}$, отвечающее флагу $X \supset C \ni P$. Другими словами, ввиду того, что мы находится в размерности 2, у нас теперь имеется диаграмма длины не 1, а 2.

Как все эти поля определяются? Давайте я начну с самого локального объекта — поля $K_{P,C}$. Это — двумерное локальное поле. Точнее говоря, оно будет таковым (в смысле того определения, которое я дал), если дополнительно предположить, что P — неособая точка и на кривой C, и на поверхности X, т. е. если мы для простоты предположим, что у нас сильно неособая ситуация: поверхность неособая в точке P и кривая тоже неособая в точке P. Если сделать такие предположения, то конструкция выглядит следующим образом.

Начнём с обычного локального кольца $\hat{\mathcal{O}}_{X,P}$, которое, как я уже сказал, на самом деле не является локальным объектом. У нас будет вполне инвариантная конструкция, не зависящая от выбора координат, но одновременно я напишу, как она выглядит в тех координатах u, t, которые мы выбрали.

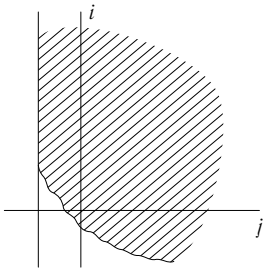
Итак, кольцо $\hat{\mathcal{O}}_{X,P} = k[[u, t]]$ — это ряды Тейлора от двух переменных. Возьмем теперь кривую C и предположим дополнительно, что уравнение $t = 0$ задаёт ее локально, в некотором аффинном множестве. Имеется идеал $(t) = \wp$, который я буду рассматривать в разных кольцах, но обозначать одной и той же буквой. В частности, это идеал в кольце $\hat{\mathcal{O}}_{X,P}$, и, следовательно, мы можем взять локализацию этого кольца вдоль этого идеала. Результат обозначим, как обычно, $(\hat{\mathcal{O}}_{X,P})_P$ и спросим, как его вычислить в координатах. Локализация вдоль идеала $\wp$ состоит в том, что мы разрешаем себе добавить отрицательные степени u. Получается кольцо, которое является кольцом дискретного нормирования и, хотя

исходное кольцо было двумерным (по Круллю), новое кольцо будет уже одномерным. Оно будет неполным и его можно пополнить по степеням идеала $\wp$. И это будет ещё один шаг, независимый от предыдущего. Итак, мы сделали два шага: сначала локализацию и затем пополнение.

После того как это сделано, результат $\widehat{(\hat{\mathcal{O}}_{X,P})_P}$ приобретает вполне простой вид $k((u))[[t]]$, т. е. это — кольцо рядов Тейлора по t с такими коэффициентами. Наконец, последний шаг состоит в том, что мы берём поле частных $\mathrm{Frac}(\widehat{\hat{\mathcal{O}}_{X,P})_P}$. После того как это сделано, мы получим поле $K_{P,C} = k((u))((t))$, т. е. поле итерированных рядов Лорана от переменных u и t *). Оно состоит из элементов вида

$$f = \sum_{j \geqslant j_0} \left(\sum_{i \geqslant i(j)} a_{ij} u^i \right) t^j.$$

У каждого такого элемента f есть носитель $\mathrm{Supp}(f) \subset \mathbb{Z}^2$. Это в точности те целочисленные точки (i, j), для которых $a_{ij} \neq 0$. В терминах носителей условие, что какой-то ряд от двух переменных представляет элемент двумерного локального поля, выглядит так. На плоскости (i, j) должна быть такая вертикальная прямая, что носитель лежит справа от неё. Кроме того, в каждой полоске, где j фиксировано, носитель лежит выше некоторой точки, своей для каждой полоски, (рис. 2); это означает, что коэффициент при t^j является элементом уже одномерного локального поля.



Р и с. 2. Носитель

Мы построили явно пример двумерного локального поля. В самом деле, поле $k((u))((t))$ содержит кольцо $k((u))[[t]]$, которое является кольцом дискретного нормирования. Его поле вычетов является обычным одномерным локальным полем $k((u))$, которое я только что рассматривал в случае алгебраических кривых. Индуктивная структура здесь очевидна.

Я показал, во-первых, что на каждой поверхности, если выбран флаг $X \supset C \ni P$, то естественным образом возникает двумерное локальное поле $K_{P,C}$. Во-вторых, инвариантное определение показывает, что эта конструкция не зависит от выбора u и t, т. е. она вполне каноническая и, более того, обладает хорошими функториальными свойствами. С другой стороны, поле $K_{P,C}$ легко вычислить, если задана какая-то система координат, т. е. оно вполне представимо явно.

*) Заметим, что переход к полю частных тоже является локализацией (подумайте, по какому идеалу?).

В дополнение к определению локального поля посмотрим ещё, какие они бывают. Самое простые поля, которые возникают в геометрии, — это поля итерированных степенных рядов $k((t_1))\ldots((t_n))$. Только что мы разобрали случай $n=2$. Наша конструкция такова, что, если дано n-мерное локальное поле, то есть и $(n-1)$-мерное (его поле вычетов) и т. д.; в конце концов доходим до того, что называется последним полем вычетов. Если последнее поле вычетов $k=\mathbb{F}_q$ конечно, то мы находимся в арифметической ситуации. Я не буду здесь почти ничего говорить об арифметике, но всё-таки скажу, что когда ситуация одномерна, т. е. $n=1$, и $k=\mathbb{F}_q$, то тогда есть локальные поля двух типов. Либо это поля степенных рядов $\mathbb{F}_q((t))$, либо это конечные расширения поля p-адических чисел: $K\supset\mathbb{Q}_p$. Оказывается, что для n-мерных полей тоже можно дать теорему классификации. Например, при $n=2$ есть итерированные ряды Лорана $\mathbb{F}_q((u))((t))$. Затем есть ряды $K((t))$, где K — конечное расширение поля p-адических чисел; достаточно очевидно, что это тоже будет локальным полем. Оказывается, что есть и третье, довольно любопытное, поле

$$K\{\{t\}\}=\left\{\sum a_i t^i\colon |a_i|\leqslant C,\ a_i\to 0 \text{ при } i\to\infty\right\}.$$

Эти три типа исчерпывают (в разумном смысле) все арифметические двумерные локальные поля.

Таким образом, новые конструкции развивают одномерную ситуацию в более высокой размерности, но пока являются просто определением, непонятно зачем нужным. Хорошо бы попробовать, что с ними можно делать. Чтобы это выяснить, вернемся сначала к одномерному случаю и поговорим о хорошо известных вещах, т. е. о том, что можно делать с обычными одномерными локальными полями.

Самая ходовая вещь в алгебраической геометрии, с которой все начинают, — это когомологии пучков. Поэтому первое, что я расскажу, как можно интерпретировать пучки в терминах локальных полей. Пусть у нас есть кривая C и на ней поле рациональных функций и локальные поля. Можно ввести кольцо аделей $\mathbb{A}=\prod\limits_{p\in C} K_p$; здесь берётся адельное произведение локальных полей по всем точкам кривой. Оно состоит из таких наборов элементов $f_p\in K_p$, что $f_p\in\hat{\mathcal{O}}_P$ почти для всех p, где $\hat{\mathcal{O}}_P$ — кольцо дискретного нормирования, которое есть в каждом локальном поле. Напомню, что K_P — пополнение поля K.

Рассмотрим самые простые пучки, которые отвечают линейным расслоениям. Каждый такой пучок $\mathcal{L}=\mathcal{O}_C(D)$ связан с некоторым дивизором $D=\sum\limits_{P\in C} n_P P$ кривой C. И каждому дивизору можно сопоставить некоторое подпространство $\mathbb{A}(D)\subset\mathbb{A}$. А именно, рассмотрим не все адели,

а только те, для которых $\nu_P(f_P) \geqslant -n_P$ для любой точки P. Поскольку $n_P = 0$ для почти всех P, это вполне согласуется с определением аделей. Иными словами, это означает, что особенности адельных векторов могут находиться только в конечном числе точек и ещё порядок полюса в каждой из таких точек мы ограничиваем.

Дивизоры образуют фильтрующееся множество: можно взять больше точек, или большие кратности. Когда всё это растет, то вы тем самым исчерпываете всё множество $\mathbb{A}$, т. е. каждый элемент множества $\mathbb{A}$ принадлежит подпространству $\mathbb{A}(D)$ для некоторого дивизора D.

Теперь можно рассмотреть комплекс $K \oplus \mathbb{A}(D) \to \mathbb{A}$, в котором отображение представляет собой сумму естественных вложений. Замечательный факт состоит в том, что имеет место канонический изоморфизм

$$H^*(K \oplus \mathbb{A}(D) \to \mathbb{A}) = H^*(C, \mathcal{L}),$$

т. е. локальные поля позволяют вычислять когомологии пучков. Дальнейшее их применение связано с теорией двойственности. Для когомологий когерентных пучков имеется теория двойственности, которая основана на существовании фундаментального класса. Эту теорию тоже можно изложить с помощью локальных полей. Я скажу лишь как построить фундаментальный класс. До сих пор мне было неважно, какую кривую мы имеем, аффинную или проективную. Но раз мы переходим к рассмотрению двойственности, то естественно предположить, что C — гладкая проективная кривая. Что такое фундаментальный класс на кривой C?

Мы должны взять на кривой пучок дифференциальных форм Ω^1_C. Основная теорема состоит в том, что есть каноническое отображение $H^1(C, \Omega^1_C) \to k$, которое является изоморфизмом. С точки зрения обычной топологии это понятно. По разложению Ходжа группа $H^{1,1}$, которая совпадает с $H^1(C, \Omega^1_C)$, действительно является топологическим $H^2(C, \mathbb{C}) = \mathbb{C}$. Тем не менее, я хочу подчеркнуть, что, во-первых, этот изоморфизм — чисто алгебраический и имеет место над любым полем, а во-вторых, теория локальных полей доставляет его совершенно замечательное объяснение.

Я уже написал адельный комплекс, который вычисляет когомологии любого пучка. В частности, он вычисляет когомологии пучка дифференциальных форм. Чтобы это сделать, нужно записать этот пучок в виде пучка, отвечающего какому-то дивизору. Для этого нужно фиксировать ненулевую рациональную дифференциальную форму ω. Тогда её дивизор будет как раз нужным нам дивизором. Для любой такой формы у нас имеется представление пучка дифференциальных форм в виде $\Omega^1_C = \mathcal{O}_C((\omega))$; здесь $\omega \in \Omega^1_K$, $\omega \neq 0$. Но можно подойти к вычислению когомологий пучка Ω^1_C и по-другому.

Дело в том, что адельный комплекс можно записать не только для аделей, компоненты которых являются элементами локального поля; мы можем взять дифференциальные формы поля рациональных функций, а также дифференциальные формы любого локального поля. И можно взять точно такое же адельное произведение $\prod\limits_P \Omega^1_{K_P}$, относительно подпространств форм, регулярных в P. Иными словами, эта конструкция переносится на случай дифференциальных форм, и абсолютно без всяких изменений; изменяются только обозначения. Каждое локальное пространство $\Omega^1_{K_P}$ выглядит так: $\Omega^1_{K_P} = K_P\,dt$. Взяв адельное произведение таких пространств, мы затем берём рациональные дифференциальные формы степени 1, глобально определённые на всей кривой, и добавляем произведение пространств регулярных дифференциальных форм $\prod\limits_P \Omega^1_{\hat{\mathcal{O}}_P}$, где $\Omega^1_{\hat{\mathcal{O}}_P} = \hat{\mathcal{O}}_P\,dt$. Получаем адельный комплекс

$$\Omega^1_K \oplus \prod_P \Omega^1_{\hat{\mathcal{O}}_P} \to \prod_P \Omega^1_{K_P}$$

и его когомологии как раз и будут когомологиями пучка дифференциальных форм. В частности, нас интересует $H^1(C, \Omega^1_C)$, т. е. коядро отображения $\Omega^1_K \oplus \prod\limits_P \Omega^1_{\hat{\mathcal{O}}_P} \to \prod\limits_P \Omega^1_{K_P}$.

Теперь я введу новое понятие, понятие *вычета*. Оно даёт возможность построить фундаментальный класс, а потом доказать его свойства. Именно, нужно доказать, что имеется каноническая точная последовательность

$$\Omega^1_K \oplus \prod_P \Omega^1_{\hat{\mathcal{O}}_P} \to \prod_P \Omega^1_{K_P} \to k \to 0.$$

Она строится с помощью вычетов. Понятие вычета состоит в том, что для каждой точки P определено отображение $\operatorname{res}_P\colon \Omega^1_{K_P} \to k$. Это отображение происходит из классического анализа: вычет формы $\omega = \sum a_i t^i dt$ равен $\operatorname{res}_P(\omega) = a_{-1}$ и не зависит от выбора переменной t.

Если $k = \mathbb{C}$, то вычет задается интегралом

$$\operatorname{res}_P(\omega) = \frac{1}{2\pi i} \int\limits_\gamma \omega,$$

где γ — маленькая петля вокруг точки P.

Все, кто изучал теорию вычетов, знают, что есть два основных факта.

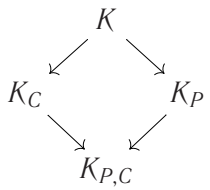
1. $\operatorname{res}|_{\Omega^1_{\hat{\mathcal{O}}_P}} = 0$, потому что у регулярных форм нет коэффициента a_{-1} (это соотношение локальное).

2. Глобальное соотношение для рациональных форм состоит в том, что $\sum\limits_{P \in C} \operatorname{res}_P(\omega) = 0$ для $\omega \in \Omega^1_K$.

Совершенно ясно, что отсюда вытекает то, что нам нужно: отображение $\prod\limits_P \Omega^1_{K_P} \to k$ можно определить как сумму вычетов $\sum\limits_P \operatorname{res}_P$. Заметьте, что наше определение аделя даёт, что для почти всех точек P форма $\omega_P \in \Omega^1_{K_P}$ регулярна. Поэтому возникающая сумма по P хотя и бесконечна, но определена корректно. Это первое. Второе, что нужно проверить, что так определённое отображение обращается в нуль на образе отображения $\Omega^1_K \oplus \prod\limits_P \Omega^1_{\hat{\mathcal{O}}_P} \to \prod\limits_P \Omega^1_{K_P}$. Проверка этого состоит из двух независимых действий. Одна проверка, локальная для каждого P, другая, глобальная, для образа пространства Ω^1_K. Обе немедленно вытекают из сформулированных выше свойств вычета.

Из того, что я сказал, ещё не вытекает, что наше отображение дает изоморфизм $H^1(C, \Omega^1_C) \to k$. Но это уже несложная работа. И тогда прообраз 1 и будет фундаментальным классом в группе $H^1(C, \Omega^1_C)$ и мы видим, что фундаментальный класс на всей кривой представляется в виде суммы локальных фундаментальных классов.

Теперь давайте посмотрим, что получается для двумерного случая. На поверхности X есть поле рациональных функций K, всё ещё не определённые мной «поля» K_P и K_C, связанные с точками и кривыми, и есть чисто локальный объект $K_{P,C}$ — двумерное локальное поле, отвечающее флагу $C \ni P$:



О том, что не было определено, я пока говорить не буду. А о том, что было определено, можно и поговорить.

Что касается двумерных вычетов, то первым написал соответствующий интеграл, по-видимому, Пуанкаре. Потом появилась много других вычетов, вычет Лере, Гротендика и т. д. То, о чём я буду говорить, это казалось бы ещё один вычет, но, на самом деле, он тот же самый. Невозможно придумать разные вычеты в принципе. Другое дело, что каждый из этих вычетов определён только в своей ситуации и обладает своим разумным набором свойств.

Одномерный вычет определялся для мероморфной локальной (в том смысле, что её коэффициенты лежали в некотором одномерном локальном поле) формы $\Omega^1_{K_P} = K_P\, dt$ степени 1. Теперь, когда у нас есть двумерное локальное поле, мы можем ввести дифференциальные формы степени 2.

Это будут суммы вида

$$\sum_{i,j} a_{ij} u^i t^j \, du \wedge dt \in \Omega^2_{K_P} = K_P \, du \wedge dt.$$

Как всегда, бывают определения инвариантные и определения, зависящие от выбора координат. Уже с одномерным вычетом доказательство того, что коэффициент a_{-1} не зависит от выбора t, нетривиально.

Для формы степени 2 вычет естественно определить так: $\mathrm{res}_{P,C}(\omega) = a_{-1,-1}$. Дальше возникает вопрос: что про него можно сказать?

Прежде всего, покажем, что вычет тоже вычисляется как интеграл, если $k = \mathbb{C}$. Именно,

$$\mathrm{res}_{P,C}(\omega) = \frac{1}{(2\pi i)^2} \int\limits_{\gamma} \omega,$$

где γ — двумерный цикл на поверхности X (топологически она — четырёхмерное многообразие). Чтобы этот цикл построить, возьмем небольшой шар около точки P. Его граница будет трехмерной сферой, и C пересекается со сферой по конечному числу непересекающихся, но, вообще говоря, заузленных, окружностей. Окружая каждую окружность небольшой трубкой, получим требуемый цикл.

Вернемся к алгебраической ситуации произвольного основного поля. Имеются ли для нашего вычета соотношения такого же типа, как и для одномерного вычета? Оказывается, что этих соотношений уже не два, а три. И вообще, в n-мерной ситуации количество соотношений равно $n+1$.

В двумерном локальном поле $K_{P,C}$ содержится подкольцо дискретного нормирования $\mathcal{O}_{P,C}$, которое состоит из рядов с неотрицательными j. Заметьте, что вычет строится итерированным способом. Сначала мы берём коэффициент при t^{-1} и получаем ряд по u. А затем у этого ряда берём коэффициент при u^{-1}. Получается такая двуступенная конструкция. Я забыл сказать важный факт, что результат не зависит от выбора u и t. Причём, что любопытно, в этой конструкции промежуточный шаг, вообще говоря, *зависит* от выбора u и t.

Теперь, какие имеются соотношения для вычетов?

1. Если $\omega \in \Omega^2_{\mathcal{O}_{P,C}}$, то тогда вычет равен нулю. Это понятно: если у вас нет отрицательных степеней t, то уже на первом шаге получается нуль.

2. Фиксируем точку P и рассмотрим (неприводимые) кривые $C \ni P$, бегающие вокруг этой точки. Для простоты возьмём $\omega \in \Omega^2_K$, т. е. дифференциальную форму, рациональную на всей поверхности. Тогда имеет место соотношение

$$\sum_{C \ni P} \mathrm{res}_{P,C}(\omega) = 0.$$

3. Если мы фиксируем кривую C и рассмотрим точку P, бегающую вдоль этой кривой, то есть двойственное соотношение, которое состоит в том, что

$$\sum_{P\in C} \operatorname{res}_{P,C}(\omega) = 0.$$

Это соотношение также имеет место для $\omega \in \Omega^2_K$.

Заметьте, что соотношение 3 не очень удивительно. Уравнение $t = 0$ задаёт кривую C (в каком-то открытом множестве). Фактически, на первом шаге мы взяли обычный одномерный вычет по отношению к C. Получилась форма степени 1 на C. Потом мы взяли её вычеты во всех точках. Это уже почти доказательство; нужно только его аккуратно записать.

А соотношение 2 a priori ни из чего не следует. Мораль состоит в том, что эти соотношения нужны оба. Они двойственны друг другу, и одно из другого не вытекает.

Все же я вкратце скажу, из чего получается соотношение 2. Можно свести форму к самому простому случаю, когда её особенностями будут две трансверсальные прямые. Это — простейший случай, к которому всё сводится, и тогда у нас будет форма $\frac{du\wedge dt}{ut}$. Точка P фиксирована, и нужно учесть вычеты только относительно двух кривых C_1 и C_2, проходящих через эту точку: в сумме вычетов по всем кривым, вертящимся вокруг точки P, нетривиальный вклад дают только две координатные оси (все остальные кривые будут давать нуль). Возникают два локальных поля K_{P,C_1} и K_{P,C_2}. В одном из них форма имеет вид $\frac{du\wedge dt}{ut}$, а в другом $-\frac{du\wedge dt}{ut}$, потому что переменные в локальных полях переставляются. Поэтому сумма будет равна нулю. А дальше с помощью раздутий, например сигма-процесса, всё можно свести к такому случаю. Это не лучшее доказательство, но оно, по крайней мере, объясняет суть дела.

Я замечу, что здесь проявляется двойственность между точками и кривыми: их можно переставлять. Это именно та самая двойственность между точками и прямыми, которая есть в элементарной проективной геометрии. Здесь, правда, не прямые, а кривые, но это не так важно.

Что можно делать с помощью теории вычетов? С её помощью можно, как и для кривых, построить теорию двойственности и фундаментальный класс. Здесь я немножко перескочил. В одномерном случае, прежде чем строить теорию двойственности, я сначала объяснил, как строить резольвенты для пучков (или адельные комплексы), а уже потом перешёл к этой конструкции. Я скажу, без всякого объяснения, что аналоги такой резольвенты можно построить в любой n-мерной ситуации. Иначе говоря,

локальные поля позволяют всё это перенести и построить теорию, которая даёт вычисление когомологий когерентных пучков на многообразиях любой размерности.

Можно пытаться перенести приведенные выше определения в более высокие размерности прямолинейным образом, примерно так, как мы доказывали соотношение вычетов для точки на поверхности. Давайте всё сведём к простейшему случаю... давайте раздуем... Довольно обычная идеология для человека, привыкшего работать в «конвенциональной» алгебраической геометрии. Когда я этим занимался, у меня был период сотрудничества с Сашей Бейлинсоном, который, не очень-то имея опыт работы с классической алгебраической геометрией, умел зато весьма лихо работать с высокими конструкциями в области гомологической алгебры и симплициальных множеств. Он и придумал общее определение адельного произведения n-мерных локальных полей на любых схемах, т. е. даже не только на многообразиях. Определение формализует то определение локального поля, которое мы обсуждали выше, в случае флага на поверхности. Помните, там было четыре шага, в которых все время переставлялись пополнение и локализация. Это и легло в основу его определения. Замечательное свойство конструкции Бейлинсона состоит в том, что адели и адельные комплексы определяются для любой схемы, она совершенно равнодушна к особенностям. И теория когомологий и двойственности имеет место в совершенно общей ситуации. К сожалению, Саша не стал участвовать в дальнейшем развитии теории аделей (даже доказательства его теорем об аделях опубликовала немецкая математичка Аннетта Хюбер [Hu]).

Сама эта теория адельных комплексов для когерентных пучков — вполне законченная теория, причем она закончена в самой максимальной общности. Больше в ней, пожалуй ничего нельзя сделать или добавить. Поэтому, может быть, не стоит мне здесь об этом подробно рассказывать. Этим надо было бы весь час занять (подробно это все изложено в [PF]). Чтобы это совсем не повисло в воздухе, я все же небольшой кусок объясню.

Так же, как для кривых у нас был адельный комплекс для дифференциальных форм и мы могли вычислить старшую группу $H^1(C, \Omega^1_C)$, исходя из этого комплекса, в двумерной ситуации тоже имеется адельный комплекс, который я, однако, целиком писать не буду.

Поскольку теперь мы работаем с поверхностью, если основное поле $k = \mathbb{C}$, то мы имеем вещественное четырёхмерное многообразие. Поэтому согласно общей теории в разложении Ходжа для когомологий старшая группа $H^{2,2}$ нетривиальна и одномерна. Но это как раз 2-мерные

когомологии пучка дифференциальных форм степени 2, $H^2(X, \Omega^2_X)$. Соответственно, алгебраический аналог топологического утверждения состоит в том, что пространство $H^2(X, \Omega^2_X)$ канонически является одномерным.

Теперь нужно написать часть того адельного (более длинного на поверхности) комплекса, который я не определил. Сюда, прежде всего, входит адельное произведение $\prod\limits_{P,C} \Omega^2_{K_{P,C}}$ (т. е. не просто произведение, а с условиями, нетривиально обобщающими условия, имевшиеся на кривых). И тогда пространство H^2 является его фактором, и в качестве ядра уже будут не две подгруппы, как это было в одномерном случае, а три:

$$\prod_{P,C} \Omega^2_{\mathcal{O}_{P,C}} \oplus \prod_{P\in X} \Omega^2_{K_P} \oplus \prod_{C\subset X} \Omega^2_{K_C} \to \prod_{P,C} \Omega^2_{K_{P,C}} \xrightarrow{\sum\limits_{P,C} \mathrm{res}_{P,C}} H^2 \to k$$

Сумма вычетов имеет смысл, как и выше, ибо вычеты равны нулю почти для всех флагов P, C. Чтобы определить отображение из H^2 в k, нужно проверить, что оно равно нулю на каждой из этих трёх подгрупп. Для регулярных форм это первое из сформулированных выше свойств вычета. А остальные два свойства, если на них внимательно посмотреть, дают обращение в нуль на двух других компонентах *).

Для построения комплекса мы используем отображения диагональных вложений $\Omega^2_{K_P} \to \Omega^2_{K_{P,C}}$ и $\Omega^2_{K_C} \to \Omega^2_{K_{P,C}}$, индуцированных вложениями полей K_P в $K_{P,C}$, когда C бегает вокруг фиксированной точки P, и $K_C P$ в $K_{P,C}$, когда P пробегает вдоль фиксированной кривой C. При таком диагональном вложении в формуле получаются как раз те суммы вычетов, которые я написал раньше.

Мне осталось определить объекты, зависящие только от точек или только от кривых. После этого сюжет будет закончен. У нас есть поле рациональных функций на поверхности: $K = k(X)$. Что такое поле, которое отвечает неприводимому дивизору? Это объект, который хорошо известен. Имеется нормирование $\nu_C : K^* \to \mathbb{Z}$, отвечающее дивизору (т. е. порядок полюса или нуля вдоль этого дивизора), и по по нему можно пополнить K и получить поле K_C. Посмотрим, как оно устроено. У кривой C в каком-то открытом множестве имеется глобальное уравнение $t = 0$. Заметьте, глобальное, но в аффинном множестве. Совсем глобального уравнения, на всей поверхности X, вообще говоря, нет. Но нормирование этого не чувствует, оно зависит только от общей точки кривой; можно выкинуть из

*) Одна тонкость: мы сформулировали свойства вычета лишь для форм ω приходящих со всей поверхности X, а в комплексе их нужно применять к формам из, соответственно, $\Omega^2_{K_P}$ и $\Omega^2_{K_C}$.

нее сколько угодно точек, порядок полюса от этого не изменится. Пополнение теперь легко вычисляется: $K_C = k(C)((t))$.

Заметьте, что с нашей точки зрения, это поле смешанное. У него две части: одна глобальная (одномерная), вдоль кривой C, а другая чисто локальная (она тоже одномерная), по нормали к кривой. Поэтому в диаграмме это поле занимает промежуточное положение между K и чисто локальным полем $K_{P,C}$.

Ещё нам нужен объект K_P, зависящий только от точки. К сожалению, это не поле; это только кольцо. Устроено оно вот как: начать нужно с двумерного локального кольца $\hat{\mathcal{O}}_{K,P}$, которое связано с точкой. Как мы уже говорили, это кольцо рядов Тейлора от двух переменных в окрестности точки P. Первое, что приходит в голову, это просто взять его поле отношений. Это, конечно, можно сделать, но это неправильный ответ. А правильный ответ более хитрый. Он состоит вот в чем.

Пусть $f,\ g \in \hat{\mathcal{O}}_{K,P}$. Поле отношений — это поле, состоящее из дробей f/g. У этой дроби есть нули и есть особенности — полюсы (нули знаменателя). Из-за того, что g задается, вообще говоря, степенным рядом, нули (неприводимые компоненты $g = 0$) будут определять не настоящие кривые на всей поверхности, а ростки кривых. Правильный ответ состоит в том, что в этой науке такие ростки запрещаются и вводится условие на знаменатель g. Оно состоит в том, что кривые $g = 0$ должны быть настоящими кривыми на поверхности X. При таком ограничении ясно, что поля вы уже не получите. В числителе стоят любые ряды, а в знаменателе не любые. Поэтому нули могут быть любыми ростками, а полюсы нет и, следовательно, обратный элемент определён не всегда.

Вспоминая нашу диаграмму полей, связанных с поверхностью, мы видим, что окрестности точек и кривых, связанные, соответственно, с K_P и K_C, являются не самыми локальными объектами на поверхности. Локальный объект — это флаг $P \in C \subset X$ и как каждая кривая $C \subset X$ состоит из всех флагов $P \in C \subset X$, так и каждая точка $P \in X$ «состоит» из всех флагов $P \in C \subset X$.

Еще одно общее замечание. Мы видим, что точки и кривые входят в теорию симметричным образом. Эту симметрию можно объяснить еще и так. В теории схем Гротендика, как вы знаете, точки могут быть не только замкнутыми. Поэтому мы должны рассматривать все точки, любых коразмерностей. В нашем примере P коразмерности 2, C коразмерности 1, X коразмерности 0. И в каком-то приближении все они неразличимы: их можно переставлять, не нарушая симметрии. А потом они уже распадаются на точки коразмерности 0, 1, и т. д. Как говорят в физике, происходит спонтанное нарушение симметрии. Но это так, вольный образ.

Я рассказал свой первый сюжет. Эта вещь была придумана мною много лет назад и много лет нигде особенно не использовалась и не применялась *). Каково же было моё изумление, когда вышла книга [АК] В. И. Арнольда и присутствующего здесь Бори Хесина, где было дано определение голоморфного числа зацепления и доказана его инвариантность, и это вычисление фактически было соотношением для вычетов, только не для поверхности, а для трёхмерного многообразия в некотором специальном случае. Во всяком случае, оно там явно появлялось. По-видимому, это первый случай, когда то, что я здесь рассказал, получило реальный выход в совсем другую область математики.

Теперь я перейду к теме, которая параллельна рассказанной, но формально от нее независима. Так что то, что я рассказал, будет новую тему немножко объяснять. Дело в том, что в алгебраической геометрии те конструкции, которые здесь были (вычеты, резольвенты), это всё аддитивные конструкции. Реально мы здесь используем только аддитивную группу (двумерного) локального поля. Но поскольку оно — кольцо, то с ним связана группа по умножению. И мультипликативная группа n-мерного локального поля несет на себе совершенно замечательную конструкцию, которая называется *ручным символом*.

Пусть K — n-мерное локальное поле с последним полем вычетов k. Тогда можно определить n-мультипликативное (я буду все же говорить n-линейное, или билинейное, когда $n=2$) отображение $\underbrace{K^* \times \ldots \times K^*}_{n+1} \to k^*$. Если $n=0$, то это просто тождественное отображение. Наука начинается в одномерном случае. В одномерном случае получается билинейная форма $K^* \times K^* \to k^*$, которая называется ручным символом. Пусть $K = k((t))$ — поле степенных рядов, его мультипликативная группа $K^* = k((t))^*$ состоит из ненулевых степенных рядов. Пусть $f,\ g \in k((t))^*$ и порядки их полюсов (нулей) равны, соответственно, m и n. По определению ручной символ равен

$$(f,\ g) = (-1)^{mn} f^{-n} g^m(0) \in k^*.$$

Особенности функций $f,\ g$ сокращаются и то, что получится, будет степенным рядом по t, начинающимся с ненулевого постоянного члена. Поэтому можно взять его значение в нуле; это будет корректно определённое число из k^*.

*) Я не говорю, конечно, об арифметике, например, теории полей классов. См. обзор [PF].

Ручной символ удовлетворяет нескольким соотношениям. По каждой переменной он мультипликативен, и еще удовлетворяет соотношению символа $(f, 1-f)=1$. Подробно я об этом говорить не буду. Если мы рассмотрим ручной символ не в чисто локальной ситуации, а в глобальной, когда у нас есть кривая C, и для каждой ее точки P определен локальный символ, то оказывается, что он удовлетворяет соотношениям, которые удивительно похожи на соотношения для вычетов.

Пусть $K=k(C)$ — поле рациональных функций на кривой C и для каждой точки P имеется вложение $K\to K_P$. Мультипликативная группа K_P^* содержит группу $\mathcal{O}_P^*$.

1. Если $f,\ g\in\mathcal{O}_P^*$, то $(f,\ g)_{K_P}=1$.
2. Если $f,\ g\in K^*$, то $\prod\limits_{P\in C}(f,\ g)_{K_P}=1$.

Первое свойство локально и очевидно, ибо в этом случае $m=0$ и $n=0$. Второе свойство, глобальное, и в высшей степени неочевидное; оно называется *законом взаимности Вейля*. Это те самые законы взаимности, которые идут из XIX-го века, от Гильберта и других классиков.

Если посмотреть на свойства одномерных вычетов, то увидим, что свойства символа формулируются весьма параллельным образом и это не случайное совпадение *). Более того, так же, как можно перейти к n-мерным вычетам, существуют символы для любого n-мерного локального поля и они зависят от $n+1$ аргумента.

В частности, пусть дана поверхность X и флаг $X\supset C\ni P$. Тогда мы имеем двумерное локальное поле $K_{P,C}$ и для любых трёх ненулевых функций $f,\ g,\ h\in K_{P,C}^*$ можно определить символ

$$(f,\ g,\ h)_{K_{P,C}}\in k^*.$$

Каждому элементу $f=\sum\limits_{j\geqslant j_0}a_{ij}u^it^j$ можно сопоставить пару целых чисел $\nu(f)=(p,\ m)\in\mathbb{Z}^2$. Вспомним, что носитель f устроен так: имеется вертикальная прямая, левее которой ничего нет. Возьмем самую правую из таких прямых, она определяет число m. Теперь возьмём первую вертикальную полосу после этой прямой. В ней до какого-то p-го места стоят нули, а на p-м месте стоит ненулевой элемент. Это дает второе целое число. Заметьте, что пара $\nu(f)$ не инвариантна: число m определено инвариантно, а число p зависит от выбора координат $(t,\ u)$. Тем не менее,

*) Более того, их можно включить в единую формулу, если рассмотреть кривые не над полем, а над артиновым локальным кольцом, т. е. немного ее продеформировав. См. Anderson G., Pablos Romo F. Simple proofs of classical explicit reciprocity laws on curves using determinant groupoids over an artinian local ring. Препринт `math.NT/0207311`

используя эти числа, можно построить нечто, от выбора координат не зависящее. Это делается следующим образом. Пусть даны два элемента из $\mathbb{Z}^2$. Для них определено внешнее произведение $\mathbb{Z}^2 \wedge \mathbb{Z}^2 \to \mathbb{Z}$ (определитель). Тогда $\nu(f) \wedge \nu(g)$ уже не зависит от выбора координат (догадайтесь, почему?). Теперь можно попытаться определить тройной символ. Рассмотрим произведение:

$$f^{\nu(g)\wedge\nu(h)} g^{-\nu(f)\wedge\nu(h)} h^{\nu(f)\wedge\nu(g)} \in K^*_{P,C}.$$

Как и в одномерном случае, элементы поля возводятся в некоторую целочисленную степень. А дальше оказывается, что так же, как в одномерном случае, особенности сократятся: и по u и по t. Поэтому мы можем взять значение в нуле (когда $t = 0$ и $u = 0$). Это значение корректно определено, и оно будет элементом из k^* (ненулевым элементом основного поля). Получилась трилинейная форма. У неё есть ещё знак. Для знака есть явная формула, но она немножко длинновата; позвольте мне её не писать, поскольку эта формула в данный момент ничему нас не научит.

