НЕЗАВИСИМЫЙ МОСКОВСКИЙ УНИВЕРСИТЕТ

globus ГЛОБУС

Общематематический семинар. Выпуск 1

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

Москва
Издательство МЦНМО
2004

УДК 51(06)
ББК 22.1я5
Г54

Издание осуществлено при поддержке РФФИ (издательский проект № 01-01-14022).

РФФИ

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

Г54

Вып. 1. — 2004. — 264 с. — ISBN 5-94057-068-2.

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

Первый выпуск включает доклады В. И. Арнольда, А. А. Болибруха, В. А. Васильева, С. И. Гельфанда, А. В. Зелевинского, В. Я. Иврия, Ю. С. Ильяшенко, С. К. Ландо, Ю. И. Манина, Й. Меннике, Я. Г. Синая, Б. Л. Фейгина, А. Я. Хелемского и М. А. Цфасмана.

УДК 51(06), ББК 22.1я5

ГЛОБУС

Общематематический семинар. Выпуск 1

Редактор *В. В. Прасолов*
Научный редактор *М. А. Цфасман*
Тех. редактор *В. Радионов*

Лицензия ИД №01335 от 24.03.2000 г. Подписано в печать 1.11.2004 г. Формат $70 \times 100\ {}^{1}/_{16}$. Бумага офсетная №1. Печать офсетная. Печ. л. 16,5. Тираж 800 экз.
Заказ №

Издательство Московского центра непрерывного математического образования.
121002, Москва, Большой Власьевский пер., 11. Тел. (095) 241–72–85.

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 5-94057-068-2 (Вып. 1)

© НМУ, 2004
© МЦНМО, 2004.

Предисловие

Мужайтесь, о други! Боритесь прилежно!
Хоть бой и неравен, борьба безнадежна.

Ф. И. Тютчев

Двадцатый век внес в историю математики множество замечательных достижений и совершенно преобразил ее лицо. Мир сегодняшней математики превосходит самые смелые мечты ученых прошлых эпох, и нам безусловно есть, чем гордиться. Но одна проблема, существующая на Земле с момента разрушения Вавилонской башни, а скорее даже со дня изгнания из Рая, — проблема человеческой разделенности и утраты единства — вызывает все большую тревогу. Всего пару столетий назад жили люди, понимавшие всю существовавшую в то время математику, да пожалуй и физику. Сто лет назад — почти всю. Сейчас, как бы лучшие ученые мира ни стремились уподобиться в этом величайшим из своих предшественников, это стало совершенно невозможно. Более того, даже математики, специализирующиеся в той или иной области, зачастую совсем не отдают себе отчета в том, чем занимаются их коллеги. Как ни печально, разобщение и специализация — одна из тенденций развития всей современной науки, и было бы наивно надеяться, что математика станет исключением из общего правила.

И вот, именно поэтому, у профессоров Независимого Московского университета возникло заведомо безнадежное желание с этой тенденцией побороться. Мы создали общеуниверситетский семинар по математике как единому органическому целому *) и назвали его «Глобусом». Цель семинара — преодолевать разбегание областей математики, призыв к докладчикам — рассказывать так, чтобы было понятно и интересно не только специалистам, но и тем, кто от их области очень далек.

За три года проведения семинара на «Глобусе» выступили А. В. Алексеевский, С. Алескер, В. И. Арнольд, С. Н. Артемов, В. В. Батырев, А. А. Бейлинсон, А. А. Белавин, А. А. Болибрух, О. И. Богоявленский, А. М. Бородин, В. М. Бухштабер (дважды), В. А. Васильев, А. М. Вершик, А. М. Виноградов, Э. Б. Винберг, С. Г. Влэдуц, С. И. Гельфанд, С. Г. Гиндикин, А. А. Глуцюк, Р. И. Григорчук, С. М. Гусейн-Заде, В. И. Данилов, П. Делинь (трижды), С. Ю. Доброхотов, Е. Б. Дынкин, Ю. Г. Зархин, А. В. Зелевинский, В. Я. Иврий, Ю. С. Ильяшенко, М. Э. Казарян,

*) См. Н. О. Лосский. «Мир, как органическое целое». — Петроград, 1917 (см. также переиздание в: Н. О. Лосский. Избранное. — М.: «Правда», 1991. — С. 338—480).

В. А. Калошин, С. Б. Каток, Б. Келлер, М. Л. Концевич, С. К. Ландо, Ж. Лашо, Л. Лафорг (трижды), Л. А. Левин, Д. А. Лейтес, Г. Л. Литвинов, А. С. Лосев, Ю. И. Манин (трижды), В. П. Маслов, О. Матьё, Р. А. Минлос, М. Муссауи, Н. С. Надирашвили, С. К. Нечаев, Ю. А. Неретин, В. В. Никулин, С. П. Новиков, Г. А. Ольшанский, С. Ю. Оревков, А. Н. Паршин, А. Н. Рыбко, В. Серганова, А. Г. Сергеев, Я. Г. Синай, А. Б. Сосинский, Я. Стинстра, Б. Л. Фейгин, М. В. Финкельберг, Д. Б. Фукс, Ж.-М. Фонтен, А. Я. Хелемский, А. Г. Хованский (дважды), М. А. Цфасман (дважды), П. Шапира, М. Шлихенмайер, О. В. Шварцман, В. Б. Шехтман, В. В. Шехтман, С. Б. Шлосман, М. А. Шубин. Большая часть этих докладов записана и издается.

Как и любое другое безнадежное предприятие, семинар «Глобус» своей благородной цели совершенно не достиг. То же, что получилось — перед Вами.

В первый выпуск вошли 15 докладов. Ю. С. Ильяшенко рассказывает о предельных циклах дифференциальных уравнений, в частности, о второй половине 16-й проблемы Гильберта — проблеме конечности числа предельных циклов уравнений с дробно-полиномиальной правой частью; об истории ошибок и достижений на пути её решения, о различных её модификациях, и о свянных с ней многочисленных открытых проблемах.

Доклад В. А. Васильева по топологической теории классических интегралов посвящен ветвлению контуров интегрирования — теории Пикара–Лефшеца и различным ее обобщениям. Покрываемый круг вопросов чрезвычайно широк, тут и многомерные теоремы Ньютона о неинтегрируемости плоских овалов, и классическая задача о потенциалах, и теория лакун Петровского для волновых уравнений, и вопрос о числе независимых гипергеометрических функций Гельфанда, и комплексифицированная теория Морса, и теория особенностей гладких отображений.

Конформным теориям поля посвящен доклад Б. Л. Фейгина, точнее говоря, речь в нем идет о вертексных операторных алгебрах, — объекту, точное математическое определение которого дать очень непросто. Вертексные алгебры, кроме физики, возникают в теории представлений алгебр Ли и в алгебраической геометрии (модули расслоений, в частности, инстантонов).

Я рассказываю о замечательном параллелизме между полями алгебраических чисел и алгебраическими кривыми над конечным полем и о классической задаче плотной упаковки равных непересекающихся шаров в n-мерном евклидовом пространстве, точнее говоря, о том, как строить плотные упаковки шаров методами алгебраической геометрии и теории чисел.

В докладе В. Я. Иврия изучаются задачи классической вариационной теории колебаний — задачи Дирихле и Неймана для эллиптических операторов в некоторой области. Гипотеза Г. Вейля утверждает, что для задачи Дирихле главный член асимптотики для числа частот пропорционален объему области и не зависит от ее формы. Излагаются геометрические подходы к решению этой и схожих задач.

Первый из двух докладов Ю. И. Манина посвящен некоммутативной геометрии, точнее говоря, попыткам определить квантовый вариант абелево многообразия. В отличие от устоявшеося понятия квантовой группы (квантования линейных алгебраических групп), абелевы многообразия традиционным способом не квантуются. Основная идея — рассмотреть накрывающий алгебраический тор и прокварновать его и решетку мультипликативных периодов, иными словами, построить квантовые тэта-функции. Завершается доклад очень интересными филосовскими рассуждениями о том, какой же должна быть некоммутативная геометрия и какова ее связь с коммутативной.

Я. Г. Синай излагает вероятностный подход ко второму началу термодинамики, основанный на модели двух газов в сосуде, разделенном поршнем. Оказывается, что при достаточно естественных предположениях система совершенно не обязательно приходит в равновесное состояние. Иногда возникают незатухающие осциляции. То есть, второе начало термодинамики в привычном нам виде не только не обосновывается, но и вовсе не выполняется. Следует отметить, что некоторые физики это тоже подозревали.

Многим кажется, что они знают, сколько решений может быть у приведенного квадратного уравнения $x^2 + py + q$. Доклад С. И. Гельфанда показывает, что если x, p и q — матрицы 2×2, задача становится глубоко нетривиальной. При достаточно общих параметрах имеется 6 (а не 16, как можно было бы подумать) решений, при необщих бывает и 0 и ∞, а полное решение человечеству пока неизвестно.

Второй доклад Ю. И. Манина — о геометрии и арифметике гладких кубических поверхностей. Проблема переноса на поверхности теоремы Морделла–Вейля о конечнопорожденности наталкивается на целый ряд трудностей, в первую очередь на отсутствие хорошего закона композиции точек. Так, для уравнения $x^3 + 2y^3 + 3z^3 + 4t^3 = 0$ мы не знаем, можно ли все его рациональные решения получить из одного методом Диофанта (проведением секущих и касательных). Задача эта могла бы быть поставлена если не Диофантом, то уж заведомо Эйлером, а мы даже не знаем, имеется ли конечное порождающее множество. Излагаются результаты компьютерных вычислений, некоторые отрицательные результаты —

кубические поверхности устроены совсем не так, как кривые, и сильно вырожденный «положительный» результат, который впрочем не особенно убеждает.

Доклад А. В. Зелевинского посвящен комбинаторному вычислению коэффициентов Литтлвуда–Ричардсона, задающих умножение в басисе Шура в пространстве симметричных многочленов, и их обобщений на произвольные полупростые алгебры Ли. В качестве основного комбинаторного объекта употребляются trail'ы (лыжни), соединяющие веса в конечномерном представлении. Техника же доказательств использует очень красивые идеи: во-первых, q-деформацию алгебры Ли и базис Люстига, а во-вторых, геометрическую интерпретацию при помощи тропического (идемпотентного) анализа.

В. И. Арнольд рассказывает о распространении волн в различных средах и связанных с этим задачах. Начало доклада исключительно наглядно: что мы видим, выжигая при помощи лупы, можно ли разговаривать в четномерном пространстве, почему и какую радугу мы можем наблюдать, как образуется мираж. Дальше появляется вариационной исчисление, группы отражений и алгебры Ли, теория особенностей, контактная и симплектическая геометрия, квантовая теория катастроф. Сюжет необыкновенно интересен и очень важен для математики и физики. Имеется также большое количество занимательных отступлений об истории математики, о ее преподавании, о конкретных ученых прежних и наших времен; с многими из таких утверждений автор этих строк решительно не согласен.

В докладе А. А. Болибруха рассказывается о проблеме Римана–Гильберта, то есть о 21-й проблеме Гильберта: существует ли линейное дифференциальное уравнение фуксова типа с заданными особенностями и монодромией? У этой проблемы, как и у 16-й (см. выше), была непростая история, связанная с ошибочными результатами начала двадцатого века. Докладчику удалось построить контрпример к проблеме и, наоборот, доказать это утверждение при некоторых дополнительных условиях. Но самое интересное, что задача эта оказалась задачей алгебраической геометрии, в контексте которой естественно появляются расслоения со связностью на алгебраических многообразиях.

Доклад С. К. Ландо посвящен инвариантам узлов и связанным с ними инвариантам графов. Теория узлов — одна из самых наглядных задач маломерной топологии, с которой каждый знакомится в детстве. Доклад начинается с понятного рассказа о хордовых диаграммых, весовых системах и инвариантах Васильева. Описывается связанная с ними алгебра Хопфа (она коммутативна и кокоммутативна) и ставится вопрос о системе примитивных инвариантов узлов. Поскольку на этот вопрос мы отвечать

не умеем, докладчик изучает несколько более простой, но очень близкий вопрос об инвариантах графов. Предположительно, алгебра Хопфа узлов отображается на алгебру Хопфа графов эпиморфно. Обсуждаются инварианты, связанные с хроматическим числом графа, и инварианты, происходящие из алгебр Ли.

А. Я. Хелемский рассказывает о гомологической теории банаховых алгебр и о гармоническом анализе на группах. К изучению банаховых алгебр можно подходить с точки зрения классического функционального анализа — меры, аменабельность, ядерность. А можно и с точке зрения гомологической алгебры; надо рассматривать категории модулей, определив правильные аналоги понятий проективности, инъективности и плоскости. Синтез этих подходов приводит к доказательству неаменабельности алгебры мер на непрерывной локально компактной группе.

В качестве дополнения публикуется доклад Й. Меннике о целочисленных линейных группах, прочитаный на предшественнике «Глобуса» — «Студенческих чтениях» *), и не вошедший по техническим причинам в соответствующие сборники. В нем рассказывается о конгруэнц-проблеме: верно ли, что любая нормальная подгруппа конечного индекса в группе $\mathrm{SL}_n(\mathbb{Z})$ содержит конгруэнц-подгруппу. При $n=2$ ответ отрицателен, а при бóльших n — положителен.

Спектр книги, как видите, достаточно широк.

Сборник этот создан беззаветными трудами Виктора Васильевича Прасолова, записавшего и отредактировавшего при помощи авторов все доклады.

Хотелось бы надеяться, что чтение этой книги поможет читателю стать значительно тоньше и шире **), по меньшей мере в том, что касается математики.

М. А. Цфасман

*) См.: Студенческие чтения МК НМУ. Вып. 1. — М.: МЦНМО, 2000; Вып. 2, 2001.

**) Прихожанин один в Ланкашире
Всю премудрость черпал из Псалтири.
Но попавши разок
Под дорожный каток,
Стал значительно тоньше и шире.
(Английский лимерик, перевод И. Грингольца.)

Ю. С. Ильяшенко

СТОЛЕТНЯЯ ИСТОРИЯ 16-Й ПРОБЛЕМЫ ГИЛЬБЕРТА

1.
Исторический обзор

Первая половина 16-й проблемы Гильберта относится к алгебраическим кривым. Сам Гильберт и несколько последующих поколений математиков очень эффективно её продвинули. Вторая половина относится к предельным циклам дифференциальных уравнений. По-видимому, Гильберт несколько недооценивал её сложность. Он предлагал некоторые идеи, как следует изучать алгебраические кривые. По поводу же дифференциальных уравнений Гильберт сказал следующее: «Используя тот же метод непрерывной деформации коэффициентов, следует узнать, что можно сказать о числе и расположении предельных циклов уравнения $\frac{dy}{dx} = \frac{P_n(x, y)}{Q_n(x, y)}$.» Здесь $P_n(x, y)$ и $Q_n(x, y)$ — многочлены степени n. Гильберт записывал это уравнение на проективной плоскости, инвариантно относительно выбора карты на проективной плоскости. Но я буду записывать это уравнение в аффинной карте.

Эту проблему ещё до Гильберта рассматривал Пуанкаре. Как только Пуанкаре в начале 1880-х годов поставил общую задачу создать геометрическую теорию дифференциальных уравнений, он сразу же начал с уравнений такого вида. Пуанкаре ввёл понятие предельного цикла; в формулировке 16-й проблемы Гильберта идёт речь о «предельных циклах Пуанкаре».

Предельный цикл — это изолированная замкнутая фазовая кривая дифференциального уравнения. Пуанкаре поставил вопрос о том, верно ли, что любое полиномиальное векторное поле *) имеет конечное число предельных циклов. Он получил положительный ответ в случае векторных полей общего положения, т. е. векторных полей, удовлетворяющих условиям типа неравенства.

*) Указанному выше дифференциальному уравнению можно сопоставить векторное поле $\dot{x} = P_n(x, y)$, $\dot{y} = Q_n(x, y)$.

Гильберт поставил проблему в 1900 г. Он не внёс никакого вклада в решение этой проблемы, помимо нескольких слов о том, что нужно пользоваться непрерывной деформацией коэффициентов и смотреть, как будет при этом изменяться фазовый портрет.

Обычно проблему Гильберта интерпретируют следующим образом: «Какое максимальное число предельных циклов может иметь полиномиальное векторное поле степени n на плоскости?» Первый вопрос здесь такой: «Верно ли, что любое индивидуальное полиномиальное векторное поле на плоскости имеет конечное число предельных циклов?» В 1923 г. Дюлак опубликовал большой мемуар, в котором было доказано, что ответ на этот первый вопрос положителен; для краткости это утверждение называют *теоремой конечности*. В 1980 г. мемуар Дюлака был переведён (горьковскими математиками старшего поколения) на русский язык и снабжён восторженным предисловием, в котором говорилось, что это, быть может, лучшая работа по качественной теории дифференциальных уравнений за последние 50 лет.

В 1955—57 гг. Петровский и Ландис опубликовали две статьи, в которых развивали теорию комплексифицированного полиномиального уравнения и давали оценку максимального количества предельных циклов. В частности, для квадратичного векторного поля давалась оценка $H(2) \leqslant 3$ ($H(n)$ — максимальное число предельных циклов полиномиального векторного поля степени n). В начале 60-х годов выяснилось, что работа Петровского и Ландиса содержала существенные пробелы, до сих пор не заполненные. Оценка $H(2) \leqslant 3$ оказалась неверной: был построен пример квадратичного векторного поля с четырьмя предельными циклами.

В 1981 г. были обнаружены существенные пробелы в работе Дюлака.

К 1981 г. наши познания о числе предельных циклов были лишь незначительно больше, чем в то время, когда Гильберт формулировал свою проблему.

Я сегодня расскажу о положительных результатах, которых довольно много, и о том, как идеи из разных областей математики входили в исследование проблемы Гильберта.

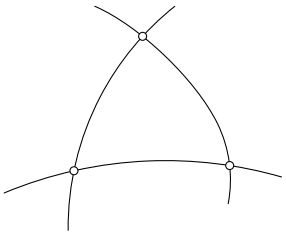
2.

Нормальные формы и теорема Дюлака

Нормальные формы — это первая, и самая специфическая, идеология, которая происходит собственно из дифференциальных уравнений и восходит к Пуанкаре. Эта идеология состоит в том, что при исследовании дифференциального уравнения его нужно не решать, а приводить заме-

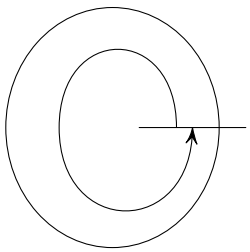
нами координат к возможно более простой, так называемой *нормальной*, форме.

Я уже упомянул, что Пуанкаре доказал теорему конечности в случае векторных полей общего положения. Его рассуждения были следующими. Если предельных циклов бесконечно много, то им нужно к чему-то накапливаться *). Они могут накапливаться, например, к периодической траектории. Они также могут накапливаться к многоугольнику, составленному из периодических траекторий — так называемому *сепаратрисному многоугольнику* (рис. 1). Но малым возмущением векторного поля сепаратрисный многоугольник можно разрушить: траектории, идущие из одной особой точки в другую, всегда могут быть разведены. Поэтому случай, когда предельные циклы накапливаются к сепаратрисному многоугольнику, Пуанкаре не рассматривает. Таким образом, в случае общего положения предельные циклы могут накапливаться либо к замкнутой траектории, либо к невырожденной особой точке (вырожденная особая точка — это тоже случай не общего положения).



Р и с. 1. Сепаратрисный многоугольник

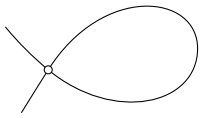
Невозможность накопления предельных циклов к замкнутой траектории следует из элементарных свойств аналитических функций — из теоремы единственности. Для этого используется конструкция, придуманная Пуанкаре. Эта конструкция заключается в том, чтобы рассмотреть отображение первого возвращения. Возьмём отрезок, трансверсальный замкнутой траектории, и посмотрим, как траектории, из него исходящие, на него возвращаются (рис. 2). Получится так называемое *отображение Пуанкаре P*. Его неподвижные точки соответствуют замкнутым траекториям; изолированные неподвижные точки — предельным циклам. Из общей теории следует, что отображение Пуанкаре полиномиального векторного поля аналитично. А из аналитичности следует, что нули функции $P(x) - x$ не могут накапливаться к внутренней точке области определения. Вот и всё доказательство теоремы Пуанкаре. Те же самые рассуждения годятся и для аналитичного векторного поля.



Р и с. 2. Отображение возвращения Пуанкаре

*) Внутри каждого предельного цикла есть особая точка, поэтому согласно теореме Безу количество предельных циклов, ни один из которых не содержится внутри другого, не превосходит n^2.

Дюлак рассматривал возможность накопления предельных циклов к сепаратрисным многоугольникам. В этом случае отображение Пуанкаре определено не по обе стороны от точки замкнутой траектории, а лишь по одну сторону. Поэтому при подходе к граничной точке отображение может не быть аналитическим. Простейший пример — петля сепаратрисы гиперболического седла (рис. 3). Для «большинства» сёдел отображение Пуанкаре, соответствующее петле сепаратрисы, имеет вид $f(x^\lambda)$, где f — аналитическая функция, λ — произвольное вещественное число.



Р и с. 3. Петля сепаратрисы гиперболического седла

Для того, чтобы доказывать, что отображение Пуанкаре не может иметь счётного множества изолированных неподвижных точек, Дюлаку понадобились нормальные формы в окрестности особых точек. Ему понадобилась также теорема о разрешении особенностей. Она была сформулирована Бендиксоном в той же самой работе (1901), в которой доказана теорема Пуанкаре—Бендиксона. Теорема о разрешении особенностей Бендиксоном была только высказана; он даже не попытался написать её законченное доказательство. Она была доказана в аналитическом случае Зайденбергом в 1968 г.; в гладком случае — Дюмортье в 1977 г., а прозрачное доказательство было придумано Ван ден Эссеном и опубликовано Маттеи и Муссю в 1980 г. Так что когда Дюлак пользовался этой теоремой, ему пришлось добавить много соображений, похожих на доказательство, но всё равно доказательства он не дал.

Подробно формулировать теорему о разрешении особенностей я не буду. Эта теорема утверждает, что в задаче о накоплении предельных циклов к сепаратрисному многоугольнику достаточно рассматривать сепаратрисные многоугольники не с произвольными сколь угодно сложными особыми точками в качестве вершин, а только с сёдлами и седлоузлами. Сёдла — это, как всем известно, особые точки с противоположными по знаку собственными значениями линеаризации. А седлоузлы — это особые точки, у которых одно из собственных значений линеаризации равно нулю, а другое не равно.

Сепаратрисные многоугольники, вершинами которых служат элементарные особые точки, называют *элементарными полициклами*.

Теория нормальных форм тогда только-только начиналась. Сам Дюлак внёс в неё очень большой вклад. Дюлак доказал следующую теорему, в которой не только не было обнаружено пробелов, но и для которой впоследствии было получено гораздо более короткое доказательство, чем у Дюлака.

Т е о р е м а (Дюлак). *Для элементарного полицикла отображение Пуанкаре P либо плоское* (*т. е. все производные в точке* 0 *равны* 0, *как у* $e^{-1/x}$), *либо вертикальное* (*обратное плоскому*), *либо полурегулярное.*

Термин *полурегулярное* весьма таинствен. Дюлак даёт определение полурегулярного отображения, а затем пишет, что к числу полурегулярных мы будем относить и другие отображения, встречающиеся в дальнейшем. Поэтому извлечь из его работы точное определение полурегулярного отображения не так уж просто. Полурегулярное отображение — это гомеоморфизм полуинтервала на другой полуинтервал с той же вершиной 0, гладкий вне вершины и допускающий асимптотическое разложение в нуле:

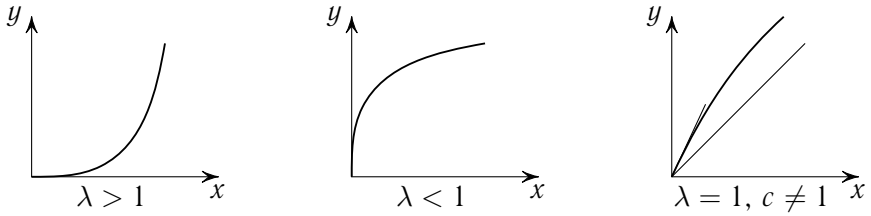
$$\hat{P}(x) = cx^{\lambda} + \sum p_j(\ln x)x^{\nu_j} \qquad (c \neq 0,\ \ \lambda \neq 0,\ \ \nu_j \to \infty). \tag{1}$$

Это означает, что с точностью до любой степени x отображение может быть приближено подходящей частной суммой этого ряда.

Из этой теоремы (повторяю — совершенно верной) Дюлак очень просто выводит теорему конечности. Сейчас я повторю этот вывод (неверный).

Л е м м а (Дюлак). *Если полурегулярное отображение P имеет бесконечно много неподвижных точек, то $P = \mathrm{id}$ — тождественное отображение.*

Д о к а з а т е л ь с т в о. Посмотрим на формулу (1). Если в главном члене показатель λ больше 1, то график функции $P(x)$ касается оси x (рис. 4). Поэтому вблизи нуля нет неподвижных точек отображения P.



Р и с. 4. График функции $P(x)$

Если λ меньше 1, то график касается оси y.

Если $\lambda = 1$, но $c \neq 1$, то график касается некоторой прямой, трансверсальной биссектрисе. В таком случае тоже не может быть счётного множества неподвижных точек.

Наконец, если $\lambda = 1$ и $c = 1$, то

$$P(x) - x = P_1(\ln x)x^{\nu_1}\bigl(1 + O(1)\bigr).$$

Вблизи нуля правая часть равенства не обращается в нуль. □

Наличие пробела в доказательстве было очень сильно замаскировано Дюлаком, который ничего не говорил об асимптотических рядах. Определение с асимптотическими рядами извлекается при формализации того, что он говорил. После того как это определение дано явно, пробел более или менее выходит на поверхность. Весь пафос состоит в том, что у асимптотического ряда может не быть первого члена. Если все коэффициенты асимптотического ряда равны нулю, т. е. если $\hat{P}(x) = x + O(x^N)$ при любом N, то рассуждение валится: в нашем разложении нет первого члена. Примером функции, формально удовлетворяющей этому условию и имеющей счётное множество неподвижных точек, служит функция $P(x) = x + e^{-1/x} \sin \frac{1}{x}$. Тем самым, полурегулярности отображения Пуанкаре не достаточно для того, чтобы получить теорему конечности.

Самое сильное впечатление, которое я пережил, когда я продумал короткое доказательство теоремы Дюлака, состояло в том, что я понял, что все рассуждения Дюлака годятся и для бесконечно гладких векторных полей. Тем самым, эти рассуждения доказывают теорему, которая не может быть верна. То, что гладкое векторное поле на плоскости не может иметь счётного числа предельных циклов, так же неверно, как и то, что гладкая функция не может иметь счётного числа нулей. Дюлак, как оказалось, пользовался только бесконечно гладкими соображениями.

Асимптотический ряд может быть тривиален: бывают сепаратрисные многоугольники, у которых отображение Пуанкаре имеет вид $P(x) = x + e^{-1/x}(1 + O(x))$. Это бывает даже для сепаратрисного многоугольника из двух седлоузлов *).

3.
Теорема конечности

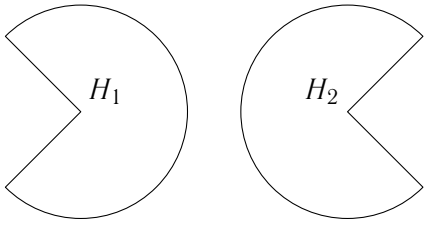
Теорему конечности доказали Экаль и я; в 1992 г. вышла книга Экаля (меньше чем в книгу доказательство этой теоремы не умещается). В 1991 г. в издательстве AMS вышла моя книга «Finiteness theorem for limit cycles» с доказательством той же теоремы.

Я расскажу только об одном новом объекте, вошедшем в комплексный анализ. Он был обнаружен независимо несколькими группами математиков: аспирантом МГУ Сергеем Ворониным и самим Экалем вместе с Мартине и Рамисом во Франции. Это — так называемые *функциональные коцепи*. Когда мы с Ворониным обдумывали задачу об аналитической

*) Простейшая модель седлоузла (точки, которая наполовину седло, а наполовину узел) задаётся системой уравнений $\dot{x} = x^2$, $\dot{y} = -y$.

классификации так называемых параболических особых точек, мы совсем не думали о проблеме конечности. Когда Мартине и Рамис обдумывали задачу об аналитической классификации особых точек типа седлоузел ($\dot{x} = x^2 + \ldots$, $\dot{y} = -y + \ldots$), они, как мне говорил потом Рамис, хорошо понимали, что делают первый шаг к доказательству теоремы конечности.

Вопрос об аналитической классификации отображений вида $z \mapsto$ $\mapsto z + z^2 + \ldots$ состоит в следующем. Напишем два разных продолжения многочлена $z + z^2$ до голоморфного отображения. Нужно узнать, при каких условиях существует аналитическая замена координат вблизи нуля, переводящая одно отображение в другое. Вот здесь и появились функциональные коцепи. А для того, чтобы выяснить, существует или нет аналитическая эквивалентность, нужно вычислить некоторые функции комплексного переменного. Инварианты аналитической классификации — это не числа, а функции.



Р и с. 5. Функциональная коцепь

Функциональные коцепи — это такие заместители ростков аналитических функций одного комплексного переменного, которые обладают похожими свойствами и которые, возможно, пригодятся и в других вопросах комплексного анализа, а не только в локальной теории аналитических дифференциальных уравнений. Простейший пример функциональной коцепи — пара аналитических функций H_1 и H_2, каждая из которых определена в секторе с вершиной в нуле; эти секторы полностью покрывают проколотую окрестность нуля, $|H_1 - H_2| = O(e^{-c/|z|})$ и обе функции допускают асимптотический ряд Тэйлора в секторе.

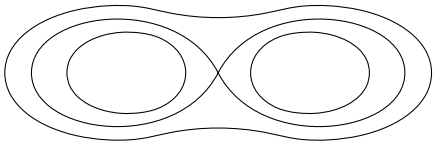
Оказывается, что в выражении для отображения Пуанкаре функциональные коцепи неизбежно возникают и они однозначно определяются своим рядом Тэйлора; описание отображения Пуанкаре требует подробного изучения таких функциональных коцепей. Для них доказываются теоремы типа теоремы единственности для аналитических функций (в этом, собственно, и состоит доказательство теоремы конечности), но технически они чрезвычайно громоздки.

Я рассказал о том, как в исследования вошли нормальные формы и новый объект одномерного комплексного анализа — функциональные коцепи. Теперь я расскажу о так называемой инфинитезимальной проблеме Гильберта, которая возникает, когда мы пытаемся облегчить себе задачу. Недавно на Математическом Обществе я рассказывал элементар-

ное доказательство теоремы Колмогорова об инвариантных торах. Грубо говоря, эта теорема обсуждает следующий вопрос. Имеем задачу, которую мы умеем решать. Что можно сказать о близких задачах? В 16-й проблеме Гильберта соответствующая постановка выглядит несравненно проще, но именно её я сейчас буду обсуждать.

4.
Римановы поверхности

Уравнения, которые мы умеем решать, это уравнения в полных дифференциалах. Если задан многочлен H_{n+1}, то уравнение $dH_{n+1}(x, y) = 0$ переписывается в виде гамильтоновой системы $\dot{x} = H_y$, $\dot{y} = -H_x$. Интегральные кривые первого уравнения, как и фазовые кривые второго уравнения, — это кривые $H = \text{const}$. Тем самым мы попадаем в область первой части 16-й проблемы Гильберта, где надо изучать линии уровня многочленов. Про эти линии уровня мы никаких вопросов не задаём; они бывают критическими, бывают овалами, бывают и некомпактные компоненты (рис. 6). На рис. 6 не изображено ни одного предельного цикла. Уравнение $\dot{x} = H_y$, $\dot{y} = -H_x$ — того же класса, с которым мы имеем дело; оно задаётся полиномиальным векторным полем (гамильтоновым). У этого векторного поля предельных циклов нет. Если фазовая кривая замкнута, то все близкие к ней фазовые кривые тоже замкнуты. Но предельные циклы присутствуют здесь неявно, они получаются при негамильтоновом возмущении. Если есть полиномиальная дифференциальная форма $\omega = A_n\,dx + B_n\,dy$ и мы рассматриваем уравнение $dH + \varepsilon\omega = 0$, то большинство замкнутых кривых разрушается, но некоторые выживают. Вопрос в том, какая из замкнутых кривых семейства $H = \text{const}$ порождает предельный цикл. Кривая, которая порождает предельный цикл, т. е. кривая, которая при возмущении остаётся замкнутой и близкой самой себе, в то время как остальные кривые размыкаются, называется *порождающей кривой*.



Р и с. 6. Линии уровня гамильтониана

Пусть γ_h — замкнутая кривая, принадлежащая множеству уровня $H = h$. Достаточное условие того, что кривая γ_h порождающая, заключается в том, что $I(h) = 0$ и $I'(h) \neq 0$, где $I(h) = \int\limits_{\gamma_h} \omega$. Эти интегралы — объект алгебраической геометрии. С их помощью вводится связность

Гаусса—Манина. Но их изучали и гораздо раньше. Это — так называемые абелевы интегралы, зависящие от параметра (параметром здесь является значение многочлена на кривой). Обычно фиксируют одну кривую, т. е. линию уровня многочлена для одного значения h, например $h=0$, и рассматривают интегралы от форм. А здесь мы рассматриваем интегралы от форм как функции от параметра.

Чем больше нулей абелева интеграла мы найдём, тем больше предельных циклов сумеем построить. Поэтому исследование абелевых интегралов позволяет что-то говорить о числе и расположении предельных циклов, помимо оценки сверху для их числа. Нас интересует следующее:

1) число (оценка снизу) и расположение нулей абелева интеграла;

2) оценка сверху для числа нулей (этот вопрос наиболее интересен).

Если гамильтоново векторное поле имеет степень n, т. е. многочлен H имеет степень $n+1$, и возмущающая форма ω тоже имеет степень n, то $I(n)$ — максимальное возможное число нулей интеграла $I(h)$ по всем вещественным семействам овалов.

Бо̀льшую часть времени я буду говорить об оценке числа нулей сверху. Но сначала поговорим об оценке снизу и о расположении овалов. В игру входит продолжение абелевых интегралов в комплексную область, теорема Пикара—Лефшеца, теорема Римана—Роха. При этом получаются чисто алгебраические утверждения типа следующей теоремы о точности: «Для типичного многочлена H_{n+1} (я не поясняю, каковы требования типичности) условие $I(h)\equiv 0$ влечёт, что $d\omega=0$.» Эту теорему для случая интегрирования по семейству так называемых исчезающих циклов, растущих из максимума или минимума многочлена H, я доказал в 1969 г. А для случая произвольных вещественных овалов теорему доказала И. Пушкарь в 1997 г. Из этой теоремы следует, что порядка $N=n^2/2$ предельных циклов можно разбросать свободно по плоскости, а именно, разместить вблизи N овалов многочлена H.

Самый интересный вопрос — оценка для числа нулей абелева интеграла сверху. Я думаю, что из сказанного уже понятно, что проблема оценки числа нулей абелева интеграла близка к 16-й проблеме Гильберта. Если мы решим 16-ю проблему Гильберта, то автоматически научимся и оценивать сверху число нулей абелевых интегралов, потому что каждый нуль порождает предельный цикл полиномиального векторного поля. Если же мы научимся оценивать сверху число нулей абелевых интегралов, то проблему Гильберта мы не решим. Однако есть все основания назвать эту проблему ослабленной, или облегчённой, проблемой Гильберта. Сейчас для этой проблемы есть два конкурирующих названия — *касательная*, или *инфинитезимальная*, проблема Гильберта.

Ещё раз подчеркну, что мы не знаем, существует ли равномерная оценка даже числа $H(2)$. Но мы знаем (и с этого началась теория малочленов), что для любого n существует равномерная оценка числа $I(n)$. Это доказали Варченко и Хованский (1984). Совсем недавно (1999) Любомир Гаврилов доказал, что $I(2)=2$. Это совпадает с оценкой снизу, которая получается тем самым способом, о котором я говорил раньше. Пока явной оценки для $I(n)$ при $n\geqslant 2$ не получено.

5.
Ограниченные варианты 16-й проблемы Гильберта

Не от хорошей жизни, но совсем не без интереса рассматриваются ограниченные варианты 16-й проблемы Гильберта, когда вопрос ставится не обо всех уравнениях вообще, а обо всех уравнениях из некоторого компакта.

В теореме о точности нужно, чтобы многочлен H был типичным. Множество типичных многочленов открыто по Зарисскому. Это открытое множество можно исчерпывать компактами. Пусть K — компакт в множестве типичных гамильтонианов H. С помощью идей, о которых я сейчас скажу несколько слов, задачу оценки числа $I(n, K)$ сверху решили Д. Новиков и Яковенко, развивая серию работ, которую 5 лет назад мы с Яковенко начали. Оценка даётся некоторой нереалистической функцией — башней из 5 экспонент, которую авторы не выписывают. Но их методы позволяют выписать эту оценку точно — так, чтобы не осталось никаких констант, про которые говорится только то, что они существуют. Коэффициенты, которые входят в эту оценку зависят от K. Точнее говоря, если K содержится в шаре радиуса R в пространстве многочленов, то в оценку входит R.

В начале этого подхода, так же, как и в основе других результатов, о которых я сейчас скажу, лежит теорема о нулях и росте голоморфных функций. Эта теорема весьма близка к известному неравенству Йенсена. Она формулируется следующим образом. Рассмотрим голоморфную функцию f в односвязном множестве U; в этом множестве есть метрика Пуанкаре. Рассмотрим компактное подмножество $K\subset U$. Пусть ρ — диаметр K в метрике Пуанкаре множества U. Тогда число нулей голоморфной функции f в множестве K не превосходит

$$e^{\rho}\ln\frac{\max\limits_{U}|f|}{\max\limits_{K}|f|}.$$

Путь, который привёл Новикова и Яковенко к башне из 5 экспонент, длинен и извилист; я не буду о нём ничего говорить.

Я хочу сказать ещё несколько слов о другой ограниченной задаче. В 70-е годы Смейл и Пью организовали семинар, который работал над 16-й проблемой Гильберта. В результате работы этого семинара появилось немного частных результатов и несколько постановок новых задач. Одним из результатов явилась проблема, которую Смейл включил в свой список проблем XXI века. В этот список Смейл включил вторую половину 16-й проблемы Гильберта и две её спецификации, два облегчённых варианта, один из которых называется *уравнение Абеля*.

6.
Уравнение Абеля

Рассмотрим уравнение, в котором расположение предельных циклов напрашивается само собой. Это — уравнение на цилиндре, и нас интересуют только те предельные циклы, которые охватывают цилиндр один раз. Уравнение будет такое:

$$\dot{x} = x^n + \sum_{j=0}^{n-1} a_j(t)x^j, \quad t \in S^1, \ \ x \in \mathbb{R}.$$

Спрашивается, что можно сказать о максимальном числе предельных циклов такого уравнения при заданном n. Ответы забавные:

n	2	3	4
$A(n)$	2	3	∞

(здесь $A(n)$ — максимальное число предельных циклов).

Если на коэффициенты не налагается никаких ограничений, кроме периодичности, то при $n \geqslant 4$ число предельных циклов уравнения Абеля может быть сколь угодно велико. Но оказалось, что если $|a_j| < C$, то тогда число предельных циклов уравнения Абеля оценивается так: $A(n, C) \leqslant e^{e^{aC^n}}$. Эта оценка тоже нереалистичная. Она очень просто получается как следствие предыдущей теоремы. Для этого нужно выйти в комплексную область по переменной x, оставив t вещественным. Тогда отображение Пуанкаре становится голоморфной функцией, несмотря на то, что коэффициенты только непрерывны. К этой голоморфной функции можно применять теорему о росте и нулях. Уравнения Абеля хорошо приспособлены для того, чтобы контролировать рост отображения Пуанкаре, поскольку отображение Пуанкаре оценивается по правой части без решения уравнения.

7.
Теория бифуркаций

Последнее, о чём я хочу рассказать, это то, как в исследование проблемы Гильберта вторглась теория бифуркаций. Я расскажу ещё одно неправильное доказательство теоремы о том, что если индивидуальное полиномиальное векторное поле имеет конечное число предельных циклов, то гильбертово число $H(n)$ существует. На этот раз обман зарыт очень неглубоко.

Пространство уравнений $\frac{dx}{dy} = \frac{P_n}{Q_n}$ отождествляется с пространством коэффициентов многочленов P_n и Q_n и тем самым получается аффинное пространство размерности $2n$. На самом деле это пространство не аффинное, а проективное, потому что уравнение не изменяется при умножении всех коэффициентов на константу. Проективное пространство компактно. Рассмотрим семейство уравнений с базой параметров B; про каждое уравнение известно, что у него есть только конечное число предельных циклов. Я предположу, что это число не является равномерно ограниченным и как бы получу противоречие. Предположим, что есть последовательность уравнений α_n, причём у уравнения α_n число предельных циклов не меньше n. У этой последовательности есть предельная точка. К этому предельному уравнению накапливаются уравнения с растущим числом предельных циклов, поэтому по теореме о непрерывной зависимости у этого уравнения бесконечно много предельных циклов.

Здесь обман уже на уровне словесного говорения. Ниоткуда не следует, что у предельного уравнения будет бесконечно много предельных циклов. Число предельных циклов могло расти, а в пределе все они могли сесть на один сепаратрисный многоугольник.

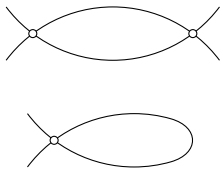
В связи с этим возникает понятие цикличности. Сепаратрисный многоугольник для краткости называют *полициклом*. *Цикличность* полицикла — это максимальное число предельных циклов, которые из него рождаются. Удобнее всего говорить о цикличности в определённом семействе; тогда речь идёт о максимальном числе при вариации параметров внутри семейства.

Гипотеза конечной цикличности (гипотеза о том, что любой полицикл имеет конечную цикличность) равносильна существованию числа $H(n)$. С этой гипотезой дела обстоят не лучше, чем с основной гипотезой Гильберта. Продвинут только случай $n = 2$. Дюмортье, Руссари и Руссо нарисовали все топологически возможные фазовые портреты векторных полей степени 2, у которых бывают полициклы. Их оказалось 121. Если дока-

зать, что для всех этих полициклов цикличность конечна, то существование $H(2)$ будет доказано.

Под влиянием некоторых гипотез Арнольда, высказанных в 1985 г., стало понятно, что правильно изучать семейства векторных полей не полиномиальные, а типичные *). Во времена Пуанкаре и Гильберта помимо полиномиальных векторных полей других почтенных конечнопараметрических семейств векторных полей наука не знала. А теперь такие семейства известны. Это — типичные семейства. Они не менее популярны, чем полиномиальные.

Рассмотрим типичное n-параметрическое семейство векторных полей на сфере (или на проективной плоскости): $X = v(x, \varepsilon)$, $x \in S^2$, $\varepsilon \in B$ (база параметров B компактна). Нас интересует, полициклы какой цикличности могут возникнуть в таких семействах. В 1-параметрическом семействе векторных полей встречаются не только векторные поля общего положения; если векторное поле выделяется каким-то одним условием на фазовый портрет, то оно может встретиться в 1-параметрическом семействе неустранимым образом. Например, в типичном 1-параметрическом семействе полицикл, изображённый на рис. 7 внизу, может встретиться, а изображённый вверху — не может.



Р и с. 7. Полициклы в 1-параметрическом семействе

Итак, разумен вопрос о том, какие бывают явления в типичных k-параметрических семействах векторных полей. Так же естествен вопрос о том, полициклы какой полицикличности встречаются в типичных k-параметрических семействах. Максимальную возможную цикличность в типичных k-параметрических семействах обозначают $B(k)$. Максимальную возможную цикличность элементарного (т. е. такого, что все его вершины — элементарные особые точки) полицикла в типичных k-параметрических семействах обозначают $E(k)$. До сих пор никто не знает, существует ли число $B(k)$ для любого k. Но для $E(k)$ доказано следующее утверждение.

Т е о р е м а (Ильяшенко и Яковенко, 1985). *Для любого k число $E(k)$ существует, т. е. в типичном k-параметрическом семействе*

*) Существуют разные определения типичности, формализующие понятие «большинства точек» в функциональном пространстве. «Топологически типичные свойства» определяются как присущие счётному пересечению открытых всюду плотных множеств. «Метрически типичные свойства» определяются как присущие почти всем точкам из гладких конечнопараметрических «типичных» семейств; существует несколько неэквивалентных определений (Колмогоров, Арнольд, ...), в которых по-разному формализуется понятие типичных семейств.

может возникнуть только полицикл, при возмущении которого родится не более $E(k)$ элементарных предельных циклов.

Недавно Вадим Калошин получил оценку $E(k) \leqslant 2^{25k^2}$ (The existential Hilbert 16-th problem and an estimate for cyclicity of elementary polycycles // Invent. Math. — 2003. — V. 151, № 3. — P. 451—512).

16 марта 2000 г.

В. А. Васильев

ВЕТВЯЩИЕСЯ ИНТЕГРАЛЫ И ТЕОРИИ ПИКАРА—ЛЕФШЕЦА

Важнейшие спецфункции математической физики имеют интегральные представления, т. е. задаются интегралами, зависящими от параметра. Таковы, в частности, фундаментальные решения большинства классических уравнений в частных производных, потенциалы Ньютона—Кулона, интегральные преобразования Фурье, гипергеометрические функции, интегралы Фейнмана, начальные данные обратных задач томографии, и т. д.

Качественное и аналитическое поведение этих функций определяется ветвлением контуров интегрирования подходящим образом комплексифицированной задачи. Это ветвление изучает теория *Пикара—Лефшеца* (являющаяся одной из важнейших компонент и источников мотивировок *теории особенностей гладких отображений*) и её различные обобщения.

В докладе рассказывается об этих теориях и их приложениях. Среди обсуждаемых результатов и тем:

— многомерные обобщения теоремы Ньютона о неинтегрируемости плоских овалов;

— обобщения теоремы Ньютона—Айвори—Арнольда об алгебраичности поверхностных потенциалов;

— вычисление количества линейно независимых общих гипергеометрических функций И. М. Гельфанда;

— теория Адамара—Петровского—Лерэ—Атиа—Ботта—Гординга резких фронтов и регулярных решений гиперболических уравнений;

— стратифицированная теория Пикара—Лефшеца как комплексификация стратифицированной теории Морса;

— ветвление интегралов многозначных форм и проблема регуляризации несобственных интегральных представлений (задаваемых интегралами по некомпактным циклам).

Во всех этих темах очень много нерешённых задач.

Более подробное изложение всего нижеследующего см. в [1], там же содержатся все необходимые ссылки.

1.
Введение

Многие специальные функции математической физики и прочих прикладных наук задаются интегральными представлениями по одной и той же схеме. Эта схема такова. Имеется гладкое локально тривиальное расслоение $E \to T$, и на пространстве этого расслоения имеется дифференциальная форма ω, которая замкнута вдоль слоёв, но не замкнута на всём пространстве. Кроме того, задана отмеченная точка $b_0 \in T$ и в слое над этой точкой фиксирован класс гомологий той же размерности, что и форма. Тогда на базе возникает функция, которая строится следующим образом. Реализуем наш класс гомологий слоя циклом. Этот цикл можно по непрерывности перетаскивать в близкие слои и интегрировать форму ω по получающимся циклам. Конечно, это перетаскивание строится неоднозначно, но классы гомологий полученных циклов определены однозначно. А следовательно и результат интегрирования зависит только от исходного класса гомологий, от способа перетаскивания он не зависит. Возникающая функция локально определена однозначно, но глобально она может ветвиться. Если фундаментальная группа базы нетривиальна, то, совершив обход вдоль нестягиваемого пути, можно прийти к другому классу гомологий. Это представление фундаментальной группы базы в гомологиях слоя называется монодромией, а его образ — группой монодромии нашего расслоения. Многие аналитические свойства возникающей функции (однозначность, алгебраичность, количество листов, регулярность) сводятся к изучению группы монодромии. Многие конкретные вопросы, о которых я буду говорить, решаются в терминах этой группы.

Стандартный случай, когда возникает эта схема, таков. Имеется отображение $\bar{E} \to \bar{T}$ чуть бóльших пространств (аналитических многообразий). По-прежнему на пространстве расслоения задана форма ω, замкнутая вдоль слоёв. Для почти всех точек базы слои топологически устроены одинаково, а следовательно над множеством таких точек циклы можно перетаскивать из слоя в слой. Но есть какое-то *дискриминантное* подмножество Σ в $\bar{T}$ (множество критических значений, слои над которыми вырождаются). Если задача комплексная (т. е. $\bar{E}$ и $\bar{T}$ — комплексные многообразия), то множество Σ имеет положительную комплексную коразмерность, а следовательно вещественную коразмерность $\geqslant 2$. Тогда можно обходить вокруг Σ вблизи его регулярных точек (где Σ устроено как гладкая гиперповерхность в $\bar{T}$) и получать при этом какое-то локально определённое ветвление групп гомологий слоя. Интересные и содержательные результаты получаются уже при изучении таких локаль-

ных ветвлений групп гомологий, а следовательно и интегральных функций, получаемых интегрированием по элементам этих групп. Рассматривавшееся выше локально тривиальное расслоение $E \to T$ — это ограничение отображения $\bar{E} \to \bar{T}$ на прообраз неособого множества $T = \bar{T} \setminus \Sigma$.

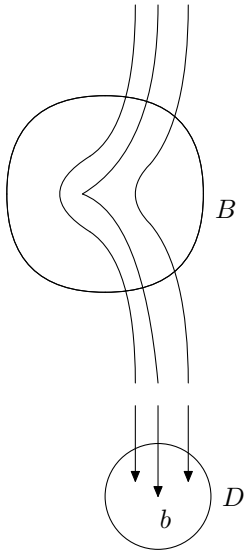
Часто возникает задача о том, что происходит с интегральной функцией при подходе к дискриминантному множеству: каково её асимптотическое поведение. Дискриминантные множества в разных науках называются по-разному — волновые фронты, видимые контуры, множества Ландау. Естественный вопрос: будет ли у интегральной функции (точнее, у её аналитического продолжения) особенность при подходе к дискриминанту. Во многих прикладных задачах возникающие функции априори вещественные (т. е. и расслоение $E \to T$ является расслоением вещественных аналитических многообразий, и дифференциальная форма интегрирования вещественнозначна). Тогда всю задачу нужно комплексифицировать и вместо того, чтобы честно интегрировать и изучать аналитическое поведение интеграла при подходе к критическому множеству, достаточно выйти в комплексную область и обойти вокруг этого множества. Если мы придём к другому циклу, причём интеграл по разности двух циклов нетривиален, то ясно, что есть ветвление. Немного повозившись, можно оценить, насколько сильна эта нерегулярность. Например, можно выяснить, будет ли функция алгебраической.

Имеется много разных задач, в которых возникает эта схема. Например: проблема о регулярности фундаментальных решений гиперболических уравнений в частных производных (теория лакун Петровского), теория поверхностных потенциалов, теория гипергеометрических функций Гаусса и современных их обобщений, и многие другие. Для их изучения требуются разные *теории Пикара—Лефшеца*, которые изучают ветвление групп гомологий слоёв при подходе к разного рода особенностям. Они опираются на следующее важное общее замечание.

Принцип локализации. Предположим, что ситуация такова, как было сказано, т. е. имеется комплексно аналитическое многообразие $\bar{E}$ (возможно с особенностями) и его аналитическое отображение $\pi\colon \bar{E} \to \bar{T}$ на другое аналитическое многообразие. *Особая точка* этого отображения—это точка в $\bar{E}$, в которой оно не сюръективно, т. е. ранг первого дифференциала не равен размерности $\bar{T}$. (Для отображений негладкого аналитического многообразия $\bar{E}$ это определение надо немного изменить: $\bar{E}$ всегда можно правильным образом *стратифицировать*, т. е. разбить на гладкие многообразия разных размерностей — страты — и назвать особыми точки, в которых вырождается дифференциал ограничения нашего отображения на соответствующие страты.) Дискриминантное множество

$\Sigma \subset \bar{T}$ — это образ множества всех особых точек. Всюду вне множества особых точек наше отображение устроено как локально тривиальное расслоение. Глобально это свойство выражается следующей *леммой Тома об изотопии*: если наше отображение $\pi\colon \bar{E} \to \bar{T}$ — собственное (т. е. прообраз любого компакта компактен), то ограничение его на прообраз множества $T = \bar{T} \setminus \Sigma$ является локально тривиальным расслоением.

Как правило, вырождение каждого фиксированного слоя $\pi^{-1}(b)$, $b \in \Sigma$, происходит в одной точке или в дискретном множестве особых точек. Будем считать для простоты, что вырождение происходит в одной точке, т. е. слой содержит ровно одну особую точку. Если вблизи её проекции комплексная коразмерность множества Σ не меньше двойки, то вблизи неё ветвления не происходит. Поэтому будем считать, что Σ — комплексная гиперповерхность, более того, будем предполагать что вблизи проекции b нашей особой точки эта гиперповерхность неособа. Пусть $D \subset \bar{T}$ — маленький вложенный комплексный диск с центром в этой точке b, трансверсальный дискриминантному множеству Σ. Если диск D достаточно мал, то все его точки, кроме центральной, недискриминантные, и слои над ними топологически устроены одинаково. Мы изучаем, что происходит в гомологиях слоёв при обходе вдоль границы этого диска: как задаваемый ею оператор монодромии действует в этих гомологиях.



Р и с. 1. Расслоение Милнора

Чтобы разобраться в этом, достаточно рассмотреть малую шаровую окрестность B критической точки в $\bar{E}$, т. е. достаточно изучить, что происходит на пересечениях слоёв с B. При этом предполагается, что диск D в базе ещё гораздо меньше чем B, так что в частности все слои $\pi^{-1}(\lambda)$, $\lambda \in D$, трансверсально пересекают границу ∂B шара B, см. рис. 1. Вне шара B слои $F_\lambda = \pi^{-1}(\lambda)$ устроены одинаково: дополнения $F_\lambda \setminus B$ этих слоёв до шара B образуют тривиальное расслоение над диском D. В частности, если рассматриваемый («монодромируемый») цикл в слое над точкой границы ∂D гомологичен в этом слое циклу, лежащему в дополнении к B, то с ним (с его классом гомологий) при обходе вдоль ∂D ничего произойти не может. Всё, что с ним происходит при таком обходе, зависит от того, как он пересекается с шаром B. Поэтому нужно рассматривать класс цикла по модулю дополнения к шару, т. е. его класс в группе относительных

гомологий слоя $H_*(F_\lambda, F_\lambda \setminus B)$. Ветвление зависит только от этого класса, который мы называем *локализацией* исходного цикла.

Цикл $\gamma \subset F_\lambda$ при обходе λ вдоль ∂D переходит более или менее в самого себя; исключение составляет лишь его маленькая часть — пересечение с шаром B, которое переходит в (вообще говоря) другой контур, также лежащий в B и совпадающий с γ вблизи ∂B. Поэтому разность между γ и классом, полученным из γ при монодромии, — некоторый абсолютный цикл, заключённый внутри B. Итак, оператор монодромии добавляет к циклу γ цикл, представляющий элемент группы $H_*(F_\lambda \cap B)$, причём класс его в этой группе зависит только от класса исходного цикла γ в группе $H_*(F_\lambda, F_\lambda \setminus B) \cong H_*(F_\lambda \cap B, F_\lambda \cap \partial B)$. Этот добавок называется *локальной вариацией* цикла γ (или определяемого им класса в последней группе).