Как я уже сказал, можно построить и $(n+1)$-линейную форму для любого n-мерного локального поля.

Оказывается, что совершенно параллельно соотношениям между вычетами для двумерного случая (на поверхности) есть законы взаимности для трилинейных форм в двумерной ситуации (и вообще в любой n-мерной ситуации). Давайте я напишу хотя бы часть из них.

1. Если фиксирована точка P и рассматриваются кривые C, содержащие эту точку, то $\prod\limits_{C \ni P} (f,\ g,\ h)_{K_{P,C}} = 1$.

2. Наоборот, если фиксирована кривая C и рассматриваются точки $P \in C$, то $\prod\limits_{P \in C} (f,\ g,\ h)_{K_{P,C}} = 1$.

В данный момент это выглядит некоей конструкцией, непонятно для чего существующей. Скажу только, что есть теория полей классов, которая занимается описанием (вычислением группы Галуа) абелевых расширениий как локальных, так и глобальных полей. В одномерной теории полей классов, которая имеет место для кривых над конечным полем или для полей алгебраических чисел, ручные символы (и некоторые их обобщения — символы норменного вычета) играют фундаментальную роль. И закон взаимности, который здесь написан, занимает центральное место в теории полей классов. Существует обобщение теории полей классов на n-мерный случай, в котором эти символы тоже играют фундаментальную роль. Это, по крайней мере, объясняет, почему они существенны в арифметике. Закон взаимности нужен для глобальной теории, а в локальной теории полей классов символы дают двойственность Куммера, которая

есть в любой размерности. Я не буду об этом сейчас говорить (см. изложение в [PF]).

По отношению к теории полей классов это уже старые вещи. А совсем недавно Аскольд Хованский нашёл применение высших ручных символов к вычислению числа решений систем полиномиальных уравнений. Он это неоднократно рассказывал, поэтому этот сюжет многим знаком.

Я хочу рассказать довольно свежие вещи о новом доказательстве закона взаимности. Есть разные способы его доказательства, в том числе и способы, не очень идейно привлекательные. Скажем, чисто вычислительные. Недавно мне вместе с Денисом Осиповым удалось получить доказательство закона взаимности Вейля для случая кривой, использующее конструкцию монодромии. Об этом я и расскажу в заключение *).

Рассмотрим локальное поле K_P и опустим на время индекс P. В локальном поле K фиксируем подпространство рядов Тейлора $\mathcal{O} = k[[t]]$ и будем рассматривать K как бесконечномерное подпространство над полем k. В нём имеется интересное отношение эквивалентности подпространств, называемое *соизмеримостью*:

$A \sim B$, если $A \cap B$ имеет конечную коразмерность внутри каждого из пространств A и B, т. е. A и B имеют конечную размерность над $A \cap B$.

Если есть два соизмеримых пространства A и B, то можно определить одномерное векторное пространство $(A|B)$. Замечу, что все одномерные пространства, конечно, изоморфны, но не канонически. Поэтому между двумя одномерными пространствами нет единственного изоморфизма. Их много, и это важный момент. В частности, если можно в одномерном пространстве выбрать канонически какой-то ненулевой элемент, то это значит, что оно канонически изоморфно k. А вот автоморфизмы одномерного векторного пространства описываются элементами из поля k^* вполне каноническим образом (почему?).

Векторное пространство $(A|B)$ определяется следующим образом: $(A|B) = \det(A/A \cap B) \otimes \det(B/A \cap B)^*$.

Детерминант векторного пространства — это его максимальная внешняя степень. Для одномерных векторных пространств есть такие операции:

---

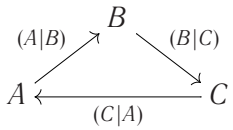
*) Для тех, кто знает конструкцию Делиня ручного символа как монодромии пучков со связностью (опубликованную в Publications Mathématiques IHES за 1991 г., но придуманную, как у него обычно, за много лет до этого), сразу скажу, что это — совсем другая конструкция. К тому же конструкция Делиня комплексно-аналитическая, в ней присутствуют интегралы, и она годится только когда $k = \mathbb{C}$, а наша конструкция чисто алгебраическая, над любым полем k. Сравнение этих конструкций — очень интересная задача.

E^{-1} — это то же самое, что переход к пространству, двойственному к E; тензорное произведение $E \otimes F$. Затем скажем, что какое-то пространство равно 1, если это пространство канонически изоморфно k. Более общо, когда мы пишем $E = F$, то это означает, что задано каноническое отображение $E \to F$, являющееся изоморфизмом.

Построенная скобка обладает такими свойствами:

1. $(A|B) = (B|A)^{-1}$.
2. $(A|B)(B|C) = (A|C)$ (стягивание).
3. Мультипликативная группа K^* действует на K умножением: $K^* \times K \to K$, т. е. каждый элемент K^* — это линейный оператор в K и все они сохраняют соизмеримость. Точнее, имеется канонический изоморфизм $(A|B) \xrightarrow{\times f} (fA|fB)$.

Второе свойство можно записать довольно любопытным образом; у него есть геометрический (симплициальный) смысл. Давайте нарисуем треугольник:



В вершины треугольника поместим подпространства A, B и C того типа, который мы рассматриваем. Затем мы соединим вершины рёбрами и каждому ребру сопоставим одномерное пространство. Если мы пройдём по границе симплекса (треугольника), то такое тройное произведение равно 1. Такая интерпретация подсказывает, конечно, возможные обобщения, где появятся симплексы высшей размерности.

Оказывается, что с помощью такой конструкции можно определить ручные символы. То, что я сейчас расскажу, — это вещи старые, но фольклорные. Их, наверное, никто никогда не опубликовал. Я думаю, что первым это написал Делинь в 70-е годы в письме Ларри Брину. У него, правда, этих скобок не было, он использовал более старый язык. Но фактически там всё было сказано. Затем независимо это сделал де Кончини в 1990 году. Перед этим у него была совместная работа с Арбарелло и Виктором Кацем [ADK] про такие скобки, про символы, про законы взаимности, где они придумали, как построить ручной символ исходя из этих скобок. Но конструкция была очень сложная, там была масса проверок, было много вычислений. Потом, через год, де Кончини придумал очень простое определение. Когда он мне его рассказал, я сразу вспомнил то, давнее письмо Делиня, в котором это определение было. Тем не менее, наш разговор не пропал втуне, потому что после этого началась деятельность, которая привела к тому, что я сейчас вам расскажу.

Как же выглядит описание Делиня и де Кончини? Фиксируем подпространство $\mathcal{O}$. Можно фиксировать и другое пространство, но раз уж у нас есть пространство $\mathcal{O}$, то фиксируем его. И будем работать с подпространствами в K, соизмеримыми с $\mathcal{O}$, т. е. $A \sim \mathcal{O}$.

Рассмотрим диаграмму отображений

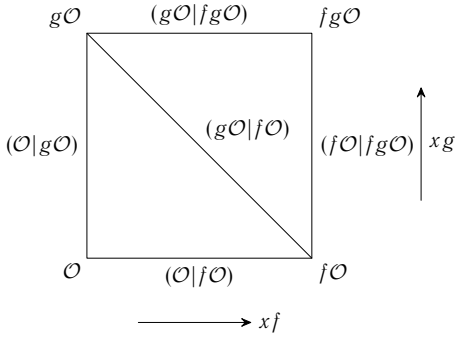
$$\begin{array}{ccc} (f\mathcal{O}|\mathcal{O})(\mathcal{O}|g\mathcal{O}) & = & (f\mathcal{O}|g\mathcal{O}) \\ {\scriptstyle \times g}\downarrow{\scriptstyle \times f} & & \| \\ (f\mathcal{O}|fg\mathcal{O})(gf\mathcal{O}|g\mathcal{O}) & = & (f\mathcal{O}|g\mathcal{O}) \end{array}$$

Верхнее равенство появляется из свойства стягивания. Используя свойство 3, можно преобразовать скобки с помощью умножений на f и g. У нас возникают отображения $(f\mathcal{O}|\mathcal{O}) \xrightarrow{\times g} (gf\mathcal{O}|g\mathcal{O})$ и $(\mathcal{O}|g\mathcal{O}) \xrightarrow{\times f} \xrightarrow{\times f} (f\mathcal{O}|fg\mathcal{O})$. С их помощью строится левая стрелка диаграммы. Поле коммутативно, поэтому $fg = gf$ и нижнее равенство тоже появляется из свойства стягивания.

Мы получили диаграмму и можно спросить: коммутативна ли она? Оказывается, что она коммутативна с точностью до ручного символа (f, g). Иными словами, проходя по этой диаграмме, мы получим отображение одномерного пространства $(f\mathcal{O}|g\mathcal{O})$ в себя, т. е. число (элемент основного поля). Это число и есть ручной символ.

Когда я это увидел, у меня появилась мысль, что в этом есть какой-то геометрический смысл, что тут явно возникает какая-то монодромия. Это долго не удавалось реализовать, но сейчас Денис Осипов этим заинтересовался. Мы об этом подумали, и благодаря его энергии все удалось понять.

Рассмотрим квадрат, в вершинах которого написаны подпространства $\mathcal{O}$, $f\mathcal{O}$, $g\mathcal{O}$, $fg\mathcal{O}$, (рис. 3). Рёбрам сопоставим одномерные пространства, способом о котором я уже говорил. Проведём ещё диагональ, которой сопоставим пространство $(g\mathcal{O}|f\mathcal{O})$; автоморфизм этого пространства и будет ручным символом. В силу свойства треугольника пространство $(g\mathcal{O}|f\mathcal{O})$ является как произведением двух пространств, отвечающих сторонам треугольника, лежащего над диагональю, так и произведением двух



Р и с. 3. Квадрат

пространств, отвечающих сторонам треугольника, лежащего под диагональю. Таким образом это пространство представимо двумя разными способами в виде произведения.

С другой стороны, на этом графе действует группа, связанная с мультипликативной группой нашего поля. А именно, можно взять и всё умножить на f. Тогда пространства переходят друг в друга и, соответственно, скобки тоже как-то отобразятся. Умножение на f на квадрате действует слева направо. И есть отображение, которое действует снизу вверх — умножение на g. Эти отображения переводят одно представление пространства $(g\mathcal{O}|f\mathcal{O})$ в виде произведения в другое и тем самым дают его автоморфизм!

Исходя из этих замечаний мы построим пространство и пучок (локальную систему) на нём, монодромия которого даст нам ручной символ.

Определение будет, конечно, звучать довольно абстрактно, тем не менее, для тех, кто занимается топологией, оно вполне естественно. Наше пространство будет симплициальным множеством $S_\bullet$. Симплициальное множество представляет собой набор множеств симплексов; у него есть вершины S_0, есть рёбра S_1, есть треугольники S_2 и т. д. У нас будут только S_0, S_1 и S_2. Они определяются следующим образом:

$$S_0 = \{A \sim \mathcal{O},\ A \subset K\},$$
$$S_1 = \{A,\ B \sim \mathcal{O},\ x \in (A,\ B)'\},$$
$$S_2 = \{A,\ B,\ C \sim \mathcal{O},\ x \in (A,\ B)',\ \ y \in (B|A)',\ z \in (C|A)',\ 1 = xyz \in (A,\ A)'\}.$$

Здесь мы используем вместо $(A|B)$ множество $(A|B)' = [(A|B) - (0)]/\{\pm 1\}$ по причине, о которой я скажу ниже. Ребро x может быть любым элемент множества $(A|B)'$; заметьте, что любые две вершины могут быть соединены многими рёбрами (например, если k — поле комплексных чисел, то любые две вершины соединены континуальным множеством рёбер). На рёбра x, y, z натянут треугольник, если $xyz = 1$. Никакой топологической структуры на этом множестве я не ввожу. С точки зрения людей, привыкших к обычной топологии, это множество выглядит экзотически. Тем не менее, для тех, кто занимается алгебраической K-теорией, это вполне обычная вещь.

Итак, у нас есть симплициальное множество. Теперь я хочу построить на нем пучок. Пучок будет устроен следующим образом: в каждой вершине $\sigma \in S_0$ будет одномерное пространство *) $\mathcal{F}_\sigma$, а именно, $\mathcal{F}_\sigma = (\mathcal{O}|A)'$, где A — подпространство, соответствующее вершине σ. Рассмотрим две вершины A и B. Слои пучка в этих точках — одномерные пространства

*) Мы используем, для краткости, такую неточную терминологию.

$\mathcal{F}_A = (\mathcal{O}|A)'$ и $\mathcal{F}_B = (\mathcal{O}|B)'$. Свойство $(\mathcal{O}|A)(A|B) = (\mathcal{O}|B)$ показывает, что если имеются подпространства A и B и ребро $x \in (A|B)'$, то элемент x определяет отображение $\mathcal{F}_A \to \mathcal{F}_B$ (умножение на элемент x). Иными словами, над каждым ребром сидит линейное отображение (изоморфизм) одного слоя пучка в другой слой *).

Заметьте, что, если в нашем симплициальном множестве задана петля и вершина на ней, то они определяют линейное отображение слоя (одномерного пространства) над вершиной в себя и, следовательно, определяют число.

Далее, я утверждаю, что можно построить петлю $\gamma(f, g) \in S_\bullet$, относительно которой монодромия пучка $\mathcal{F}$ как раз и есть ручной символ (f, g). Чтобы написать такую петлю, нужно взять точку, из которой петля будет исходить. Пусть это будет точка $\mathcal{O}$. Затем мы будем использовать приведенную выше конструкцию с квадратом, но очень содержательно её изменяя:

$$\begin{array}{ccc} g\mathcal{O} & \xleftarrow{gy^{-1}} & fg\mathcal{O} \\ {\scriptstyle x^{-1}}\downarrow & & \uparrow{\scriptstyle fx} \\ \mathcal{O} & \xrightarrow[y]{} & f\mathcal{O} \end{array}$$

Мы нарисовали четыре вершины и соответствующие рёбра, задающие петлю $\gamma_P(f, g)$. Заметьте, что у петли $\gamma_P(f, g)$ есть дополнительные параметры: x и y можно выбирать произвольно. Ответ не зависит от их выбора. Утверждение состоит в том, что голономия (монодромия) нашего пучка относительно этой петли и есть ручной символ (f, g) **).

Скажем теперь два слова, как возникает закон взаимности. Конструкция, которую мы сейчас получили, чисто локальная: всё делается с одним локальным полем. Дальше нужно взять пространство аделей и построить такое же симплициальное множество $S_\bullet(\mathbb{A})$, с ним связанное. В качестве вершин мы выбираем в пространстве аделей подпространства, соизмеримые с подпространствами $\mathbb{A}(D)$. Здесь $\mathbb{A}(D)$ — любой дивизор на кривой, от его выбора ничего не зависит (можно взять, например, $D = 0$).

Далее, слово в слово так же определяются ребра и такой же пучок $\mathcal{F}$. Всё, что я рассказал для локальной ситуации, переносится и в глобальную ситуацию. Так же строится петля $\gamma_{\mathbb{A}}(f, g)$.

*) Это то, что чисто технически можно назвать связностью.

**) На самом деле так получается не весь ручной символ, а лишь символ с точностью до знака, так как определяя симплициальное множество, мы факторизовали главное однородное пространство $(A|B) - (0)$ по группе $\{\pm 1\}$.Чтобы получить также и знак, нужно более сложная конструкция, которую я здесь не обсуждаю.

Затем нужно доказать (и это отдельная работа, довольно приятная), что $\gamma_{\mathbb{A}}(f, g) = \prod_P \gamma_P(f, g)$, где $f, g \in K^*$, т. е. глобальная петля есть произведение локальных петель. Из этого вытекает, что $\mathrm{Hol}_{\gamma_{\mathbb{A}}}(f, g)(\mathcal{F}) = \prod_P \mathrm{Hol}_{\gamma_P}(\mathcal{F})$, т. е. голономия (монодромия) относительно глобальной петли есть произведение локальных голономий. А локальные голономии — это локальные символы. Таким образом, произведение локальных символов по всем P мы интерпретируем как монодромию пучка на глобальном объекте.

Почему эта монодромия тривиальна? Ведь мы же хотим доказать закон взаимности, т. е. хотим показать, что $\prod_P \mathrm{Hol}_{\gamma_P}(\mathcal{F}) = 1$, если петля задается парой функций из поля K. Ответ очень интересный. Дело в том, что можно написать новый пучок $\mathcal{L}$:

$$\mathcal{L}_{\langle A\rangle} = \mathrm{Det}(A \oplus K \to \mathbb{A}),$$

где $A \subset \mathbb{A}$, $A \sim A(0)$ пробегают вершины симплициального множества $S_\bullet(\mathbb{A})$. На $S_\bullet(\mathbb{A})$ действует мультипликативная группа K^* поля рациональных функций на кривой. Как в локальном случае мы могли преобразовывать скобки, умножая на элементы поля, так и здесь. Тогда:

• пучок $\mathcal{L}$ K^*-эквивариантен, т. е. имеется согласованная система изоморфизмов слоев

$$\mathcal{L}_{\langle A\rangle} \xrightarrow{\times f} \mathcal{L}_{\langle fA\rangle},$$

где $f \in K^*$;

• $\mathcal{L} = \mathrm{Det}\ H^\bullet(C, \mathcal{O}_C) \otimes \mathcal{F}$.

Второе свойство показывает, что монодромии пучков $\mathcal{L}$ и $\mathcal{F}$ совпадают (они отличаются на постоянный пучок). А из первого свойства следует, что пучок $\mathcal{L}$ спускается на фактор-множество $S_\bullet(\mathbb{A})/K^*$. Обозначим через $\pi\colon S_\bullet(\mathbb{A}) \to S_\bullet(\mathbb{A})/K^*$ естественную проекцию. Тогда $\mathcal{L} = \pi^*\mathcal{G}$ и монодромия пучка $\mathcal{L}$ относительно петли $\gamma_{\mathbb{A}}(f, g)$ будет равна монодромии пучка $\mathcal{G}$ относительно петли $\pi(\gamma_{\mathbb{A}}(f, g))$. Последняя равна произведению четырех *петель* $\bar{x}^{-1} \cdot \bar{y}^{-1} \cdot \bar{x} \cdot \bar{y}$, где $\bar{x} = \pi(x)$, $\bar{y} = \pi(y)$ (см. рис. выше), откуда и следует тривиальность монодромии (группа k^* абелева, а произведение петель является коммутатором!) *).

*) Мы получили закон взаимности с точностью до знака. Чтобы его учесть, нужно использовать более тонкую конструкцию, например, ввести ориентацию в одномерные векторные пространства $(A|B)$ или рассматривать суперпространства.

Литература

[ADK] E. Arbarello, C. De Concini, V. G. Kac. The Infinite Wedge Representation and the Reciprocity Law for Algebraic Curves // Proc. Symp. Pure Math., 1989. **49**, 1. P. 171–190.

[AK] Arnold V. I., Khesin B. A., Topological Methods in Hydrodynamics. Springer-Verlag, 1998.

[Hu] Huber A. On the Parshin-Beilinson adeles for schemes // Abhandl. Mathem. Seminar Univ. Hamburg, 1991. P. 249–273.

[PF] Parshin A. N., Fimmel T. Introduction to higher adelic theory. Preprint, 1999.

24 января 2002 г.

А. Б. Сосинский

МОЖЕТ ЛИ ГИПОТЕЗА ПУАНКАРЕ БЫТЬ НЕВЕРНОЙ?

В лекции будет обсуждаться гипотеза Пуанкаре, которая в современной формулировке выглядит следующим образом: если трёхмерное замкнутое компактное многообразие M^3 односвязно, то оно обязательно должно быть сферой: $M^3 \approx S^3$.

Я хочу уточнить класс многообразий, для которых это утверждение делается. Этот класс я обозначу $\mathcal{M}^3$. Здесь рассматриваются замкнутые компактные трёхмерные многообразия. *Замкнутое* будет означать — без края и связное. Указанная импликация верна для любого односвязного $M^3 \in \mathcal{M}^3$. Это называется сегодня *гипотезой Пуанкаре*. Доказательство этой гипотезы является одной из «семи проблем тысячелетия», предложенных Институтом Клея. Положительное решение стоит 1 миллион долларов. Но, чтобы вы меня не заподозрили в алчности, я сразу хочу сказать, что я никогда не пытался эту гипотезу доказывать, а сейчас буду рассказывать почему. А именно, я попытаюсь объяснить, почему я думаю, что её можно опровергнуть. Я не умею этого делать, но в ходе доклада я сформулирую и докажу некоторые простые результаты, которые наводят на мысль, что она действительно неверна *).

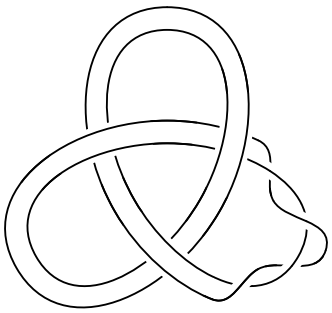
Прежде чем это сделать, я хочу немножко остановиться на истории вопроса, потому что она довольно содержательная. Гипотезу Пуанкаре действительно сформулировал Пуанкаре в 1895 г., но иначе. А именно, он предположил, что если многообразие $M^3 \in \mathcal{M}^3$ является гомологической сферой, то тогда оно является сферой. Что означает «гомологическая сфера»? Гомологическая сфера — это топологическое пространство, которое имеет такие же гомологии, как сфера. (Впрочем, Пуанкаре не знал, что такое гомологические группы; но он прекрасно понимал, что такое гомологии; в каком-то смысле он сам их придумал.) В своей знаменитой книге «Analysis Situs» он высказал именно такое утверждение в виде гипотезы.

*) Между тем, как этот доклад был сделан, и его публикацией появилось доказательство гипотезы Пуанкаре, принадлежащее Г. Перельману. Результаты настоящей статьи имеют вид «Если A, то гипотеза Пуанкаре неверна»; теперь доказано, что «A» неверно, и, тем самым, абсолютный результат об алгоритмической разрешимости одной из проблем топологии. — *Прим. ред.*

Но довольно быстро он сам же и обнаружил, что это неверно, и придумал два замечательных контрпримера, которые показывают, что гомологическая сфера может не быть настоящей сферой.

Первый пример сейчас часто называется *сферой Пуанкаре*. Он основан на склейке додекаэдра. Берётся додекаэдр; потом каждая грань поворачивается на угол $2\pi/10$ и склеивается с противоположной гранью. Получается, как легко понять, некое трёхмерное многообразие M^3_P. Легко найти его гомологии; они действительно такие же, как у сферы. Но это не сфера. Для того чтобы доказать, что это не сфера, Пуанкаре сделал замечательную вещь — он изобрёл фундаментальную группу. (Замечу, что во Франции фундаментальная группа называется группой Пуанкаре.) Для многообразия M^3_P не очень сложно посчитать фундаментальную группу. Оказывается, что она не равна нулю (а у сферы она равна нулю), и поэтому это многообразие не является сферой.

Кроме этого примера Пуанкаре построил ещё и другой, совершенно замечательный пример, основанный на других геометрических идеях. Второй пример в течение сегодняшнего доклада будет играть главенствующую роль. Этот пример строится так. Возьмём какой-нибудь узел в трёхмерной сфере, например, трилистник (рис. 1). Возьмём его трубчатую окрестность, вырежем её и вклеим её обратно, притом определённым образом. Чтобы описать вклейку, достаточно понять, куда попадёт край меридианной полоски (горизонтального сечения) полного тора, который мы будем вклеивать обратно. Он будет вклеиваться следующим образом. Край меридианной полоски приклеивается к краю полоски, изображённой на рис. 1. Такая полоска (или задающее её целое число) называется *оснащением*. На рис. 1 изображена полоска, соответствующая оснащению 1. Если рассмотреть две кривые в трёхмерном пространстве, которые здесь нарисованы, то их индекс зацепления как раз равен 1. Одна из них обматывается вокруг другой ровно один раз. Хотя, казалось бы, всё выглядит слишком сложно, и кривая обматывается много раз. Но если не делать никаких дополнительных оборотов вокруг кривой, а просто провести на рис. 1 «параллельные» линии, то индекс зацепления будет бóльшим.



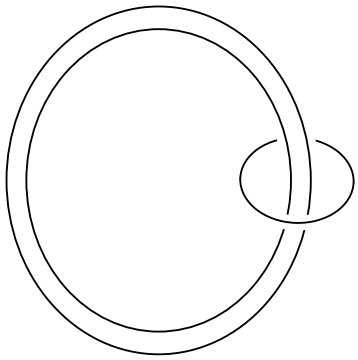
Р и с. 1. Перестройка по узлу

Есть общая теорема, которая утверждает следующее.

Т е о р е м а 1. *Если взять любой узел с оснащением* ± 1, *то тогда при соответствующей переклейке получится гомологическая сфера.*

Так строится второй контрпример, и даже целая серия контрпримеров.

Увидеть это своими глазами трудно. Естественно, мы не можем увидеть никакое трёхмерное многообразие. С другой стороны, я приведу более простой пример. Если взять тривиальный узел (окружность) и оснащение, равное нулю, и сделать соответствующую склейку, то в результате получится $S^1 \times S^2$. Доказательство этого — хорошее упражнение для студентов 1-го курса, которые начинают изучать топологию. (Такие перестройки называют *ортогональными*; мы вклеиваем полноторие как бы наоборот.) Если эту окружность охватить ещё одной окружностью (рис. 2) и там сделать такую же перестройку, то мы вернёмся обратно к трёхмерной сфере.



Р и с. 2. Двойная перестройка

Гипотеза о том, что любая трёхмерная гомологическая сфера является обычной сферой, оказалась неверной. Тогда Пуанкаре, в знаменитом добавлении к книге «Analysis Situs», сформулировал тот вариант своей гипотезы (это было в 1905 г.), с которого я начал эту лекцию. Сегодня его можно переформулировать так: если трёхмерное многообразие является гомотопической сферой, то оно является настоящей сферой.

Трёхмерная гомотопическая сфера — это трёхмерное многообразие, у которого и гомологические и гомотопические группы такие же, как у сферы. На самом деле не нужно требовать, чтобы все гомотопические группы были такие же, как у сферы. Достаточно потребовать, чтобы первая (фундаментальная) гомотопическая группа была такая же, как у сферы, т. е. была равна нулю. Из односвязности следует, что это гомотопическая сфера. Это постепенно доказывалось в начале века. Главную роль в этом сыграла теорема Гуревича, которая утверждает, что фундаментальная группа, если её прокоммутировать, будет первой гомологической группой. Потом нужно ещё пользоваться двойственностями. С 20-х годов проблема стала формулироваться таким совершенно элементарным образом.

Решение этой проблемы стало очень популярным видом спорта среди математиков. Не таким популярным, как теоремы Ферма, но очень много народу её решали, и очень много народу её «решили». Среди моих собственных хороших знакомых есть четыре человека, которые «решили» гипотезу Пуанкаре. Один из них даже несколько раз. Я не буду называть фамилии.

Это действительно нечто очень притягательное, которым многие люди довольно долго занимались. Без всякого успеха. Пока что, на сегодняшний день, насколько мне известно, нет сколько-нибудь серьёзных заявок на то, чтобы она была доказана *).

Эта третья формулировка мне представляется довольно интересной, потому что она естественно обобщается на многомерный случай. Начиная примерно с 30-х годов появилась обобщённая гипотеза Пуанкаре, которая утверждает, что для любого n-мерного многообразия $M^n \in \mathcal{M}^n$ (где $\mathcal{M}^n$ — множество замкнутых компактных n-мерных многообразий) из условия, что M^n — гомотопическая сфера (т. е. у него такие же гомологические и гомотопические группы, как у сферы), следует, что M^n действительно сфера: $M^n \approx S^n$.

Эта гипотеза была доказана в 1960 году, причём в двух вариантах. Смейл доказал эту гипотезу для гладких многообразий размерности $n \geqslant 5$. В том же году Столлингс рассматривал класс не обязательно гладких многообразий, а PL-многообразий (т. е. многообразий, обладающих PL-триангуляцией) при $n \geqslant 7$. Конструкции Смейла и Столлингса существенно различны. Смейл использовал в первую очередь теорию Морса. Попутно он доказал одну из самых фундаментальных теорем теории многообразий, так называемую Handlebody Theorem, которая говорит, что любое многообразие можно получить из пустого множества последовательным приклеиванием ручек. Я не буду давать соответствующие определения, поскольку это не относится к моей теме. Столлингс же доказал обобщенную гипотезу Пуанкаре с помощью принципиально комбинаторного метода, который называется *метод поглощения*.

Наибольший успех был получен в обобщённом случае. Это доказательство привело к ещё большей активности на поприще попыток доказать гипотезу Пуанкаре в размерности три.

Я хочу немножко прокомментировать своё отношение к гипотезе Пуанкаре. Когда она впервые формулировалась, я думаю, что её понимали как первый шаг в вопросе классификации многообразий. Первое, что мы хотим уметь, — это уметь узнавать тривиальное многообразие, самое простое (сферу). Для этого придумали инварианты — гомологии. Считая их, мы можем узнать, сфера у нас или не сфера. Гомологии оказались недостаточными для этих целей. К ним прибавили фундаментальную группу и хотели получить такой же результат, т. е. $H_* M^3 = 0$ и $\pi_1 M^3 = 0$ влечет $M^3 \approx S^3$. (Вместо «односвязность» я буду писать $\pi_1 M^3 = 0$.) Но странным

*) В настоящее время имеется ряд препринтов Г. Перельмана, в которых намечен ход доказательства. (*Добавлено автором при корректуре*.)

образом сам этот результат мало чего даёт. Многие топологи, занимаясь классификацией трёхмерных многообразий, доказывали разные теоремы типа того, что что-то верно по модулю гипотезы Пуанкаре. Таких теорем накопилось довольно много. Но они не слишком интересны и наука мало продвинулась с их помощью в область классификации трёхмерных многообразий. Эта проблема до сих пор стоит. Зато есть замечательная и очень простая теорема Маркова, которая утверждает, что не существует алгоритма, который классифицирует многообразия размерности $n \geqslant 4$, т. е. нет алгоритмической процедуры, которая для любых двух многообразий позволяет установить, например, комбинаторно эквивалентны они или нет.

Однако сравнительно недавно тот вопрос, который я считаю первоначальным и более важным, чем вопрос о справедливости гипотезы Пуанкаре (как распознать сферу), был решён. Рассмотрим следующую проблему: дано многообразие $M \in \mathcal{M}^3$; верно ли, что M гомеоморфно S^3? Это глобальная проблема. Дело в том, что все топологические трёхмерные многообразия триангулируемы (это теорема конца 40-х годов, очень тяжёлая; она принадлежит американскому математику Эдвину Моису). Поэтому компактные трёхмерные многообразия — комбинаторный объект; их можно задавать с помощью слов в фиксированном алфавите. Для этого нужно просто перечислить все вершины, все рёбра, треугольники и тетраэдры — и тем самым закодировать трёхмерное многообразие. Для таких объектов можно ставить алгоритмическую задачу сравнения их со сферой. Эта задача алгоритмически разрешима.

Это результат работы двух математиков, Рубинштейна (Rubinstein) и Томпсон (Thompson). Работу Рубинштейна разобрать достаточно сложно, я, во всяком случае, не сумел. Что касается работы Томпсон (тоже сложной), то её, во всяком случае, разобрал Сергей Матвеев из Челябинска и написал текст, который человек, знакомый с трёхмерной топологией, может прочитать и понять. Так что это вполне установленная теорема. Это была крупная сенсация. И в каком-то смысле идеологически, во всяком случае среди тех людей, которые настроены на алгоритмической волне, этот результат гораздо более интересен, чем утверждение гипотезы Пуанкаре.

Тут естественно возникает такой вопрос. А как обстоит дело с фундаментальной группой? Можно ли так же алгоритмически выяснить, является ли фундаментальная группа тривиальной или нет?

Я напомню классический результат, принадлежащий Петру Сергеевичу Новикову, который состоит в следующем. Рассмотрим множество всех заданий (или, как иногда говорят, копредставлений) конечнопредставимых

групп:

$$\mathcal{G} \ni G = \langle g_1, \ldots, g_n \,|\, r_1 = \ldots = r_k = 1 \rangle.$$

Здесь $g_1, \ldots, g_n$ — образующие группы G, $r_1, \ldots, r_k$ — соотношения. Соотношение — это слово в алфавите, состоящем из букв

$$g_1,\ g_1^{-1},\ \ldots,\ g_n,\ g_n^{-1}.$$

Запись $r_i = 1$ означает, что слово r_i задает тривиальный элемент группы G. Формально такой записи, такому копредставлению, однозначно соответствует некоторая группа. А именно, если мы возьмём свободную группу, натянутую на образующие $g_1, \ldots, g_n$, и профакторизуем её по нормальной подгруппе, натянутой на слова $r_1, \ldots, r_k$, то тогда то, что получится, будет некоторой группой. Это и есть группа $[G]$, заданная копредставлением G.

Можно рассмотреть подмножество

$$\mathcal{G} \supset \mathcal{G}_{\mathcal{M}} = \{G \in \mathcal{G}\colon \text{существует } M^3 \in \mathcal{M}^3,\ \pi_1(M^3) \cong [G]\}.$$

Есть две естественных больших проблемы.