Суммируя, получаем, что оператор монодромии, соответствующий пути ∂D, действует на гомологии слоя F_λ как сумма тождественного отображения и оператора, определяемого как композиция трёх операторов $i \circ \mathrm{Var} \circ j$:

$$H_*(F) \xrightarrow{j} H_*(F, F \setminus B) \cong H_*(F \cap B, F \cap \partial B) \xrightarrow{\mathrm{Var}} H_*(F \cap B) \xrightarrow{i} H_*(F). \quad (1)$$

Здесь j — приведение цикла по модулю дополнения к B, Var — оператор локальной вариации (мы переносим относительный цикл над ∂D так, что он оказывается в некотором естественном смысле тривиализованным на границе, т. е. протаскиваем его в соответствии с тривиализацией расслоения над границей диска; разность того, что было, и того, что получилось, называется вариацией), i — оператор, заданный тождественным вложением $F_\lambda \cap B \to F_\lambda$.

Все теории Пикара—Лефшеца занимаются изучением операторов локальной вариации, связанных с самыми разнообразными случаями вырождения. Из этих операторов локальной вариации складывается набор операторов, которые соответствуют разным образующим фундаментальной группы дополнения дискриминантного множества. Действительно, если $\bar{T}$ односвязно (как это обыкновенно случается), то фундаментальная группа множества T некритических значений порождается *малыми петлями*, идущими от отмеченной точки к той или иной неособой точке множества Σ, обходящими это множество по сколь угодно малой петле наподобие рассмотренной выше границы диска ∂D, и возвращающимися обратно по старому пути. Изучая соответствующие им операторы монодромии (сводящиеся к локальной вариации над центральным отрезком пути) можно выяснить, как фундаментальная группа базы действует на гомологиях слоя, а следовательно, выяснить аналитическое поведение функции,

заданной интегральным представлением (т. е. интегрированием замкнутой вдоль слоёв формы ω по циклам $\gamma_\lambda \subset F_\lambda$).

В самом деле, *аналитическое продолжение такой функции вдоль некоторого замкнутого пути в T равно интегралу этой же формы ω по циклу, который получается из γ при обходе вдоль этого пути.*

Принцип локализации работает не всегда. Например, он не всегда работает в теории общих гипергеометрических функций И. М. Гельфанда, потому что там встречаются нелокальные вырождения.

Пока мы будем рассматривать только операторы локальной вариации. Модельный пример таких операторов доставляют голоморфные функции с изолированными особенностями. Пусть $f\colon (\mathbb{C}^n, 0) \to (\mathbb{C}, 0)$ — аналитическая функция от n переменных с изолированной особенностью в нуле, т. е. $df(0) = 0$ и $df \neq 0$ в некоторой проколотой окрестности нуля $B \setminus 0$. Рассмотрим также в $\mathbb{C}^1$ диск D с центром в 0, очень маленький даже по сравнению с B. Прообраз $f^{-1}(\lambda)$ любой точки λ множества $D \setminus 0$ — неособое многообразие. Для изучения локальной монодромии достаточно понять, что происходит с их пересечениями с B, т. е. нас интересуют множества $F_\lambda \cap B = f^{-1}(\lambda) \cap B$, $\lambda \in D \setminus 0$. Топология у всех многообразий F_λ одинаковая, но они вырождаются, когда значение λ стремится к нулю. Если же мы будем обходить вокруг нуля, то на гомологии F_λ действует оператор монодромии, который сводится к оператору локальной вариации, связанному с особой точкой функции f. Многообразие $F_\lambda \cap B$ имеет вещественную размерность $2n-2$. Известно (теорема Милнора), что его гомологии сосредоточены в средней размерности $n-1$. Более того, оно гомотопически эквивалентно букету нескольких сфер размерности $n-1$; количество этих сфер обозначают $\mu(f)$ и называют *числом Милнора* особенности. В частности, $H_{n-1}(F_\lambda \cap B) \cong \mathbb{Z}^{\mu(f)}$ за исключением случая $n=1$, когда эта группа изоморфна $\mathbb{Z}^{\mu(f)+1}$, поэтому для единообразия полезно рассматривать группу $\tilde{H}_*(F_\lambda \cap B)$ гомологий, приведённых по модулю точки. Эта группа всегда изоморфна $\mathbb{Z}^{\mu(f)}$ и сосредоточена в размерности $n-1$. По двойственности Пуанкаре, группа относительных гомологий $\tilde{H}_*(F_\lambda \cap B, F_\lambda \cap \partial B)$, приведённая по модулю фундаментального цикла, также сосредоточена в размерности $n-1$ и изоморфна $\mathbb{Z}^{\mu(f)}$. Оператор локальной вариации действует так:

$$\tilde{H}_{n-1}(F_\lambda \cap B,\, \partial B) \to \tilde{H}_{n-1}(F_\lambda \cap B). \tag{2}$$

Известно, что этот оператор в нашем случае является изоморфизмом.

Например, если функция f морсовская, т. е. её квадратичная часть невырожденная, то многообразие $F_\lambda \cap B$, $\lambda \in D \setminus 0$, гомотопически эквивалентно одной сфере размерности $n-1$ (более того, оно диффеоморф-

но пространству T_1S^{n-1} касательных векторов длины $\leqslant 1$ в касательном расслоении к $(n-1)$-мерной сфере). Локальная вариация в этом случае задаётся *стандартной формулой Пикара—Лефшеца*:

$$\mathrm{Var}(\nabla) = (-1)^{n(n+1)/2}\langle \nabla,\, \Delta\rangle \Delta, \tag{3}$$

где ∇ — образующая относительной группы гомологий, Δ — образующая абсолютной группы гомологий. Δ называется *исчезающим циклом*, поскольку группа гомологий обращается в 0 когда $\lambda \to 0$. Порождающий её цикл (сфера) наглядным образом стягивается в точку. Действительно, по *лемме Морса* в некоторых локальных голоморфных координатах функция f записывается в виде $f = x_1^2 + \ldots + x_n^2$. Рассмотрим вещественную плоскость $\mathbb{R}^n$, заданную этими координатами (т. е. условием $\mathrm{Im}\, x_j = 0$, $j = 1, \ldots, n$). Тогда при $\lambda > 0$ исчезающий цикл реализуется сферой радиуса $\sqrt{\lambda}$ в $\mathbb{R}^n$ и стягивается в точку, когда λ стремится к 0.

Эта же формула (3) описывает монодромию в (глобально определённом) слое F_λ, соответствующую обходу вокруг критического значения функции f, принимаемого в морсовской критической точке.

Предположим, что n нечётно. Тогда билинейная форма, заданная в гомологиях слоя F_λ индексом пересечения, симметрична. В этом случае оператор монодромии $H_{n-1}(F_\lambda) \to H_{n-1}(F_\lambda)$, заданный формулой (3), можно рассматривать как отражение относительно гиперплоскости, ортогональной (в смысле этой билинейной формы) исчезающему циклу Δ.

Хотя в формуле (3) ∇ рассматривают как класс относительной группы гомологий, но думать про это не надо. Всегда исходная задача — это вычисление действия оператора монодромии на абсолютные классы гомологий (нелокализованного) слоя F_λ. Но для того чтобы разобраться, что происходит с таким классом при обходе (насколько он изменяется — варьируется), его нужно локализовать, т. е. достаточно рассмотреть его класс в группе $H_{n-1}(F_\lambda,\, F_\lambda \setminus B)$. Если его индекс пересечения с Δ равен 0, то $\mathrm{Var}(\nabla) = 0$, т. е. все циклы, ортогональные Δ, остаются на месте, как это и должно быть при отражении (впрочем, они остаются на месте и при чётных n). А сам цикл Δ при нечётном n переходит в $-\Delta$, как это легко проверить с помощью формулы (3), если учесть, что эйлерова характеристика чётномерной сферы равна 2.

Имеется много параллельных (и эквивалентных) задач, в которых вместо гомологий слоя часто приходится рассматривать другие, тесно связанные с ними группы гомологий. А именно, иногда интересен следующий оператор вариации:

$$H_n(B,\, F_\lambda \cup \partial B) \to H_n(B,\, F_\lambda), \tag{4}$$

который возникает при аналогичной локализации оператора монодромии, действующего на относительной группе $H_n(\mathbb{C}^n, F_\lambda)$. Есть и ещё один вариант:

$$H_n(B \setminus F_\lambda, \partial B) \to H_n(B \setminus F_\lambda). \tag{5}$$

Он является локализацией оператора монодромии, действующего на группе $H_n(\mathbb{C}^n \setminus F_\lambda)$; впрочем в обоих случаях вместо $\mathbb{C}^n$ может стоять любое другое n-мерное многообразие M^n, на котором может быть определена функция $f\colon M^n \to \mathbb{C}^1$, монодромия слоёв (или их дополнений) которой может быть нам интересна. Эти глобальные задачи, приводящие к операторам (4), (5), чуть более сложны, чем задача о ветвлении для групп гомологий слоя. Но их локализованные формы все изоморфны, потому что есть изоморфизмы, сопрягающие все операторы (2), (4), (5): эти изоморфизмы определяют вертикальные стрелки в следующей диаграмме, в которой эти 3 оператора являются горизонтальными стрелками:

$$\begin{array}{ccc} H_n(B,\, F_\lambda \cup \partial B) & \longrightarrow & H_n(B,\, F_\lambda) \\ \downarrow{\scriptstyle \partial_*} & & \downarrow{\scriptstyle \partial_*} \\ H_{n-1}(F_\lambda \cap B,\, \partial B) & \longrightarrow & H_{n-1}(F_\lambda \cap B) \\ \downarrow{\scriptstyle \phi} & & \downarrow{\scriptstyle \phi} \\ H_n(B \setminus F_\lambda,\, \partial B) & \longrightarrow & H_n(B \setminus F_\lambda) \end{array} \tag{6}$$

Здесь ∂_* — граничные операторы из подходящих точных последовательностей троек и пар, а ϕ — трубочные (кограничные) операторы Лере. Трубочный оператор Лере задаётся следующим образом. Пусть есть какой-то $(n-1)$-мерный цикл, лежащий в комплексной гиперповерхности. Тогда надо взять расслоение трубчатой окрестности этой комплексной гиперповерхности (слой его — маленький диск в $\mathbb{C}^1$, снабжённый стандартной ориентацией, определённой комплексной ориентацией нормального расслоения к гиперповерхности). Для каждой точки цикла нужно взять соответственно ориентированную границу слоя над этой точкой. Все эти окружности вместе заметут корректно определённый n-мерный цикл. Его класс в гомологиях дополнения к гиперповерхности и будет результатом трубчатого оператора Лере, применённого к нашему циклу (к его классу гомологий).

Легко видеть, что диаграмма (6) коммутативна, а следовательно, исследование всех трёх операторов, задающих её горизонтальные стрелки, — это одна и та же задача.

2.
Задача Ньютона об интегрируемости овалов

Эту задачу нашёл В. И. Арнольд в «Математических началах натуральной философии» Ньютона. Ньютон решил её в двумерном случае (по тогдашним меркам строгости). Задача такова. Пусть в $\mathbb{R}^N$ задано тело, для начала выпуклое. Рассмотрим $\mathbb{P}$ — пространство аффинных гиперплоскостей в $\mathbb{R}^N$. На $\mathbb{P}$ возникает функция объёма (как минимум двузначная) $V(L)$, которая на гиперплоскости L принимает значение, равное объёму, отсечённому от тела гиперплоскостью L. Задача заключается в том, чтобы выяснить, какова эта функция $V(L)$. Например, может ли она быть алгебраической? Может ли её аналитическое продолжение не быть бесконечнолистным?

Если исходное тело — шар в нечётномерном пространстве, то эта функция алгебраична; это доказал Архимед (при $N=3$). А для шара в чётномерном пространстве — не алгебраична.

Теорема 1. *В $\mathbb{R}^{2k}$ для выпуклых тел с гладкой границей функция объёма всегда не алгебраична, а в $\mathbb{R}^2$ не алгебраична и для невыпуклых.*

Для доказательства этого утверждения нужно выйти в комплексную область и заметить, что мы оказались в ситуации, описанной во введении. Легко видеть, что поверхность тела должна быть полуалгебраической (т. е. задаваться условием $f=0$, где f — ненулевой полином, который может обращаться в 0 и где-то ещё): иначе, в силу соображений проективной двойственности, функция объёма $V(L)$ заведомо не алгебраична. Обозначим комплексификацию этой поверхности через $A_{\mathbb{C}}$: это — множество комплексных нулей того же полинома, задающего границу тела. В качестве базы T выберем $\mathbb{P}_{\mathbb{C}} \setminus \Sigma$, где $\mathbb{P}_{\mathbb{C}}$ — множество комплексных гиперплоскостей в $\mathbb{C}^N$, Σ — множество гиперплоскостей не в общем положении с комплексной гиперповерхностью $A_{\mathbb{C}}$. Форма ω — это форма объёма $dx_1 \wedge \ldots \wedge dx_N$; она очевидно продолжается в комплексную область и является голоморфной формой в $\mathbb{C}^N$. Слой расслоения E над точкой $\lambda \in \mathbb{P}_{\mathbb{C}}$, символизирующей комплексную плоскость $L_{\mathbb{C}} \subset \mathbb{C}^N$, — это не одно пространство, а пара топологических пространств $(\mathbb{C}^N, L_{\mathbb{C}})$. Соответственно, группа гомологий слоя, которой принадлежат контуры интегрирования, — это относительная группа гомологий

$$H_N(\mathbb{C}^N, L_{\mathbb{C}} \cup A). \tag{1}$$

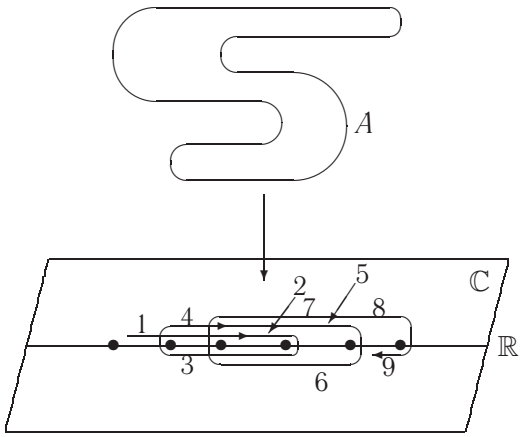
Интеграл от формы ω по относительному циклу в $\mathbb{C}^N$ $(\mathrm{mod}\ L_{\mathbb{C}} \cup A_{\mathbb{C}})$ зависит только от гомологического класса этого цикла в группе (1),

потому что форма голоморфная. Воспользуемся теперь теорией Пикара—Лефшеца. Рассмотрим однопараметрическое семейство гиперплоскостей, а именно, выберем гиперплоскость общего положения в $\mathbb{R}^N$, пересекающую тело, и будем двигать её параллельно самой себе; при этом условие общности положения состоит в том, что вещественная линейная функция $\phi(x)$, множества уровня которой $\{\phi(x)=\tau\}$ задают гиперплоскости из этого семейства, имеет только морсовские критические точки в ограничении на поверхность тела, причём критические значения в них различны. (Если тело строго выпуклое, то имеется лишь две такие критические точки — минимум и максимум.) Когда мы сдвинем плоскость параллельно себе от одного критического положения до другого, объём отсечённой части тела будет изменяться от 0 до полного объёма тела. Если дойти по полного объёма, а затем начать двигаться обратно, то объём будет уменьшаться. Но если перед тем, как начать двигаться обратно, заставить параметр τ нашего семейства выйти в комплексную область и обойти вокруг критического значения, то оказывается, что в силу формулы Пикара—Лефшеца (при чётном N) объём будет не уменьшаться, а снова увеличиваться — ровно на столько же, на сколько он уменьшался бы, если бы мы не обходили вокруг критического значения.

Действительно, мы должны «комплексифицировать» задачу, т. е. рассматривать не только вещественные гиперплоскости в $\mathbb{R}^N$, но и их комплексификации, т. е. гиперплоскости в $\mathbb{C}^N$, заданные теми же уравнениями $\phi(x)=\tau$, и, более того, такие же гиперплоскости, соответствующие невещественным значениям τ. Ветвление аналитического продолжения (первоначально вещественной) функции объёма по параметру τ определяется оператором монодромии групп (1) при обходе τ вокруг критического значения. Но в силу формулы Пикара—Лефшеца и верхней части диаграммы (6), при таком обходе вокруг критического значения, соответствующего морсовской критической точке, при чётном N исчезающий цикл переходит не в себя, а в себя со знаком минус.

Затем можно дойти до другого критического значения, обойти вокруг него в комплексной области и пойти обратно. Функция снова будет возрастать. Так можно набрать бесконечно много значений в любой точке и получить бесконечнозначную функцию. Такая функция не может быть алгебраической. Например, для окружности получим арккосинус.

В двумерном случае нетрудно доказать то же самое и для невыпуклых тел, см. рис. 2. Начнём со значения τ, чуть большего чем абсолютный минимум функции ϕ_A. Тогда прямая $\{\phi(x)=\tau\}$ пересекает кривую A в двух близких точках $\alpha(\tau)$, $\beta(\tau)$. Отрезок кривой A, соединяющий эти две



Р и с. 2. Монодромия функции объёма для плоской кривой

точки $\alpha(\tau)$ и $\beta(\tau)$, в этот момент также очень мал. Мы будем доказывать утверждение с помощью индукции по величине этого отрезка.

Будем увеличивать значение τ до тех пор, пока одна из этих двух точек, скажем $\beta(\tau)$, не дойдёт до точки локального максимума функции ϕ_A. Если этот максимум — глобальный, т. е. другая точка $\alpha(\tau)$ также приближается к нему с другой стороны, то оставшаяся часть доказательства — та же, что раньше. В противном случае обернём параметр τ в комплексной области вокруг критического значения. Этот оборот переведёт точку $\beta(\tau)$ в другую точку (назовём её $\beta_1(\tau)$) кривой A, находящуюся на том же уровне τ. Важно, что отрезок кривой, соединяющий точки $\alpha(\tau)$ и $\beta_1(\tau)$, больше, чем близкий к нему отрезок $[\alpha(\tau), \beta(\tau)]$, полученный по непрерывности из начального маленького отрезка. Контур интегрирования при этом *убывает* на класс маленькой области в $\mathbb{R}^2$, ограниченной малой дугой кривой A и маленьким отрезком прямой $\{\phi = \tau\}$, кончающимися в точках $\beta(\tau)$ и $\beta_1(\tau)$. Далее мы уменьшаем параметр τ, пока точка $\beta_1(\tau)$ не приблизится к локальному минимуму функции ϕ_A, после чего она обносится вокруг соответствующего критического значения, и т. д. После того, как наши две точки обойдут вокруг всех локальных (но не глобальных) экстремумов функции ϕ_A, они наконец приблизятся с двух сторон к точке глобального максимума. Отрезок $[\alpha, \beta]$, за которым мы следим, при этом равен почти всей кривой A. Обойдём вокруг критического значения в этой точке, пройдём обратно по старому пути, и дополнительно обернёмся вокруг значения в глобальном минимуме. Тогда мы вернёмся в исходную точку, но аналитическое продолжение функции площади при этом возрастёт на удвоенную площадь, ограниченную кривой A.

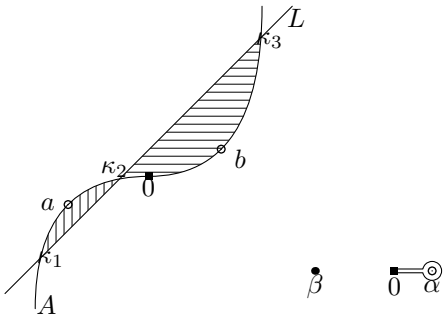
Непонятно, как обобщить эту конструкцию на невыпуклые поверхности в $2k$-мерном пространстве при $k>1$. Конечно, для гиперповерхностей общего положения, у комплексификаций которых нет особенностей в $\mathbb{CP}^N$, всё получается автоматически. Действительно, с комплексной точки зрения мы получаем ту же самую задачу, потому что ветвление функции объёма (точнее, её аналитического продолжения в $\mathbb{P}_{\mathbb{C}}$) задаётся представлением монодромии, отображающим фундаментальную группу множества неособых гиперплоскостей $L_{\mathbb{C}}$ (трансверсальных к $A_{\mathbb{C}}$) в группу автоморфизмов группы гомологий (1). Но для всех типичных алгебраических поверхностей $A_{\mathbb{C}}$ данной степени эти представления одинаковы, независимо от того, каково их пересечение A с вещественной плоскостью: это следует из того, что пространство всех комплексных типичных алгебраических гиперповерхностей данной степени линейно связно. В классе вещественных гиперповерхностей, заданных уравнениями той же самой степени, всегда есть гиперповерхности, имеющие выпуклые компоненты. Более того, они занимают некоторую открытую область в пространстве таких гиперповерхностей, а следовательно содержат и некоторые типичные (в комплексном смысле) гиперповерхности. Можно посмотреть на группу монодромии для такой типичной гиперповерхности с выпуклой вещественной компонентой. Для неё мы уже знаем, что этой группы достаточно, чтобы набрать сколько угодно значений. (В вещественном случае мы при этом использовали тот факт, что интеграл формы объёма по добавляемому контуру не равен нулю: действительно, это просто объём тела в $\mathbb{R}^N$. В общем случае это заранее не гарантировано, и это условие надо добавить к определению «гиперповерхности общего положения».) Следовательно, то же верно и для любой типичной гиперповерхности, имеющей непустое пересечение с $\mathbb{R}^N$. Таким образом, задача остаётся только для гиперповерхностей не общего положения. В этом случае также известны некоторые препятствия к алгебраичности, но я не умею доказывать, что их достаточно, чтобы обобщить теорему Ньютона на любые невыпуклые тела в пространствах чётной размерности $N>2$.

В нечётномерном случае, как я уже говорил, ситуация гораздо сложнее, потому что там есть шары и эллипсоиды, для которых алгебраичность есть. По-видимому, других таких примеров нет. Препятствия к построению таких примеров следующие. Рассмотрим комплексную гиперповерхность $A_{\mathbb{C}}$. Возьмём какую-нибудь параболическую точку гиперповерхности $A_{\mathbb{C}}$, т. е. точку, в которой гиперповерхность гладкая, но форма кривизны вырожденная. Если сказать совсем просто, то это означает следующее. Если в этой точке провести аффинную касательную гиперплоскость, то в локальных комплексных аффинных координатах эту гиперплоскость можно

задать уравнением $z_N = 0$, а наша гиперповерхность в этих же координатах будет задаваться каким-то уравнением вида $z_N = f(z_1, \ldots, z_{N-1})$, где f — гладкая функция. Тогда в рассматриваемой точке функция f будет иметь особенность (т. е. $df = 0$). Параболичность гиперповерхности в этой точке эквивалентна тому, что эта особенность не морсовская. Допустим, что такая точка есть; с другой стороны потребуем чтобы эта особенность была не *бесконечно вырожденная*, т. е. множество касания гиперповерхности и гиперплоскости $\{z_N = 0\}$ вблизи этой точки состояло только из неё самой; это эквивалентно тому, что особенность функции f — изолированная. Тогда сразу можно утверждать что алгебраичности нет. Действительно, мы начнём с какого-то маленького кусочка, отсекаемого гиперплоскостью вблизи эллиптической точки вещественной гиперповерхности (такие точки всегда есть). Выйдем в комплексную область. Для нашей комплексной гиперповерхности множество точек, в которых она эллиптична (форма кривизны невырожденная), линейно связно. Поэтому гиперплоскость можно двигать так, чтобы она оставалась трансверсальной к гиперповерхности $A_{\mathbb{C}}$, но «скользила» вдоль неё (т. е. в любой момент была близка в $\mathbb{P}_{\mathbb{C}}$ к некоторой гиперплоскости, касательной к $A_{\mathbb{C}}$), и подвести к той самой параболической точке касания. При этом цикл интегрирования, будучи маленьким с самого начала, таким и останется, и перейдёт в маленький относительный цикл, заключённый в малой окрестности D этой параболической точки. После этого нужно воспользоваться *локальной теорией Пикара—Лефшеца для функций с изолированной особенностью*, у которых число Милнора больше 1. Эта теория является хорошо развитой областью теории особенностей гладких функций.

Рассмотрим например простейшую параболическую (неморсовскую) изолированную особенность (типа A_2). В случае $N = 2$ (который, честно говоря, мы непосредственно сейчас не рассматриваем, интересуясь случаем нечётных N) поверхность $A_{\mathbb{C}}$ вблизи такой особенности в аффинных координатах x, y задаётся условием $y = f(x)$, где $f(x) = x^3 + O(x^4)$; соответствующая вещественная особенность изображена на рис. 3 слева.

Подходящим образом подводя гиперплоскость (прямую) L к положению, указанному на этом рисунке, мы можем добиться того, что цикл интегрирования, участвовавший в задаче Ньютона, превратится в относительный цикл, заданный вертикально заштрихованной областью. Слегка сдвигая эту прямую параллельно себе в $\mathbb{C}^2$, мы будем получать топологически одинаковые множества $A \cup L$, за исключением ровно двух положений этой прямой — когда она будет касательна к $A_{\mathbb{C}}$ в точке a (это положение соответствует исчезновению цикла интегрирования Δ) или в точке b (она соответствует исчезновению другого, горизонтально заштрихованного



Р и с. 3. Монодромия гиперплоских сечений вблизи параболических точек

относительного цикла Δ'). Множество возможных параллельных сдвигов образует комплексную прямую с координатой τ, на которой критические положения соответствуют точкам α и β соответственно, см. рис. 3, справа.

Что будет происходить с циклом Δ если мы будем сдвигать гиперплоскость L параллельно себе, перемещая соответствующие точки в нормальной к ней прямой $\mathbb{C}^1$ вдоль замкнутых путей в множестве недискриминантных точек $\mathbb{C}^1 \setminus \{\alpha, \beta\}$ этой прямой?

Несложно показать (используя граничные операторы, изображённые верхними вертикальными стрелками диаграммы (6)) что для этого достаточно понять, как будет преобразовываться $(N-2)$-мерный цикл в множестве $L \cap A$, лежащий на «углу» цикла Δ, т. е. получающийся из него при помощи сквозного гомоморфизма

$$\tilde{H}_N(B,\, B \cap (L \cup A)) \to \tilde{H}_{N-1}(B \cap (L \cup A)) \to$$
$$\to \tilde{H}_{N-1}(B \cap (L \cup A),\, B \cap L) \to \tilde{H}_{N-1}(B \cap A,\, B \cap A \cap X) \to \tilde{H}_{N-2}(B \cap A \cap X);$$

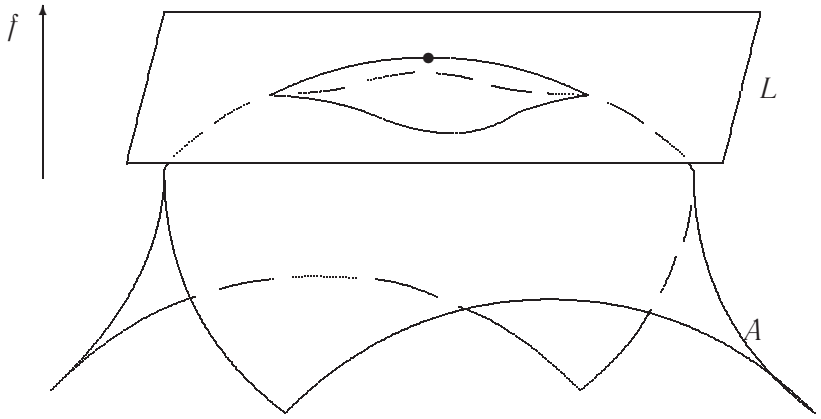
здесь первая и последняя стрелки — граничные операторы, третья — вырезание, а вторая взята из точной последовательности пары; легко видеть, что все они являются изоморфизмами. На нашей картинке этот цикл задаётся парой точек κ_1, κ_2.

Обход вдоль петли, обходящей значение α, при чётном N превращает цикл Δ в $-\Delta$ (именно этим мы пользовались при доказательстве обобщения теоремы Ньютона в чётномерном случае; при $N=2$ это можно без труда увидеть из рисунка). При нечётном же N он оставляет Δ на месте. Обход вокруг точки β при любом N переводит Δ в $\Delta \pm \Delta'$, где знак $\pm$ зависит от N и от выбора ориентации цикла Δ; это следует из формулы Пикара—Лефшеца, поскольку индекс пересечения $\langle \Delta, \Delta' \rangle$ в нашем случае равен ± 1. При $N=2$ это соответствует тому, что этот обход оставляет точку κ_1 множества $A \cap L$ на месте, но меняет местами

точки κ_2 и κ_3. Пройдём вдоль этой же петли ещё раз: что произойдёт с полученным только что циклом $\Delta \pm \Delta'$? Если N чётно, то этот цикл перейдёт обратно в Δ. Действительно, при этом монодромия определяется теорией Пикара—Лефшеца для функций от нечётного числа переменных $N-1$ (как например функция f), а мы помним, что в этом случае любой оператор Пикара—Лефшеца является отражением, в частности инволюцией. Напротив, если N нечётно, то, обходя m раз в одном направлении вокруг β, мы переведём Δ в $\Delta \pm m\Delta'$. Поскольку интеграл формы объёма по циклу Δ' (рассматриваемый как функция на $P_{\mathbb{C}}$) не равен тождественно нулю (это легко усмотреть, когда сам цикл Δ' близок к исчезновению) мы получаем логарифмическое ветвление интегральной функции. Аналогичные рассуждения проходят вблизи любой другой параболической не бесконечно вырожденной точки поверхности $A_{\mathbb{C}}$ и доказывают, что наличие таких точек препятствует алгебраичности функции объёма.

Параболических точек у гиперповерхности $A_{\mathbb{C}}$ может и не быть. Например, их нет для невырожденной поверхности второго порядка. Для гиперповерхности более высокого порядка в общем положении параболические точки должны быть. Но иногда так случается, что их всё-таки нет. Например, возьмём любую гладкую алгебраическую гиперповерхность в $\mathbb{CP}^n$ и рассмотрим проективно двойственную ей гиперповерхность, т. е. подмножество в двойственном проективном пространстве, состоящее из гиперплоскостей нашего пространства, касательных к нашей гиперповерхности. Если исходная поверхность гладкая, то у проективно двойственной параболических точек нет. Но зато у неё есть много особенностей. (Вообще, методами алгебраической геометрии доказывается, что если гиперповерхность степени $\geqslant 3$ в $\mathbb{CP}^N$ неособа, то у неё обязательно есть параболические не бесконечно вырожденные точки — даже в $\mathbb{C}^N$, т. е. они не могут содержаться все в несобственной плоскости. Существенная часть доказательства этого — теорема Ф. Л. Зака.) С этими особенностями тоже можно разбираться. Скорее всего, у такой гиперповерхности будут *стандартные рёбра возврата*. Например, при проективной двойственности такие рёбра получаются из стандартных параболических точек типа A_2, рассмотренных в примере выше. Существование стандартных рёбер возврата тоже является препятствием к алгебраичности. Доказательство этого примерно такое же как для параболических точек, но более сложное, потому что приходится изучать оператор локальной вариации не в обычной, а в *стратифицированной* теории Пикара—Лефшеца. Эта теория изучает ветвление групп гомологий гиперплоских сечений комплексных алгебраических множеств *с особенностями* при изменении этих секущих гиперплоскостей. Как обычно, это

ветвление определяется операторами локальной вариации, соответствующими маленьким петлям, обходящим дискриминантное подмножество в $\mathbb{P}_{\mathbb{C}}$, состоящее из нетипичных по отношению к $A_{\mathbb{C}}$ положений гиперплоскости. Первый пример таких положений (изучаемый по существу стандартной теорией Пикара—Лефшеца) — это гиперплоскости, касающиеся $A_{\mathbb{C}}$ в её неособых точках. Другая ситуация показана на рис. 4: изображённая на нём поверхность A *стратифицируется*, т. е. разбивается на неособые компоненты (в нашем случае — размерности 2 (неособую часть) и размерности 1 (ребро возврата)). Плоскость L близка к *стратифицированно вырожденному* положению: подняв его немного, мы сделаем её касательной к одномерному страту. Если вместо этого мы обнесём её в комплексной области вокруг вырожденного положения, то на соответствующие группы вида $H_N(\mathbb{C}^N, A_{\mathbb{C}} \cup L)$, $H_{N-1}(A_{\mathbb{C}}, A_{\mathbb{C}} \cap L)$ или $H_{N-2}(A_{\mathbb{C}} \cap X)$ и т. п. также подействует некоторый оператор локальной монодромии. Стратифицированная теория Пикара—Лефшеца изучает именно такие операторы.



Р и с. 4. Стратифицированно морсовская функция

Основной результат этой теории заключается в том, что задача сводится (впрочем, довольно непростым образом) к изучению аналогичной задачи для трансверсального к страту среза пространства $\mathbb{C}^N$. Например, в случае ребра возврата в трансверсальном срезе возникает полукубическая парабола; поскольку страт $(N-2)$-мерен, а его трансверсальный срез двумерен, то все сводится к некоторой задаче в двумерном пространстве, которая, конечно же, не может быть сложной.

3.
Задача о потенциалах

Эта задача тоже восходит к Ньютону. Имеется следующая его знаменитая теорема. Пусть есть однородный сферический слой, который притягивает тела. Тогда для тел внутри него притяжения нет, а тела, расположенные снаружи, он притягивает так, как притягивала бы их материальная точка той же совокупной массы, помещённая в его центре. Айвори (1809 г.) обобщил эту теорему на случай эллипсоидов. Он доказал, что если есть эллипсоид, на поверхности которого некоторым естественным (точное описание см. ниже) способом размазана притягивающая масса, то для точек внутри притяжения нет. А для притяжения точек снаружи им было доказано, что оно одинаково для всех конфокальных эллипсоидов той же совокупной массы.

В. И. Арнольд обобщил первую часть этой теоремы на случай произвольных *гиперболических* гиперповерхностей степени d в $\mathbb{R}^n$. Алгебраическая гиперповерхность A степени d в $\mathbb{R}^n$ (или в $\mathbb{RP}^n$) называется гиперболической, если существует такая точка x, не принадлежащая A, что любая проходящая через неё вещественная прямая пересекает A ровно в d вещественных точках (некоторые из которых могут быть бесконечно удалёнными). Связная компонента множества точек x, для которых это условие выполнено, называют *областью гиперболичности* нашей гиперповерхности. Гладкими гиперболическими гиперповерхностями степени 2 в $\mathbb{R}^n$ являются только эллипсоид и двухкомпонентный гиперболоид.

Областей гиперболичности у гладкой гиперповерхности может быть не больше двух (а в проективном пространстве даже не более одной), и такая область всегда выпукла. Две области гиперболичности в $\mathbb{R}^n$ могут быть только в том случае, когда несобственная гиперплоскость разделяет область гиперболичности в $\mathbb{RP}^n$ на две части.

Напомним, что сила притяжения всегда равна минус градиенту некоторой функции — потенциала притяжения (который мы опишем чуть ниже).

Теорема 2 (Арнольд). *Для неособой гиперболической поверхности степени d в $\mathbb{R}^n$ распространяемый ею стандартный потенциал силы притяжения равен постоянной функции внутри области гиперболичности, т. е. сила притяжения внутри области гиперболичности равна нулю.*

Этот стандартный потенциал $I(x)$ определяется формулой

$$I(x) = \int \sigma\, G(x - z)\, dV/dF, \tag{1}$$

где интегрирование ведётся по z, пробегающим нашу гиперповерхность, $G(x)$ — ядро оператора Лапласа (с точностью до коэффициента, равного объёму единичной сферы, оно равно $\ln|x|$ при $n=2$ и $1/|x|^{n-2}$ при $n>2$), dV — форма объёма $dz_1\wedge\ldots\wedge dz_n$ в $\mathbb{R}^n$, F — уравнение поверхности. Смысл этого интеграла (в частности, выражения dV/dF) таков: надо брать слои между данной гиперповерхностью $\{F=0\}$ и гиперповерхностью $\{F=\epsilon\}$ и в этих слоях массу распределить равномерно; затем силу притяжения каждым таким слоем нужно разделить на ϵ и устремить ϵ к нулю.

Ещё в формулировке теоремы Арнольда нужно взять слои с учётом знаков $\sigma=\pm$. А именно, слой, ближайший к фиксированной компоненте области гиперболичности, нужно взять со знаком плюс, следующий — со знаком минус, затем снова со знаком плюс и т. д.

Доказательство теоремы Арнольда почти такое же, как доказательство теоремы Ньютона. Нужно рассматривать узкие конуса с вершинами в точке x и смотреть, как притягивают эту точку пересечения конусов с разными компонентами слоя. То, что в этом случае притяжения ни в какую сторону не будет, непосредственно следует из теоремы о том, что сумма вычетов рациональной функции равна нулю.

Давайте займёмся другой частью теоремы Ньютона—Айвори. Будем рассматривать поведение силы притяжения в других компонентах дополнения до гиперповерхности (отличных от области гиперболичности). Например, будет ли там она (или её потенциал (1)) алгебраической?

Пусть $n=2$ и $d=2$. В этом случае в потенциал входит логарифм, поэтому будем рассматривать не потенциал, а саму силу притяжения. Для окружности всё очень просто: сила притяжения вне её — однозначная рациональная вектор-функция $\left(\frac{x}{x^2+y^2},\,\frac{y}{x^2+y^2}\right)$. Для разноосного эллипса функция однозначной не будет, но, как мы увидим ниже, она будет алгебраической 4-значной.

Будет ли сила притяжения алгебраической при $d>2$? Ответ таков:

Теорема 3. *Если $n=2$ или $d=2$, то сила притяжения алгебраична во всех областях дополнения до гиперповерхности, если же $n>2$ и $d>2$, и притягивающая гиперповерхность — общего положения (в некотором точном смысле), то алгебраичности не будет ни в какой области кроме области гиперболичности.*

Чтобы разобраться с этой задачей, вновь комплексифицируем её. В частности, рассмотрим комплексификацию $A_{\mathbb{C}}$ гиперповерхности A (заданную тем же уравнением, но не в $\mathbb{R}^n$, а в $\mathbb{C}^n$). Затем воспользуемся теорией Пикара—Лефшеца. А именно, возьмём точку x и будем её двигать,

выходя и в комплексную область. Будем смотреть, что происходит, когда мы переносим точку x по замкнутому пути и возвращаемся обратно. Как будет изменяться цикл интегрирования и как в соответствии с этим будет изменяться интеграл?

Что может происходить с циклом? В любой момент цикл (контур интегрирования) — это вещественная часть гиперповерхности $A_{\mathbb{C}} \subset \mathbb{C}^n$. Этот цикл лежит в группе гомологий $H_{n-1}(A_{\mathbb{C}} \setminus S(x))$, где $S(x)$ — множество точек в $\mathbb{C}^n$, лежащих на расстоянии 0 от точки x. Расстояние в $\mathbb{C}^n$ задаётся обычной формулой $\rho(x, z) = \sqrt{(x_1 - z_1)^2 + \ldots + (x_n - z_n)^2}$, таким образом $S(x)$ — это множество точек z, на которых имеет особенность ядро оператора Лапласа с центром в x, участвующее в формуле (1). Это расстояние при каждом фиксированном $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$ является двузначной алгебраической функцией. Но если n чётно и больше двух, то ядро G оператора Лапласа обратно пропорционально расстоянию в чётной степени, поэтому получается однозначная рациональная функция. Если $n = 2$, то однозначные рациональные функции получаются для интегральных выражений обеих компонент вектора силы притяжения. В любом случае, с определением особого множества $S(x)$ не возникает проблем ни при каком n, только при некоторых n оно будет только полюсом подынтегральной формы, а при других — ещё и множеством ветвления.

На группе гомологий $H_{n-1}(A_{\mathbb{C}} \setminus S(x))$ группа монодромии может действовать нетривиально, потому что, если мы выведем точку x в комплексную область и совершим обход в множестве таких $x \in \mathbb{C}^n$, что множества $A_{\mathbb{C}}$ и $S(x)$ пересекаются типичным образом, то конус $S(x)$ может пересечь исходный цикл интегрирования, состоящий из вещественных точек множества $A_{\mathbb{C}}$. Чтобы пересечения не было, цикл придётся деформировать. За счёт этого появляется монодромия и ветвление интегралов.

Если точка x расположена в области гиперболичности, то группа монодромии тривиально действует на класс гомологий соответствующего цикла интегрирования: он является инвариантным вектором представления монодромии. Это эквивалентно утверждению теоремы Арнольда о том, что сила притяжения равна нулю.

Действительно, инвариантность цикла означает что эта сила является голоморфной однозначной вектор-функцией; но в силу определения она имеет не более чем полиномиальный рост на бесконечности в $\mathbb{C}^n$, значит это полином. К тому же она стремится к нулю на бесконечности в $\mathbb{R}^n$, стало быть равна 0. Обратно, если бы цикл интегрирования был неинвариантен, это означало бы, что на него нетривиально действует некоторый *локальный* оператор Пикара—Лефшеца (в подходящей теории Пикара—Лефшеца, связанной с нашей задачей), прибавляющий к нему,

как обычно, некоторый *исчезающий цикл*, интеграл по которому не может быть тождественно нулевым, а следовательно аналитическое продолжение силы притяжения (тождественно нулевой в силу предположения) было бы ненулевым.

Конечно, утверждение об инвариантности этого цикла интегрирования для x из области гиперболичности без большого труда проверяется и непосредственно геометрически, без ссылки на теорему Арнольда.

Вообще, для x из любой фиксированной компоненты множества $\mathbb{R}^n \setminus A$ соответствующие классы интегрирования $\Delta(x)$ (вещественные циклы, ориентированные как указано в определении интеграла (1)) имеют одинаковые алгебраические свойства по отношению к группе монодромии: в частности орбиты её действия на этих элементах естественно отождествляются при помощи любого пути в $\mathbb{R}^n \setminus A$, соединяющего эти точки.

Пусть например точка x расположена в области, соседней с областью гиперболичности. Тогда помимо старого контура интегрирования $\Delta(x)$ (т. е. вещественного цикла), рассматриваемого как элемент новой группы гомологий ($H_{n-1}(A_{\mathbb{C}} \setminus S(x))$, где x принадлежит новой компоненте дополнения к $A_{\mathbb{C}}$), интересен ещё один очень похожий на него элемент $\Delta'(x)$ этой же группы, который получается следующим образом. Мы считаем что x находится очень близко к стенке, отделяющей нашу компоненту от области гиперболичности; пусть x' — точка в области гиперболичности, находящаяся очень близко к x с другой стороны от этой стенки. Проведём через x и x' комплексную прямую в $\mathbb{C}^n$, и соединим x и x' в ней маленькой дугой, лежащей в недискриминантном множестве, т. е. в множестве таких её точек $\tilde{x}$ что соответствующие пространства $A_{\mathbb{C}} \setminus S(\tilde{x})$ все топологически одинаковы (образуют локально тривиальное расслоение). Тогда их группы гомологий, содержащие интересующие нас циклы интегрирования, естественно отождествляются. (Это отождествление называется гомологической связностью Гаусса—Манина над нашей дугой.) Искомый элемент $\Delta'(x)$ получается при этом отождествлении из вещественного цикла $\Delta(x')$, где x' — точка из области гиперболичности. (Нетрудно проверить, что результат не будет зависеть от того, какую из двух существенно различных дуг, обходящих в нашей прямой дискриминантную точку, соответствующую пересечению с $A_{\mathbb{C}}$, мы будем использовать.)

Поскольку точки x и x' очень близки, получающиеся классы $\Delta(x)$ и $\Delta'(x)$ очень похожи: их можно реализовать циклами, различающимися лишь в маленькой (общей) окрестности этих точек x и x'. Например, в случае $n=2$ их разность $\Delta(x) - \Delta'(x)$ устроена следующим образом. Конус $S(x)$ с вершиной в точке x — это пара комплексных прямых. Каждая из них пересекает кривую $A_{\mathbb{C}}$ в d комплексных точках, одна из которых

лежит в маленькой окрестности точки x. Тогда цикл $\Delta(x) - \Delta'(x)$ равен сумме двух маленьких окружностей, лежащих в комплексификации $A_{\mathbb{C}}$ гиперболической поверхности A и обходящих вокруг этих близких к x точек пересечения с двумя прямыми $S(x)$.

Начнём двигать точку x в комплексной области. Каково будет ветвление интегралов по циклу $\Delta(x)$ при этом движении? Разложим $\Delta(x)$ в сумму $\Delta'(x) + (\Delta(x) - \Delta'(x))$ и соответственно разложим интегральную функцию в сумму интегралов по циклам, получающимся при перенесении над нашими путями из этих двух слагаемых. Первое слагаемое ветвиться не будет, потому что оно не ветвилось в том случае, когда движение начиналось из области гиперболичности. Всё ветвление связано с тем, что происходит с двумя окружностями. Таким образом, всё определяется тем, как перемещаются две точки пересечения конуса с комплексификацией A гиперболической поверхности.

Для первой точки есть не более d вариантов, куда она может перейти: это все точки пересечения соответствующей прямой с $A_{\mathbb{C}}$. Интегралы по окружностям, обходящим вокруг этих точек пересечения, дают не более чем d-значную аналитическую функцию. Интегралы по окружностям, соответствующим точкам пересечения с другой прямой, тоже дают не более чем d-значную функцию. Всего получаем не более чем d^2-значную функцию. Это означает, что в области множества $\mathbb{R}^2 \setminus A$, соседней с областью гиперболичности, сила притяжения совпадает с аналитической вектор-функцией, аналитическое продолжение которой не более чем d^2-значно. Эта функция имеет не более чем степенной рост при приближении к особым точкам, поэтому она совпадает с некоторой алгебраической функцией.

В следующей области будет не более чем $\binom{d}{2}^2$-значная алгебраическая функция, а в k-й от области гиперболичности зоне функция будет не более чем $\binom{d}{k}^2$-значная.

Для окружности $A = \{x^2 + y^2 = 1\}$ функция притяжения во внешней области имеет меньше значений, чем даётся такой оценкой: одно значение вместо четырёх. Это объясняется следующим. Главная однородная часть уравнения окружности совпадает с произведением уравнений прямых, составляющих множество $S(x)$. Поэтому одна из двух точек пересечения каждой из этих прямых с $A_{\mathbb{C}}$ бесконечно удалённая; при перемещении x в конечной области эта точка должна переходить сама в себя. Тем самым и оставшаяся точка, вокруг которой обходит цикл интегрирования, переходит только сама в себя. А если вместо окружности взять эллипс, то четвёрка получается.

Обобщая это рассуждение, получаем забавное наблюдение, которого, кажется, никто раньше не замечал. Пусть d чётно и к тому же для любого $i = 1, \ldots, d/2$ однородная часть степени $d/2 + i$ полинома, задающего гиперболическую поверхность A, делится на $(x^2 + y^2)^i$. Тогда у A будет $d/2$ овалов в $\mathbb{R}^2$.

У т в е р ж д е н и е. *У таких A в самой «внешней» области множества $\mathbb{R}^2 \setminus A$ сила притяжения вновь будет однозначной* (*т. е. рациональной*) *алгебраической функцией*.

Действительно, для x из этой области цикл $\Delta(x) - \Delta'(x)$ (где $\Delta'(x)$ по-прежнему «принесён» из области гиперболичности) равен сумме $d = 2 \times d/2$ окружностей в $A_{\mathbb{C}} \setminus S(x)$ — по $d/2$ вокруг точек пересечения каждой из двух прямых $S(x)$ с $A_{\mathbb{C}}$. Но в силу условия на старшие члены уравнения A это — все такие точки пересечения, лежащие в конечной области. Поэтому, перемещая x, мы можем только переставлять между собой эти окружности, сумма же их (и интеграл по этой сумме) не будет ветвиться.

Всю эту теорию можно обобщить на случай $n > 2$. Теория, которая позволяет в этом случае всё посчитать, немножко более сложная, чем просто теория монодромии изолированных особенностей гладких функций. Здесь нужна *теория монодромии изолированных особенностей полных пересечений*. Полное пересечение — это набор функций $(f_1, \ldots, f_k)\colon \mathbb{C}^n \to \mathbb{C}^k$, $k \leqslant n$, для которого коразмерность множества общих нулей равна k. Особенность такого набора (т. е. точка, в которой $f_1 = \ldots = f_k = 0$, но ранг матрицы Якоби $\{\partial f_j / \partial x_i\}$ меньше k) называется изолированной, если в её проколотой окрестности других особых точек нет (в частности, множество общих нулей в проколотой окрестности — гладкое многообразие).

Для таких наборов тоже есть развитая теория монодромии, которую в основном изучала немецкая школа Э. Брискорна (Х. Хамм, В. Эбелинг, Г.-М. Грёль и др.). Оказалось, что эта наука очень хорошо подходит к решению нашей задачи. А именно, рассмотрим пару функций (f_1, f_2), где $f_1 = 0$ — уравнение гиперболической поверхности, а f_2 — уравнение множества особенностей ядра оператора Лапласа или какого-то другого ядра, если мы решаем задачу о других потенциалах, отличных от потенциала Ньютона—Кулона. Такую пару функций будем рассматривать как полное пересечение. Тогда изучение рассмотренной выше группы гомологий и её монодромии хорошо укладывается, хотя и не совсем стандартным образом, в разработанную школой Брискорна теорию. Эта теория позволяет всё до конца посчитать в случае общего положения, когда главная однородная часть функции f_1 находится в общем положении с символом (главной

частью) ядра. В этом случае ответ следующий. Если размерность n произвольна, но степень d равна 2, то по-прежнему сила притяжения (а также и её потенциал) будет алгебраической и вне области гиперболичности. Группа монодромии в этом случае — достаточно сложная группа, порождённая отражениями. Она бесконечна, но орбита цикла интегрирования, который нас интересует, расположена на цилиндре, т. е. на гиперповерхности в пространстве $H_{n-1}(A_{\mathbb{C}} \setminus S(x), \mathbb{C})$, равной произведению линейного подпространства на эллипсоид. На образующих этого линейного подпространства интегральная функция обращается в нуль, поэтому всё сводится к задаче с конечной группой монодромии (орбита которой бегает по эллипсоиду).

Если же $n > 2$ и $d > 2$, то алгебраичности в общем случае не будет. В этом случае группа монодромии достаточно велика, что позволяет получить бесконечно много значений.

При нечётном n форма интегрирования не однозначна, а двузначна: при её определении использовалось извлечение квадратного корня. Поэтому в этом случае нужно рассматривать *группы гомологий с коэффициентами в локальной системе*. Такие локальные системы — это средство корректно интегрировать многозначные дифференциальные формы. Например, в нашем случае нужна локальная система со слоем $\mathbb{C}^1$ на $A_{\mathbb{C}} \setminus S(x)$, на слои которой обход вокруг $S(x)$ действует как умножение на -1.

Именно поэтому (в отличие от задач, рассматриваемых в предыдущем и следующем параграфах) задача о потенциалах имеет сходные решения в пространствах $\mathbb{R}^n$ с n любой чётности: при нечётных n теория монодромии таких подкрученных гомологий очень похожа на «неподкрученную» теорию гомологий для чётных n, и наоборот.

4.

Теория лакун Петровского для гиперболических операторов

Гиперболические уравнения в $\mathbb{R}^n$ — это уравнения в частных производных $Pf(x) = \phi(x)$, задача Коши для которых корректна в наиболее сильном смысле: такие уравнения обладают фундаментальным решением, носитель которого принадлежит собственному конусу в положительном полупространстве, а следовательно решение $f(x)$ этой задачи в точке x зависит от поведения начальных данных и правых частей лишь в некотором компактном подмножестве этого полупространства.

Это понятие тесно связано с понятием гиперболической гиперповерхности, рассматривавшимся в разделе 3. А именно, рассмотрим базисные

векторы $\partial/i\partial x_j$ комплексификации $\mathbb{C}^n$ исходного пространства $\mathbb{R}^n$ как базисные координатные функции ξ_j двойственного пространства $\check{\mathbb{R}}^n$. Тогда оператору в частных производных с постоянными коэффициентами, т. е. полиному P от переменных $\partial/i\partial x_j$, соответствует обычный полином P от координат ξ_j в $\check{\mathbb{R}}^n$. Если определяемое его главной однородной частью подмногообразие $A(P) \subset \check{\mathbb{R}}\mathbb{P}^{n-1}$ — неособая гиперповерхность, то эта гиперповерхность гиперболична (в смысле старого определения) относительно точки $\boldsymbol{\xi} \in \check{\mathbb{R}}\mathbb{P}^{n-1} \setminus A(P)$ тогда и только тогда, когда соответствующий оператор $P(\partial/i\partial x)$ гиперболичен (в новом смысле) для задачи Коши в полупространстве $\mathbb{R}^n_+$, отделяемом гиперплоскостью нулей ковектора с направлением $\boldsymbol{\xi}$. Для полиномов с особой главной частью соответствие немного более сложно.

Наиболее известным гиперболическим уравнением является волновое уравнение

$$\frac{\partial^2}{\partial x_1^2} = c^2 \sum_{j=2}^{n} \frac{\partial^2}{\partial x_j^2}, \tag{1}$$

описывающее, в частности, распространение звуковых волн со скоростью c. Фундаментальное решение этого уравнения (т. е. элементарная волна, возникающая из мгновенного точечного возмущения в начале координат нашего пространства-времени) имеет особенности на его *волновом фронте*, т. е. на конусе, заданном условиями

$$c^2 x_1^2 = \sum_{j=2}^{n} x_j^2, \qquad x_1 \geqslant 0. \tag{2}$$

В дополнении к фронту качественное поведение фундаментального решения существенно зависит от чётности n. В пространстве-времени чётной размерности $n > 2$ это фундаментальное решение тождественно равно 0 вне фронта: неподвижный наблюдатель слышит сигнал лишь в один момент, когда волна проходит через него. Напротив, в нечётномерном случае звук продолжает раздаваться (медленно затихая) всё время после момента первой встречи. Первое обстоятельство позволяет нам общаться при помощи звука в нашем 4-мерном пространстве-времени, а в силу второго «акустический слой» в океане, являясь прекрасным проводником отдельных сигналов, непригоден для быстрой передачи сложной информации.

Аналоги обоих этих вариантов качественного поведения распространения звуковых волн встречаются у произвольных гиперболических уравнений. Согласно терминологии общей теории (введённой И. Г. Петровским) говорят, что в чётномерном случае обе компоненты дополнения к волновому фронту являются *лакунами*, а в нечётномерном случае имеет место

диффузия волн со стороны внутренней компоненты. Внешняя компонента дополнения до фронта всегда является лакуной: *до* момента первой встречи звук не слышен.

Для любого гиперболического оператора соответствующее фундаментальное решение совпадает с аналитической функцией вблизи любой точки, лежащей вне некоторой конической гиперповерхности в $\mathbb{R}^n$ (которая, по аналогии с волновым уравнением, называется *волновым фронтом* и на которой сосредоточены все особенности этого решения). Именно, волновой фронт — это проективно двойственная поверхность к множеству $A(P)$, т. е. множество таких точек $x \in \mathbb{R}^n_\downarrow$, что проективизованная ортогональная к нему гиперплоскость $X(x) \subset \check{\mathbb{R}}\mathbb{P}^{n-1}$ находится не в общем положении с $A(P)$ (например, касается $A(P)$ в её неособой точке). В отличие от случая уравнений 2-го порядка, для операторов старших порядков волновой фронт может иметь особенности даже вдали от начала координат. Простейшими из них являются стандартные рёбра возврата, вблизи которых фронт устроен как произведение гладкого $(n-2)$-мерного многообразия на полукубическую параболу.

Имеется далеко продвинутая классификация особых точек волновых фронтов *строго* гиперболических операторов (строгость означает, что множество нулей главного символа неособо вне начала координат). Эта классификация почти точно соответствует классификации особых точек гладких функций: точке фронта соответствует особенность функции, графиком которой задаётся двойственная неособая гиперповерхность. Наиболее часто встречающиеся (и во многих отношениях наиболее замечательные) особенности функций и фронтов — это так называемые *простые* особенности, расклассифицированные В. И. Арнольдом: классы A_k $(k \geqslant 1)$, D_k $(k \geqslant 4)$, E_6, E_7, E_8. Например, неособым точкам фронтов соответствуют морсовские особенности A_1, простейшим рёбрам возврата — A_2. Вообще, класс особенностей с нижним индексом k на типичном фронте определяет особое множество коразмерности $k-1$.

Успех качественной теории гиперболических уравнений обеспечивается тем, что фундаментальные решения этих уравнений имеют интегральные представления, а следовательно их можно исследовать описанными выше методами.

Для точки x вне волнового фронта значение фундаментального решения (или по крайней мере достаточно высоких его частных производных) в этой точке задаётся явной интегральной формулой — *интегралом Гергломтца—Петровского—Лерэ* по *циклу Петровского*, представляющему собой элемент группы $H_{n-1}(X_{\mathbb{C}}(x) \setminus A_{\mathbb{C}}(P))$. Например, в случае чётных n этот цикл задаётся трубкой Лерэ в $X_{\mathbb{C}}(x) \setminus A_{\mathbb{C}}(P)$ вокруг множества ве-

щественных точек подмногообразия $A_{\mathbb{C}}(P) \cap X_{\mathbb{C}}(x)$; при нечётных n определение чуть более сложно.

Условие Петровского для точки x (и для всей содержащей её компоненты дополнения к фронту) состоит в том, что этот цикл гомологичен нулю: если оно выполнено, то в этой компоненте решение обязано совпадать с голоморфной функцией, определённой на всём $\mathbb{C}^n$ (и даже с полиномом, если оператор P однороден), иными словами, эта компонента является лакуной.

Атиа, Ботт и Гординг придумали локальный вариант этого условия, позволяющий объяснять и предсказывать *локальную* регулярность решений. А именно, если мы приближаемся к точке y волнового фронта со стороны компоненты его дополнения вдоль пути $x(\tau)$, $\tau \in [\epsilon, 0)$, то соответствующие множества $X_{\mathbb{C}}(x(\tau)) \cup A_{\mathbb{C}}(P)$ вырождаются в последний момент (соответствующий точке $x(0) = y$). Если, тем не менее, циклы Петровского, соответствующие точкам этого пути, стремятся к некоторому циклу в $X_{\mathbb{C}}(y) \setminus A_{\mathbb{C}}(P)$ *), то вблизи точки y ограничение фундаментального решения на эту компоненту совпадает с аналитической функцией, определённой на всей окрестности точки y. Последнее обстоятельство называется локальной (голоморфной) *резкостью* в точке y со стороны нашей компоненты.

Обратное утверждение (сформулированное Атиа, Боттом и Гордингом в качестве «сложной нерешённой проблемы») «почти верно»: для почти всех гиперболических операторов вблизи всех точек их волновых фронтов из резкости вытекает локальное условие Петровского, см. [1]. (Впрочем, имеются примеры очень вырожденных операторов, для которых это всё же не так.)

Доказательство этого основано на теории монодромии. Действительно, допустим, что плоскость $X_{\mathbb{C}}(y)$ имеет лишь конечное число точек нетрансверсальности с гиперповерхностью $A_{\mathbb{C}}(P)$ — для простоты, всего одну точку a. Тогда локальное гомологическое условие Петровского можно переформулировать как тривиальность некоторой локализации класса Петровского, а именно класса определяемого им относительного цикла в группе $H_{n-1}(B \cap X_{\mathbb{C}}(x) \setminus A_{\mathbb{C}}(P), \partial B \cap X_{\mathbb{C}}(x) \setminus A_{\mathbb{C}}(P))$, где B — маленький шар с центром в точке нетрансверсальности a. Если этот относительный класс нетривиален, то методами локальной теории монодромии особенностей

*) Это условие проще сформулировать так: при малых τ класс цикла Петровского гомологий в группе $H_{n-1}(X_{\mathbb{C}}(x(\tau)) \setminus A_{\mathbb{C}}(P))$ лежит в образе группы $H_{n-1}(X_{\mathbb{C}}(y) \setminus A_{\mathbb{C}}(P))$ при естественной проекции $X_{\mathbb{C}}(y) \to X_{\mathbb{C}}(x(\tau))$, отображающей любое фиксированное компактное подмножество множества $X_{\mathbb{C}}(y) \setminus A_{\mathbb{C}}(P)$ в множество $X_{\mathbb{C}}(x(\tau)) \setminus A_{\mathbb{C}}(P)$ с достаточно малым τ. Это условие Атиа, Ботт и Гординг назвали *локальным условием Петровского*.

функций (см. в частности рис. 3) удаётся доказать, что, двигая плоскость $X_{\mathbb{C}}(x)$ вдоль подходящих замкнутых путей в множестве близких к $X_{\mathbb{C}}(y)$ гиперплоскостей в $\check{\mathbb{C}}\mathbb{P}^{n-1}$, мы сможем добавить к циклу Петровского нетривиальный исчезающий цикл, интеграл по которому не равен 0, а следовательно ветвление аналитического продолжения нашей функции нетривиально в сколь угодно малой окрестности точки y.

Задача о локальной регулярности (резкости) вблизи *неособых* точек волновых фронтов была решена задолго до Атиа, Ботта и Гординга (напомним, что такие точки соответствуют при проективной двойственности непараболическим точкам множества нулей главного символа оператора). Ответ называется *критерием Давыдовой—Боровикова* и формулируется в терминах индексов инерции второй квадратичной формы фронта. Поскольку резкость одновременно выполнена или не выполнена для всех точек фронта, лежащих на одном луче, можно «проективизовать» задачу и говорить о резкости в точке y^* проективизации волнового фронта $W^* \subset \mathbb{R}\mathbb{P}^{n-1}$ со стороны той или иной компоненты дополнения к нему в $\mathbb{R}\mathbb{P}^{n-1}$. Зафиксируем локальные аффинные координаты $z_0, z_1, \ldots, z_{n-2}$ в $\mathbb{R}\mathbb{P}^{n-1}$ с центром в точке y^* таким образом, что гиперплоскость $\{z_0 = 0\}$ касается фронта в этой точке, а базисный вектор $\partial/\partial z_0$ направлен внутрь интересующей нас компоненты L^*. Вблизи этой точки фронт задаётся условием вида $z_0 = g(z_1, \ldots, z_{n-2})$, где f — гладкая функция, $g(0) = 0$, $dg(0) = 0$. Очевидно, сигнатура квадратичной формы $d^2g(0)$ не зависит от выбора таких координат.

Определение. Локальная компонента L^* дополнения к фронту удовлетворяет условию Давыдовой—Боровикова в точке y^*, если отрицательный индекс инерции этой квадратичной формы $d^2g(0)$ чётен.

Теорема 4. *Вблизи неособой точки фронта локальная компонента его дополнения является локальной лакуной* (*т. е. со стороны этой компоненты имеется резкость*) *тогда и только тогда, когда она удовлетворяет условию Давыдовой—Боровикова.*

Часть «только тогда» была доказана в 1945 г. в диссертации ученицы Петровского А. М. Давыдовой, более сложная часть «тогда» доказана в конце 1950-х учеником И. М. Гельфанда В. А. Боровиковым при помощи явных и трудных аналитических выкладок. При помощи же топологического локального критерия Петровского (—Атиа—Ботта—Гординга) она доказывается без каких-либо формул: нужно лишь убедиться в том, что критерий Давыдовой—Боровикова эквивалентен тривиальности локализованного класса Петровского.

Л. Гординг рассмотрел также поведение фундаментального решения вблизи рёбер возврата волнового фронта, в частности нашёл все локаль-

ные компоненты дополнения к фронту вблизи таких рёбер, со стороны которых имеется резкость. Ответ оказался очень прост. Во-первых, со стороны «бóльшей» компоненты дополнения к фронту вблизи ребра возврата резкость никогда не наблюдается. Во-вторых, если n чётно, то резкости также не бывает. В оставшемся случае нечётного n и внутренней компоненты есть ещё два подслучая, различающиеся тем, какова чётность индекса инерции второй квадратичной формы поверхности фронта в близких неособых точках. В одном из них резкости очевидно не может быть, поскольку в силу критерия Давыдовой—Боровикова её нет уже при приближении из этой компоненты к соседним с ребром возврата неособым точкам. И только в оставшемся случае резкость есть! (Заметим, что в случае другой чётности этих индексов (также при нечётном n) имеется локальная резкость при подходе из другой («большой») компоненты к любой неособой точке её границы, но при подходе к самому ребру возврата решение всё же начинает ветвиться.) Гординг получил также полное описание локальных лакун вблизи следующего по сложности класса особенностей волновых фронтов — A_3.

После того как было замечено, что вся эта теория включается в теорию особенностей гладких функций, с помощью методов этой теории удалось получить полное описание лакун вблизи всех простых (т. е. классов A_k, D_k, E_k) и многих непростых особенностей волновых фронтов. В частности, имеется следующий геометрический критерий локальной резкости вблизи простых особенностей.

Теорема 5. *Локальная компонента дополнения к фронту вблизи его простой особенности является локальной лакуной* (*т. е. фундаментальное решение резко с её стороны*) *тогда и только тогда, когда*

а) *вблизи всех неособых компонент фронта, ограничивающих эту компоненту, выполнено условие Давыдовой—Боровикова, и*

б) *граница компоненты не содержит рёбер возврата, по отношению к которой эта компонента являлась бы «бóльшей» компонентой дополнения.*

Необходимость обоих этих условий очевидна из сказанного выше.

5.
Гипергеометрические интегралы

Общие гипергеометрические функции — это очень широкий класс аналитических функций, обобщающий гипергеометрическую функцию Гаусса (а также Г- и В-функции Эйлера). Они задаются интегральными

представлениями следующего вида. Рассмотрим в $\mathbb{C}^n$ набор голоморфных функций (как правило, полиномов) $f_\lambda = (f_{1,\lambda}, \ldots, f_{k,\lambda})$, зависящих от параметра $\lambda \in \mathbb{C}^M$ (т. е. имеется набор функций $F_j(x, \lambda)\colon \mathbb{C}^n \times \mathbb{C}^M \to \mathbb{C}$, $j = 1, \ldots, k$, и для любого $\lambda \in \mathbb{C}^M$ полагаем $f_{j,\lambda} \equiv F(\cdot, \lambda)$). Обозначим через $A(f_\lambda)$ множество нулей функции $f_{1,\lambda} \cdot \ldots \cdot f_{k,\lambda}$ в $\mathbb{C}^n$ и через $\bar{A}(f_\lambda) \subset \mathbb{CP}^n$ его объединение с несобственной плоскостью $\mathbb{CP}^{n-1}_\infty$ в $\mathbb{CP}^n$. Для почти всех значений параметра λ соответствующие пары топологических пространств $(\mathbb{CP}^n, \bar{A}(\lambda))$ устроены топологически одинаково. Множество таких недискриминантных значений λ будет базой искомого расслоения $E \to T$ (см. введение), слоем его над точкой λ будет пространство $\mathbb{C}^n \setminus A(\lambda)$ (или, если угодно, $\mathbb{CP}^n \setminus \bar{A}(\lambda)$), а интегральная функция на T задаётся интегралом

$$\Gamma(\lambda;\, \alpha;\, \Omega) = \int\limits_\Omega f_{1,\lambda}^{\alpha_1}(x) \cdot \ldots \cdot f_{k,\lambda}^{\alpha_k}(x) dx_1 \wedge \ldots \wedge dx_n \tag{1}$$

по некоторому замкнутому циклу Ω в $\mathbb{C}^n \setminus A(\lambda)$ (непрерывно зависящему от λ, так чтобы всё время оставаться циклом); здесь α_j — некоторые комплексные числа.

Задача о поведении таких интегральных функций чрезвычайно содержательна уже в одномерной ситуации. В этом случае интеграл (1) сводится к виду

$$\Gamma(A;\, \alpha;\, \Omega) = \int\limits_\Omega (z - a_1)^{\alpha_1} \cdot \ldots \cdot (z - a_k)^{\alpha_k} dz, \tag{2}$$

где интегрирование идёт, например, по гладкому интервалу, соединяющему пару точек a_i, a_j (и не проходящему через другие точки a_l), или по линейной комбинации таких интервалов, или по такому интервалу, соединяющему одну из точек a_j с бесконечностью, или по некоторому компактному пути в $\mathbb{C}^1 \setminus \bigcup \{a_j\}$. Точки конфигурации $A = \bigcup \{a_j\}$ являются и параметрами этой системы (играющими роль параметров λ в интеграле (1)). Дробно-линейным преобразованием можно превратить параметр a_1 в 0, а a_2 в 1, тогда при $k = 3$ мы получим классическую гипергеометрическую функцию Гаусса.

Обычно задачи об интегралах (1), (2) имеют вещественную мотивировку: исходно, все полиномы F_j вещественны (т. е. принимают вещественные значения на $\mathbb{R}^n \times \mathbb{R}^M$), и при вещественном λ соответствующий цикл $\Omega(\lambda)$ — это некоторая компонента множества $\mathbb{R}^n \setminus A(\lambda)$.

С этими интегралами связаны следующие основные вопросы.

0. Правильное определение контуров интегрирования: элементами какой группы гомологий их следует считать?

1. Изучение особенностей интеграла как функции показателей α_j. Зафиксируем функции $f_{j,\lambda}$ и будем менять набор показателей $\alpha = (\alpha_1, \ldots, \alpha_k)$. Тогда (если только все компоненты $\{f_{j,\lambda}\}$ и $\mathbb{CP}^{n-1}_\infty$

множества $\bar{A}(\lambda)$ не имеют общих точек в $\mathbb{CP}^n$) для почти всех наборов α интеграл (1) корректно определён (хотя его определение иногда связано с трудностями), а именно он является мероморфной функцией от α, множество полюсов которой пробегает конечный набор бесконечных семейств параллельных гиперплоскостей в $\mathbb{C}^k$: именно, r-е семейство характеризуется набором целых чисел $\theta_{r,j}$, $j=1, \ldots, k$, и плоскости из этого семейства задаются условием, что линейная комбинация $\sum_{j=1}^{k} \theta_{r,j}\alpha_j$ является достаточно малым целым числом. Наборы коэффициентов $\Theta_r = \{\theta_{r,j}\}$ называются *резонансами* нашей задачи и зависят только от набора функций f_j. Они всегда включают в себя $k+1$ тривиальный набор $(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1), (-1, \ldots, -1)$, отвечающий всем $k+1$ компонентам множества $\bar{A}(\lambda)$. Например, при $n=1$ множество резонансов исчерпывается этими стандартными наборами.

Задача состоит в явном описании этого набора полюсов (и корректном определении интегральной функции от показателей α, при котором она действительно становится мероморфной).

2. Ветвление интегралов, как функций от λ. Можно, напротив, зафиксировать набор показателей α и начать менять параметр $\lambda \in T = \mathbb{C}^M \setminus \Sigma$, соответственно непрерывно изменяя контур интегрирования $\Omega = \Omega(\lambda)$. Как будет изменяться при этом функция (1)? Даже если мы вначале рассматривали «естественную» вещественную задачу с интегрированием по областям дополнения к $A(\lambda)$ в $\mathbb{R}^n$ (когда проблем типа задачи 0 не возникает), для изучения асимптотических свойств соответствующего интеграла (например, характера роста коэффициентов его разложения в степенной ряд, тесно связанного со свойствами аналитического продолжения этой функции) полезно выйти в комплексную область и посмотреть, как преобразуется цикл $\Omega(\lambda)$ при λ обходящем вдоль замкнутых путей: аналитическое продолжение интеграла (1) по параметру λ — это просто тот же интеграл (1) по новому циклу, полученному в результате такого преобразования.

3. Размерность пространства интегральных функций. Рассмотрим пространство всех существенно различных контуров интегрирования $\Omega(\lambda)$ (для какого-то фиксированного значения $\lambda = \lambda_0$) и, соответственно, пространство ростков в точке λ_0 интегральных функций вида (1). Какова будет размерность последнего пространства? Конечно, она ограничена сверху размерностью пространства контуров, но, во-первых, как посчитать эту размерность, и во-вторых будет ли эта оценка достигаться?

Эти вопросы важны в специфических областях теории дифференциальных уравнений — теории так называемых голономных систем, D-мо-

дулей и пр. Действительно, интегральные функции (1) удовлетворяют таким системам уравнений (первым примером которых является опять-таки гипергеометрическое уравнение Гаусса), и эти вопросы интерпретируются как вопросы о размерности пространства решений таких уравнений. *)

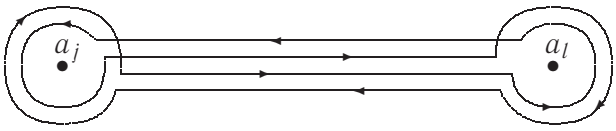
Например, хорошо известно, что для обычного гипергеометрического уравнения Гаусса с типичным набором показателей α эти размерности равны 2; более общим образом, для интеграла (2) с произвольным k они равны $k-1$. Ещё одна интересная проблема здесь — какие наборы показателей являются нетипичными для данной задачи, т. е. размерности наших пространств изменяются.

Конечно, можно пытаться ответить на эти вопросы с помощью тупого разложения в степенные ряды и сравнения нескольких первых коэффициентов. Однако наиболее красивый (и, по-видимому, единственный эффективно обобщающийся на многомерные варианты этих задач) метод состоит в использовании соображений монодромии, то есть результатов решения задачи 2. В наиболее важных ситуациях группа монодромии настолько богата, что, начиная с одного-единственного цикла интегрирования, мы можем породить всё пространство таких циклов (т. е. полную систему его образующих), и, что ещё важнее, доказать, что интеграл по любой нетривиальной линейной комбинации таких циклов не равен тождественно нулю.

Мы подробно рассмотрим все эти вопросы для случая одномерных интегралов (2), а затем наметим его обобщение на многомерный случай (1). Эта часть в основном следует работе [2], более развёрнутое её изложение см. в последней главе книги [1].

Решение вопроса 0 зависит от того, какую задачу — 1 или $2+3$ — мы собираемся решать в дальнейшем. В любом случае, полезно начать с *гомологического $\mathbb{Z}^k$-листного накрытия* $\Xi(A)$ над проколотой комплексной прямой $\mathbb{C}^1 \setminus A$, на которое любая дифференциальная форма, стоящая под интегралом в (2), поднимается как однозначная. Это накрытие определяется условием, что замкнутая петля в $\mathbb{C}^1 \setminus A$ поднимается до замкнутой петли в его пространстве тогда и только тогда, когда её класс в $H_1(\mathbb{C}^1 \setminus A)$ равен 0. Контуры интегрирования, рассматриваемые в задаче 1, — это

*) A priori, интегралы (1) могут давать лишь часть решений соответствующей системы уравнений, но в интересных ситуациях, таких как, например, *общее многомерное гипергеометрическое уравнение Гельфанда на открытом страте грассманиана* (что бы ни означал этот набор слов) можно получить верхнюю оценку числа решений системы дифференциальных уравнений (И. М. Гельфанд, С. И. Гельфанд, 1986) и его же нижнюю оценку (предъявив нужное количество функций, заданных интегралами (1), см. [2]) и убедиться, что обе оценки совпадают!



Р и с. 5. Одномерная двойная петля

1-мерные циклы, т. е. элементы группы 1-мерных гомологий этого накрытия. Рассматривают две такие группы гомологий: группу $H_*(\Xi(A))$ обычных гомологий комплекса конечных сингулярных цепей (с коэффициентами в $\mathbb{C}$), и группу $H_*^{\mathrm{rf}}(\Xi(A))$ гомологий комплекса локально конечных сингулярных цепей, проекции которых в $\mathbb{C}^1 \setminus A$ также локально конечны. Типичные примеры элементов последней группы доставляют интервалы, соединяющие точки a_i, $a_j \in A$ или точки a_j и ∞ (и поднятые на тот или иной лист накрытия). Именно они являются контурами интегрирования, рассматриваемыми в классической теории гипергеометрических функций Гаусса.

Метод регуляризации несобственных интегралов, решающий задачу 1, состоит в замене несобственных интегралов по таким некомпактным контурам интегралами по подходящим компактным контурам из группы $H_1(\Xi(A))$.

Например, для интервала (a_j, a_l) его регуляризующий цикл изображён на рис. 5: этот цикл надо поднять в накрытие $\Xi(A)$ так, чтобы его горизонтальный отрезок, изображённый вторым снизу, попал на тот же лист, что и исходное поднятие интервала (a_j, a_l). Тогда для всех таких $\alpha = (\alpha_1, \ldots, \alpha_k)$, для которых интеграл (2) по интервалу сходится, (т. е. при $\operatorname{Re}\alpha_j > -1 < \operatorname{Re}\alpha_l$) интегралы по старому (некомпактному) и новому (компактному) циклам тождественно совпадают с точностью до коэффициента

$$\frac{1}{(1-e^{2\pi i\alpha_j})(1-e^{2\pi i\alpha_l})}. \tag{3}$$

Но с интегралом по компактному циклу ничего плохого не может происходить ни при каких α: он определяет целую голоморфную функцию на всём пространстве $\mathbb{C}^k$ показателей α. Следовательно, интеграл по интервалу аналитически продолжается из области сходимости до мероморфной функции, которая может иметь полюса только там, где их имеет коэффициент (3), т. е. при целых значениях α_j или α_l. В действительности, если эти значения неотрицательны, то полюса у интеграла нет: это означает, что интеграл по соответствующему компактному циклу должен иметь нули.

Любой элемент группы $H_1^{\mathrm{rf}}(\Xi(A))$ можно свести к конечной линейной комбинации подобных интервалов. Регуляризуя каждый из них по отдельности при помощи своей «двойной петли» как на рис. 5 мы получим доказательство мероморфности аналитического продолжения любого интеграла (2) по параметру α. А именно, мы докажем, что множество его полюсов принадлежит множеству *резонансных* наборов показателей, т. е. таких что либо одно из его чисел α_j, либо сумма всех этих чисел является целым числом.

Регуляризующий цикл, аналогичный рис. 5, удаётся построить и в многомерной задаче, связанной с интегралом (1). Точную комбинаторную конструкцию этого цикла (в [2] он был назван двойной петлёй) см. в § I.10.7 книги [1]. Однако коэффициенты, с которыми они переходят в соответствующие некомпактные циклы, вообще говоря зависят от показателей α_j гораздо сложнее, чем в формуле (3).

В частности, из этого следует, что множество резонансных значений α не всегда исчерпывается условиями $\alpha_j \in \mathbb{Z}$, $\sum_{j=1}^{k} \alpha_j \in \mathbb{Z}$, как в одномерной задаче. Это сохраняется только если множество $\bar{A} \subset \mathbb{CP}^n$ является *дивизором с нормальными пересечениями*. Если же это не так, то, используя алгоритм Хиронаки, можно разрешить его сложные особенности и всё же превратить его в алгебраическое множество с нормальными пересечениями. При этом может потребоваться вклеить несколько новых компонент этого множества. Вблизи неособой точки любой такой компоненты, поднятая дифференциальная форма (1) в некоторых локальных координатах запишется в виде $z_1^{\beta} dz_1 \wedge \ldots \wedge dz_n$, где $\beta = \beta(\alpha)$ — некоторая линейная функция с целыми коэффициентами от показателей α_j. В эту линейную комбинацию с ненулевыми коэффициентами могут входить лишь показатели α_j, соответствующие компонентам множества $\bar{A}$, содержащим множество, при раздутии которого возникла эта новая компонента. (Если эта компонента — старая, т. е. собственный прообраз компоненты, определяемой уравнением f_j, то $\beta = \alpha_j$, а если это собственный прообраз несобственной плоскости $\mathbb{CP}^{n-1}_{\infty}$, то $\beta = -(\sum \alpha_j + n + 1)$.) Тогда резонансными окажутся и все такие наборы $\alpha = (\alpha_1, \ldots, \alpha_j)$, что эта линейная комбинация $\beta(\alpha)$ будет целым числом для хотя бы одной компоненты любого разрешения множества $\bar{A}$.

Первая общая теорема мероморфности такого рода (для вещественной постановки задачи), высказанная в качестве гипотезы И. М. Гельфандом, была доказана И. Н. Бернштейном и С. И. Гельфандом с одной стороны и М. Атиа с другой немедленно после доказательства Хиронакой его теоремы.

Поскольку мы теперь займёмся задачей $2+3$, то самое время заметить, что рассматривавшиеся выше группы гомологий

$$H_*(\Xi(A)), \quad H_*^{\mathrm{rf}}(\Xi(A)) \tag{4}$$

чрезвычайно избыточны. Действительно, рассмотрим два цикла в пространстве накрытия $\Xi(A)$, лежащие точно друг над другом на разных листах этого накрытия, так что их проекции в $\mathbb{C}^1 \setminus A$ совпадают. Путь в $\Xi(A)$, соединяющий соответствующие точки этих циклов, проектируется в петлю, индексы зацепления которой с точками a_j равны каким-то числам p_j, $j=1,\ldots,k$. Тогда интегралы формы (2) по этим циклам отличаются множителем

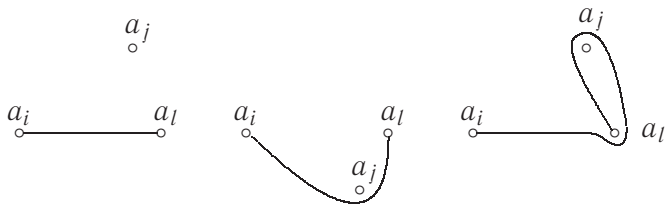
$$\exp(2\pi i(\alpha_1 p_1 + \ldots + \alpha_k p_k)). \tag{5}$$

Более того, это же соотношение сохранится если мы начнём двигать конфигурацию точек A, соответственно меняя контуры интегрирования. Поэтому число линейно независимых интегральных функций (1) (рассматриваемых как функции от A при фиксированных α) будет не больше, чем размерность пространства циклов в $\Xi(A)$, профакторизованных не только по отношению гомологичности, но и по действию группы $\mathbb{Z}^k$ (переводящему всякий цикл во все остальные циклы, имеющие ту же проекцию и взятые с подходящими коэффициентами вида (5)).

Удобное определение такого пространства существенно различных циклов интегрирования даёт *теория гомологий с коэффициентами в локальной системе групп*, связанной с ветвлением дифференциальной формы (2). Одно из её эквивалентных определений *) состоит в том, что мы рассматриваем комплекс цепей в $\Xi(A)$, факторизуем его по описанному действию группы $\mathbb{Z}^k$, и считаем гомологии полученного факторкомплекса. Если мы начинали с группы конечных цепей в $\Xi(A)$, то полученная группа обозначается $H_*(\mathbb{C}^1 \setminus A, L_\alpha)$, а если с группы локально конечных цепей, проекция которых в $\mathbb{C}^1 \setminus A$ также локально конечна, то $H_*^{\mathrm{lf}}(\mathbb{C}^1 \setminus A, L_\alpha)$. При любом фиксированном α корректно определён интеграл (2) по любому элементу первой из этих групп, а если α нерезонансно — то и второй.

З а м е ч а н и е. Эта факторизация была бы преждевременной когда мы занимались задачей 1. Действительно, эти группы не обладают никакой естественной связностью по α, т. е. способом отождествлять между собой элементы таких групп, соответствующих близким но различным значениям α. С другой стороны, если α фиксировано, но слегка меняется конфигурация A, то имеется естественная связность Гаусса—Манина, отождествляющая такие группы с близкими A.

*) Другое см., например, в [1].



Р и с. 6. Монодромия гипергеометрического интеграла

Из сказанного ясно, что очень важна задача о вычислении размерностей групп

$$H_*(\mathbb{C}^1 \setminus A,\, L_\alpha), \quad H_*^{\mathrm{lf}}(\mathbb{C}^1 \setminus A,\, L_\alpha). \tag{6}$$

Эта задача (и, более общим образом, задача о вычислении аналогичных групп $H_*(\mathbb{CP}^n \setminus \bar{A},\, L_\alpha)$, $H_*^{\mathrm{lf}}(\mathbb{CP}^n \setminus \bar{A},\, L_\alpha)$ для многомерных интегралов) легко решается в случае нерезонансных наборов коэффициентов.

Т е о р е м а 6. *Если набор α — нерезонансный* *), *то естественное отображение*

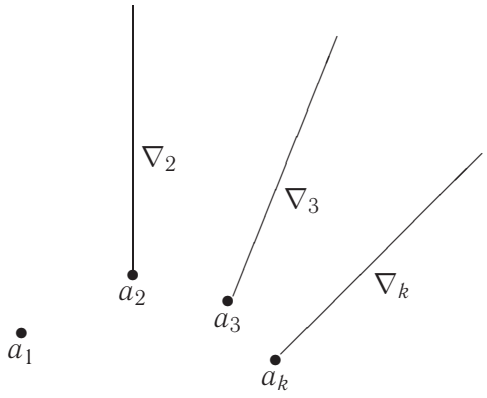
$$H_*(\mathbb{CP}^n \setminus \bar{A},\, L_\alpha) \to H_*^{\mathrm{lf}}(\mathbb{CP}^n \setminus \bar{A},\, L_\alpha) \tag{7}$$

является изоморфизмом.

Но $\mathbb{CP}^n \setminus A$ — многообразие Штейна, откуда следует, что левая группа (7) может быть нетривиальна только в размерностях, не превосходящих n, а правая — только в размерностях, не меньших n. Следовательно, обе группы равны 0 во всех размерностях, кроме n, а в размерности n изоморфны $\mathbb{C}^{|\chi|}$, где χ — эйлерова характеристика множества $\mathbb{CP}^n \setminus \bar{A}$.

Например, в одномерном случае для почти всех α обе группы (6) изоморфны $\mathbb{C}^{k-1}$. В этом случае пространством параметров подынтегральных функций является пространство $\mathbb{C}^k$ всевозможных положений точек $a_1, \ldots, a_k$. Дискриминант Σ — это в точности множество наборов, в которых какая-то пара точек a_i, a_j совпадает. Фундаментальная группа оставшегося множества $\mathbb{C}^k \setminus \Sigma$ называется группой крашеных кос с k нитями, она порождена всевозможными петлями в пространстве конфигураций, при которых некоторая точка a_j обходит вокруг другой (скажем, a_l) и возвращается обратно, остальные же точки неподвижны; см. рис. 6. Представление монодромии этой группы описывается этим же рисунком: например, к интервалу (a_i, a_l) такой обход добавит класс интервала (a_j, a_l), поднятого на некоторый лист накрытия и взятого с коэффициентом $(1 - e^{2\pi i \alpha_j})$. Если число α_j не целое, то этот добавок нетривиален

*) В смысле, указанном на стр. 54.



Р и с. 7. Базис для подкрученных гомологий

(а если α_l нецелое, то нетривиально действие этого же обхода на интервал (a_i, a_j)).

Это даёт решение задачи 2 в простейшей одномерной ситуации. Этого достаточно и для решения задачи 3. Действительно, в случае нерезонансных α_j в качестве образующих группы $H_1^{\mathrm{lf}}(\mathbb{C}^1 \setminus A, L_\alpha) \sim \mathbb{C}^{k-1}$ можно взять классы интервалов $\nabla_2, \ldots, \nabla_k$, соединяющих точки $a_2, \ldots, a_k$ с ∞, см. рис. 7.

П р е д л о ж е н и е. *Для любого фиксированного нерезонансного набора показателей* $\alpha_1, \ldots, \alpha_k$, *число линейно независимых интегральных функций* (2) *в любой односвязной области конфигурационного пространства* $\mathbb{C}^k \setminus \Sigma$ *равно* $k-1$.

Иными словами, интегралы по линейно независимым элементам группы $H_1^{\mathrm{lf}}(\mathbb{C}^1 \setminus A, L_\alpha)$ являются линейно независимыми функциями от $a_1, \ldots, a_k$.

Д о к а з а т е л ь с т в о. Отрицание предложения состоит в следующем: найдётся нетривиальная линейная комбинация $\sum \nu_j \nabla_j$ этих интервалов (коэффициенты ν_j которой суть сечения локальной системы L_α) такая, что интеграл (2) по этой линейной комбинации тождественно равен 0 при всех конфигурациях, близких к $(a_1, \ldots, a_k)$ (а значит, в силу аналитичности, и при всех конфигурациях вообще).

Рассмотрим петлю в конфигурационном пространстве, вдоль которой эти точки $a_2, \ldots, a_k$ остаются неподвижными, а точка a_1 подходит к одной из них (не пересекая при этом интервалов ∇_j), затем обходит вокруг этой точки a_j и возвращается назад по старому пути. Согласно предыдущему, при этом контур $\sum \nu_j \nabla_j$ увеличится на класс интервала (a_1, a_j), взятого с коэффициентом типа $\nu_j(e^{2\pi i \alpha_j} - 1)$. Интеграл по добавленному конту-

ру должен тождественно равняться нулю: действительно, в силу нашего предположения он равен приросту *тождественно нулевой* функции при её аналитическом продолжении вдоль некоторого замкнутого пути в пространстве аргументов. Но интеграл по интервалу (a_1, a_j) не равен тождественно нулю: это достаточно проверить в случае, когда a_1 и a_j достаточно близки. Следовательно, коэффициент ν_j должен равняться 0, и так для каждого j. □

Аналогичные результаты имеют место и для многомерных интегралов (1). Одним из интереснейших примеров являются *общие гипергеометрические функции И. М. Гельфанда—Аомото*, получающиеся когда все функции f_j — линейные, и множества их нулей находятся в общем положении в $\mathbb{CP}^n$. В этом случае размерность группы $H_n^{\mathrm{lf}}(\mathbb{CP}^n \setminus A, L_\alpha)$ при типичных α равна $\binom{k-1}{n}$. Методы *теории Пикара—Лефшеца с подкрученными коэффициентами* в совокупности с обобщением предыдущих рассуждений позволяют доказать (см. [2]), что интегралы по линейно независимым элементам этой группы линейно независимы.

С другой стороны, для тесно связанной с этой задачей системы *обобщённых гипергеометрических уравнений Гаусса* методами теории D-модулей было доказано (И. М. Гельфанд, С. И. Гельфанд, 1986) что число решений этой системы не превосходит $\binom{k-1}{n}$. A priori известно, что любая интегральная функция даёт решение этой системы уравнений. Таким образом, получается, что для нерезонансных α имеется ровно $\binom{k-1}{n}$ решений этой голономной системы уравнений, причём *все* они даются интегральными представлениями.

В действительности, теорема невырожденности (утверждающая что интегралы по линейно независимым циклам дают линейно независимые интегральные функции) имеет место во всех интересных ситуациях (кроме специально подобранных экзотических контрпримеров); о некоторых других её проявлениях см. также [2], [1].

Литература

[1] В. А. Васильев. Ветвящиеся интегралы. — М.: МЦНМО, 2000. — 432 с.

[2] В. А. Васильев, И. М. Гельфанд, А. В. Зелевинский. Общие гипергеометрические функции на комплексных грассманианах // Функц. анализ и его прилож. — 1987. — Т. 21, вып. 1. — С. 23—38.

6 апреля 2000 г.

Б. Л. Фейгин

КОНФОРМНЫЕ ТЕОРИИ ПОЛЯ

Речь пойдёт о таком новом алгебраическом объекте, как Vertex Operator Algebra (VOA). С теорией представлений таких алгебр связано много достижений в математике последнего времени. В определённом смысле примыкающими к ним можно назвать и теорию узлов, и даже деформационное квантование Концевича — оно может быть перетолковано как некое утверждение такого типа. Лет 5 назад было популярно изучение связи вертексных операторных алгебр с конечными простыми группами.

1.
Алгебра

Теперь, после краткого введения, я скажу, о каком алгебраическом объекте идёт речь. Аккуратного определения мне дать не удастся; к этому я и стремиться не буду. Я дам только «определение-описание» вертексных операторных алгебр. Прежде всего это алгебра (над $\mathbb{C}$). Эту алгебру удобно описывать, имея одновременно её представление в градуированном пространстве $W = \bigoplus W_i$, $i \in \mathbb{Z}$, $i \geqslant 0$. На этом пространстве действуют элементы, порождающие алгебру. Предполагается, что есть набор индексов α, и каждому α сопоставлен формальный символ (производящая функция)

$$a^\alpha(z) = \sum_{n \in \mathbb{Z}} a_n^\alpha z^{-n}.$$

Это означает, что на пространстве W действуют операторы a_n^α: $W_i \to W_{i+n}$; другими словами, оператор a_n^α имеет градуировку n. Эти операторы a_n^α являются элементами алгебры; больше об алгебре пока ничего не сказано.

Если есть два индекса α и β, то (по крайней мере формально) можно записать произведение

$$a^\alpha(z_1)a^\beta(z_2) = \sum_{n_1, n_2} a_{n_1}^\alpha a_{n_2}^\beta z_1^{-n_1} z_2^{-n_2}.$$

Буква z должна иметь содержательный смысл. Посмотрим, как устроены матричные элементы операторов $a^\alpha(z_1)a^\beta(z_2)$. Для этого нужно фиксировать вектор $w \in W$ и ковектор $\check{w} \in W^*$ и рассмотреть число $\langle \check{w}, a^\alpha(z_1)a^\beta(z_2)w\rangle$ (матричный элемент); физики обычно используют обозначение $\langle \check{w}|a^\alpha(z_1)a^\beta(z_2)|w\rangle$. Матричный элемент является формальным рядом от z_1 и z_2. Основное требование состоит в том, чтобы это был не просто какой-то формальный ряд, а разложение в ряд некоторой функции $F^{\alpha,\beta}(z_1, z_2)$ (эта функция зависит от w и $\check{w}$). Эта функция должна (при любых w и $\check{w}$) удовлетворять следующим условиям:

1) F — рациональная функция от z_1 и z_2, у которой полюс может быть только при $z_1 = z_2$ (в большинстве примеров порядок полюса ограничен некой константой, не зависящей от w и $\check{w}$);

2) $F^{\alpha,\beta}(z_1, z_2) = F^{\beta,\alpha}(z_2, z_1)$ («локальность» или «коммутативность»; иногда допускается косокоммутативность).

Обсудим, что такое локальность (коммутативность). Для простоты рассмотрим $F^{\alpha,\alpha}(z_1, z_2)$. Вместо $a^\alpha(z)$ я буду писать $a(z)$. Напомню, что $a(z) = \sum a_n z^{-n}$. Предположим, что для любых w и $\check{w}$ у функции F полюса нет. Тогда из локальности (коммутативности) и того, что в данном случае проблемы разложения в ряд не существует), следует, что операторы a_n и a_m коммутируют. Если же полюс есть, то эти операторы коммутировать не будут; что будет происходить в этом случае, я скажу чуть позже.

Давайте усилим требование: предположим, что для любых w и $\check{w}$ не только нет полюса, но и функция $F(z_1, z_2)$ обращается в нуль при $z_1 = z_2$. Из симметричности этой функции следует, что она делится на $(z_1 - z_2)^2$. Из этого легко вывести тождество $a^2(z) = 0$. Это означает, что операторы a_n и a_m не только коммутируют, но и выполняется равенство $\sum\limits_{n_1+n_2=m} a_{n_1}a_{n_2} = 0$, которое можно рассматривать как усиление коммутативности. Написанную бесконечную сумму нужно понимать следующим образом. Пространство $W = W_0 + W_1 + \ldots$ градуированное, а оператор a_n повышает градуировку на n. Поэтому если применить оператор $\sum\limits_{n_1+n_2=m} a_{n_1}a_{n_2} = 0$ к вектору из W_i, то сумма будет конечной.

Если потребовать, чтобы функция $F(z_1, z_2)$ на диагонали обращалась в нуль с кратностью 4 (для любых векторов w и $\check{w}$), то к написанному выше соотношению прибавится ещё соотношение $\sum\limits_{n_1+n_2=m} a_{n_1}a_{n_2}n_1n_2 = 0$.

Более сложный случай, подробности которого мне придётся пропустить, возникает, когда функция $F(z_1, z_2)$ при $z_1 = z_2$ имеет полюс не выше 2-го порядка (для любых векторов w и $\check{w}$). В этом случае можно доказать, что существует система операторов b_i (иначе говоря, существует

$b(z)=\sum b_i z^{-i}$), для которых выполняется условие

$$[a_{n_1},\, a_{n_2}]=(n_1-n_2)b_{n_1+n_2}.$$

И это условие равносильно наличию полюса не выше 2-го порядка. Если допустить наличие полюса 4-го порядка, то получится более сложная формула; зависимость от индексов будет не линейной, а 3-й степени.

П р и м е р. Алгебра Вирасоро — это алгебра Ли с образующими C и $\{L_i\}$, $i\in\mathbb{Z}$. Коммутаторы такие:

$$[L_i,\, L_j]=(j-i)L_{i+j}+\delta_{i+j,0}\frac{i^3-i}{12}C,\quad [L_i,\, C]=0.$$

Чтобы привести это к той ситуации, которую я описывал, нужно сделать так. В качестве пространства W нужно взять модуль Верма. Иначе говоря, это пространство, в котором есть вектор v и на него образующие алгебры Вирасоро действуют так: $L_i v=0$ при $i>0$; $L_0 v=hv$, где h — число; $Cv=cv$, где c — число. Тогда модуль Верма порождён действием на вектор v остальных образующих:

$$W=\mathbb{C}[L_{-1},\, L_{-2},\, \dots]=W_0+W_{-1}+W_{-2}+\dots$$

Эти стандартные соглашения о нумерации противоположны тем, которые были приняты в начале лекции. Но это не существенно.

На таком пространстве действуют операторы L_i. Поэтому можно образовать величину $T(z)=\sum L_i z^{-i}$ и вычислить матричные элементы $\langle\check{w}|T(z_1)T(z_2)|w\rangle$. В результате получится, что функция $F(z_1,\, z_2)$ имеет полюс 4-го порядка:

$$F(z_1,\, z_2)=(z_1-z_2)^{-4}G(z_1,\, z_2).$$

Функция $G(z_1,\, z_2)$ не имеет особенностей на диагонали. Более того, $G(z,\, z)$ — константа (которая зависит от w и $\check{w}$).

П р и м е р. Пусть есть $a(z)$ и для любых w и $\check{w}$ соответствующая функция $F(z_1,\, z_2)$ имеет полюс не выше 2-го порядка. В этом случае $\langle\check{w}|a(z_1)a(z_2)a(z_3)|w\rangle$ — рациональная функция

$$H(z_1,\, z_2,\, z_3)=\prod(z_i-z_j)^{-2}G(z_1,\, z_2,\, z_3),$$

где функция G симметрична и полюсов не имеет.

Имеет место следующий факт. Пусть есть $\sum a_n z^{-n}$ и мы рассматриваем алгебру, порождённую элементами $\{a_n\}$. Предположим, что существуют b_i, для которых $[a_n,\, a_m]=(n-m)b_{n+m}$. Из этого вытекает, что образующие a_n удовлетворяют квадратичным соотношениям

$$\frac{[a_n,\, a_m]}{n-m}-\frac{[a_{n+1},\, a_{m-1}]}{n-m+2}=0.$$

Предположим, что a_i включаются в алгебру Ли гамильтоновых векторных полей на плоскости. Напомню, что её элементы — это функции $H(p, q)$; коммутатор устроен следующим образом:

$$[H_1(p, q), H_2(p, q)] = \frac{\partial H_1}{\partial p}\frac{\partial H_2}{\partial q} - \frac{\partial H_1}{\partial q}\frac{\partial H_2}{\partial p}.$$

Я буду рассматривать подпространство $\mathcal{L}$, состоящее из функций вида $p^i f(q)$, где $i \geqslant 2$, $f(q)$ — полином Лорана от q. Ясно, что $\mathcal{L}$ — алгебра Ли, она порождается своим подпространством $\{p^2 f(q)\}$.

Пусть имеется представление алгебры Ли $\mathcal{L}$ в некотором пространстве W. Пусть, далее, $a_i = p^2 q^i$. Ещё есть такие требования: $W = W_0 + W_1 + W_2 + \ldots$ и $a_i \colon W_j \to W_{i+j}$. В этом пространстве можно вычислить матричный элемент

$$F(z_1, z_2, z_3) = \langle \check{w} | a(z_1)a(z_2)a(z_3) | w \rangle.$$

Эта функция имеет вид

$$F(z_1, z_2, z_3) = \prod (z_i - z_j)^{-2} G(z_1, z_2, z_3),$$

где $G(z_1, z_2, z_3)$ — симметрический полином (Лорана) от z_1, z_2, z_3, причём $G(z, z, z) = 0$. Кроме того, G удовлетворяет ещё одному условию (см. вторую часть (2′)).

Алгебра Ли $\mathcal{L}$ своими свойствами напоминает максимальные нильпотентные подалгебры в алгебрах Каца—Муди. В частности, она может быть описана как алгебра Ли, порождённая образующими $\{a_i\}$, удовлетворяющими некоторому аналогу серровских соотношений. Эти соотношения выглядят так:

(1) пусть $i + j = k + l$, тогда $(l - k)[a_i, a_j] = (j - i)[a_k, a_l]$ — квадратичные соотношения;

(2) пусть $i + j + k = i_1 + j_1 + k_1$, тогда

$$(i_1 + j_1 - 2k_1)(j_1 - i_1) \cdot [a_k, [a_i, a_j]] = (i + j - 2k)(j - i) \cdot [a_{k_1}, [a_{i_1}, a_{j_1}]]$$

— соотношения.

Напомним, что подобным образом описывается максимальная нильпотентная подалгебра в $\mathfrak{gl}(n)$ — посредством квадратичных и кубических серровских соотношений.

В терминах матричных элементов (или, как принято говорить, «корреляционных функций») соотношения (1) и (2) переписываются так:

(1′) $\langle \check{w} | a(z_1)a(z_2) | w \rangle = (z_1 - z_2)^{-2} G(z_1, z_2)$, где $G(z_1, z_2)$ — симметрический полином Лорана;

(2′) $\langle \check{w} | a(z_1)a(z_2)a(z_3) | w \rangle = (z_1 - z_2)^2 (z_2 - z_3)^2 (z_1 - z_3)^2 F(z_1, z_2, z_3)$, где $F(z_1, z_2, z_3)$ — симметрический полином Лорана, такой что:

а) $F(t, t, t) = 0$, б) $2U'_x(t, t) - U'_y(t, t) = 0$, где $U(x, y) = \dfrac{F(x, x, y)}{(x-y)^2}$.

Универсальная обёртывающая алгебра $U(\mathcal{L})$ алгебры Ли $\mathcal{L}$ может быть описана «алгебраически» — то есть посредством образующих и соотношений (1) и (2). С другой стороны, $U(\mathcal{L})$ определяет операторную алгебру. Это и означает, что мы описываем $\mathcal{L}$ в терминах аналитических свойств функций $\langle \check{w}|a(z_1)a(z_2)\ldots a(z_n)|w\rangle$. Вертексная алгебра, отвечающая $\mathcal{L}$, может быть продеформирована (при этом деформируются соотношения (1) и (2) или (1′) и (2′)). В результате получаются так называемые W-алгебры — новый математический объект, интенсивно изучаемый в последнее время. Деформированные (1′) и (2′) — это как бы серровские соотношения для W-алгебр.

Имеется целая наука о попытках описать все (в разумном классе) вертексные алгебры, порождённые одним семейством операторов $a(z)$, и имеется очень много красивых примеров.

Проблема с вертексными алгебрами вот в чём. Если разбирать примеры, то ясно, о чём идёт речь. Но если попробовать написать систему аксиом, то получится целая книга. И такие книги существуют.

2.
Геометрия

Я уже сказал, что теория вертексных алгебр — наука алгебраическая, похожая на теорию представлений алгебр Ли. А ещё есть геометрия. Геометрия возникает из-за того, что в алгебре соотношения и всё остальное записывается в терминах генераторов $a(z)$, зависящих от переменной z. По этим алгебрам можно строить различные объекты, в которых z — точка алгебраической кривой.

Если есть кривая, то на ней можно взять любую точку и рассмотреть её окрестность (формальную) и считать, что z — элемент формальной окрестности. Все явления по этому z разложены в ряд. Поэтому если есть какая-то вертексная операторная алгебра и её представление, то это представление можно поместить в точку z, считая, что вертексная операторная алгебра выражена через z. Это похоже на стандартные конструкции теории автоморфных форм: в каждой точке алгебраической кривой сидит некоторое пространство, и эти пространства как-то друг с другом взаимодействуют.

Так можно построить некоторое пространство, являющееся аналогом пространства автоморфных форм. Я сейчас приведу некую простую алгебро-геометрическую конструкцию, дающую алгебраическое многообразие.

Пусть $\mathscr{E}$ — алгебраическая кривая, ν — двумерное расслоение на $\mathscr{E}$. Предположим, что пространство сечений $H^0(\nu)$ достаточно велико, а $H^1(\nu)=0$. Пусть M — двумерное многообразие прямых в слоях расслоения ν (линейчатая поверхность). Тогда имеется расслоение $M\to\mathscr{E}$ со слоем $\mathbb{C}P^1$. При этом на M есть линейное расслоение μ, для которого

$$H^0(\mu)|_M\cong H^0(\nu)|_{\mathscr{E}}.$$

Следующую конструкцию можно провести для любого многообразия M и линейного расслоения μ. Я буду строить коммутативную градуированную алгебру. Пусть $N=\dim H^0(\mu)$. Рассмотрим N-кратное внешнее произведение $\mu\otimes\mu\otimes\ldots\otimes\mu$ — линейное расслоение на $M\times\ldots\times M$ (N раз). В той алгебре, которая строится, степени 0 соответствует $\mathbb{C}$; степени 1 тоже $\mathbb{C}$ (единственное кососимметрическое сечение расслоения $\mu\otimes\ldots\otimes\mu$, т. е. единственное сечение внешней степени). Далее каждому n сопоставляется пространство сечений расслоения $\mu^n\otimes\ldots\otimes\mu^n$ на $M\times\ldots\times M$, которые имеют n-кратный нуль на подмногообразиях $(m_1,\ldots,m_N)$: $m_i=m_j$ (т. е. на диагоналях); кроме того, сечения должны быть кососимметрическими при нечётном n и симметрическими при чётном n.

Каждому n сопоставлено пространство S_n. Имеется произведение $S_n\otimes S_m\to S_{n+m}$ (сечения поточечно перемножаются). В результате получается алгебра $\{\bigoplus S_i\}$. Меня интересует проективный спектр этой алгебры. Этот объект очень близок к пространству модулей SL_2-расслоений на кривой $\mathscr{E}$.

Имеется замечательный способ вычисления размерности пространства S_k в терминах так называемой алгебры Верлинде. Пусть A — кольцо конечномерных представлений алгебры Ли $\mathfrak{sl}_2$, $\varepsilon_0,\varepsilon_1,\ldots$ — базис в A (ε_s отвечает $(s+1)$-мерному неприводимому представлению). Алгебра Верлинде для $\widehat{\mathfrak{sl}}_2$ (уровня k) — это фактор $A_k=A/(A\cdot\varepsilon_{k+1})$. Введём также групповую алгебру C_{2k} группы $\mathbb{Z}/2k\mathbb{Z}$, и пусть $\nu_0,\nu_1,\ldots,\nu_{2k-1}$ — естественный базис в ней: ν_0 — единица в $\mathbb{Z}/2k\mathbb{Z}$, ν_1 — образующая. Введём в алгебре $A_k\otimes C_{2k}$ $\mathbb{Z}_2$-градуировку, положив $\deg(\varepsilon_0\otimes\nu_1)=1$ и $\deg(\varepsilon_1\otimes\nu_0)=1$. Пусть $(A_k\otimes C_{2k})^0$ — подалгебра в $A_k\otimes C_{2k}$, состоящая из элементов градуировки нуль. Элемент $\varepsilon_{k+1}\otimes\nu_k$ имеет порядок 2 в $(A_k\otimes C_{2k})^0$. Положим

$$V=(A_k\otimes C_{2k})^0/(A_k\otimes C_{2k})^0(\varepsilon_k\otimes\nu_k-1).$$

V — факторалгебра $(A_k\otimes C_{2k})^0$ по идеалу, порождённому $\varepsilon_k\otimes\nu_k-1$. Базис в V — представители элементов $\varepsilon_i\otimes\nu_j$, где $i+j$ чётно и $i<k$. Размерность пространства S_k вычисляется следующим образом. Пусть

g — род кривой $\mathscr{E}$,

$$\left(\sum(\varepsilon_i \otimes \nu_j)^2\right)^g = \sum a_g(i,\ j) \cdot \varepsilon_i \otimes \nu_j.$$

Здесь $\varepsilon_i \otimes \nu_j$ — базис в V. То есть мы g-ю степень суммы квадратов базисных элементов в V разлагаем по образующим. Оказывается, что коэффициент $a_g(0, 0)$ равен размерности пространства S_k.

Этот чисто алгебро-геометрический результат получается посредством применения теории VOA. А именно, в теории VOA имеется так называемый «модулярный функтор». Этот функтор сопоставляет алгебраической кривой и операторной алгебре некоторое пространство. Наше пространство сечений S_k получается как раз таким способом (из некой вертексной алгебры, описать которую нет возможности).

Это простейший пример, когда работает следующий подход. Имеется семейство вертексных операторных алгебр (похожих на то, что я объяснял раньше), зависящих от параметра i. Каждая из этих алгебр порождает пространство матричных элементов некоторых представлений. Они появляются примерно так же, как функции $F(z_1, \ldots, z_n)$, но только теперь z имеют более сложную природу. Для этих пространств можно рассмотреть сечения. Возникают объекты, которые простым и естественным образом можно перемножать и брать проективный спектр. В результате получается алгебраическое многообразие, которое близко к пространству модулей каких-то интересных объектов. Есть предположение, что все (или во всяком случае многие) пространства модулей интересных объектов могут быть описаны такого сорта манипуляциями. Известные конструкции касаются, скорее, пространства модулей расслоений. Одно из самых интересных пространств модулей — пространство модулей инстантонов. Это — комплексное многообразие. Но те его описания, которые мне известны, не описывают его как проективный спектр чего-нибудь. Найти такого сорта описание было бы очень интересно (понять, какие вертексные операторные алгебры вовлечены в игру, можно).

Ситуация даже более интересна. Алгебраической кривой $\mathscr{E}$ можно сопоставить пространство $\mathrm{Mod}(\mathrm{SL}_2, \mathscr{E})$ (пространство модулей SL_2-расслоений на этой кривой). Функтор $\mathscr{E} \to \mathrm{Mod}(\mathrm{SL}_2, \mathscr{E})$ обладает некоторыми замечательными свойствами, которые позволяют сопоставить ему операторную алгебру. Таким образом, тут до некоторой степени есть полная обратимость. Операторных алгебр очень много, пространств модулей тоже. В этой науке наиболее интересен вопрос о взаимоотношениях между ними.

20 апреля 2000 г.

М. А. Цфасман

АЛГЕБРАИЧЕСКАЯ ГЕОМЕТРИЯ, ТЕОРИЯ ЧИСЕЛ И ИХ ПРИЛОЖЕНИЯ К ПЛОТНЫМ УПАКОВКАМ

В математике существуют периоды и области, которые вдруг начинают бурно развиваться и потом входят в сознание каждого. В теории чисел XX века таким обстоятельством является удивительная связь между числами и функциями, иными словами, между теорией чисел и алгебраической геометрией.