1) Дано копредставление $G \in \mathcal{G}$. Нужно выяснить, верно ли, что $[G] = 0$. Ответ тут такой: эта проблема алгоритмически неразрешима. Это теорема П. С. Новикова, которая известна в литературе как теорема Рабина—Адяна, потому что Адян в 1955 г. доказал некую более общую теорему, которая содержит эту теорему в качестве частного случая, а потом Рабин эту более общую теорему передоказал в 1958 г.

2) Рассмотрим ту же самую задачу на меньшем классе копредставлений. А именно, пусть $G \in \mathcal{G}_{\mathcal{M}}$, т. е. G — копредставление группы, которая на самом деле является фундаментальной группой трёхмерного многообразия. Нужно выяснить, верно ли, что $[G] = 0$. Естественный вопрос: выяснить, эта проблема алгоритмически разрешима или нет. Ответ на этот вопрос не известен. Однако имеет место следующая теорема, очень простенькая, доказанная мной примерно два года назад.

Теорема 2. *Если эта задача тоже алгоритмически неразрешима, то тогда гипотеза Пуанкаре неверна* *).

Я сейчас приведу доказательство. Оно совсем не сложное, хотя и использует много разных вещей из разных частей математики. Прежде чем доказывать эту теорему, я введу некоторые обозначения. Я уже говорил, что все трёхмерные многообразия триангулируемы. Поэтому не составляет никакого труда построить последовательность $M_1, M_2, M_3, \ldots, M_l, \ldots$ всех трёхмерных многообразий. Точнее, существует алгоритм, который

*) Тем самым, если гипотеза Пуанкаре все же доказана, то эта задача алгоритмически разрешима, что довольно удивительно. — *Прим. ред.*

перечисляет все трёхмерные многообразия. Как этот алгоритм построить? Трёхмерные многообразия можно считать триангулированными, т. е. с фиксированной триангуляцией. Нетрудно придумать процедуру, которая перечисляет все конечные трёхмерные полиэдры. Есть простой алгоритм, который позволяет среди всех трёхмерных полиэдров обнаружить, какие из них являются трёхмерными многообразиями. Для этого нужно взять линк каждой вершины, т. е. границу её комбинаторной звезды, и проверить, что это двумерная сфера. То, что это двумерная сфера, проверяется подсчетом эйлеровой характеристики. Таким образом, есть алгоритм, который перечисляет все трёхмерные многообразия, с колоссальным количеством повторений; любое трёхмерное многообразие будет повторяться бесконечное число раз. Это — очень неэкономичный способ перечисления трёхмерных многообразий, существуют гораздо более экономичные способы. Но сейчас мне это неважно; мне важен принципиальный момент — существование алгоритма, перечисляющего все трехмерные многообразия.

Дальше я делаю следующее. Есть алгоритм, который каждому триангулированному многообразию сопоставляет копредставление его фундаментальной группы. Этот алгоритм описан в любом учебнике по алгебраической топологии. Давайте составим бесконечную матрицу многообразий и копредставлений

$$
\begin{array}{cccccc}
M_1 & M_2 & M_3 & \ldots & M_l & \ldots \\
G_1^1 & G_2^1 & G_3^1 & \ldots & G_l^1 & \ldots \\
G_1^2 & G_2^2 & G_3^2 & \ldots & G_l^2 & \ldots \\
G_1^3 & G_2^3 & G_3^3 & \ldots\ldots & & \\
\ldots\ldots & & & & & \\
\ldots\ldots & & & & G_l^m & \ldots \\
\ldots\ldots & & & & &
\end{array}
$$

Она строится так. Напомню, что для групп, заданных образующими и соотношениями, верна теорема Титце, которая утверждает, что два копредставления задают изоморфные группы тогда и только тогда, когда от одного копредставления можно перейти к другому с помощью так называемых преобразований Титце. Я не буду говорить, что такое преобразования Титце. Замечу только, что они конечно определены. Поэтому, как следствие теоремы Титце, получается следующее утверждение: существует алгоритм, который для любого копредставления группы перечисляет все копредставления групп, изоморфных данной группе. Для этого нужно систематически применять преобразования Титце к заданному копредставлению. Далее, в

l-й столбец матрицы записываются последовательно все копредставления фундаментальной группы многообразия M_l (полученные посредством этого алгоритма).

Теперь я доказываю теорему 2. Мне нужно показать, что если проблема распознавания тривиальности групп $[G]$ для $G \in \mathcal{G}$ алгоритмически неразрешима, то гипотеза Пуанкаре неверна. Я буду, естественно, рассуждать от противного. Пусть гипотеза Пуанкаре верна. Тогда я приду к противоречию следующим образом: я предъявлю разрешающую процедуру для ответа на эту проблему распознавания.

Можно сказать так: мы с вами играем в такую игру. Вы мне даёте некоторое копредставление $G \in \mathcal{G}_M$, т. е. копредставление группы, которая является фундаментальной группой некоторого трёхмерного многообразия. Я обязан сказать, тривиальна эта группа или нет.

Как я это делаю? Очень просто. Я запускаю мой алгоритм и нахожу многообразие M_1. Затем, также алгоритмически, я нахожу копредставление G_1^1 и сравниваю это копредставление с заданным копредставлением. (Копредставления сравниваются как два слова: они одинаковы, если состоят из одинаковых букв, написанных в одинаковом порядке; никакого выяснения изоморфности групп я не провожу, а я просто сравниваю два слова.) Конечно, с первого раза мне не повезло; копредставление G_1^1 не годится. Тогда я на время забываю про копредставление G_1^1 (впрочем, сохраняю его в памяти). Вычисляю M_2 и G_2^1; смотрю на копредставление G_2^1. Опять не повезло. Далее понятно, что я буду делать: я буду обходить матрицу. В какой-то момент я обязательно встречу слово G. Почему? Вы мне дали копредставление некоторого трёхмерного многообразия. Значит, где-то это многообразие живёт, потому что мы перечислили все трёхмерные многообразия. Далее, в столбце, который стоит под этим многообразием, расположены все копредставления его фундаментальной группы. Значит, то копредставление, которое вы мне дали, здесь есть. Я найду это копредставление.

Что я делаю дальше? Я получил многообразие, например, M_l. Я беру телефонную трубку, звоню Абигейл Томпсон в университет Калифорнии в Дэвисе и говорю ей, что у меня есть многообразие M_l, оно устроено так-то и так-то и прошу её сказать (используя алгоритм Рубинштейна—Томпсон), сфера это или нет. Она говорит: «Подожди минуточку, я сейчас перезвоню». Перезванивает (лет через 10 000 — алгоритм у неё медленный) и либо говорит, что M_l — сфера, либо говорит, что M_l — не сфера. Притом правильно говорит: алгоритм Рубинштейна—Томпсон работает правильно.

Пусть M_l — это сфера. Тогда $\pi_1(M_l) = 0$. Следовательно, $G = 0$. И я в нашей игре отвечаю, что группа G нулевая. Теперь пусть M_l — это

не сфера. Я напомню, что мы доказываем теорему от противного, т. е. предполагаем, что гипотеза Пуанкаре верна. Поэтому если M_l — не сфера, то согласно гипотезе Пуанкаре $\pi_1(M_l) \neq 0$. Следовательно, $G \neq 0$.

Такое совсем простое доказательство показывает, что из предположения, что проблема тривиальности для $G \in \mathcal{G}$ алгоритмически неразрешима, вытекает, что гипотеза Пуанкаре неверна.

Естественно, когда я доказал эту теорему, я сразу стал смотреть, не могу ли я как-нибудь доказать, что эта проблема действительно алгоритмически неразрешима. В этом я не преуспел. Более того, я убедился в том, что это дело довольно безнадёжное. Сужение класса рассматриваемых групп до класса фундаментальных групп трёхмерных многообразий — очень сильное, и метод доказательства теоремы Адяна—Рабина не проходит.

Однако, как ни странно, теорема 2 отнюдь не является пустым результатом, даже если нельзя доказать её посылку. Дело в том, что для того, чтобы доказывать неверность гипотезы Пуанкаре, мы всегда рассуждаем от противного и поэтому можно предположить, что гипотеза Пуанкаре верна. Это позволяет пользоваться теми утверждениями, которые следуют из гипотезы Пуанкаре.

Вообще говоря, когда есть гипотеза, могут возникнуть разные ситуации. Она может быть неверной и опровержимой, т. е. можно найти доказательство, которое её опровергает. Она может быть верной и доказуемой, т. е. её можно доказать. Каков статус гипотезы Пуанкаре? Она не аналогична, скажем, континуум-гипотезе. Континуум-гипотеза независима от аксиоматики Цермело—Френкеля. Этот факт — теорема Гёделя—Коэна. К аксиоматике Цермело—Френкеля можно добавлять либо отрицание континуум-гипотезы, либо саму эту гипотезу в качестве аксиомы. Это не повлияет на непротиворечивость системы аксиом. Наша ситуация несколько иная, потому что если гипотеза Пуанкаре неверна, то тогда заведомо есть алгоритм, который предъявляет контрпример. Тут одно из двух: либо гипотеза Пуанкаре неверна, и тогда существует алгоритм, предъявляющий контрпример, либо она верна. Третьего не дано. Разумеется, я за кадром оставляю вопрос о том, что, может быть, как доказательство, так и опровержение настолько велики по объему, что они не помещаются, скажем, в нашей вселенной.

Я не объяснил, почему наследники господина Клея платят 1 миллион долларов за решение и ничего не объявили по поводу того, сколько стоит опровержение. Дело в том, что может оказаться, что будет найдено опровержение, которое по существу, концептуально, действительно не будет стоить ломаного гроша.

Я сейчас объясню, каким образом можно опровергать гипотезу Пуанкаре с помощью компьютера. Нужно взять компьютер и заказать замечательную программу SNAPPEA. Я настоятельно советую всем топологам с этой программой поиграть. Это «дружественная» программа, созданная Уиксом (Weeks) под влиянием Тёрстона. Она специально настроена на то, чтобы классифицировать, находить, изучать и считать объёмы гиперболических трёхмерных многообразий по Тёрстону. Программа устроена следующим образом. Вы задаёте этой программе узел и его оснащение (так как мы опровергаем гипотезу Пуанкаре, то оснащение может быть равно только ± 1). После этого программа строит многообразие, которое получается, если осуществить перестройку трёхмерной сферы по этому узлу с этим оснащением. Про это многообразие она сообщает нам следующую информацию:

1. Гиперболическое это многообразие или нет. Эксперименты показывают, что в подавляющем большинстве случаев многообразия получаются гиперболические.

2. Если многообразие гиперболическое, то программа вычисляет его объём. (Есть каноническое понятие объёма гиперболического многообразия.)

3. Программа выписывает образующие и соотношения фундаментальной группы этого многообразия.

Эта программа находится по адресу

`www.geom.umn.edu/software/download/snappea.html`

Кроме того, в другом месте (адрес я, к сожалению, не помню) находятся программы по компьютерной алгебре, которые как бы смотрят, является ли группа, заданная образующими и соотношениями, тривиальной или нет. Эти программы, конечно, иногда «работают вечно», так и не дав ответа, потому что не существует алгоритма, который это выясняет. Но есть алгоритм, который считает, считает и, если группа нулевая, то он рано или поздно (может быть, очень поздно) сообщит, что группа нулевая.

Для поиска контрпримера далее нужно систематически нарисовать все возможные узлы. Для этого не нужно решать проблему классификации узлов; узлы можно задавать и с повторениями. А с повторениями несложно указать алгоритм, рисующий все узлы. Нужно сделать такую программу, и последовательно подставлять её результаты в SNAPPEA.

Но далее с распознаванием сферы есть тонкость. Дело в том, что это не программа. Это алгоритм, который, однако, запустить на компьютере нельзя. По крайней мере, при современном развитии вычислительной техники, это выше способностей какого-либо программиста или какого-либо компьютера. Но программа SNAPPEA сообщает вам, гиперболическое

многообразие или нет. А сфера — не гиперболическое многообразие. Это уже помогает. А потом, если вы обнаружили какое-то сомнительное многообразие, которое имеет тривиальную фундаментальную группу, то тогда можно попытаться руками выяснить, сфера это или нет. Так что тут какой-то элемент творчества, может быть, и возникнет. А может быть, и не возникнет, если вы найдёте многообразие с тривиальной фундаментальной группой и программа SNAPPEA скажет, что оно гиперболическое. Тогда гипотеза Пуанкаре опровергнута.

Довольно понятно, что по этой причине, т. е. по причине того, что можно найти совершенно неконцептуальное опровержение гипотезы Пуанкаре, большой научной ценности с точки зрения Института Клея такой результат не имеет.

Я спрашивал людей, которые находятся рядом с этой программой, делает ли кто-нибудь то, о чём я говорю. Я так понял, что на самом деле нет, но есть исследователи, которые что-то похожее делают, но немножко поумнее, чем тот прямолинейный подход, который я изложил.

Давайте пойдём дальше. Пусть $A \subset \mathbb{N}$ — перечислимое неразрешимое множество. «Перечислимое» означает, что существует алгоритм, который последовательно перечисляет (несущественно, с повторениями или нет) все элементы этого множества. «Неразрешимое» означает, что не существует такого алгоритма, который для какого-то $x \in \mathbb{N}$ сможет правильно ответить на вопрос, принадлежит ли x множеству A или нет. Такие множества существуют. Это едва ли не самый главный факт в теории алгоритмов: существуют перечислимые неразрешимые множества. Это отвечает такой общефилософской ситуации. Если вы что-то ищете на компьютере, и компьютер это не нашёл, и вы долго-долго ищете и знаете, что если эта вещь есть, то он её найдёт, но не знаете, есть она или нет, а он ищет, ищет и не даёт ответа, то вы оказываетесь в глупой ситуации. Вы не знаете: он не нашёл потому, что её вообще нет, или не нашёл потому, что вы не дали ему достаточно времени искать. Эта основная неприятность формализуется следующей теоремой.

Множество трёхмерных многообразий $\{M(n)\}$ назовем *семейством тест-многообразий*, если $\pi_1\big(M(n)\big) = 0$ тогда и только тогда, когда $n \in A$ (и существует алгоритм построения многообразия $M(n)$ для каждого n). Это семейство позволяет тестировать, принадлежит ли точка множеству A или нет.

Имеет место следующее совсем простенькое утверждение.

Теорема 3. *Если для некоторого перечислимого неразрешимого множества A существует семейство тест-многообразий, то гипотеза Пуанкаре неверна.*

Доказательство этой теоремы, в сущности, тавтология. Будем вести доказательство от противного. Предположим, что гипотеза Пуанкаре верна. Тогда проблема из теоремы 2 не может быть алгоритмически неразрешимой, потому что если бы она была алгоритмически неразрешима, то гипотеза Пуанкаре была бы неверна. Значит, эта проблема алгоритмически разрешима. Построим теперь семейство тест-многообразий. Тогда множество фундаментальных групп этого семейства является подмножеством множества фундаментальных групп всех трёхмерных многообразий. А раз проблема алгоритмически разрешима для большего семейства, то она алгоритмически разрешима и для меньшего семейства. Значит, я могу разрешить это неразрешимое множество следующим образом. Вы мне даёте число n, по этому числу n я строю многообразие $M(n)$, затем я выписываю образующие фундаментальной группы. Пользуясь тем, что проблема распознавания тривиальной фундаментальной группы разрешима, я выясняю, фундаментальная группа нулевая или нет. После этого, если фундаментальная группа нулевая, то я вам (правильно) говорю, что число n принадлежит A, а если фундаментальная группа ненулевая, то я вам (правильно) говорю, что число n не принадлежит A. Вот и всё. Это чисто тавтологическая теорема. Но она гораздо лучше, чем теорема 2, потому что она даёт надежду, что можно построить семейство тест-многообразий.

В оставшееся время я кое-что скажу о построении некоего семейства многообразий, которое претендует на эту роль. Я не умею доказывать, что это действительно семейство тест-многообразий. Как оптимист, я, естественно, предполагаю, что это так. Но это предложение может быть неверным. Так что я не знаю, насколько интересно то, что я сейчас расскажу. Но я думаю, что это интересно хотя бы вот почему. Попутно мне придётся объяснить теорему Адяна—Рабина (теорему Петра Сергеевича Новикова). Может быть, не все её знают, а она связана с очень красивой конструкцией, придуманной Рабиным. (Этой конструкции не было явным образом у Новикова и Адяна.) А именно, конструкцией того, что Рабин называет тест-группой.

Пётр Сергеевич Новиков доказал одну из самых великих теорем XX века, состоящую в том, что существует конкретная группа, явно заданная своим копредставлением

$$G_0 = \langle x_1, \ldots, x_n \,|\, r_1 = \ldots = r_k = 1 \rangle$$

с образующими $x_1, \ldots, x_n$ и соотношениями $r_1, \ldots, r_k$, в которой проблема слов неразрешима. Что это значит? Если я напишу произвольное слово в алфавите $x_1, \ldots, x_n, x_1^{-1}, \ldots, x_n^{-1}$, то это может быть единичный элемент

группы или нет. Как это узнать? Естественный способ состоит в том, чтобы играть с этим словом, пользуясь этими соотношениями, вставляя их, вычёркивая их и т. д. Тут возникает та же самая ситуация. Некоторое время поиграл, нуль не получил. А почему? Потому что недоиграл или потому, что это вообще не нуль? Оказывается, что здесь как раз не существует алгоритма, который для этой конкретной группы отвечает на этот вопрос. По поводу доказательства этой теоремы я ничего говорить не буду.

Теорема Рабина состоит в том, что алгоритмическая неразрешимость проблемы тривиальности группы, заданной образующими и соотношениями, является несложным следствием неразрешимости проблемы тождества слов. У Рабина это делается следующим образом. Он строит тест-группу. По каждому слову $w \in G_0$ (т. е. слову в этих образующих) алгоритмически строится копредставление $G_0(w)$, которое обладает следующим свойством: группа $[G_0(w)]$, соответствующая копредставлению $G_0(w)$, тривиальна тогда и только тогда, когда слово w является тривиальным словом в группе G_0. Как это делается? Просто явным образом пишутся образующие и соотношения. Я мог бы написать образующие и соотношения для группы G_0; там хватает шести образующих и примерно тридцати соотношений, очень коротких. Но я не буду это делать; это неинтересно. А копредставление $G_0(w)$ я явно выпишу. Набор образующих группы $G_0(w)$ состоит из исходных образующих $x_1, \ldots, x_n$, ещё одной образующей x_{n+1} и образующих t, a, s, b, c, d. Введём обозначение $u = x_{n+1} w x_{n+1}^{-1} w$. Соотношения в группе $G_0(w)$ следующие: соотношения $r_1 = \ldots = r_k = 1$, которые были раньше, и новые соотношения

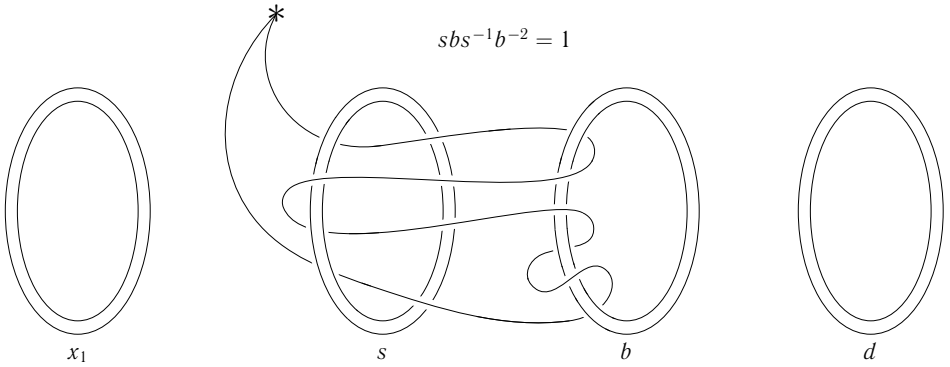
$$ut = t^2 u, \quad us = s^2 u, \quad ta = a^2 t, \quad sb = b^2 s, \quad a = c,$$
$$x_i b^i a b^{-1} = d^i c d^{-1} \quad \text{для } i = 1, \ldots, n+1,$$
$$b^{n+2} a b a^{-1} b^{-n-2} = d^{n+2} c d c^{-1} d^{-n-2}.$$

Совсем несложно доказывается, что это действительно тест-группа, т. е. она нулевая в том и только том случае, если слово w является тривиальным элементом группы G_0. В самом деле, пусть w — тривиальный элемент группы G_0. Подставим его в выражение для u и получим, что $u = 1$. Значит, $t = 1$ и т. д. Я не буду это детально разбирать; это совсем просто. Несколько более сложно доказать обратное утверждение: если элемент w не нуль, то тогда группа тоже ненулевая. Но это несложно: нужно поиграть с этими соотношениями.

Группа G_0 — очень знаменитая группа. Ей занимался такой известный алгебраист, как Хигман (Higman). Она обладает целым рядом замечательных свойств. Это некоммутативная группа с двумя образующими. Люди

много играли с такими вещами в 30-е и 40-е годы, когда было модно изучение некоммутативных групп просто как вид спорта.

Теперь я предлагаю некоторую конструкцию. Я буду строить некоторое семейство многообразий $\{MG(w)\}$. Строить я их буду следующим образом. Я начну с той же самой группы $G(w)$, с которой начинал Рабин. Я посмотрю на эту группу и на её образующие. Я начну с того, что нарисую вертикальные окружности (точнее, узенькие полнотория), соответствующие образующим (рис. 3). Первым делом я вырежу эти полнотория и сделаю



Р и с. 3. Построение многообразия

ортогональную перестройку: вклею их обратно. Тогда то, что получится, будет иметь фундаментальную группу, изоморфную свободной группе с соответствующим количеством образующих. Действительно, нетрудно видеть, что одна такая перестройка даёт $S^1 \times S^2$, а когда окружности далеко друг от друга, то получится просто связная сумма конечного числа экземпляров $S^1 \times S^2$. Затем для каждого из выписанных выше соотношений я построю некоторую трубку. На рис. 3 такая трубка нарисована для соотношения $sbs^{-1}b^{-2}=1$. Звёздочка — это начальная точка всех петель. Я считаю, что одна из образующих фундаментальной группы дополнения к этим трубкам — это обход вокруг этой трубки. Проход под окружностью s соответствует образующей s, проход над окружностью s — образующей s^{-1}. Такие картинки я нарисую для всех соотношений. Это я буду делать в некотором смысле аккуратно, чтобы совсем сильно все эти трубки не перепутались дополнительно между собой. Все эти трубки я ортогонально переклею. На первый взгляд может показаться, что переклейка трубки сразу даст соотношение $sbs^{-1}b^{-2}=1$ в фундаментальной группе многообразия. (Представим себе на минутку, что было только одно соотношение.) К сожалению, это неверно. Не получаются в точности эти соотношения. Получается нечто чуть-чуть модифицированное. Топологи, которые

когда-нибудь считали так называемые копредставления Вертингера узла или зацепления, понимают, как тут нужно действовать. Нужно обозначить отрезки разными буквами, потом написать соотношения, соответствующие каждому перекрёстку, и т. д. Имеется стандартная техника, которая позволяет выписать образующие и соотношения той группы, которая получится. Образующих будет больше, чем в этой группе. Соотношения будут очень похожи на все имеющиеся соотношения, но будут модифицированы. Это всё можно сделать явно. Если задаться целью конкретно выписать эту группу, то можно реально написать все соотношения, которые здесь возникают. Это сложная и достаточно неприятная работа.

У нас получилось некоторое многообразие $MG(w)$, фундаментальная группа $\pi_1(MG(w))$ которого имеет какое-то количество образующих (большее, чем число образующих исходной группы) и какое-то количество соотношений (тоже довольно большое). Вопрос состоит в следующем. Верно ли, что $\pi_1(MG(w)) = 0$ тогда и только тогда, когда $G(w) = 0$? Это вопрос. Зато мы знаем, что $G(w) = 0$ тогда и только тогда, когда $w = 1$.

Вроде бы, у меня есть доказательство в одну сторону (естественно, в лёгкую сторону). В трудную сторону у меня доказательства нет. К сожалению, великолепная наука теория представлений, которой здесь, казалось бы, нужно пользоваться, не помогает. Нам нужно доказать, что некоторая группа — не нуль. Как это обычно делается? Группу нужно как-то представить. Но здесь теория представлений очень мало помогает. В ситуации, когда есть группа, заданная образующими и соотношениями, теория представлений мало что даёт. Поэтому здесь нужно действовать как бы руками.

Я, как оптимист, надеюсь, что это, может быть, и можно доказать. Но очень может быть, что это, например, неверно. Это бы меня очень удивило, потому что группа становится как бы «у́же», чем та, с которой мы стартуем. Но даже если это неверно или это не удаётся доказать, то остаётся ещё такая надежда. Определение тест-многообразия совершенно не привязано непосредственно к теореме Рабина и к теореме Новикова. Может быть, здесь могут сработать совсем другие идеи. Дело в том, что все доказательства алгоритмической неразрешимости чего бы то ни было устроены так: данная проблема сводится к другой проблеме, про которую раньше было установлено, что она алгоритмически неразрешима. Поэтому совершенно не обязательно рассматривать группу и конструкцию Рабина. Может быть, можно использовать какую-нибудь алгоритмически неразрешимую полугрупповую проблему. Какие-то надежды на то, что эта абсолютно детская проблема может оказаться результативной, всё-таки сохраняются.

Моё убеждение таково. Здесь мы находимся на очень тонкой грани между алгоритмической разрешимостью и алгоритмической неразрешимостью разных проблем. Оказалось, что в некоммутативной алгебре, грубо говоря, все содержательные проблемы алгоритмически неразрешимы. С другой стороны, какие-то проблемы трёхмерной топологии алгоритмически разрешимы. Например, чудовищно сложная проблема распознавания трёхмерной сферы алгоритмически разрешима. А другие проблемы неразрешимы. Неверность гипотезы Пуанкаре (если это действительно имеет место) вытекает из тонкого несовпадения между алгоритмической неразрешимостью в алгебре и алгоритмической разрешимостью в геометрии. *)

14 февраля 2002 г.

*) См. примечание на стр. 88. — *Прим. ред.*

С. Алескер

ТЕОРИЯ ПРЕДСТАВЛЕНИЙ В ВЫПУКЛОЙ И ИНТЕГРАЛЬНОЙ ГЕОМЕТРИИ

Основной предмет моего доклада — валюации на выпуклых множествах. Теория валюаций обобщает, с одной стороны, классическую теорию меры, а с другой стороны, теорию эйлеровой характеристики.

Я начну с некоторых формальных определений и обозначений. Пусть V — конечномерное векторное пространство над $\mathbb{R}$, $\mathcal{K}(V)$ — семейство всех выпуклых компактных подмножеств в V.

Определение. *Валюация* — это функционал $\varphi\colon \mathcal{K}(V) \to \mathbb{C}$, который обладает следующим свойством аддитивности:

$$\varphi(K_1 \cup K_2) = \varphi(K_1) + \varphi(K_2) - \varphi(K_1 \cap K_2)$$

для любых K_1, $K_2 \in \mathcal{K}(V)$, для которых $K_1 \cup K_2 \in \mathcal{K}(V)$.

Примеры. 1. Любая мера на V является валюацией.

2. Эйлерова характеристика является валюацией. Действительно, если K — выпуклое компактное множество, то $\chi(K) = 1$. Поэтому свойство аддитивности выполняется.

Позже я дам некоторые другие примеры валюаций, более интересные с точки зрения геометрии. Изучать просто валюации, которые удовлетворяют этому определению, — трудная задача. Естественное ограничение, которое будет наложено, состоит в непрерывности.

Определение. Валюацию φ называют *непрерывной*, если φ непрерывна относительно метрики Хаусдорфа на $\mathcal{K}(V)$.

Метрика Хаусдорфа определяется следующим образом: расстояние между двумя компактами A и B равно

$$d_H(A, B) = \inf\{\varepsilon\colon \varepsilon > 0,\ A \subset (B)_\varepsilon,\ B \subset (A)_\varepsilon\},$$

где $(A)_\varepsilon$ — ε-окрестность A. Легко проверить неравенство треугольника и всё, что положено.

Непрерывность в метрике Хаусдорфа — это очень сильное ограничение на валюацию. Оно гораздо сильнее, чем может показаться на первый взгляд. Но тем не менее, оно выполняется во многих геометрически интересных ситуациях.

Мы будем рассматривать только непрерывные валюации и только трансляционно инвариантные валюации. Трансляционная инвариантность — это, вообще говоря, не необходимое предположение. Изучаются и другие валюации, среди них есть важные. Но для наших целей мы ограничимся классическим случаем трансляционно инвариантных валюаций. Очевидно, что пространство всех непрерывных трансляционно инвариантных валюаций — это линейное пространство, поскольку валюации можно складывать и умножать на скаляры. Это линейное пространство мы будем обозначать $\mathrm{Val}(V)$.

Основные примеры такие.

1. Мера Лебега (объём).

2. Эйлерова характеристика.

Третий пример обобщает эти два примера.

3. Фиксируем множество $A \in \mathcal{K}(V)$ и рассмотрим $\varphi(K) = \mathrm{vol}(K + A)$, где $K + A = \{k + a\colon\ k \in K,\ a \in A\}$ — сумма Минковского выпуклых множеств.

Неявно валюации возникли в 1900 г. в работе Макса Дена, в которой он решил третью проблему Гильберта, доказав неравносоставленность многогранников в трёхмерном пространстве. Если говорить современным языком, то он построил некую валюацию, инвариантную относительно вращений, которая отлична от объёма и принимает различные значения на кубе и тетраэдре одинакового объёма. Потом эта тематика была на какое-то время забыта. В 1930-е годы Вильгельм Бляшке вернулся к этой тематике в связи с проблемами интегральной геометрии. Бляшке считается одним из основателей современной интегральной геометрии, и его мотивировка рассматривать валюации была именно такая. Но сколько-нибудь важных результатов в этой области он не получил. Систематическое исследование валюаций было начато в 1940-е годы Хьюго Хадвигером. Наиболее классические результаты в этой области принадлежат ему. Я упомяну некоторые из них.

Первый результат, который я хочу упомянуть, — это теорема Хадвигера, опубликованная в 1957 г. в его книге. Она классифицирует все непрерывные валюации, инвариантные относительно не только сдвигов, но и вращений. Эта классификационная теорема особенно важна при доказательстве формул интегральной геометрии. Пусть V — евклидово пространство. Тогда любая непрерывная валюация на V, инвариантная относительно всех изометрий, единственным образом представляется в следующем виде: $\varphi(K) = \sum\limits_{i=0}^{n} c_i V_i(K)$, где $n = \dim V$, а V_i — валюации специального вида, которые часто возникают в аффинной и интегральной геометрии. А именно, $V_0 = \mathrm{vol}(D)\chi(K)$, где $\mathrm{vol}(D)$ — объём единичного

шара, $V_n(K) = \mathrm{vol}(K)$, а если $0 < i < n$, то

$$V_i(K) = c \int\limits_{\partial K} \sigma_{n-1-i}(k_1(s), \ldots, k_{n-1}(s))\, ds,$$

где $k_1(s), \ldots, k_{n-1}(s)$ — главные кривизны в точке s, σ_l — элементарный симметрический многочлен степени l, а c — нормализационная константа, зависящая только от n и i, которую мы не будем явно выписывать.

Я сейчас приведу простейший результат, иллюстрирующий применение теоремы Хадвигера. Его доказательство можно получить и без использования этой теоремы Хадвигера, но с её помощью доказательство получается моментально. Это — иллюстрация того, как такие классификационные теоремы возникают в интегральной геометрии.

П р и м е р. Пусть K — выпуклое компактное множество в $\mathbb{R}^n$. Рассмотрим следующий интеграл: $\int\limits_{F \in \mathrm{Gr}_k(\mathbb{R}^n)} \mathrm{vol}_k(\mathrm{Pr}_F K)\, dF$ (усреднённый объём ортогональных проекций множества K на все k-мерные подпространства относительно меры Хаара на грассманиане). Утверждается, что

$$\int\limits_{F \in \mathrm{Gr}_k(\mathbb{R}^n)} \mathrm{vol}_k(\mathrm{Pr}_F K)\, dF = \kappa_{k,n} V_k(K).$$

Это стандартный результат из интегральной геометрии. Есть много более интересных формул такого типа. Все они могут быть легко доказаны с помощью теоремы Хадвигера. Кстати, сам Хадвигер является автором части этих теорем.

Как это доказать? Выражение в левой части равенства — это, очевидно, валюация $\varphi(K)$, которая непрерывна и инвариантна относительно сдвигов и вращений. Следовательно, по теореме Хадвигера она может быть представлена в указанной в этой теореме форме. С другой стороны, $\varphi(\lambda K) = \lambda^k \varphi(K)$ для любого $\lambda > 0$. А среди валюаций V_i только одна имеет данную степень однородности.

Валюации, которые не инвариантны относительно вращений, тоже часто возникают в задачах интегральной и аффинной геометрии. Был вопрос о том, как описать непрерывные валюации, которые трансляционно инвариантны, без предположения о дополнительных симметриях. В каких терминах это можно сделать? Задача описания трансляционно инвариантных валюаций была рассмотрена ещё Бляшке, но точная формулировка гипотезы появилась гораздо позже. А именно, была гипотеза Питера Макмюллена, который её предложил в 1980 г. Это — некая слабая форма того, как описать непрерывные трансляционно инвариантные валюации. Гипотеза состоит в следующем. Линейные комбинации валюаций вида $\mathrm{vol}(K + A)$, где A фиксировано, плотны в пространстве $\mathrm{Val}(V)$.

Оказывается, что эта гипотеза верна. Я сформулирую более сильный результат, который даёт более точное описание пространства валюаций. Но для этого мне понадобится более старый факт о валюациях.

Теорема 1 (Макмюллен, 1977). *Любая валюация* φ *имеет единственное представление в виде* $\varphi = \sum_{i=0}^{n} \varphi_i$, *где* $n = \dim V$ *и* $\varphi_i(\lambda K) = \lambda^i \varphi_i(K)$ *для всех* $\lambda > 0$.

Другими словами, на пространстве всех валюаций есть некая естественная градуировка:

$$\mathrm{Val}(V) = \bigoplus_{i=0}^{n} \mathrm{Val}_i(V).$$

О пространствах $\mathrm{Val}_i(V)$ известны следующие утверждения.

Теорема 2. а) $\mathrm{Val}_0(V) = \mathbb{C} \cdot \chi$, *т. е. все валюации однородной степени* 0 *пропорциональны эйлеровой характеристике*. (Это тривиальный факт.)

б) $\mathrm{Val}_n(V) = \mathbb{C} \cdot \mathrm{vol}$, *т. е. все валюации максимальной однородной степени* n *пропорциональны объёму*. (Этот факт был доказан Хадвигером в 50-е годы.)

в) *Можно описать валюации* $\mathrm{Val}_{n-1}(V)$ *однородной степени* $n-1$. (Я не буду давать точную формулировку этого результата. Он был доказан Макмюлленом в 1980 г. Из полученного им описания сразу ясно, что его гипотеза верна для таких валюаций.)

г) *Валюации* $\mathrm{Val}_1(V)$ *настолько явно описать нельзя, но можно описать некое плотное подпространство. Для них гипотеза Макмюллена верна*. (Это доказали Гуди (Goody) и Вейль (Weil) в 1984 г.)

Ещё одно (тривиальное) наблюдение о структуре этих пространств состоит в том, что любую валюацию можно единственным образом представить в виде суммы чётной и нечётной части: $\varphi = \varphi^{\mathrm{ev}} + \varphi^{\mathrm{odd}}$; здесь $\varphi^{\mathrm{ev}}(-K) = \varphi^{\mathrm{ev}}(K)$ и $\varphi^{\mathrm{odd}}(-K) = -\varphi^{\mathrm{odd}}(K)$. Тем самым у нас получается разложение

$$\mathrm{Val}(V) = \bigoplus_{i=0}^{n} (\mathrm{Val}_i^{\mathrm{ev}}(V) \oplus \mathrm{Val}_i^{\mathrm{odd}}(V)).$$

Последнее наблюдение перед формулировкой первого основного результата состоит в том, что группа $\mathrm{GL}(V)$ естественным образом действует на $\mathrm{Val}(V)$ и сохраняет это разложение. А именно, если есть элемент $g \in \mathrm{GL}(V)$, то мы полагаем $(g\varphi)(K) = \varphi(g^{-1}K)$. Очевидно, что это — некое линейное представление. На пространстве $\mathcal{K}(V)$ есть метрика Хаусдорфа. Относительно этой метрики оно является локально компактным пространством, и там есть топология сходимости, равномерной по компактам.

Т е о р е м а 3 (о неприводимости). *Действие* $\mathrm{GL}(V)$ *на пространстве* $\mathrm{Val}_i^{\mathrm{ev}}(V)$ *и на пространстве* $\mathrm{Val}_i^{\mathrm{odd}}(V)$ *неприводимо, т. е. любое замкнутое инвариантное подпространство — это либо нуль, либо всё пространство.*

Доказательство этой теоремы опубликовано мной в журнале GAFA в 2001 г. Это доказательство довольно нетрадиционное для выпуклой геометрии. Оно использует теорию представлений редуктивных групп, теорию D-модулей, теорию Бейлинсона—Бернштейна.

Из этой теоремы сразу следует гипотеза Макмюллена. Всё пространство разбивается в прямую сумму этих подпространств, поэтому гипотезу Макмюллена достаточно доказать только для них. Но легко видеть, что подпространство, порождённое валюациями такого типа, инвариантно относительно группы $\mathrm{GL}(n)$. Поэтому оно либо нуль, либо всюду плотно. Легко видеть также, что оно не нуль. Из этого немедленно следует гипотеза Макмюллена.

З а м е ч а н и е. Пространство чётных валюаций устроено проще, чем пространство нечётных валюаций. Про чётные валюации можно сказать гораздо больше. Можно явно описать структуру K-типов пространства чётных валюаций. Что это значит? С точки зрения чистой теории представлений это есть некое неприводимое представление группы $\mathrm{GL}(n)$. С точки зрения теории представлений это представление хорошее (как говорят, допустимое). И если мы рассмотрим ограничение представления группы $\mathrm{GL}(n)$ на подгруппу $\mathrm{SO}(n)$, то можно описать разложение этого пространства относительно действия этой группы в терминах старших весов. Оказывается, что кратность каждого представления специальной ортогональной группы не больше 1, и можно явно описать эту структуру. Эта дополнительная информация следует из неких чисто теоретико-представленческих результатов, опубликованных Хау (Howe) и Ли (Lee) в 1999 г. Они, естественно, не занимались никакими валюациями. Они решали некую чисто теоретико-представленческую задачу и провели довольно технические вычисления. (K-типы — это представления максимальной компактной подгруппы.) Эта информация будет использована позже для описания унитарно инвариантных валюаций, которые я сегодня ещё буду обсуждать.

Следующий результат про структуру чётных валюаций, который я хочу рассказать, — это аналог сильной теоремы Лефшеца для валюаций. Рассмотрим на пространстве валюаций оператор $\Lambda\colon\ \mathrm{Val}(V) \to \mathrm{Val}(V)$, который определяется следующим образом:

$$(\Lambda\varphi)(K) = \frac{d}{d\varepsilon}\Big|_0 \varphi(K + \varepsilon \cdot D),$$

где D — единичный евклидов шар. По одной из теорем Макмюллена выражение $\varphi(K+\varepsilon\cdot D)$ является многочленом от ε, если φ — непрерывная трансляционно инвариантная валюация. Поэтому можно взять производную. Нетрудно показать, что Λ понижает степень однородности валюации на 1, т. е. $\Lambda\colon\ \mathrm{Val}_i(V)\to\mathrm{Val}_{i-1}(V)$. Сузим оператор Λ на чётные валюации. Тогда имеет место следующая теорема, которую можно назвать аналогом сильной теоремы Лефшеца.

Теорема 4. *Пусть $i>n/2$, где $d=\dim V$. Тогда отображение*

$$\Lambda^{2i-n}\colon\ \mathrm{Val}_i^{\mathrm{ev}}(V)\to\mathrm{Val}_{n-i}^{\mathrm{ev}}(V)$$

является изоморфизмом, по крайней мере на уровне $\mathrm{O}(n)$-конечных векторов. Другими словами, это отображение не имеет ядра и его образ плотен.

$\mathrm{O}(n)$-конечные векторы — это векторы, орбиты которых относительно максимальной компактной подгруппы содержатся в некотором конечномерном пространстве. Такие векторы составляют как бы алгебраическую основу представления. В каком-то смысле они наиболее существенны.

Основная часть доказательства теоремы 4 основана на моей совместной статье с Иосифом Бернштейном «Range characterization of the cosine transform on higher Grassmanians». В этой работе решается другая задача из интегральной и выпуклой геометрии о так называемом преобразовании косинусов. Есть некое естественное преобразование между функциями на различных грассманианах, которое естественно возникает в геометрии. Впервые это преобразование косинусов изучалось Матероном (Matheron) в 1974 г. Потом был ряд работ, посвящённых описанию образа преобразования косинусов в некоторых частных случаях. В нашей статье найдена связь между преобразованием косинусов и валюациями. В частности, там используется связь с преобразованием Радона. Доказательство теоремы 4 использует также некоторые результаты Гельфанда, Граева и Росу.

Теперь я расскажу о приложениях этих теорем к новым классификационным результатам об унитарно инвариантных валюациях на $\mathbb{C}^n$. Пространство $\mathbb{C}^n$ будет у нас эрмитовым пространством, т. е. мы фиксируем эрмитову метрику. Обозначим $\mathrm{Val}_k^{\mathrm{U}(n)}(\mathbb{C}^n)$ пространство унитарно инвариантных валюаций на $\mathbb{C}^n$ степени однородности k и введём стандартный набор валюаций

$$C_{k,l}(K)=\int\limits_{F\in\mathrm{Gr}_l(\mathbb{C}^n)}V_k(\mathrm{Pr}_F K)\,dF,$$

где $0\leqslant k\leqslant 2n$ и $k/2\leqslant l\leqslant n$. Легко видеть, что это непрерывные трансляционно инвариантные и унитарно инвариантные валюации.

Т е о р е м а 5. *Пусть* $0 \leqslant k \leqslant 2n$. *Тогда валюации* $C_{k,l}$, *где* $\frac{\max\{k,\ 2n-k\}}{2} \leqslant l \leqslant n$, *являются базисом в пространстве* $\mathrm{Val}_k^{\mathrm{U}(n)}(\mathbb{C}^n)$.

В частности, если индекс l меняется в других пределах, то соответствующие валюации по этой теореме выражаются как линейные комбинации таких валюаций, и каждый такой результат есть некая интегрально-геометрическая формула. Это формулы Черна и Сантало.

Из теоремы 5 можно получить ряд приложений к формулам интегральной геометрии. Я сформулирую одно из таких приложений.

Т е о р е м а 6. *Пусть* $K \subset \mathbb{C}^n$ *— выпуклое компактное множество с гладкой границей* ∂K. *Тогда мера Хаара всех комплексных подпространств размерности* i *в* $\mathbb{C}^n$, *пересекающих данное множество* K, *выражается следующим образом*:

$$\mathrm{mes}\{F \in \mathrm{Gr}_i(\mathbb{C}^n) \colon F \cap K \neq \varnothing\} = \sum_{\max\{i,\, n-i\} \leqslant l \leqslant n} \alpha_l C_{2(n-i),l}(K).$$

Если бы мы рассматривали все вещественные подпространства, пересекающие данное выпуклое множество, то по теореме Хадвигера мы получили бы некое выражение V_i. А сейчас, когда мы рассматриваем комплексные подпространства, пересекающие данное выпуклое множество, можно написать это выражение в виде линейной комбинации валюаций $C_{k,l}$. Этот результат сразу следует из классификационной теоремы 5, поскольку то, что написано в левой части, является унитарно инвариантной валюацией. Получается, что доказательство этой элементарной формулы интегральной геометрии сильно неэлементарно и значительно использует теорию представлений, в частности, теорию представлений редуктивных групп, в то время как здесь рассматриваются только унитарно инвариантные валюации. Я, вообще говоря, не могу посчитать явно коэффициенты α_l. Только для $n = 2$, в $\mathbb{C}^2$, это можно сделать руками. Посчитать эти коэффициенты, по-видимому, интересно, но я не знаю, как это сделать.

Есть ещё один пример унитарно инвариантных валюаций, который возникает не из выпуклой геометрии и не из интегральной геометрии. Он возникает из комплексного анализа. Впервые, насколько я знаю, он был рассмотрен в 1982—84 гг. в двух работах Бориса Казарновского. Мотивировкой этого примера была не теория валюаций, а некие обобщения теоремы Давида Бернштейна и Кушниренко о нулях многочленов. Теорема Бернштейна—Кушниренко считает число нулей системы полиномиальных уравнений от n неизвестных в терминах многогранника Ньютона этой системы. Обобщение, доказанное Казарновским, считает асимптотическое число нулей системы экспоненциальных сумм. Для каждой конечной экспоненциальной суммы определяется многогранник Ньютона, и ответ

давался в терминах псевдообъёмов. Если в вещественном случае, в теореме Бернштейна—Кушниренко, многогранник Ньютона лежал в вещественном пространстве $\mathbb{R}^n$, то для экспоненциальных сумм многогранник Ньютона, определённый Казарновским, лежит в комплексном пространстве $\mathbb{C}^n$. И если в качестве ответа в теореме Бернштейна—Кушниренко рассматривался объём или смешанный объём многогранника Ньютона, то комплексный аналог — это псевдообъём многогранника Ньютона, который определялся следующим образом. Пусть K — компактное выпуклое множество в $\mathbb{C}^n$. Определим его опорную функцию следующим образом:

$$h_K(x) = \sup_{y \in K} \langle x, y \rangle.$$

Определим $P(K)$ так:

$$P(K) = \int\limits_{D \subset \mathbb{C}^n} (dd^c h_K)^n = c \int\limits_{D \subset \mathbb{C}^n} \left(\frac{\partial h_k}{\partial z_i \partial \bar{z}_j} \right) \mathrm{vol}.$$

Нужно пояснить, что здесь написано. Когда мы имеем дело с многогранниками, опорная функция негладкая, и нужно определить, что это такое. Но выражение в первом интеграле можно определить как меру и записать в координатах (второй интеграл). Стандартный результат из теории плюрисубгармонических функций комплексного переменного состоит в том, что можно определить это выражение и доказать, что так определённый псевдообъём непрерывен. Псевдообъём P является унитарно инвариантной валюацией степени однородности n (половинной степени однородности): $P \in \mathrm{Val}_n^{U(n)}(\mathbb{C}^n)$. Тем самым, из классификационной теоремы об унитарно инвариантных валюациях можно вывести следующую интегрально-геометрическую формулу для псевдообъёма.

Теорема 7. $P = \sum\limits_{l=[n/2]}^{n} \beta_l C_{n,l}$.

Эта формула опять-таки совершенно элементарна, но её доказательство использует более или менее всё, что известно о чётных валюациях: $\mathrm{GL}(n)$-структуру, теоремы о K-типах, преобразование косинусов и аналог сильной теоремы Лефшеца. Я не знаю другого доказательства этой формулы.

Другая проблема состоит в том, что я, вообще говоря, не могу посчитать коэффициенты β_l, кроме случая $n = 2$.

Теперь я расскажу идею доказательства основной теоремы о неприводимости, которая используется во всех других результатах. Я расскажу, как там возникают D-модули; теория представлений понятно, как возникает. Напомню, что нам нужно доказать, что действие группы $\mathrm{GL}(n)$

на $\mathrm{Val}_i^{\mathrm{ev}}(V)$ и на $\mathrm{Val}_i^{\mathrm{odd}}(V)$ неприводимо. Для простоты рассмотрим случай $\mathrm{Val}_i^{\mathrm{ev}}(V)$ (случай чётных валюаций), который чуть проще. У нас есть некоторое представление. Хотелось бы использовать стандартные методы теории представлений, чтобы его изучать. Сначала нужно сделать это представление более стандартным. Оказывается, что можно построить два различных коммутирующих с действием $\mathrm{GL}(n)$ вложения $\mathrm{GL}(n)$-модуля $\mathrm{Val}_i^{\mathrm{ev}}(V)$:

$$\begin{array}{ccc} \mathrm{Val}_i^{\mathrm{ev}}(V) & \longrightarrow & \Gamma(\mathbb{P}(V),\, T) \\ \downarrow & & \\ \Gamma(\mathrm{Gr}_i(\mathbb{R}^n),\, L) & & \end{array}$$

Здесь $\Gamma(\mathbb{P}(V),\, T)$ — пространство непрерывных сечений некоторого конечномерного расслоения T над проективным пространством, а $\Gamma(\mathrm{Gr}_i(\mathbb{R}^n),\, L)$ — пространство непрерывных сечений некоторого линейного расслоения L над многообразием Грассмана. Что нам это даёт? Оба эти представления более стандартны с точки зрения чистой теории представлений, потому что они индуцированы из некоторых конечномерных представлений параболических подгрупп. Из вложения в $\Gamma(\mathbb{P}(V),\, T)$ можно заключить, что функциональная размерность нашего пространства не больше, чем функциональная размерность пространства $\Gamma(\mathbb{P}(V),\, T)$. Другими словами, $\dim \mathbb{P}(V) = n - 1$, поэтому пространство $\mathrm{Val}_i^{\mathrm{ev}}(V)$ зависит не более чем от $n - 1$ параметра. Это в каком-то смысле маленькое пространство, хотя и бесконечномерное. Более формально это можно сказать так. Можно рассмотреть модуль Хариш-Чандры этого представления, т. е. заменить всё пространство на $\mathrm{O}(n)$-конечные векторы. На нём действует алгебра Ли $\mathfrak{gl}_n$. Размерность Гельфанда—Кириллова этого модуля это и есть функциональная размерность этого пространства. Как только мы установили, что размерность Гельфанда—Кириллова не больше $n - 1$, мы можем забыть про вложение в $\Gamma(\mathbb{P}(V),\, T)$ и рассматривать только вложение в $\Gamma(\mathrm{Gr}_i(\mathbb{R}^n),\, L)$.

Следующий шаг доказательства состоит в том, чтобы доказать, что в пространстве $\Gamma(\mathrm{Gr}_i(\mathbb{R}^n),\, L)$ есть не более одного неприводимого подфактора с такой маленькой размерностью Гельфанда—Кириллова. Из этого будет следовать, что представление $\mathrm{Val}_i^{\mathrm{ev}}(V)$ неприводимо. Как это делается? Как считать размерность Гельфанда—Кириллова? Тут используется понятие, которое было введено около 20 лет назад Иосифом Бернштейном, понятие ассоциированного многообразия для модуля Хариш-Чандры. Это есть некое алгебраическое подмногообразие в нильпотентном конусе над

многообразием нильпотентных матриц, и его размерность равна размерности Гельфанда—Кириллова. Чтобы описать ассоциированное многообразие (геометрический объект), нужно воспользоваться теорией локализации Бейлинсона—Бернштейна. А именно, заменить модуль Хариш-Чандры неким пучком D-модулей на комплексном многообразии Грассмана, глобальные сечения которого — это векторы модуля Хариш-Чандры. Преимущество этого подхода состоит в том, что мы можем использовать теорию пучков, локальные методы. Связь ассоциированного многообразия с этим D-модулем состоит в следующем. У каждого D-модуля есть сингулярный носитель. Образ этого сингулярного носителя при отображении моментов (которое я не буду определять) равен ассоциированному многообразию. Теперь осталось посчитать сингулярный носитель, но это делается уже более или менее руками; это несложная задача из алгебраической геометрии. Можно доказать, что размерности всех сингулярных носителей, кроме одного, очень велики, и их образы при отображении моментов очень большие. Нужная нам размерность может быть не более, чем у одного подмодуля. Подфакторы нашего представления соответствуют подмодулям в этом D-модуле. Это доказывает теорему в чётном случае.

Инъективность вложения $\mathrm{Val}_i^{\mathrm{ev}}(V)$ в пространство $\Gamma(\mathrm{Gr}_i(\mathbb{R}^n), L)$ доказал в основном Клайн (Kline) в 1995 г. Нечётный случай чуть более технический. Вложение в $\Gamma(\mathbb{P}(V), T)$ более или менее такое же, а во втором вложении вместо многообразия Грассмана возникают частичные флаги. Инъективность вложения следует из результатов Шнайдера (Schneider), тоже полученных в 1995 г. В этом случае вычисления с D-модулями производятся не на многообразии Грассмана, а на частичных флагах.

21 февраля 2002 г.

М. А. Цфасман

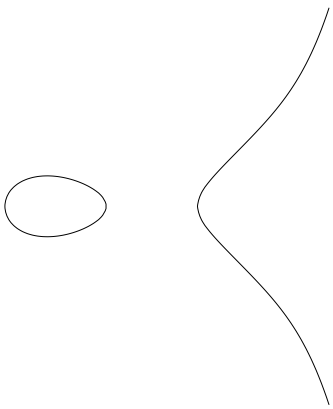
ГЕОМЕТРИЯ НАД КОНЕЧНЫМ ПОЛЕМ

Я попробую в очередной раз сделать невозможное — осуществить основной замысел семинара «Глобус» и рассказать то, что я собираюсь рассказать, так, чтобы это было всем понятно. То, что это у меня не получится, это мне ясно априори. Но вот насколько сильно не получится, это мы увидим. Поэтому я заранее прошу извинения за то, что (особенно в начале) будут вещи, которые многие присутствующие знают очень хорошо. А я их, мало того что буду объяснять, но и буду объяснять довольно приблизительным образом.

Когда мы начинаем изучать геометрию в курсах, скажем, дифференциальной геометрии или топологии (я уж не говорю про геометрию в школе), то всегда геометрический объект мыслится для нас с некоторой протяжённостью. Когда я говорю *кривая*, то обычно имеется в виду нечто, обладающее следующими интуитивными особенностями: во-первых, она одномерна, во-вторых, она непрерывна, и в-третьих, на ней можно выбрать направление и в этом направлении двигаться (в каком смысле двигаться — это отдельный вопрос). В той ситуации, о которой я буду рассказывать сегодня, заведомо ничего этого не получится. А именно, моя цель такова: взять конечное поле $\mathbb{F}_r$ из r элементов, где r — степень простого числа (иначе поля не существует), и попробовать изучать системы алгебраических уравнений над этим полем. Например, изучать одно уравнение $f(x, y) = 0$. Именно про это уравнение мне бы хотелось сказать, что оно определяет плоскую кривую над полем из r элементов.

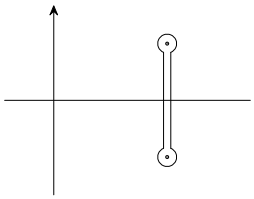
Первое, что мы замечаем, это то, что как эти уравнения ни писать, решений у них будет конечное число по той простой причине, что у нас есть только r возможностей для x и r возможностей для y. Поэтому кривая — это выбор из r^2 точек некоторого количества точек. И геометрии мы не видим совсем. Однако геометрия там присутствует. Но переход к этой геометрии происходит в несколько этапов. Первый этап состоит в том, что мы резко уменьшаем количество возможностей. Когда мы рисуем кривую на плоскости, то обычно имеется в виду либо непрерывное отображение, либо дифференцируемое отображение; их очень много. К сожалению,

единственное, для чего сегодня существует разумная теория над другими полями и особенно над конечными полями, — это ситуация, когда кривая задаётся алгебраическим уравнением, т. е. плоская кривая есть множество нулей уравнения $P_n(x, y) = 0$, где $P_n(x, y)$ — полиномом степени n от двух переменных. Здесь сразу возникает вопрос: «Каких нулей?» Потому что даже во вполне классической ситуации очень естественно переходить от поля вещественных чисел (даже если коэффициенты самого полинома вещественны) к полю комплексных чисел. Это нужно в первую очередь потому, что над полем комплексных чисел многие вещи проще формулируются. В частности, любой полином от одной переменной степени выше 1 имеет корень и разлагается на линейные множители. Над $\mathbb{R}$ это уже не так. И, если вы вспомните, на первом курсе университета нас учили предмету, который назывался *аналитическая геометрия*. И там была классификация конических сечений. При изучении конических сечений возникают ситуации, когда вещественные точки пропадают. С одной стороны, первая мысль — от этого нужно избавиться, и введением поля комплексных чисел мы от этого избавляемся: сразу появляется некоторая униформность теории. А с другой стороны, избавляться от этого очень жалко, потому что нас далеко не всегда интересуют только комплексные решения, а часто именно вещественные. Тем не менее, к одному мы должны привыкнуть. Давайте я приведу пример, который для многих является стандартным: рассмотрим кубическую кривую $y^2 = x^3 + ax + b$. Тогда, с одной стороны, если у полинома справа есть три вещественных корня, то кривая выглядит так, как на рис. 1: у кривой есть овал и уходящая на бесконечность часть. А если вещественный корень один, то овала нет. С другой стороны, мы знаем (сейчас я не буду объяснять, почему), что если мы рассмотрим комплексные решения этого уравнения, то это будет тор с одной выколотой точкой, которая соответствует бесконечному решению. Тут сразу возникает разнобой в терминологии, потому что этот объект, который называется римановой поверхностью, для алгебраических геометров называется алгебраической кривой. Мы, конечно, скажем, что этот объект одномерный, по следующей причине. У нас есть над полем комплексных чисел плоскость — двумерное образование, и одним уравнением мы из него высекли одномерное образование, которое естественно называть кривой.



Р и с. 1. Кубическая кривая

Говоря про вещественный и комплексный случай, нужно сказать ещё вот что. Давайте посмотрим на стандартную картинку (рис. 2). На этом рисунке изображена аффинная прямая над полем комплексных чисел. Если я рассматриваю её над полем вещественных чисел, у неё есть точки: вещественная ось. Кроме того, у неё есть комплексные точки, которые в каком-то смысле не лежат над основным полем, и мы с этим столкнёмся в других ситуациях. Но если есть какая-то комплексная точка, то есть и комплексно сопряжённая с ней. И правильно сделать следующее. Если меня интересуют вещественные вопросы, то правильно говорить, что над полем вещественных чисел эти две точки (образующие орбиту относительно группы порядка 2, порождённой комплексным сопряжением) — это одна комплексная точка степени 2. Эта терминология очень полезна, потому что с точки зрения поля вещественных чисел эти две сопряженные точки никак не различаются. Если у вас есть полином с вещественными коэффициентами, имеющий комплексный корень, то он имеет и комплексно сопряжённый корень. Без каких-то дополнительных структур отличить один корень от другого нельзя.



Р и с. 2. Аффинная прямая над $\mathbb{C}$

Итак, первый шаг, который мы сделали, — мы перешли от геометрии вообще к алгебраической геометрии, т. е. к геометрии чего-то, задаваемого алгебраическими уравнениями. Второй шаг: мы сказали, что всегда, независимо от того, над каким полем мы изучаем основной объект, нас интересуют решения соответствующих уравнений не только над этим полем, но и над всеми его алгебраическими расширениями. Поле $\mathbb{C}$ было выбрано не просто так, а потому, что оно есть алгебраическое замыкание поля $\mathbb{R}$. Если же у меня исходные полиномы с рациональными коэффициентами, то мне нужно алгебраически замкнутое поле, содержащее рациональные числа, т. е. либо поле комплексных чисел, либо просто алгебраическое замыкание поля рациональных чисел. И в этой ситуации, когда коэффициенты рациональны, у нас будут точки из $\mathbb{Q}$ — это точки степени 1; будут точки типа корня $\sqrt{2}$, который сопряжён $-\sqrt{2}$ (эти точки сопряжены, потому что они являются корнями одного и того же неприводимого многочлена с рациональными коэффициентами) — это будет точка порядка 2. Наряду с этим могут быть точки порядка 3, и точки порядка 4 и т. д. Любой неприводимый многочлен степени n задаст n точек, которые объединены в одну орбиту относительно действия группы Галуа замыкания поля над самим полем, и эта орбита будет точкой степени n. Иными словами, неприводимый полином степени n задаёт некую точку степени n.

Эта же ситуация будет и на кривых и на более сложных алгебраических многообразиях. Первая вещь, которую я делаю, состоит в том, что для того, чтобы изучать что-то над конечным полем, я вкладываю это поле в его алгебраическое замыкание: $\mathbb{F}_r \subset \overline{\mathbb{F}}_r$. При этом $\overline{\mathbb{F}}_r = \bigcup \mathbb{F}_{r^n}$. Это специфика конечных полей. Алгебраическое замыкание конечного поля — это объединение всех его конечных расширений, а каждое конечное расширение поля $\mathbb{F}_r$ состоит из r^n элементов, и число n однозначно задаёт это поле. Это связано со следующим обстоятельством: поле $\mathbb{F}_{r^n}$ задаётся как множество корней уравнения $x^{r^n} - x = 0$ над полем $\mathbb{F}_r$. Тем самым, если мы фиксировали алгебраически замкнутое поле, то его подполе $\mathbb{F}_{r^n}$, состоящее из r^n элементов, задано в нём однозначно. Подполя $\mathbb{F}_{r^n}$ вложены друг в друга в следующем смысле: $\mathbb{F}_{r^n} \subset \mathbb{F}_{r^m}$ тогда и только тогда, когда n делит m. Это видно из того, на что делится многочлен $x^{r^m} - x$. Если n делит m, то один многочлен делится на другой, а если n не делит m, то нет.

Здесь мы сталкиваемся ещё с одним феноменом, который довольно существен. Наше поле $\mathbb{F}_r$ не только не является полем вещественных чисел и не только не алгебраически замкнуто, но ещё имеет конечную характеристику. У нас $r = p^l$, где p — простое число, и $px = 0$ для любого элемента $x \in \mathbb{F}_r$. В частности, на всех геометрических объектах, которые мы будем рассматривать, действует гомоморфизм Фробениуса, который действует на координаты следующим образом: $x \mapsto x^r$. Такое же отображение действует на любом пространстве, например, на аффинном пространстве $\mathbb{A}^n$. Оно же действует и на кривых, заданных уравнениями с коэффициентами в поле $\mathbb{F}_r$. Неподвижные точки гомоморфизма Фробениуса — это элементы поля $\mathbb{F}_r$, а неподвижные точки его n-й степени — это элементы поля $\mathbb{F}_{r^n}$.

Поясню это чуть подробнее. Я определил действие гомоморфизма Фробениуса на элементе поля $\overline{\mathbb{F}}_r$. Это действие распространяется до действия на несколько координат: каждая координата возводится в r-ю степень.

Теперь давайте поговорим, что такое аффинное пространство. Оказывается, что даже если нас интересует только вопрос о том, какие есть решения над исходным полем $\mathbb{F}_r$, рассматривать всё нужно всегда вместе с замыканием. Поэтому, когда я пишу аффинное пространство $\mathbb{A}^n$, подразумевается пространство $\overline{\mathbb{F}}_r^n$, на которое я смотрю под специальным углом зрения. Я не просто рассматриваю пространство $\overline{\mathbb{F}}_r^n$, но для каждой точки я беру её орбиту: есть точки степени 1, точки степени 2, точки степени 3 и т. д. Иными словами, когда я говорю об аффинном пространстве, то теоретико-множественно это $\overline{\mathbb{F}}_r^n$, но на самом деле там есть ещё действие группы Галуа поля $\overline{\mathbb{F}}_r$ над полем $\mathbb{F}_r$, которое объединяет точки в орбиты.

Объединяются точки, которые сопряжены над полем $\mathbb{F}_r$. Точками $\mathbb{A}^n$ являются орбиты. При этом у каждой точки есть инвариант — её степень.

Здесь есть два языка, и вы можете говорить на любом из них. На самом деле, полезно говорить на обоих языках сразу. Один язык: точками являются настоящие точки, определённые над алгебраическим замыканием, но вы всегда помните, что на них действует группа Галуа. Другой язык: точками являются орбиты относительно действия группы Галуа. Если вы встречаете выражение «Точка степени 5», то, значит, мы говорим на втором языке. А если вы встречаете выражение «Точка A сопряжена с точкой B», то, значит, мы говорим на первом языке.

Необходим ещё и следующий шаг, который необходим и в классической геометрии. Он связан с тем, что очень неудобно рассматривать некомпактные объекты. Поэтому кривые естественно каким-то образом компактифицировать. Компактификация в алгебраической геометрии производится разными способами. Простейший способ такой: аффинное пространство $\mathbb{A}^n$ вкладывается в проективное пространство $\mathbb{P}^n$. Соответственно, кривые (и поверхности), лежащие в $\mathbb{A}^n$, компактифицируются: они каким-то образом продолжаются в проективное пространство. А именно, многочлен от двух переменных, который задаёт кривую, делается однородным посредством введения третьей переменной. Рассматривая переменные с точностью до пропорциональности, получаем кривую в проективном пространстве.

Итак, вместо того чтобы рассматривать кривые как конечные множества точек, я их пополнил. Точек стало гораздо больше. Теперь это можно рисовать как непрерывную кривую, хотя на самом деле это условность, потому что никакой непрерывности на этом поле нет. Но, по крайней мере, точек уже бесконечно много. В частности, если у уравнения было только одно решение над исходным полем, то до того, как я перешёл к алгебраическому замыканию, я просто никак не мог различить две такие кривые. А теперь через одну точку уже проходит очень много разных кривых, у которых все остальные точки не определены над основным полем.

Но это ещё не всё. В геометрии есть ещё целый ряд полезных понятий. Например, есть понятие касательной. Далее, странно говорить о кривых, не говоря о функциях на этих кривых. Соответственно, нужно понять, с какими функциями мы имеем дело. В алгебраической геометрии поступают следующим образом. Мы замечаем такое обстоятельство: когда есть кривая, то, какие бы функции мы на ней ни рассматривали, если мы зададим дополнительное условие, что функции обращаются в нуль в некоторой заданной точке, то эти функции образуют идеал (функция,

умноженная на что-то, продолжает обращаться в нуль). Более того, этот идеал оказывается максимальным.

Сначала я должен сказать, что такое функция на алгебраической кривой. Ничего кроме полиномов у нас нет, просто потому, что мы не умеем работать не с полиномами. Поэтому если у нас есть произвольное поле k, и над ним задано проективное или аффинное пространство (не очень важно, какое именно), и задана система полиномиальных уравнений, которая в $\mathbb{A}^n$ определяет подмногообразие $X \subset \mathbb{A}^n$, то функции на X определяются следующим образом: нужно рассмотреть всевозможные отношения $\frac{P_1(x_1, \ldots, x_n)}{P_2(x_1, \ldots, x_n)}$ с точностью до уравнений, задающих X. Иными словами, функции P_1/P_2 и Q_1/Q_2 равны, если $P_1Q_2 - P_2Q_1$ делится на какую-нибудь линейную комбинацию уравнений, задающих многообразие X. Это очень естественное определение, потому что, если что-то делится на уравнение, то его значения в точках X нулевые. Поэтому задача состоит в том, чтобы написать такую систему функций, которая точки X различала бы, но при этом не было бы избыточных функций.

Итак, многообразию X ставится в соответствие поле $k(X)$, называемое полем рациональных функций на X. Если есть некоторое подмножество $U \subset X$ (подмножества разумно рассматривать не любые, а только дополнения до меньших алгебраических многообразий), то U можно сопоставить набор функций, которые не имеют на U полюсов. Эти функции называются *регулярными* на U; множество регулярных функций обозначается $k[U]$.

Оказывается, что почти все геометрические понятия, которые мы используем в интуитивной геометрии, или в геометрии дифференциальной, пересказываются в алгебраических терминах. Я скажу, например, что такое касательное пространство. Во первых, что такое точка? Точке $x \in U$ однозначно соответствует максимальный идеал m_x в кольце $k[U]$ — это идеал тех функций, которые обращаются в нуль в точке x. Идеал m_x максимален, поэтому фактор $k[U]/m_x$ — это поле. Следовательно, точке x можно сопоставить поле $K = k[U]/m_x$. Это поле, вообще говоря, не совпадает с исходным полем k; оно может быть больше. Степень точки $\deg x$ — это, как выясняется, как раз и есть степень поля K над исходным полем k. Оказывается, что касательное пространство $T_{x,X}$ есть не что иное, как $(m_x/m_x^2)^*$. Это не удивительно по следующей причине. Если у вас есть касательная прямая, то вы можете рассмотреть дифференциалы функций на этой кривой. Дифференциал функции, вообще говоря, может меняться. Например, если к функции мы прибавили какую-то комбинацию уравнений нашего многообразия, то он изменится. Но его ограничение на

касательное пространство оказывается определённым однозначно. Дифференциал, как мы знаем, убивает константы, а в самой точке не чувствует квадратов: он как бы выделяет линейную часть. Вот мы и выделили у функции линейную часть, и то, что на этом пространстве действует, это и есть касательное пространство. И масса всего другого, что есть в дифференциальной геометрии, пересказывается в геометрию алгебраическую достаточно просто. В каком-то смысле, это связано с тем, что мы сильно ограничили ситуацию. Мы рассматриваем только полиномы. У полиномов есть производная и вообще всё, что нужно.