1.
Числовые и функциональные поля

Теория чисел начинается с рассмотрения кольца целых чисел $\mathbb{Z}$. В этом кольце есть простые числа и простые идеалы (p). Все простые идеалы главные; любой простой идеал максимален. (Нулевой идеал следует рассматривать отдельно.) Фактор $\mathbb{Z}/(p)$ конечен, это — конечное поле.

Какие ещё кольца обладают такими же свойствами? Можно, например, обратить число 2, т. е. рассмотреть кольцо $\mathbb{Z}[1/2]$. Тогда исчезает идеал (2), а все остальные идеалы остаются. Можно обратить два числа: $\mathbb{Z}[1/3, 1/7]$. Тогда исчезнут два идеала.

Такие кольца в некотором смысле меньше $\mathbb{Z}$; точнее говоря, сами они больше, но идеалов у них меньше. Кольцо $\mathbb{Z}[i]$ тоже удовлетворяет свойству, что фактор по любому простому идеалу является конечным полем. Точно так же можно рассмотреть, например, кольца $\mathbb{Z}[\sqrt{2}]$ и $\mathbb{Z}[\sqrt[10]{1}]$.

Далеко не для всех таких колец любой простой идеал будет главным. Например, это неверно для кольца $\mathbb{Z}[\sqrt{-5}]$.

Теперь можно выделить следующий тип колец. Пусть R — коммутативное кольцо с единицей и без делителей нуля (целостное кольцо). Будем также предполагать, что кольцо нётерово, т. е. любая строго возрастающая цепочка идеалов (не обязательно простых) $\mathfrak{a}_1 \subsetneq \mathfrak{a}_2 \subsetneq \dots$ конечна. Кроме того, кольцо одномерно, т. е. если есть последовательность простых идеалов $(0) \subsetneq \mathfrak{p} \subset \mathfrak{p}'$, то $\mathfrak{p} = \mathfrak{p}'$. Одномерность кольца эквивалентна тому, что любой ненулевой простой идеал максимален. Будем также предполагать, что кольцо R имеет конечный тип над $\mathbb{Z}$, т. е. порождено над $\mathbb{Z}$ конечным

числом элементов. Наконец, я потребую, чтобы фактор $R/\mathfrak{p}$ был не просто полем, а конечным полем.

Соответствующий кольцу R геометрический объект называется *спектром* кольца. Как множество спектр — все простые идеалы. На этом множестве есть топология (топология Зариского); замкнуты все конечные объединения замкнутых точек, а замкнутыми являются точки, соответствующие максимальным идеалам. Это можно представлять себе как кривую, точками которой являются максимальные идеалы. Кроме того, в стороне есть ещё точка (0), в замыкании которой лежит вся кривая (так называемая *общая точка*).

Какие ещё кольца удовлетворяют выписанным выше условиям? Например, этим условиям удовлетворяет кольцо $\mathbf{Z} = \mathbb{F}_q[t]$ — кольцо многочленов от одной переменной t над конечным полем.

Я напомню основные свойства конечных полей:

$|\mathbb{F}_q| = q = p^n$, т. е. число элементов конечного поля является степенью простого числа;

любое конечное тело является полем (т. е. коммутативно);

мультипликативная группа $\mathbb{F}_q^*$ циклическая;

для заданного q поле $\mathbb{F}_q$ единственно;

любое его подкольцо также является полем, если в кольце есть единица, причём та же самая (естественно сказать, что $\mathbb{F}_q$ — поле размерности 0).

группа Галуа алгебраического замыкания $\mathrm{Gal}(\bar{\mathbb{F}}_q/\mathbb{F}_q)$ есть $\hat{\mathbb{Z}}$ — проконечное пополнение группы целых чисел; из этого следует, что расширение данной степени единственно и всегда циклично;

алгебраическое замыкание $\bar{\mathbb{F}}_q = \bigcup\limits_{m \geqslant 1} \mathbb{F}_{q^m}$.

З а д а ч а 1. Содержит ли $\mathbb{F}_8$ подполе $\mathbb{F}_4$?

Простые идеалы в $\mathbf{Z} = \mathbb{F}_q[t]$ имеют вид $(P(t))$, где P — многочлен, неприводимый над $\mathbb{F}_q$. Если многочлен P неприводим и $\deg P = m$, то $\mathbf{Z}/(P(t)) = \mathbb{F}_{q^m}$, т. е. фактор — конечное поле.

Если присоединить две независимые переменные, то фактор уже не обязательно будет конечным. Например, $\mathbb{F}_q[t, x]/(x) = \mathbb{F}_q[t]$ — бесконечное кольцо (даже не поле). Но если взять неприводимый многочлен $Q(t, x)$ от двух переменных, то факторкольцо $\mathbb{F}_q[t, x]/(Q(t, x))$ обладает всеми теми свойствами, о которых говорилось раньше. В этом кольце не любой идеал будет главным, но такого требования и не было.

Обсудим подробнее последний пример. Уравнение $Q(t, x) = 0$ задаёт некоторую плоскую кривую. Кольцо $\mathbb{F}_q[t, x]/(Q(t, x))$ — это кольцо регулярных рациональных функций на этой кривой (функция $R(t, x)/S(t, x)$

регулярна на кривой, если S не обращается в нуль в точках кривой; заметим, что точка кривой — это точка не обязательно с координатами из $\mathbb{F}_q$, а и из любого его алгебраического расширения). Здесь мы уже дошли до геометрии. Например, кольцо $\mathbb{F}_q[t, \sqrt{t^3-1}]$ соответствует эллиптической кривой $y^2 = t^3 - 1$.

О кривых над конечным полем часто бывает удобно думать, как о кривых над полем рациональных чисел с некоторыми небольшими отличиями. Когда есть кривая с рациональными коэффициентами, для нас естественно сначала рассмотреть её комплексные точки, затем вещественные, а уже затем рациональные. Для конечного поля это соответствует рассмотрению точек над его замыканием.

Ф а к т 1. Пусть кольцо R удовлетворяет тем свойствам, которые сформулированы на с. 66; K — его поле частных. Тогда либо $\operatorname{char} K = 0$ и $[K : \mathbb{Q}] < \infty$, т. е. поле K является конечным расширением поля рациональных чисел, либо $\operatorname{char} K = p$ и $[K : \mathbb{F}_q(t)] < \infty$, тогда поле K является конечным расширением поля рациональных функций от одной переменной над полем $\mathbb{F}_q$. Более того, во втором случае существует единственное число q и единственная кривая X (гладкая и проективная) над полем $\mathbb{F}_q$, для которой $K = \mathbb{F}_q(X)$ — поле рациональных функций на X.

Для тех, кто привык к комплексной геометрии, нужно сказать следующее. Можно взять поле $\bar{\mathbb{F}}_q(X)$ и рассмотреть только те функции, которые инвариантны относительно действия группы Галуа $\operatorname{Gal}(\bar{\mathbb{F}}_q/\mathbb{F}_q)$. Равным образом, то, что на алгебраическом языке называется точкой степени m, на геометрическом языке есть орбита группы Галуа, состоящая из m точек.

О п р е д е л е н и е. Поля K, описанные в факте 1, называют *глобальными*. В первом случае поля называют *числовыми*, а во втором *функциональными*.

Кроме того, имеет место факт, который описывает подкольца в K интересующего нас типа. Чтобы его сформулировать, нужно напомнить определение целого алгебраического числа. Число $a \in K$ называют *целым*, если a — корень многочлена $x^m + a_1 x^{m-1} + \ldots + a_m$, где $a_i \in \mathbb{Z}$.

Ф а к т 2. Пусть $\mathcal{O}$ — кольцо целых в поле K, S — конечное множество простых идеалов в $\mathcal{O}$. Если обратить идеалы из S, то получим кольцо $\mathcal{O}_S$. В числовом случае кольца вида $\mathcal{O}_S$ в K — это в точности подкольца, удовлетворяющие условиям, сформулированным на с. 66.

З а д а ч а 2. Сформулируйте аналогичный результат в функциональном случае, как на алгебраическом, так и на геометрическом языке.

Аналогия между числами и функциями не только удивительна, но и очень продуктивна. По этому поводу можно приводить много разных соображений. Я ограничусь одним примером.

В анализе известна дзета-функция Римана

$$\zeta_{\mathbb{Z}}(s) = \zeta_{\mathbb{Q}}(s) = \sum_{n>0} \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}.$$

Ряд сходится при $\operatorname{Re} s > 1$. Задаваемую этим рядом функцию можно продолжить на всю плоскость; в точке 1 эта функция имеет простой полюс. Дзета-функция Римана удовлетворяет следующему функциональному уравнению. Рассмотрим функцию

$$\Lambda_{\mathbb{Q}}(s) = \pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta_{\mathbb{Q}}(s).$$

Тогда

$$\Lambda_{\mathbb{Q}}(1-s) = \Lambda_{\mathbb{Q}}(s).$$

Риман высказал гипотезу, что все нули дзета-функции, кроме тривиальных нулей в целых точках $-n$, лежат на прямой $\operatorname{Re} s = 1/2$. Эта гипотеза не доказана. Для теории чисел эта гипотеза (и её различные обобщения) чрезвычайно важна.

Для любого кольца R описанного выше типа тоже можно построить аналогичную дзета-функцию

$$\zeta_R(s) = \sum_{\mathfrak{a}} \frac{1}{N(\mathfrak{a})^s} = \prod_{\mathfrak{p}} \left(1 - \frac{1}{N(\mathfrak{p})^s}\right)^{-1}.$$

Здесь $N(\mathfrak{a})$ — норма идеала $\mathfrak{a}$, т. е. количество элементов кольца $R/\mathfrak{a}$.

Функция $\zeta_R(s)$ явно зависит от кольца R: если обратить один из идеалов, то один из множителей выпадет. А вот соответствующая функция $\Lambda_R(s)$ уже будет зависеть только от поля частных K. Эту функцию можно обозначить $\Lambda_K(s)$.

Например, для поля $\mathbb{F}_q(t)$ функция Λ выглядит очень просто:

$$\Lambda_{\mathbb{F}_q(t)}(s) = \frac{1}{(1-q^{-s})(1-q^{1-s})}.$$

Более того, в функциональном случае $\Lambda_k(s)$ всегда имеет вид

$$\frac{P_X(q^{-s})}{(1-q^{-s})(1-q^{1-s})},$$

где P_X — многочлен степени $2g$ (удвоенный род гладкой кривой X, задающей функциональное поле).

В функциональном случае дзета-функция имеет весьма простой вид, для неё гипотеза Римана доказывается. Это типичная ситуация: в функциональном случае обычно задачи решаются проще, чем в числовом.

Аналогии между числовыми полями и функциональными не всегда прямые. Прежде всего ясно, что кольцу $\mathbb{Z}$ соответствует кольцо $\mathbf{Z} = \mathbb{F}_q[t]$; для каждого q кольцо своё. Простому идеалу (p) соответствует простой идеал $\big(P(t)\big)$, где P — неприводимый многочлен. Для чисел есть функция $\operatorname{sgn}\colon \mathbb{Z} \setminus \{0\} \to \{\pm 1\}$. Её аналог $\operatorname{sgn}\colon \mathbf{Z} \setminus \{0\} \to \{\mathbb{F}_q^*\}$ устроен следующим образом: многочлену $Q(t)$ сопоставляется его старший коэффициент. В функциональном случае тоже можно определить аналоги чисел e и π; можно определить гамма-функцию. Этим очень много занимался L. Carlitz и его школа.

2.
Упаковки шаров

Упаковки шаров меня интересуют не столько сами по себе, сколько как приложение той связи между теорией чисел и геометрией, о которой было только что рассказано.

Рассмотрим в евклидовом пространстве $\mathbb{R}^N$ непересекающиеся открытые шары одинакового радиуса. Такого рода объект называют *упаковкой шаров*. Чтобы задать упаковку шаров, достаточно задать радиус шаров и множество их центров $P \subset \mathbb{R}^N$. При этом радиус даже не обязательно задавать — мы всегда будем предполагать, что он максимальный допустимый, т. е. диаметр шаров равен $d = \min\limits_{x \neq y,\ x,y \in P} \|x - y\|$.

Есть задача о том, как плотнее всего паковать шары. Эта задача фигурирует в двух видах: для произвольного множества центров и для множества центров $L \subset \mathbb{R}^N$, которое является решёткой (дискретной аддитивной подгруппой в $\mathbb{R}^N$ полного ранга).

Плотность упаковки $\Delta(P)$ определяется как предел

$$\lim \frac{\text{объём шаров в большой коробке}}{\text{объём коробки}}.$$

Если такого предела не существует, то можно рассмотреть верхний (или нижний) предел.

В случае решётки фундаментальный параллелепипед содержит несколько частей разных шаров, из которых параллельными переносами складывается в точности один шар. Поэтому

$$\Delta(L) = \frac{V_N\, d^N}{2^N \operatorname{covol}(L)},$$

где V_N — объём шара радиуса 1, $\operatorname{covol}(L)$ — объём фундаментального параллелепипеда решётки L (кообъём решётки).

В малых размерностях плотнейшими решётками являются решётки, связанные с системами корней: A_1, A_2, A_3, D_4, D_5, E_6, E_7, E_8. Однако бесконечные семейства A_n и D_n (а также B_n и C_n, которые к ним сводятся) асимптотически плохи. Для них Δ ведёт себя как n^{-n}. Не исключено, что у них есть интересные подрешётки, но мы их не знаем.

В основном меня будет интересовать случай, когда N велико. Об асимптотике говорить проще. Если дано семейство решёток $\{L_N \subset \mathbb{R}^N\}$, $N \to \infty$, то для него можно определить предельную плотность. При этом нельзя просто взять предел плотностей, потому что такой предел заведомо равен нулю. Сначала нужно угадать асимптотику зависимости максимальной плотности от размерности. А именно, для каждой решётки $L \subset \mathbb{R}^N$ определим число $\lambda(L)$ формулой

$$\Delta(L) = 2^{-\lambda(L)N}$$

и назовём его *логарифмической* плотностью. Теперь предельную логарифмическую плотность семейства можно определить так:

$$\lambda(\{L_N\}) = \limsup \lambda(L_N).$$

Семейство решёток $\{L_N\}$ называют *асимптотически хорошим*, если $\lambda(\{L_N\}) < \infty$.

Такое определение оправдано по следующим причинам. Известно, что для любого семейства $\{L_N\}$ выполняется неравенство $\lambda(\{L_N\}) \geqslant 0{,}5990\ldots$ Кроме того, известно, что существуют семейства $\{L_N\}$, для которых $\lambda(\{L_N\}) = 1$ (теорема Минковского). Эта теорема доказывается методом интегрирования по всем решёткам. Средняя асимптотическая плотность равна 1. Более того, в некотором смысле почти любое семейство решёток имеет логарифмическую плотность 1, т. е. 2^{-N} — типичная плотность решётки.

Вопрос о том, существуют ли семейства $\{L_N\}$, для которых $\lambda(\{L_N\}) < 1$, до сих пор открыт.

Придумать семейство решёток с конечной логарифмической плотностью не так уж просто. Те примеры, которые сразу приходят в голову, дают семейства решёток с бесконечной логарифмической плотностью. Первые в том или ином смысле явные примеры семейств решёток с конечной логарифмической плотностью были построены методами алгебраической геометрии и алгебраической теории чисел. Наиболее важны для этих целей конструкции из алгебраической теории чисел, предложенные, с совершенно иными целями, Дирихле и Дедекиндом.

Пусть K — числовое поле, причём $[K : \mathbb{Q}] = n = s + 2t$, где числа s и t определяются следующим образом. Число s равно количеству вложений

$\sigma_i\colon K\to\mathbb{R}$, а число t равно количеству пар комплексно сопряжённых вложений $\tau_j\colon K\to\mathbb{C}$ (образы которых не лежат в $\mathbb{R}$). Нетрудно доказать, что размерность K над $\mathbb{Q}$ действительно равна $s+2t$.

Рассмотрим теперь все вложения одновременно: $K\subset\mathbb{R}^s\times\mathbb{C}^t$. Отождествим $\mathbb{C}$ с $\mathbb{R}^2$, переводя $a+bi$ в $(a+b,\,a-b)$. Тогда $\mathbb{R}^s\times\mathbb{C}^t$ отождествляется с $\mathbb{R}^n$. Поле K вложено в $\mathbb{R}^n$ как аддитивная подгруппа, $\sigma\colon K\to\mathbb{R}^n$. Оказывается, что образ кольца целых $\mathcal{O}_K$ в $\mathbb{R}^n$ — решётка полного ранга. Чтобы оценить плотность этой решётки, нужно знать кообъём и минимальное расстояние между её точками. Для решётки минимальное расстояние — это минимальная длина вектора.

Ф а к т 3. $\operatorname{covol}\mathcal{O}_K=\sqrt{|D_K|}$, где $|D_K|$ — абсолютная величина дискриминанта поля K. (Для нас это — определение абсолютной величины дискриминанта.)

Дискриминант — это некий очень важный параметр поля. Позже я скажу, в каких формулах он встречается.

Теперь поговорим о минимальной норме. Для простоты будем предполагать, что $t=0$. Пусть $a\in\mathcal{O}_K$. Тогда

$$|\sigma(a)|=\sqrt{\sum\sigma_i(a)^2}\geqslant\sqrt{n\Big(\prod\sigma_i(a)\Big)^{2/n}}\geqslant\sqrt{n}.$$

Последнее неравенство следует из того, что число $\prod\sigma_i(a)$ целое, поскольку a — целое алгебраическое. Действительно, имеет место следующий факт.

Ф а к т 4. Если a — корень многочлена $x^n+a_1x^{n-1}+\ldots+a_n$ с целыми коэффициентами, то

$$(-1)^n a_n=N_{K/\mathbb{Q}}(a)=\prod\sigma_i(a)\prod(\tau_j(a)\bar{\tau}_j(a)).$$

Полученная оценка минимальной нормы точная; она достигается для целого числа 1. Та же самая оценка верна и точна для любых полей (проверьте!). Векторы минимальной длины суть образы корней из единицы в поле K.

Итак, плотность решётки оценена через некоторые числовые инварианты поля (дискриминант и размерность над $\mathbb{Q}$). Теперь возникает задача построения семейств полей, для которых эта плотность велика. Пока я про эту задачу говорить не буду, потому что там используется существенно более тонкая теория чисел.

В теории алгебраических чисел есть фундаментальная теорема Дирихле о единицах; она описывает все обратимые элементы в кольце целых для любого конечномерного поля алгебраических чисел. Она доказывается с помощью рассмотрения некоторой другой решётки. Для построения

предыдущей решётки использовалась аддитивная структура поля. Теперь нам понадобится мультипликативная структура. Чтобы перевести мультипликативную структуру в аддитивную, нужно взять логарифм. А именно, рассмотрим отображение $K^* \to \mathbb{R}^{s+t}$, заданное формулой

$$a \mapsto (\ln |\sigma_1(a)|, \ldots, \ln(\tau_1(a)\bar{\tau}_1(a)), \ldots).$$

Чтобы получить решётку, ограничим это отображение на $\mathcal{O}_K^*$ — множество единиц поля K. (Число $a \in K$ называют *единицей*, если $a \in \mathcal{O}_K$ и $a^{-1} \in \mathcal{O}_K$.) Оказывается, что образ множества $\mathcal{O}_K^*$ лежит в гиперплоскости $H = \{\sum x_i = 0\}$ и является в этой гиперплоскости решёткой полного ранга.

Нам нужно ещё одно важное теоретико-числовое понятие — *регулятор* поля K. Пусть L — решётка, которая только что была построена. Тогда регулятор R_K определяется формулой

$$\operatorname{covol}(L) = \sqrt{s+t}\, R_K.$$

Отображение $K^* \to \mathbb{R}^{s+t}$ имеет ядро, которое состоит их корней из единицы, лежащих в поле K.

Длину минимального вектора сейчас я считать не буду, потому что это сложно. Кроме того, я хочу рассказать несколько иную конструкцию, позволяющую это сделать довольно просто.

В этом месте нельзя не сделать отступление и не рассказать о следующем поразительном факте. Оказывается, что все встретившиеся нам числа связаны между собой следующим образом:

$$\operatorname*{Res}_{s=1} \zeta_K(s) = \frac{R_K h_K}{w_K \sqrt{|D_K|}}.$$

Здесь h_K — число классов идеалов поля K, которое определяется следующим образом. Идеалы в кольце целых поля K образуют мультипликативную группу. Фактор этой группы по подгруппе, состоящей из главных идеалов, конечен. Число элементов факторгруппы — это и есть h_K. Далее, w_K — число корней из единицы, лежащих в K.

Эта формула показывает, что значительная часть нужной нам информации содержится в дзета-функции. Именно работа с дзета-функцией позволяет всё посчитать.

Теперь нужно вспомнить то, с чего я начинал: алгебраические числа и алгебраические функции — это одно и то же. Мне будет проще говорить на геометрическом языке. Вам придётся мне поверить, что то, что я называю прямыми аналогиями, действительно таковыми является.

Рассмотрим гладкую проективную неприводимую кривую X над полем $\mathbb{F}_q$. Пусть n — количество точек кривой, определённых над основным

полем. Мы будем предполагать, что $n > 0$. Действовать по прямой аналогии здесь плохо. По прямой аналогии нужно было бы сказать, что K — рациональные функции, K^* — ненулевые рациональные функции, а нас интересуют целые рациональные функции, т. е. не имеющие полюсов. Но таких функций немного — только константы. Здесь нужно вспомнить, что некоторые идеалы можно обратить.

Пусть $S = X(\mathbb{F}_q)$ — множество точек кривой, определённых над основным полем. Рассмотрим кольцо $\mathcal{O}^*_{K,S}$, состоящее из рациональных функций на X, все нули и полюса которых лежат в S. При этом нужно помнить, что если расширить поле $\mathbb{F}_q$, то на кривой появятся точки, отличные от S. Аналог этой ситуации таков: у кривой может быть мало вещественных точек, но много комплексных.

Пусть $S = \{P_1, \ldots, P_n\}$. Зададим отображение

$$\mathcal{O}^*_{K,S} \to \mathbb{Z}^n \subset \mathbb{R}^n,$$

сопоставив функции f набор $(a_1, \ldots, a_n)$ коэффициентов её *дивизора* $\sum a_i P_i$, где

$$a_i = \begin{cases} k, & \text{если в точке } P_i \text{ функция } f \text{ имеет нуль кратности } k; \\ -k, & \text{если в точке } P_i \text{ функция } f \text{ имеет полюс кратности } k; \\ 0, & \text{если в точке } P_i \text{ у функции } f \text{ нет ни нуля, ни полюса.} \end{cases}$$

Как и в комплексном случае, выполняется равенство $\sum a_i = 0$, т. е. сумма кратностей нулей равна сумме кратностей полюсов. Это означает, что образ функции f лежит в гиперплоскости $H = \{\sum x_i = 0\}$. В числовом случае в образе отображения лежали только корни из единицы, и теперь в образе отображения лежат только корни из единицы, поскольку функции без нулей и полюсов — ненулевые константы.

Образ функции — это главный идеал; дивизоры и идеалы это более или менее одно и то же. Кообъём covol выражается через кообъём пересечения целочисленной решётки с гиперплоскостью H и через индекс нашей решётки в этой целочисленной решётке. (Решётки являются группами; индекс одной решётки в другой — это то же самое, что индекс подгруппы в группе.) Индекс не превосходит числа классов идеалов, поэтому

$$\operatorname{covol} \leqslant h_K \operatorname{covol}(\mathbb{Z}^n \cap H) = h_K \sqrt{n}.$$

Из алгебраической геометрии известно, что h_K равно числу точек на $J_X(\mathbb{F}_q)$ (якобиане кривой, определённом над основным полем). Якобиан — это абелево многообразие (тор), так сказать, комплексной размерности g. Точек, определённых над конечным полем, на нём конечное число.

Мы оценили кообъём. Оценим теперь длину минимального вектора. Нам нужно оценить снизу число $\sqrt{\sum a_i^2}$. Число a_i целое, поэтому $a_i^2 \geqslant |a_i|$. Следовательно,

$$\sqrt{\sum a_i^2} \geqslant \sqrt{\sum |a_i|} = \sqrt{2 \sum_{a_i>0} |a_i|} = \sqrt{2 \deg f}.$$

Равенство $\sum\limits_{a_i>0} |a_i| = \deg f$ можно объяснить так. Число $\sum\limits_{a_i>0} |a_i|$ равно количеству прообразов нуля при отображении f, посчитанному с учётом кратностей. Это число равно степени отображения f.

Мы приходим к задаче об оценке минимально возможной степени функции f. Прямую на прямую можно отобразить со степенью 1. Для любой кривой рода $g \geqslant 1$ требуется более высокая степень. Минимальную такую степень нам и нужно оценить.

Эту оценку можно получить следующим образом. На $\mathbb{P}^1$ есть точки с координатами из основного поля; ясно, что $|\mathbb{P}^1(\mathbb{F}_q)| = q+1$. Предположим, что $|X(\mathbb{F}_q)| = N$. Функция f определена над $\mathbb{F}_q$, поэтому она отображает $X(\mathbb{F}_q)$ в $\mathbb{P}^1(\mathbb{F}_q)$. Следовательно, $N \leqslant (q+1) \deg f$. Оценка для степени получена. Она, в свою очередь, даёт оценку минимальной длины вектора.

По полю K (по кривой X) построена $(n-1)$-мерная решётка L_K (возможно, не единственная). Для этой решётки получена оценка плотности $\Delta(L_K)$ снизу. Эта оценка включает некоторые параметры поля K.

Дальше начинается самое интересное — построение примеров решёток с большой плотностью. Сначала нужно понять, чего мы хотим от кривой (чтобы плотность соответствующей решётки была велика). Оказывается, что мы хотим, чтобы число $|X(\mathbb{F}_q)|$ было возможно больше при заданном роде. Пусть $|X(\mathbb{F}_q)| = n_X$. Известно следующее неравенство

$$n_X \leqslant q + 1 + 2g_X\sqrt{q}.$$

Оно вытекает из работ А. Вейля, связанных с дзета-функцией. В. Г. Дринфельд и С. Г. Влэдуц доказали более сильную асимптотическую оценку:

$$\lim_{g_X \to \infty} \frac{n_X}{g_X} \leqslant \sqrt{q} - 1.$$

Факт очень неожиданный, но его доказательство весьма просто. Более того, если q — квадрат, то есть примеры семейств кривых, для которых

$$\lim_{g_X \to \infty} \frac{n_X}{g_X} = \sqrt{q} - 1.$$

Эти кривые берутся уже из более тонкой алгебраической геометрии. Задача об оценке числа точек кривой трудна потому, что точки на кривой

неотличимы друг от друга. Чтобы точки стало легче считать, нужно каждой точке придать нечто индивидуальное. Это можно сделать, рассматривая не просто кривые, а модулярные кривые. Идея модулярных кривых, и вообще модулярных многообразий, заключается в следующем. Пусть E — кривая рода 1. Все кривые рода 1 параметризуются прямой $\mathbb{P}^1$. Можно рассматривать пары, состоящие из кривой E плюс некоторые дополнительные данные. Тогда возникает накрытие прямой $\mathbb{P}^1$ — множества кривых E — пространством параметров, точками которого служат эллиптические кривые с дополнительными данными, которые называют *структурой уровня*. Так получается много разных кривых, род которых при изменении структуры уровня может стремиться к бесконечности. В случае конечной характеристики среди всех эллиптических кривых выделяются *суперсингулярные кривые*. Оказывается, что для любой естественной (не буду уточнять, что это такое) структуры уровня все точки, которые лежат над суперсингулярными эллиптическими кривыми, определены хотя и не над самим исходным полем, но над его расширением степени 2. А именно, если исходное поле $\mathbb{F}_{\sqrt{q}}$, то все точки модулярных кривых будут определены над $\mathbb{F}_q$. Оказывается, что этих точек будет очень много. Таков один из способов построения кривых с большим количеством $\mathbb{F}_q$-точек.

В числовом случае этот способ не действует. Одна из важных нерешённых задач в этой области связана с построением башен полей, которые в каком-то смысле были бы аналогичны модулярным кривым.

В числовом случае приходится обходиться другими средствами — *теорией полей классов*. В теории полей классов строят так называемые *башни Голода—Шафаревича*. Это название связано с тем, что Е. С. Голод и И. Р. Шафаревич разрешили в отрицательном смысле проблему Гильберта, построив бесконечную последовательность неразветвлённых абелевых расширений

$$K_0 \subset K_1 \subset K_2 \subset \dots$$

Например, для $K_0 = \mathbb{Q}(\cos 2\pi/11, \sqrt{-46})$ такая башня даёт семейство решёток с $\lambda(\{L_N\}) \approx 2{,}22$. Это число больше 1, но этот мой результат не так уж плох. Когда начинали заниматься построением плотных упаковок, то без использования алгебраической геометрии и теории чисел семейства решёток с конечным λ построить вообще не удалось, а для нерешётчатых упаковок было построено семейство с $\lambda = 6$.

В геометрическом случае, используя модулярные кривые, можно добиться лучших результатов. Там есть два типа конструкций. Одна — буквально та, о которой я говорил. А в другой рассматривается некоторая её подрешётка, связанная с дивизором. В первом случае нужно взять $q = 9$,

а во втором $q = 47^2$. Соответственно, получим $\lambda \approx 1{,}87$ и $\lambda \approx 1{,}39$. (Эти результаты получены мной совместно с М. Ю. Розенблюмом.)

В математике известно много других решёток. Например, если взять эллиптическую кривую над числовым или функциональным полем, то её целые точки образуют решётку. Элкис и Шиода посчитали плотность возникающих решёток. Оказалось, что асимптотически эти решётки плохи, но поразительно хороши в размерностях между 100 и 1000. Многие табличные рекорды плотности оказались таким образом побиты.

Здесь собрались представители различных областей математики. Мой призыв: в вашей области наверняка встречаются решётки — не поленитесь проверить их на плотность, может быть, вы найдёте новые интересные семейства.

4 мая 2000 г.

В. Я. Иврий

ВСЁ НАЧАЛОСЬ С ВЕЙЛЯ

Я хочу рассказать об одной очень старой задаче. Я назвал лекцию «Всё началось с Вейля». На самом деле родоначальником был даже не Г. Вейль, а Дебай. В начале века он задал вопрос о том, как распределены частоты у прямоугольной пластинки.

С прямоугольной пластинкой всё довольно просто. Мы рассматриваем уравнение $-\Delta u = \lambda_j u$; здесь λ_j — собственные значения. Граничные условия $u|_{\partial X} = 0$ (Дирихле) или $\frac{\partial u}{\partial n}\Big|_{\partial X} = 0$ (Нейман). Прямоугольник — одна из немногих фигур, для которых эта задача допускает разделение переменных и ответ известен (другие такие фигуры — прямоугольный параллелепипед, шар и эллипс). Для прямоугольника собственные функции для задачи Дирихле имеют вид

$$u_{mn}(x,\, y) = \sin\, \frac{\pi m x}{a}\, \sin\, \frac{\pi n y}{b}$$

(для задачи Неймана синусы заменяются на косинусы). Собственные значения имеют вид

$$\lambda_{mn} = \frac{\pi^2 m^2}{a^2} + \frac{\pi^2 n^2}{b^2}.$$

Чтобы определить, как распределены собственные значения, можно применить считающую функцию

$$N(\lambda) = \#\{\lambda_j < \lambda\} = \#\Big\{(m,\, n)\colon \frac{\pi^2 m^2}{a^2} + \frac{\pi^2 n^2}{b^2} < \lambda\Big\}.$$

Таким образом, $N(\lambda)$ — количество целочисленных точек в четвертинке эллипса. Нужно оценить $N(\lambda)$ при $\lambda \to \infty$. Это уже задача теории чисел. Я не хочу разбирать её во всей глубине; известен довольно точный результат. Для нас важно то, что это число примерно равно площади. Поэтому Дебай сказал, что $N(\lambda) \sim$ площадь. Площадь четвертинки эллипса равна $\frac{\pi}{4}\left(\frac{a}{\pi}\right)\left(\frac{b}{\pi}\right)\lambda = \frac{1}{4\pi} \cdot (\text{площадь прямоугольника}) \cdot \lambda$.

Дебай высказал предположение, что в d-мерном случае для всех областей $N(\lambda)$ примерно равно $c(\mathrm{Vol}\, X)\lambda^{d/2}$.

Первый математический шаг был сделан Г. Вейлем в 1911 г. Он доказал, что для краевых условий Дирихле (это очень существенно!) предположение Дебая верно для любой ограниченной области. Это была совершенно замечательная работа. Я хочу рассказать метод. Он очень прост.

Возьмём от Дебая подсчёт для прямоугольника — для условий Дирихле и для условий Неймана. Рассмотрим квадратичную форму

$$Q_\lambda(u) = \int (|\nabla u|^2 - \lambda |u|^2)\, ds$$

(это почти форма Дирихле). Можно показать, что

$$N(\lambda) = \max_{\substack{\mathscr{L} \text{ — отрицательное} \\ \text{подпространство } Q_\lambda(u)}} \dim \mathscr{L}$$

(отрицательное подпространство квадратичной формы — это пространство, на котором форма отрицательна). Точнее говоря, для Неймана никаких дополнительных условий не требуется: краевое условие Неймана для оператора Лапласа в случае квадратичных форм автоматически получается из вариационных соображений. А для условий Дирихле нужно дополнительно потребовать, чтобы выполнялось равенство $u|_{\partial X} = 0$.

Рассмотрим теперь произвольную область. Мы интересуемся максимальной размерностью пространства $\mathscr{L}$. На границе функции должны обращаться в нуль, поэтому можно считать, что и за границей области функции тоже обращаются в нуль. Рассмотрим, далее, покрытие области маленькими прямоугольниками (или d-мерными параллелепипедами).

Немного изменим нашу задачу. Изначально есть система функций на прямоугольниках и на эти функции есть два условия: за границей области функции обращаются в нуль и функции на границах прямоугольников должны сшиваться в непрерывную функцию. Первая модификация задачи такова. Давайте потребуем, чтобы на границах прямоугольников функции обращались в нуль, и на всех прямоугольниках, не лежащих целиком внутри области, функции тоже обращались бы в нуль. Это приводит к увеличению числа требований. Тем самым размерность подпространств уменьшается. Значит, считающая функция для новой задачи может только увеличится. Пусть $N_1(\lambda)$ — новая считающая функция. Тогда

$$\begin{aligned} N(\lambda) \geqslant N_1(\lambda) &= \sum N_{D_i}(\lambda) = \\ &= c\lambda^{d/2} \sum \operatorname{Vol} X_i + o(\lambda^{d/2}) \geqslant \\ &\geqslant c\lambda^{d/2}(\operatorname{Vol} X - \varepsilon) + o(\lambda^{d/2}); \end{aligned}$$

здесь $N_{D_i}(\lambda)$ — считающая функция для задачи Дирихле в i-м прямоугольнике (в каждом маленьком прямоугольнике у нас есть задача Дирихле).

Чтобы получить оценку числа $N(\lambda)$ сверху, рассмотрим вторую модификацию задачи. Будем предполагать лишь, что в нуль обращаются функции на прямоугольниках, вообще не пересекающихся с данной областью. Никакой согласованности функций на границах прямоугольников не предполагается. При этом количество функций только увеличивается. Краевые условия Неймана для квадратичной формы выполняются автоматически, поэтому

$$N(\lambda) \leqslant \sum N_{N_i}(\lambda) \leqslant c\lambda^{d/2}(\mathrm{Vol}\ X + \varepsilon) + o(\lambda^{d/2}).$$

Это верно для любого $\varepsilon > 0$. Поэтому в результате получаем $N(\lambda) = c\lambda^{d/2}\ \mathrm{Vol}\ X + o(\lambda^{d/2})$; единственное условие — мера границы области равна нулю.

Эта теорема была доказана только для условий Дирихле. Если взять задачу Неймана и не накладывать никаких условий на гладкость границы, то можно получить даже непрерывный спектр. Тем самым, никакой оценки получить нельзя.

Эти вариационные методы в дальнейшем бурно развивались и обобщались. С моей точки зрения это самые общие методы. Но проблема в том, что точных результатов они дать не в состоянии.

После того как Вейль получил это доказательство, он вернулся к той самой задаче, с которой начал Дебай. Вейль заметил, что для прямоугольника можно получить гораздо более точную оценку. Во-первых, погрешность порядка площади полоски ширины примерно 1 вблизи границы:

$$N(\lambda) = c\lambda^{d/2}\ \mathrm{Vol}\ X + O(\lambda^{(d-1)/2}). \tag{1}$$

На самом деле можно посмотреть более аккуратно и заменить O большое на o малое:

$$N(\lambda) = c\lambda^{d/2}\ \mathrm{Vol}\ X + o(\lambda^{(d-1)/2}).$$

Но нужно помнить, что для условий Дирихле или Неймана это не совсем так, потому что либо считаются, либо не считаются целочисленные точки, лежащие на осях эллипса. Поэтому получается

$$N(\lambda) = c\lambda^{d/2}\ \mathrm{Vol}\ X \mp c'\lambda^{(d-1)/2}\ \mathrm{Vol}(\partial X) + o(\lambda^{(d-1)/2}) \tag{2}$$

(минус для Дирихле, плюс для Неймана).

Вообще, это интересная задача. Берём какую-либо область, раздуваем её и считаем число целочисленных точек. Если область — эллипсоид, то оценки получены очень точные. Случай эллипса существенно сложнее, чем случай эллипсоида размерности $\geqslant 3$. С другой стороны, если область не слишком выпуклая (например, раздувается прямоугольник), то оценки с

o малым быть не может, потому что в какой-то момент сторона прямоугольника проходит через очень много точек.

Вейль высказал гипотезу, что (2) имеет место для произвольных областей.

В 1924 г. Курант, используя метод Вейля, доказал, что для областей с гладкой границей

$$N(\lambda) = c_0 \lambda^{d/2} \operatorname{Vol} X + O(\lambda^{(d-1)/2} \ln \lambda).$$

От гипотезы Вейля это отличается тем, что стоит O большое, а не o малое, и ещё стоит логарифм.

Затем теория развивалась бурно: были получены многочисленные обобщения. Например, можно было рассматривать оператор Шрёдингера с потенциалом, растущим на бесконечности; можно было рассматривать более общие операторы — в ограниченных областях и неограниченных. Но для конкретно этой задачи после Куранта прогресса не наблюдалось в течение многих лет. Первый шаг после Куранта сделал в 1968 г. Хёрмандер. Для многообразий без края (и не для операторов Лапласа, а для операторов Лапласа—Бельтрами) он получил оценку с $O(\lambda^{(d-1)/2})$.

Следующий шаг сделали Дюйстермаат и Гийемин примерно в 1975 г. Они тоже рассматривали многообразия без края и получили асимптотику

$$N(\lambda) = c_0 \lambda^{d/2} \operatorname{Vol} X + o(\lambda^{(d-1)/2}). \tag{3}$$

Но им потребовалось некоторое геометрическое условие. Дело в том, что если мы не потребуем никакого условия, то результат будет неверен. Контрпример — сфера S^d. Для простоты рассмотрим S^2. Собственными значениями оператора Лапласа—Бельтрами на S^2 являются числа вида $n(n+1)/2$; кратность такого собственного значения равна $2n+1$, т. е. кратность собственного значения λ имеет порядок $\lambda^{(d-1)/2}$. Когда λ проскакивает это чудовищно вырожденное собственное значение, формула (3) верна быть не может, потому что изменение $N(\lambda)$ имеет порядок $\lambda^{(d-1)/2}$ и правая часть не может компенсировать такое большое изменение.

Геометрическое условие, о котором идёт речь, таково. Рассмотрим расслоение единичных сфер S^*X. На нём задан геодезический поток Φ_t. Он сохраняет меру. Условие состоит в том, что мера множества периодических траекторий равна нулю.

На сфере геодезические замкнуты, поэтому мера множества периодических траекторий не равна нулю. Для сферы условие не выполняется.

Помимо сферы есть, например, поверхности Золля, на которых все траектории периодические. Эти поверхности являются поверхностями

вращения. На них все траектории, за исключением одной, имеют один и тот же период, а исключительная траектория имеет вдвое меньший период.

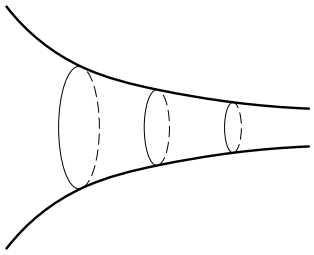
Откуда в этой задаче появились периодические траектории, я расскажу немного позже. Сначала расскажу о результатах.

Следующий шаг сделал Роберт Сили примерно в 1978 г. Он рассматривал только трёхмерный случай, но это несущественно, и только собственные числа оператора Лапласа, но это тоже несущественно (его результаты были сразу же обобщены). Сили доказал, что

$$N(\lambda) = c_0 \lambda^{d/2} \operatorname{Vol} X + O(\lambda^{(d-1)/2}), \quad \partial X \neq \varnothing. \tag{4}$$

В это время М. А. Шубин и Б. М. Левитан предложили мне заняться доказательством гипотезы Вейля. Я это сделал в 1979 г. Было доказано то, что уже писалось. Но нужно было какое-то условие вроде периодических геодезических. Чем заменяются периодические геодезические на областях с границей? Пусть геодезическая попадает на границу и отражается по закону геометрической оптики (угол падения равен углу отражения). Что будет, если геодезическая касается границы? Такие бильярды могут вести себя очень плохо. Но мера плохих точек кокасательного пространства (плохих бильярдов) равна нулю. Траектории хороших бильярдов касаются границы трансверсально и не устраивают скачков.

Таким образом, в касательном расслоении сфер существует множество полной меры, на котором задан геодезический поток. Условие состоит в том, что множество периодических траекторий имеет меру нуль.



Р и с. 1. Касп

Утверждение верно как для условий Дирихле, так и для условий Неймана.

Что будет, если область неограниченная? Ответ простой. Если мы говорим о граничной задаче Дирихле, то для того, чтобы результат (полная асимптотика со вторым членом) сохранился, нужно чтобы область имела конечный объём и конечную боковую поверхность плюс ещё некоторая регулярность на бесконечности. Например, касп (рис. 1) вполне возможен (при условии, что он имеет конечный объём и конечную боковую поверхность).

Для граничной задачи Неймана ситуация совсем другая. Например, если область имеет касп, уходящий на бесконечность, то, вообще говоря, возникает непрерывный спектр, за исключением того случая, когда этот касп очень тонкий (убывает быстрее экспоненты). Для тонкого каспа результат сохраняется.

В прошлом году я получил ещё следующий результат (в некотором смысле с подачи молодого польского математика Леха Зелинского). Если область ограниченная и коэффициенты оператора (метрика) и граница области имеют гладкость C^σ, $\sigma > 1$, т. е.

$$|(\nabla f)(x) - (\nabla f)(y)| \leqslant C|x-y|^{\sigma-1},$$

то асимптотика (2) имеет место. Первоначально это результат был доказан для областей с бесконечно гладкой границей.

Теперь я расскажу о методе. Рассмотрим семейство операторов $\cos\sqrt{-\Delta}\,t = U(t)$, зависящее от параметра t. Поставим задачу

$$\begin{cases} U''_{tt} - \Delta U = 0, \\ U|_{t=0} = I, \\ U_t|_{t=0} = 0. \end{cases}$$

Если $u(x, y, t)$ — ядро Шварца оператора $U(t)$ (любой оператор в обобщённом смысле может быть записан в виде $Uf = \int u(x, y, t) f(y)\, dt$), то мы имеем просто решение волнового уравнения с соответствующими краевыми условиями на границе и соответствующими начальными условиями.

Задача о построении решения $U(t)$ — задача на уравнения с частными производными. С другой стороны, можно написать по-другому: $U(t) = \int \cos \lambda t\, d_\lambda E(\lambda^2)$, где E — спектральный проектор (мы имеем самосопряжённый оператор, у него есть спектральный проектор). Теперь запишем след: $\int u(x, x, t)\, dx = \operatorname{Tr} U(t) = \int \cos \lambda t\, d_\lambda N(\lambda^2)$; здесь N — след спектрального проектора. Но след спектрального проектора — это не что иное, как количество собственных значений, не превосходящих λ^2. Таким образом, след оператора $U(t)$ связан с тем, что мы хотим найти косинус-преобразованием Фурье. Поэтому если бы мы знали $U(t)$ для всех t, то задача была бы тривиальной: мы бы просто взяли обратное преобразование Фурье. Однако ситуация более сложная.

Вопрос первый: как построить $U(t)$? Хёрмандер для этой цели использовал так называемый интегральный оператор Фурье. Для конечных времён пропагатор $U(t)$ — некоторый осциллирующий интеграл. Эта конструкция обладает, однако, двумя недостатками. Во-первых, она даёт ответ не точно, а по модулю бесконечно гладкой функции. Во-вторых, бесконечно гладкая функция на больших интервалах может становиться очень большой. Поэтому мы знаем след оператора $U(t)$ только на конечных интервалах. Из этого следует, что функцию $N(\lambda)$ можно восстановить только с точностью до $O(\lambda^{(d-1)/2})$. Это была конструкция Хёрмандера.

Дюйстермаат и Гийемин сделали следующее. Если мы знаем след оператора $U(t)$ на очень больших интервалах, то мы можем более точно

оценить функцию $N(\lambda)$; чем больше интервал, тем более точная оценка получается. Это, в общем-то, правда. Если след известен на интервале длины T, то оценка остатка будет $(c/T)\lambda^{(d-1)/2}$ (число $\lambda^{(d-1)/2}$ связано с ростом некоторых вспомогательных констант). Но только это полуправда: можно привести контрпример. На сфере оценка остатка не зависит от T.

Важно вот что: функция $\operatorname{Tr} U(t)$ гладкая на маленьком интервале за исключением нуля, т. е. особенность в нуле изолированная. Что же касается других особенностей, то есть замечательный результат, согласно которому сингулярности $\operatorname{Tr} U(t)$ содержатся во множестве всех периодов (эти множества совпадают не всегда, но почти всегда).

Эту теорему можно доказать следующим образом. Есть понятие волнового фронта. Волновой фронт — это обобщение понятия сингулярного носителя. Давайте рассмотрим функцию $u(x)$. Будем говорить, что эта функция *пренебрежима* на некотором множестве, если она на этом множестве бесконечно гладкая. Возьмём точку $\bar{x}$, умножим её на шапочку и сделаем преобразование Фурье. Это преобразованием Фурье может плохо вести себя на бесконечности (тогда функция негладкая), может быстро убывать. А может быть и нечто промежуточное, когда функция быстро убывает в неких конусах.

Пусть есть направление $\bar{\xi}$, есть конус, содержащий направление $\bar{\xi}$, и в этом конусе после срезки преобразование Фурье убывает быстрее любой степени. Тогда мы говорим, что функция в окрестности точки $(\bar{x}, \bar{\xi})$ *бесконечно гладкая*, а объединение всех точек, где функция не бесконечно гладкая, называют *волновым фронтом*. Волновой фронт объясняет, где функция негладкая и куда эта негладкость направлена. Например, дельта-функция негладка в нуле и негладкость направлена во все стороны. Дельта-функция на поверхности негладка на этой поверхности, но её негладкость направлена только по нормали к поверхности; по всем другим направлениям функция гладкая. Можно построить функцию, негладкую только в одной точке и только в одном направлении.

Если мы находимся внутри области, то волновыми фронтами распространение особенностей описывается очень хорошо. А именно, особенности распространяются вдоль гамильтоновых потоков; в данном случае — вдоль геодезических потоков. Скажем это подробнее. Волновой фронт лежит в фазовом пространстве (кокасательном расслоении). У каждого оператора есть символ $a(x, \xi)$; для данного класса операторов символ — квадратичная форма, которая связана с оператором Лапласа—Бельтрами. Этот оператор порождает гамильтонов поток. Волновые фронты распространяются вдоль гамильтонова потока.

Возьмём произвольную функцию f и будем её использовать в качестве начального условия (второе начальное условие, например, нуль) для нашей функции. Вопрос, где лежит волновой фронт решения. Нужно рассмотреть соответствующий гамильтониан и написать траектории. Мы берём волновой фронт f, находим τ из уравнения $\tau^2 - a(x, \xi) = 0$ и сдвигаемся по времени t вдоль этого гамильтониана. Мы получим волновой фронт решения уравнения.

Доказать этот результат можно многими разными способами. Первый способ — рассмотрение осцилляционных интегралов, которыми задаётся волновой фронт. Но я предпочитаю другой метод, который, с моей точки зрения, удобнее. Известны так называемые псевдодифференциальные операторы. Псевдодифференциальные операторы — это результат квантования символов, т. е. результат некоторой операции перехода, скажем, от классической механики к квантовой. Если есть хорошая функция от двух переменных x и ξ, то ей ставится в соответствие некий оператор, который называется квантованием. Это и есть псевдодифференциальный оператор. При такой операции скобке Пуассона символов соответствует почти точно (по модулю более «слабых» операторов) коммутатор операторов.

Оператор $U(t)$ лучше рассматривать в пространстве $\mathscr{L}^2$, чтобы иметь унитарную полугруппу, порождённую оператором $e^{\sqrt{-\Delta}it}$. Оператор $U(t)$ удовлетворяет задаче $U'_t = iAU$. Посмотрим, как этот оператор преобразует оператор $Q_t = U(-t)QU(t)$. Дело в том, что в квантовой механике есть два подхода. Один их них введён Шрёдингером, а другой — Гейзенбергом. По Гейзенбергу волновая функция остаётся постоянной, но меняются наблюдаемые. Наблюдаемые соответствуют псевдодифференциальным операторам.

Оператор $Q_t = U(-t)QU(t)$ будет оставаться псевдодифференциальным оператором, если Q — псевдодифференциальный оператор. Заметим его основное свойство: $Q'_t = i[A, Q]$. Если перейти к символам, то коммутатор заменится скобкой Пуассона. Таким образом, символ будет удовлетворять уравнению с частными производными первого порядка, поэтому он будет постоянен вдоль траекторий.

Есть второе описание волнового фронта. Попробуем подействовать на функцию разными псевдодифференциальными операторами, которые переводят её в гладкую функцию. Такие псевдодифференциальные операторы обязаны на каких-то точках кокасательного пространства быть равными нулю. Те точки, на которых они заведомо равны нулю, образуют волновой фронт. Например, если умножать дельта-функцию на функцию, которая в нуле не равна нулю, то мы не сможем получить гладкую функцию. Чтобы получить гладкую функцию, нужно умножать на функцию,

которая в нуле равна нулю; она не обязана быть равна нулю ни в какой другой точке. Поэтому нуль — сингулярный носитель дельта-функции. Аналогичный результат имеет место для волновых фронтов. Только там нужно не умножать на функцию, а применять псевдодифференциальный оператор.

Если взять волновой фронт и спроектировать его на x, т. е. сделать ξ равным нулю, то получится тот же сингулярный носитель.

Таким образом, если посчитать волновой фронт u, то он распространяется вдоль потока, т. е.

$$WF(u(x,\, y,\, t)) \subset \{(x,\, \xi) = \Phi_t(y,\, -\eta)\}.$$

Положим $x = y$. Тогда

$$WF(u(x,\, x,\, t)) \subset \{(x,\, \xi - \eta) \colon (x,\, \xi)\Phi_t(x,\, -\eta)\}.$$

Теперь проинтегрируем по x:

$$WF\left(\int u(x,\, x,\, t)\, dx\right) \subset \{(t,\, \tau) \colon (x,\, \xi) = \Phi_t(x,\, \xi)\}.$$

Вот откуда появилась периодическая траектория. Таким образом, сингулярный носитель спектрального ядра тесно связан с периодическими траекториями. Представим себе, что периодических траекторий нет вообще (такого быть не может, но представим это). Тогда можно сделать вот что. Поскольку у функции u особенность только в нуле, то на сколь угодно большом интервале мы знаем функцию u с точностью до гладкой функции. На основании этого мы восстанавливаем функцию $N(\lambda)$ с точностью до

$$\frac{\lambda^{(d-1)/2}}{T} + o(\lambda^{(d-1)/2}).$$

А поскольку T произвольное, то оценка будет с точностью до $o(\lambda^{(d-1)/2})$.

Теперь обратимся к реальной жизни. Допустим, что периодических траекторий мало, но они всё-таки есть. Фиксируем какое-то большое T. Тогда периодические траектории с периодом не больше T образуют замкнутое нигде не плотное множество, потому что мы предположили, что периодические траектории образуют множество меры нуль. Разобьём фазовое пространство на две части. Одна часть — малая окрестность периодических траекторий с периодом не больше T. Другая часть — всё остальное. В соответствии с этим можно разбить пропагатор; можно разбить спектральный проектор; можно разбить его след. Если взять бо́льшую часть, то к ней можно применит те же самые аргументы, что и выше. Если же взять маленькую часть, то T будет фиксированным и конечным, потому что есть окрестность нуля, вне которой нет особых точек.

Но возникнет ещё один коэффициент — мера множества; этот коэффициент можно сделать сколь угодно малым. Таким образом, оценка сверху будет

$$\left(\frac{C}{T}+C_T\varepsilon\right)\lambda^{(d-1)/2}+o(\lambda^{(d-1)/2}).$$

Эта оценка имеет место для любого ε и любого T, поэтому окончательная оценка будет $o(\lambda^{(d-1)/2})$.

Теперь обратимся к областям (когда граница непуста). Этот случай оказался очень сложным. Вся проблема в том, что построить параметрикс (пропагатор) в виде осциллирующих интегралов для областей с границами нельзя. Вообще-то, для некоторых типов областей границей можно, но чем более общий случай рассматривается, тем более сложно это делается. Такой путь достаточно бесперспективен.

Вся проблема в том, что есть лучи, касательные к границе. Тот случай, когда параметрикс построить можно — это случай, когда касание не очень сильное, когда есть невырожденность.

Мне удалось придумать некий способ, как построить след параметрикса для областей с границей. Мне будет более удобно рассматривать немного другую задачу, потому что для неё метод гораздо более прозрачный, чем он был в самом начале. Вместо того чтобы рассматривать число собственных значений оператора Лапласа, не превосходящих λ, я буду рассматривать число собственных значений оператора $H=-h^2\Delta-V$, не превосходящих нуля. Если в качестве V взять единичный оператор, а в качестве h^2 взять λ^{-1}, то это будет та же самая задача. Таким образом, это — задача о квазиклассической асимптотике собственных значений оператора Шрёдингера ($h\to 0$).

Нам нужно рассмотреть задачу $hU_t=iHU$, где $U=e^{ith^{-1}H}$. Давайте попробуем исследовать, где находятся сингулярности функции u. Допустим на мгновение, что сингулярности распространяются вдоль гамильтоновых траекторий, как и раньше. Тогда гамильтонова траектория не стоит на месте по x за исключением того случая, когда потенциал равен нулю. Таким образом, сингулярность должна быть изолированной. Доказать это с помощью осциллирующих интегралов или с помощью того метода Гейзенберга, который я описывал, для областей с границей невозможно. Однако есть очень грубый метод энергетических оценок. Он не даёт картинки, не даёт представления о решении. Но он хорош тем, что работает во многих ситуациях невзирая на частности. С помощью этого метода можно доказать, что если есть малое время T, то при $V(x)\neq 0$ функция $u(x,\,x,\,t)$ пренебрежимо мала при $t=T$. В обычном смысле функция пренебрежима, если она меньше любой степени h. Вообще-то, функцию нужно немножко

подреза́ть вблизи уровня потенциальной энергии, равной нулю; но это не существенно.