После этого можно поговорить о вещах более конкретных, связанных именно с конечным полем. Почему нас интересует конечное поле? По нескольким разным причинам. Одна причина: конечное поле — это, в каком-то смысле, простейший пример. Но это не очень хороший аргумент, потому что геометрия над $\mathbb{C}$ проще. Вторая причина более существенна. Многие задачи теории чисел сводятся к задачам геометрии над конечным полем. Например, пусть над кольцом $\mathbb{Z}$ задан полином $P(x, y)$. Вас интересует, есть ли у него нули над $\mathbb{Z}$. Но для этого необходимо, чтобы он имел нули по модулю любого простого числа. Мы рассматриваем редукцию многочлена над полем $\mathbb{Z}/(p) = \mathbb{F}_p$, т. е. превращаем коэффициенты многочлена в элементы поля $\mathbb{F}_p$, и изучаем соответствующую кривую над конечным полем. Наконец, есть ещё третий смысл, который, как мне кажется, самый главный. В алгебраической геометрии в определённый момент оказывается, что у многообразия инвариантов недостаточно. Мы бы хотели знать про многообразия что-то ещё, но только мы не знаем, как вопрос задать. В ситуации, когда основное поле конечно, появляется масса численных инвариантов. Я приведу пару примеров. Один пример простейший. Пусть у вас есть $X/\mathbb{F}_r$ (алгебраическое многообразие X над полем $\mathbb{F}_r$). На нём всегда конечное число точек. Вопрос: «Сколько точек на этом многообразии, т. е. чему равно число $|X(\mathbb{F}_r)|$?» Когда я задаю такой вопрос, я имею в виду точки, определённые над основным полем. По кривой X над произвольным полем, так же, как над полем комплексных чисел, строится её якобиан J_X (я сейчас не буду говорить, что это такое). И можно задать вопрос, сколько точек на якобиане этой кривой (т. е. чему равно число $|J_X(\mathbb{F}_r)|$). Этот вопрос связан с вопросом об однозначности или неоднозначности разложения на простые в арифметических аналогах этой ситуации. На поверхности может быть только конечное число кривых определённого типа. Поэтому можно задавать, например, вопрос, сколько на данной поверхности существует кривых степени n. И опять это будет число. Это даёт нам некоторую возможность различать алгебраические многообразия. Если у нас были, скажем, многообразия

с целыми коэффициентами (что достаточно часто встречается в жизни), то для того, чтобы их различить, мы можем, например, перейти в конечную характеристику, сделать редукцию, и посмотреть, одинаковое там число точек или разное. Или какие-то другие инварианты совпадают или нет. И этот количественный характер этой науки позволяет ставить массу вопросов, которые невозможно задать в алгебраической геометрии над $\mathbb{C}$.

Чем кривые лучше поверхностей? Это важный момент для всей алгебраической геометрии, и для того, что я буду рассказывать, он тоже важен. Оказывается, что среди кривых есть самые хорошие. Самая хорошая кривая отличается следующим свойством. Она должна быть, во-первых, *неприводимой* (это означает, что она не есть объединение двух других кривых); во-вторых, *гладкой* (мы знаем, что такое касательное пространство; в особой точке касательное пространство имеет размерность не 1, а выше, потому что касательная прямая — это прямая, которая пересекает кривую с кратностью 2 и выше, и прямыми, касательными к кривой в её особенности, заполнена вся плоскость; для гладкой кривой касательная — прямая, для гладкой поверхности касательная — плоскость и т. д.; поэтому мы знаем, что такое гладкая точка); следующее условие: мне хочется, чтобы эта кривая была *полной* (т. е., интуитивно, компактной); для этой цели мы считаем, что кривая всегда лежит в проективном пространстве, и в этом проективном пространстве задаётся какими-то уравнениями без изъятий (потом, вообще говоря, какие-то точки мы можем выкинуть, кривая станет неполной). По любой кривой X мы построили поле $k(X)$ — поле функций на X. Оно строится однозначно. Оказывается, что обратный переход от $k(X)$ к X становится однозначным, если мы хотим получить неприводимость, гладкость и полноту. Для любого поля, которое является полем функций для некоторой кривой, существует единственная гладкая полная неприводимая кривая (с точностью до естественно определяемого изоморфизма), которая имеет своим полем функций это поле. Верно также, что если поле K имеет степень трансцендентности 1 над полем k, то по полю K обязательно строится такая кривая; эта кривая называется *моделью* поля K. Неприводимая гладкая полная модель у поля K единственна. Более того, у каждого поля конечной степени трансцендентности n тоже есть неприводимая гладкая полная модель, но эта модель уже не будет единственной. И когда я перейду к рассказу о поверхностях, я покажу, почему. Для существования модели размерность не важна. Размерностью, с одной стороны, естественно называть степень трансцендентности поля функций. С другой стороны, оказывается, что размерность можно определять как размерность касательного пространства в гладкой точке. Здесь вы меня спросите,

откуда я заранее знаю, что имеется гладкая точка? Ответ такой: можно взять касательное пространство в общей точке над алгебраическим замыканием.

Итак, кривые гораздо проще, чем всё остальное. Потом, кривые в том или ином смысле можно задавать плоским уравнением, т. е. одним уравнением $P(x, y) = 0$. Правда, в этом случае мы не можем гарантировать, что кривая будет гладкая: не у любой кривой есть плоская гладкая модель; это скорее исключение, чем правило. Но хоть как-то мы можем задать кривую одним уравнением; иногда это бывает полезно.

Теперь вернёмся к конечному полю. Пусть X — гладкая полная неприводимая кривая над конечным полем $\mathbb{F}_r$. Первое, что нас интересует, это число точек $N = N_1 = |X(\mathbb{F}_r)|$. Однако у кривой есть и другие численные инварианты, а именно, для любого натурального числа m у неё есть естественный инвариант $N_m = |(X \otimes_{\mathbb{F}_r} \mathbb{F}_{r^m})(\mathbb{F}_{r^m})|$. С точки зрения геометрической интуиции я не сделал ничего: просто я разрешил рассматривать решения над расширением степени m, и их считать точками степени 1. Раньше я считал любое решение из расширения точкой более высокой степени, а теперь я стал его считать точкой степени 1. Мы теперь считаем орбиты, состоящие из одной точки, но не относительно группы Галуа $\mathrm{Gal}\ \overline{\mathbb{F}}/\mathbb{F}_r$, а относительно её подгруппы — группы Галуа $\mathrm{Gal}\ \overline{\mathbb{F}}/\mathbb{F}_{r^m}$.

Откуда взялись дополнительные точки? Возьмём кривую над $\overline{\mathbb{F}}$. У неё есть точки степени 1, определённые над $\mathbb{F}_r$. Теперь меня интересует всё, что определено над полем $\mathbb{F}_{r^m}$. Поэтому если $d|m$ и у меня была орбита из d точек, то при переходе от исходной кривой к кривой над полем $\mathbb{F}_{r^m}$ вместо одной точки степени d у меня появляется d точек степени 1. Тем самым, $N_m = \sum\limits_{d|m} dB_d$, где B_d — число точек степени d на исходной кривой X. Действительно, каждая точка степени d — это орбита, а орбита даёт мне d точек. Я получил формулу, которая связывает информацию о точках на самой кривой с информацией о точках на той же самой кривой, но рассматриваемой над расширением исходного поля.

* * *

Пусть M — компактное многообразие, на котором задано действие $f\colon M \to M$, достаточно хорошее (например, неподвижные точки изолированные и, по возможности, кратности 1). То, что я сейчас сформулирую, это не теорема, а тип теорем, каждая из которых в своей ситуации своя. Поэтому бóльших деталей давать, по-видимому, не надо. Теорема Лефшеца в геометрической ситуации говорит следующее. Пусть нас интересует правильно посчитанное число неподвижных точек (точки

надо считать со знаком плюс или минус, кратные точки нужно считать с учётом кратности). Оказывается, что по многообразию M можно построить набор линейных пространств $H^i(M)$ над каким-то полем k, которые называются его *когомологиями* и отличаются рядом полезных свойств: 1) эта конструкция функториальна по M; 2) H^i конечномерны над некоторым полем k, $\operatorname{char} k = 0$. При этом «число» неподвижных точек отображения f равно $\sum(-1)^i \operatorname{Tr}(f : H^i)$. Теперь я поясню эту формулу. Поскольку когомологии — это функтор, то действие $f\colon M \to M$ определяет действие f на $H^i(M)$. Это — линейное пространство, на нём действует линейный оператор, можно посчитать след этого линейного оператора и вычислить альтернированную сумму следов для всех i. Это и есть теорема Лефшеца.

Теорему Лефшеца можно переформулировать в несколько более удобной для меня форме. Рассмотрим график $\Gamma_f \subset M \times M$ отображения f и пересечём этот график с диагональю. Пересечение этого графика с диагональю (опять-таки, правильно определённое), которое обозначается $(\Gamma_f \cdot \Delta)$, как раз равно числу Лефшеца $\sum(-1)^i \operatorname{Tr}(f : H^i)$.

Достаточно удивительным образом оказывается, что эта идея вполне применима к многообразиям над конечным полем по следующей причине: точки над основным полем это те и только те точки, которые неподвижны относительно гомоморфизма Фробениуса F. Иными словами, для гомоморфизма Фробениуса F, действующего на многообразии X, мне нужно посчитать количество его неподвижных точек. Гомоморфизм Фробениуса переводит многообразие X в себя, потому что его уравнения определены над основным полем. Действительно, пусть $\sum a_i x^i = 0$, причём a_i лежат в основном поле. Тогда $a_i^r = a_i$, а значит,

$$\sum a_i x^{ir} = \sum a_i^r x^{ir} = \left(\sum a_i x^i\right)^r = 0.$$

Тем самым, $N = (\Gamma_F \cdot \Delta) = \sum(-1)^i \operatorname{Tr}(F : H^i(X))$. Об этом люди догадались довольно давно. В случае эллиптической кривой такая формула была доказана Хассе. Потом она была доказана Андре Вейлем для произвольной кривой. Для произвольного многообразия (и это была некоторая сенсация) эту формулу доказал в 70-е годы Делинь. Казалось бы, я всё объяснил. Возникает вопрос, что здесь доказывать? Ответ таков. Если над полем характеристики p задано многообразие, то все естественные когомологии являются пространствами над этим же полем. Поэтому вместо равенства между числами вы получите сравнение по модулю p. А для того, чтобы получить настоящее равенство, вам нужно построить такую теорию когомологий, которая была бы не над полем характеристики p,

а над каким-то полем характеристики 0, и при этом удовлетворяла бы всем обычным свойствам теории когомологий: формула Кюннета, функториальность. Наконец, эта теория когомологий должна удовлетворять тому свойству, что для неё действительно можно доказать теорему Лефшеца. Конструкция этих когомологий, которые называются *l-адическими* когомологиями, как раз и была предложена Делинем. Главное, конечно, не сама конструкция, а доказательство этих свойств. Это довольно тонкая вещь.

Итак, в некотором смысле мы вычислили число точек. Давайте посмотрим, что получилось для кривой. Среди прочих свойств когомологий существует так называемая *теорема сравнения*, которая говорит, что хорошая теория когомологий это такая, что если у вас было достаточно хорошее многообразие с целыми коэффициентами, а потом вы взяли редукцию в характеристику p, то размерности когомологий, как правило, не должны измениться. Поэтому очень часто, если вы хотите просто вспомнить, чему равны когомологии того или иного многообразия над интересующим вас полем, то вы просто смотрите на обычные когомологии де Рама аналогичного многообразия над полем комплексных чисел. По теореме сравнения они совпадают. В частности, про неприводимую гладкую полную (компактную) кривую $X/\mathbb{F}_r$ мы знаем следующее. В силу неприводимости кривой $\dim H^0 = 1$, поскольку $\dim H^0$ — это число связных компонент. Есть также теорема двойственности, которая говорит, что для d-мерного алгебраического многообразия пространство H^i двойственно H^{2d-i} (поле у нас замкнутое, т. е. в каком-то смысле комплексное, поэтому когомологии есть в размерностях от 0 до $2d$; двойственность тоже соответствующая). В силу двойственности $\dim H^0 = \dim H^2$. Кроме того, $\dim H^1 = 2g$ есть удвоенный род кривой. Как определять род кривой, я подробно говорить не буду. Можно определять род как число ручек соответствующей комплексной кривой; пространство первых когомологий редуцированной по модулю p кривой (по одному из свойств когомологий) имеет ту же самую размерность $2g$. Но более правильный способ такой. Нужно взять когомологии когерентные с коэффициентами в структурном пучке. Первые когомологии с коэффициентами в структурном пучке — это то же самое, что глобально регулярные дифференциальные формы. Их будет g штук (т. е. g-мерное пространство). А раз мы знаем, что такое касательное пространство, то мы знаем, что такое дифференциал и что такое дифференциальная форма *).

Имеет место следующее сильное утверждение, которое не совсем верно, когда кривая некомпактна.

*) Обратите внимание, что «правильных» l-адических когомологий вдвое больше, чем когерентных, не говоря уже о том, что когерентные — линейное пространство над полем характеристики p.

Утверждение. *Собственные значения F на $H^i(X)$ лежат в $\overline{\mathbb{Q}}$, и при вложении $\overline{\mathbb{Q}} \subset \mathbb{C}$ для собственного значения ω имеет место равенство $|\omega| = r^{i/2}$.*

Это утверждение называется *гипотезой Римана* по следующим причинам. Обычная дзета-функция Римана определяется либо как сумма $1/n^s$ по всем целым числам, либо как эйлеровское произведение. Что такое сумма по целым числам в данном случае? По целому числу n мы строим идеал (n) и $n = |\mathbb{Z}/(n)|$. Поэтому можно сказать, что мы рассматриваем сумму по идеалам:

$$\zeta(s) = \sum \frac{1}{n^s} = \sum_{\text{идеалы } \mathfrak{a}} \frac{1}{N(\mathfrak{a})^s},$$

где $N(\mathfrak{a})$ равно числу элементов в факторкольце по $\mathfrak{a}$. Это имеет свой аналог в любом кольце. В частности, это определение обобщается на кривую X:

$$\zeta_X(s) = \sum_{\text{дивизоры } D = \sum n_i P_i} \frac{1}{N(D)^s}.$$

Здесь n_i — целые числа, P_i — точки (не обязательно степени 1), $N(D) = \prod r^{n_i \deg P_i}$.

Удивительным образом оказывается, что дзета-функция кривой помимо того, что она обладает всеми свойствами, которыми обладает обыкновенная дзета-функция (продолжается на всю плоскость, имеет функциональное уравнение), записывается следующим образом. Пусть $\zeta_X(s) = Z_X(r^s)$. Тогда $Z_X(t)$ — рациональная функция следующего вида:

$$Z_X(t) = \frac{\prod_{i=1}^{2g}(1 - \omega_i t)}{(1-t)(1-rt)},$$

где ω_i — собственные значения оператора Фробениуса на первых когомологиях. В числителе записан характеристический многочлен действия Фробениуса на первых когомологиях, а в знаменателе записано произведение характеристических многочленов действия Фробениуса на одномерных нулевых и вторых когомологиях. Для кривой никаких других когомологий нет. Для произвольных многообразий все нечётные когомологии пишутся в числителе, а чётные в знаменателе. Гипотеза Римана, которая говорит, что все нетривиальные нули дзета-функции должны лежать на прямой с вещественной частью $1/2$, равносильна тому, что все ω_i должны лежать на окружности радиуса $\sqrt{r}$. В общем случае, для произвольного алгебраического многообразия, собственные значения действия оператора

Фробениуса на i-х когомологиях должны лежать на окружности радиуса $r^{i/2}$. Это и есть гипотеза Римана в функциональном случае, которая доказана, в отличие от классической гипотезы Римана. (Хассе для эллиптических кривых, Андре Вейлем для произвольных кривых и Делинем для алгебраических многообразий произвольной размерности.)

Это обстоятельство позволяет нам вычислить число точек. А именно, $N = r + 1 - \sum \omega_i$; $r + 1$ — это число точек на проективной прямой. Более того, $N_m = r^m + 1 - \sum \omega_i^m$. Кроме того, мы знаем, что по абсолютной величине ω_i равно $\sqrt{r}$. Поэтому

$$|N - r - 1| \leqslant 2g\sqrt{r}. \tag{1}$$

Это неравенство называется *границей Вейля* для числа точек на кривой. Первый вопрос: бывает ли так, что она достигается? Ответ: да, бывает.

На таком уровне знаний мы были в самом начале 80-х годов. Дальше оказалось следующее. С одной стороны, в этот момент возникла алгебро-геометрическая теория кодирования, которую придумал Гоппа и которая потребовала изучения алгебраических кривых над конечным полем с некоторых точек зрения, с которых они раньше не изучались. А с другой стороны, в это же время Ихара в Японии обнаружил одно замечательное обстоятельство. Удивительно, что в одном и том же номере журнала появилась статья Манина, в которой он рассказывал про наши работы по изучению точек на кривых, связанные с кодами, и заметка Ихары о том феномене, о котором я сейчас расскажу. Часто в математике бывает, что разные люди просыпаются одновременно. Феномен такой. Давайте зададимся вопросом, достигается ли максимальное число точек, которое равно $r + 1 + 2g\sqrt{r}$? Меня интересует ситуация, когда $N = r + 1 + 2g\sqrt{r}$. Во-первых, r должно быть квадратом. Но это мелочи. Давайте посмотрим, где должны лежать ω_i, если достигается равенство $N = r + 1 + 2g\sqrt{r}$. В выражение $N = r + 1 - \sum \omega_i$ все ω_i входят со знаком минус, поэтому они должны лежать на отрицательной полуоси. Давайте теперь думать, не противоречит ли что-нибудь здравому смыслу в этой ситуации? С одной стороны, нет, потому что есть примеры, когда это именно так. А с другой стороны, немножко противоречит по следующей причине. Все ω_i^2 положительны. Но тогда $N_2 = r^2 + 1 - 2gr$. Если род велик, то $N_2 < 0$. А так не бывает, потому что N_2 — это число точек. Более того, если $m|n$, то $N_m \leqslant N_n$, потому что точки, определённые над полем, определены и над его расширением. А уж про положительность я и не говорю. Поэтому мораль такая: максимальные по Вейлю кривые существуют только для небольших значений рода. Для больших значений рода они существовать не могут.

Теперь я приведу пример максимальной кривой, которая имеет максимальный возможный род. Это так называемая *эрмитова* кривая. Пусть $\bar{x} = x^{\sqrt{r}}$ (я предполагаю, что r — квадрат). Эту операцию естественно назвать комплексным сопряжением по той причине, что $\bar{\bar{x}} = x^r = x$ для $x \in \mathbb{F}_r$. Рассмотрим эрмитову форму

$$x\bar{x} + y\bar{y} + z\bar{z} = 0,$$

т. е.

$$x^{\sqrt{r}+1} + y^{\sqrt{r}+1} + z^{\sqrt{r}+1} = 0.$$

Род этой кривой равен $g = \frac{r-\sqrt{r}}{2}$, а число точек на ней $N = r^{3/2}+1$. Утверждение про число точек — это элементарное упражнение, а утверждение про род кривой — задача из простой топологии или алгебраической геометрии над $\mathbb{C}$: нужно посчитать, сколько ручек у кривой, заданной гладким уравнением данной степени. Если мы сопоставим эти два числа, то увидим, что кривая максимальна. И можно доказать, что для бóльших родов кривая максимальной быть не может.

Есть гипотеза, что любая кривая, которая является максимальной по Вейлю, накрывается эрмитовой кривой. Доказывать это мы не умеем.

Если r — нечётная степень числа p, то в неравенстве (1) мы можем поставить целую часть, поскольку у нас числа целые. Но оказывается, что мы можем сделать намного лучше. Это придумал Серр в 1983 г. На самом деле можно написать такое неравенство: $|N - r - 1| \leqslant g[2\sqrt{r}]$. Это довольно существенное усиление неравенства Вейля, потому что, например, для $r = 2$ и для большого рода g число $g[2\sqrt{r}]$ гораздо меньше, чем число $[2g\sqrt{r}]$. Но ещё большее усиление заключено в следующем результате, который принадлежит В. Дринфельду и С. Влэдуцу. Это тоже результат 82—84-го года. Он доказывается за несколько минут, но у меня сейчас нет на это времени. Они доказали следующее. Посмотрим на максимум отношения N/g при $g \to \infty$. Это означает, что мы рассматриваем семейство кривых растущего рода над фиксированным полем. Из формулы Серра сразу видно, что $N/g \lesssim [2\sqrt{r}]$. Но оказывается, что асимптотически имеет место гораздо более сильное неравенство $N/g \lesssim \sqrt{r} - 1$. Идея доказательства — в чистом виде обобщение рассуждения Ихары, но с учётом всех неравенств $N_m \leqslant N_n$ для $m \mid n$. Во-первых, оказывается, что эти неравенства избыточные. Что учитывать их все, что учитывать только неравенства $N_1 \leqslant N_m$, это одно и то же. Как учитывать все неравенства $N_1 \leqslant N_m$ сразу? Я расскажу сейчас некую более общую схему. Теорему Дринфельда—Влэдуца можно доказать совсем элементарно, но мне эта более общая схема важна для дальнейшего. Она называется *теорией*

явных формул. Мы можем взять систему уравнений

$$N_m = r^m + 1 - r^{m/2} 2 \sum_1^g \cos\, m\varphi_i.$$

Здесь я учёл, что ω_i лежит на окружности радиуса $r^{1/2}$, и ещё есть сопряжённое число $\bar{\omega}_i$; их сумма равна $r^{1/2} 2 \cos\, \varphi_i$. Я хочу учесть то обстоятельство, что такая формула есть для всех m. В качестве следующего шага я умножу обе части на некоторый коэффициент v_m, а после этого просуммирую по m:

$$\sum_{m=1}^{\infty} v_m N_m = \sum_{m=1}^{\infty} \left(r^m + 1 - r^{m/2} 2 \sum_1^g \cos\, m\varphi_i \right) v_m.$$

Затем я учту то обстоятельство, что $N = N_1 \leqslant N_m$:

$$N \sum_{m=1}^{\infty} v_m \leqslant \sum_{m=1}^{\infty} v_m N_m = \sum_{m=1}^{\infty} \left(r^m + 1 - r^{m/2} 2 \sum_1^g \cos\, m\varphi_i \right) v_m.$$

Дальше мне нужно разобраться с этим неравенством. Оказывается, что если выполнены два условия, а именно, $v_m \geqslant 0$ и $f_v(\varphi) = 1 + 2 \sum v_m \cos m\varphi \geqslant$ $\geqslant 0$, то тогда все косинусы можно неким образом заменить нулём и получить некоторое неравенство. Я это неравенство выписывать не буду, но смысл его такой: любой выбор набора коэффициентов v_m, удовлетворяющих этим двум условиям, даёт некоторое неравенство на величину N. Дальше возникает вопрос о том, как умно выбрать эти коэффициенты; это зависит от задачи, которую вы ставите. Такой набор условий встречался в анализе в XIX веке. Это называлось *двояко положительные ядра интегрирования*. Дальше так. Для кривых всё, оказывается, довольно просто. А когда мы работаем с поверхностями (я сейчас об этом скажу), работа выглядит таким образом. Открываешь учебник анализа XIX века, находишь там какое-нибудь двояко положительное ядро Валле Пуссена, и после этого смотришь, оно хорошо или плохо с этой точки зрения. Потом берёшь другое ядро и сравниваешь результаты. Никакой алгебраической геометрии здесь нет, здесь есть 60 крупноформатных страниц анализа, хотя и называется всё это «Явные формулы для числа точек на многообразиях над конечным полем» (это моя с Ж. Лашо совместная работа в журнале Крелля). Ж. Остерле доказал, что если род кривой находится в некотором диапазоне, то в этом диапазоне на число точек есть некоторая оценка, и лучше этим аналитическим методом не сделаешь. Иными словами, Остерле выжал из этого аналитического метода для кривых абсолютный максимум возможного. Откуда мы знаем, что это

максимум? Мы можем поставить ту же самую задачу, забыв про то, что ω_i — алгебраические числа, помня лишь их метрические свойства. Тогда границы Остерле есть объективно оптимальные границы. А учитывать то, что ω_i — алгебраические числа, очень трудно. Границы Серра это учитывают, но в асимптотиках это не очень нужно.

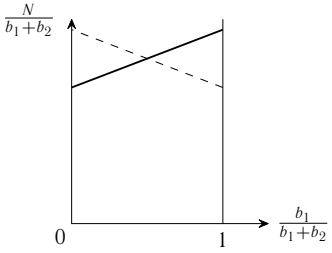
Правда, здесь есть одно но. А именно, если $r=2$, то границы Остерле работают не всегда. И что работает на самом деле, никто не знает. Точнее говоря, есть диапазоны для родов, когда мы не знаем оптимума даже в этом аналитическом смысле. Когда $r>2$, там, как это часто бывает в анализе, происходит некое чудо. Одна величина совпадает с другой, хотя априори не было ясно, что они должны совпадать, и из-за этого всё получается.

Нам интересно знать, кривые с каким числом точек существуют. Я рассказал, как получать верхние границы. И есть несколько десятков хороших математиков во всём мире, которые занимаются тем, что строят кривые с достаточно большим числом точек. В интернете существуют таблицы того, что мы знаем для небольших r и небольших родов. Например, мы знаем, что если $r=2$ а род равен 50, то существует кривая с 40 точками, а кривой с 41 точкой не существует. Но в большей части случаев мы знаем лишь диапазон возможных значений. Есть пример или абстрактно доказанная нижняя граница (теорема существования) и есть верхняя оценка; между ними имеется определённое расстояние. Это один из видов деятельности: попытка понять, насколько эти границы близки к тому, что имеет место на самом деле. А вторая вещь — это типичное для математики слепое пятно. Вся эта наука была сделана в районе 1983–84 года. В построении примеров с тех пор был значительный прогресс, и ещё кое в чём, но границы все были известны уже тогда. Почему-то никто на протяжении 10 лет после этого не задал себе вопрос: «А что же имеет место для поверхностей?» Чтобы просто задать вопрос, потребовалось десять лет паузы.

Что происходит для поверхностей? Все эти методы работают и для поверхностей тоже. Для поверхностей тоже имеют место аналогичные формулы. Про формулу для дзета-функции это как раз было известно. Это сделал Дворк: нужно все нечётные когомологии собрать в числителе, а все чётные — в знаменателе. В формуле для числа точек вместо одной суммы с минусом появляется много сумм, соответствующих всем промежуточным когомологиям. Для (двумерной) поверхности формула такая:

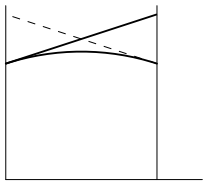
$$N_m = r^{2m} + 1 - (r^{m/2} + r^{3m/2}) \sum_{i=1}^{b_1} \alpha_i^m + r^m \sum_{i=1}^{b_2} \beta_i^m.$$

Здесь сумма $r^{m/2}+r^{3m/2}$ соответствует первым и третьим когомологиям, r^m соответствует вторым когомологиям; b_1 и b_2 — числа Бетти; $|\alpha_i|=1$ и $|\beta_i|=1$. Может быть, вопрос не ставился в том числе и потому, что было непонятно, что стремить к бесконечности здесь. Для кривых понятно: род стремится к бесконечности, это и есть асимптотика. Оказывается, что достаточно, чтобы хотя бы одно из чисел Бетти стремилось к бесконечности. Если мы хотим асимптотически смотреть на задачу, то нам нужно чтобы сумма b_1+b_2 стремилась к бесконечности, а как именно она стремится к бесконечности, это всё равно.



Р и с. 3.

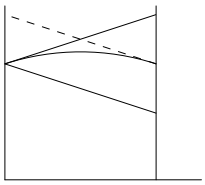
В этой задаче правильно рисовать график. Этот график выглядит следующим образом (рис. 3). По оси абсцисс я нарисую отношение $\dfrac{b_1}{b_1+b_2}$, а по оси ординат я буду рисовать отношение $\dfrac{N}{b_1+b_2}$. Граница Вейля — это сплошная прямая. Она использует только тот факт, что мы знаем абсолютные величины $|\alpha_i|=1$ и $|\beta_i|=1$. Если я заменяю все α_i и β_i на ± 1, то я получаю границу Вейля. Она явно выписывается. Оказывается, что можно взять трюк Ихары в чистом виде: если была точка в районе отрицательной вещественной полуоси, то её квадрат находится в районе положительной полуоси. Оказывается, что при больших b_1, т. е. когда отношение $\dfrac{b_1}{b_1+b_2}$ близко к 1, эту границу можно существенно уменьшить. Вторая граница — это пунктирная прямая. Эту границу придумал я, и с того момента, как я задался этим вопросом, мне понадобилась примерно неделя. Это ситуация, когда вопрос подразумевает ответ. Для этой границы я использовал в чистом виде метод Дринфельда–Влэдуца в его элементарной формулировке. Оказывается, что применение более хитрого метода двояко положительных ядер позволяет построить много других прямых. Огибающая этого семейства прямых — это как раз та верхняя граница, которая нам сегодня известна. Огибающая выглядит следующим образом (рис. 4). Мы сначала какое-то время идём по границе Вейля, а потом есть непрерывная кривая, которая в самом конце касается моей границы. Это мы сделали с Ж. Лашо. Было сразу понятно, что что-то получится, но счёта там было необыкновенно много.


Р и с. 4.

Тут есть масса неотвеченных вопросов. Один из вопросов состоит в том, что, в отличие от кривых,

когда мы даже на конечном уровне (по крайней мере, когда $r \neq 2$) знаем, какая граница аналитически оптимальна (мы не знаем, реализуется ли она алгебраически, но мы знаем аналитику), здесь этого чуда не происходит, и даже в асимптотике мы не знаем, что эта огибающая оптимальна, и даже уверены, что она — не самое лучшее, что можно построить. Выбирая ядра другим образом, можно построить огибающую лучше. Это чистый анализ, с которым мы, в каком-то смысле, не справились, хотя и написали 60 страниц текста. Оптимального двояко выпуклого ядра подобрать не удаётся. Удаётся подобрать какие-то ядра, которые хороши то в одной точке, то в другой. И при этом даже в этих точках мы не знаем, что они оптимальны.

Совершенно другая задача — попробовать построить примеры. Я сходу построил нижнюю границу, которая ведёт себя так, как показано на рис. 5. Для этого нужно перемножить две кривые, а потом раздуть на них некоторое количество точек. Разброс довольно большой. Дальше идёт тоже довольно простая, но опирающаяся на много алгебраической геометрии, попытка построить какие-нибудь интересные примеры. Оказывается, что интересных примеров строится не так уж много. Один пример: можно рассмотреть плоскость $\mathbb{P}^2$ и раздуть на ней точки, т. е. произвести моноидальные преобразования, вклеивающие вместо точки прямую. Можно брать поверхности Делиня–Люстига, которые возникают из алгебраических групп, и их раздувать. Есть ещё кое-что. Когда я говорю «раздувать», это всё равно, что сказать: «А потом мы делаем нечто очень простое.» А интересно не раздувать. То есть, рассматривать минимальные поверхности. Для минимальных поверхностей есть классификация. И можно попытаться пробежаться по классификации и позадавать себе разные вопросы. Например, сколько может быть точек на поверхности типа $K3$? Это отчасти написано в моей статье.



Р и с. 5.

Основная мораль в том, что это совершенно непаханный край. Этой задачей почти никто не занимался, потому что это требует знания алгебраических поверхностей, с одной стороны, а с другой стороны, это требует работы. У меня такое ощущение, что возьмёшься — и сразу получится.

28 марта 2002 г.

В. М. Бухштабер

АЛГЕБРАИЧЕСКИЕ МНОГООБРАЗИЯ ПОЛИСИММЕТРИЧЕСКИХ ПОЛИНОМОВ И КОЛЬЦА ДИФФЕРЕНЦИАЛЬНЫХ ОПЕРАТОРОВ

Сегодняшняя лекция, с одной стороны, тесно связана с моей предыдущей лекцией (см. [1]), с другой стороны, она тесно связана с тем, что я рассказывал на юбилейной конференции Независимого Московского Университета (см. также [2], [4]). Но акцент будет немножко другой, потому что если первая лекция ставила целью ввести в проблему, то сейчас я хочу обратить внимание на следующее:

С точки зрения классической теории инвариантов задача о многообразии полисимметрических полиномов представляет собой один из первых и важнейших *специальных* случаев задачи о действии конечной группы $G \subset \mathrm{GL}(d, \mathbb{C})$ на пространстве $\mathbb{C}^d$. Эта задача имеет геометрическую часть — описать пространство орбит $\mathbb{C}^d/G$, и алгебраическую часть — описать множество порождающих инвариантных полиномов и соотношения между ними.

Общие подходы и результаты, в том числе алгоритм решения алгебраической части общей задачи на основе базисов Грёбнера, детально описаны в [3] (глава 7).

Но в том то и дело, что общие задачи приводят к теоремам существования, которые бывают полезны, но редко приводят к эффективным ответам в важнейших частных случаях. Замечательно, что решения, отвечающие специфике этих случаев, как правило, открывают взаимосвязи разделов математики, казавшихся до этого далекими.

Я благодарен Виктору Васильевичу Прасолову за профессиональную запись моей лекции. Редактируя предоставленный им текст я несколько перепланировал свою лекцию с учетом новых результатов и вставил ссылки на публикации.