Изменением масштаба по t сделаем малую константу T равной 1. Соответственно нужно зафиксировать y, а x изменить по t. Тогда $u(x, x, T)$ будет пренебрежимо, но при этом вместо h будет h/T. Нужно оставить h/T маленьким, например, $h/T \leqslant h^{\delta}$. Таким образом, если $h^{1-\delta} \leqslant T \leqslant T_0$, то $u(x, x, T)$ остаётся пренебрежимым. Теперь мы знаем не только то, что сингулярность в нуле изолирована, но и то, в каких рамках она изолирована.

Давайте теперь попробуем построить решение нашего уравнения методом последовательных приближений. Любой человек вам скажет, что для таких задач метод последовательных приближений бессмыслен. Это правда на 99 %. Но давайте запишем $hu_t = iH(x, D_x)u$. В качестве невозмущённого оператора я запишу $H(y, D_x)$ (вспомните, что при $t = 0$ функция u равнялась дельта-функции). Если есть время t, то x отличается от y на величину порядка t, потому что скорость распространения конечна. Таким образом, ошибка метода последовательных приближений будет порядка t. Затем к ошибке можно применить обратный оператор, который записывается формулой Дюамеля: $h^{-1} \int\limits_0^T u(x, y, t) f(T, t)\, dt$. Норма обратного оператора примерно равна T/h. Если бы мы взяли конечное T, то каждая итерация имела бы всё бóльшую и бóльшую норму — вполне бессмысленный результат. Но если брать T очень маленьким, то можно получить осмысленный результат.

Обратный оператор $h^{-1} \int\limits_0^T U(T - t) f(t)\, dt$ надо применять к невязке

$$hu_t - iH(y, D_x)u = i(H(x, D_x) - H(y, D_x)).$$

Теперь мы пишем $u = R(H(x, D_x) - H(y, D_x))$, где R — обратный оператора. Равенство $u|_{t=0} = \delta(x - y)$ даёт y. Для каждого y мы применяем свою последовательность приближений. Когда мы решаем уравнение методом последовательных приближений, получаются итерации $u = (R(H(x, D_x) - H(y, D_x)))^n$.

При методе последовательных приближений важно, чтобы оператор имел маленькую норму; тогда последовательность приближений сходится. Маленькая норма здесь означает, что она должна быть меньше h^{δ}; тогда далёкие члены можно выбросить из рассмотрения, поскольку они будут меньше h в очень большой степени. Произведение нормы обратного оператора на норму того, к чему он применяется, примерно

равно $(T/h)T = T^2/h$. Нужно, чтобы это произведение было меньше h^δ, т. е. нужно, чтобы выполнялось неравенство $T \leqslant h^{1/2+\delta}$. Таким образом, этот метод работает не для всех времён, а только для достаточно маленьких.

Меня не интересует сам пропагатор, меня интересует только его след. След тоже можно построить от $h^{1/2-\delta}$ до $h^{1/2+\delta}$. Но я знаю, что начиная с $h^{1-\delta}$ дальше ничего нет до времени T_0. Таким образом, если построить решение методом последовательных приближений на довольно маленьком интервале и ещё воспользоваться совсем другими аргументами, о которых шла речь выше, мы автоматически получим оценки на некотором конечном интервале.

Дальнейшее расширение на большие интервалы идёт проще, потому что там можно заботиться только о почти всех направлениях, т. е. там можно избегать касательных к границе направлений. Самое сложное — построить решение именно вблизи $t = 0$. Именно там касательные траектории играют большую роль. Если мы находимся вдали от нуля, то там уже разница, по существу, между O большим и o малым, т. е. мы должны строить распространение вдоль траекторий за исключением некоторого малого множества. Это малое множество берёт на себя все касательные траектории.

Вот, собственно говоря, этот метод и сработал. Этот метод оказался весьма универсальным. Он работает во многих разных задачах.

Какова ситуация с негладкими задачами? С негладкими задачами дело обстоит примерно так. Писать негладкие псевдодифференциальные операторы — хуже не придумаешь. Но можно пытаться аппроксимировать негладкий оператор гладкими. Но аппроксимация — вещь хитрая, потому что мы выбираем параметр аппроксимации тоже как малый параметр. На самом деле нужно брать чуть больше, чем h.

Здесь есть две разные ошибки — ошибка аппроксимации и оценка остатков гладких операторов. Для остатков нужно получать равномерную оценку, хотя и аппроксимация у нас неравномерная, потому что она содержит малый параметр сглаживания.

Примерно тем же способом можно доказать, что несмотря на то, что аппроксимация неравномерно гладкая, оценки остатков будут равномерно гладкими. Здесь даже внутри области никакие конструкции с интегральными операторами Фурье нехороши по следующей причине. Все эти конструкции с осциллирующими интегралами или с представлением Гейзенберга начинают смешивать x и ξ. Но когда они смешивают x и ξ, шкалы по x и ξ нужно брать одинаковыми, равными параметру аппроксимации ε. А по принципу неопределённостей произведение этих шкал должно быть

не меньше h, т. е. $\varepsilon \geqslant h^{1/2}$. Это плохой выбор; для нужной оценки остатков нужна гладкость как минимум 2.

Однако метод энергетических оценок грубый; он не требует, чтобы по x и ξ была одна и та же шкала. По x можно взять чуть больше, чем h, а по ξ — чуть меньше, чем 1; тогда будет всё хорошо.

Это первое. Второе связано с границей. С границей дело оказалось сложнее. Если мы начнём границу выпрямлять, то первая производная от границы влезает в коэффициенты. Поэтому, чтобы всё прошло по-прежнему, нужно, чтобы граница была как минимум дважды непрерывно дифференцируема; это уже чрезмерно. Здесь сработали некие игры со старым доказательством Роберта Сили. Это доказательство заключалось в следующем. Нам нужно избежать границы. Пока время меньше расстояния до границы, то о границе мы ничего не знаем. Поэтому нужно взять шарик, радиус которого равен некой константе, делённой на время. Получим оценку $\int \frac{\lambda^{(1-d)/2}}{d}$, где d — расстояние до границы. Этот интеграл, к сожалению, расходится (логарифмически). Однако если бы d стояло в степени чуть меньше 1, то этот интеграл сходился бы. Сили понял, что есть две возможности. Либо точка движется к границе круто. Тогда она отразится нормально и всё будет хорошо. Либо она движется почти параллельно границе. Тогда она хотя бы в одном из двух противоположных направлений долго не дойдёт до границы. Поэтому если мы выбираем направление плюс-минус по времени, то до границы мы не дойдём в гладком случае за время порядка $\sqrt{d}$. В негладком случае это не так, но всё равно d будет в степени меньше 1. После этого все эти аргументы нужно перевести с классического случая на квантовый.

Если нас интересует только след, то выбор направления времени в нашей воле.

Если гладкость чуть больше 1, то получаем оценку $O(\lambda^{(d-1)/2})$.

Для многогранников верна оценка $\lambda^{\frac{d-1}{2}-\delta}$.

18 мая 2000 г.

Ю. И. Манин

НЕКОММУТАТИВНАЯ ГЕОМЕТРИЯ И КВАНТОВЫЕ ТЭТА-ФУНКЦИИ

В Независимом университете я читаю лекцию уже не впервые. В этот раз я хочу рассказать о своей очередной попытке понять, что может называться квантованной версией абелевых многообразий. Вся история квантования математических объектов, в отличие от физических, занимала добрую половину прошедшего века. Событием, происшедшем уже на моей памяти, было изобретение квантовых групп. Вероятно, многие из вас знают, что это такое. Но я вкратце напомню определение квантовой группы GL(2).

Мы начинаем с квантования пространства матриц M(2). В обычной интерпретации это пространство состоит из матриц $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. На этом пространстве a, b, c, d — функции (коммутирующие координаты). Если k — основное поле, то M(2) можно рассматривать как Spec $k[a, b, c, d]$ — спектр кольца многочленов от четырёх переменных, которые объединяются в матрицы. Пространство квантовых матриц $\mathrm{M}_q(2)$ представлено чем-то, что условно можно назвать «некоммутативным спектром». Это — объект, двойственный кольцу $k\langle a, b, c, d\rangle/(*)$. Здесь имеется в виду факторкольцо кольца, свободно порождённого над k переменными a, b, c, d, по идеалу, порождённому следующими соотношениями. Будем считать, что $a \to b$ обозначает соотношение $ab = q^{-1}ba$, где q — некоторый элемент основного поля (параметр квантования). Рассмотрим четыре стрелки:

$$\begin{array}{ccc} a & \longrightarrow & b \\ \downarrow & & \downarrow \\ c & \longrightarrow & d. \end{array}$$

Каждой из этих четырёх стрелок соответствует соотношение указанного выше вида. Кроме того, будем считать, что переменные b и c коммутируют: $bc = cb$. Имеется также более сложное соотношение $ad - da = (q^{-1} - q)bc$.

Некоммутативное кольцо, которое получается при факторизации по такому идеалу соотношений, объявляется кольцом некоммутативных функций на квантованном пространстве матриц. Это кольцо обозначают $F(\mathrm{M}_q(2))$.

Здесь присутствует идея о том, что квантование, в частности, включает замену коммутирующих наблюдаемых на некоммутирующие. Но главное, что оправдывает введение такого определения, состоит в том, что сохраняются некоторые основные свойства кольца матриц.

Обычные матрицы образуют полугруппу: если взять две матрицы $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ и $\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$ с элементами из коммутативного кольца, то их произведение тоже будет матрицей с элементами из того же коммутативного кольца. Верно ли такое утверждение для квантованных матриц? Нет, это неверно. Если взять произведение двух матриц, элементы которых удовлетворяют указанным выше коммутационным соотношениям, то его элементы уже не обязательно будут удовлетворять этим коммутационным соотношениям. Однако эти коммутационные соотношения будет выполняться в том случае, когда элементы a, b, c, d попарно коммутируют с элементами a', b', c', d'.

Обычно это говорят на более формальном языке. Рассмотрим диагональное отображение (коумножение)

$$\Delta\colon F(\mathrm{M}_q(2)) \to F(\mathrm{M}_q(2)) \otimes F(\mathrm{M}_q(2)),$$

при котором матрица $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ переходит в свой тензорный квадрат $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \otimes \otimes \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Это отображение является гомоморфизмом колец. Есть также операция умножения

$$\mu\colon F(\mathrm{M}_q(2)) \otimes F(\mathrm{M}_q(2)) \to F(\mathrm{M}_q(2)).$$

Вместе операции Δ и μ образуют структуру, которую называют *биалгеброй*.

Это ещё не матричная группа, поскольку есть вырожденные квантованные матрицы. Чтобы получить матричную группу, нужно написать квантовый детерминант и формально его обратить. Тогда получится некоммутативное кольцо, которое является кольцом функций на матричной группе. Это очень интересный объект, изобретение которого привело к осознанию того, что огромные части классической математики допускают квантованные варианты (что бы это ни означало), и эти квантованные варианты заслуживают изучения.

Эта деятельность превратилась в одну из глав некоммутативной геометрии. Некоммутативная геометрия сейчас представляет собой бурно развивающуюся область математики, состоящую из нескольких плохо взаимодействующих частей. Мы не вполне понимаем, что такое некоммутативная геометрия. Но несомненно, что квантовые группы являются одной из её глав и составляют один пакет основных идей некоммутативной

геометрии. Позже я скажу немножко больше о философии. Сейчас я перейду к конкретной проблеме квантования абелевых многообразий, а в конце лекции расскажу о конкурирующих и взаимодействующих идеях, которые сейчас обсуждаются в некоммутативной геометрии.

Нужно ещё сказать, что все коммутационные соотношения для матричных элементов $M_q(2)$ (которые были известны много лет назад) можно получить из более простой идеи. А именно, рассмотрим квантованную плоскость, в которой переменные x и y удовлетворяют основному хорошо известному соотношению $xy = qyx$. Рассмотрим, далее, матрицы с некоммутирующими коэффициентами, для которых

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix},$$

где x' и y' удовлетворяют тому же самому коммутационному соотношению $x'y' = qy'x'$. Иными словами, рассматриваются матрицы как бы автоморфизмов квантованной плоскости. Тогда из предположения, что a, b, c и d коммутируют с x и y, вытекают коммутационные соотношения. Таким образом мы получаем чрезвычайно наглядное выражение того, что квантовая группа, не будучи группой, тем не менее является описанием симметрий нового вида.

Если рассматривать матрицы с коэффициентами из коммутативного кольца, то обыкновенных симметрий очень мало — только диагональные матрицы и перестановки. А вот квантовых симметрий ровно столько, сколько и классических. Квантовые группы — это симметрийный объект.

Известно, что в алгебраической геометрии есть два основных класса групп. Это линейные алгебраические группы, реализуемые как подгруппы в группе матриц, задаваемые системами алгебраических уравнений, и абелевы многообразия, которые являются коммутативными, но зато не аффинными. Абелевы многообразия — это интересные алгебраические многообразия, которые вкладываются в проективное пространство. Простейшими примерами абелевых многообразий являются неособые кубические кривые на плоскости. На абелевых многообразиях есть групповые законы, но все эти законы коммутативны.

Естественный вопрос о том, нельзя ли проквантовать абелевы многообразия, стоит давно. Я могу сразу сказать, что окончательный ответ на этот вопрос не известен; не известно правильное определение. Я сделал попытку дать определение квантового абелева многообразия 10 лет назад, а сейчас вернулся к ней с некоторыми добавлениями и усовершенствованиями.

При квантовании алгебраических групп понятен и общепринят, по крайней мере, ответ на вопрос о том, в какую вселенную попадают квантованные линейные алгебраические группы. Они попадают во вселенную алгебр Хопфа. Алгебра Хопфа A — это линейное пространство над полем k с ассоциативным умножением μ и коумножением Δ. Кроме того есть отображение $A \to k$, которое на групповом языке означает вычисление значения в единице (аугментация); есть также вложение постоянных функций $k \to A$. Наконец, есть инволюция $i\colon A \to A$, которой на группе соответствует переход от элемента b к элементу b^{-1} (на языке алгебр Хопфа эту инволюцию называют *антиподом*). В результате получаем следующую диаграмму отображений:

$$\begin{array}{ccccc} & & k & & \\ & & \downarrow & & \\ A\otimes A & \xrightarrow{\mu} & A & \underset{\circlearrowleft}{\overset{\Delta}{\rightrightarrows}} & A\otimes A. \\ & & \downarrow & & \\ & & k & & \end{array}$$

Известно, что квантовые алгебраические группы суть некоторые алгебры Хопфа. Вопрос только в том, как в мире алгебр Хопфа выбрать те, которые близки к алгебрам функций на линейных группах.

Если говорить о квантованных абелевых многообразиях, то совершенно непонятно, в какой мир они должны попадать. Это определённо не алгебры Хопфа по той банальной причине, что поскольку абелево многообразие проективное, а не аффинное, то у него нет кольца алгебраических функций и, соответственно, нет естественно связанной с ним коммутативной и кокоммутативной алгебры Хопфа. Поэтому уже обыкновенное абелево многообразие в таком виде представить нельзя.

Много лет назад Дэвид Мамфорд в замечательной работе «Уравнения, определяющие абелевы многообразия» занимался проблемой, куда попадают в алгебраическом смысле абелевы многообразия самые обыкновенные. Он рассуждал примерно так. Пусть $\mathscr{A}$ — некоторое абелево многообразие. Если у него нет аффинного кольца функций, то во всяком случае есть какое-нибудь проективное. Добавим какой-нибудь очень обильный пучок L, рассмотрим все его степени и возьмём однородное кольцо, которое представляет это абелево многообразие (т. е. проективный спектр этого однородного кольца является данным абелевым многообразием):

$$\bigoplus_{n=0}^{\infty} \Gamma(\mathscr{A}, L^{\otimes n}).$$

Обозначим полученное кольцо $F(\mathscr{A}, L)$. Может быть оно является алгеброй Хопфа? Нет, это кольцо тоже не алгебра Хопфа. Оно не является алгеброй Хопфа потому, что если написать отображение умножения (сложения) в группе $\mathscr{A} \times \mathscr{A} \xrightarrow{m} \mathscr{A}$, заданное формулой $(x, y) \mapsto x + y$, и потом применить контравариантное отображение в кольцо однородных функций на абелевом многообразии, то его нельзя будет связать с самим собой. Неверно, что m^* отображает L в $L \boxtimes L$. Это неверно потому, что $m^*(L) \neq L \boxtimes L$. Когда вы переносите обратимый пучок с помощью операции сложения, вы получаете нечто, не выражающееся через этот обратимый пучок.

Идея Мамфорда (точнее говоря, он реализовал идею, известную раньше) состояла в том, что вместо отображения чистого сложения нужно взять комбинированное отображение $\mathscr{A} \times \mathscr{A} \xrightarrow{M} \mathscr{A} \times \mathscr{A}$, заданное формулой $(x, y) \mapsto (x + y, x - y)$. Тогда $M^*(L \boxtimes L) = L^2 \boxtimes L^2$ (здесь тензорные произведения внешние). Таким образом, M^* индуцирует некое отображение, которое нужно построить, из этого кольца. Однако получившаяся структура не будет алгеброй Хопфа; она будет чем-то другим. Это что-то другое до сих пор не аксиоматизировано. Мамфорд изучил получившуюся структуру (диаграммы градуированных колец) с такой степенью подробности, что он сумел в достаточно явном виде написать координаты на пространстве модулей этих структур и уравнения, которые их выделяют; т. е. он получил алгебраическое описание пространства модулей абелевых многообразий и самих абелевых многообразий, живущих над точками пространства модулей. Но то, что он получил, есть настолько конкретное описание, что оно не вкладывается ни в какую бóльшую вселенную. Поэтому опять непонятно, как это квантовать.

Когда я впервые стал задумываться об этом сюжете, моя идея была такая. Рассмотрим абелево многообразие $\mathscr{A}^d$. Временно будем работать над полем комплексных чисел (значительная часть последующего будет применима к полным нормированным полям типа поля p-адических чисел). Группа комплексных точек $\mathscr{A}^d(\mathbb{C})$ — это вещественный $2d$-мерный тор. У него есть универсальное накрытие и промежуточное накрытие:

$$\begin{array}{ccccc} \mathbb{R}^{2d} & \longrightarrow & \mathbb{C}^{*d} & \longrightarrow & \mathscr{A}^d(\mathbb{C}) = \mathbb{C}^{*d}/B. \\ \cup & & \cup & & \\ \Omega & \longrightarrow & B & & \end{array}$$

Здесь Ω — группа периодов, B — её образ, который является группой мультипликативных периодов.

В подходящем базисе группа периодов может быть записана как матрица

$$\begin{pmatrix} 1 & 0 & \dots & 0 & \tau_{11} & \dots & \tau_{1d} \\ 0 & 1 & \dots & 0 & \tau_{21} & \dots & \tau_{2d} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \tau_{d1} & \dots & \tau_{dd} \end{pmatrix}.$$

В этом базисе получаем нечто вроде экспоненциального отображения: как обычно алгебра Ли отображается на соответствующую группу Ли; возникает комплексная структура. Таким образом, можно выбрать мультипликативную униформизацию, которая представляет комплексный тор в виде фактора произведения нескольких комплексных мультипликативных групп по группе мультипликативных периодов. Этот фактор — $\mathbb{C}$-точки алгебраического тора. Выбор группы периодов в естественном смысле однозначен, тогда как выбор того, что получается на полдороге, неоднозначен. Но этот выбор часто бывает существен при исследовании абелевых многообразий; я всегда буду его осуществлять.

Самая классическая версия мультипликативной униформизации — представление эллиптической кривой в виде фактора

$$E = \mathbb{C}^*/(q^{\mathbb{Z}}), \quad q = e^{2\pi i\tau}.$$

Если на абелевом многообразии $\mathscr{A}$ дополнительно выделить обратимый пучок L, то на алгебраическом уровне у него есть конечномерное пространство сечений $\Gamma(L)$, а на уровне униформизации сечения превращаются в пространство тэта-функций данного типа. Это — комплексно аналитические функции на универсальном накрытии, которые почти периодичны относительно сдвигов на Ω (точный характер периодичности я напомню позже). В точности периодичными эти функции быть не могут, потому что тогда они были бы ограничены на всём пространстве. При сдвиге на период они умножаются на некоторый экспоненциальный множитель.

Как квантовать алгебраические торы, имеется давно известный стандартный рецепт. Их нужно квантовать, слегка обобщая соотношение $xy = q^{-1}yx$. Моя идея состояла в том, что можно попробовать построить аналог абелева многообразия, воображая его себе как фактор квантованного тора по мультипликативной решётке периодов, и строя алгебраические структуры, которые наполняют это воображение алгебраическим содержанием, непосредственно из построения квантовой версии соответствующего пространства тэта-функций. Ближайшая часть лекции будет посвящена описанию деталей этой конструкции. А потом я начну философствовать и объяснять, что в этой конструкции хорошо и что плохо.

1.
Категория некоммутативных торов

Пусть K — основное поле (в самой содержательной части конструкции это $\mathbb{C}$ или $\mathbb{Q}_p$ или какое-то другое полное нормированное поле). Некоторое время можно работать с произвольным полем нулевой характеристики.

Пусть (H, α) — пара, состоящая из свободной абелевой группы H конечного ранга и кососимметрического спаривания $\alpha\colon H \times H \to K^*$ (спаривание не в аддитивной группе, а в мультипликативной, — тут присутствуют экспоненты). Формально это записывается так:

$$\alpha(g, h) = \alpha(h, g)^{-1},$$
$$\alpha(h_1 + h_2, g) = \alpha(h_1, g)\alpha(h_2, g).$$

Отметим, что хотя спаривание антисимметрично, оно может случайно оказаться и симметричным, если все значения α равны ± 1. С этим связан ряд тонкостей, о которых нужно не забывать. По модулю ± 1 спаривание действительно антисимметрично.

Характер $\varepsilon(h) = \alpha(h, h)$ принимает значения ± 1, хотя и выглядит как квадратичная функция.

Морфизм таких пар $(H_1, \alpha_1) \to (H_2, \alpha_2)$ — это гомоморфизм групп $f\colon H_1 \to H_2$, для которого $\alpha_2^2(f(h), f(g)) = \alpha_1^2(h, g)$. У морфизма есть характеристика (билинейная форма)

$$\alpha_2(f(h), f(g))^{-1}\alpha_1(h, g) = \varepsilon_f(h, g) \in \{\pm 1\}.$$

Некоммутативным, или *квантовым*, тором с группой характеров (H, α) называют нечто, представленное кольцом функций. Функции я рассматриваю трёх родов:

— алгебраические $\mathrm{Al}(H, \alpha) = \bigoplus\limits_{h \in H} K\, e(h)$, т. е. это — линейное пространство, натянутое на набор формальных символов $e(h)$, $h \in H$, а умножение задаётся формулой $e(h)e(g) = \alpha(h, g)e(h + g)$ *);

— формальные $\mathrm{Af}(H, \alpha) = \prod\limits_{h \in H} K\, e(h)$. Они не образуют кольца, но образуют модуль над алгебраическими функциями;

— аналитические

$$\mathrm{An}(H, \alpha) = \left\{ \sum a_h e(h)\colon a_h \in K,\ |a_h| = O(\|h\| + 1)^{-N} \text{ для любого } N \right\}.$$

*) Здесь e относится к разным H и α. Если мне нужно будет подчеркнуть, к каким именно, я буду писать более подробно $e_{H,\alpha}$. Например, правила коммутации зависят от H и α.

Эти функции определены только над полным нормированным полем. Они образуют кольцо, содержат алгебраические функции и содержатся в формальных. Аналитические функции являются модулем над алгебраическими.

Сам тор будем обозначать $T(H, \alpha)$. Торы образуют категорию, дуальную категории функций. На этой лекции — дуальную категории алгебраических функций Al. Другое понятие получится, если взять формальные или аналитические функции. Морфизм торов $T(H_2, \alpha_2) \to T(H_1, \alpha_1)$ — это то же самое, что обратный морфизм колец алгебраических функций

$$F^*\colon \mathrm{Al}(H_1, \alpha_1) \to \mathrm{Al}(H_2, \alpha_2).$$

Следующее простое предложение показывает, что морфизмы таких колец тесно связаны с морфизмами характеров.

Предложение. а) *Обратимые элементы в* $\mathrm{Al}(H, \alpha)$ *— это в точности множество* $\{ae(h)\colon a \in K^*, h \in H\}$.

б) *Для любого морфизма выполняется равенство* $F^*(e(h)) = a_h(ef(h))$. *При этом* $f\colon H_2 \to H_1$ *— гомоморфизм групп характеров, для которого выполняется соотношение*

$$a_h a_g a_{h+g}^{-1} = \alpha_1(h, g)\alpha_2^{-1}(f(h), f(g)). \tag{1}$$

Более того, f *— морфизм* $(H_2, \alpha_2) \to (H_1, \alpha_1)$.

в) *Каждый морфизм торов определяет морфизм групп характеров* f, *и если фиксировать* f, *то множество всех морфизмов торов* F, *проектирующихся в* f, *либо пусто, либо является главным однородным пространством над гомоморфизмами*

$$\mathrm{Hom}(H_1, K^*) = T(H_1, 1)(K).$$

Здесь $T(H_1, 1)(K)$ *— K-точки обычного коммутативного тора с группой характеров* H_1.

Утверждение а) проще всего доказать с помощью того, что любую свободную абелеву группу с конечным числом образующих можно вполне упорядочить так, что сложение совместимо с этим порядком (для этого свободную абелеву группу нужно представить как решётку в $\mathbb{R}^n$ и выбрать гиперплоскость, пересекающую эту решётку только по нулю; всё, что лежит по одну сторону от гиперплоскости, нужно объявить положительным, а всё, что по другую, отрицательным). Теперь уже легко перечислить обратимые элементы в кольце алгебраических функций. Действительно, у любого элемента этого кольца есть старший моном и младший. Они никогда не будут совпадать, за исключением того случая, когда этот элемент сам

является мономом. При перемножении старшие мономы перемножаются, младшие тоже перемножаются. Стало быть, обратимы только мономы.

Утверждение б) вытекает из а). Равенство $F^*(e(h)) = a_h(ef(h))$ следует из того, что единица переходит в единицу. То, что f — гомоморфизм групп характеров, следует из формулы умножения. Соотношение (1) тоже следует из формулы умножения; на это раз нужно проследить за константами a_h. Левая часть равенства (1) симметрична по g и h, а правая кососимметрична. Поэтому их значения равны ± 1. После возведения в квадрат получим 1. Стало быть, отображение f совместимо с α_1^2 и с α_2^2, т. е. f не просто гомоморфизм групп, а морфизм $(H_2, \alpha_2) \to (H_1, \alpha_1)$.

По каждому морфизму торов мы построили морфизм групп характеров f и получили соотношение, которое должно выполняться, если гомоморфизм групп характеров можно достроить до морфизма торов. Если его вообще можно достроить, то его можно достроить многими разными способами. А именно, можно выбрать произвольное отображение $a_h\colon H \to K^*$, которое мультипликативно по отношению к h и g. Морфизм торов однозначно определяется выбором этого отображения a_h. Это доказывает в).

2.
Стандартные морфизмы торов

Понимание того, как устроены морфизмы некоммутативных торов, позволяет построить некоторые стандартные морфизмы торов. Следует отметить, что при нетривиальном α некоммутативный тор сам по себе, вообще говоря, не является группой. Мы избежали упомянутой выше проблемы. У нас нет алгебр Хопфа, связанных с абелевыми многообразиями, и теперь уже нет алгебр Хопфа, связанных с некоммутативными торами: они не суть группы. Абелевы многообразия суть группы. Мы как будто бы теряем на этом обстоятельстве возможность говорить о многих структурных объектах, связанных с абелевыми многообразиями, потому что у нас нет групповых отображений. Тем не менее, морфизмы, которые заменяют соответствующие групповые отображения, есть. Вот список некоторых стандартных морфизмов некоммутативных торов, которые впоследствии переносятся на морфизмы квантованных абелевых многообразий. (Напомню, что сейчас мы находимся на уровне $(\mathbb{C}^*)^d \to \mathscr{A}^d(\mathbb{C})$, т. е. речь идёт об объектах, которые являются, так сказать, полууниверсальными накрывающими или универсальными полунакрывающими; я ещё не ввёл решётку периодов, но скоро введу.)

1. Действие обыкновенного коммутативного тора $T(H, 1)$ на некоммутативном торе $T(H, \alpha)$ с той же самой группой характеров H, но с

нетривиальным α. Это действие таково:

$$b \in T(H,\ 1)(K) \mapsto b^*(e(h)) = h(b)e(h),$$

где $h(b) = e_{H,1}(h)(b)$.

2. У нас нет сложения, но есть умножение $[n]$ на любое целое число n, правда, меняющее параметр квантования α. Морфизм

$$[n]\colon T(H,\ \alpha) \to T(H,\ \alpha^{n^2})$$

задаётся формулой

$$[n^*]e_{H,\alpha^{n^2}}((h)) = e_{H,\alpha}(nh).$$

3. Есть даже умножение

$$T(H,\ \alpha) \times T(H,\ \beta) \xrightarrow{m_{\alpha,\beta}} T(H,\ \alpha\beta).$$

Оно задаётся формулой

$$m^*_{\alpha,\beta}(e_{H,\alpha\beta}(h)) = e_{H,\alpha}(h) \otimes e_{H,\beta}(h).$$

4. Наконец, есть мамфордовский морфизм

$$T(H \oplus H,\ \alpha \oplus \alpha) \xrightarrow{M} T(H \oplus H,\ \alpha^2 \oplus \alpha^2),$$

который в будущем фрагменте теории должен играть основную роль. Он задаётся формулой

$$M^*(e(h,\ g)) = e(h + g,\ h - g).$$

То, что некоммутативные торы не являются группами, не означает, что с ними ничего нельзя делать. Есть масса морфизмов, которые являются квантовыми версиями соответствующих классических свойств.

Сейчас я введу решётку периодов. Некоммутативный тор $H(T,\ \alpha)$ не является группой, но на него действует $T(H,\ 1)$. Решётка периодов — это дискретная группа автоморфизмов B, которая действует на $H(T,\ \alpha)$. Морально я хочу построить факторпространство $H(T,\ \alpha)/B$. В классическом случае можно дополнительно сказать, что я могу взять группу периодов максимального ранга, который может иметь дискретная подгруппа в торе. И даже после этого фактор $(\mathbb{C}^*)^d$ ещё не обязан быть абелевым многообразием. Нужно ещё, чтобы на этом факторе было достаточно много мероморфных функций. В классическом случае существование достаточно многих мероморфных функций эквивалентно существованию достаточно многих тэта-функций. Сначала я буду изучать факторы на весьма формальном уровне, а потом, когда введу тэта-функции, я буду изучать квантовый аналог того, что на факторе имеется достаточно много независимых функций.

Построение факторпространства $H(T, \alpha)/B$ примерно равносильно тому, что берётся одна из версий колец функций — алгебраических Al, формальных Af или аналитических An — и находятся B-инвариантные элементы. В алгебраическом случае B-инвариантных элементов просто нет; этот случай исключается. В формальном случае нужно говорить не о B-инвариантных функциях (это слишком сильное условие), а о B-ковариантных функциях, которые при сдвигах на B приобретают множитель. Часть теории относится к изучению формальных B-ковариантных функций, а другая часть — к изучению B-ковариантных аналитических функций.

Пусть задан коммутативный тор $T(H, 1)$ с решёткой периодов B, которая на данный момент является просто подгруппой точек. Он действует на квантованный тор $T(H, \alpha)$. Я хочу написать формальные функции на квантованном торе, которые удовлетворяют определённому уравнению относительно сдвигов на элементы группы периодов.

О п р е д е л е н и е 1. *Двусторонний мультипликатор* для пары $(T(H, \alpha), B)$ — это система $\mathscr{L} = (h_l, h_r, \psi, (\ ,\ ))$, состоящая из четырёх объектов:

(i) гомоморфизмы групп $h_l, h_r: B \to H$ (левый h и правый h); мне часто будут нужны их сумма и разность $h^{\pm} = h_l \pm h_r$; я часто буду писать $h_l(b) = h_{b,l}$;

(ii) гомоморфизм групп $\psi: B \to K^*$;

(iii) симметрическое скалярное произведение $(\ ,\ ): B \times B \to K^*$, но опять же экспоненцированное, со значениями в мультипликативной группе. (Это связано с тем уровнем, на котором мы работаем. На следующем уровне оно становится настоящим скалярным произведением.)

Эта система должна удовлетворять следующему уравнению связи:

$$h^-_{b_2}(b_1) = (b_1, b_2)^2 \alpha(h_{b_1,l}, h_{b_2,l})\alpha(h_{b_1,r}, h_{b_2,r})^{-1}$$

для любых $b_1, b_2 \in B$.

О п р е д е л е н и е 2. *Фактор автоморфности*, отвечающий мультипликатору $\mathscr{L}$, — это набор линейных отображений, определённых для каждого элемента $b \in B$:

$$j_{\mathscr{L}}(b): \mathrm{Af}(H, \alpha) \to \mathrm{Af}(H, \alpha).$$

(Af можно заменить на Al или An.) Эти отображения определяются следующим образом:

$$\Phi \mapsto \psi(b)(b, b)e(h_{b,l})\Phi e(h_{b,r})^{-1}$$

Л е м м а. *Если $\mathscr{L}$ фиксировано, то отображение $b \mapsto j_\alpha(b) \circ b^*$ — гомоморфизм групп.*

Таким образом определяется действие группы периодов на пространстве функций любого типа — алгебраических, формальных, аналитических. В частности, можно поставить задачу об описании подпространств, инвариантных относительно действия этой группы. Эта задача решается чистыми вычислениями, которые оправдывают упомянутое выше соотношение. Оно берётся из естественного необходимого условия: сдвиги на периоды, подправленные на фактор автоморфности, должны определять групповое действие. Сам же выбор действия, с неизвестными ψ и e, является довольно примитивным обобщением классического случая. В классическом случае e, конечно, коммутируют между собой, поэтому никакого резона разбивать e на левые и правые нет. Все их можно писать слева. Нет никакого резона вводить что-нибудь кроме h^-, потому что если в классическом варианте записать всё слева, то возникнет только h^-. А множители $\psi(b)$ и $(b,\, b)$ появляются и в классической версии. Имеется экспоненцированная квадратичная форма от периода и экспоненцированная линейная форма от периода. Они естественным образом сохраняются.

В старом варианте своей статьи 90-го года я поставил только одну экспоненту и написал, что её можно ставить как слева или справа. (Это означает переход к тору с противоположным умножением.) Но так или иначе некоммутативность факторов автоморфности приводит к тому, что тэта-функции нельзя умножать. Потом мне пришло в голову, что тэта-функции можно будет хотя бы частично умножать, если я позволю себе ставить экспоненты с двух сторон.

О п р е д е л е н и е 3. Формальная тэта-функция с мультипликатором $\mathscr{L}$ — это элемент $\Phi \in \mathrm{Af}(H,\, \alpha)$, который обладает тем свойством, что группа действует на него тривиально, т. е. $b^*(\Phi) = j_{\mathscr{L}}(b)(\Phi)$ для любого $b \in B$.

В частности, в классическом случае, когда $\alpha = 1$, уравнение становится таким:

$$b^*(\Phi) = \psi(b)(b,\, b)e(h_b)^{-1}(\Phi).$$

Уравнение связи принимает вид $h^-_{b_2}(b_1) = (b_1,\, b_2)^2$. Поэтому не нужно даже задавать h^-.

В определение формальных тэта-функций вошли римановы условия симметрии; они оправданы леммой. Теперь я напишу теорему о квантовании римановых условий положительности в версии, которая годится для $\mathbb{C}$ и для $\mathbb{Q}_p$; при специализации она, в частности, даёт обычную теорию Якоби—Тэйта p-адических тэта-функций.

Т е о р е м а о к в а н т о в а н и и у с л о в и й п о л о ж и т е л ь н о с т и. а) *Пусть $\Gamma(\mathscr{L})$ — пространство всех тэта-функций с мультипликатором $\mathscr{L}$. Тогда* $\dim \Gamma(\mathscr{L}) = [H : h^-(B)]$.

б) *Пусть основное поле K полное нормированное. Тогда если* $[H : h^-(B)] < \infty$, $\ln|(b, b)\alpha(h_{b,l}h_{b,r}^{-1})|$ *— положительно определённая квадратичная форма на* B *и* $B \subset T(H, 1)(K)$ *дискретно, то все тэта-функции сходятся, т. е. лежат в* $\mathrm{An}(H, \alpha)$.

В частности, если α унитарно, т. е. $|\alpha| = 1$, то условие положительности становится таким же, как в классическом случае. В p-адическом случае модуль p-адический. Тогда условие положительности то же самое, что возникает в теории p-адических тэта-функций.

Морально вы уже понимаете, что тип тэта-функций — это нечто вроде квантованного обратимого пучка на будущем абелевом многообразии, а $\Gamma(\mathscr{L})$ — нечто вроде квантованного пространства его сечений, реализованное после подъёма в накрывающее пространство. В классическом случае это верно буквально: если есть обратимый пучок, то у него есть подъём как тэта-тип, состоящий из двух компонент, а пространство его сечений превращается в пространство тэта-функций, удовлетворяющих определённому функциональному уравнению. Поэтому я выбрал обозначения так, чтобы они напоминали ситуацию в классическом случае.

Доказательство теоремы о квантовании условий положительности в достаточной мере тупое. Записывается функциональное уравнение и рассматриваются коэффициенты. Вы обнаруживаете, что на каждом представителе класса смежности H по $h^-(B)$ коэффициенты можно задать как угодно, а остальные функции по ним однозначно восстанавливаются. Стало быть, количество линейно независимых функций найдено. Затем нужно посмотреть, как восстанавливаются остальные функции. Они умножаются примерно на экспоненту плюс нечто, зависящее от b только линейно. Поэтому квадратичная часть должна обеспечивать убывание быстрее любой степени.

Теперь нужно сделать замечание об умножении. Пусть есть два тэта-типа (мультипликатора) $\mathscr{L} = (h_l, h_r, \psi, (\ ,\ ))$ и $\mathscr{L}' = (h'_l, h'_r, \psi', (\ ,\ )')$. Предположим, что $h'_l = h_r$. Тогда их произведение определяется так: $\mathscr{L}\mathscr{L}' = (h_l, h'_r, \psi\psi', (\ ,\ )(\ ,\ )')$. Утверждение состоит в том, что имеется частичное отображение $\Gamma(\mathscr{L}) \otimes \Gamma(\mathscr{L}') \to \Gamma(\mathscr{L}\mathscr{L}')$. Это наводит на мысль, что такое отображение является морфизмом в некоторой категории. К этому мы ещё вернёмся. Временно можно представлять себе, что сумма всех $\Gamma(\mathscr{L})$ — нечто вроде универсального градуированного кольца, связанного с абелевым многообразием. В данном случае есть только полукольцо. В классическом случае есть кольцо

$$\bigoplus_{\mathscr{L}} \Gamma(\mathscr{L}) \approx \bigoplus_{\alpha \in \mathrm{Pic}\ \mathscr{A}} \Gamma(\mathscr{A}, \alpha).$$

Более правильна интерпретация тэта-функций как морфизмов в некоторой категории, но у меня нет правильного определения объектов этой категории. Морфизмы есть, умножение тоже есть, а что такое объект категории, непонятно. Но нет никакого сомнения в том, что то, что здесь написано, является квантовой деформацией категории Пикара, связанной с абелевым многообразием.

3.
Функториальные свойства тэта-функций

Морфизмы торов, совместимые с периодами, действуют на квантованные торы. Формулировка этого утверждения очевидна, а доказательство немножко занудное.

Начнём с морфизма накрывающих торов. В частности, он действует на характерах $F\colon T(H_2, \alpha_2) \to T(H_1, \alpha_1)$ и индуцирует морфизм $F^*(e_{H_1,\alpha_1}(h)) = a_h(f(h))$. Для коммутативных торов получаем морфизм $F^*(e_{H_1,1}(h)) = (f(h))$. Предположим, что для коммутативных торов коммутативна диаграмма

$$\begin{array}{ccc} F\colon T(H_2,\,1) & \to & T(H_1,\,1) \\ \cup & & \cup \\ B_2 & \xrightarrow{F} & B_1, \end{array}$$

т. е. морфизм совместим с периодами. В этой ситуации тэта-функции одного типа можно перетаскивать в тэта-функции другого типа при дополнительном предположении, что a_h — характер. Как вы помните, a_{h+g} отличается от $a_h a_g$ на некий множитель, равный ± 1, который я называл характеристикой морфизма. Мы предполагаем, что характеристика морфизма равна 1.

Утверждение 1. *Для системы* $\mathscr{L}_1 = (h_l, h_r, \psi, (\ ,\ ))$, *связанной с* $T(H_1, \alpha_1)$ *и* B_1, *морфизм* F *определяет обратный образ* $\mathscr{L}_2 = F^*\mathscr{L}_1 = (h'_l, h'_r, \psi', (\ ,\ )')$.

Здесь $h'_{l,r}$ определяется коммутативной диаграммой

$$\begin{array}{ccc} B_1 & \xrightarrow{h_{l,r}} & H_1 \\ {\scriptstyle F}\uparrow & & \downarrow{\scriptstyle f} \\ B_2 & \xrightarrow{h'_{l,r}} & H_2, \end{array}$$

т. е. $h'_{l,r} = f \circ h_{l,r} \circ F$. Другие элементы системы $\mathscr{L}_2$ определяются так: $\psi'(b) = \psi(F(b))a_{h^-_{F(b)}}$ и $(b_1,\, b_2)' = (F(b_1),\, f(b_2))$.

У т в е р ж д е н и е 2. а) *Система* $\mathscr{L}_2$ *— мультипликатор.*
б) *Если* $\theta \in \Gamma(\mathscr{L}_1)$, *то* $F^*(\theta) \in \Gamma(F^*(\mathscr{L}_2))$.

4.
Немного философии

То, о чём я рассказал, — базис, на котором можно играть в игру, довольно длинную. Можно взять вышеупомянутую статью Мамфорда. Значительный её кусок квантуется, и с помощью этой системы объектов и определений перетаскивается в некоммутативную область. Интересно, что при этом получится.

Вторая игра, в которую с этим можно играть, связана с тем, что когда есть некоммутативный тор и простые коммутационные соотношения $xy = qyx$, для них есть простые представления псевдодифференциальных операторов: умножение на x — это оператор умножения на экспоненту, а умножение на y — оператор умножения на экспоненту $\frac{\partial}{\partial x}$, т. е. сдвиг на 1. Обычно каждая некоммутативная теория существует в двух версиях: одна как некоммутативная, а другая как коверсия коммутативной теории. Классический пример — q-биномиальные коэффициенты. Их можно записать с помощью формулы бинома Ньютона, применённого к некоммутативным переменным:

$$(x+y)^n = \sum \binom{n}{i}_q x^i y^{n-i}, \quad \text{где} \quad xy = qyx.$$

С другой стороны, q-биномиальные коэффициенты можно записать, оставаясь в рамках коммутативной алгебры:

$$\prod (1 + q^n y x^{-1}) = \sum \binom{n}{i}_q (yx^{-1})^{n-i}.$$

Точно так же все эти алгебраические игры существуют в двух вариантах: в некоммутативном, который я изложил, и в коммутативном, который получится, если все фигурирующие экспоненты заменить на соответствующие операторы, одна часть которых соответствует сдвигам, а другая — умножениям на экспоненту. Что за q-версии тэта-функций при этом получаются — интересный вопрос, который я совсем не изучал.

Наконец, третья игра, в которую можно играть, — сменить идеологию некоммутативной геометрии. Но прежде чем её менять, нужно понять, какая идеология некоммутативной геометрии была неявной здесь. Это была, конечно, деформационная идеология. Как и в примере, с которого я начал (что такое квантовое GL(2)?), мы начинаем с кольца функций

на обыкновенной группе GL(2), затем вводим параметры деформации и плоским образом деформируем коммутативное кольцо в некоммутативное. Изначальная идеология здесь была такой же. У нас было универсальное градуированное кольцо, связанное с абелевым многообразием (прямая сумма сечений всех обратимых пучков). Я ввожу параметр деформации и начинаю деформировать умножение, пытаясь сохранить при этом какие-то разумные свойства.

В некоммутативной геометрии есть дополнительная идеология, связанная с именем Алэна Конна. Эта идеология совсем не такая. Конн тоже работает с некоммутативными кольцами и модулями над ними. Значительная часть его теории связана, скорее, с некоммутативным функциональным анализом, чем с алгебраической геометрией. Но это не принципиально. Принципиален сдвиг в понимании того, как нужно связывать коммутативную геометрию с некоммутативной. Этот сдвиг связан в первую очередь с тем, что в некоммутативной геометрии возникает знаменитый эффект Мориты, которого не было в коммутативной геометрии. Именно, категория модулей над коммутативным кольцом однозначно определяет это кольцо. Категория модулей над некоммутативным кольцом вовсе это кольцо однозначно не определяет.

В некоммутативной геометрии нет понятия функции на данном пространстве. Матрицы от функций на данном пространстве играют совершенно такую же роль, как сами функции. В любой глобальной версии мы теряем понятие структурного пучка. Стало быть, идеология деформаций и идеология Конна в некотором смысле несовместимы. В идеологии деформаций мы нажимаем на данное кольцо, на данный структурный пучок. Идеология деформаций совмещается с идеологией Конна, если мы будем деформировать не кольцо, а категорию модулей над этим кольцом; не многообразие, представленное каким-то коммутативным объектом, а категорию пучков на этом многообразии, и т. п. Здесь есть свои подводные камни, потому что нет окончательных определений деформации категории и даже категории, которую нужно деформировать.

Я говорил, что частичное отображение $\Gamma(\mathscr{L}) \otimes \Gamma(\mathscr{L}') \to \Gamma(\mathscr{L}\mathscr{L}')$ нужно рассматривать как морфизм в некоторой категории. Это открывает возможности для продумывания той же самой конструкции в рамках другой деформационной идеологии: у нас нет функций, а есть только пучки на воображаемом пространстве. Кроме того, это даёт возможность связать эту картину с очень интересными явлениями, которые давно известны в категории квантовых торов. Например, казалось бы, о двумерном квантовом торе

$$k[x, y, x^{-1}, y^{-1} : xy = qyx]$$

нельзя сказать ничего интересного. Но если мы рассмотрим

$$\operatorname{Spec} k[x, y, x^{-1}, y^{-1}: xy = qyx] = T_q,$$

где $q = e^{2\pi i\tau}$, $\tau \in \mathbb{R}$, то торы T_q и $T_{q'}$ Морита-эквивалентны тогда и только тогда, когда $\tau' = \dfrac{a\tau + b}{c\tau + d}$ (дробно-линейное преобразование с целыми коэффициентами). Это дробно-линейное преобразование является элементом модулярной группы. С классической точки зрения модулярная группа естественно действует на верхней полуплоскости. Таким образом получаем продолжение действия модулярной группы, действующей на верхней полуплоскости, имеющее чрезвычайно простой смысл. А именно, эквивалентность эллиптических кривых превращается в Морита-эквивалентность, когда мы садимся с верхней полуплоскости на вещественную ось. Таким образом, в некотором смысле слова сам тор является эллиптической кривой, у которой при мультипликативной униформизации решётка периодов становится недискретной. Когда $|q| \neq 1$, то получается настоящая эллиптическая кривая, точнее говоря, её представитель в мире некоммутативной геометрии. Когда же $|q| = 1$, получается чисто некоммутативный тор. Он является объектом, который морально эквивалентен предельной эллиптической кривой.

Тот эффект, что на пространствах модулей классических коммутативных объектов имеются граничные области, при приближении к которым мы, по-видимому, попадаем в мир некоммутативных объектов, сейчас становится всё более осознанным. Физики его давно поняли, а математики только сейчас начинают понимать. Есть очень интересная статья Капустина, Кузнецова и Орлова. Она связана с описанием аналогичных эффектов в пространстве модулей инстантонов.

Вообще, это явление широко распространено, и понимание того, что коммутативный мир и некоммутативный мир связаны не только посредством деформаций, но и посредством существования коридоров, по которым из коммутативного мира естественно попадаем в некоммутативный мир, было, вероятно, одним из самых интересных явлений, которые были осознаны математиками в последнее время (открыто это явление было уже лет 10 тому назад).

Более того, поскольку некоммутативный мир морально включает в себя коммутативный мир, то, вероятно, вся геометрия должна рассматриваться как некоммутативная. Нужно сделать глобальный пересмотр с некоммутативной точки зрения всего того, о чём мы привыкли думать в коммутативных терминах. Конечно, в коммутативной геометрии, в частности в алгебраической геометрии, есть очень сильная машина, которая

позволяет справляться с многими задачами, которые плохо решает некоммутативная геометрия. Одна из основных задач — сделать факторы по плохим отношениям эквивалентностями. Алгебраические геометры справляются с этим, изучая стэки.

У меня такое ощущение, что переход к буквально некоммутативной точке зрения доставляет совершенно новую интуицию, очень важную и заслуживающую изучения. С этой точки зрения то, что я здесь делаю, если слегка обобщить термины, есть изучение не деформированных абелевых многообразий, а торов, отфакторизованных по очень скверным решёткам периодов, которые содержат сильную недискретную часть. Некоммутативность, по-видимому, — это способ уловить механизмы такой факторизации.

29 мая 2000 г.

Я. Г. Синай

ДИНАМИКА АДИАБАТИЧЕСКОГО ПОРШНЯ (НАРУШЕНИЕ ВТОРОГО НАЧАЛА ТЕРМОДИНАМИКИ)

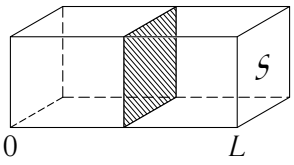
Прошлый раз я здесь выступал примерно год назад в той же самой аудитории. Сегодня я хочу рассказать про три работы: Пясецкий, Синай [1], Синай [2], Лебовиц, Пясецкий, Синай [3], а также их продолжние: Лебовиц, Синай, Чернов [4] и Нейштадт, Синай [5]. Дж. Лебовиц — американский математик и физик, один из лидеров статистической механики в мире. Н. И. Чернова многие могут знать по его замечательным работам в теории бильярдов. Ярослав Пясецкий — физик из Варшавы, известный специалист в неравновесной статистической механике.

Второй закон термодинамики в самой простейшей форме состоит в том, что если мы имеем замкнутую систему, в которой нет никаких внешних связей, то тогда с течением времени она монотонно приходит в состояние равновесия так, что если мы введём величину, которую естественно назвать *энтропией* (хотя единого определения энтропии и нет), то энтропия будет возрастать и окончательно в предельном состоянии будет принимать максимальное значение, а система будет в состоянии полного равновесия. Таково самое грубое описание второго закона термодинамики. Уже если вы посмотрите книгу Ландау и Лифшица по статистической механике, то там вы увидите различные комментарии на эту тему, смысл которых состоит в том, что в таком виде второй закон термодинамики совершенно неясен и является весьма тёмным утверждением, а кроме того, для того, чтобы процесс выхода на равновесие действительно можно было бы описать, необходимо привлекать и общую теорию относительности, и квантовую механику, т. е. в рамках классической механики второй закон термодинамики несправедлив, не выполняется и не является столь универсальным.

Одним из результатов этих работ, особенно третьей, будет пример того, что в ситуации, в которой, казалось бы, должен происходить выход на равновесие, по крайней мере некоторое время происходит осциллирующий режим. Что будет дальше, я не знаю. Это, по-видимому, трудный и интересный вопрос, но совершенно открытый.

Используя эти предварительные замечания, я хочу объяснить задачу. Задача элементарная; она должна была бы рассматриваться физиками

ещё в начале XX века. Но тогда физики были заняты открытием квантовой механики и многими другими проблемами; на эту проблему они особого внимания не обратили. А проблема вполне классическая. Представим себе, что имеется прямоугольный параллелепипед, все рёбра которого имеют одинаковый порядок. Например, на рис. 1 одно ребро имеет длину L, два другие ребра имеют длины L_1 и L_2. При этом L_1 и L_2 тоже порядка L. В дальнейшем L будет стремиться к бесконечности. Таким образом, мы будем рассматривать систему, в которой допустим *термодинамический предельный переход*.



Р и с. 1. Поршень

Представьте себе, что имеется перегородка, которая в этом докладе будет называться *поршнем*. Она расположена перпендикулярно одной из осей, и её площадь равна площади сечения параллелепипеда S. Эта перегородка представляет собой адиабатический поршень в том смысле, который я сейчас объясню.

Предположим, что слева и справа имеются два разных газа, которые находятся при разных термодинамических условиях. Я потом более точно объясню, что это значит.

Если давление слева больше, чем справа, то поршень, естественно, начнёт двигаться. Под словом «газ» будет пониматься ансамбль большого числа частиц. Поршень движется под действием столкновений с этими частицами. Слово «адиабатичность» в данном контексте означает, что у поршня нет никаких внутренних степеней свободы. Это — чисто механическое тело, которое передвигается в результате того, что частота столкновений с одной стороны превышает частоту столкновений с другой. Спрашивается, как найти траекторию поршня как функцию от параметров газов.

Я сейчас уточню оба понятия: что значит, что слева и справа заданы два разных газа, и сделаю предположение относительно массы поршня. Мы будем считать, что слева находится $N^{(-)}$ частиц, а справа $N^{(+)}$. Вообще, в дальнейшем обозначение $(-)$ будет относится к объектам слева от поршня, о обозначение $(+)$ будет относиться к объектам справа. Пусть, далее, общее число частиц равно N, а координата поршня равна $X(t)$. Тогда под плотностью газа слева и справа естественно понимать величины

$$\rho^{(-)} = \frac{N^{(-)}}{X(t)S}, \quad \rho^{(+)} = \frac{N^{(+)}}{(L - X(t))S}.$$

Такое число частиц, естественно, термодинамически большое, имеет некое случайное распределение. Мы будем рассматривать довольно широкий

класс распределений. Самое простое распределение слева от поршня, которое можно себе представить в этой ситуации, — это распределение, задаваемое функцией, скажем, $\pi^{(-)}(x, v)$ (первой корреляционной функцией). Смысл этой функции состоит в том, что интеграл

$$\int\limits_{C_1 \times C_2} \pi^{(-)}(x, v)\, dx\, dv \tag{1}$$

по какому-то множеству (C_1 лежит внутри параллелепипеда) есть среднее число частиц в этом множестве. Если C_1 — полоса, высекаемая каким-то интервалом на оси L, то тогда интегрирование происходит по отрезку, умноженному на площадь сечения. В этом случае C_1 будет пропорционально L^2, если интервал порядка 1. Скорости частиц принимают произвольные значения, поэтому по v множество C_2 — это обычное множество в трёхмерном пространстве.

Простейшее предположение состоит в том, что если мы знаем функцию $\pi^{(-)}(x, v)$, то тогда мы знаем всё распределение вероятностей. Например, в случае пуассоновского распределения, если мы возьмём множество $C_1 \times C_2$ и обозначим через $\nu(C_1 \times C_2)$ случайное число частиц в этом множестве, то тогда это число частиц имеет распределение Пуассона с параметром, который задаётся интегралом (1). Ещё более упрощающее предположение состоит в том, что функция $\pi^{(-)}(x, v)$ есть произведение функции от x на функцию от v; это означает, что координаты и скорости независимы. И ещё более упрощающее предположение состоит в том, распределение по v задаётся гауссовской плотностью; в этом случае мы имеем дело с максвелловским распределением. Если же функция $\pi^{(-)}(x, v)$ не зависит от x, то мы имеем дело с однородным распределением — распределение частиц вдоль x одно и то же.

На самом деле, мне понадобятся некоторые предположения об этой функции, но достаточно широкие. Я их сформулирую немножко позже.