Геометрическая часть

Рассмотрим m-мерное комплексное линейное пространство $\mathbb{C}^m$, $m \geqslant 1$. Возьмём $n \geqslant 1$ и образуем новое пространство

$$\mathrm{Sym}^n(\mathbb{C}^m) = \underbrace{\mathbb{C}^m \times \ldots \times \mathbb{C}^m}_{n} / S_n,$$

где S_n — группа перестановок. Нам будет важна и другая реализация этого пространства. Рассмотрим пространство $M(m, n)$ матриц с m строками и n столбцами. Пусть v_{ij}, где $1 \leqslant i \leqslant m$ и $1 \leqslant j \leqslant n$, — общий элемент такой матрицы. Линейное пространство $M(m, n)$, конечно, можно отождествить с пространством $\underbrace{\mathbb{C}^m \times \ldots \times \mathbb{C}^m}_{n}$, имея в виду, что у нас имеется как раз n столбцов. Действие группы S_n тоже понятное: справа умножаем на матрицу перестановок; она будет переставлять столбцы. Поэтому, если мы рассмотрим пространство $M(m, n)/S_n$, т. е. профакторизуем $M(m, n)$ по действию матриц перестановок порядка n, то получим пространство $\mathrm{Sym}^n(\mathbb{C}^m)$.

Сразу возникает вопрос, какова геометрическая природа этого объекта. До тех пор пока мы не факторизовали, у нас было линейное пространство размерности mn. После факторизации возникает многообразие той же размерности: пространство орбит действия симметрической группы S_n. Если бы группа S_n действовало свободно, мы получили бы гладкое многообразие, накрываемое пространством $\mathbb{C}^{nm}$. Но ясно, что S_n действует не свободно, и мы имеем дело с разветвленным накрытием. Как же описать пространство орбит? В наших работах с Элмером Рисом мы получили описание этого геометрического объекта как аффинного алгебраического многообразия, используя рекурсию Фробениуса (см. [5]). Я напомню через некоторое время эту рекурсию. А пока давайте подготовимся.

Рассмотрим кольцо полиномов от m переменных $\mathbb{C}(m) = \mathbb{C}[u_1, \ldots, u_m]$ и введём объект $\mathbb{C}(m)^*$, который определяется следующим образом. Надо рассмотреть $\mathrm{Hom}_{\mathbb{C}}(\mathbb{C}[u_1, \ldots, u_m], \mathbb{C})$, т. е. все линейные над полем $\mathbb{C}$ гомоморфизмы кольца $\mathbb{C}(m)$ в $\mathbb{C}$. Мы как бы забываем, что $\mathbb{C}(m)$ это кольцо и рассматриваем его только как линейное пространство над $\mathbb{C}$. Потом мы вспомним об умножении благодаря Фробениусу. Кольцо $\mathbb{C}[u_1, \ldots, u_m]$, рассматриваемое как (бесконечномерное) линейное пространство над $\mathbb{C}$, имеет канонический базис u^ω. Здесь $\omega = (i_1, \ldots, i_m)$ — мультииндекс, а $u^\omega = u_1^{i_1} \ldots u_m^{i_m}$. После этого очень естественно ввести градуировку. Будем считать, что $|\omega| = \sum i_k$, и положим $\deg u^\omega = |\omega|$. Пусть меня простят строгие алгебраисты, которые предпочли бы, чтобы я написал $\deg u^\omega = 2|\omega|$, чтобы не встречались нечетные числа. Но у нас коммутативная алгебра, и мы не будем обращать на это внимания.

Теперь можно объяснить, что такое Hom в нашем случае. Когда есть бесконечномерное линейное пространство, нужно объяснять, что такое Hom. Как только у нас введено понятие градуировки, то кроме всего прочего можно ещё сказать, что в этом пространстве есть топология. Мы будем считать, что $u^\omega \to 0$ при $|\omega| \to \infty$. Это даёт нам возможность

говорить про топологию (это — известная топология, задаваемая градуировкой). Когда я говорю про Hom, я имею в виду непрерывные гомоморфизмы в этой топологии.

Я специально не хочу всё уточнять, потому что по ходу дела у нас многое будет проясняться. Но я хочу заранее сказать, что пространство $\mathbb{C}(m)^*$ я буду рассматривать как линейное топологическое пространство с базисом u_ω, двойственным базису u^ω:

$$\langle u_{\omega'},\, u^{\omega''}\rangle = \delta_{\omega',\omega''}.$$

Теперь мы полностью вошли в привычное русло.

Итак, у нас есть пространство $\mathrm{Sym}^n(\mathbb{C}^m)$, есть пространство $\mathbb{C}(m)^*$, и мы можем написать отображение эвалюации $\mathrm{ev}\colon \mathrm{Sym}^n(\mathbb{C}^m) \to \mathbb{C}(m)^*$:

$$[v_1,\, \ldots,\, v_n] \mapsto \mathrm{ev}[v_1,\, \ldots,\, v_n], \quad [v_1,\, \ldots,\, v_n](u^\omega) = \sum_{k=1}^{n} v_k^\omega,$$

здесь $[v_1, \ldots, v_n]$ — неупорядоченный набор m-мерных векторов, $v_k^\omega = v_{1k}^{i_1} \ldots v_{mk}^{i_m}$.

Теперь у нас есть две задачи. Первую задачу я уже давал в [1] и повторю ещё раз.

1. Доказать, что ev — вложение (многообразия с особенностями в линейное пространство).

2. Описать образ.

Что значит «описать образ»? Благодаря первой задаче мы можем реализовать это конечномерное многообразие как подмножество в линейном пространстве $\mathbb{C}(m)^*$. А описание его — это и есть один из главных наших результатов. Мы опишем это подмножество как алгебраическое подмногообразие в $\mathbb{C}(m)^*$, указав все уравнения. Мы сейчас построим систему алгебраических уравнений на $\mathbb{C}(m)^*$, такую, что точка принадлежит образу нашего многообразия тогда и только тогда, когда она удовлетворяет этим уравнениям. Единственная тонкость в том, что пространство $\mathbb{C}(m)^*$ бесконечномерное. Назовем это решением задачи 2(1). Потом будет следующая задача 2(2): показать, как это сделать уже в конечномерном пространстве. Если задачу 2(1) мы решили уже давно, и я об этом много раз рассказывал, то сегодня я буду рассказывать, как решить задачу 2(2). Я надеюсь, что в конце моей лекции станет понятно, насколько мы продвинулись в этой задаче.

Задачу 1 вы можете решить достаточно быстро. Она решается элементарными средствами алгебраической геометрии. Если бы мы имели дело не с $\mathrm{Sym}^n(\mathbb{C}^m)$, а с симметрической степенью $\mathrm{Sym}^n(X)$ топологического

пространства X, то нам для доказательства того, что ev является вложением, пришлось бы использовать методы функционального анализа (см. [6]). А здесь — чистая алгебраическая геометрия.

Вторая задача нетривиальна уже хотя бы потому, что сначала нужно выяснить, откуда берутся уравнения. А потом ещё нужно доказать, что предъявленное множество уравнений полное.

Разрешите мне перейти к задаче 2(1.1): откуда берутся уравнения?

Здесь мы вспомним, что работаем не просто с линейным пространством $\mathbb{C}(m)^*$, а с непрерывными линейными функционалами на алгебре. Поэтому, для того чтобы описать уравнения, я напомню такое понятие, как *фробениусовы n-гомоморфизмы*. Это понятие относится к линейным отображениям алгебры A с единицей в алгебру B. Сегодня мне достаточно будет ограничиться случаем $B=\mathbb{C}$. Мы будем рассматривать линейные $\mathbb{C}$-гомоморфизмы $f\colon A\to\mathbb{C}$. Определим по индукции гомоморфизмы

$$\Phi_k(f)\colon \underbrace{A\otimes\ldots\otimes A}_{k}\to\mathbb{C}$$

(здесь берутся тензорные произведения $\mathbb{C}$-модулей над $\mathbb{C}$) следующим образом:

$$\begin{gathered}\Phi_1(f)=f,\\ \Phi_2(f)(a_1,\,a_2)=f(a_1)f(a_2)-f(a_1a_2),\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \Phi_{k+1}(f)(a_1,\,\ldots,\,a_{k+1})=f(a_1)\Phi_k(f)(a_2,\,\ldots,\,a_{k+1})-\\ -\sum_{l=2}^{k+1}\Phi_k(f)(a_2,\,\ldots,\,a_1a_l,\,\ldots,\,a_{k+1}).\end{gathered}$$

О п р е д е л е н и е. Линейный гомоморфизм $f\colon F\to\mathbb{C}$ называется *фробениусовым n*-гомоморфизмом, если:

1) $f(1)=n$;

2) $\Phi_{n+1}(f)\equiv 0$.

Рекурсия обрывается в точности на $(n+1)$-м шаге; если $f(1)=n$, то раньше рекурсия не оборвется. Для того чтобы с этим разобраться, я даю следующую задачу.

З а д а ч а. Классифицировать все фробениусовы n-гомоморфизмы из $\mathbb{C}[u]$ в $\mathbb{C}$.

Очень советую эту задачу решить. Сразу многое станет понятно.

Фактически мы уже приблизились к тому, чтобы описать семейство алгебраических уравнений, которое нам нужно. У нас есть отображение

ev: $\mathrm{Sym}^n(\mathbb{C}^m) \to \mathbb{C}(m)^*$. Пространство $\mathbb{C}(m)^*$ — это пространство всех линейных отображений. В нём есть подмножество $\Phi_n(m) \subset \mathbb{C}(m)^*$, состоящее из фробениусовых n-гомоморфизмов. Это подмножество алгебраическое: если вы распишете рекурсию, то увидите, что условие обрыва есть алгебраическое уравнение.

Напомню, что когда у вас есть пространство, двойственное пространству функций, то на этом пространстве естественные координаты — это сами функции. В качестве координатных векторов в пространстве $\mathbb{C}(m)^*$ возьмем мономы u^ω. Тогда координаты будут занумерованы всеми разбиениями $\omega = (i_1, \ldots, i_m)$. Как эти координаты задаются? Если вы возьмете вектор $f \in \mathbb{C}(m)^*$, то координата f_ω определяется следующим образом: $f_\omega \stackrel{\text{def}}{=} f(u^\omega)$.

Итак, любой линейный гомоморфизм на кольце функций $\mathbb{C}(m)$ может быть записан как вектор с координатами f_ω, где $\omega \in \mathbb{Z}^m_{\geqslant 0}$. Давайте посмотрим, что такое кольцевые гомоморфизмы, т. е. 1-гомоморфизмы. Условие обрыва на втором шаге запишется следующим образом. Для любых a_1 и a_2 должно выполняться равенство $f(a_1 a_2) = f(a_1) f(a_2)$. Мы знаем, что $u^{\omega_1} u^{\omega_1} = u^{\omega_1 + \omega_2}$. Значит, в этих координатах уравнение запишется так: $f_{\omega_1} f_{\omega_2} - f_{\omega_1 + \omega_2} = 0$. Это — полный список алгебраических уравнений (когда ω пробегает решетку), задающих кольцевые гомоморфизмы в этом бесконечномерном пространстве. Узнаете уравнение гиперболы $xy = z$?

У нас появились решётки, соотношения на решетках. В дальнейшем всё это будет играть большую роль. Но сначала я хочу обратить ваше внимание на то, что 1-гомоморфизмы образуют алгебраическое подмногообразие, которое задаётся в бесконечномерном пространстве такими гиперболическими уравнениями.

З а д а ч а. Написать уравнения для $\Phi_2(m)$, задающие многообразия 2-гомоморфизмов.

После того как я ввёл соглашения о координатах, каждый из вас легко заметит, что рекурсия Фробениуса даёт нам настоящие (те, к которым мы привыкли в алгебраической геометрии) алгебраические уравнения в этих координатах.

Теперь я объясню, как от этой задачи перейти к вложению многообразия $\mathrm{Sym}^n(\mathbb{C}^m)$ в конечномерное линейное пространство. Я буду изучать канонические отображения $\mathbb{C}(m)^* \to \mathbb{C}^N$ (бесконечномерного пространства в конечномерное). Для этого я фиксирую функции $\varphi_1, \ldots, \varphi_N$ из $\mathbb{C}(m)$ и буду сопоставлять линейному функционалу $f \in \mathbb{C}(m)^*$ набор его координат $f_{\varphi_1}, \ldots, f_{\varphi_N}$. Здесь, как и выше, имеется в виду, что $f_{\varphi_k} = f(\varphi_k)$.

Если бы у нас было не $\mathbb{C}^m$, а какое-то многообразие, мы могли бы сделать всё так же, но потом в качестве функций взяли бы те, которые разделяют координатные окрестности, покрывающие это многообразие. И получилась бы классическая теорема о вложении многообразия в линейное пространство. Эта теория удивительным образом имеет смысл даже тогда, когда $n=1$. В случае общего компактного гладкого m-мерного многообразия мы сначала рассматриваем вложение $M\subset\mathbb{C}(M)^*$, а потом, для соответствующего набора функций на многообразии, получаем вложение в конечномерное пространство. В традиционных курсах анализа мы обычно не обращаем внимание на это. А вообще, классический анализ устроен именно так. Сначала берется достаточно широкий класс так называемых основных функций и строится каноническое отображение в линейное пространство, двойственное к пространству таких функций, а потом берутся пробные функции, которые уже сажают нас на конечномерное пространство. Здесь мне потребовалось напомнить это с самого начала.

Итак, взяв набор функций $\varphi_1, \ldots, \varphi_N$, мы попадаем в $\mathbb{C}^N$. А дальше возникает такая интересная задача, в полном соответствии с классическим анализом: Надо найти набор функций $\varphi_1, \ldots, \varphi_N$, такой, чтобы композиция ev с проекцией на $\mathbb{C}^N$, задаваемой этим набором, была вложением, и переписать условия обрыва в координатах пространства $\mathbb{C}^N$. Как мы будем это делать, я расскажу позже.

Посмотрим ещё раз внимательно на рекурсию Фробениуса. Буквально глядя на неё мы должны увидеть, что если взять в качестве N число $\binom{n+m}{n}$, которое многие специалисты по теории представлений тут же узнают, то имеется набор $\varphi_1, \ldots, \varphi_N$, для которого композиция будет вложением. Мы можем рассматривать образ этой композиции как аффинное алгебраическое многообразие в конечномерном пространстве размерности $N=\binom{n+m}{n}$. Откуда взялось это число? Что оно собой представляет? Давайте вспомним, что мы выбираем в линейном пространстве $\mathbb{C}(m)^*$ базис u^ω, координаты вектора $f\in\mathbb{C}(m)^*$ дают набор $\{f_\omega,\ |\omega|\geqslant 0\}$, где $f_\omega=f(u^\omega)$ и $|\omega|=\sum\limits_{k=1}^{m} i_k$. Если мы подсчитаем, сколько существует разбиений ω (включая пустое множество), для которых $|\omega|\leqslant n$, то мы как раз и получим число $\binom{n+m}{n}$.

В этих терминах очень легко объяснить это отображение. Мы просто берем вектор f из $\mathbb{C}(m)^*$ и отбрасываем все его координаты, у которых норма индекса ω превосходит n. Почему мы получаем вложение? Потому что если $f\in\Phi_n(m)$, то формула рекурсии Фробениуса показывает следующее. Условие обрыва показывает, что любая координата, для которой

$|\omega| > n$, алгебраически выражается через координаты, для которых $|\omega| \leqslant n$. Давайте попробуем это увидеть в классическом случае. Для кольцевых гомоморфизмов $f_{\omega_1+\omega_2} = f_{\omega_1} f_{\omega_2}$. Если $|\omega| \geqslant 2$, то всегда можно представить ω в виде суммы $\omega = \omega_1 + \omega_2$, где $|\omega_1| < 2$. Это — как блуждание по решётке: если расстояние от точки решётки до начала координат больше или равно 2, то всегда до неё можно дойти каким-то путём, проходя через ω_1 и ω_2, и значит, выразить координату с индексом ω через координаты с индексами, у которых норма индекса меньше $|\omega|$. Вот и всё доказательство.

Точно так же можно сделать для любых m и n.

Рекурсия Фробениуса сопоставляет каждому линейному гомоморфизму $f\colon \mathbb{C}(m) = \mathbb{C}[u_1, \ldots, u_m] \to \mathbb{C}$ линейные гомоморфизмы $\Phi_k(f)\colon \mathbb{C}(m) \otimes \ldots \otimes \mathbb{C}(m) \to \mathbb{C}$, $k = 1, 2, \ldots$ Таким образом, для данного k мы имеем преобразование

$$\Phi_k\colon \mathbb{C}(m)^* \to \mathbb{C}(m)^* \mathbin{\widehat{\otimes}} \ldots \mathbin{\widehat{\otimes}} \mathbb{C}(m)^* = \mathrm{Hom}(\mathbb{C}(m) \otimes \ldots \otimes \mathbb{C}(m),\ \mathbb{C}),$$

представляющее собой алгебраическое отображение линейных пространств. Здесь $\widehat{\otimes}$ — это символ пополненного тензорного произведения топологических линейных пространств.

В координатах $\{f_\omega\}$ вектора $f \in \mathbb{C}(m)^*$

$$\Phi_k(\{f_\omega\}) = \{\Phi_{\omega_1,\ldots,\omega_k}\}.$$

где $\Phi_{\omega_1,\ldots,\omega_k} = \Phi_k(u^{\omega_1} \otimes \ldots \otimes u^{\omega_k})$. Например, при $k = 2$

$$\Phi_{\omega_1,\omega_2} = f_{\omega_1} f_{\omega_2} - f_{\omega_1+\omega_2},$$

при $k = 3$

$$\Phi_{\omega_1,\omega_2,\omega_3} = f_{\omega_1} f_{\omega_2} f_{\omega_3} - f_{\omega_1} f_{\omega_2+\omega_3} - f_{\omega_1+\omega_2} f_{\omega_3} - f_{\omega_2} f_{\omega_1+\omega_3} + 2 f_{\omega_1+\omega_2+\omega_3}.$$

В общем случае, решая рекурсию Фробениуса (см. [5]), получаем следующий результат: Фиксируем разложение перестановки $\sigma \in S_k$ в произведение циклов $\gamma_1 \ldots \gamma_q$. Для $\gamma_i = (i_1, \ldots, i_s)$ положим $f_{\gamma_i} = f_{\omega_{i_1}+\ldots+\omega_{i_s}}$ и $f_\sigma = f_{\gamma_1} \ldots f_{\gamma_q}$. Имеет место формула

$$\Phi_{\omega_1,\ldots,\omega_k} = \sum_{\sigma \in S_k} \varepsilon(\sigma) f_\sigma,$$

где $\varepsilon(\sigma)$ — знак перестановки σ.

Таким образом, мы получили композицию отображений

$$\mathrm{Sym}^n(\mathbb{C}^m) \xrightarrow{\mathrm{ev}} \mathbb{C}(m)^* \xrightarrow{\Phi_{n+1}} \mathbb{C}(m)^* \mathbin{\widehat{\otimes}} \ldots \mathbin{\widehat{\otimes}} \mathbb{C}(m)^*,$$

где отображения ev и Φ_{n+1} заданы явными формулами.

Т е о р е м а 1 (Бухштабер, Рис, см. [5]). *Образ вложения* ev *задается уравнением* $\Phi_{n+1}(f)=0$, *где* $f_\varnothing = f(1) = n$.

В координатах $\{f_\omega\}$ вектора $f \in \mathbb{C}(m)^*$ алгебраическое многообразие $\mathrm{Sym}^n(\mathbb{C}^m)$ задается системой уравнений

$$f_\varnothing = n,\ \Phi_{\omega_1,\dots,\omega_{n+1}} = 0, \quad \text{где } |\omega_l| > 0,\ l = 1, \dots, n+1.$$

Я закончил программу геометрическую. Теперь я хочу перейти к алгебраической части — полисимметрическим полиномам.

Алгебраическая часть

Мы можем уменьшить на 1 размерность пространства вложения $\mathbb{C}^N$, так как функция $f_\varnothing$ является постоянной на $\Phi_n(m)$. Перейдем теперь к описанию $\mathrm{Sym}^n(\mathbb{C}^m)$ как алгебраического многообразия в $\mathbb{C}^N$, где $N = \binom{n+m}{n} - 1$. Для этого мы построим каноническую алгебраическую замену координат в $\mathbb{C}(m)^*$. Начнем со следующей общей конструкции.

Рассмотрим пространство $\mathbb{C}^\infty$ с координатами $s = (s_1, \dots, s_n, \dots)$. Введем алгебраическое обратимое преобразование (замену координат)

$$E\colon \mathbb{C}^\infty \to \mathbb{C}^\infty : E(s) = (e_1, \dots, e_k, \dots),$$

которая в терминах производящих рядов задается формулой:

$$1 + \sum_{k=1}^{\infty} e_k t^k = \exp\left(\sum_{l=1}^{\infty} (-1)^{l-1} s_l \frac{t^l}{l} \right).$$

Имеем $e_1 = s_1$ и

$$(-1)^k s_k + k e_k = \sum_{q=1}^{k-1} (-1)^{q-1} s_q e_{k-q} \quad \text{при } k > 1. \tag{1}$$

Положим $\deg s_k = \deg e_k$, тогда соотношение (1) становится однородным. Это соотношение дает рекуррентные формулы для полиномов

$$e_n = e_n(s_1, \dots, s_n) \quad \text{и} \quad s_n = s_n(e_1, \dots, e_n),$$

которые являются однородными в указанной градуировке. Таким образом, мы получаем взаимообратные алгебраические отображения E и $S\colon \mathbb{C}^\infty \to \mathbb{C}^\infty$:

$$E(s) = (e_1(s_1), \dots, e_n(s_1, \dots, s_n), \dots)$$

и

$$S(e) = (s_1(e_1), \dots, s_n(e_1, \dots, e_n), \dots).$$

Л е м м а 1 (см. [4]). *Отображение E полностью определяется следующими двумя свойствами*:

1. $\frac{\partial e_n}{\partial s_1} = e_{n-1}$, *где* $e_0 = 1$ *и* $e_n(0) = 0$ *при* $n > 0$.
2. *Положим* $d = \sum\limits_{r \geqslant 2} r s_{r-1} \frac{\partial}{\partial s_r}$. *Тогда*

$$de_n = -(n-1)e_{n-1}, \quad n = 1, 2, \ldots$$

Доказательство леммы опирается на следующие факты:

1. $\left[\frac{\partial}{\partial s_k}, d\right] = \frac{\partial}{\partial s_k} d - d \frac{\partial}{\partial s_k} = (k+1) \frac{\partial}{\partial s_{k+1}}$.
2. Пусть $\frac{\partial g(s)}{\partial s_1} = 0$ и $dg(s) = 0$. Тогда $g(s) = g(0)$.

Полезно знать явные формулы для полиномов $e_n = e_n(s_1, \ldots, s_n)$:

$$e_n = \frac{1}{n!} \det \begin{pmatrix} s_1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ s_2 & s_1 & 2 & 0 & \ldots & 0 & 0 \\ s_3 & s_2 & s_1 & 3 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & n-2 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & s_1 & n-1 \\ s_n & s_{n-1} & \ldots & \ldots & \ldots & s_2 & s_1 \end{pmatrix}$$

и для $s_n = s_n(e_1, \ldots, e_n)$:

$$s_n = (-1)^n n \sum_{i_1 + 2i_2 + \ldots + ki_k = n} (-1)^{i_1 + i_2 + \ldots + i_k} \frac{(i_1 + i_2 + \ldots + i_k - 1)!}{i_1! \ldots i_k!} e_1^{i_1} \ldots e_k^{i_k}. \quad (2)$$

Эти формулы хорошо известны специалистам и широко используются в классической теории симметрических полиномов. В случае $m = 1$ формула для e_n дает выражение элементарной симметрической функции через полиномы Ньютона, а формула для s_n дает выражение полинома Ньютона через элементарные симметрические функции. В полисимметрическом случае ($m > 1$) элементарные полисимметрические функции перестают быть алгебраически независимыми. Тем более замечательно, что эти формулы все же продолжают прекрасно работать.

Далее мы увидим, что при $m > 1$ эти формулы описывают связь *алгебраически независимых* функций на линейном пространстве, содержащем $\mathrm{Sym}^n(\mathbb{C}^m)$ как алгебраическое подмногообразие, и только *ограничение* на $\mathrm{Sym}^n(\mathbb{C}^m)$ возвращает нас к полисимметрическим функциям.

Обозначим через $\mathbb{C}(m)_0^* \subset \mathbb{C}(m)^*$ подпространство, натянутое на координатные оси, соответствующие мономам u^ω, $\omega = (i_1, \ldots, i_m)$ с $|\omega| > 0$. Построим требуемую замену координат в $\mathbb{C}(m)_0^*$ при помощи *поляризации* отображения E.

Сопоставим вектору $f=\{f_\omega\}\in\mathbb{C}(m)_0^*$ набор однородных полиномов $(s_1(f),\ldots,s_k(f),\ldots)$, $\deg s_k(f)=k$ в $\mathbb{C}(m)=\mathbb{C}[u_1,\ldots,u_m]$, $\deg u_i=1$, где

$$s_k(f)=\sum_{|\omega|=k}\binom{k}{\omega}\,f_\omega u^\omega,\quad \text{где } \binom{k}{\omega}=\frac{k!}{i_1!\ldots i_m!}.$$

Заметим теперь, что так как $e_n=e_n(s_1,\ldots,s_n)$ является однородным полиномом степени n от переменных $s_1,\ldots,s_n$, то

$$e_n(f)=e_n(s_1(f),\ldots,s_n(f))$$

— однородный полином степени n от вектора переменных $u=(u_1,\ldots,u_m)$. Следовательно, имеет место разложение

$$e_n(f)=\sum_{|\omega|=n}e_\omega(f)u^\omega,$$

которое *однозначно* определяет функции $e_\omega(f)$, представляющие собой полиномы от переменных $\{f_{\omega'},\ |\omega'|\leqslant|\omega|\}$.

Например, при $n=m=2$:

$$\begin{aligned}
s_1(f)&=f_{(1,0)}u_1+f_{(0,1)}u_2,\\
s_2(f)&=f_{(2,0)}u_1^2+2f_{(1,1)}u_1u_2+f_{(0,2)}u_2^2,\\
e_1(f)&=e_{(1,0)}(f)u_1+e_{(0,1)}(f)u_2=s_1(f),
\end{aligned}$$

т. е.

$$\begin{aligned}
&e_{(1,0)}(f)=f_{(1,0)},\quad e_{(0,1)}(f)=f_{(0,1)},\\
&e_2(f)=e_2(s_1(f),\,s_2(f))=\frac{1}{2}(s_1(f)^2-s_2(f))=\\
&\quad=\frac{1}{2}(f_{(1,0)}^2-f_{(2,0)})u_1^2+(f_{(1,0)}f_{(0,1)}-f_{(1,1)})u_1u_2+\frac{1}{2}(f_{(0,1)}^2-f_{(2,0)})u_2^2,
\end{aligned}$$

т. е.

$$\begin{aligned}
e_{(2,0)}(f)&=\frac{1}{2}(f_{(1,0)}^2-f_{(2,0)}),\quad e_{(1,1)}(f)=(f_{(1,0)}f_{(0,1)}-f_{(1,1)}),\\
e_{(0,2)}(f)&=\frac{1}{2}(f_{(0,1)}^2-f_{(2,0)}).
\end{aligned}$$

Таким образом, мы построили алгебраическое обратимое преобразование

$$\mathcal{E}\colon\mathbb{C}(m)_0^*\to\mathbb{C}(m)_0^*,\quad \mathcal{E}(f)=\{e_\omega(f),\ |\omega|>0\}.$$

Для построения обратного отображения

$$\mathcal{S}\colon\mathbb{C}(m)_0^*\to\mathbb{C}(m)_0^*,\quad \mathcal{S}(\{e_\omega\})=\{f_\omega\}$$

надо подставить в формулу для однородного полинома $s_n(e_1,\ldots,e_n)$ вместо e_k выражения

$$\widehat{e}_k=\sum_{|\omega|=k}e_\omega(f)u^\omega$$

и получить набор координат $\{f_\omega,\ |\omega|=n\}$ из однородного по вектору u полинома $s_n(\widehat{e}_1, \ldots, \widehat{e}_n)$ при помощи разложения вида

$$s_n(f)=\sum_{|\omega|=n}\binom{n}{\omega}\ f_\omega u^\omega.$$

Вернемся теперь к отображению

$$\mathrm{ev}\colon\ \mathrm{Sym}^n(\mathbb{C}^m)\to\mathbb{C}(m)^*\ :\ \mathrm{ev}([v_1, \ldots, v_n])=\{f_\omega,\ |\omega|\geqslant 0\},$$

где $f_\omega=\sum\limits_{k=1}^{n} v_k^\omega$ — полисимметрический полином Ньютона.

Вычислим композицию

$$\mathrm{Sym}^n(\mathbb{C}^m)\xrightarrow{\mathrm{ev}}\mathbb{C}(m)_0^*\xrightarrow{\mathcal{E}}\mathbb{C}(m)_0^*.$$

Имеем:

$$\exp\sum_{l=1}^{\infty}(-1)^{l-1}s_l(\mathrm{ev}([v_1, \ldots, v_n]))\frac{t^l}{l}=$$

$$=\exp\sum_{l=1}^{\infty}(-1)^{l-1}\sum_{|\omega|=l}\binom{l}{\omega}\left(\sum_{k=1}^{n}v_k^\omega u^\omega\frac{t^l}{l}\right)=$$

$$=\exp\sum_{k=1}^{n}\left(\sum_{l=1}^{\infty}(-1)^{l-1}\langle v_k,\ u\rangle^l\frac{t^l}{l}\right)=\prod_{k=1}^{n}(1+\langle v_k,\ u\rangle t),$$

где $\langle v_k,\ u\rangle=\sum\limits_{i=1}^{m} v_{ik}u_i$. Следовательно,

$$1+\sum_{k=1}^{\infty}\left(\sum_{|\omega|=k}e_\omega(\mathrm{ev}([v_1, \ldots, v_n]))u^\omega\right)t^k=\prod_{k=1}^{n}\left(1+\langle v_k,\ u\rangle t\right).$$

Заметим теперь, что $e_\omega(\mathrm{ev}([v_1, \ldots, v_n]))$ — это элементарные симметрические функции $e_\omega(v_1, \ldots, v_n)$. В частности, $e_\omega(\mathrm{ev}([v_1, \ldots, v_n]))=0$ если $|\omega|>n$. Таким образом, мы получили, что композиция отображений $\mathcal{E}\cdot\mathrm{ev}$ сопоставляет точке $[v_1, \ldots, v_n]\in\mathrm{Sym}^n(\mathbb{C}^m)$ набор элементарных полисимметрических функций $\{e_\omega(v_1, \ldots, v_n)\}$.

Рассмотрим в пространстве $\mathbb{C}(m)_0^*$ с координатами $e=\{e_\omega,\ |\omega|>0\}$ подпространство $\mathbb{C}^N$, выделяемое условиями $e_\omega=0$ при $|\omega|>n$. Обозначим через $\mathcal{S}_N\colon\mathbb{C}^N\to\mathbb{C}(m)^*$ алгебраическое вложение, задаваемое композицией

$$\mathbb{C}^N\xrightarrow{i_N}\mathbb{C}(m)^*\xrightarrow{\mathcal{S}}\mathbb{C}(m)^*,$$

где i_N — вложение в координатах $\{e_\omega,\ |\omega|>0\}$.

Суммируя приведенные выше достаточно простые конструкции, мы получаем совершенно нетривиальный результат:

Т е о р е м а 2. *Отображение*

$$\mathrm{ev}\colon\ \mathrm{Sym}^n(\mathbb{C}^m)\to\mathbb{C}(m)^*$$

разлагается в композицию

$$\mathrm{Sym}^n(\mathbb{C}^m)\xrightarrow{\mathrm{ev}_E}\mathbb{C}^N\xrightarrow{\mathcal{S}_N}\mathbb{C}(m)^*,$$

где $\mathrm{ev}_E([v_1,\ \ldots,\ v_n])=\{e_\omega(v_1,\ \ldots,\ v_n)\}$.

Ясно, что ev_E — вложение, и мы имеем

С л е д с т в и е. *В координатах* $\{e_\omega\}$ *вектора* $g\in\mathbb{C}^N$ *алгебраическое многообразие* $\mathrm{Sym}^n(\mathbb{C}^m)$ *задается системой уравнений*

$$\widetilde{\Phi}_{\omega_1,\ldots,\omega_{n+1}}(g)=\Phi_{\omega_1,\ldots,\omega_{n+1}}(\mathcal{S}_N(g))=0,$$

где $|\omega_j|>0,\ j=1,\ \ldots,\ n+1$.

Обозначим через $\Phi_{n+1,N}$ композицию отображений

$$\mathbb{C}^N\xrightarrow{\mathcal{S}_N}\mathbb{C}(m)^*\xrightarrow{\Phi_{n+1}}\mathbb{C}(m)^*\mathbin{\widehat{\otimes}}\ldots\mathbin{\widehat{\otimes}}\mathbb{C}(m)^*.$$

Отождествим $\mathrm{Sym}^n(\mathbb{C}^m)$ с его образом при ev_E и введем факторпространство $X=\mathbb{C}^N/\ \mathrm{Sym}^n(\mathbb{C}^m)$, т. е. стянем в точку $\mathrm{Sym}^n(\mathbb{C}^m)$ как замкнутое подмножество в $\mathbb{C}^m$.

Композиция отображений

$$\mathrm{Sym}^n(\mathbb{C}^m)\xrightarrow{\mathrm{ev}_E}\mathbb{C}^N\xrightarrow{\Phi_{n+1,N}}\mathbb{C}(m)^*\mathbin{\widehat{\otimes}}\ldots\mathbin{\widehat{\otimes}}\mathbb{C}(m)^*$$

согласно теоремам 1 и 2 индуцирует вложение

$$i_X\colon X\hookrightarrow\mathbb{C}(m)^*\mathbin{\widehat{\otimes}}\ldots\mathbin{\widehat{\otimes}}\mathbb{C}(m)^*.$$

В поддержку утверждения о нетривиальности теоремы 2, достаточно сказать, что из нее вытекает следующая геометрическая интерпретация второй фундаментальной теоремы теории инвариантов (в терминологии Д. Гильберта) в рассматриваемом случае:

Для построенного выше конечномерного пространства X существует проекция π_X пространства $\mathbb{C}(m)^*\mathbin{\widehat{\otimes}}\ldots\mathbin{\widehat{\otimes}}\mathbb{C}(m)^*$ на конечномерное пространство $\mathbb{C}^{N_1}$, такая, что композиция $\pi_X i_X\colon X\to\mathbb{C}^{N_1}$ является вложением.