Мне не обязательно требовать, чтобы распределение вероятностей было пуассоновским; можно требовать немножко меньшего. А именно, если область $C_1 \times C_2$ достаточно большая, то число частиц в этой области есть математическое ожидание плюс флуктуация, которая имеет порядок квадратного корня из объёма находящихся справа от поршня частиц. И это может быть сформулировано в терминах так называемой второй корреляционной функции, т. е. функции от четырёх переменных, которая задаёт плотность вероятности того, что в двух точках есть частицы.

Точно так же функцией $\pi^{(+)}(x, v)$ мы задаём распределение для числа частиц, находящихся справа. Тогда возникает статистика частиц. Динамика этих частиц — это динамика идеального газа, состоящая в том, что

каждая частица не взаимодействует с остальными частицами, а только сталкивается с поршнем. Через поршень некоторый вариант взаимодействия всё-таки происходит.

Хочу сразу заметить, что здесь можно думать о дальнейшем развитии всех этих исследований, состоящем в том, что можно отказаться от требований идеальности газа и рассматривать так называемый газ Кнудсена, когда длина свободного пробега частиц имеет тоже порядок L. То есть можно считать, что частицы имеют определённые размеры (частицы представляют собой твёрдые шарики); скорости частиц порядка 1, поэтому они сталкиваются на временах порядка L. В этой ситуации проблем типа эргодической теории для газа из твёрдых тел не возникает, поскольку нас интересует поведение через конечное, но большое число столкновений; мы не переходим к пределу по числу столкновений.

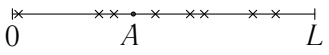
Я возвращаюсь к идеальному газу. Ещё один параметр, который я не указал и который играет чрезвычайно большую роль, — это масса поршня. Таким образом, даны $\pi^{(\pm)}(x, v)$ и $M_{\text{поршня}} = m$.

Первый случай, который разобран в нашей работе с Пясецким, таков.

С л у ч а й 1. Масса поршня m равна массе одной частицы идеального газа.

Это, конечно, не физический случай, но этот случай допускает полное исследование. Чтобы объяснить, что в этом случае происходит, мы перейдём от трёхмерной картинки к одномерной, потому что эффективно вся динамика здесь является одномерной; нас интересует только проекция скорости на направление, параллельное оси x. Поэтому рассмотрим просто отрезок длины L, на котором есть выделенная частица A, которую будем называть поршнем (рис. 2).

Слева и справа имеется идеальный газ, состоящий из невзаимодействующих частиц. Задача состоит в том, чтобы описать динамику точки A. Все частицы имеют одинаковую массу. По этой причине закон столкновений имеет очень простой вид: при каждом столкновении частицы обмениваются скоростями. Вот это и есть свойство одномерной динамики, которое сейчас будет работать. Кроме того, я для простоты предположу, что начальное распределение не зависит от x, т. е. $\pi^{(\pm)}(x, v) = \pi^{(\pm)} q^{(\pm)}(v)$.



Р и с. 2. Одномерная динамика

В этой ситуации наиболее интересен случай, когда давление справа равно давлению слева, а температуры разные. Сейчас я объясню, что это означает. Под давлением слева в данном случае естественно понимать среднюю энергию частиц, находящихся слева, а под давлением справа —

среднюю энергию частиц, находящихся справа:

$$E^{(-)} = \pi^{(-)} X(t) \int v^2 q^{(-)}(v)\, dv,$$
$$E^{(+)} = \pi^{(+)} (L - X(t)) \int v^2 q^{(+)}(v)\, dv.$$

Мы предположим, что в начальный момент $E^{(-)} = E^{(+)}$.

Температура, согласно правилам термодинамики, есть средний квадрат кинетической энергии одной частицы, т. е. с точностью до пропорциональности

$$T^{(-)} = \int v^2 q^{(-)}(v)\, dv, \quad T^{(+)} = \int v^2 q^{(+)}(v)\, dv.$$

Мы будем предполагать, что функции $q^{(-)}$ и $q^{(+)}$ непрерывны вместе с первой производной. Спрашивается, какой будет в этой ситуации динамика поршня.

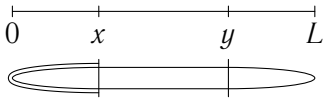
Первая теорема, которая здесь имеет место, состоит в том, что при такого рода динамике поршень движется и при $t \to \infty$ он асимптотически приходит в такое положение, при котором плотность на всём отрезке будет постоянной, т. е. среднее число частиц в предельном положении слева будет равно среднему числу частиц в предельном положении справа. Тем самым температура слева тоже будет равна температуре справа. Хотя в этом месте есть некоторые нюансы. Если под температурой понимать среднее значение квадрата скорости, то она действительно будет одной и той же слева и справа. Но под температурой нельзя понимать написанный выше интеграл, потому что распределение будет меняться.

Это как бы идеальный случай, где предсказания статистический механики выполняются именно в таком виде, как хотелось бы. Сейчас я постараюсь объяснить, откуда эта теорема получается. Для доказательства используется довольно простой аргумент из эргодической теории, или из теории вероятностей. Он состоит в следующем. Происходит динамика, и поршень двигается. Возьмём какую-то точку с координатой x на нашем отрезке и обозначим через $\mathcal{N}(x, t)$ количество частиц, находящихся левее точки x в момент времени t. Частицы двигаются, поэтому величина $\mathcal{N}(x, t)$ является случайной; для разных реализаций она может быть разной. Пусть $N(x, t) = E\mathcal{N}(x, t)$ — математическое ожидание этой величины. Напишем уравнение $N(x, t) = N^{(-)}$. Инвариант нашей динамики состоит в том, что когда поршень двигается, всегда слева от него остаётся одно и то же число частиц $N^{(-)}$. Частицы друг через друга как бы проскакивают, но от поршня они отражаются. Поэтому число частиц слева от точки A остаётся неизменным и число частиц справа от точки A тоже не меняется. Поэтому мы составляем уравнение $N(x, t) = N^{(-)}$, и решение этого уравнения даёт нам функцию $x(t)$, стремящуюся к пределу $x(t) \to x^{(\infty)}$ при $t \to \infty$.

Предел соответствует равномерному распределению. Истинное положение поршня отличается от $x(t)$ на величину порядка $\sqrt{N}$. Это частный случай утверждения о гауссовском распределении, что в таких ситуациях случайная величина отличается от своего математического ожидания на величину порядка $\sqrt{N}$.

Посмотрим, как мы будем исследовать эту случайную величину. Во всех задачах такого рода, когда скорости не меняются, естественно следить не за траекториями частиц, а за траекториями скоростей. Частицы испытывают столкновения, а скорости нет. Если частица сталкивается с поршнем, то они обмениваются скоростями, и после этого поршень начинает двигаться со скоростью той частицы, с которой он столкнулся. Поэтому скорость не чувствует столкновений, в том смысле, что носитель этой скорости движется равномерно. Поэтому каждая скорость v имеет такую траекторию: сначала она движется по отрезку длины L, затем отражается и движется назад со скоростью $-v$. Тем самым, она совершает периодическое движение на окружности длины $2L$ со скоростью v. Такова траектория каждой точки.

Поэтому удобнее считать, что у нас есть не один отрезок, а двойной отрезок (рис. 3). Каждая скорость движется по этому двойному отрезку (окружности длины $2L$). Теперь допустим, что на отрезке есть точка x, которая выделяет на окружности длины $2L$ дугу длины $2x$. Число частиц $\mathcal{N}(x, t)$ — это число тех скоростей, которые находятся в момент времени t внутри этой дуги.



Р и с. 3. Двойной отрезок

Чтобы правильно понимать эту задачу, нужно постоянно иметь в виду, что мы интересуемся временами t порядка L, т. е. такими временами, за которые каждая скорость делает достаточно большое число оборотов.

Я хочу посчитать число $\mathcal{N}(x, t)$. Для этого я возьму на отрезке некоторую координату y. Я буду считать, что положительные скорости находятся на одной (например, верхней) стороне двойного отрезка, а отрицательные на другой. Я хочу найти среднее число частиц, которые в начальный момент находились в точке y, затем сделали k оборотов по окружности, и в момент времени t оказались внутри дуги длины $2x$.

Результатом моих вычислений будет некий отрезок на оси v, который зависит от точки y. Длина этого отрезка будет себя вести как величина порядка $1/k$.

Мы имеем неравенства

$$L - y + 2Lk + L - x < vt < L - y + 2Lk + L + x,$$

т. е.

$$\frac{2L-y+2Lk}{t}-\frac{x}{t}<v<\frac{2L-y+2Lk}{t}+\frac{x}{t}.$$

Когда меняется k, у нас получаются разные отрезки с шагом $2L/t$. Когда t стремится к бесконечности по отношению к L, длина этого отрезка стремится к нулю. Внутри этого отрезка меня интересуют точки, которые образуют только его часть длины $2x/t$. Частицы, которые удовлетворяют нашему условию, имеют скорости, лежащие в арифметической прогрессии отрезков, идущих с шагом порядка $1/t$. Чтобы вычислить среднее число таких частиц, нужно взять плотность и умножить её на сумму длин этих отрезков. Но плотность по предположению является непрерывной функцией, поэтому эта сумма есть доля маленького отрезка внутри большого. Эта доля равна x/L. Иными словами, при $t\to\infty$ доля частиц, которые оказываются внутри этого отрезка, ведёт себя как x/L. Поэтому при $t\to\infty$ приближённо получаем уравнение $\frac{x}{L}=\frac{N^{(-)}}{N}$.

Сейчас происходит исследование такого уточнения этой задачи. Как я уже сказал, поршень движется с постоянной скоростью и приходит в положение равновесия. Интересно найти следующий член этой асимптотики, т. е. написать положение поршня в виде $X(t)=x(t)+\xi(t)$, где $x(t)$ — детерминированная компонента, $\xi(t)$ — случайная компонента; для этой случайной компоненты интересно найти предельное распределение.

Случай 2. Представьте себе, что есть отрезок [0, 1]; на нём есть одна большая частица массы M и некоторое (конечное) число частиц слева от неё и справа от неё. Предположим, что в начальный момент скорость массивной частицы (частицы с массой M) равна нулю. Тогда энергия этой системы не зависит от M. Рассмотрим динамику, при которой маленькие частицы не взаимодействуют между собой и сталкиваются с массивной частицей. Благодаря тому, что энергия системы сохраняется, для скорости массивной частицы получаем оценку $|V_M|\leqslant\sqrt{2H/M}$. Введём медленное время $\tau=t/\sqrt{M}$ и запишем скорость и координату массивной частицы как функции медленного времени: $V_M(\tau)$, $X_M(\tau)$. Спрашивается, какова будет динамика частицы в медленном времени.

Теорема. *При $M\to\infty$ динамика массивной частицы сходится к динамике одномерной классической частицы с гамильтонианом*

$$\mathcal{H}=\frac{p^2}{2}+\frac{C_1}{x^2}+\frac{C_2}{(1-x)^2},$$

где константы C_1 и C_2 зависят от начальных условий.

Эта одномерная гамильтонова система совершает осцилляции, совершает периодические движения. Тем самым, эта теорема означает, что при

$M \to \infty$ движение массивной частицы становится периодическим в неком отрезке, который создаётся динамически. Это не топологическая граница, а динамическая. Отрезок движется до некоторого места, затем возвращается, потом доходит до другого места и опять возвращается.

Как вы увидите, эта задача связана с третьей частью моего доклада. Швейцарский физик Кристиан Грубер показал мне очень красивый пример динамики такого типа. А именно, представьте себе, что имеется массивная частица, справа от которой есть одна частица, и слева тоже есть одна частица. В его примере массивная частица движется влево, потом встречается с левой частицей. Эти частицы некоторое время взаимодействуют; постепенно левая частица останавливает массивную частицу, и массивная частица начинает двигаться вправо. При движении вправо возникает последний момент, когда массивная частица соприкасается с левой частицей. В этот момент левая частица останавливается, а массивная частица движется вправо. Она встречает правую частицу, тоже с ней как-то взаимодействует, меняет направление скорости, и в последний момент лёгкая правая частица останавливается. Возникает такое забавное периодическое движение, которое тоже описывается приведённой выше теоремой.

Эта теорема относится к теории усреднения. В нашей недавней работе с Нейштадтом метод усреднения был использован для рассмотрения более общих систем лёгких частиц и с его помощью было показано, что периодические колебыния имеют весьма общий характер.

Сейчас будет рассматриваться наиболее реалистичный и наиболее интересный в этой теории случай.

С л у ч а й 3. Мы будем иметь дело с кубом со стороной L и будем считать, что масса поршня растёт как L^2, т. е. $M_{\text{поршня}} \equiv M_L \sim L^2$. Это как бы означает, что у поршня имеется поверхностная плотность. Массы частиц равны 1. Их распределения слева и справа разные. В частности, давление справа может быть больше давления слева, и поэтому поршень может двигаться. Из-за того что частиц много, статистика приводит к тому, что движение поршня в главном порядке детерминировано. Разность давлений порядка 1, поэтому скорость движения поршня порядка 1. Нас будет также интересовать, что происходит с поршнем за время порядка L, когда он проходит макроскопическое расстояние. И это совсем другой вопрос.

Я напишу формулы упругого столкновения, которые здесь используются, и по существу весь анализ, вся математика нашей задачи основаны на анализе этих формул.

Пусть $\varepsilon = \dfrac{2}{M_L + 1}$; мы будем считать, что масса стремится к бесконечности так, что $\varepsilon L^2 = a$ — положительная константа. Пусть, далее, V и v —

скорости поршня и частицы до столкновения, V' и v' — скорости поршня и частицы после столкновения. Закон столкновения частицы с поршнем выглядит так:

$$V' = (1-\varepsilon)V + \varepsilon v,$$
$$v' = -(1-\varepsilon)v + (2-\varepsilon)V.$$

Я сразу хочу сделать замечание, что поскольку ε — величина порядка $1/L^2$, мы будем пользоваться приближённой формулой $v' \approx -v + 2V$. При этом приближённом выражении не соблюдается закон сохранения энергии. Но ошибка, которую мы делаем, имеет следующий порядок малости. Ошибка скорости имеет порядок L^{-2}. Мы будем рассматривать времена немного больше L. Число столкновений будет при этом немного больше 1. Ошибка в положении частицы тоже несущественна.

Я хочу придерживаться обозначений нашей работы. В ней мы обозначаем первую корреляционную функцию для частиц слева и справа $p^{(\mp)}(x, v, t)$.

Поскольку поршень может двигаться только в направлении оси x, во всей динамике для нас существенна только проекция скорости на ось x. Другие проекции никакой роли не играют, поэтому ими можно пренебречь. По этой причине с самого начала можно считать, что газ состоит из частиц, у которых есть только горизонтальные скорости. Это будет та же самая задача. Таким образом, v можно считать одномерной переменной.

То, что происходит на поверхности поршня, тоже однородно. Неважно, где частица сталкивается с поршнем: вверху, внизу, посредине. По этой причине x тоже можно считать одномерной переменной.

При этом нужно иметь в виду, что для того, чтобы найти число частиц в интервале dx, dv, нужно дополнительно функцию q умножить на L^2; этот множитель даёт правильный порядок величины.

Сначала я выпишу результаты, а затем расскажу, как они получаются. Координату и скорость поршня будем обозначать $X_L(t)$ и $V_L(t)$. Основную роль будет играть функция $p^{(\mp)}(X_L, v, t)$, задающая распределение скоростей сталкивающихся частиц на самом поршне.

Введём следующие величины:

$$Q_0(t) = \int \operatorname{sgn} v\, p(X_L, v, t)\, dv,$$
$$Q_1(t) = \int |v|\, p(X_L, v, t)\, dv,$$
$$Q(t) = \int v^2 \operatorname{sgn} v\, p(X_L, v, t)\, dv.$$

Здесь $p = p^{(-\operatorname{sgn} v)}$.

Сделаем также ещё предположение, что $p^{(\mp)}(x, v, 0) = 0$, если $|v| < v_0$ или $|v| > v_1$. Это означает, что в начальный момент нет частиц, у которых

скорость достаточно маленькая, и нет частиц, у которых скорость достаточно большая. Это упрощение состоит в том, что в том режиме, который мы будем рассматривать, скорость поршня будет достаточно маленькой; поэтому, когда частица сталкивается с поршнем, она от него отражается и летит до левой границы и не может столкнуться с поршнем на коротких временах; перестолкновение (новое столкновение) происходит после того, как частица отразилась от стенки. Это некое упрощение. Но есть ощущение, что если рассматривать перестолкновения на коротких временах, то они должны вносить относительно небольшой вклад.

Во всех написанных выше интегралах существенно, что мы рассматриваем первые корреляционные функции на поршне (у этой функции есть два значения). Кроме того, они соответствуют значениям v до столкновения. Поэтому, скажем, когда v положительно, это означает, что мы рассматриваем частицу слева и берём функцию $p^{(-)}$. А когда v отрицательно, берётся функция $p^{(+)}$.

Интеграл $Q(t)$ очень важный. Он равен среднему значению энергии слева минус среднее значение энергии справа. Этот интеграл есть как раз разность давлений.

Первое утверждение состоит в том, что $V_L(t)$ допускает представление $V_L(t) = W_L(t) + \xi_L(t)$, где $W_L(t)$ — главный член, а погрешность (флуктуация) $\xi_L(t)$ стремится к нулю по вероятности. Главный член есть решение уравнения (2):

$$\frac{W_L(t)}{dt} = -2aQ_1(t)W_L(t) + a_0Q_0(t)W_L^2(t) + aQ(t). \tag{2}$$

Константа a здесь та самая, что была определена выше.

Это уравнение типа уравнения Риккати; квадратичное. Его нужно понимать так. Здесь Q_0, Q_1 и Q — функционалы от p, а p зависит от W_L. Поэтому уравнение неполное. Мы должны ещё написать уравнения для первых корреляционных функций.

Поскольку мы имеем дело с идеальным газом, для первых корреляционных функций очень просто пишутся дифференциальные уравнения внутри объёма и чуть сложнее пишутся условия на границе. Дифференциальное уравнение внутри объёма имеет вид

$$\frac{\partial p^{(\mp)}}{\partial t} + \frac{\partial p^{(\mp)}}{\partial x}v = 0. \tag{3}$$

Это можно называть одномерным кинетическим уравнением, уравнением для свободной динамики — как угодно. Это уравнение описывает динамику свободных частиц.

Граничные условия такие:

$$p^{(-)}(0,\, v,\, t) = p^{(-)}(0,\, -v,\, t). \tag{4}$$

Граничные условия на поршне такие:

$$p^{(\mp)}(X_L,\, v,\, t) = p^{(\mp)}(X_L,\, 2W_L(t) - v,\, t). \tag{5}$$

Система уравнений (2), (3), (4), (5) уже является полной. Она описывает динамику первых корреляционных функций, затем динамику поршня (в которой коэффициенты зависят от первых корреляционных функций). Кроме того, она содержит граничные условия.

Давайте рассмотрим следующий частный случай. Предположим, что наша система является стенкой между двумя пространственно однородными газами (полубесконечными), т. е. $p^{(+)}$ и $p^{(-)}$ зависят только от v. Это означает, что слева есть газ при одном давлении и справа тоже есть газ при каком-то давлении. В этом случае интегралы Q_0, Q_1 и Q не будут зависеть от времени, и мы получим уравнение с постоянными коэффициентами. Чтобы понять, что происходит, предположим, что величина Q_0 маленькая. Тогда главная часть уравнения (2) состоит из двух членов. Из этого выражения видно, что при $t \to \infty$ (в обычном времени, без всякой нормировки) $W_L(t)$ выходит на предел, который является корнем этого уравнения. Иными словами, когда слева находится один газ, а справа находится другой газ, то поршень через какое-то небольшое время начинает двигаться с постоянной скоростью, причём значение этой скорости находится как корень уравнения (2), где $W_L(t) = 0$. Оказывается, что это обстоятельство имеет гораздо более общий смысл.

Утверждение состоит в следующем. Произведём преобразование масштаба $X/L = \tilde{x}$ (новая координата меняется в пределах $0 \leqslant \tilde{x} \leqslant 1$) и перенормируем время: $t/L = \tilde{t}$. Это означает, что у нас есть ящик размера 1, в котором имеется стенка, разделяющая газы с разной плотностью. Спрашивается, как будет двигаться эта стенка в нормальных переменных. Это, конечно, наиболее важный вопрос. Чтобы ответить на этот вопрос, мы должны предположить, что первые корреляционные функции тоже являются функциями медленных переменных, т. е. функции $p_L^{(\mp)}\left(\frac{X}{L},\, v,\, t\right)$ имеют предел при $L \to \infty$. Тогда эти предельные функции действительно описывают положение поршня. Но если это так, то производные по настоящему времени от $p_L^{(\mp)}$ являются величинами порядка $1/L$; они являются производными по медленному времени (при любом масштабе). Из этого вытекает чрезвычайно важное обстоятельство. Обозначим через

$W_L^{(0)}(t)$ один из двух корней квадратного уравнения (2) с нулевой левой частью:

$$W_L^{(0)}(t) = \frac{Q_1}{Q_0}\left(1 - \sqrt{1 - \frac{QQ_0}{Q_1^2}}\right) = \frac{Q}{2Q_1}\left(1 + O\left(\frac{QQ_0}{Q_1^2}\right)\right).$$

В главном порядке величина $W_L^{(0)}(t)$ пропорциональна разности давлений. Утверждение таково: $W_L(t)$ можно записать в виде $W_L(t) = W_L^{(0)}(t) + W_L^{(1)}(t)$, где величина $W_L^{(1)}(t)$ имеет порядок $1/L$. Иными словами, благодаря структуре этого уравнения оказывается, что скорость поршня есть однозначно определённый функционал от первых корреляционных функций, точнее говоря, функционал от коэффициентов Q_0, Q_1 и Q. Это утверждение — обобщение того утверждения, которое я сделал в том случае, когда коэффициенты Q_0, Q_1 и Q — константы. Если эти коэффициенты константы, то решение выходит на некий предел. В общем случае они зависят от t, но являются медленно меняющимися функциями от t. Поэтому главная часть решения есть $W_L^{(0)}(t)$, а погрешность имеет порядок $1/L$. Когда мы переходим к перенормированному времени, погрешность стремится к нулю. Поэтому в перенормированном времени координата поршня есть однозначная функция от Q_0, Q_1 и Q. В нашей работе с Лебовицем и Черновым предложен строгий вывод упомянутой выше формулы для динамики поршня, работающий на временах порядка нескольких столкновений с поршнем.

Когда давления справа и слева равны, скорость поршня имеет порядок не 1, а меньше. Только когда есть разность давлений, скорость поршня имеет порядок 1. Ответ на вопрос, какова будет скорость поршня в случае, когда давления справа и слева равны, а температуры не равны, не известен. Имеется работа Лебовица, в которой утверждается, что в этом случае скорость будет иметь порядок $1/L^3$.

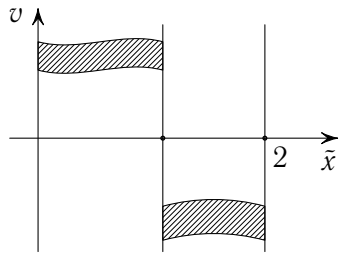
Теперь я хочу применить весь этот аппарат к объяснению того, с чего я начал. А именно, неверно думать, что в этой ситуации при $t \to \infty$ поршень будет монотонно стремиться к положению равновесия, которое будет отвечать равномерному распределению плотности. Для этого нет никаких причин, по крайней мере на временах порядка L или немножко больше.

Динамика поршня, как я сейчас пытаюсь объяснить, в каком-то смысле ближе к динамике конечного числа частиц, которую я упомянул раньше. Пусть давление слева больше, чем давление справа. Поршень движется направо. При этом он сжимает газ, который находится справа. Из-за этого давление справа растёт. В какой-то момент скорость поршня меняет знак.

Он начинает двигаться налево; сжимает газ, который находится слева; там давление начинает расти. Таким образом возникают флуктуации. Нет никаких результатов, которые объясняли бы асимптотическое поведение поршня в этом случае. Нет примера типа примера Грубера, в котором динамика поршня была бы периодична.

Кстати говоря, Пясецкий нашёл старый учебник термодинамики, учебник Кэллена начала XX века, в котором задача о поршне обсуждается, и там прямо есть такая фраза, что нет никаких причин, по которым поршень должен приходить в положение равновесия. Мы пришли к этой же точке зрения, а потом уже нашли это высказывание.

Сейчас я сделаю некую конструкцию и поясню, каким образом можно увидеть эти осцилляции. Мне не нравится трёхмерность оставшейся ситуации. Я хочу описать одномерную ситуацию гидродинамического типа, которая возникает в результате некой редукции исходной задачи. Рассмотрим плоскость, в которой есть координата $\tilde{x}$ и скорость v (рис. 4). Исходная перенормированная длина была равна 1; я рассмотрю отрезок длины 2.



Р и с. 4. Редукция к одномерному случаю

Конструкция, которую я хочу сделать, состоит в следующем. Координата поршня на рассматриваемой плоскости есть $2W$, где W — координата поршня на исходной плоскости. То, что я сейчас хочу построить, — это изоморфизм исходной динамической системы на некую другую динамическую систему. Проведём вертикальную черту, соответствующую положению поршня. Слева и справа от неё имеется газ. Слева частицы будут иметь только положительные скорости. Для этого и нужна конструкция. Я хочу сделать такое изменение системы координат. Пусть частица слева столкнулась с поршнем и стала двигаться направо. В момент столкновения я хочу считать, что её координата равна нулю, а скорость равна минус её истинной скорости. Иными словами, закон столкновения (закон взаимодействия с поршнем) такой. Если частица прошла через поршень со скоростью v, то в следующий момент времени у неё будет скорость $v - 2W$.

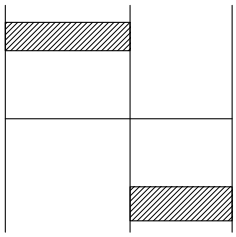
Динамика системы такова. Предположим, что у нас есть область слева и область справа, занятые газом. Представьте себе, что это непрерывный газ (идеальный газ, которым занимается газовая динамика). В каждой точке есть плотность газа. Газ движется с известной скоростью слева направо. На границе частицы газа меняют скорость v на скорость $-v + 2W$,

где W — скорость поршня. Точно так же частицы справа перескакивают в точку с координатой v.

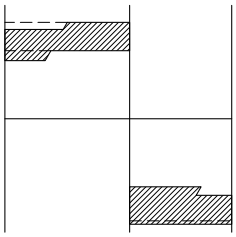
Чтобы найти координату поршня, нужно взять плотности газа на этой прямой, для v положительных и для v отрицательных; посчитать величины Q_0 и Q_1, а затем найти скорость перегородки. Это и есть динамика замкнутой системы, которая эквивалентна динамике поршня.

Мы рассматриваем динамику на достаточно больших временах, при которых скорость поршня меньше скорости каждой частицы, но порядок один и тот же. Этот порядок времён больше времени нескольких столкновений одной частицы с границей.

Рассмотрим случай, когда газ был равномерно распределён внутри одного прямоугольника слева и внутри другого прямоугольника справа (рис. 5). Проследим за частицами слева. Они начинают сталкиваться.



Р и с. 5. Распределение газа



Р и с. 6. Распределение газа после одного столкновения

Затем они имеют меньшую скорость. Через некоторое время у нас не будет частиц около максимума, потому что частицы, которые переходят, имеют меньшую скорость. Наоборот, появляются частицы, которые имеют скорость меньше минимума, потому что W положительно. По той же причине у частиц справа скорость будет увеличиваться. Поскольку поршень движется вправо, частицы, которые здесь добавляются, достигнут поршня раньше. Поэтому в какой-то момент времени появится ситуация, когда все эти частицы достигли поршня. Тогда давление частиц справа станет преобладать над давлением частиц слева. Поэтому можно определить момент времени, в который давление справа равно давлению слева; в этот момент меняется направление скорости поршня. После этого он начинает двигаться влево. Справа появляются частицы с маленькими скоростями и исчезают частицы с большими скоростями, поэтому давление справа уменьшается.

Рисуя диаграммы, можно посчитать несколько первых осцилляций поршня для перенормированных времён порядка 1. Асимптотика движения поршня на больших временах неизвестна.

Литература

[1] J. Piasecki, Ya. G. Sinai. A model of non-equilibrium statistical mechanics // Dynamics: models and kinetic methods for non-equilibrium many body systems (Leiden, 1998). — Dordrecht: Kluwer Acad. Publ., 2000. — P. 191—199. — (NATO Sci. Ser. E Appl. Sci.; V. 371.).

[2] Я. Г. Синай. Динамика массивной частицы, окружённой конечным числом лёгких частиц // Теорет. и мат. физика. — 1999. — Т. 121, № 1. — С. 110—116.

[3] J. L. Lebowitz, J. Piasecki, Ya. Sinai. Scaling dynamics of a massive piston in an ideal gas // Hard ball systems and Lorentz gas. — Berlin: Springer, 2000. — P. 217—227. — (Encycl. Math. Sci.; V. 101.).

[4] Дж. Л. Лебовиц, Я. Г. Синай, Н. И. Чернов. Динамика массивного поршня, поруженного в идеальный газ // Успехи матем. наук. — 2002. — Т. 57, вып. 6(348). — С. 3—86.

[5] A. I. Neishtadt, Ya. G. Sinai. Adiabatic piston as a dynamical system // J. Statist. Phys. — 2004. — Vol. 116, № 1—4. — P. 815—820.

7 июня 2000 г.

С. И. Гельфанд

О ЧИСЛЕ РЕШЕНИЙ КВАДРАТНОГО УРАВНЕНИЯ

1.

Квадратное уравнение для матриц второго порядка

Квадратное уравнение $x^2 + px + q = 0$, $p, q \in \mathbb{C}$, в общем случае (если $p^2 - 4q \neq 0$) имеет два корня, а если $p^2 - 4q = 0$, то оно имеет один корень. Я не буду вдаваться в тонкости алгебраической геометрии, что здесь один корень — это на самом деле тоже два корня, но они совпадают. Кроме того, есть формула, выражающая корни через p и q. Но это пока не очень важно, потому что если степень уравнения более высокая, то факты о количестве корней всё равно имеются, а формулы нет.

Теперь рассмотрим такое же уравнение $X^2 + PX + Q = 0$, где P, Q и X — матрицы 2×2 над полем $\mathbb{C}$. Первый вопрос — сколько корней может иметь такое уравнение? Здесь есть два разных вопроса. Во-первых, какие вообще есть возможности (для скалярного квадратного уравнения число корней может быть равно 1 или 2). Во-вторых, сколько корней в случае общего положения (для скалярного квадратного уравнения в случае общего положения 2 корня).

Для начала я разберу два примера. Рассмотрим уравнение $X^2 = 0$. Решениями этого уравнения являются матрицы $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, для которых $a + d = 0$ и $ad - bc = 0$. Таких матриц бесконечно много; они образуют 2-мерный конус. Этот пример показывает, что бывает бесконечно много решений.

Рассмотрим теперь уравнение $X^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Это уравнение не имеет решений. Если квадрат матрицы имеет только нулевые собственные значения, то она сама тоже имеет только нулевые собственные значения. Это означает, что матрица нильпотентна. А квадрат любой нильпотентной матрицы 2-го порядка равен нулю.

Итак, может быть 0 решений и может быть бесконечно много решений. Но видно, что это случаи специальные. Один из способов попытаться выяснить, что будет в случае общего положения, такой. Перепишем урав-

нение в скалярном виде. Тогда будет 4 неизвестных; для них выполняются 4 уравнения 2-й степени. Согласно теореме Безу, если уравнения в общем положении, то должно быть $2^4 = 16$ решений. Но в действительности уравнения не в общем положении, и число решений не 16.

Сейчас я попробую объяснить, как решать квадратное уравнение для матриц. Я должен сказать, что то решение, которое я приведу, в каком-то смысле неудовлетворительно. Меня и многих других оно не удовлетворяет. Например, Арнольд тоже считает это решение неудовлетворительным. Решение алгебраическое, а геометрически непонятно, почему ответ такой, а не другой.

Вторая вещь, о которой я хочу сказать, совершенно удивительная. Это то, что меня поразило, и почему я решил здесь об этом рассказать. Задача в таком виде — решать квадратные уравнения для матриц — кажется абсолютно классической. Было бы естественно, что если уж не раньше, то хотя бы в XIX веке её должны были бы по крайней мере сформулировать, да и решить, потому что там ничего не требуется. Тем не менее, я спрашивал многих людей об этой задаче. Мне давали разные советы, где это можно было бы поискать. Кое-где я смотрел, но я нигде не смог найти даже упоминания об этой задаче. Единственное место, которое я ещё не посмотрел, — такого рода задачи, может быть, возникают в теории оптимального управления и в теории обыкновенных дифференциальных уравнений.

Даже если были и более ранние решения, то решение, которое я сейчас расскажу, придумал независимо от всех более ранних попыток Дмитрий Борисович Фукс два или три года тому назад. Он тоже решал не эту задачу, а чуть другие задачи, о которых я потом упомяну.

Этот метод решения годится и для алгебраических уравнений любой степени над квадратными матрицами произвольного порядка. Можно отдельно рассматривать уравнение $X^2 + XP + Q = 0$; одно сводится к другому транспонированием. Наиболее общее уравнение $X^2 + P_1X + XP_2 + Q = 0$. Тот метод, о котором я буду рассказывать, к такому уравнению неприменим. (Точнее говоря, уравнение $X^2 + P_1X + XP_2 + Q = 0$, вероятно, можно преобразовать к виду $X^2 + PX + Q = 0$. Но для уравнений более высокого порядка этого нельзя сделать.) Это одна из причин, почему этот метод неудовлетворителен.

Пусть $\lambda^2E + \lambda P + Q = A(\lambda)$; здесь E — единичная матрица, λ — комплексное число. Напишем скалярное уравнение $\det A(\lambda) = 0$. Это уравнение в общем положении имеет 4-ю степень относительно λ. В общем положении у него есть 4 корня λ_1, λ_2, λ_3, λ_4. Число λ является корнем уравнения $\det A(\lambda) = 0$ тогда и только тогда, когда матрица $A(\lambda)$ вырожденная, т. е. когда у неё есть нетривиальное ядро. Мы будем считать, что

все 4 корня различны и для каждого корня $\dim \ker A(\lambda_i) = 1$. Это — некоторые условия общности положения. Выберем в каждом ядре по одному ненулевому вектору: v_1, v_2, v_3, v_4. Будем считать, что все эти векторы не пропорциональны, т. е. $\mathbb{C}v_i \neq \mathbb{C}v_j$. Это ещё одно условие общности положения. Оно выполняется не автоматически; четыре числа λ_i могут быть различными, но два из соответствующих им векторов могут быть одинаковыми. То, что векторы могут совпадать, как раз и даёт вырождения в числе решений.

Итак, получилось четыре различных комплексных числа и четыре попарно линейно независимых вектора, соответствующих этим числам. Если матрицы P и Q таковы, что выполняются сформулированные выше условия общего положения, то существует ровно $C_4^2 = 6$ решений квадратного уравнения, соответствующих выборам двух пар (λ_i, v_i) из данных четырёх пар.

Решения квадратного уравнения строятся следующим образом. Пусть мы выбрали пары (λ_i, v_i) и (λ_j, v_j). Построим матрицу $X_{(ij)}$ 2-го порядка, для которой $X_{(ij)}v_i = \lambda_i v_i$ и $X_{(ij)}v_j = \lambda_j v_j$. Эти условия однозначно задают матрицу $X_{(ij)}$, поскольку векторы v_i и v_j образуют базис, а линейное преобразование однозначно задаётся образами базисных векторов. Явный вид матрицы $X_{(ij)}$ таков. Пусть $v_i = \begin{pmatrix} v_{i1} \\ v_{i2} \end{pmatrix}$ и $v_j = \begin{pmatrix} v_{j1} \\ v_{j2} \end{pmatrix}$. Тогда

$$X_{(ij)} = \begin{pmatrix} v_{i1} & v_{j1} \\ v_{i2} & v_{j2} \end{pmatrix} \begin{pmatrix} \lambda_i & 0 \\ 0 & \lambda_j \end{pmatrix} \begin{pmatrix} v_{i1} & v_{j1} \\ v_{i2} & v_{j2} \end{pmatrix}^{-1}.$$

Легко проверить, что матрица $X_{(ij)}$ является корнем нашего квадратного уравнения. Для этого достаточно проверить, что матрица $X_{(ij)}^2 + PX_{(ij)} + Q$ переводит линейно независимые векторы v_i и v_j в нуль. Это следует из того, что действие матрицы $X_{(ij)}$ на векторы v_i и v_j известно и что по условию $A(\lambda_i)v_i = 0$. Выбирая разные пары (λ_i, v_i), (λ_j, v_j), мы получим 6 решений квадратного уравнения.

У п р а ж н е н и е. Доказать, что если условия общего положения выполнены, то других решений нет.

Для решения этого упражнения нужно использовать то, что у любой матрицы есть хотя бы один собственный вектор, соответствующий собственному значению.

Посмотрим, как всё это выглядит геометрически. Введём два алгебраических многообразия. Многообразие E (от слова equation) — множество всех уравнений 2-го порядка над матрицами 2-го порядка. Оно состоит из пар матриц (P, Q) и представляет собой 8-мерное аффинное пространство $\mathbb{C}^8$. Второе многообразие S (от слова solution) состоит из троек (X, P, Q), где X — матрица 2-го порядка, для которой $X^2 + PX + Q = 0$.

Это многообразие нелинейное. Оно лежит в $\mathbb{C}^{12}$ и задаётся там четырьмя уравнениями. Есть также естественная проекция $S \to E$ (мы забываем X). Задача состоит в том, чтобы описать слои над разными точками. Это эквивалентно задаче о числе решений квадратного уравнения.

Проекция $S \to E$ устроена довольно хитро. У общей точки есть ровно шесть прообразов. Однако есть точки, у которых нет прообразов, и есть точки, у которых прообразов бесконечно много. Я ещё хочу отметить, что бесконечные слои бывают разных типов. Я приводил пример, когда слой — конус. Но есть примеры, когда слой — гиперболоид, т. е. гладкое многообразие. (Точнее говоря, получается гиперболоид плюс две точки.) По-видимому, конус является вырождением семейства гиперболоидов.

Из-за того, что ситуация здесь аффинная, а не проективная, здесь возможны разные интересные явления.

Хотелось бы иметь какой-то разумный алгоритм, который позволяет сказать, сколько есть решений у данного уравнения. Для скалярного уравнения алгоритм простой: если дискриминант равен нулю, то решение одно, а если дискриминант не равен нулю, то решений два. Есть и другой алгоритм, для решения более близкой задачи, — для определения жордановой нормальной формы матрицы. Там тоже можно написать нечто вроде $A(\lambda)$ и есть элементарные преобразования, которые приводят такую матрицу к каноническому виду; этот канонический вид определяет жорданову нормальную форму. Но, конечно, таких простых инвариантов быть не может. Там должны быть какие-то более тонкие инварианты.

Насколько я понимаю, ответ на вопрос, что здесь вообще может быть, такой. Я, правда, не всё умею доказывать. Но если этим не занимались классики английской алгебраической школы середины или конца XIX в., то этого никто и не знает. По-видимому, ситуация здесь такая: может быть любое число решений от 0 до 6, а также два типа бесконечности (конус и гиперболоид). В частности, обе бесконечности двумерные. Я не смог построить пример, когда слой одномерен. Но я не знаю, почему слой не может быть одномерным.

Сейчас я приведу примеры с числом решений 0, 2, 4, 5 и 6, а также примеры, когда решения — конус и гиперболоид. Примеров с числом решений 1 и 3 я не построил, но они, по-видимому, есть. Если уравнение имеет конечное число решений, то это число не больше 6. Это следует из общих результатов алгебраической геометрии: если есть отображение двух аффинных многообразий, и у общей точки 6 прообразов, то подскок может быть только в размерности, а не в числе прообразов.

Как из шести решений получить пять? У нас есть числа λ_1, λ_2, λ_3, λ_4 и им соответствуют векторы v_1, v_2, v_3, v_4. Предположим (и этого дей-

ствительно можно добиться), что $v_1 = v_2$, а все остальные векторы не пропорциональны (числа λ_i по-прежнему все различны). Тогда не существует матрицы $X_{(12)}$, потому что она должна один и тот же вектор одновременно умножать на λ_1 и на λ_2. Поэтому из шести решений одно пропадает (уходит на бесконечность; в проективной картинке оно остаётся).

Дальше можно написать, что одновременно $v_1 = v_2$ и $v_3 = v_4$ (числа λ по-прежнему все различны). Тогда останется четыре решения.

Можно также попытаться написать $v_1 = v_2 = v_3$, но тогда придётся искать три различных корня у одного и того же скалярного квадратного уравнения. Поэтому для матриц второго порядка ситуаций, когда $v_1 = v_2 = v_3$, не бывает. По этой причине я не умею строить примеры, когда решений ровно три.

Пример с двумя решениями строится вручную. Рассмотрим уравнение $X^2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Его решениями служат матрицы $\begin{pmatrix} \pm 1 & 0 \\ 0 & 0 \end{pmatrix}$. Других решений нет, потому что у каждого решения должны быть два различных собственных значения: 0 и ± 1. Такая матрица должна быть диагональна.

Пример с гиперболоидом и двумя отдельными точками строится так. Рассмотрим уравнение $X^2 = E$. Решение X имеет собственные значения ± 1. Если оба собственных значения совпадают, то решение не может быть жордановой клеткой. Поэтому решения с совпадающими собственными значениями — это $X = E$ и $X = -E$. Остальные решения — матрицы с собственными значениями 1 и -1, не обязательно диагональные. Такие матрицы сопряжены с матрицей $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Если квадрат матрицы сопряжён с единичной матрицей, то он сам равен единичной матрице.

Всё это делалось достаточно кустарно. Недостаток всего этого состоит в том, что геометрически это понять нельзя.

2.

Уравнение произвольной степени для матриц произвольного порядка

Рассмотрим уравнение $X^m + A_1 X^{m-1} + \ldots + A_m = 0$, $A_i, X \in \mathrm{M}_n(\mathbb{C})$. Тем же способом, о котором я говорил, легко доказать следующее утверждение.

Теорема 1. *В общем случае число решений равно* C_{mn}^n.

Доказательство такое. Уравнение $\det A(\lambda) = 0$ имеет степень mn, и каждый выбор n корней этого уравнения даёт одно решение рассматриваемого уравнения.

3.
Квадратное уравнение над некоммутативной алгеброй

Теперь я немного расскажу о некоторых вещах, которыми в последнее время занимались Израиль Моисеевич Гельфанд и Володя Ретах, иногда вместе с другими людьми. Речь идёт о некоммутативных аналогах простейших алгебраических конструкций, связанных с решением уравнений от одного переменного, с линейной алгеброй и т. д. Понятно, что когда работают не с числами, а с элементами некоммутативной алгебры, многое из того, что делается в коммутативном случае, автоматически перестаёт работать в некоммутативной ситуации. На первый взгляд — почти всё. Удивительная вещь, которую они обнаружили, — это то, что существует много утверждений, которые можно сохранить, переформулировав некоторым образом.

Основным моим объектом по-прежнему будет квадратное уравнение $x^2 + px + q = 0$ (потом — уравнение более высокой степени), но на этот раз p, q, $x \in A$, где A — некоммутативная (но ассоциативная) алгебра над $\mathbb{C}$. Если всё делать строго, то на алгебру A нужно наложить некоторые ограничения; для неё должны выполняться некоторые теоремы о существовании алгебр частных. Я, насколько мне это удастся, буду жульничать и заметать всё это под ковёр.

Прежде всего посмотрим, какие есть результаты, которые мы могли бы надеяться сохранить. Первый результат — у квадратного у равнения не может быть больше двух корней. Мы видели, что это неверно. В случае, когда алгебра A некоммутативна, вполне может быть больше двух корней. Однако есть вариант этой теоремы, который верен. Я не знаю автора этой теоремы и даже не знаю доказательства (в том доказательстве, которое я знал, обнаружился пробел). Но я знаю ссылку: верное доказательство этой теоремы есть в книге [1].

Т е о р е м а 2. *Пусть x_1, x_2, x_3 — три корня уравнения $x^2 + px + q = 0$, причём выполняются следующие условия: $x_1 \neq x_2$, $x_1 - x_3$ и $x_2 - x_3$ обратимы в A. Тогда у этого уравнения есть бесконечно много корней.* (*Алгебра A здесь не обязательно должна быть над алгебраически замкнутым полем*; *но поле обязательно должно быть бесконечным*.)

Посмотрим, как эта теорема соотносится с задачей о матрицах, которую мы обсуждали перед этим. Там были решения $X_{(ij)}$, i, $j = 1, \ldots, 4$. Посмотрим, можем ли мы построить противоречащий пример. Рассмотрим разность $X_{(ij)} - X_{(kl)}$. Если один из индексов i, j совпадает с одним из индексов k, l, то у матриц $X_{(ij)}$ и $X_{(kl)}$ есть общий собственный вектор с

одним и тем же собственным значением, поэтому разность этих матриц необратима. Матрица $X_{(ij)} - X_{(kl)}$ обратима тогда и только тогда, когда индексы k, l дополнительны к индексам i, j. Поэтому противоречия с теоремой 2 нет.

Теорема, аналогичная теореме 2, есть и для уравнений произвольной степени n. Она состоит в том, что если у уравнения степени n есть $n+1$ решение, причём первые n решений различны, а их разности с $(n+1)$-м решением обратимы, то у уравнения есть бесконечно много решений. Здесь тоже видно, что задача о числе решений матричного уравнения степени n для матриц порядка m не приводит к противоречию.

Это первая вещь, которая является некоммутативным аналогом известной коммутативной теоремы. Она была известна уже давно.

4.
Теорема Виета

Дальше я буду рассматривать квадратные уравнения. У скалярного квадратного уравнения над полем $\mathbb{C}$ есть три свойства. Одно свойство состоит в том, что у квадратного уравнения не больше двух корней. Как это обобщается, я уже сказал. Другие свойства таковы.

1) Если x_1 и x_2 — корни квадратного уравнения $x^2 + px + q = 0$, то $x_1 + x_2 = -p$ и $x_1 x_2 = q$ (теорема Виета).

2) Если x_1 и x_2 — корни, то $x^2 + px + q = (x - x_1)(x - x_2)$.

В оставшейся части лекции я расскажу, как эти свойства переносятся на уравнения с некоммутирующими неизвестными и коэффициентами. Здесь при формулировках всегда нужно иметь в виду, что обратимо, а что необратимо. Это обычно довольно сложно. Простейший способ всего этого избежать — работать в свободной алгебре, т. е. считать, что необратим только 0. Чтобы не попасть в противоречие, нужно использовать результаты Коэна о реализации алгебр и т. д. Я этого делать не буду, а буду просто считать, что когда я делю на какой-то элемент, этот элемент обратим. А если он необратим, то всегда можно формально расширить алгебру, добавив обратный к нему элемент. В некоммутативном случае, правда, возникают всякие сложности, например, должно выполняться условие Оре. Тем не менее, будем считать, что никаких сложностей нет.

Чтобы сформулировать первый результат, относящийся к теореме Виета, введём обозначения

$$x_{1;2} = (x_1 - x_2)x_2(x_1 - x_2)^{-1} \quad \text{и} \quad x_{2;1} = (x_1 - x_2)x_1(x_1 - x_2)^{-1}.$$

Для уравнений высокой степени возникают аналогичные более сложные

выражения. В любой такой записи правильный индекс — это тот, который стоит после точки с запятой; если бы алгебра была коммутативна, то только он бы и выжил. Все такие формулы нужно проверять, стирая всё то, что стоит перед точкой с запятой. Эти обозначения ввели Гельфанд и Ретах.

Т е о р е м а 3. *Если x_1 и x_2 — корни квадратного уравнения, то*

$$x_1 + x_{1;2} = -p = x_2 + x_{2;1},$$
$$x_{1;2}x_1 = \ q \ = x_{2;1}x_2.$$

В этой теореме есть ещё утверждение о том, что $x_1 + x_{1;2} = x_2 + x_{2;1}$ и $x_{1;2}x_1 = x_{2;1}x_2$. Это не зависит от p и q.

Д о к а з а т е л ь с т в о. Известно, что $x_1^2 + px_1 + q = 0$ и $x_2^2 + px_2 + q = 0$. Вычтем одно равенство из другого: $x_1^2 - x_2^2 - p(x_1 - x_2) = 0$. Домножив на $(x_1 - x_2)^{-1}$, получим $(x_1^2 - x_2^2)(x_1 - x_2)^{-1} = -p$. Затем в первом выражении в скобках нужно добавить и вычесть x_1x_2 или x_2x_1 (в одном случае получится одно выражение для p, а в другом другое). Например, добавим и вычтем x_1x_2:

$$\begin{aligned} -p &= (x_1^2 - x_1x_2 + x_1x_2 - x_2^2)(x_1 - x_2)^{-1} = \\ &= (x_1(x_1 - x_2) + (x_1 - x_2)x_2)(x_1 - x_2)^{-1} = \\ &= x_1 + (x_1 - x_2)x_2(x_1 - x_2)^{-1}. \end{aligned}$$

Если одно из выражений для p подставить в одно уравнение, то получится одно выражение для q, а если другое выражение для p подставить в другое уравнение, то получится другое выражение для q. □

Если взять произвольные x_1 и x_2 и определить p и q такими формулами, то получим квадратное уравнение, корнями которого являются x_1 и x_2. Поэтому формулы, которые верны для корней, будут верны и для произвольных x_1 и x_2.

Теорема 3 обобщается и на уравнения высокой степени над матрицами. Именно это делалось в работе Д. Б. Фукса и А. С. Шварца. Это у них была не самоцель; они применяли это к какой-то физике, по-видимому, по предложению Шварца. Их работа есть на сети (`math.RA/9410207`). Они доказали часть теоремы Виета для матриц, а затем использовали результат Амицура о том, что если что-то верно для матриц любых порядков, то это верно всегда. Хотя эта идеология противоречит реальному положению дел. Неверно, что матрицы высокого порядка — это то же самое, что свободная алгебра. В матрицах высокого порядка квадратное уравнение имеет много решений, а в свободной алгебре — не больше двух. Поэтому чем больше мы увеличиваем порядок матриц, тем дальше отдаляемся от свободной алгебры, по крайней мере в этой ситуации.

В общем случае теорема Виета формулируется следующим образом. Пусть $x_1, \ldots, x_m$ — корни уравнения $x^m + a_1 x^{m-1} + \ldots + a_m = 0$ над некоммутативной алгеброй. Введём элементы $x_{F;i}$, где $F \subset \{1, \ldots, m\}$ и $i \notin F$, следующими рекуррентными соотношениями. Положим $x_{\varnothing;i} = x_i$, а дальше

$$x_{F\cup j;i} = (x_{F;i} - x_{F;j}) x_{F;i} (x_{F;i} - x_{F;j})^{-1}.$$

Можно проверить, что $x_{F;i}$ не зависит от порядка элементов F. Правильное определение $x_{F;i}$ использует теорию того, что Гельфанд и Ретах называют квазидетерминантами. (Это теория о том, как определять детерминанты для матриц с некоммутирующими элементами. Это отдельная очень красивая теория, но на неё у меня сейчас нет времени.) Напишем теперь $y_k = x_{1,\ldots,k-1;k}$ (в коммутативном случае $y_k = x_k$).

Теорема 4 (обобщённая теорема Виета). *Если у уравнения* $x^m + a_1 x^{m-1} + \ldots + a_m = 0$ *есть корни* $x_1, \ldots, x_m$, *то*

$$a_l = (-1)^l \sum_{1 \leqslant k_1 < \ldots < k_l \leqslant m} y_{k_l} y_{k_{l-1}} \ldots y_{k_1}.$$

Как следствие получаем, что такие суммы не зависят от того, в каком порядке берутся корни уравнения, хотя сами числа y_k зависят от порядка. Упорядочения корней можно понимать как действие симметрической группы порядка m. Теорема о том, что эти суммы инвариантны относительно действия симметрической группы, есть более или менее часть теоремы о симметрических функциях в некоммутативном случае. У этой теоремы есть и вторая более сложная часть: любой симметрический многочлен есть многочлен от этих элементарных симметрических функций. Некий вариант этой теоремы верен и в некоммутативном случае. Можно явно охарактеризовать те симметрические функции, который являются полиномами от коэффициентов соответствующего уравнения. Это будут не все симметрические функции, но они образуют подкольцо, которое как линейное пространство изоморфно кольцу обычных коммутативных симметрических функций. А как устроены все симметрические функции, не известно.

5.
Разложение на множители

Последнее, о чём я хочу рассказать, — как разлагать на множители. Из того, что у теоремы Виета для квадратного уравнения есть два варианта (можно брать x_1 и $x_{1;2}$, а можно — $x_{2;1}$ и x_2), видно, что и разложений должно быть два. К сожалению, универсальное разложение на множители, когда ничто ни с чем не коммутирует, выглядит не очень красиво. Поэтому

я сначала сформулирую утверждение в частном случае, когда переменная x коммутирует с алгеброй коэффициентов A, т. е. мы работаем в алгебре полиномов $A[x]$.

Теорема 5. *Если переменная x коммутирует с коэффициентами, то*

$$x^2 + px + q = (x - x_{1;2})(x - x_1) = \\ = (x - x_{2;1})(x - x_2).$$

Формально эту теорему нужно понимать так. Пусть B — алгебра, порождённая элементами x, p, q, x_1, x_2 и соотношениями $x_1^2 + px_1 + q = 0$, $x_2^2 + px_2 + q = 0$, $xp = px$, $xq = qx$. Тогда в алгебре B с добавленным частным $(x_1 - x_2)^{-1}$ выполнены соотношения из условия теоремы 5. Как всегда в таких вещах, трудный вопрос — доказать, что эта алгебра ненулевая.

Если x не коммутирует с A, то тоже есть некий аналог разложения. Он переходит в то, что было написано, если наложить условия коммутирования. Но выглядит он чуть более сложно.

Теорема 6. *В общем случае $x^2 + px + q = (z^{(1)} - x_{1;2})(x - x_1)$, где $z^{(1)} = (x - x_1)x(x - x_1)^{-1}$.*

Я здесь выписал одно из двух разложений. Аналогичные результаты есть и для уравнений произвольной степени m. Для уравнения степени m получается $m!$ разложений. Удобнее всего их выписывать на основе теории квазидетерминантов; рекуррентные соотношения для z записываются сложно.

Единственный известный мне случай, когда всё это имеет какие-то приложения, связан с работой в кольце дифференциальных операторов, а особенно с работой в кольце дифференциальных операторов с матричными коэффициентами. Тогда результаты такого сорта полезны для исследования интегрируемых систем.

Литература

[1] T. Y. Lam. First course in non-commutative algebra. — New York: Springer-Verlag, 1991. — (Graduate Texts in Mathematics; V. 131).

7 сентября 2000 г.

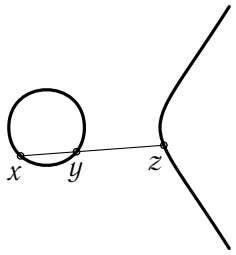
Ю. И. Манин

ПРОБЛЕМА МОРДЕЛЛА—ВЕЙЛЯ ДЛЯ КУБИЧЕСКИХ ПОВЕРХНОСТЕЙ

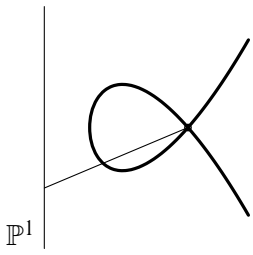
Я хочу рассказать вам сюжет, который меня волнует уже лет 30 и в котором медленно капают продвижения, причём обычно не те, которых я ожидаю и даже не те, которые я хотел бы. Этот сюжет постепенно развивается, но тем не менее основные задачи остаются нерешёнными. Часть этой истории описана в моей старой книге «Кубические формы». Другая часть — компьютерный эксперимент, который никогда нигде толком не описывался. Ещё одна часть описана в недавнем препринте Димы Каневского и моём [1].

Я начну с общих обозначений. Можно рассматривать произвольные размерности, но интересные результаты есть только для кубических кривых и для кубических поверхностей.

Если задана плоская кубическая кривая и на этой кривой есть три коллинеарные (лежащие на одной прямой) точки x, y и z (рис. 1), то мы это будем записывать так: $x = y \circ z$ (любая точка есть композиция двух других), или $x = t_y(z)$ (точка x получается из точки z «отражением» относительно точки y). Так определённое t_y — некое бирациональное отображение кривой в себя; оно удовлетворяет соотношению $t_y^2 = 1$. Конечно, нужно рассматривать разные частные случаи. Например, если имеет место касание, т. е. точка x слилась с точкой y, то нужно написать $x = x \circ z$, и т. д.



Р и с. 1. Кубическая кривая



Р и с. 2. Рациональная параметризация

Если кубическая кривая негладкая (рис. 2), то вся эта игра тоже имеет смысл. Но нужно избегать, чтобы одна из коллинеарных точек была двойной. Для двойной точки возникает произвол в выборе направления соответствующей прямой. Как известно, в этом случае возникает стереографическая проекция. Если рассмотреть прямую $\mathbb{P}^1$ на бесконечности

и проводить прямые через точки $\mathbb{P}^1$ и двойную точку, то получится рациональная параметризация кубической кривой с двойной точкой. Таким образом, вне двойной точки будут иметь место обыкновенные соотношения коллинеарности и возникают обычные бирациональные отображения, а сама двойная точка используется для доказательства того, что кубическая кривая с двойной точкой рациональная, а не эллиптическая.

Всё это — сводка определений коллинеарности трёх точек (включая вырожденные случаи), закона композиции и бирационального отображения. Нужно отметить, что если y, $z \in V(k)$ (т. е. эти точки рациональны над некоторым полем k), то $x \in V(k)$.

Рассмотрим теперь кубическую поверхность. Кубические поверхности я буду рассматривать только гладкие. В случае общего положения ситуация та же самая. Если на кубической поверхности выбрать две точки x и y, то прямая, проходящая через них, пересекает поверхность ещё в одной точке z. Можно использовать точно такие же обозначения $x = y \circ z$ и $x = t_y(z)$. Здесь t_y — бирациональное отображение $V \to V$. Оно тоже удовлетворяет соотношению $t_y^2 = 1$. Снова выполняется свойство, что если две точки рациональны, то третья точка пересечения тоже будет рациональна.

Однако здесь возникает сложность в случае, когда одной из коллинеарных точек разрешается быть двойной. Если взять произвольную точку кубической поверхности и провести через неё касательную плоскость к поверхности, то в сечении, вообще говоря, получится кубическая кривая с двойной точкой. Эта двойная точка согласно конструкции будет рациональна. В этом случае разумно считать, что $x \circ x$ — любая точка кривой, по которой пересекает кубическую поверхность касательная плоскость в точке x. Операция композиции перестаёт быть однозначной в том случае, когда она применяется к совпадающим точкам. Но в общей точке бирациональное отображение определено.

1.
Алгебра

Кривые. Для случая неособых кривых хорошо известно, что если фиксировать произвольную точку $u \in V(k)$ и ввести операцию $x + y = u \circ (x \circ y)$, то получится коммутативная группа. Например, если уравнение кривой записано в форме Вейерштрасса, то кривая симметрична относительно оси абсцисс и на бесконечности есть точка с тройным касанием; если фиксировать эту бесконечно удалённую точку, то получим обычное описание сложения точек на кубической кривой. Но можно взять и произвольную точку.

Операция $\circ$ не есть групповой закон; но если её заменить указанным образом, то получится групповой закон с нулём u. Классическая теорема Морделла—Вейля имеет несколько почти эквивалентных формулировок (предполагается, что поле k конечно порождённое).

1. Группа $V(k)$, в которой задан групповой закон $+$, конечно порождённая.

2. В $V(k)$ можно выбрать конечный набор точек $(x_1, \ldots, x_k)$ так, что все остальные точки $V(k)$ получаются из них методом секущих и касательных.

3. Группа бирациональных автоморфизмов $\mathrm{Bir}\, V$ конечно порождена и содержит подгруппу конечного индекса, порождённую всеми t_x, $x \in V(k)$.

Вопрос, который я сначала хочу поставить в неопределённой форме, а затем его последовательно уточнять, таков: «В какой мере эти утверждения могут быть распространены на кубические поверхности?» Для эллиптических кривых эти три утверждения почти эквивалентны; различие между ними совсем небольшое. Сразу нужно сказать, что когда кривая имеет простейшую двойную точку, теорема о конечной порождённости перестаёт быть верной. Как я уже сказал, стереографическая проекция даёт изоморфизм кривой (минус двойная точка) с $\mathbb{P}^1$ без двух точек, соответствующих касательным в двойной точке. Эти две точки можно считать нулём и бесконечностью; получается изоморфизм с однородным пространством без одной точки. Если вы перенесёте закон композиции с кривой на это однородное пространство, то получится просто умножение. Таким образом, если на кривой есть двойная точка, то в смысле этого закона композиции группа $V(k)$ есть просто k^*, а группа k^* не конечно порождённая, даже когда $k = \mathbb{Q}$. Конечная порождённость резко нарушается в случае, когда есть двойная точка.

Поверхности. Можно было бы надеяться на аналог конечной порождённости группы $(V(k), \circ)$ в случае поверхностей, но какова бы ни была его формулировка (а мы сейчас увидим, что есть несколько гипотетических формулировок), она должна что-то делать с особыми кубическими кривыми на поверхности. Тот случай, который для кубических кривых был нетипичен, на кубических поверхностях прямо непосредственно живёт: там есть много подструктур, замкнутых относительно композиции $\circ$ и определённо не являющихся конечно порождёнными: это пересечения $V(k)$ с касательными плоскостями к $x \in V(k)$. Нужно либо считать, что операция $\circ$, применённая к одной и той же точке, многозначна (т. е. $x \circ x$ имеет много значений), либо сделать факторизацию $(V(k)/\mathscr{U}, \circ)$ так, чтобы операция $\circ$ стала однозначной, и тогда уже спрашивать о конечной порождённости отфакторизованного множителя, либо нужно что-то ещё. Гипотетический

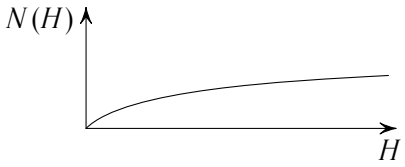
аналог конечной порождённости требует исследования, как его правильно сформулировать.

С формулировкой (2) всё обстоит примерно так же, а формулировка (3) интересным образом совершенно перестаёт быть верной. Точнее говоря, элементы t_x порождают группу бирациональных автоморфизмов $\mathrm{Bir}\, V$, но она не является конечно порождённой. В двумерном случае помимо образующих t_x известна полная система соотношений: $t_x^2 = 1$, $(t_x t_y t_{x\circ y})^2 = 1$. Последнее соотношение доказывается так. В нём фигурируют три коллинеарные точки. Поэтому если вы примените бирациональное отображение $t_x t_y t_{x\circ y}$ к четвёртой точке z, то вы будете работать с четырьмя точками, которые лежат в одной плоскости. Эта плоскость высекает некую кубическую кривую, поэтому требуемое соотношение достаточно проверить для кубической кривой. С точностью до некоторых мелочей (точки могут оказаться на одной из 27 прямых, лежащих на кубической поверхности) это полная система соотношений. Группа, которая получается таким образом, доказуемо не является группой конечного типа. Грубо говоря, мы получаем бесконечное свободное произведение групп $\mathbb{Z}/2\mathbb{Z}$; если точек бесконечно много, то группа заведомо не может быть конечно порождённой. Поэтому теорема Морделла—Вейля в такой форме неверна.

Теперь, прежде чем исследовать этот сюжет теоретически, я расскажу численный эксперимент, который производился на протяжении многих лет. Его целью было выяснить, можно ли верить в конечную порождённость. Я работал с кубической поверхностью $x_1^3 + 2x_2^3 + 3x_3^3 + 4x_4^3 = 0$, постепенно наращивая количество имеющейся информации. Началось всё с того, что на каком-то стареньком компьютере в 70-м или 71-м году я стал считать, сколько на этой поверхности есть точек ограниченной высоты. Высота точки P с целыми координатами (без общего делителя), лежащей на этой поверхности, равна $h(P) = \sum |x_i|$. Меня интересовало количество точек множества $\{P\colon h(P) \leqslant H\}$ в зависимости от H. Прежде всего я начал составлять список точек, лежащих на этой поверхности. Получилась таблица, которая начиналась следующим образом.

Номер решения	x_1	x_2	x_3	x_4	Высота
1	1	0	1	−1	3
2	1	1	−1	0	3
3	−1	1	1	−1	4
4	3	1	1	−2	7
5	−5	4	−1	0	10

Из классического способа доказательства теоремы Морделла—Вейля вытекает, что на кубической гладкой кривой количество решений высоты, не превосходящей H, приблизительно равно $c(\log H)^{r/2}$, где r — ранг свободной части (свободной абелевой группы). Поэтому график функции $N(H)$, где $N(H)$ — количество элементов множества $\{P\colon h(P) \leqslant H\}$, должен был бы выглядеть так, как на рис. 3. Тогда очень быстро точки перестали бы появляться, и компьютерный эксперимент пришлось бы закончить. Но оказалось, что функция растёт линейно. За 30 лет, прошедших с тех пор, линейный рост $N(H)$ стал чёткой гипотезой; известна даже теоретическая константа C, для которой $N(H) \approx CH$. Эта константа получается при экстраполяции константы из метода Харди—Литтлвуда (тригонометрических сумм) назад на заданное малое число переменных; в некоторых примерах иногда появляются ещё дополнительные неклассические множители, связанные с порядком группы Брауэра. Но сама эта гипотеза до сих пор не доказана.



Р и с. 3. График $N(H)$ для кривой

Гипотеза о линейном росте потом была обобщена на произвольные многообразия Фано. В каких-то случаях она доказана. Для очень точной формулировки (с точной степенью логарифма, в размерности 3 или больше) есть контрпример. Была довольно большая волна работ, связанных с оценкой порядка роста. Но, повторяю, ни для исходного примера и вообще ни для одной кубической поверхности такая асимптотика не доказана.

Вопросу о конечной порождённости вся эта информация не помогает, поскольку в таблице не содержится информации о композициях точек. Чтобы извлечь какую-то информацию о конечной порождённости, нужно добавить информацию о композициях. Здесь можно поступать грубо. Допустим, что вы решили, что всё порождается первыми тремя точками. Тогда можно написать программу, которая складывает эти точки, затем складывает их с полученными результатами и т.д., и посмотреть, будет ли это всё покрывать. Так, в принципе, можно сделать и даже так делалось, но это совершенно не поучительно, потому что не видно, что на самом деле происходит.

Я делал по-другому. Самая ранняя попытка была такая. Предположим, что я хотел бы экспериментально установить справедливость теоремы Морделла—Вейля для кубических поверхностей, пользуясь этим методом. Тогда я попытался бы представить каждую точку достаточно большой высоты в виде композиции точек меньшей высоты. На самом

деле, доказательство теоремы Морделла—Вейля так и идёт. Центральный факт, который доказывается, таков: если у вас есть точка высоты больше некоторой константы, то её можно представить в виде композиции двух точек меньшей высоты. Тогда по индукции вы сваливаетесь к конечному числу точек.

Поэтому я попытался записать в таблицу данные следующего типа. Будем кодировать точки поверхности их номерами. Возьмём точку с номером N и вычислим все точки $N \circ m$, $m \leqslant N$. Как правило, вы будете залетать куда-то дальше. Но время от времени вы будете получать что-то, что уже встречалось раньше. Если вы получили соотношение $N \circ m = n$, где $n < N$, то тогда в силу симметрии отношения коллинеарности $N = m \circ n$, где $m \leqslant N$ и $n < N$. Все такие разложения запишем в таблицу. Начало таблицы выглядит следующим образом.

N	
1	неразложимо
2	неразложимо
3	неразложимо
4	$1 \circ 2$ и $3 \circ 3$
5	$2 \circ 2$ и $3 \circ 3$
6	$1 \circ 1$ и $3 \circ 3$
...	...
30	неразложимо

Здесь имеется в виду, что первая после $N = 3$ точка, не имеющая разложений такого типа, имеет номер $N = 30$. Эта точка сопротивлялась разложению примерно 10 лет. Раз уж она такая упорная, я выпишу её координаты: $(-15, 37, -5, 29)$.

Если бы я переставил точку с номером 3 на первое место, то оказалось бы, что две другие точки через неё выражаются. Тем самым, все точки до 29-й включительно выражаются через точку с номером 3. Есть гипотеза, что точка 3 — базис.

Таблицу в такой форме сделал по моей просьбе Цагир. В этой таблице одновременно увеличивалось количество разложений (если они есть; встречались очень длинные строчки разложений) и количество неразложимых точек. Количество неразложимых точек увеличивается примерно линейно. Сразу стало видно, что доказывать теорему о конечной порождённости для кубических поверхностей тем же самым способом, как и для кубических кривых (т. е. представляя точку большой высоты в виде композиции *двух* точек меньшей высоты), по-видимому, безнадёж-

но. Таблица не показывает никакой тенденции к уменьшению количества неразложимых точек.

На некоторое время я застрял в этой ситуации, не зная, что делать. Потом мне пришло в голову, что можно применить следующий трюк. Повторяю, что мне не хотелось прямым перебором составлять полную таблицу всех разложений. Имеющуюся таблицу можно обрабатывать так. Предположим, что в ней появилась первая неразложимая точка (например, номер 30). Я стал смотреть, не встречается ли она где-нибудь дальше в двойных разложениях. Допустим, что имеет место равенство $287 \stackrel{?}{=} 27 \circ 30 \stackrel{?}{=} 16 \circ 18$ (я привожу это равенство в качестве гипотетического примера). Тогда получаем $30 = 27 \circ (16 \circ 18)$, т. е. точка 30 представляется в виде комбинации не двух точек меньшей высоты, а трёх. Но чтобы извлечь это, нужно посмотреть на таблицу гораздо дальше того места, где находится точка с номером 30. Оказалось, что этот метод чрезвычайно плодотворно работает. Если у вас есть достаточно большой список такого типа, то огромный кусок начала всегда оказывается разложимым. Но иногда встречаются очень длинные минимальные разложения. Нужно также учитывать, что слова неассоциативные, т. е. нужно правильно расставлять скобки.

Этим способом мне в конце концов удалось разложить точку с номером 30. Мораль стала такая: по-видимому, невозможно делать индукцию по высоте. Вы получаете длинные разложения, в которые входят только точки меньшей высоты, но промежуточные неассоциативные слова выкидывают вас далеко вперёд. Разглядывание этой таблицы убедило меня в том, что стандартная стратегия, работающая для эллиптических кривых, здесь неприменима. Это, конечно, не доказательство. Но, по-видимому, это верно.

Я убил довольно много времени, пытаясь как-то сымитировать стандартное вейлевское доказательство, но ничего, кроме каких-то тривиальных вещей, получить не смог. Дело в том, что вейлевское доказательство в его самой точной форме основано, как вы знаете, на том обстоятельстве, что на множестве точек эллиптической кривой функция $x \mapsto \log h(x)$ является приблизительно евклидовой метрикой. На самом деле, с точностью до константы её можно изменить так, что она станет в точности евклидовой метрикой на пространстве $V(k) \otimes R$. На этом, собственно, и основано всё доказательство. Это означает, что вы понимаете, как ведёт себя высота (или логарифм высоты), когда вы складываете две точки. Грубо говоря, получается сумма, умноженная на косинус угла. Ничего похожего для кубических поверхностей, по-видимому, не имеет места.

Для кубической поверхности, указанной выше, помимо гипотезы о линейном росте $N(H)$ имеется следующая гипотеза: всё $V(\mathbb{Q})$ порождено

одной точкой $(1, -1, -1, 1)$. Эти две совершенно конкретные теоретико-числовые гипотезы до сих пор не доказаны. Я считаю это крупным вызовом в теории диофантовых уравнений, потому что, с одной стороны, они относятся к простейшим диофантовым уравнениям после эллиптических кривых, а с другой стороны, они относятся почти что к рациональным поверхностям (кубическая поверхность становится рациональной после конечного расширения основного поля). Кубическая поверхность — это почти что проективная плоскость. Наличие для таких поверхностей диофантовых вопросов конкретного вида, которые вполне классические и могли бы быть поставлены в XIX веке и на которые мы не умеем отвечать, свидетельствует о том, что в нашем образовании есть пробелы.

2.
Отрицательные результаты

Я выпишу для сравнения теоремы, связанные со свойством конечности, для кубических кривых и для кубических поверхностей.

Кривые. Пусть V — гладкая кубическая кривая.

а) Группа бирациональных преобразований $\mathrm{Bir}\, V$ является полупрямым произведением конечной группы и подгруппы, порождённой $\langle t_x t_y \colon x, y \in V(k)\rangle$. (Я мог бы взять и группу, порождённую $\langle t_x \colon x \in V(k)\rangle$, но тогда t_x попадает в первую группу. А так мы получаем полупрямое произведение.)

б) Имеют место соотношения $t_x^2 = (t_x t_y t_{x\circ y})^2 = 1$.

в) Если поле k конечно порождённое, то группа $\mathrm{Bir}\, V$ конечно порождённая (теорема Морделла—Вейля).

г) Все точки $V(k)$ можно получить методом секущих и касательных из некоторого конечного набора точек.

Поверхности. Пусть V — минимальная гладкая кубическая поверхность. Минимальность означает следующее. Известно, что над алгебраическим замыканием основного поля на кубической поверхности есть 27 различных прямых. Бирациональным преобразованием каждую из них можно стянуть в точку. Минимальность означает, что никакой набор этих прямых нельзя стянуть над основным полем, т. е. группа Галуа действует таким образом, что у каждой прямой есть сопряжённая, которая её пересекает.

Для поверхностей первый факт тоже верен.

а′) Группа $\mathrm{Bir}\, V$ — полупрямое произведение конечной группы (группы проективных автоморфизмов) и подгруппы, порождённой следующими вещами: $\{t_x \colon x \in V(k)\}$ и $\{s_{u,v} \colon u, v \in V(K),\ u = \bar{v},\ [K:k] = 2\}$, где

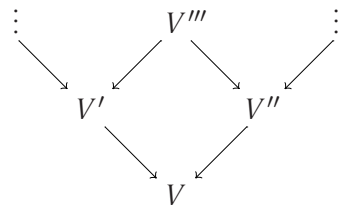
$s_{u,v} = t_u t_{uov} t_v$. (Сами точки u и v определены над квадратичным расширением K, но автоморфизм $s_{u,v}$ определён над исходным полем k.)

б′) Имеют место соотношения $t_x^2 = (t_x t_y t_{xoy})^2 = (s_{u,v})^2 = 1$.

в′) Система соотношений из б′) полная. Поэтому $\mathrm{Bir}(V)$ не является конечно порождённой группой, если $V(k)$ бесконечно.

Полнота системы соотношений и конечная порождённость в некотором смысле дополнительны друг другу. Для кривых абелевость и конечная порождённости влекут неполноту системы соотношений. Для поверхностей полнота системы соотношений влечёт невозможность группе быть конечно порождённой.

Я скажу об этом всего несколько слов, потому что это результат старый. Доказательство полноты системы соотношений приведено в моей книге «Кубические поверхности»; оно занимает страниц 10. Это доказательство достаточно прямолинейное, но довольно муторное. Доказательство комбинаторное. Оно основано на рассмотрении естественного бесконечномерного представления группы бирациональных отображений. Для группы группы бирациональных отображений нет очевидных линейных объектов, на которых она была бы представлена. Неочевидный объект бесконечного типа получается так. Нужно рассмотреть многообразие V и рассматривать всё бóльшие морфизмы сверху вниз:



Отображение $V \to V$ можно представить в виде композиции раздутия и стягивания. Рассмотрев бесконечную башню, можно изготовить нечто, на чём действует группа бирациональных отображений. В данном случае это предел $\varprojlim \mathrm{Pic}(V')$ относительно взятия обратных образов. Это будет бесконечномерная решётка, на которой действует группа автоморфизмов. Чтобы эту решётку описать какой-то разумной системой образующих, нужно применить классический язык линейных систем с предписанными особенностями. Нужно взять группу $\mathrm{Pic}\ V \oplus \left(\bigoplus_{x \in V} \mathbb{Z}x\right)$ и посмотреть, куда переводит бирациональное отображение b стандартную линейную систему: $b^*(\omega_x^{-1}) \in \mathrm{Pic}\ V \oplus \left(\bigoplus_{x \in V} \mathbb{Z}x\right)$. У вас получится линейная комбинация какой-то кратности самого ω и нескольких точек. После этого вы смотрите, что происходит с коэффициентами этой ли-

нейной комбинации, выделяете максимальную особенность и т. д. Это стандартная процедура, которая теперь разработана и в многомерной ситуации.

Рассматривая это бесконечномерное представление, мы в его терминах можем для любого бирационального автоморфизма написать некое каноническое слово. Для этого нужно выделить максимальную особенность и подкрутить её с помощью образующих. В результате получается нечто более простое; затем процедура повторяется. Потом доказывается, что если есть два элемента группы, то каноническое слово для их произведения получается из произведения канонических слов применением только этих соотношений.

Причина того, что на этом уровне нет теоремы конечности для $V(k)$, состоит в том, что мы не проводим факторизации, не считаем получающимися за один приём сразу все точки рациональных кривых, о которых шла речь выше. Нужно учитывать не только сами эти рациональные кривые (сечения касательными плоскостями), но и всё, что получается из них применением полной группы бирациональных преобразований. Получается огромная сеть рациональных кривых, которые в каком-то смысле достаются дёшево. Но нужно рассматривать ещё и образы всех этих неизвестных априори точек. Из-за этого получается сложная сеть, которая имеет инфинитный характер.

3.
«Положительный результат»

Положительный результат здесь в кавычках, потому что это просто некая идея и притянутый за уши пример того, что она в каком-то смысле работает.

Вместо того чтобы представлять теорему ad hoc, я начну с противоположного конца — с рассмотрения некой теоремы конечности, в которой в явном виде фигурируют такие большие рациональные кривые. В этой форме теорема почему-то немногим известна, хотя она заслуживает того, чтобы её знали.

Рассмотрим множество точек проективной плоскости $\mathbb{P}^2(k)$ над полем k. Применим следующую конструкцию. Пусть есть конечное подмножество $S \subset \mathbb{P}^2(k)$. Начнём с ним делать примерно то же самое, что мы делали с точками кубической поверхности. Через две точки $a, b \in S$ проведём прямую l_{ab}, через другие две точки $c, d \in S$ проведём прямую l_{cd}, и добавим к S их точку пересечения. Затем эту операцию можно повторить и т. д. Это четверной закон композиции: $l_{ab} \cap l_{cd} = *(a, b; c, d)$.

Теорема 1. *Если поле k конечного типа, то можно выбрать конечное подмножество $S \subset \mathbb{P}^2(k)$ так, что в результате получим все точки $\mathbb{P}^2(k)$, т. е. $\mathbb{P}^2(k)$ конечно порождено в смысле этого закона композиции.*

Эта теорема — переформулировка результата о классификации абстрактных проективных плоскостей. Абстрактная проективная плоскость — это множество точек и множество прямых. Каждая прямая — подмножество в множестве точек. Любые две прямые пересекаются ровно по одной точке. Через любые две точки проходит ровно одна прямая. Должны также выполняться аксиомы Дезарга и Паппа. Это — определение абстрактной проективной плоскости. Классическая теорема состоит в том, что любая абстрактная проективная плоскость реализуется как конкретная проективная плоскость $\mathbb{P}^2(k)$ для некоторого поля k. В классической теореме обычно явно не формулируется, что морфизм проективных плоскостей индуцируется морфизмом векторных пространств. Но это нужно; это там неявно содержится.

Как из теоремы классификации выводится теорема 1? Это делается чрезвычайно просто. Возьмём какое-нибудь конечное множество $S \subset \mathbb{P}^2(k)$ так, чтобы в нём было по крайней мере 4 неколлинеарных точки. Породим этим множеством всё, что можно. Понятно, что то, что в результате получится, будет абстрактной проективной плоскостью. Стало быть, оно получается из конкретной проективной плоскости посредством вложения $\mathbb{P}^2(k') \subset \mathbb{P}^2(k)$. Это означает, что есть вложение $k' \subset k$, и какое-то линейное отображение координат осуществляет вложение проективных плоскостей. Для k есть конечное число образующих над простым полем. Будем последовательно добавлять по одной точке, первая аффинная координата которой — нужная образующая. Легко видеть, что в результате всё будет получено.

Эта приятная теорема конечности явно немножко похожа на то, что хочется здесь сделать. Меня некоторое время мучил вопрос, нельзя ли придумать какой-то трюк. Такой трюк я придумал. Это, конечно, обман. Но этот обман, на мой взгляд, заслуживает внимания и дальнейшего рассмотрения. Обман состоит в том, что я меняю определения. Пусть V — гладкая кубическая поверхность. Я введу на ней новый закон композиции, зависящий от параметра. Прежде чем его формулировать, давайте проанализируем, как устроена композиция $x \circ y$. Сначала я проведу через точки x и y произвольную плоскость. В сечении получится кубическая кривая, содержащая эти точки. Затем на этой плоскости через точки x и y я проведу прямую. Ясно, что закон композиции не зависит от того, какую именно плоскость я выбираю.

Это можно сформулировать следующим образом. Выберем пару точек $(x, y) \in (C, p)$, где C — кубическая кривая, заданная вместе с плоским вложением p (плоской структурой), и определим операцию $x \circ y$ как операцию $x \circ y$ по отношению к этой плоской структуре. До сих пор я работал со стандартными плоскими структурами. В качестве C я выбирал плоское сечение. Туда кривая укладывалась, и в этой плоской структуре я работал. Теперь я рассмотрю случай, когда V не минимально и даже допускает бирациональный морфизм $V \to \mathbb{P}^2$. Пусть есть какое-то плоское сечение $C \subset V$. Тогда его можно отобразить в $\mathbb{P}^2$ посредством бирационального морфизма, который мы фиксируем раз и навсегда. Будем складывать точки кубической поверхности в изменённой плоской структуре: для точек $x, y \in C$ будем рассматривать композицию, получая $\mathbb{P}^2$ не из естественного вложения, а из отображения, связанного с проекцией. Результат теперь уже будет зависеть от того, какую именно кривую C мы выберем для точек x, y. Такую композицию будем обозначать $x \circ_{(C,p)} y$.

Понятно, что я делаю подгонку под теорему конечности, которая у меня есть.

Теорема 2. *$V(k)$ конечно порождено относительно закона композиции $\circ_{(C,p)}$.*

Доказательство этой теоремы состоит в том, что операция, которая была на проективной плоскости, с помощью морфизма p переносится на V. Там получается операция над четырьмя точками. Можно подробно посмотреть, во что превращаются прямые на плоскости, поднятые вверх на V. Если они не проходят через раздуваемые точки, то они превращаются в кубики (пространственные, косые) на V. Четверная операция на плоскости превращается в четверную операцию на точках $V(k)$, состоящую в следующем. Через точки $a, b \in V(k)$ проводим единственную косую пространственную кубику, затем проводим кубику через точки $c, d \in V(k)$, и находим точку пересечения этих кубик. В действительности эту же самую точку можно представить в виде $a \circ_{(C,p)} b$. Поэтому из конечной порождённости относительно четверной операции следует конечная порождённость относительно композиции $\circ_{(C,p)}$. Получается даже несколько более сильное утверждение: мы не используем применение композиции к одной и той же точке, т. е. не рассматриваем $a \circ_{(C,p)} a$.

Очень бы хотелось следующим шагом это хозяйство как-то сократить. Мы получаем, что слова вида $(x_{i_1} \circ_{(C_1,p)} x_{i_2}) \circ_{(C_2,p)} (\ldots)\ldots$ исчерпывают все наши точки. А хотели бы получить, что их исчерпывают такие же слова, но с обыкновенным законом композиции, а не с законом композиции в смысле (C, p). Здесь есть разные трюки, простейший из которых состоит в том, что $(x \circ_{(C,p)} y) \circ_{(C,p)} z = (x \circ y) \circ z$, т. е. если стоящие рядом C_1 и

C_2 совпадают, то мы получаем обычный закон композиции. Таким способом можно добиться того, чтобы во всех местах, кроме одного, стояли обычные законы композиции, а в это одно место были бы загнаны все неприятности. Я не хочу это точно формулировать, потому что это всё равно неудовлетворительный результат. Но он показывает, что какие-то резервы во всём этот хозяйстве есть.

У меня есть подозрение, что существует комбинаторная теория абстрактных кубических поверхностей такого же уровня красоты, как и теория абстрактных проективных плоскостей, где в аксиоматику заложены все те геометрические свойства, которые, с одной стороны, можно доказать, а с другой стороны, их достаточно для реконструкции всего этого хозяйства. Может быть, такая теория абстрактных кубических поверхностей, со всей её алгеброй, позволила бы доказать теорему о конечной порождённости для кубических поверхностей.

Таким образом, для аналога г) вопрос остаётся открытым, и даже его точная формулировка допускает несколько неэквивалентных вариантов; подробности см. в [1].

Литература

[1] D. Kanevsky, Yu. Manin. Composition of points and Mordell—Weil problem for cubic surfaces // Rational points on algebraic varieties / E. Peyre, Yu. Tschinkel, Eds. — Basel: Birkhäuser, 2001. — P. 199—219. — (Progress in Math.; V. 199.) — E-print http://arxiv.org/math.AG/0011198.

11 сентября 2000 г.

А. В. Зелевинский

ОБОБЩЁННЫЕ КОЭФФИЦИЕНТЫ ЛИТТЛВУДА—РИЧАРДСОНА, КАНОНИЧЕСКИЕ БАЗИСЫ И ПОЛНАЯ ПОЛОЖИТЕЛЬНОСТЬ

Это — совместная работа с Аркадием Беренштейном, которую мы начали ещё в Москве 12 лет назад. В прошлом году всё постепенно собралось вместе, и, в какой-то степени, те задачи, которые мы поставили 12 лет назад, удалось решить.

Сначала я расскажу об исходной задаче. Напомню, что такое классические коэффициенты Литтлвуда—Ричардсона. Эти коэффициенты впервые возникли в работе Якоби в середине XIX века в контексте симметрических многочленов. В кольце симметрических многочленов $\mathbb{C}[x_1, \ldots, x_n]^{S_n}$ есть замечательный базис — базис многочленов Шура (такие многочлены до Шура рассматривал Якоби). Самый естественный способ построить симметрический многочлен — взять моном и его симметризовать. А чтобы получить многочлен Шура, нужно взять моном, антисимметризовать его, а потом поделить на антисимметричный моном минимальной степени (определитель Вандермонда), чтобы вернуть многочлен из антисимметрических в симметрические. Многочлены Шура определяются следующим образом:

$$s_\lambda = \frac{\sum\limits_{w \in S_n} \varepsilon(w) x^{w(\lambda+\delta)}}{\sum\limits_{w \in S_n} \varepsilon(w) x^{w\delta}}.$$

Здесь λ — разбиение на n неотрицательных слагаемых, т. е. целочисленная последовательность $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \lambda_n \geqslant 0$, а $\delta = (n-1, n-2, \ldots, 0)$. Когда λ пробегает все такие разбиения, получается базис в пространстве симметрических многочленов. Естественный вопрос — изучить, как устроено умножение в этом базисе:

$$s_\lambda s_\nu = \sum_\mu c_{\lambda\nu}^\mu s_\mu.$$

Коэффициенты $c_{\lambda\nu}^\mu$ как раз и называют (классическими) коэффициентами Литтлвуда—Ричардсона.

Впервые комбинаторные выражения для этих коэффициентов нашли Литтлвуд и Ричардсон в 1934 г. Строгое доказательство было получено

в 70-е годы. Я не буду его приводить. Оно основано на комбинаторике таблиц Юнга. Это то, чем занимается алгебраическая комбинаторика, можно сказать, комбинаторика типа A. Она имеет дело с симметрической группой, с полной линейной группой. Явное комбинаторное выражение для коэффициентов Литтлвуда—Ричардсона — один из самых важных результатов этой ветви алгебраической комбинаторики.

Если вы изучали теорию представлений, то когда вы глядите на определение s_λ, становится ясно, что оно имеет представленческий смысл. Одним из первых этот смысл заметил Шур; поэтому многочлены s_λ и называются многочленами Шура. Это — характеры конечномерных неприводимых представлений группы SL_n или GL_n. Значит, это можно обобщить на любые полупростые группы, и можно спросить, как обобщить классическое выражение для коэффициентов Литтлвуда—Ричардсона на теорию представлений произвольной полупростой группы. Это и есть то, что мы называем обобщёнными коэффициентами Литтлвуда—Ричардсона.

Я напомню основные факты про полупростые алгебры Ли. В своей основе полупростая алгебра Ли $\mathfrak{g}$ — это тоже комбинаторный объект. Он задаётся квадратной матрицей $A=(a_{ij})$, $i,j=1,\ldots,r$, с целочисленными элементами (эта матрица называется матрицей Картана). Матрица Картана должны обладать следующими свойствами: $a_{ii}=2$, $a_{ij}\in\mathbb{Z}_{\leqslant 0}$ (так мы будем обозначать неположительные целые числа); наконец, из равенства $a_{ij}=0$ следует, что $a_{ji}=0$ (рудиментарное свойство симметрии). Чтобы матрица Картана задавала конечномерную полупростую алгебру Ли, все её главные миноры должны быть положительны.

Все такие (неразложимые) матрицы классифицируются знаменитой классификацией Картана—Киллинга. Получаются серии A_n, B_n, C_n, D_n и исключительные матрицы E_6, E_7, E_8, F_4, G_2.

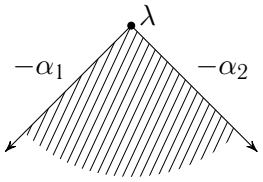
Как только вы задали матрицу Картана, вы сразу же можете задать полупростую алгебру Ли образующими и соотношениями. Она задаётся тремя типами образующих: $\mathfrak{g}=\langle e_i,\alpha_i^\vee,f_i\rangle$, $i=1,\ldots,r$; здесь r — ранг алгебры Ли (размер матрицы Картана), $\alpha_i^\vee$ — простые кокорни (вместо них часто пишут h_i). Эти образующие удовлетворяют соотношениям, которые явно записываются в терминах матрицы Картана. Например, $[\alpha_i^\vee,e_j]=a_{ij}e_j$. Выполняется также соотношение $[\alpha_i^\vee,\alpha_j^\vee]=0$, т. е. простые кокорни порождают коммутативную подалгебру $\mathfrak{h}=\langle\alpha_1^\vee,\ldots,\alpha_r^\vee\rangle$, которую называют подалгеброй Картана.

В двойственном пространстве к подалгебре Картана $\mathfrak{h}^*$ лежит решётка весов $P=\{\gamma\in\mathfrak{h}^*\colon\gamma(\alpha_i^\vee)\in\mathbb{Z}\text{ для всех }i\}$. Таким образом, веса — это линейные формы на $\mathfrak{h}$, которые принимают целочисленные значения на простых кокорнях. Среди всех весов выделяются доминантные, или старшие. Они

принимают неотрицательные значения на простых кокорнях. Теория представлений полупростой алгебры Ли очень проста: имеется взаимно однозначное соответствие между старшими весами и конечномерными неприводимыми представлениями. А именно, старшему весу λ соответствует представление V_λ, весовое разложение которого имеет вид $V_\lambda = \bigoplus_{\gamma \in P} V_\lambda(\gamma)$; под весовым разложением мы понимаем разложение по собственным подпространствам картановской подалгебры. Здесь

$$V_\lambda(\gamma) = \{v \in V_\lambda \colon hv = \gamma(h)v \text{ для всех } h \in \mathfrak{h}\}.$$

Известно, что картановская подалгебра диагонализуется во всех конечномерных представлениях, поэтому есть прямая сумма. Представление V_λ однозначно выделяется тем, что вес λ в нём самый старший. Это означает следующее. Мы вводим частичный порядок на весах: $\lambda \geqslant \mu$, если $\lambda - \mu$ есть целочисленная неотрицательная линейная комбинация простых корней, т. е. $\lambda - \mu \in \sum \mathbb{Z}_{\geqslant 0}\alpha_j$. (Простые корни — это линейные формы на картановской подалгебре, для которых $\alpha_j(\alpha_i^\vee) = a_{ij}$.) Поскольку матрица Картана обратима, простые корни образуют базис в двойственном пространстве. На них натянут конус полной размерности. Обычно рисуют такую картинку. Вы берёте старший вес, вычитаете из него простые корни — идёте вниз (рис. 1). Это и есть частичный порядок. По отношению к этому порядку V_λ имеет единственный максимальный вес, а именно, тот самый вес λ. Это соответствие есть взаимно однозначное соответствие между всеми старшими весами и всеми конечномерными неприводимыми представлениями.



Р и с. 1. Частичный порядок

Возникает естественный вопрос: что произойдёт, когда вы тензорно перемножите два конечномерных неприводимых представления и разложите это тензорное представление по неприводимым представлениям: $V_\lambda \otimes V_\nu = \bigoplus_\mu c_{\lambda\nu}^\mu V_\mu$. (Можно доказать, что для полупростых алгебр Ли тензорное произведение конечномерных неприводимых представлений раскладывается в прямую сумму неприводимых представлений.) В случае, когда берётся алгебра Ли $\mathfrak{sl}_n$, коэффициенты $c_{\lambda\nu}^\mu$ — это в точности классические коэффициенты Литтлвуда—Ричардсона.

Наша исходная задача состояла в том, чтобы найти комбинаторное выражение для обобщённых коэффициентов Литтлвуда—Ричардсона. Мы поставили перед собой задачу найти «естественные» выражения в том смысле, что они должны иметь ясный смысл с точки зрения

теории представлений. То есть должно быть некое векторное пространство, размерность которого равна этому коэффициенту, и то выражение, которое мы найдём, должно быть числом каких-то объектов, и эти объекты должны быть естественно связаны с базисными векторами. Таким образом, объяснение этого факта должно быть не комбинаторным, а должно вытекать из конструкций теории представлений. Кроме того, хотелось бы получить такое представление не отдельно в каждом случае картановской классификации, а в общем виде, чтобы оно имело смысл для любой полупростой алгебры Ли. Это не каприз. Здесь есть любопытный вызов с точки зрения комбинаторики. Комбинаторика типа A, которая имеет дело с классическими коэффициентами Литтлвуда—Ричардсона, с этой общей точки зрения, должна вся сидеть в картановской матрице для типа A. Вся мешанина диаграмм Юнга, таблиц Юнга, всевозможных соответствий Кнута и всех мыслимых вещей, которые можно с ними сделать, всё это должно в каком-то смысле содержаться в картановской матрице для типа A. Напомню, что такая матрица имеет вид

$$\begin{pmatrix} 2 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 2 \end{pmatrix}.$$

И если таблицы Юнга каким-то образом скрываются за этой матрицей, то спрашивается, как извлечь что-то такое, имеющее смысл для любой полупростой алгебры Ли, т. е. какого рода комбинаторика должна заместить комбинаторику таблиц Юнга.

Постепенно мы приближались к решению этого вопроса. Пока мы раскачивались, в 1994 г. Литтельман, в сущности, эту задачу решил. В его работе получено некое общее комбинаторное выражение для этих кратностей. Но оно получено комбинаторно и в тех терминах, которые нас не совсем устраивали. В частности, связь его языка с теорией представлений не очень ясна. Поэтому мы продолжали трудиться, и в 1999 г. наконец получили, как мне кажется, удовлетворительное решение этой задачи. Ответ совершенно не такой, как у Литтельмана. Мы с ним несколько раз разговаривали на эту тему. Каждый раз мы пытались разобраться, как свести вместе эти два ответа, и пока совершенно неясно, как это сделать. Это, на мой взгляд, интересный открытый вопрос. Ответ Литтельмана в наиболее современной формулировке использует его *модель путей* (paths model). Наш ответ использует формально близкие, но на самом деле совсем другие вещи, которые мы сначала тоже называли путями. Чтобы

отличать их от литтельмановских путей, мы их переименовали в $\boldsymbol{i}$-trail'ы. Сначала я попытался придумать какой-то русский перевод, но когда я сюда приехал и обнаружил, что русский язык обогатился словами типа «лизинг» и «холдинг», я решил оставить это название. Его смысл — трассы для горных лыж, лыжни.

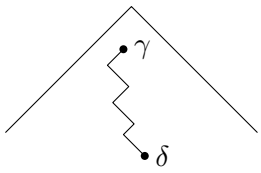
Я хочу сформулировать теорему, как найти обобщённые коэффициенты Литтлвуда—Ричардсона в терминах нового комбинаторного образования, которое называется $\boldsymbol{i}$-trail'ы. Если останется время, скажу несколько слов про доказательство. Оно сложилось довольно интересное, потому что используются те вещи, которые нам и не снились в 1988 году. В основном используются две ключевые идеи Люстига. Одна идея — о существовании канонического базиса и его свойствах, и ещё более новая идея о связи теории представлений с геометрией вполне положительных многообразий. Сначала я сформулирую ответ, который ничего этого не использует. А то единственное доказательство, которое мы знаем, включает в себя все эти ингредиенты. *)

Понятие $\boldsymbol{i}$-trail'ов нельзя отрывать от $\boldsymbol{i}$; они должны быть вместе. Здесь $\boldsymbol{i} = (i_1, \ldots, i_m)$ — последовательность индексов, которую образует приведённое разложение элемента $w_0 \in W$ максимальной длины в группе Вейля. Я сейчас напомню соответствующие определения. Главный комбинаторный объект, который ассоциируется с матрицей Картана, — это группа Вейля. Она строится как конечная группа, порождённая отражениями, действующими в подалгебре Картана $\mathfrak{h}$ или в двойственном пространстве весов. Для типа A это просто симметрическая группа. А в общем случае — это красивая группа, порождённая отражениями. На каждый простой корень (или простой кокорень) приходится по одной образующей. Например, на простые корни образующие группы $W = \langle s_1, \ldots, s_r \rangle$ действуют следующим образом: $s_i \alpha_j = \alpha_j - a_{ij} \alpha_j$. Таким образом, в базисе из простых корней действие задаётся с помощью матрицы Картана, а дальше распространяется по линейности. Эти отражения порождают группу Коксетера. Как и в каждой группе Коксетера, в этой группе есть понятие приведённого разложения. Это представление элемента группы в виде произведения образующих, которое имеет минимальную длину. Вместо того чтобы писать $w_0 = s_{i_1} \ldots s_{i_m}$, я буду записывать только последовательность индексов в этом произведении: $\boldsymbol{i} = (i_1, \ldots, i_m)$. Такую последовательность индексов называют приведённым словом для данного элемента. Известно, что в

*) A. Berenstein, A. Zelevinsky. Tensor product multiplicities, canonical bases and totally positive varieties // Invent. Math. — 2001. — V. 143. — P. 77—128. На моей страничке `www.math.neu.edu/~zelevinsky` можно найти ссылки на эту и примыкающие к ней работы.

каждой группе Коксетера есть ровно один элемент, для которого приведённое слово имеет самую большую длину. Для симметрической группы это такая перестановка, когда все индексы переставляются в обратном порядке.

Приведённых разложений много. Мы зафиксируем раз и навсегда одно из них. После этого введём понятие $\boldsymbol{i}$-trail'а. Он живёт в конечномерном представлении V и идёт от одного веса γ до другого веса δ. Представлять его нужно как лыжную трассу. Вы рисуете весовую диаграмму для пространства V, отмечаете в ней веса γ и δ. Нетривиальный trail получится только в том случае, когда $\gamma > \delta$ в этом частичном порядке, потому что интереснее катиться вниз, а не вверх (рис. 2).



Р и с. 2. Trail

Направления поворотов диктуются приведённым словом. Вы берёте последовательность весов $\gamma_0 = \gamma \geqslant \gamma_1 \geqslant \ldots \geqslant \gamma_m = \delta$, начинающуюся с γ и спускающуюся к δ так, что направление спуска от каждого корня к следующему диктуется соответствующим местом в приведённом слове: $\gamma_{k-1} - \gamma_k = c_k \alpha_{i_k}$, где $c_k \in \mathbb{Z}_{\geqslant 0}$. Есть ещё одно требование, гарантирущее, что мы не вылетим за пределы весовой диаграммы. Это требование такое. Возьмём повышающие операторы $e_i \colon V(\delta) \to V(\delta + \alpha_i)$; они играют роль подъёмников. Композиция $e_{i_1}^{c_1} \ldots e_{i_m}^{c_m}$ возвращает вас обратно, т. е. переводит $V(\delta)$ в $V(\gamma)$. Требование состоит в том, что $e_{i_1}^{c_1} \ldots e_{i_m}^{c_m} \neq 0$. Из этого автоматически вытекает, что по дороге мы не могли выйти за пределы диаграммы весов, потому что все промежуточные весовые пространства тоже должны быть ненулевыми. Но наше требование более сильное.

Последнее свойство делает это понятие не совсем комбинаторным, потому что мы должны проанализировать, когда эта композиция операторов отлична от нуля. На первый взгляд это обесценивает такое понятие, потому что проанализировать такие вещи — это, в сущности, задача того же порядка, что и проанализировать сами кратности. Но на самом деле это не так, потому что та формула, которую я сейчас напишу, использует $\boldsymbol{i}$-trail'ы только в очень ограниченном числе. Нам нужно получить ответ для любой тройки старших весов. Оказывается, что форма ответа следующая. Коэффициент $c_{\lambda\nu}^{\mu}$ равен числу целочисленных решений некой системы линейных неравенств, причём левые части этих неравенств будут заготовлены раз и навсегда; они не будут зависеть от λ, μ и ν. Левые части неравенств как раз и будут вычисляться по $\boldsymbol{i}$-trail'ам, причём по конечному их числу (только по $\boldsymbol{i}$-trail'ам для фундаментальных представлений, более того, только по вполне определённым $\boldsymbol{i}$-trail'ам для фундаменталь-

ных представлений). Веса λ, μ и ν будут фигурировать только в правых частях.

Чтобы сформулировать ответ, с каждым $\boldsymbol{i}$-trail'ом $\pi = (\gamma_0 = \gamma \geqslant \gamma_1 \geqslant \ldots \ldots \geqslant \gamma_m = \delta)$ помимо чисел

$$c_k(\pi) = \frac{\gamma_{k-1} - \gamma_k}{2}(\alpha_{i_k}^\vee)$$

свяжем ещё числа

$$d_k(\pi) = \frac{\gamma_{k-1} + \gamma_k}{2}(\alpha_{i_k}^\vee).$$

(Для них я не знаю «лыжной» интерпретации.)

Т е о р е м а 1. *Пусть λ, μ, ν фиксированы. Тогда для любого приведённого разложения $\boldsymbol{i} = (i_1, \ldots, i_m)$ для w_0 кратность $c_{\lambda\nu}^\mu$ равна числу наборов t целых чисел $(t_1, \ldots, t_m)$, удовлетворяющих следующим условиям.*

1. $\sum\limits_{k=1}^{m} d_k(\pi) t_k \geqslant 0$ *для любого индекса i и любого $\boldsymbol{i}$-trail'а π из $\omega_i^\vee$ в $w_0 s_i \omega_i^\vee$ в фундаментальном представлении $V_{\omega_i^\vee}$ двойственной по Ленглендсу алгебры Ли ${}^L\mathfrak{g}$.*

2. $\sum t_k \alpha_{i_k} = \lambda + \nu - \mu$.

3. $\sum d_k(\pi) t_k \geqslant -\lambda(\alpha_i^\vee)$ *для любого i и любого $\boldsymbol{i}$-trail'а π в $V_{\omega_i^\vee}$, ведущего из $s_i \omega_i^\vee$ в $w_0 \omega_i^\vee$.*

4. $t_k + \sum\limits_{l>k} a_{i_k i_l} t_l \leqslant \nu(\alpha_{i_k}^\vee)$ *для всех $k = 1, \ldots, m$.*

Необходимые пояснения: через ω_i обозначены фундаментальные веса, образующие двойственный базис к простым кокорням, т. е. $\omega_i(\alpha_j^\vee) = \delta_{ij}$. Обозначение $\omega_i^\vee$ показывает, что берётся фундаментальный вес не для исходной алгебры Ли $\mathfrak{g}$, а для алгебры Ли ${}^L\mathfrak{g}$ — двойственной к ней по Ленглендсу. Здесь ${}^L\mathfrak{g}$ — это точно такая же алгебра, только соответствующая транспонированной матрице Картана. В условии 1 trail'ы ведут из старшего веса $\omega_i^\vee$ в вес $w_0 s_i \omega_i^\vee$, непосредственно предшествующий младшему весу $w_0 \omega_i^\vee$. Что касается условия 3, то, в отличие от условия 1, мы начинаем не с самого старшего веса, а чуть ниже, но зато теперь сваливаемся в самый младший вес.

Теперь я попробую объяснить, что стоит за доказательством. Как я уже говорил, есть две идеи. Обе в значительной степени инициированы Люстигом. Первая идея — это идея канонического базиса. А именно, кратность $c_{\lambda\nu}^\mu$ можно интерпретировать как число элементов некоторой части векторов канонического базиса.

Люстиг построил свой канонический базис в алгебре $U_q(\mathfrak{n})$, которая является q-деформацией универсальной обёртывающей алгебры от нильпотентной части нашей алгебры Ли; здесь $\mathfrak{n} = \langle e_1, \ldots, e_r \rangle \subset \mathfrak{g}$ — подалгебра,

порождённая элементарными повышающими операторами («подъёмниками») $e_1, \ldots, e_r$. Алгебру $U_q(\mathfrak{n})$ тоже можно задать, совершенно явно, образующими и соотношениями. А именно, $U_q(\mathfrak{n}) = \langle E_1, \ldots, E_r \rangle$ с некими конкретными соотношениями, которые пишутся по матрице Картана. Например, если $a_{ij} = a_{ji} = 0$, то $E_i E_j = E_j E_i$. Если $a_{ij} = a_{ji} = -1$, то

$$E_i^2 E_j - (q + q^{-1}) E_i E_j E_i + E_j E_i^2 = 0.$$

Здесь q — просто буква, переменная. Алгебра $U_q(\mathfrak{n})$ — это алгебра над полем $\mathbb{C}(q)$ рациональных функций от переменной q.

Для алгебр Ли типа A, D, E никаких других соотношений нет. Для других алгебр Ли появляются более длинные соотношения, но они имеют ту же самую природу; они тоже пишутся по матрице Картана.

Когда q полагается равным 1, возникает ассоциативная алгебра над полем комплексных чисел, которая естественно отождествляется с универсальной обёртывающей ассоциативной алгеброй для $\mathfrak{n}$, порождённой $e_1, \ldots, e_r$.

В алгебре $U_q(\mathfrak{n})$ Люстиг построил замечательный $\mathbb{C}(q)$-базис B. Определение базиса B довольно техническое. Вместо того чтобы давать это определение, я скажу, какие основные свойства этого базиса нам нужны.

1. Базис B состоит из однородных элементов относительно естественной градуировки. Алгебра $U_q(\mathfrak{n})$ градуирована решёткой весов $\bigoplus_{i=1} \mathbb{Z}_{\geqslant 0}\alpha_i$; градуировка такая: $\deg E_i = \alpha_i$.

2. Базис совместим с естественными фильтрациями. А именно, для любого i и любого целого неотрицательного числа n пространство $E_i^n U_q(\mathfrak{n})$ порождено частью базиса B. Это свойство весьма нетривиально: бесконечное семейство подпространств можно одновременно сделать координатными подпространствами. Далеко не любое семейство подпространств обладает таким свойством. Это свойство означает, что семейство подпространств, при меняющихся i и n, порождает дистрибутивную решётку. Самое простое доказательство этого факта состоит в люстиговской конструкции базиса.

3. Базис устойчив относительно симметрии ($\mathbb{C}(q)$-линейного антиавтоморфизма) $E \to E^\iota$. Этот антиавтоморфизм разворачивает все произведения в обратном порядке, но все образующие оставляет на месте. Из устойчивости относительно этой симметрии следует, например, что пространство $U_q(\mathfrak{n})E_i^n$ тоже порождено частью базиса.

Зная эти свойства, легко интерпретировать коэффициенты $c_{\lambda\nu}^{\mu}$ как размерности некоторых подпространств, которые строятся по базису Люстига. Этот факт хорошо известен в классической теории представлений, а

Люстиг подметил, что в этом варианте теория представлений квантовой группы совпадает с классической. Поэтому когда мы добавляем q, в этом месте решительно ничего не меняется.

Для краткости введём обозначение $U_q(\mathfrak{n}) = U^+$.

У т в е р ж д е н и е.

$$c_{\lambda\nu}^{\mu} = \dim \left[U^+ \Big/ \sum_i \left(E_i^{\lambda(\alpha_i^\vee)+1} U^+ + U^+ E_i^{\nu(\alpha_i^\vee)+1} \right) \right]^{(\lambda+\nu-\mu)},$$

т. е. коэффициент $c_{\lambda\nu}^{\mu}$ равен размерности однородной компоненты степени $\lambda+\nu-\mu$ в таком бесконечномерном факторпространстве.

Тем самым, задача естественности комбинаторного описания получает вполне определённый смысл. Мы знаем, что $c_{\lambda\nu}^{\mu}$ равно размерности пространства, которое порождено частью канонического базиса. Поэтому нужно некоторым естественным способом занумеровать векторы канонического базиса и посчитать, какие метки отвечают этой части. Нужно только понять, что эти метки выделяются в точности системой уравнений и неравенств из теоремы 1.

Нужно сообразить, какая именно нумерация канонического базиса приводит к этой теореме. Есть два основных способа параметризовать канонический базис. Один по Люстигу, другой, логично было бы сказать, по Кашиваре, но у меня такое впечатление, что Аркадий Беренштейн и я придумали его немного раньше Кашивары, в 1992 г. Поэтому второй способ мы будем называть струнными параметризациями. Я буду использовать именно их. Мне неловко произносить «струнные», поскольку наши струнные параметризации никак не связаны с теорией струн. Они скорее связаны со струнами, которые возникают в алгебре, — со струнами положительных корней. Это в некотором смысле струны чисел.

Струнные параметризации естественно связаны с подходом Кашивары, основанном на понятии кристалла. Но объяснить их легче, потому что они не требуют кристаллов. Струнами удобнее параметризовать не сам канонический базис, а двойственный к нему. Рассмотрим двойственное (в конечномерном смысле) пространство $(U^+)^* = \bigoplus_\gamma U^+(\gamma)^*$. В этом пространстве двойственный базис $B^* = \{b^*\}$ образуют линейные формы, для которых $b^*(b') = \delta_{bb'}$. Струнные параметризации возникают в следующей ситуации. Пусть V — локально конечный U^+-модуль; локальная конечность означает, что если мы возьмём любую образующую и начнём применять её к любому вектору, то забьём его в нуль (это не значит, что образующая действует как нильпотентный оператор, но локально она дей-

ствует как нильпотентный оператор). Применяться это будет к $V = (U^+)^*$. Кстати сказать, геометрический смысл этой двойственной алгебры состоит в том, что она является q-деформацией кольца регулярных функций на соответствующей группе Ли. Как всегда, универсальная обёртывающая — это дифференцирования, локальный объект; всё происходит в единице. А двойственный объект — это кольцо регулярных функций на соответствующей группе Ли. Тогда сама алгебра действует на двойственной к ней. Алгебраически это описать проще всего так. Алгебра U^+ действует на $(U^+)^*$ следующим образом: $(Ef)(E') = f(E^\iota E')$. Понятно, что раз при умножении на E^ι в U^+ градуировка повышается, то сам оператор будет понижающим градуировку. Поэтому действие будет локально нильпотентным.