Я уверен, что знатоки алгебраической топологии и гомологической алгебры уже узнали в этой интерпретации построение геометрической реализации резольвенты кольца полисимметрических полиномов.

Приложение колец дифференциальных операторов

В этом разделе в качестве иллюстрации плодотворных связей разных разделов математики мы приведем в терминах кольца дифференциальных операторов формулировку нашего с Элмером Рисом результата о соотношениях между полисимметрическими полиномами.

Рассмотрим алгебру формальных дифференциальных операторов $D(m) = \mathbb{C}[[\partial_1, \ldots, \partial_m]]$, действующих на кольце полиномов $\mathbb{C}(m) = \mathbb{C}[u_1, \ldots, u_m]$, где $\partial_k = \dfrac{\partial}{\partial u_k}$. Каждому линейному гомоморфизму

$$f : \mathbb{C}[u_1, \ldots, u_m] \to \mathbb{C}$$

сопоставим оператор

$$d(\partial;\ f) = d(f) = \sum_{|\omega| \geqslant 0} f_\omega \frac{\partial^\omega}{\omega!},$$

где $\partial = (\partial_1, \ldots, \partial_m)$ и $\partial^\omega = \partial_1^{i_1} \ldots \partial_m^{i_m}$ для $\omega = (i_1, \ldots, i_m)$. Имеем $d(f)u^\omega\big|_{u=0} = f_\omega$. Таким образом, мы получаем изоморфизм в координатах $\{f_\omega\}$

$$d : \mathbb{C}(m)^* \to D(m), \qquad f \to d(f).$$

В координатах $\{e_\omega\}$ пространства $\mathbb{C}(m)^*$ сопоставим вектору $g = \{e_\omega\}$ набор дифференциальных операторов

$$\delta_l = \delta_l(\partial;\ g) = \sum_{|\omega| = l} e_\omega \partial^\omega, \qquad l = 1, 2, \ldots$$

Тогда, как легко видеть, в случае n-гомоморфизмов имеет место формула

$$d(\partial;\ \mathcal{S}(g)) = n + \sum_{k=1}^{\infty} \frac{1}{k!} s_k(\delta_l, \ldots, \delta_k).$$

Используя формулу (2), мы получаем явное описание композиции гомоморфизмов

$$\mathbb{C}(m)^* \xrightarrow{\mathcal{S}} \mathbb{C}(m)^* \xrightarrow{d} D(m),$$

которая задает изоморфизм в координатах $\{e_\omega\}$. Дифференциальный оператор

$$d(\mathrm{ev}([v_1, \ldots, v_n])) = \sum_{|\omega| \geqslant 0} p_\omega(v_1, \ldots, v_n) \frac{\partial^\omega}{\omega!}$$

можно рассматривать одновременно и как производящий ряд для полисимметрических полиномов Ньютона.

Положим $V=[v_1, \ldots, v_n]\in \mathrm{Sym}^n(\mathbb{C}^m)$. Имеем

$$d_V = d(\mathrm{ev}\ V) = \sum_{k=1}^{n} \exp\langle v_k, \partial\rangle,$$

где $\partial=(\partial_1, \ldots, \partial_m)$. В частности, при $n=1$, $d_V=\exp\langle v, \partial\rangle$ — оператор сдвига на вектор v: $d_V p(u)=p(u+v)$. В общем случае (т. е. при $n>1$) оператор d_V — оператор многозначного (n-значного) сдвига (см. [7], [8]).

Теорему 1 теперь можно переформулировать в виде:

Т е о р е м а 3. *Формальный дифференциальный оператор* $d\in D(m)$ *задает* n*-гомоморфизм тогда и только тогда, когда* $d(1)=n$ *и* $d=\sum\limits_{k=1}^{n} \exp\langle v_k, \partial\rangle$ *для некоторой точки* $V\in \mathrm{Sym}^n(\mathbb{C}^m)$.

Обозначим через $D(m, n+1)$ алгебру формальных дифференциальных операторов $\mathbb{C}[[\nabla_1, \ldots, \nabla_{n+1}]]$, действующих на кольце полиномов $\mathbb{C}(m, n+1)=\mathbb{C}[U_1, \ldots, U_{n+1}]$, где $U_k=(u_{ik})$, $i=1, \ldots, m$, $k=1, \ldots, n+1$, и $\nabla_k=(\partial_{1k}, \ldots, \partial_{mk})$. Имеет место изоморфизм

$$\mathbb{C}(m)^* \mathbin{\widehat{\otimes}} \ldots \mathbin{\widehat{\otimes}} \mathbb{C}(m)^* \to D(m, n+1).$$

В терминах колец дифференциальных операторов отображению

$$\Phi_{n+1}\colon \mathbb{C}(m)^* \to \mathbb{C}(m)^* \mathbin{\widehat{\otimes}} \ldots \mathbin{\widehat{\otimes}} \mathbb{C}(m)^*$$

соответствует алгебраическое отображение

$$\Phi_{n+1}\colon D(m) \to D(m, n+1).$$

Используя данную выше формулу для Φ_{n+1} в координатах $\{f_\omega\}\in\mathbb{C}(m)^*$, получаем следующий результат:

Т е о р е м а 4. *Для* $g=\{e_\omega\}\in\mathbb{C}^N$ *оператор* $\Phi_{n+1,N}(g)$ *имеет вид*

$$\sum_{\sigma\in S_{n+1}} \varepsilon(\sigma) d_\sigma(\nabla_1, \ldots, \nabla_{n+1};\ g),$$

где $d_\sigma(\nabla_1, \ldots, \nabla_{n+1};\ g)=d_{\gamma_1}(\mathcal{S}_N g)\ldots d_{\gamma_q}(\mathcal{S}_N g)$ — *произведение операторов* $d_{\gamma_i}(\mathcal{S}_N g)=d(\nabla_{i_1}+\ldots+\nabla_{i_s};\ \mathcal{S}_N g)$ *для* $\gamma_i=(i_1, \ldots, i_s)$.

Здесь $\sigma=\gamma_1\ldots\gamma_n$ — как и выше, разложение перестановки $\sigma\in S_{n+1}$ в произведение циклов.

Примеры:

Для $f=\mathcal{S}_N g$, где $N=\binom{n+m}{n}$,

$$\Phi_{2,N}(g)=d(\nabla_1+\nabla_2;\ f)-d(\nabla_1;\ f)d(\nabla_2;\ f),$$

$$\begin{aligned}\Phi_{3,N}(g)=2d(\nabla_1+\nabla_2+\nabla_3;\ f)&-d(\nabla_1;\ f)d(\nabla_2+\nabla_3;\ f)-\\ &-d(\nabla_1+\nabla_2;\ f)d(\nabla_3;\ f)-d(\nabla_2;\ f)d(\nabla_1+\nabla_3;\ f)+\\ &+d(\nabla_1;\ f)d(\nabla_2;\ f)d(\nabla_3;\ f).\end{aligned}$$

Теорема 4 завершает описание всех соотношений между элементарными полисимметрическими функциями $e_\omega(v_1, \ldots, v_n)$:

Дифференциальный оператор $\Phi_{n+1,N}(g)$ является производящим рядом для образующих таких соотношений.

З а д а ч а. Найти замену переменных в $D(m, n+1)$, при которой ряд $\Phi_{n+1,N}(g)$ переходит в полином.

Указание: Найти соответствующую поляризацию описанной выше алгебраической замены переменных $\mathcal{E}\colon \mathbb{C}(N)_0^* \to \mathbb{C}(N)_0^*$.

Литература

[1] Бухштабер В.М., Симметрические полиномы многих векторных аргументов. Классические задачи и современные приложения // ГЛОБУС, выпуск 2, Общематематический семинар НМУ, М.: МЦНМО—НМУ. 2005. С. 126—145.

[2] Бухштабер В.М., Многообразия полисимметрических полиномов. Классические задачи, современные приложения // Математика. Механика. Информатика., Труды конференции, посвященной 10-летию РФФИ. — М.: ФИЗМАТЛИТ. 2004. С. 129—145.

[3] Кокс Д., Литтл Дж., О'Ши Д., Идеалы, многообразия и алгоритмы., — М.: Мир, 2000.

[4] Бухштабер В.М., Рис Э.Г., Кольца непрерывных функций, Симметрические произведения и алгебры Фробениуса // Успехи матем. наук. 2004. Т. 59. № 1. С. 125—144.

[5] Buchstaber V.M., Rees E.G., The Gel'fand map and symmetric products., — Selecta Math. (N.S.), 2002. V. 8. № 4. P. 523—535.

[6] Бухштабер В.М., Рис Э.Г., Конструктивное доказательство обобщенного изоморфизма Гельфанда // Функц. анализ и его прил. 2001. Т. 35, № 4. С. 20—25.

[7] Buchstaber V.M., Veselov A.P., Integrable correspondences and algebraic representation of multivalued groups // IMRN. 1996. № 8. P. 381—400.

[8] Buchstaber V.M., Rees E.G., Multivalued groups, their representations and Hopf algebras // Transformation groups. 1997. V. 2. № 4. Birkhauser-Boston. P. 325—349.

11 апреля 2002 г.

Пьер Делинь

О ζ-ФУНКЦИЯХ МНОГИХ ПЕРЕМЕННЫХ

Пусть $s_1, \ldots, s_r \geqslant 1$ — целые числа, а

$$\zeta(s_1, \ldots, s_n) = \sum_{n_1 > n_2 > \ldots > n_r} \frac{1}{n_1^{s_1} \ldots n_r^{s_r}}.$$

Впервые ζ-функции рассматривал Эйлер, при $r = 1$. В этом случае получается ζ-функция Римана. Я сначала напомню, что сделал в этой области Эйлер; это интересно и с точки зрения алгебраической геометрии. Эйлер получил следующую формулу для суммы обратных квадратов натуральных чисел:

$$\sum_n \frac{1}{n^2} = \frac{\pi^2}{6} = -\frac{(2\pi i)^2}{24}.$$

Число $2\pi i$ встречается часто; а числа 6 и 24 связаны так: если рассмотреть число, взаимно простое с 6, то его квадрат всегда дает остаток 1 при делении на 24.

Эйлер получил также формулу для суммы произвольных четных степеней чисел, обратных к натуральным:

$$\sum_n \frac{1}{n^{2k}} = \frac{1}{2}(2\pi)^k \frac{B_{2k}}{(2k)!}$$

и понял, что он не может сделать ничего для нечетных показателей.

Эйлер дает этому факту следующее дерзкое доказательство. Он рассматривает функцию $\frac{\sin \pi s}{\pi s}$ как многочлен бесконечной степени. Корни этого многочлена — целые числа, кроме нуля, а производная в нуле равна единице, откуда Эйлер получает следующую формулу:

$$\frac{\sin \pi s}{\pi s} = \prod_{\substack{n \in \mathbb{Z} \\ n \neq 0}} \left(1 - \frac{s}{n}\right).$$

Разумеется, этому нужно придать смысл; если собрать вместе члены с n и $-n$, то получится сходящееся произведение. Если вычислить логарифмическую производную от этой функции, нетрудно получить формулу Эйлера.

Эйлер также рассмотрел значения ζ-функции в отрицательных целых числах. Сумма положительных степеней целых чисел не имеет смысла, но существует формула, связывающая значения ζ-функции в точках s и $1-s$, также угаданная Эйлером:

$$\zeta(1-s)=2(2\pi)^{-s}\Gamma(s)\cos\frac{\pi s}{2}\zeta(s),$$

откуда вытекает, что при четных k

$$\zeta(1-k)=\sum_n n^{k-1}=-\frac{B_k}{k}.$$

Эйлер также рассматривал ζ-функции от 2 переменных. Его определение несколько отличалось от нашего: мы рассматриваем сумму по значениям $n_1>n_2$, а Эйлер разрешал также и равенство:

$$\zeta^+(s,\,t)=1+\frac{1}{2^s}\left(1+\frac{1}{2^t}\right)+\frac{1}{3^s}\left(1+\frac{1}{2^t}+\frac{1}{3^t}\right)+\ldots$$

Эйлер доказал также тождество $\zeta(2,\,1)=\zeta(3)$ (мы пишем его в виде, когда равенство в индексах суммирования не разрешается). В его доказательстве используется тождество

$$\zeta(p,\,q)+\zeta(q,\,p)+\zeta(p+q)=\zeta(p)\zeta(q),$$

которое следует непосредственно из определения: рассмотрим сумму

$$\sum_n\frac{1}{n^p}\sum_m\frac{1}{m^q}=\sum_{n,m}\frac{1}{n^pm^q}$$

и разобьем ее на три слагаемых: сумму по $n>m$, сумму по $n<m$ и сумму по $n=m$.

Доказательство Эйлера нестрого. Он рассматривает двойную сумму

$$\sum\frac{1}{m^p}\sum\frac{1}{n^q}=\sum_{m,n}\frac{1}{m^pn^q}$$

и делит ее на три подсуммы, соответствующие $m>n$, $m=n$ и $m<n$. В первой сумме положим $m=n+a$, где a положительно, и рассмотрим сумму по n и a. Рациональную дробь $\frac{1}{(x+a)^px^q}$ можно разложить на элементарные дроби со знаменателями, составленными из степеней $(x+a)$ и x, например,

$$\frac{1}{(x+a)^2x}=\frac{1}{a^2x}-\frac{1}{a(x+a)^2}-\frac{1}{a^2(x+a)},$$

Отсюда вытекает, что

$$\zeta(2,\,1)=\sum_{a>0,n\in\mathbb{Z}}\frac{1}{(n+a)^2n}=\zeta(2)\zeta(1)-\zeta(1,\,2)-\zeta(2,\,1)=\zeta(3),$$

согласно вышеприведенному тождеству.

Разумеется, это рассуждение некорректно, поскольку ряды для $\zeta(1)$ и $\zeta(1, 2)$ расходятся. Если, однако, мы рассмотрим не бесконечные суммы, а суммы, где m и n ограничены сверху, и вычтем правую часть из левой, мы получим не нуль, но остаток, который стремится к нулю, когда границы стремятся к бесконечности. Для этого доказательства характерно то, что получить результат, касающийся сходящихся рядов типа $\zeta(2, 1)$, можно только рассматривая и расходящиеся ряды $\zeta(1, 2)$ и $\zeta(1)$. На самом деле существует, как мы увидим, способ регуляризации этих расходящихся рядов.

На этом я закончу изложение работ Эйлера в этой области и перейду к объяснению того, почему такие формулы представляют интерес в алгебраической геометрии. Основная причина состоит в том, что значения ζ-функций можно выразить не только через итерированные суммы, но и как интегралы от определенных функций. Свойства интегралов от алгебраических величин, например, эллиптических функций, это один из источников алгебраической геометрии и одно из лучших ее приложений.

Введем вначале некоторые обозначения, относящиеся к итерированным интегралам. Рассмотрим набор голоморфных 1-форм на комплексной плоскости $\mathbb{C}$, а также путь $\gamma(t) \subset \mathbb{C}$, соединяющий точку a с точкой b. Определим итерированный интеграл $\mathrm{It}\int \omega_1 \dots \omega_d$ вдоль пути γ так. Сначала построим при помощи отображения γ обратный образ всех форм ω_i на отрезке $[0, 1]$. После этого рассмотрим выражение

$$\int_0^1 \omega_1 \dots \int_0^{t_{d-2}} \omega_{d-2} \int_0^{t_{d-1}} \omega_{d-1} \int_0^{t_d} \omega_d = \int_{1>t_1>\dots>t_d} \omega_1 \dots \omega_d,$$

которое и есть $\mathrm{It}\int\limits_\gamma \omega_1 \dots \omega_d$.

Прежде чем выразить ζ-функции как итерированные интегралы, заметим, что поскольку 1-формы $\omega_1, \dots, \omega_d$ голоморфны, интеграл инвариантен относительно гомотопий пути γ с фиксированными концами. Теперь ζ-функция от многих переменных выражается в виде итерированного интеграла следующим образом:

$$\zeta(s_1, \dots, s_r) = \mathrm{It} \int_0^1 \underbrace{\frac{dz}{z} \dots \frac{dz}{z} \frac{dz}{1-z}}_{s_1} \underbrace{\frac{dz}{z} \dots \frac{dz}{z} \frac{dz}{1-z}}_{s_2} \dots \underbrace{\frac{dz}{z} \dots \frac{dz}{z} \frac{dz}{1-z}}_{s_r},$$

так что общее число форм равно $s_1 + \dots + s_r$. Действительно, поскольку $\dfrac{dz}{1-z} = \sum\limits_{n \geqslant 0} z^n\, dz$, то, начиная интегрирование с конца, как требуется по

определению, получим

$$\int \frac{dz}{1-z} = \sum_{n\geqslant 0} \frac{z^{n+1}}{n+1} = \sum_{n\geqslant 1} \frac{z^n}{n}.$$

Умножение на $\frac{dz}{z}$ и интегрирование дает

$$\int \sum_{n\geqslant 1} \frac{z^{n-1}dz}{n} = \sum_{n\geqslant 1} \frac{z^n}{n^2},$$

затем $\sum\limits_{n\geqslant 1} \frac{z^n}{n^3}$, и так далее, вплоть до $\sum\limits_{n\geqslant 1} \frac{z^n}{n^{s_r}}$. Следующая группа форм теперь дает:

$$\int \sum_{n\geqslant 1} \frac{z^n dz}{n^{s_r}(1-z)} = \int \sum_{n\geqslant 1} \sum_{m\geqslant 0} \frac{z^{m+n}dz}{n^{s_r}} = \sum_{n_1\geqslant 1} \sum_{n_2>n_1} \frac{z^{n_2}}{n_2 n_1^{s_r}},$$

и так далее.

Из представления ζ-функций в виде итерированных интегралов вытекает, в частности, формула для произведения ζ-функций. Одну ζ-функцию можно представить как итерированный интеграл по множеству $1\geqslant t_1\geqslant t_2\geqslant\ldots\geqslant t_N\geqslant 0$, и другую — как интеграл по множеству $1\geqslant u_1\geqslant u_2\geqslant\ldots\geqslant u_M\geqslant 0$, так что в целом получается интеграл по произведению двух симплексов. Такой интеграл можно разбить в сумму интегралов по симплексам, каждый из которых соответствует определенному упорядочению переменных t и u (так, чтобы $t_1, \ldots, t_N$ и, отдельно, $u_1, \ldots, u_M$ шли в убывающем порядке). Эти интегралы также будут ζ-функциями многих переменных. К полученному равенству можно применить соображения, похожие на те, что использовались при выводе формулы $\zeta(2, 1)=\zeta(3)$, некоторые тождества между рациональными функциями. Таким образом можно получить тождество

$$\zeta(p)\zeta(q)=\zeta(p, q)+\zeta(q, p)+\zeta(p+q),$$

которое мы уже получали раньше, а также более сложные формулы для произведений ζ-функций многих переменных.

Как показывает пример с $\zeta(3)$, при работе с ζ-функциями нужно заботиться о сходимости. Выражение для $\zeta(s_1, \ldots, s_r)$ сходится при условии $s_1\neq 1$. Представление ζ-функции в виде итерированных интегралов показывает причину расходимости при $s_1=1$: интеграл в этом случае заканчивается интегрированием $\int\limits_0^1 \frac{dz}{1-z}$, которое расходится в верхнем пределе. Также это видно из представления ζ-функции в виде суммы ряда: тогда ряд имеет вид $\sum\limits_n \frac{1}{n}\times\frac{1}{\ldots}$, откуда видно, что он расходится.

Чтобы придать смысл расходящимся рядам или интегралам, используемым для определения $\zeta(1, \ldots)$, их нужно регуляризовать. Здесь возникает одно неожиданное обстоятельство: оказывается, существуют два различных типа тождеств для ζ-функций и два различных способа регуляризации. Чтобы для расходящихся ζ-функций выполнялись тождества первого типа, нужна одна регуляризация, а для сохранения тождеств второго типа — другая; так что существует два различных способа приписывать значения расходящимся ζ-функциям. Например, если исходить из представления ζ-функций рядами, то должно иметь место тождество

$$\zeta(1)\zeta(1) = \zeta(1, 1) + \zeta(1, 1) + \zeta(2),$$

тогда как с точки зрения интегрального представления естественно ожидать, что

$$\zeta(1)\zeta(1) = \zeta(1, 1) + \zeta(1, 1).$$

Рассмотрим теперь ζ-функцию $\zeta(s_1, \ldots, s_r)$, в которой $s_1, \ldots, s_r$ — произвольные комплексные числа. Ряд, определяющий $\zeta(s_1, \ldots, s_r)$, сходится в некотором подмножестве пространства $\mathbb{C}^r$ — например, когда все s_i действительны и больше 1. Нетрудно доказать, что его сумма допускает мероморфное продолжение на все пространство $\mathbb{C}^r$. При этом множество полюсов лежит в объединении гиперплоскостей, задаваемом условием: $\sum_i (s_i - 1)$ — целое неположительное число. Можно также показать, что вычет ζ-функции в полюсе опять представляет собой ζ-функцию, но от меньшего числа переменных.

Самые важные значения ζ-функций возникают при целочисленных наборах s, которые отвечают размерностям когомологий определенных алгебраических многообразий.

ζ-функции можно обобщать различными способами. Например, переменным $s_1, \ldots, s_r$ можно давать не только комплексные, но и p-адические значения. Но у таких ζ-функций отсутствует когомологическая интерпретация, что делает их не слишком интересными. Другой способ обобщения заключается в том, чтобы рассматривать ряды вида $\sum \frac{z_1^{n_1} \ldots z_r^{n_r}}{n_1^{s_1} \ldots n_r^{s_r}}$. Это приводит к функциями типа полилогарифма, которые удовлетворяют красивым тождествам и дифференциальным уравнениям. Особенно интересен случай, когда переменные z_i являются корнями из единицы. В действительности, большинство утверждений об обыкновенных ζ-функциях имеют аналоги для рядов такого типа, в которых z_i — корни из единицы. Так, из ζ-функции Римана получаются некоторые ζ-функции Дирихле; их изучение дает дополнительную информацию и об обыкновенных ζ-функциях.

Моя цель теперь — объяснить, откуда берутся итерированные интегралы. Рассмотрим пространство

$$X = \mathbb{P}^1 \setminus \{0,\ 1,\ \infty\} = \mathbb{C}^* \setminus \{1\}.$$

Итерированные интегралы возникают из интегрирования форм на этом пространстве, типа $dz/(1-z)$. Также интересно рассматривать в качестве пространства X комплексную плоскость, из которой выколоты корни определенной степени из единицы — это помогает исследовать те аналоги ζ-функций, о которых шла речь выше. Рассмотрим фундаментальную группу $\pi_1(X, a)$, где a — некоторая отмеченная точка. Эта группа является свободной группой с двумя образующими; для симметрии можно ввести 3 образующих γ_0, γ_1, γ_∞, соответствующие элементарным петлям, обходящим соответствующие точки. В этом случае возникает одно соотношение: $\gamma_0\gamma_1\gamma_\infty = 1$. Мы будем считать, что умножение петель производится справа налево; это потому, что мы в дальнейшем свяжем с каждой петлей преобразование — параллельный перенос вдоль этой петли — а преобразования принято умножать справа налево.

Мы будем рассматривать группу π_1 по модулю N-го члена $\mathfrak{Z}^N$ ее нижнего центрального ряда. Это соответствует тому, что вместо всех представлений группы π_1 мы будем рассматривать только представления, получающиеся путем последовательного расширения тривиального представления. Например, все операторы вида

$$\begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

действуют на уровне нижнего центрального ряда тривиально.

Определение нижнего центрального ряда $\mathfrak{Z}^1 \supset \mathfrak{Z}^2 \supset \dots$ таково: $\mathfrak{Z}^1$ — это сама группа π_1, $\mathfrak{Z}^2$ — это коммутатор $[\pi_1, \pi_1]$, так что фактор $\mathfrak{Z}^1/\mathfrak{Z}^2$ это абелева группа с двумя образующими, $\mathfrak{Z}^3 = [\pi_1, \mathfrak{Z}^2]$, и т. д. Теперь можно ввести на $\pi_1/\mathfrak{Z}^N$ нечто вроде системы координат, следующим образом. Рассмотрим фактор $\mathfrak{Z}^i/\mathfrak{Z}^{i+1}$. Это абелева группа; нетрудно видеть, что она имеет конечный ранг, а в данном случае еще и не имеет кручения. Возьмем теперь какой-нибудь базис в факторе $\mathfrak{Z}^1/\mathfrak{Z}^2$ и поднимем его в группу $\pi_1 = \mathfrak{Z}^1$; обозначим поднятие $e_{1,1}$, $e_{1,2}$. Затем проделаем такую же процедуру с $\mathfrak{Z}^2/\mathfrak{Z}^3$; при этом $\mathfrak{Z}^2 \subset \mathfrak{Z}^1$, так что мы получаем несколько элементов $e_{2,1}, \dots \in \pi_1$, и так далее. Теперь всякому элементу $\gamma \in \pi_1$ можно сопоставить моном вида $e_{1,1}^{n_{11}} e_{1,2}^{n_{12}} e_{2,1}^{n_{21}} \dots$ следующим образом. Прежде всего отобразим γ в фактор $\mathfrak{Z}^1/\mathfrak{Z}^2$ и разложим там по базису. Это даст однозначно определенные числа n_{11} и n_{12}. Затем домножим γ на элемент, обратный

к $e_{1,1}^{n_{11}} e_{1,2}^{n_{12}}$, и получим элемент $\mathfrak{Z}^2$. Спроектируем его в $\mathfrak{Z}^2/\mathfrak{Z}^3$, определим коэффициенты n_{21}, …, и так далее. Полученная система целочисленных показателей n_{11}, n_{12}, … образует нечто вроде координат на $\pi_1/\mathfrak{Z}^N$; возникает вопрос, как записать в этих координатах групповую операцию. Рассуждение по индукции показывает, что эта операция записывается в виде полиномиальной формулы с рациональными коэффициентами. Действительно, произведение двух элементов группы — элемент группы, поэтому рассматриваемый многочлен должен принимать в целых точках целые значения. Как хорошо известно, такой многочлен имеет рациональные коэффициенты.

Конструкция Мальцева заключалась в том, чтобы рассмотреть бóльшую группу — тензорное произведение $(\pi_1/\mathfrak{Z}^N) \otimes \mathbb{Q}$. Формально говоря, эта группа состоит из элементов группы $\pi_1/\mathfrak{Z}^N$, возведенных во всевозможные рациональные степени, и их произведений. Если рассматривать только те представления π_1, которые получаются из тривиального последовательными расширениями, то эта группа ничем не хуже самой π_1. Действительно, если рассмотреть векторное пространство V и в нем унипотентный оператор T такой, что $1 - T$ нильпотентен, получим формулу бинома Ньютона:

$$T^n = (1 + (T-1))^n = \sum \binom{n}{k} (T-1)^k,$$

в которой сумма в правой части конечна. Биномиальные коэффициенты в правой части — многочлены от n, принимающие целые значения в целых точках.

Чтобы показать, что представления π_1, полученные последовательными расширениями тривиального, эквивалентны представлениям линейной алгебраической группы $(\pi_1/\mathfrak{Z}^N) \otimes \mathbb{Q}$, удобно рассмотреть проективный предел π_1^B при $N \to \infty$.

Рассмотрим сначала группу π_1. Мы будем изучать только унипотентные представления этой группы. Это эквивалентно изучению представлений некоторой унипотентной алгебраической группы. Мы будем говорить, что это π_1 в форме Бетти. В алгебраической геометрии слова «в форме Бетти» и «в форме де Рама» означают соответственно объекты, наделенные обычной топологией, и объекты, изучаемые специфическими средствами алгебраической геометрии. Перейдем теперь к изучению представлений π_1 в форме де Рама, то есть с помощью комплексов алгебраических дифференциальных форм.

Рассмотрим, следуя Квиллену, группу π_1 вместе с ее групповой алгеброй $\mathbb{Q}[\pi_1]$. Представления группы π_1 это то же самое, что модули

над алгеброй $\mathbb{Q}[\pi_1]$. Унипотентные представления соответствуют модулям над групповой алгеброй, в которых для каждого элемента группы γ некоторая степень оператора $\gamma - 1$ равна нулю. Например, это будет так, если действие оператора γ выражается верхнетреугольной матрицей, на диагонали которой стоят единицы. Такие представления являются модулями над факторами групповой алгебры по некоторой степени I^N идеала I, порожденного элементами вида $\gamma - 1$.

Имея два представления, мы можем взять их тензорное произведение. Для этого рассмотрим отображение $\Delta\colon \mathbb{Q}[\pi_1] \to \mathbb{Q}[\pi_1] \otimes \mathbb{Q}[\pi_1]$ (коумножение), которое переводит каждый элемент $\gamma \in \pi_1$ в $\gamma \otimes \gamma$. Рассмотрим теперь два представления, т. е. модуля над $\mathbb{Q}[\pi_1]$. Их тензорное произведение является модулем над $\mathbb{Q}[\pi_1] \otimes \mathbb{Q}[\pi_1]$, и мы можем воспользоваться отображением Δ, чтобы превратить его в модуль над $\mathbb{Q}[\pi_1]$.

Коумножение Δ можно определить как отображение

$$\Delta\colon \mathbb{Q}[\pi_1]/I^{2N} \to (\mathbb{Q}[\pi_1]/I^N) \otimes (\mathbb{Q}[\pi_1]/I^N).$$

Невозможно, однако, поставить в левой части I^N. Следовательно, чтобы определить тензорное произведение унипотентных модулей, нужно перейти к пределу при $N \to \infty$. Обозначая $\widehat{\mathbb{Q}[\pi_1]}$ проективный предел факторов $\mathbb{Q}[\pi_1]/I^N$, и символом $\widehat{\otimes}$ пополненное тензорное произведение, получим отображение

$$\Delta\colon \widehat{\mathbb{Q}[\pi_1]} \to \widehat{\mathbb{Q}[\pi_1]} \,\widehat{\otimes}\, \widehat{\mathbb{Q}[\pi_1]}.$$

В алгебраической геометрии вместо группы U рассматривают алгебру $\mathcal{O}(U)$ полиномиальных функций на ней. Произведение в группе соответствует коумножению $\mathcal{O}(U) \to \mathcal{O}(U) \otimes \mathcal{O}(U)$ — при этом многочлену f на группе соответствует многочлен $F(u, v) = f(uv)$ на произведении групп.

Представлению группы соответствует комодуль над групповой алгеброй; координатами являются функции на группе. Чтобы перейти от представления к копредставлению, рассмотрим двойственный модуль

$$(\widehat{\mathbb{Q}[\pi_1]})^\vee = \varinjlim (\mathbb{Q}[\pi_1]/I^N)^\vee,$$

где в правой части стоит индуктивный предел двойственных модулей.

Для произвольной алгебраической группы возникает вопрос, что такое полиномиальная функция на ней. Определим для произвольной функции f на группе и произвольного элемента γ группы новую функцию $\Delta_\gamma f(x) = f(\gamma x) - f(x)$. Тогда полиномиальными называются все функции f, которые обнуляются достаточно длинной последовательностью произвольных операторов Δ_γ: $\Delta_{\gamma_1} \ldots \Delta_{\gamma_n} f = 0$.

С помощью теории де Рама можно получить альтернативное описание того, что такое унипотентное представление группы π_1. Произвольное представление группы π_1 это то же самое, что и локальная система векторных пространств над $\mathbb{C}$. Унипотентное представление — это результат нескольких последовательных расширений тривиальной локальной системы. Иными словами, это голоморфное векторное расслоение с голоморфной плоской связностью $\nabla = d - \omega$, где ω — функция со значениями в эндоморфизмах слоев.

Рассмотрим теперь проективную прямую с выколотыми точками $0, 1, \infty$, и на ней голоморфное расслоение с голоморфной плоской связностью. Такое расслоение полностью определяется монодромией связности в окрестности выколотых точек; поскольку мы рассматриваем унипотентные представления, эта монодромия T должна быть унипотентной. Следовательно, оператор $N = \log T$ будет нильпотентен.

Существует естественный способ выбрать базис в сечениях рассматриваемого расслоения. Мы начинаем с некоторой точки слоя, и продолжаем ее до ковариантно постоянного сечения, применяя к нему одновременно оператор $\exp \frac{\log z - N}{2\pi i}$, где z — параллельный перенос по отношению к связности. Ясно, что если сделать полный оборот вокруг выколотой точки, то получится единичный оператор, и тем самым мы получим тривиализацию — базис глобальных сечений — рассматриваемого расслоения. В выколотых точках базисные сечения будут иметь простые полюса.

Операции расширения одного представления посредством другого

$$0 \to V' \to V \to V'' \to 0$$

соответствует операция расширения векторных расслоений. Унипотентное представление получается в результате конечной последовательности расширений из тривиального представления. На проективной прямой $H^1(\mathcal{O}) = 0$, что означает, что расширение тривиального расслоения тривиальным тривиально. Таким образом, все унипотентные представления соответствуют тривиальным векторным расслоениям.

Рассмотрим расслоение $\mathcal{O}^N$ и на нем связность $d + \omega$, где ω — 1-форма с логарифмическими полюсами. На проективной прямой без трех точек базис в пространстве таких форм доставляют формы $\frac{dz}{z}$ и $\frac{dz}{1-z}$.

Рассмотрим теперь в качестве примера $\mathbb{P}^1 \setminus S$, где S — конечное множество. В этом случае, начав с унипотентного представления группы π_1, мы приходим к тривиальному расслоению на $\mathbb{P}^1$. Тривиальное расслоение на полном многообразии $\mathbb{P}^1$ представляет собой векторное пространство $V^{DR} \otimes \mathcal{O}$. Действительно, глобальные голоморфные функции на таком

многообразии являются константами, поэтому две тривиализации расслоения отличаются на константу. Если мы рассмотрим теперь, как раньше, голоморфное расслоение со связностью, и некоторый изоморфизм между слоями над двумя разными точками, то этот изоморфизм должен быть параллельным переносом.

Рассмотрим некоторый базис $\omega_1, \ldots, \omega_r$ в пространстве 1-форм с логарифмическими полюсами. Тогда 1-форма со значениями в эндоморфизмах представляется в виде $\sum e_i \omega_i$, где $e_1, \ldots, e_r$ — эндоморфизмы векторного пространства. Так, на проективной прямой без трех точек всякая связность имеет вид $d - e_0 \frac{dz}{z} - e_1 \frac{dz}{1-z}$. Таким образом, множество всех связностей является модулем над свободной алгеброй, порожденной e_0 и e_1. В силу унипотентности можно вместо свободной алгебры рассмотреть ее фактор $\mathbb{C}\langle\langle e_0, e_1\rangle\rangle / I^N$.

Таким образом, с точки зрения теории де Рама унипотентные представления группы π_1 — это модуль V над алгеброй ассоциативных некоммутативных многочленов от переменных e_0, e_1, которому соответствует векторное расслоение со связностью, причем из унипотентности следует, что модуль должен убиваться достаточно большой степенью идеала I. Это дает возможность рассматривать V как модуль над алгеброй $\mathbb{C}\langle\langle e_0, e_1\rangle\rangle$ формальных степенных рядов. Соответствие устанавливается так: рассматривается тривиальное расслоение $V \otimes \mathcal{O}$ со связностью $d - e_0 \frac{dz}{z} - e_1 \frac{dz}{1-z}$. Локальной системой будет теперь система ковариантно постоянных сечений этого расслоения.

Если имеются два векторных расслоения со связностями, $(v, d - \omega)$ и $(W, d - \eta)$, то можно определить их тензорное произведение формулой $(V \otimes W, d - \omega \otimes 1 - 1 \otimes \eta)$. Это равносильно тому, чтобы рассматривать модули над свободной алгеброй Ли $\mathrm{FL}(e_0, e_1)$ с обычным тензорным произведением модулей.

25 апреля 2002 г.

С. Б. Каток

ВСЁ, ЧТО ВАМ ХОТЕЛОСЬ БЫ УЗНАТЬ О МАТРИЦАХ ВТОРОГО ПОРЯДКА

Эквивалентные матрицы

Я буду говорить про модулярную группу $\mathrm{SL}(2, \mathbb{Z})$. Это — группа матриц $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ с целочисленными элементами ($a, b, c, d \in \mathbb{Z}$) и определителем 1 ($ad - bc = 1$). Вопрос, который я хочу задать, следующий: «Когда матрицы $A, B \in \mathrm{SL}(2, \mathbb{Z})$ сопряжены над $\mathbb{Z}$, т. е. когда существует матрица $C \in \mathrm{SL}(2, \mathbb{Z})$, для которой $B = C^{-1}AC$?» (Для матриц, сопряжённых над $\mathbb{Z}$, мы будем использовать обозначение $A \sim B$.) Понятно, что если $A \sim B$, то $\mathrm{tr}\, A = \mathrm{tr}\, B$. Обратное, вообще говоря, неверно. Если взять две целочисленные матрицы с одинаковыми следами, то они с большой вероятностью окажутся не сопряжёнными над $\mathbb{Z}$. Матрицы A и B, конечно, всегда сопряжены над $\mathbb{Q}$, потому что они имеют одинаковые собственные значения λ и $\frac{1}{\lambda}$. Но над $\mathbb{Z}$ матрицы могут быть не сопряжены.

Я формулирую четыре задачи.

1. Доказать, что матрицы $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ и $B = \begin{pmatrix} 0 & -1 \\ 1 & 3 \end{pmatrix}$ сопряжены над $\mathbb{Z}$.

2. Доказать, что любые две матрицы из $\mathrm{SL}(2, \mathbb{Z})$ со следом 3 сопряжены над $\mathbb{Z}$.

3. Пусть $A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix}$ и $D = \begin{pmatrix} 0 & -1 \\ 1 & 4 \end{pmatrix}$. Какие из этих матриц сопряжены над $\mathbb{Z}$?

4. Найти необходимое и достаточное условие того, что матрицы $A, B \in \in \mathrm{SL}(2, \mathbb{Z})$ с одинаковыми следами сопряжены над $\mathbb{Z}$.

Нас будет интересовать только случай гиперболических матриц, т. е. тот случай, когда $|\mathrm{tr}\, A| > 2$.

Первую задачу решить нетрудно. Легче найти матрицу, которая сопрягает две матрицы, чем доказать, что такой матрицы не существует. Будем искать матрицу $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ непосредственно из соотношения

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 3 \end{pmatrix}.$$

Возникает система из четырёх линейных уравнений с четырьмя неизвестными. Эта система будет ранга 2. Её нужно решить в целых числах. Это не всегда бывает легко, но если решение есть, то найти его можно. Ещё есть дополнительное условие $ad - bc = 1$. Немножко повозившись, мы находим ответ: $C = \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}$. Это не даёт нам никакого интересного взгляда на задачу.

Чтобы решить задачи 2 и 3, интересно обнаружить связь этой задачи с квадратичными формами. Каждой гиперболической матрице $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \in \mathrm{SL}(2, \mathbb{Z})$ сопоставляется знаконеопределённая квадратичная форма $Q_A(x, y) = cx^2 + (d - a)xy - by^2$ с положительным дискриминантом $D = = (a + d)^2 - 4 > 0$. Довольно скоро будет видно, как эта форма возникает. Матрицы $A, B \in \mathrm{SL}(2, \mathbb{Z})$ сопряжены в $\mathrm{SL}(2, \mathbb{Z})$ тогда и только тогда, когда соответствующие формы Q_A и Q_B эквивалентны в узком смысле (т. е. посредством матрицы из $\mathrm{SL}(2, \mathbb{Z})$). В теории чисел есть понятие *числа классов идеалов* (в узком смысле) $h(D)$; в данном случае — это просто число неэквивалентных (в узком смысле) квадратичных форм с дискриминантом D. Если $\operatorname{tr} A = a + d = 3$, то $D = 3^2 - 4 = 5$. Известно, что $h(5) = 1$, поэтому все формы сопряжены, значит, все матрицы со следом 3 эквивалентны. Тем самым задача 2 решена. В задаче 3 дискриминант равен 12, а $h(12) = 2$.

Оказывается, что задачи 3 и 4 легко решить с помощью геометрической интерпретации. Рассмотрим теперь группу $\mathrm{SL}(2, \mathbb{R})$ не только целочисленных матриц, но и всех вещественных матриц с определителем 1. Группа $\mathrm{SL}(2, \mathbb{R})$ действует на верхней полуплоскости $\mathcal{H} = \{z \in \mathbb{C}\colon \operatorname{Im} z > 0\}$ преобразованиями Мёбиуса

$$z \mapsto \gamma(z) = \frac{az + b}{cz + d}, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z}).$$

Эти преобразования являются изометриями плоскости Лобачевского $\mathcal{H}$ с метрикой $ds = \dfrac{\sqrt{dx^2 + dy^2}}{y}$. Преобразования Мёбиуса бывают трёх видов:

• эллиптические: $|a + d| < 2$ (у них есть одна неподвижная точка на верхней полуплоскости $\mathcal{H}$; другая неподвижная точка лежит на нижней полуплоскости);

• параболические: $|a + d| = 2$ (у них есть одна неподвижная точка на $\partial\mathcal{H} = \mathbb{R} \cup \{\infty\}$);

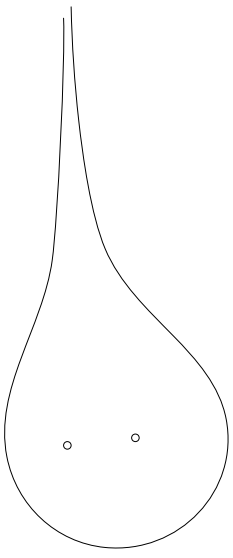
• гиперболические: $|a + d| > 2$ (у них есть две неподвижных точки на $\partial\mathcal{H} = \mathbb{R} \cup \{\infty\}$).

Нас интересует именно гиперболический случай. Чтобы найти неподвижные точки, нужно просто решить уравнение $\dfrac{az + b}{cz + d} = z$. Это даёт

нам именно ту квадратичную форму, о которой шла речь выше: $cz^2 + (d-a)z - b = 0$. Если $c \neq 0$, то получается квадратное уравнение с двумя действительными корнями $z_\pm = \frac{a-d \pm \sqrt{D}}{2c}$ и дискриминантом $D = (a+d)^2 - 4 > 0$. Если мы соединим эти два корня полуокружностью с центром на действительной оси, то эта полуокружность будет геодезической на плоскости Лобачевского, и эта геодезическая будет сохраняться под действием преобразования γ. Эта геодезическая называется *осью* гиперболического преобразования. Корни будем обозначать u и w. При этом будем считать, что точка w притягивающая, т. е. $\gamma'(w) = \frac{1}{(cw+d)^2} < 1$, а точка u отталкивающая, т. е. $\gamma'(u) > 1$.

Если $c = 0$, то вместо квадратного уравнения мы получаем линейное уравнение. В этом случае получаем две неподвижные точки: одна принадлежит $\mathbb{R}$, а другая — это точка ∞.

Если $\gamma \in \mathrm{SL}(2, \mathbb{Z})$, то его ось $c(\gamma)$ становится геодезической на модулярной поверхности $\mathrm{SL}(2, \mathbb{Z}) \setminus \mathcal{H}$. Модулярная поверхность — это фактор верхней полуплоскости по модулярной группе. Классы сопряжённых гиперболических элементов в $\mathrm{SL}(2, \mathbb{Z})$ находятся во взаимно однозначном соответствии с замкнутыми геодезическими на модулярной поверхности. Картинка 1 представляет собой модулярную поверхность. Две точки символизируют тот факт, что это не многообразие. В группе $\mathrm{SL}(2, \mathbb{Z})$ есть два эллиптических преобразования: одно порядка 2, а другое порядка 3. Они дают две специальные эллиптические точки на поверхности, где угол не 2π, а меньше. Конец, называемый каспом, символизирует тот факт, что модулярная поверхность некомпактна (но имеет конечный объём).



Р и с. 1. Модулярная поверхность

Теперь я хочу сказать несколько слов о связи с квадратичными формами. Кажется, что она идёт только в одну сторону: гиперболической матрице соответствует квадратичная форма, введённая на с. 153. Чтобы увидеть связь в обратном направлении, квадратичной форме с положительным дискриминантом D сопоставляется геодезическая в $\mathcal{H}$, соединяющая корни соответствующего квадратного уравнения. Её образ на модулярной поверхности $\mathrm{SL}(2, \mathbb{Z}) \setminus \mathcal{H}$ будет замкнут. Это вытекает из следующего несложного теоретико-числового рассмотрения. Множество всех рациональных матриц, имеющих ту же ось, что и наша геодезическая, — это действительное

квадратичное поле

$$K = \mathbb{Q}(\sqrt{D}) = \{\lambda\alpha + \mu,\ \lambda \in \mathbb{Q}^*,\ \mu \in \mathbb{Q}\},$$

где $\alpha \in \mathrm{GL}(2, \mathbb{Z})$ — некоторый элемент с той же осью. Мы всегда можем сопоставить квадратичной форме много матриц из $\mathrm{GL}(2, \mathbb{Z})$ с той же самой осью. Вопрос лишь в том, есть ли среди них матрица с определителем 1. Оказывается, что есть: она будет соответствовать единице нормы 1 в этом квадратичном поле.

Это очень удобно и приятно, потому что есть знаменитая теория Гаусса приведения квадратичных форм. Ей соответствует теория, выраженная на матричном языке.

Следующий ингредиент этой теории — это так называемая теория «$-$» цепных дробей. Рассмотрим последовательность целых чисел n_0, n_1, n_2, …, n_s, …, где $n_i \in \mathbb{Z}$ и $n_i \geqslant 2$ при $i \geqslant 1$. Тогда можно написать конечную дробь

$$(n_0, n_1, \ldots, n_s) := n_0 - \cfrac{1}{n_1 - \cfrac{1}{n_2 - \ldots \cfrac{1}{n_{s-1} - \cfrac{1}{n_s}}}}$$

и можно рассмотреть предел $(n_0, n_1, \ldots, n_s, \ldots) = \lim\limits_{s \to \infty} (n_0, n_1, \ldots, n_s) = \alpha \in \mathbb{R}$. Нетрудно видеть, что предел у этой последовательности всегда существует. Обратно, если у вас есть любое число $\alpha \in \mathbb{R}$, то так же, как его можно разложить в обычную цепную дробь, его можно разложить и в «$-$» цепную дробь. Только вместо того, чтобы на каждом шаге брать целую часть, нужно брать целую часть плюс 1: $n_0 = [\alpha] + 1$, $n_i = [\alpha_i] + 1$, где $\alpha_{i+1} = \dfrac{1}{n_i - \alpha_i}$.

В каком-то смысле эта теория удобнее обычной теории цепных дробей. В обычной теории цепных дробей подходящие дроби поочерёдно то больше, то меньше числа α. А в этой теории последовательность подходящих дробей монотонна.

Свойства «$-$» цепных дробей очень похожи на свойства обычных цепных дробей.

1. В обычной теории цепных дробей, если $\alpha \in \mathbb{Q}$, то цепная дробь конечна. Здесь же всегда будут бесконечные последовательности. Рациональному числу α соответствует последовательность, в которой начиная с некоторого номера идут все двойки (например, для числа 1 разложение состоит только из двоек).

2. Число α является квадратичной иррациональностью тогда и только тогда, когда его разложение периодично с какого-то места. Это свойство такое же, как для обычных цепных дробей.

3. Число α имеет чисто периодическую цепную дробь тогда и только тогда, когда α — квадратичная иррациональность, для которой $\alpha > 1$ и $0 < \alpha' < 1$ (α' — сопряжённое число).

4. $\alpha = C\beta$ (т. е. α и β связаны дробно-линейным преобразованием из $\mathrm{SL}(2, \mathbb{Z})$) тогда и только тогда, когда периоды разложений α и β в «$-$» цепные дроби отличаются циклической перестановкой.

Для обычных цепных дробей в свойстве (4) участвует не $\mathrm{SL}(2, \mathbb{Z})$, а $\mathrm{GL}(2, \mathbb{Z})$. Поэтому теория «$-$» цепных дробей более приспособлена для группы $\mathrm{SL}(2, \mathbb{Z})$.

Есть связь между разложениями в «$-$» цепную дробь и в обычную цепную дробь. Эта связь не очень красивая, но чёткая: по одному разложению можно написать другое, и наоборот.

Теперь мы уже довольно близки к ответу на 4-й вопрос.

Т е о р е м а 1. *Две гиперболические матрицы* $A, B \in \mathrm{SL}(2, \mathbb{Z})$ *с одинаковым следом сопряжены в* $\mathrm{SL}(2, \mathbb{Z})$ *тогда и только тогда, когда притягивающие неподвижные точки* w_A *и* w_B *имеют периоды разложения в «$-$» цепную дробь, которые отличаются циклической перестановкой.*

Д о к а з а т е л ь с т в о. Первый шаг в доказательстве следующий. Если есть две матрицы $A, B \in \mathrm{SL}(2, \mathbb{Z})$, которые имеют одну общую неподвижную точку, то их вторые неподвижные точки тоже совпадают. Это следует просто из дискретности группы.

Второй шаг. Если периоды w_A и w_B совпадают (с точностью до циклической перестановки), то согласно свойству (4) существует матрица $C \in \mathrm{SL}(2, \mathbb{Z})$, для которой $w_A = Cw_B$. Тогда матрицы CBC^{-1} и A оставляют точку w_A неподвижной, и согласно первому шагу доказательства они имеют одну и ту же ось. Так как их следы равны, $CBC^{-1} = A$ или $CBC^{-1} = A^{-1}$. Обе точки w_A и w_B притягивающие, поэтому точка w_A притягивающая для A и для CBC^{-1}, но не для A^{-1}. Поэтому A^{-1} отпадает и $CBC^{-1} = A$.

В обратную сторону понятно. Если $A \sim B$, то $CBC^{-1} = A$, а потому $w_A = Cw_B$. Согласно свойству (4) периоды w_A и w_B совпадают (с точностью до циклической перестановки). □

Теперь мы можем ответить на вопрос 3. Возьмём матрицы A, B и D, вычислим притягивающие точки и сравним их периоды разложения в «$-$» цепную дробь. Мы получим $w_A = (1, 4, 4, \ldots)$, значит, период (4); $w_B = (1, 3, 2, 3, 2, \ldots)$, значит, период (3, 2); $w_D = (0, 4, 4, \ldots)$, значит, период (4). Таким образом, $A \not\sim B$, $A \sim D$, $B \not\sim D$.

Кодирование замкнутых геодезических на модулярной поверхности

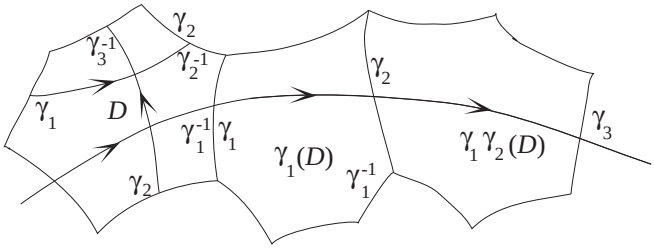
Первым делом я хочу определить арифметический код матрицы $A \in \in \mathrm{SL}(2, \mathbb{Z})$. Это как раз в точности следует из первой части моего доклада. Если есть матрица $A \in \mathrm{SL}(2, \mathbb{Z})$, то можно написать разложение в «$-$» цепную дробь её притягивающей точки w_A. Матрица у нас с целыми коэффициентами, поэтому w_A — квадратичная иррациональность. Разложение w_A в «$-$» цепную дробь периодическое:

$$w_A = (n_1, n_2, \ldots, n_k, \overline{n_{k+1}, \ldots, n_{k+m}}).$$

Этот период и будет арифметическим кодом. Арифметический код — это последовательность натуральных чисел $\geqslant 2$.

Я хочу напомнить теорию приведения Гаусса, переформулировав её в терминах матриц. Для любой гиперболической матрицы $A \in \mathrm{SL}(2, \mathbb{Z})$ существует матрица $C \in \mathrm{SL}(2, \mathbb{Z})$, для которой $w_{CAC^{-1}} = (\overline{n_{k+1}, \ldots, n_{k+m}})$ чисто периодическая дробь. Значит, $w_{CAC^{-1}} > 1$ и $0 < u_{CAC^{-1}} < 1$. Матрица CAC^{-1} называется *приведённой* (т. е. матрица называется приведённой, если её неподвижные точки обладают этим свойством). На самом деле, теория приведения Гаусса говорит нам больше: она говорит, как такое приведение выполняется алгоритмически.

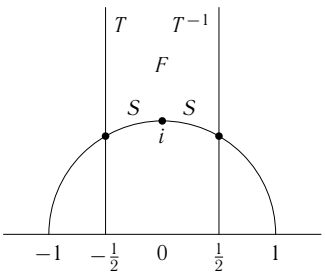
Есть ещё общая конструкция для фуксовых групп, которая сопоставляет матрице из $\mathrm{SL}(2, \mathbb{Z})$ некий другой код — геометрический. Сначала я изложу его для фуксовых групп, а потом покажу, как он становится численным кодом. Если у вас есть конечно порождённая фуксова группа 1-го рода, то у неё существует фундаментальная область D, которая является многоугольником с чётным числом сторон. (Если там есть эллиптический элемент порядка 2, то он делит сторону на две части; всегда можно считать, что число сторон чётно.) Стороны фундаментального многоугольника D отождествляются при помощи образующих элементов группы (рис. 2). Если элемент γ_i отождествляет две стороны D, мы помечаем первую сторону элементом γ_i, а вторую — элементом γ_i^{-1}.



Р и с. 2. Геометрический код

Верхнюю полуплоскость можно замостить образами фундаментальной области D. Стороны всех образов я помечу так же, как они были помечены на многоугольнике D. Пусть теперь у меня есть любая геодезическая. Я хочу её закодировать при помощи образующих γ_1, γ_2, γ_3, γ_1^{-1}, γ_2^{-1}, γ_3^{-1} (в случае, изображённом на рис. 2). Я вхожу в фундаментальную область и смотрю, какую её сторону пересекает геодезическая. Первый элемент кода будет γ_1. Потом смотрим, какую сторону пересекает геодезическая в следующей области. В данном случае это будет γ_2. Потом будет γ_3 и т. д. Заметьте, что образ фундаментальной области D, который примыкает к D по стороне, обозначенной γ_1, это будет просто $\gamma_1(D)$. Следующий образ будет $\gamma_1\gamma_2(D)$ — именно в этом порядке. И так далее. Вообще говоря, для произвольной геодезической мы получим бесконечный код. Если геодезическая замкнута, то её можно выразить как произведение элементов кода. Построенный таким образом код инвариантен относительно выбора фундаментальной области. Код определён с точностью до циклической перестановки. Если у нас есть геодезическая $\gamma = \gamma_1\gamma_2\ldots\gamma_n$, то $\gamma_1^{-1}\gamma\gamma_1 = \gamma_2\ldots\gamma_n\gamma_1$.

Эта конструкция общая для фуксовых групп. Этот код называют *кодом Морса*. Теперь можно посмотреть, что получится для модулярной группы $\Gamma = \mathrm{PSL}(2, \mathbb{Z})$. Рассмотрим стандартную фундаментальную область F (рис. 3). Её можно рассматривать как четырёхугольник. Две вертикальные стороны отождествляются при помощи преобразования $T(z) = z + 1$, а две дуги отождествляются при помощи преобразования $S(z) = -\frac{1}{z}$. Если мы проделаем нашу конструкцию и будем кодировать геодезическую при помощи этих образующих, то у нас получится последовательность, в которой встречаются несколько T, потом S, потом несколько T^{-1}, потом S, потом снова несколько T и т. д. (несколько S подряд встретиться не могут, поскольку S^2 — тождественное преобразование; из этого также следует, что $S^{-1} = S$). Мы посчитаем количество элементов T, которые стоят подряд, и напишем соответствующее целое число со знаком плюс, и посчитаем количество элементов T^{-1}, которые стоят подряд, и напишем соответствующее целое число со знаком минус. Тем самым мы сопоставим матрице из $\mathrm{SL}(2, \mathbb{Z})$ другой код, который тоже состоит из целых чисел, но они могут быть как положительными, так и отрицательными.



Р и с. 3. Стандартная фундаментальная область

Вопрос состоит в том, для каких матриц $A \in \mathrm{SL}(2, \mathbb{Z})$ эти коды совпадают: $(A) = [A]$? Ответ довольно неожиданный; он даётся в терминах арифметического кода.

Т е о р е м а 2 (С. Каток, 1996). *Для матрицы* $A \in \mathrm{SL}(2, \mathbb{Z})$ *с арифметическим кодом* $(A) = (n_1, n_2, \ldots, n_m)$ *арифметический код* (A) *совпадает с геометрическим кодом* $[A]$ *тогда и только тогда, когда* (A) *не содержит* 2 *и не содержит пар* $\{3, 3\}$, $\{3, 4\}$, $\{3, 5\}$, $\{4, 3\}$ *и* $\{5, 3\}$.

Как эти пары сюда попали, совершенно непонятно *).

Теперь я хочу немножко уточнить теорию Гаусса приведения квадратичных форм. Во-первых, любой арифметический код реализуется некоторой матрицей. Если взять набор чисел $n_1, \ldots, n_m \geqslant 2$, то матрица $A = T^{n_1} S T^{n_2} S \ldots T^{n_m} S$, если она гиперболическая, приведённая с притягивающей точкой $w_A = (\overline{n_1, \ldots, n_m})$, которая является чисто периодической «$-$» цепной дробью. Все приведённые матрицы $B \sim A$ составляют так называемый A-цикл. Это в точности матрицы, которые получаются из матрицы A стандартным сопряжением, т. е. $A_k = T^{n_{k+1}} S T^{n_{k+2}} S \ldots T^{n_{k+m}} S$, где $(n_{k+1}, \ldots, n_{k+m})$ — циклическая перестановка $(n_1, \ldots, n_m)$.

Следующая теорема немножко объясняет теорему 2.

Т е о р е м а 3. *Следующие утверждения для матрицы* A *эквивалентны*:

(1) $[A] = (A)$;

(2) *оси всех матриц в* A*-цикле пересекают стандартную фундаментальную область* F;

(3) *все отрезки замкнутой геодезической, соответствующей классу эквивалентности* A *на* F, *направлены по часовой стрелке, т. е. в положительном направлении* (*я назвала такие геодезические положительными*).

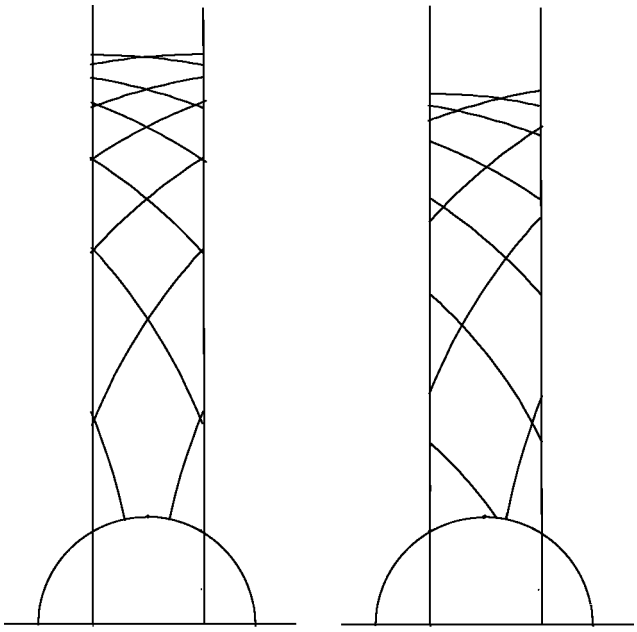
Давайте нарисуем геодезическую. Она сначала поднимается между параллельными прямыми, а затем постепенно спускается вниз и ударяется в дугу окружности. Это может произойти двумя способами (рис. 4). Случай, изображённый на рис. (б) нам не годится.

Положительным геодезическим можно дать объяснение, если рассмотреть геодезические на SM — единичном касательном расслоении над $M = \mathrm{SL}(2, \mathbb{Z}) \setminus \mathcal{H}$. Положительные геодезические — это в точности те, которые содержатся в положительном полупространстве

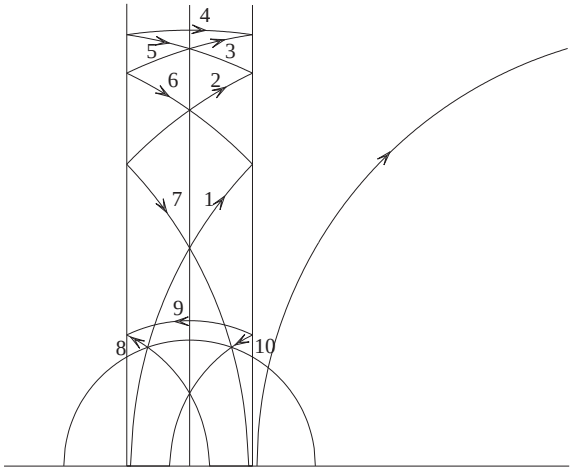
$$S^+M = \{(z, \zeta) \in SM : \ \mathrm{Re}\ \zeta > 0\},$$

где (z, ζ) — стандартные координаты на SM: $z \in \mathcal{H}$, $\zeta \in \mathbb{C}$, $|\zeta| = \mathrm{Im}\ z$.

*) Заметим, что эти пары соответствуют правильным многогранникам в $\mathbb{R}^3$. — *Прим. ред.*



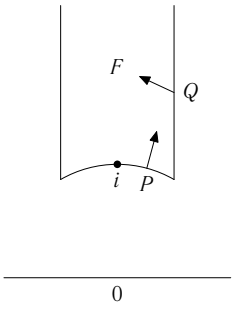
Р и с. 4. Два примера геодезических



Р и с. 5. Пример геодезической

Например, для матрицы $A = \begin{pmatrix} 15 & -8 \\ 2 & -1 \end{pmatrix}$ арифметический код (8, 2) не совпадает с геометрическим кодом [6, −2]. Если посмотреть на рис. 5, то можно увидеть, что не все отрезки геодезической имеют одинаковое направление (есть 3 отрезка, которые направлены против часовой стрелки).

Теория приведения распространяется на все ориентированные геодезические на модулярной поверхности. Можно обобщить теорию приведения Гаусса следующим образом. Нам понадобится определение приведённой геодезической. Геодезическая на $\mathcal{H}$, идущая из точки u в точку w, называется *приведённой*, если $w > 1$ и $0 < u < 1$. Эта теория изложена в нашей с Б. Гуревичем работе, опубликованной в Moscow Mathematical Journal (2001, V. 1. № 4. P. 569—582). Оказывается, что теория Гаусса связана с так называемым сечением на SM. Что такое сечение? Сечение — это такая поверхность на единичном касательном пучке, что любая геодезическая возвращается на неё бесконечно много раз. Сечение состоит из двух частей: множества P и множества Q. Любой элемент единичного касательного пучка, у которого базисная точка z принадлежит вертикальной стороне фундаментальной области, а касательный вектор направлен внутрь, принадлежит множеству Q. Множество P определяется немножко более хитрым образом. Базисная точка принадлежит маленькой дуге (рис. 6), а касательный вектор направлен внутрь. Но не любые касательные векторы, направленные внутрь, нам подходят. Подходят только следующие касательные векторы. Если мы выпустим геодезическую, то у неё обе точки, притягивающая и отталкивающая, должны быть положительными.



Р и с. 6. Сечение

Оказывается, что это — сечение. Любая геодезическая будет пересекать это множество бесконечное число раз. Пусть $\tilde{\gamma}_i$ — отрезок между возвращениями на $P \cup Q$. Если $\tilde{\gamma}_i$ начинается на P, то он автоматически будет приведённым, потому что у него одна точка больше 1, а другая между 0 и 1. Если же $\tilde{\gamma}_i$ начинается на Q, то он сопряжён посредством TS приведённому отрезку γ_i. Рассмотрим все приведённые отрезки $\{\gamma_i\}$. У этих приведённых геодезических будут притягивающие и отталкивающие точки. Пусть w_i — притягивающая точка геодезической γ_i, а u_i — отталкивающая точка. Приложим друг к другу две бесконечные дроби

$$w_i = (n_1, n_2, \ldots),$$
$$\frac{1}{u_i} = (n_0, n_{-1}, n_{-2}, \ldots), \quad n_i \geqslant 2,$$

и составим из них бесконечную в обе стороны последовательность $(\ldots, n_{-2}, n_{-1}, n_0, n_1, n_2, \ldots)$. Все элементы этой последовательности будут больше или равны 2, из-за того что геодезические приведённые.

Отрезок γ_i и отрезок γ_{i+1} отличаются только сдвигом:

$$\gamma_i \to (\ldots, n_{-2}, n_{-1}, n_0\ddagger, n_1, n_2, \ldots)$$
$$\gamma_{i+1} \to (\ldots, n_{-2}, n_{-1}, n_0, n_1\ddagger, n_2, \ldots)$$

Это нетрудно доказать.

Поэтому вся эта последовательность целиком может быть рассмотрена как арифметический код. Для замкнутой геодезической у нас была конечная последовательность, а для произвольной геодезической получается бесконечная последовательность. Конечные последовательности можно рассматривать как периодические бесконечные последовательности.

Эта кодировка позволяет рассмотреть символическую динамику геодезического потока на SM. Каждая геодезическая кодируется при помощи бесконечной в обе стороны последовательности:

$$X : \mathcal{N}^{\mathbb{Z}} = \{x = (n_i)_{i=-\infty}^{\infty},\ n_i \in \mathcal{N},\ i \in \mathbb{Z}\}$$

с тихоновской топологией. Алфавит $\mathcal{N} = \{n \in \mathbb{Z},\ n \geqslant 2\}$ состоит из чисел, больших или равных 2. Можно говорить о *положительных* геодезических, т. е. о геодезических, код которых не содержит 2 и не содержит пар $\{3, 3\}$, $\{3, 4\}$, $\{3, 5\}$, $\{4, 3\}$ и $\{5, 3\}$.

Есть довольно красивая формула времени первого возвращения на $P \cup Q$. Функция возвращения $f(x)$ гомологична функции $2 \log w(x)$, где $w(x)$ — притягивающая точка геодезической. Точная формула выглядит следующим образом. Геодезической $(\gamma) \in X$ сопоставляются притягивающая точка $w(x) = (n_1, n_2, \ldots)$ и отталкивающая точка $u(x) = (n_0, n_{-1}, \ldots)$. Тогда

$$f(x) = 2 \log w(x) + \log g(x) - \log g(\sigma x),$$

где

$$g(x) = \frac{w(x) - u(x)\sqrt{w(x)^2 - 1}}{w(x)^2\sqrt{1 - u(x)^2}};$$

здесь σ — левый сдвиг последовательности: $\gamma_{i+1} = \sigma(\gamma_i)$.

Множество положительных геодезических является инвариантным множеством геодезического потока, т. е. можно рассмотреть подпоток, который мы называем *положительным* геодезическим потоком. По естественной мере множество положительных геодезических имеет меру нуль. Для того чтобы оценить топологическую энтропию положительного геодезического потока рассматривается его символическое представление

как специального потока над левым сдвигом в пространстве положительных геодезических с функцией возвращения $2 \log w(x)$. В формуле для $f(x)$ функция g роли не играет, так как специальные потоки с гомологичными функциями возвращения топологически сопряжены и поэтому имеют ту же топологическую энтропию. Топологическая энтропия положительного геодезического потока будет меньше 1. Её можно оценить. Мы с Гуревичем получили хорошую оценку топологической энтропии положительного геодезического потока, но я об этом говорить уже не буду.

23 мая 2002 г.

Оглавление

Предисловие 3
А. Г. Хованский. Системы уравнений с многогранниками Ньютона общего положения 4
А. Г. Хованский. Проблема Арнольда о гиперболических поверхностях в проективных пространствах 28
С. Б. Шлосман. Геометрические вариационные задачи комбинаторики и статфизики 47
А. Н. Паршин. Локальные конструкции в алгебраической геометрии 60
А. Б. Сосинский. Может ли гипотеза Пуанкаре быть неверной? 83
С. Алескер. Теория представлений в выпуклой и интегральной геометрии 99
М. А. Цфасман. Геометрия над конечным полем 109
В. М. Бухштабер. Алгебраические многообразия полисимметрических полиномов и кольца дифференциальных операторов . . 127
Пьер Делинь. О ζ-функциях многих переменных 142
С. Б. Каток. Всё, что вам хотелось бы узнать о матрицах второго порядка 152