Струны определяются следующим образом. Фиксируем такую же последовательность индексов $\boldsymbol{i} = (i_1, \ldots, i_m)$, как и раньше. (При определении струны даже не важно, будет ли эта последовательность приведённым словом. Но применяться это будет к приведённым словам, поэтому я сохраню прежние обозначения.) Мы будем определять струну в направлении $\boldsymbol{i}$ от любого ненулевого вектора $v \in V$. Пусть $c_i(v)$ — максимальная степень n, для которой $E_i^n v \neq 0$. Эта максимальная степень конечна в силу локальной конечности. Построим теперь отображение $v \mapsto c_{\boldsymbol{i}}(v) = (t_1, \ldots, t_m)$, соответствующее не одному элементу i, а всей последовательности $\boldsymbol{i}$, следующим образом. Положим $t_1 = c_{i_1}(v)$, $t_2 = c_{i_2}(E_{i_1}^{t_1} v)$ и т. д. Последовательность $(t_1, \ldots, t_m)$ назовём *струной* данного вектора в данном направлении.

Непосредственно из определения не видно, параметризация это или нет. Но теперь доказывается следующая теорема.

Теорема 2 (Беренштейн и Зелевинский, Литтельман, 1998). *Если $\boldsymbol{i} = (i_1, \ldots, i_m)$ — приведённое разложение для w_0, то отображение $c_{\boldsymbol{i}}$ задаёт биекцию B^* на $C_{\boldsymbol{i}}(\mathbb{Z})$, где $C_{\boldsymbol{i}} \subset \mathbb{R}^n$ — некоторый выпуклый многогранный конус, а $C_{\boldsymbol{i}}(\mathbb{Z})$ — множество точек с целочисленными координатами в этом конусе.*

Тот факт, что это биекция, даёт естественный способ параметризовать двойственный канонический базис, а значит, и обычный канонический базис. Ровно эти струны вычислены в теореме 1. Смысл этой теоремы с такой точки зрения следующий. Условие (1) задаёт в точности струнный конус; заметьте, что задача вычисления струнного конуса тоже была не решена. Условие (2) описывает то, что в кратности вес должен быть фиксирован, и как раз равен $\lambda + \nu - \mu$. Условия (3) и (4) описывают те условия, что мы должны профакторизовать по векторам, которые делятся на определённые степени E_i слева и справа.

За оставшееся время я попробую рассказать, как это связано с полной положительностью, и откуда берутся явные формы всех этих неравенств,

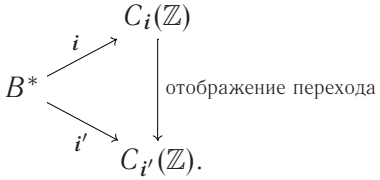
в частности, явное описание струнных конусов. Почти одновременно, год или два назад, появились две работы: работа Литтельмана и работа Саши Постникова и моего аспиранта Олега Глейзера, которые как бы дополняли друг друга. Литтельман вычислил конус C_i для всех алгебр Ли, рассматривая каждую серию отдельно, но для очень специального выбора приведённых разложений, в то время как Глейзер и Постников вычислили его только для типа A, но для произвольного приведённого разложения. После этого мы с Аркадием сообразили, что этот ответ имеет смысл для произвольной алгебры Ли и для произвольного приведённого разложения.

Идея доказательства в каком-то смысле чрезвычайно странная. Она основана на исчислении, которое имеет несколько имён. Иногда используется название *идемпотентный анализ*. Но, насколько мне известно, эта вещь была известна в прикладной математике раньше исследований по идемпотентному анализу под разными другими названиями. Моё любимое название — *тропическое исчисление*. Его используют специалисты по оптимизации, теории контроля. Основная идея такова. Когда приходится иметь дело со сложной системой линейных неравенств, вместо того чтобы писать все эти неравенства, можно сказать, что минимум большого семейства линейных форм больше или равен чему-то. Но с этими минимумами трудно работать. Тропическое исчисление даёт возможность очень эффективно работать с кусочно-линейными выражениями, содержащими минимумы линейных форм. Идея чрезвычайно простая, даже наивная. Мы вводим экзотическую структуру полукольца, скажем, на множестве целых чисел $\mathbb{Z}$. Вводим операции $\oplus$ (тропическое сложение), $\odot$ (тропическое умножение) и $\oslash$ (тропическое деление); нет операции вычитания. Эти операции определяются следующим образом: $a \odot b = a + b$, $a \oslash b = a - b$, $a \oplus b = \min(a, b)$. Лёгкая проверка показывает, что сложение и умножение коммутативны и ассоциативны; имеется обычная дистрибутивность. Вычитания нет, потому что сложение идемпотентно: $a \oplus a = a$. Все свойства, которые можно вывести для полей, не используя вычитания, выполняются. На самом деле, можно сообразить, что если что-то можно вывести, даже используя вычитание, но только так, что в конечном результате вычитания не окажется, то в такой структуре это тоже будет верно. Это как с мнимыми числами — их можно использовать, и если потом в ответе они исчезнут, то ответ всё равно будет правильным. Здесь нужно соблюдать только одно правило игры: чтобы ни в начале вычислений, ни в конце никакого вычитания не было. Тогда всё будет верно в тропической ситуации.

Если есть кусочно-линейное выражение, которое можно составить из этих операций, вы можете поднять его. Это то, что мы называем *геометрический подъём*. Он поднимает кусочно-линейные выражения в

обычные безвычитательные рациональные выражения. Возьмём, например, выражение $2\min(a, b)$. Тропически его можно записать в двух видах: $2\min(a, b) = a^2 \oplus b^2 = (a \oplus b)^2$. Идея состоит в том, что если есть какое-то кусочно-линейное выражение, которое содержит минимумы линейных форм, то его нужно поднять неким умным способом до алгебраического выражения так, чтобы это алгебраическое выражение имело разумную алгебро-геометрическую интерпретацию. Используя эту интерпретацию, мы вычисляем наши алгебраические выражения, а в конце снова заменяем обычные операции на тропические и получаем искомый ответ. Наши формулы были получены именно таким способом.

Я попробую, не развивая общей теории, сказать, где ключевое место: как осуществляется геометрический подъём и как мы переходим от комбинаторики к геометрии. Это делается следующим образом. Заметьте, что у нас есть двойственный канонический базис, и с каждым выбором приведённого разложения для w_0 связана биекция этого базиса с $C_{\boldsymbol{i}}(\mathbb{Z})$. Другое приведённое разложение определяет другую биекцию. Мы можем задать вопрос, как связаны между собой эти параметризации канонического базиса:



Если разобраться, можно увидеть, что отображение перехода кусочно-линейное.

Как вы помните, чтобы вычислить условия на векторы канонического базиса, нужно понять, каким условиям должны удовлетворять струны, чтобы соответствующий вектор делился справа на какую-то степень образующей E_i. Оказывается, что в зависимости от E_i нам иногда удобно применять струну, связанную с одним приведённым разложением для w_0, а иногда — с другим. Если вы знаете струну для приведённого разложения, где первый индекс равен в точности i, то ответить на вопрос, на какую степень этого конкретного E_i вектор делится, будет очень легко. Но зато, зная эту струну, трудно ответить на вопрос, на какую степень другой образующей делится вектор. А если мы будем знать струны во всех направлениях, то на все эти вопросы будет легко ответить. Поэтому получение ответа сводится к анализу отображений перехода. Если нам удастся явным образом написать кусочно-линейные формулы перехода, то нам удастся решить задачу. Это не так уж легко сделать, но идея состоит именно в этом.

Такие отображения перехода Люстиг предложил рассматривать для своих параметризаций. Но для параметризаций по струнам он их не рассматривал. Давайте я напишу «струнное» отображение перехода в конкретном случае. В этом примере $\mathfrak{g} = \mathfrak{sl}_3$ — матрицы порядка 3 со следом нуль. Есть только две образующие E_1 и E_2. Приведённых разложений тоже два: $\boldsymbol{i} = (1, 2, 1)$ и $\boldsymbol{i}' = (2, 1, 2)$. Пусть им соответствуют тройки (t_1, t_2, t_3) и (t'_1, t'_2, t'_3). Тогда отображения перехода выглядят следующим образом:

$$t'_1 = \max(t_3,\ t_2 - t_1) = \left[\frac{t_2 t_3}{t_1 t_3 + t_2}\right]_{\text{trop}},$$

$$t'_2 = t_1 + t_3 = [t_1 t_3]_{\text{trop}},$$

$$t'_3 = \min(t_1,\ t_2 - t_3) = \left[\frac{t_1 t_3 + t_2}{t_3}\right]_{\text{trop}}.$$

Если мы притворимся, что про тропическое исчисление мы хотим забыть, то нам нужно попытаться найти какой-то геометрический смысл в этих преобразованиях. Что бы мог означать переход от одного набора чисел (скажем, комплексных) к другому, задаваемый этими рациональными преобразованиями? Это, в сущности, одно из ключевых мест в нашей работе. Мы сообразили, что эти формулы описывают такой переход — уже в геометрической (или в алгебраической) задаче. Возьмём группу $G = \mathrm{SL}_3$ (группа матриц порядка 3 с определителем 1). В этой группе рассмотрим кривые $x_{-i}(t) = \varphi_i \begin{pmatrix} t^{-1} & 0 \\ 1 & t \end{pmatrix}$. Здесь φ_i — вложение матриц порядка 2 в матрицы порядка 3. Эти вложения заданы следующим образом: $\varphi_1 = \begin{pmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & 1 \end{pmatrix}$ и $\varphi_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{pmatrix}$; вместо звёздочек подставляется данная матрица порядка 2. Возьмём нижнюю треугольную матрицу $x_{-1}(t_1)x_{-2}(t_2)x_{-1}(t_3)$ и попробуем представить её по-другому: $x_{-1}(t_1)x_{-2}(t_2)x_{-1}(t_3) = x_{-2}(t'_1)x_{-1}(t'_1)x_{-2}(t'_3)$. Если теперь вычислить переход от (t_1, t_2, t_3) к (t'_1, t'_2, t'_3), то получатся в точности те же формулы, которые были выписаны выше, но без тропикализации. То есть геометрическим смыслом переходов струнных параметризаций в каноническом базисе оказывается переход между произведениями обычных матриц. В этом можно убедиться, перемножив матрицы; здесь нет никаких трудностей.

Как только мы это сообразили, задачу сразу же можно переформулировать геометрически — как связь между двумя произведениями матриц. Это уже задача линейной алгебры. Можно забыть про кусочную линейность, и начинать решать эту задачу. Связь с так называемой полной положительностью состоит в том, что все матрицы, которые тут появляются, и

все их произведения, имеют положительные миноры. Матрицу называют вполне положительной, если все её миноры положительны. Когда речь идёт о треугольных матрицах, положительными должны быть те миноры, которые не равны тождественно нулю для всех (нижних) треугольных матриц.

Такой формализм можно развить в любой группе, не только в SL_3. Основной технический результат, за счёт которого всё получается, состоит в следующем. Представьте себе, что элемент группы записан в виде длинного произведения, а вложения в общем случае отвечают простым корням. Тогда можно вычислить компоненты как явные рациональные функции от произведения. Это так называемая задача факторизации. После того как такой результат получен, у нас появляется контроль над функциями перехода. Для этих вычислений приходится развить детерминантное исчисление — исчисление (обобщённых) миноров в произвольной полупростой группе. Когда мы всё это сделаем, trail'ы получаются как результат вычислений каких-то странных обобщённых определителей в полупростых группах. Мы вычисляем определитель и приходим к ответу, даваемому теоремой 1. Если хотите, это какая-то своеобразная идея производящей функции. Вместо того чтобы работать с минимумом большого числа линейных форм, мы кодируем его рациональным выражением, а результат вычисления этого рационального выражения — определитель.

14 сентября 2000 г.

В. И. Арнольд

ТЕОРИЯ РАСПРОСТРАНЕНИЯ ВОЛН

Теория распространения волн — это необъятный предмет. Это примерно такое же название для часовой лекции, как «математический анализ». В годовой курс математический анализ не уложишь, это гораздо больше. Так же и с теорией распространения волн. Она включает в себя всю математику, всю физику и многие другие науки. Поэтому я буду говорить не обо всей теории распространения волн, а выберу из неё только маленькие кусочки, которые уже всё равно очень многое содержат. Эти маленькие кусочки будут в основном посвящены тому, что называется *геометрической оптикой*. Эта наука является, в сущности, частью геометрии. Имеется теория распространения ударных волн, имеется теория распространения волн, которая называется физической оптикой, есть квантовая механика и т. д. Про все эти области, хотя они тесно связаны с той, про которую я буду говорить, я избегу говорить, потому что они связаны с более сложными объектами. Я буду говорить только об объектах, которые теоретически очень просты.

Я начну с первого примера, каким образом это работало. Я долго считал, что первым эту область придумал Архимед, примерно за 250 лет до Рождества Христова. Он сжёг корабли, пытавшиеся атаковать Сиракузы, в которых он жил. Для этого Архимед установил нужным образом зеркала. Недавно мне указали (за это сообщение я благодарен Ф. Аикарди), что я заблуждаюсь, и на самом деле, как всегда бывает, когда что-то кому-то приписывается, автор не он. Оказывается, этот результат примерно за 200 или 300 лет до Архимеда упоминается в пьесе Аристофана «Облака». Там он приписывается Сократу (примерно 450 или 500 лет до Р. Х.). Там упоминается применение в практической жизни каустик (фокальных точек, т. е. как раз того, о чём я буду сегодня рассказывать). Там описывается их применение в юриспруденции. Эта пьеса Аристофана направлена против Сократа, который представлен ужасным дураком и софистом, который

Автор благодарен М. А. Цфасману за организацию этой лекции и В. В. Прасолову за её конспектирование и компьютерный набор.

ставит всякие глупые задачи и учит в какой-то спецшколе своих учеников софистов глупо рассуждать. В «Облаках» есть такое упражнение, которое Сократ даёт ученику: «Ты задолжал большую сумму денег, и у твоего врага есть расписка, где ты клянёшься, что их отдашь. Срок подошёл, ты не отдал. Он обращается к судье, судья тебя вызывает. Тебе нужно выиграть процесс. Как это сделать?» Дальше они с учеником долго рассуждают, придумывают разные софизмы. В конце концов кто-то из них придумывает решение. Нужно назначить для суда солнечный день. Затем в день суда нужно прийти и встать на солнечное место. Перед этим нужно пойти в аптеку (я не знал, что тогда у греков были такие замечательные аптеки) и там купить линзу (оказывается, в аптеках уже были линзы). Затем нужно сфокусировать этой линзой пучок световых лучей на твоей расписке в тот момент, когда противник будет её предъявлять, и сжечь её в руках у него. Так можно выиграть процесс.

Здесь, конечно, требуется серьёзная математика, потому что нужно понять, как это лучи становятся такими сильными, чтобы сжигать. Их нужно сфокусировать. Теория фокусировок, теория пучков лучей, — сложная наука, которая может излагаться по-разному. Таким образом возникают разные науки, которые по-разному называются: вариационное исчисление, оптимальное управление, механика, оптика и т. д. Есть разных много наук, и специалисты во всех этих науках имеют разные терминологии и при этом не понимают друг друга, потому что они дают разные обозначения для одних и тех же понятий. Есть ещё и причины хуже, по которым люди тут друг друга не понимают: так специально устроена наука. Например, ещё в прошлом веке у математиков и у физиков обозначения были противоположные. То, что математики обозначали p и q, то физики обозначали q и p. В течение долгого времени была борьба, о которой вы можете прочитать в книге Ф. Клейна *) «Лекции о развитии математики в XIX столетии». В конце концов физики победили. Надо сказать, что почти всегда так бывает. Математики были правы, у них были более разумные обозначения (более первичный объект обозначен более ранней буквой алфавита, вторичный — последующей), но победили физики, потому что их в 10 раз больше.

Теория, про которую я буду говорить, — это очень общая математическая теория, размером примерно с анализ или топологию — большая теория. Но объяснить её сразу было бы очень трудно, потому что если бы я стал на бурбакистский аксиоматический подход, ввёл бы определения, то

*) Немецкое *произношение* «Кляйн», но русская транскрипция «Клейн» — такова традиция! Также и «Г. Вейль» произносится «Херман Вайль», но в печатном тексте никто бы не избрал этого фонетически правильного варианта.

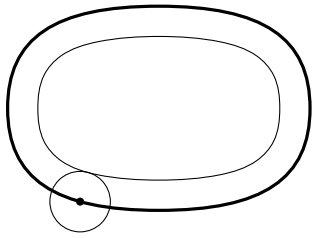
мне было бы легко, а вам трудно. Поэтому я этого делать не буду, а начну с простейших примеров. Соотношение этих примеров к общей теории примерно такое же, как отношение многочленов к анализу. Конечно, если у вас есть парабола, то вы можете провести к ней касательную. Конечно, проведение касательной к параболе — это важнейший шаг в построении теории дифференцирования. Точно так же, под параболой можно посчитать площадь, и, конечно, вычисление этой площади (и объёма шара, площади сферы) предшествовало созданию анализа. И если этого не понять, то анализ никогда и не поймёшь. Точно так же, примеры, о которых я буду говорить, должны предшествовать теории распространения волн. Поэтому об этих примерах я скажу сначала, а теория появится уже в конце.

Простейший случай — распространение волн на плоскости. Плоскость $\mathbb{R}^2$ я буду считать снабжённой евклидовой метрикой. Вместо плоскости можно взять любое риманово многообразие, но уже плоскость вполне интересна. Пусть на плоскости имеется начальный волновой фронт — какая-нибудь кривая. На нём есть какое-то возмущение — лесной пожар, эпидемия или что-нибудь ещё. Для примера в качестве волнового фронта возьмём эллипс. Волновой фронт может быть произвольным, не только эллипсом, но эллипс — это уже вполне интересно. Распространение волнового фронта осуществляется так. Предположим, что в каждой точке задан какой-то локальный закон, по которому процесс распространяется. Например, возмущения идут со скоростью 1 (плоскость евклидова, там есть метрика) во все стороны. Тогда имеется принцип Гюйгенса, который говорит, что возникает сферический волновой фронт. В данном случае сферический волновой фронт через время t будет окружностью радиуса t. Аналогичная конструкция есть для любого риманова многообразия. Можно всё это обобщить и в гораздо более общей ситуации — в вариационном исчислении, в механике и т. д. Об этом я не буду говорить. Я рассматриваю очень конкретный пример.

Образующийся через время t из исходного фронта волновой фронт — это огибающая семейства сферических фронтов радиуса t с центрами в точках исходного фронта. Для построения этой огибающей можно обойтись без конструирования всех этих сферических фронтов, воспользовавшись «лучами»: нормалями к исходному фронту. Через время t нужно по нормали сдвинуться на расстояние t. В каждой точке исходного фронта, например эллипса, нужно провести нормаль. Если волна распространяется внутрь, то внутреннюю нормаль. На этой внутренней нормали внутрь эллипса нужно отложить отрезок длины t. Тогда все эти точки образуют некоторую новую кривую. Если исходная кривая была Γ_0, то через время t будет кривая Γ_t — волновой фронт через время t (рис. 1). Можно считать,

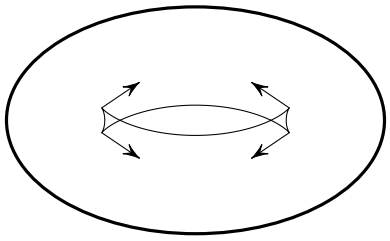
что снаружи всё сгорело, пожар пошёл внутрь, и через малое время t границей пожара будет кривая Γ_t, параллельная кривой Γ_0.

Если у вас есть возможность рисовать эллипсы и другие кривые, то это можно продолжать дальше. Раньше, в XIX веке, ещё умели рисовать. Даже школьники в гимназиях рисовали. Сейчас многими «математиками» запрещается рисовать. Бурбакистские запрещения рисунков привели к тому, что нынешним школьникам, студентам, а следовательно и профессорам, рисовать уже не под силу.



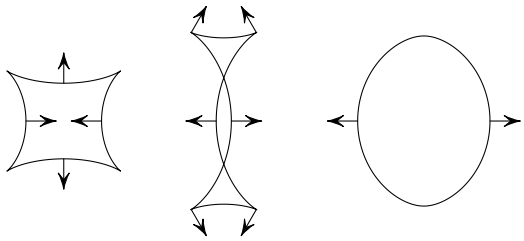
Р и с. 1. Волновой фронт

Оказывается, что если взять t побольше, то фронт через время t будет уже негладкий. Это довольно странное обстоятельство. Понять это можно только если поэкспериментировать. Этот эксперимент желательно делать в возрасте 8—10 лет. Получается волновой фронт, изображённый на рис. 2. У него помимо двух точек самопересечения есть четыре точки возврата. Около точек возврата эта кривая, оказывается, диффеоморфна полукубической параболе $x^2 = y^3$. Это уже нетривиальная теорема. Можете воспринимать её как хорошую задачу.



Р и с. 2. Негладкий волновой фронт

Дальше фронт продолжает распространяться. Если рисовать дальше, то вы можете убедиться, что в некоторый момент две части фронта коснутся друг друга и пройдут одна сквозь другую; точки самопересечения исчезнут (рис. 3). После этого через некоторое время коснутся друг друга две другие ветви. Затем исчезнут два образовавшихся треугольника, и фронт снова станет гладким. Дальше фронт будет распространяться, оставаясь гладким.



Р и с. 3. Распространение волнового фронта

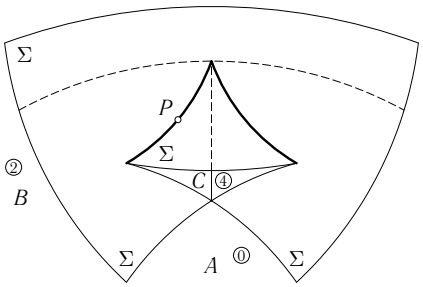
Оказывается, что явление, которое я здесь описал, устойчиво. Для эллипса это всё — алгебраическая геометрия; можно написать явные формулы для волновых фронтов. Между прочим, не такие уж простые. Современный алгебраический геометр, вообще говоря,

не справляется с такой задачей. Это ведь настоящая, реальная (т. е. вещественная) задача. Но старые алгебраические геометры такие задачи решать могли. Первым ответ опубликовал Кэли. В то же время, этот ответ был давно известен. В первом учебнике анализа Лопиталя уже всё это было. Он, правда, ничего не доказывал и не понимал. Он только находил ответы. Вероятно, он подходил как физик. Что касается Кэли, то у него уже всё в порядке, всё доказано. Эти кривые алгебраические. Для этих алгебраических кривых по теории Плюккера он находит все особые точки, всё полностью исследует. Здесь замечательным является факт устойчивости полукубической особенности $x^2 = y^3$. Если эллипс заменить какой-нибудь другой кривой, то всё равно в некоторый момент появится полукубическая особенность, и она устойчива — сохраняется в течение некоторого времени. Этот факт примерно такого же характера, как лемма Морса, которая говорит, что около минимума функция ведёт себя как квадратичная форма: если второй дифференциал невырожден, то заменой переменных функцию можно свести к квадратичной форме. Теорема об устойчивости полукубической особенности того же рода, только менее лёгкая. При распространении волн на плоскости всегда получается полукубическая парабола. Такие же ответы получаются на всех двумерных многообразиях с римановой метрикой. Локально всегда только такие особенности и есть. Более сложные особенности, конечно, бывают. Например, если вы начнёте с окружности, то она в какой-то момент столкнётся в центр. Но это вещь неустойчивая. Если вы гладко пошевелите окружность и превратите её, например, в эллипс, то эллипс в точку уже не сворачивается. Оказывается, что в общем положении есть только такие особенности, а никаких других не будет. Это уже довольно нетривиальная теорема. Но это только для размерности 2.

Если вместо плоскости мы возьмём трёхмерное пространство, то положение оказывается гораздо более сложным. В трёхмерном пространстве можно провести аналогичные эксперименты с гладкими поверхностями. Можно начать с эллипсоида. Кэли, между прочим, это сделал. Он нарисовал ответы для эллипсоида, между прочим, довольно хитрые. Волновые фронты для гладких поверхностей получаются довольно хитрые. В трёхмерном пространстве роль полукубической особенности (самой типичной особенности волнового фронта) играет специальная поверхность, которая тоже задаётся алгебраическим уравнением. Но это алгебраическое уравнение гораздо более длинное. В разных системах координат иногда 6 слагаемых, иногда 12. Я не буду писать уравнение этой поверхности, а опишу её геометрически. У этой поверхности есть около сотни разных определений, и есть около десяти тысяч теорем, которые утверждают, что все эти определения эквивалентны. Я дам несколько из этих определений,

потому что это один из замечательных объектов. В математике мало замечательных объектов: прямая, окружность, эллипс, гипербола, парабола и вот эта поверхность. Она называется *ласточкин хвост*. Не знаю, кто это придумал, и почему её так называют. Немножко на ласточку она похожа, конечно. Кажется, придумал это название Рене Том. А первым человеком, который исследовал эту поверхность, был Кронекер, который знал, что Бог создал целые числа, а остальное — дело рук человеческих. И вот, как дело рук человеческих, он исследовал эту поверхность.

Одно из определений этой поверхности таково. Возьмём многочлен 4-й степени. Член 4-й степени нормируем на единицу выбором масштаба, а кубический член убиваем сдвигом начала координат. Поэтому многочлен x^4+ax^2+bx+c — это, в сущности, произвольный многочлен 4-й степени. Такие многочлены образуют трёхмерное пространство $\mathbb{R}^3=\{x^4+ax^2+bx+c\}$; a, b и c — это координаты. Среди этих многочленов есть многочлены с кратными вещественными корнями: $\Sigma=\{(x-u)^2(x^2+2ux+v)\}$. Они образуют поверхность; эта поверхность и называется *ласточкин хвост*. Координатами на этой поверхности служат (u, v). Формула, которую я написал, задаёт поверхность параметрически. Точка с координатами (u, v) отображается в трёхмерное пространство с координатами a, b и c. Чтобы выразить a, b и c через u и v, нужно раскрыть скобки. Если проделать это, то можно сообразить, как эта поверхность выглядит. Это настоящее упражнение по анализу. И так как анализу больше не учат, то теперь ни один профессор, ни один студент, никто этого делать больше не умеет — кончено дело. А так как это к тому же полиномы, то это алгебраическая геометрия. А так как вместо алгебраической геометрии теперь учат какие-то модули и кольца, то и этого делать уже никто не умеет. Поэтому я нарисую ответ, а решение этой задачи предоставляю тем, кто действительно интересуется математикой — или физикой.



Р и с. 4. Ласточкин хвост

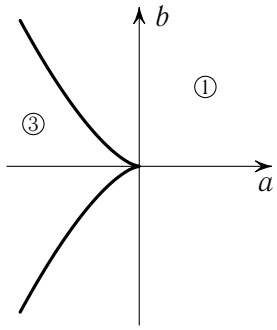
Сечение плоскостью $a=-1$ представляет собой кривую с двумя полукубическими особыми точками и одной точкой самопересечения (рис. 4). Если же взять плоскость $a=1$, то в сечении получится гладкая кривая. В сечении плоскостью $a=0$ должна получиться не очень гладкая кривая, потому что она должна быть переходной. У этой кривой есть особенность типа $x^4=y^3$. При всех отрицательных a в сечении есть треугольник; точки

возврата сами образуют ребро возврата на этой поверхности, которое имеет полукубическую особенность в начале координат. Двойные точки образуют гладкую кривую.

Эта поверхность делит трёхмерное пространство на 3 части. Так и должно быть, потому что вещественный многочлен четвёртой степени может иметь разное количество вещественных корней. Три области соответствуют случаям общего положения, а в общем положении многочлен четвёртой степени может иметь 4, 2 или 0 корней. Промежуточные случаи, 3 и 1, соответствуют границе.

З а д а ч а 1. В какой области число корней равно 4, в какой 2 и в какой 0?

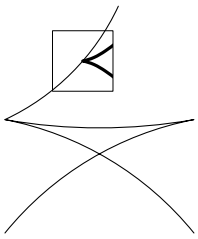
Эту задачу физики, вообще говоря, решают, а математики почти никогда. Здесь есть такие соображения. Области A, B, C имеют следующие соседства: $A \leftrightarrow B \leftrightarrow C$. (Область A расположена внизу; область C — пирамидка.) Из A, перейдя границу, всегда попадаем в B. Из B, перейдя границу, можно попасть либо в A, либо в C. А перейти из A в C, минуя B, нельзя. Конечно, перейти по рёбрам можно, но по поверхностям они не соединяются. Точно так же, 4, 2 и 0 соединяются только таким образом: $4 \leftrightarrow 2 \leftrightarrow 0$. Остальные переходы не общего положения. Мы приходим к выводу, что B — это область многочленов, имеющих 2 вещественных корня. Остаётся ещё вопрос относительно A и C. Область A большая, а область C маленькая, особенно в окрестности нуля. Что вероятнее, иметь 4 вещественных корня или не иметь их вовсе? Уравнение $x^4 + c = 0$, $c > 0$, не имеет вещественных корней. Это уравнение лежит в области A. Значит, уравнения с четырьмя вещественными корнями лежат в области C. Это, между прочим, так называемый принцип хрупкости хорошего. Если хорошо — это иметь много вещественных корней, и если мы находимся на границе (например, в точке P), то при случайном сдвиге мы скорее всего попадаем в плохую область, а не в хорошую. Чтобы наступило хорошее явление, нужно, чтобы выполнялось много условий. А при случайной деформации мы попадаем в плохую область. Поэтому точек, в которых хорошо — четыре корня, их мало, а точек, в которых плохо — нет корней, их много.



Р и с. 5. Полукубическая особенность

Нужно сказать, что особенности в задаче про эллипс устроены совершенно аналогичным образом, только нужно взять не многочлены четвёртой степени, а кубические многочлены $x^3 + ax + b$. Многочлены с кратны-

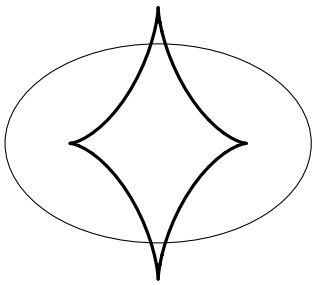
ми корнями образуют полукубическую параболу (рис. 5). У кубического многочлена в общем положении может быть 1 вещественный корень или 3 вещественных корня. Принцип хрупкости хорошего говорит, что область, где 1 вещественный корень, большая, а область, где 3 вещественных корня, маленькая.



Р и с. 6. Трансверсальное сечение ребра возврата

Между прочим, для ласточкиного хвоста тоже можно было бы сообразить, если знать, что здесь есть области с числом корней 4, 2 и 0, а на границе в трансверсальном направлении реализуется предыдущая ситуация (рис. 6). Действительно, полукубическая парабола является сечением в трансверсальном направлении, поэтому для многочленов четвёртой степени всё должно быть так же, как и для многочленов третьей степени. Если относиться к этому предмету не как логики, алгебраисты или бурбакисты, а относиться к нему как к предмету естествознания, то ответ очевиден. (Конечно, имеются другие точки зрения на естествознание.)

Прежде чем оставить этот пример, я сформулирую одну математическую теорему, доказывать которую не буду. Сейчас я хочу ввести ещё одно красивое и важное понятие, одно из самых важных в этой науке. Это понятие *каустики*. «Каустика» означает «жгущая». Это как раз та поверхность (или кривая на плоскости), где собираются лучи, где концентрируются лучи и где горит ваша расписка.



Р и с. 7. Астроида

Геометрия, которая описывается в этой ситуации, включает в себя лучи, нормали, фронты (я их уже раньше нарисовал), а ещё есть каустики. Каустика — это огибающая системы лучей. Я не буду рисовать все лучи в примере с эллипсом; это будет слишком сложная картинка. Все точки возврата, расположенные на всех фронтах, образуют кривую. Эту кривую, если вы такой хороший алгебраический геометр, как Кэли, можно найти. Можно найти её степень, посчитать особые точки и т. д. Можно её исследовать. Эта кривая называется *астроидой*; у неё 4 точки возврата (рис. 7). Это специальный случай для эллипса; если брать другие кривые общего положения, то у них тоже есть каустики. У этих каустик тоже есть точки возврата полукубического типа и они тоже устойчивы.

В трёхмерном пространстве каустики будут уже поверхностями. У этих поверхностей тоже есть типичные особенности. Между прочим, у них одна

из типичных особенностей — это тоже ласточкин хвост. Но есть ещё и две другие типичные особенности, о которых я расскажу позже. Они связаны с группами Вейля, или Коксетера.

Сейчас я хочу немножко рассказать про астроиду. В добрые старые времена, в XVII веке, астроида входила в курсы математики, и не знать того, что я сейчас расскажу, было нельзя. А нынче на первом курсе астроиду уже не проходят, и уже никто не знает тех элементарных фактов, которые, на самом деле, надо было бы уже в школе проходить, но которым почему-то не учат нигде. Я, например, для себя открыл их очень недавно, пару лет назад, и совершенно случайно, когда думал про эти самые волны.

Я сформулирую несколько математических задач. Их можно не связывать с волнами; они интересны сами по себе.

Рассмотрим пространство прямых на плоскости. Если плоскость считать проективной ($\mathbb{R}^2 \subset \mathbb{R}P^2$), то у этой плоскости есть двойственная проективная плоскость $\mathbb{R}P^{2\vee}$ — множество прямых в исходной проективной плоскости. Это общее понятие двойственности, которое, Бог знает, на первом курсе по линейной алгебре ещё проходят, а может быть, уже и не проходят больше. Но раньше, в старое время, когда ещё учили линейное программирование или оптимизацию, двойственность была. Ещё в более старые времена было преобразование Лежандра и уравнение Клеро. Все эти теории — это та же самая двойственность. Та самая, которая переставляет p и q.

В этой двойственности евклидова структура не играет никакой роли. Эта двойственность — понятие проективной геометрии. Есть две совершенно разные двойственности — евклидова двойственность (сопоставляющая каждому подпространству евклидова пространства его ортогональное дополнение) и проективная двойственность (сопоставляющая подпространствам одного пространства подпространства дополнительной размерности другого пространства). *) Проективная двойственность определяется следующим образом. Точка проективной плоскости представляет собой прямую в трёхмерном пространстве, проходящую через начало координат. У этого трёхмерного пространства есть двойственное пространство, которое тоже является трёхмерным векторным пространством. Точки этого двойственного векторного пространства — линейные функции в исходном пространстве. Если это двойственное пространство проективизировать, т. е. взять в нём прямые, проходящие через начало координат,

*) Я вставил эти тривиальные комментарии к понятию двойственности после того, как на опыте убедился, что учить этому в университетах перестали — а зря!

то эти прямые тоже образуют проективную плоскость; она называется двойственной к исходной. Чем задаётся линейная функция, отличная от нуля и заданная с точностью до умножения на ненулевую константу? Такая линейная функция задаётся своими нулями. Нули — это плоскость в трёхмерном пространстве, проходящая через начало координат. В проективных терминах — это прямая на проективной плоскости. Множество всех прямых в исходной проективной плоскости это и есть двойственная проективная плоскость.

Задача 2. Рассмотрим семейство всех нормалей к эллипсу. Это семейство является кривой в двойственном пространстве: $\Gamma^\vee \subset \mathbb{R}P^{2\,\vee}$. Что это за кривая?

Ответ такой: это *антиокружность*, которая в некоторой аффинной системе координат задаётся уравнением $x^{-2} + y^{-2} = 1$.

Для гладкой кривой на проективной плоскости можно определить проективно двойственную ей кривую. Для этого в каждой точке кривой проводится касательная, а затем все эти прямые рассматриваются как точки двойственной проективной плоскости. Та же самая конструкция позволяет построить для гиперповерхности в проективном пространстве проективно двойственную ей гиперповерхность. Двойственная кривая (гиперповерхность) не зависит от евклидовой структуры — это очень важно.

Важно также, что двукратное применение перехода к двойственному объекту возвращает к исходному объекту: например, поверхность, двойственная поверхности, которая двойственна к исходной выпуклой гладкой поверхности, есть сама эта исходная поверхность (это — основная теорема в теории двойственности, содержащая теорию уравнения Клеро и т. п.).

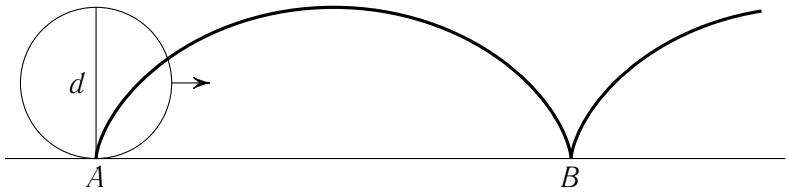
Задача 3. Возьмём антиокружность и рассмотрим проективно двойственную ей кривую. Какая кривая при этом получится?

Ответ такой: это астроида, т. е. $(x^{-2} + y^{-2} = 1)^\vee = (u^{2/3} + v^{2/3} = 1)$. Это означает, что можно выбрать координаты так, что проективно двойственная кривая будет задаваться таким уравнением.

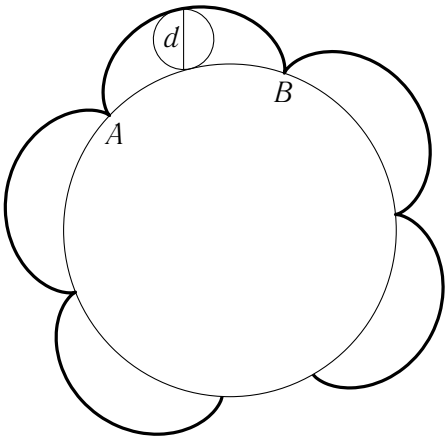
Из задач 2 и 3 следует, что огибающая системы нормалей к эллипсу есть астроида.

Про астроиду я хочу сказать ещё несколько слов. Астроиду можно ещё так определить. Раньше в школе учили, что если взять железную дорогу, взять точку на колесе и потом колесо покатить, то траекторией точки будет кривая (между прочим, с полукубическими точками возврата на рельсах), которая называлась *циклоида* (рис. 8). А если учесть, что Земля не плоская, то получится кривая, которая называется *эпициклоида* (рис. 9). Но если катить колесо внутри, то получится кривая, которая называется *гипоциклоида* (рис. 10).

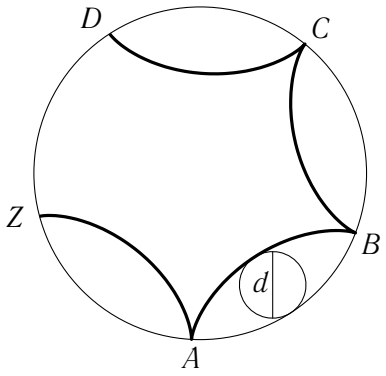
Если отношение радиусов рационально, то гипоциклоида — замкнутая алгебраическая кривая. А если отношение радиусов иррационально, то гипоциклоида плотна в некотором кольце. Есть интересные примеры. Например, если радиус внутренней окружности вдвое меньше радиуса внешней, то гипоциклоида — диаметр. Если отношение радиуса внешней окружности к радиусу внутренней окружности равно $3:1$, то гипоциклоида — кривая с тремя остриями (точками возврата).



Р и с. 8. Циклоида



Р и с. 9. Эпициклоида



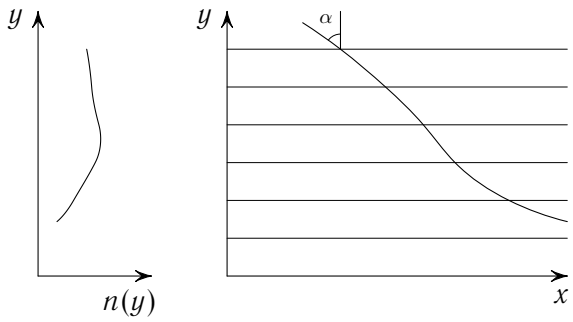
Р и с. 10. Гипоциклоида

З а д а ч а 4. Доказать, что если отношение радиуса внешней окружности к радиусу внутренней равно $4:1$, то гипоциклоида — астроида. То есть что эта циклоида задаётся в некоторой системе аффинных координат на плоскости уравнением $x^{2/3}+y^{2/3}=1$. Огибающая системы нормалей к эллипсу получается из этой кривой неодинаковыми изменениями масштабов по осям x и y.

Отсюда, между прочим, следует простая формула, параметрически задающая астроиду при помощи комплексных чисел.

1.
Неоднородные среды

До сих пор я рассматривал наиболее простой пример, когда метрика евклидова. Теперь я метрику испорчу, и рассмотрю некоторые примеры того, как распространяются лучи в неоднородных средах, когда они не прямые. Хотя понятия можно определять все те же самые, но обстоятельства получаются теперь более сложные. Простейший пример — когда имеются две среды (вода и воздух или воздух и стекло). Или немножко более общий пример, когда имеется слоистая среда, где плотность зависит от высоты (рис. 11), например, воздух над пустыней. Над пустыней именно так и обстоит дело, потому что температура над поверхностью, а следовательно и плотность воздуха, меняются, а из-за этого скорость света разная на разных высотах.



Р и с. 11. Слоистая среда

В предыдущем примере скорость распространения волн была всюду одинаковая, я её считал равной 1. А в этом примере она будет разная. Эксперименты показывают, что в воде скорость световых волн меньше, чем в воздухе. Если в воздухе скорость 1, то в воде скорость 3/4. А величина обратная к скорости называется *показателем преломления*. В прошлом веке в физике употреблялся термин, который сейчас почему-то оставлен, а очень был удобный термин, — *медленность*. Теперь её называют *лагранжиан*, ещё как-то. В общем, употребляют плохие называния, ничего не поймёшь. А на самом деле это медленность — единица, делённая на скорость.

Как будут распространяться лучи в такой ситуации? Можно говорить о геодезических; это будет принцип Ферма, который утверждает, что лучи распространяются по кратчайшим путям. Кратчайшие в том смысле, что свет проходит их за самое короткое время. Скорость известна, поэтому для каждой кривой можно сказать, сколько времени придётся идти свету по этой кривой. А потом нужно решать вариационную задачу, чтобы путь был покороче.

Уравнения можно писать по-разному. Но я сейчас немножко расскажу про экспериментальные результаты, которые подтверждаются любыми

теориями. Теорий здесь много, и все они приводят к одним и тем же результатам. Но эти результаты впервые были обнаружены экспериментально. Это называется закон Снеллиуса, или Снелла, который был открыт, по-видимому, экспериментальным путём. В задаче о слоистой среде закон Снеллиуса такой: $n(y)\sin\alpha = \text{const}$; α — угол луча с нормалью, $n(y)$ — показатель преломления.

Прежде чем доказать эту формулу, я расскажу её историю. Одним из претендентов на теорию, которую я сейчас буду рассказывать, является Декарт. Декарт является претендентом на основание всей французской науки. Он основал всю французскую науку на четырёх принципах, которые чрезвычайно сильно ей повредили. Вот эти четыре принципа.

Первый принцип. Не имеет никакого значения соответствие реальности исходных аксиом. Декарт не знал Гильберта, но говорил то же самое: исходные аксиомы — это произвольные утверждения. Имеют ли они какое-нибудь отношение к какой-нибудь реальности, не имеет никакого значения.

Второй принцип. Столь же малое значение имеет какое-либо соответствие какой-либо реальности окончательных выводов теории.

Третий принцип. Наука является своего рода обобщением умножения многозначных чисел. Это — формальная процедура, которая перерабатывает аксиомы по правилам аристотелевой логики. Важно не делать при этом ошибок. Всё остальное вообще просто чепуха. В частности, математика — не наука, потому что в ней используется нечто другое. Чтобы сделать математику наукой, надо изгнать из неё ненаучную часть, каковой являются все чертежи. Чтобы геометрия сделалась наукой, нужно изгнать из неё прежде всего чертежи. Вот в этом и есть дух Декарта, который впоследствии был принят Бурбаки. Изгнать чертежи — и всё.

Четвёртый принцип. Следует запретить все другие методы преподавания и обучения. Только мой метод является истинным настоящим хорошим методом. А причина такова: этот метод является истинно демократическим. Это означает, что при обучении моим методом самый посредственный ум получает такие же результаты, что и самый хороший. Ни при каких других методах этот результат не достигается. Поэтому все остальные методы в обучении, особенно школьном, надо запретить. Что и сделал Бурбаки. Во Франции сейчас геометрия изгнана вообще — из математики, в школе и в университетах изгнана. Старые книги, в которых были чертежи, выброшены из университетских библиотек.

Теперь я хочу сказать, чего достиг Декарт при помощи этого метода, применяя его к нашей задаче. Открытия Декарта в этой области опубликованы в его книге по оптике. Год опубликования этой книги 1996-й. Четыре

года назад работа Декарта была наконец опубликована. Он боялся её публиковать, потому что Джордано Бруно сожгли, у Галилея были неприятности. Теория Декарта категорически расходилась с теориями других, поэтому он боялся публиковать её. Его теория действительно противоречит учению Гюйгенса, о котором я рассказывал, и противоречит принципу Ферма, о котором я уже тоже говорил. Но, между прочим, ещё до Декарта его в каком-то смысле учитель Монтень объявил основные принципы не математики, а вообще всей французской науки. Они таковы.

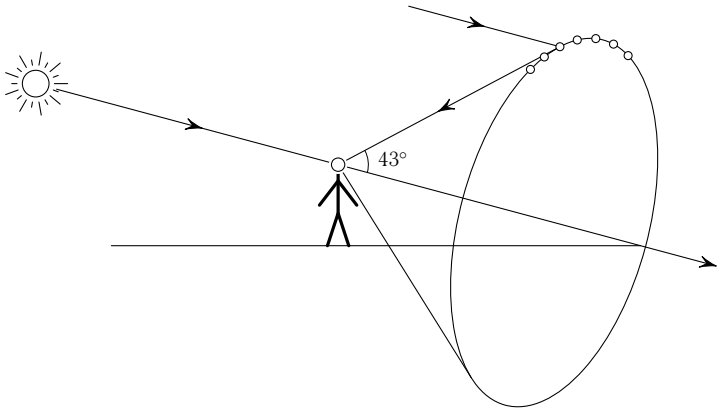
Первый принцип Монтеня. Если пишешь какую-то научную работу, то позаботься о том, чтобы никто не мог понять в ней ни одного слова. Потому что если хоть кто-нибудь что-нибудь поймёт, то все скажут, что ты ничего нового не открыл, что всё это было давно уже известно.

Второй принцип Монтеня. Для того чтобы во Франции твоя работа пользовалась успехом, ты не должен пользоваться ничьей чужой терминологией. Вся терминология должна быть введена самим тобой. Лучше всего в этой же работе, но разрешается ссылаться на твои предыдущие работы, чтобы заставить людей их читать. Это в крайнем случае тоже можно. Но категорически запрещаются любые ссылки на иностранцев.

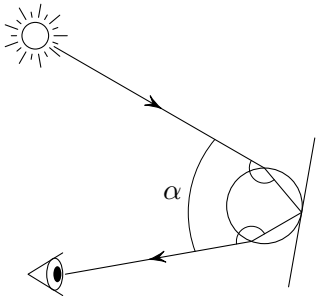
Поэтому Декарта не очень волновали противоречия, которые получились с его предшественниками. Первая теорема, которую получил Декарт, такова: «Скорость света в воде на 30% больше, чем в воздухе.» Если кто-нибудь знает уравнения Максвелла или какую-нибудь другую физику, то он понимает, что это совершенная чушь. Есть принцип относительности Эйнштейна и всякие другие принципы, согласно которым этого никак не может быть. Но Декарт не знал Эйнштейна, и к тому же это иностранец. Декарт, употребляя свои методы, которые не согласуются ни с какой реальностью, получал результаты, которые не согласуются ни с какой реальностью. Всё правильно, всё по его методу. Там только логика используется, и получается бурбакистское доказательство совершенно неверного результата. Замечательно!

На самом деле скорость света в воде на 25 % меньше, чем в воздухе. По-видимому, есть причина, по которой Декарт пришёл к такому выводу. Дело в том, что скорость звука в воде действительно больше, чем в воздухе. И довольно сильно — раз в 5. Поэтому Декарт думал: «Ну что такое свет? Это что-то вроде звука.» По-видимому, он так рассуждал. Не знаю, как он мог получить такой результат.

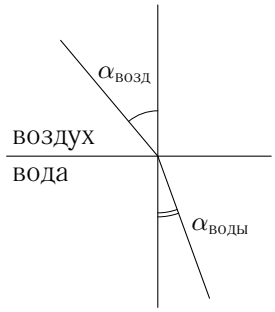
А второй результат, который получил Декарт, правильный. Этот замечательный глубокий результат — теория радуги. Теорию радуги Декарта



Р и с. 12. Радуга



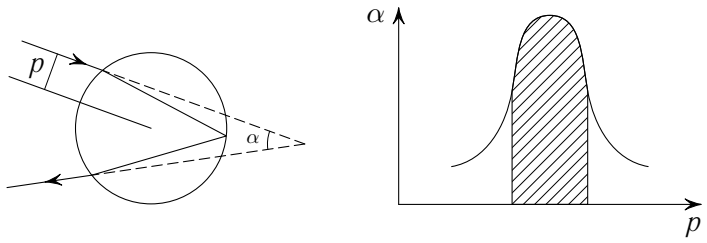
Р и с. 13. Капля воды



Р и с. 14. Преломление в воде

в школе тоже почему-то не учат. Я её сейчас расскажу. Основное утверждение этой теории таково. Пусть есть человек, наблюдающий радугу, и есть Солнце (рис. 12). Центр радуги находится в противосолнечной точке. Имеется конус с осью, проходящей через человека и центр радуги, и углом раствора (углом между осью и образующей) 43°. Точнее говоря, угол не совсем 43°; угол разный для разных цветов. Радуга разноцветная, потому что дуги разных цветов немножко разного радиуса. Но я этим пренебрегаю.

Это явление экспериментально давно известно. Оно, вероятно, описано уже Аристотелем. Хотя угла 43° у Аристотеля ещё нет. Декартово объяснение таково. Рассмотрим каплю воды. (Капля воды нужна, потому что если капель воды нет, если совсем сухо, то радуги не будет. Она бывает после дождя или перед дождём. Нужен туман — водяные капли.) Рассмотрим луч солнца, который падает на эту каплю (рис. 13). В соответствии



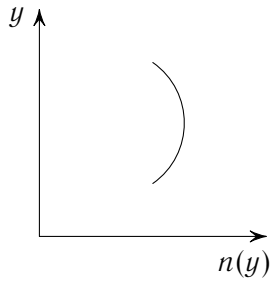
Р и с. 15. Функция $\alpha(p)$

с законами Снеллиуса происходит преломление. Для воды n больше, а произведение $n(y)\sin\alpha$ сохраняется, поэтому когда луч входит в воду $\sin\alpha$ должен уменьшиться, поэтому угол α должен уменьшиться (рис. 14). Сзади происходит полное внутреннее отражение: угол падения равен углу отражения, как у зеркала. А при выходе из воды снова происходит преломление.

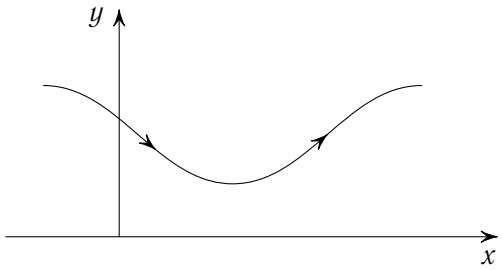
В результате выходящий луч образует с исходным лучом некоторый угол α. Этот угол можно вычислить, пользуясь нашей формулой. Но он зависит от того, в каком месте капли произошло падение луча. Если луч падает в другом месте, то геометрия будет немножко другой. Угол α является функцией от прицельного расстояния. Введём параметр p, нумерующий лучи, которые падают на одну и ту же каплю. Пусть $\alpha(p)$ — угол, под которым луч, соответствующий параметру p, отразится после всех этих операций. Декарт провёл вычисления, и оказалось, что функция $\alpha(p)$ не монотонна. У неё есть довольно резкий максимум, и высота этого максимума равна 43° (рис. 15). Энергия, которая попадает в окрестность этого максимума, очень велика. Там функция велика, а в других местах функция меньше. Поэтому интегралы, которые нужны, чтобы посчитать, сколько света куда попадёт, будут в основном определяться окрестностью этой точки. Это и означает, что те лучи, которые мы будем видеть, это те лучи, которые отражены под углом 43°.

Формулу $n(y)\sin\alpha = \mathrm{const}$ очень полезно применить в случае пустыни. В этом случае показатель преломления $n(y)$ должен иметь максимум на какой-то высоте (рис. 16). Это соответствует тому, что скорость света имеет минимум. (Для температуры это будет максимум.)

Там, где n имеет максимум, α имеет минимум — луч идёт наиболее круто. Если мы продолжим этот луч, то он выполаживается (рис. 17). А если посчитать аккуратно, то получится, что лучи поворачивают (по высоте); они идут то вверх, то вниз. Поэтому в пустыне бывают миражи. Закон Снеллиуса объясняет явление миража в пустыне. От зеркала тоже



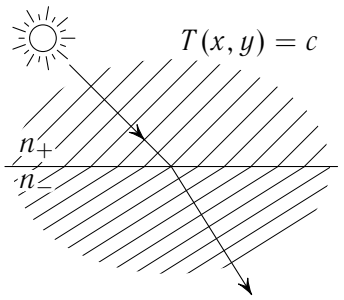
Р и с. 16. Показатель преломления над пустыней



Р и с. 17. Луч света над пустыней

луч идёт сначала вверх, потом вниз. Поэтому лучи, которые идут от неба, отражаясь и потом возвращаясь к нам, мы видим внизу; нам кажется, что там разлита вода и она отражает, как небо. Поэтому нам кажется, что там море или озеро, а там ничего нет. Просто этот слой создаёт такие отражения.

Формулу $n(y)\sin\alpha = \text{const}$, которая объясняет не только радугу, но и миражи, я докажу только для случая двух однородных сред. Можно брать n, зависящее от y, но это обычный трюк анализа с измельчением, с дифференцированием и т. д. Я сделаю это, когда есть только две среды, когда имеются n_- и n_+ (вода и воздух, например). Предположим, что есть луч, идущий из одной среды в другую. Внутри каждой однородной среды он будет, конечно, распространяться по прямой (рис. 18). Теперь сделаем такой трюк. Если есть какой-то источник, то рассмотрим функцию $T(x, y)$, которая есть оптическая длина пути от этого источника. По принципу Ферма T — минимизируемая функция. Нарисуем линии уровня функции T. Функция T гладкая в верхней полуплоскости и гладкая в нижней полуплоскости; на границе она непрерывна. Поэтому производная T по x должна быть одинакова сверху и снизу. Несложные вычисления показывают, что



Р и с. 18. Преломление луча света

$$\left.\frac{\partial T(x,\,y)}{\partial x}\right|_{y=0} = n_{\pm}\,\sin\,\alpha_{\pm}.$$

Это просто вопрос об обозначениях. Производная T — это $1/v$, потому что v — это производная x по T. Значит, производная T — это n. Подробности

я пропускаю. Декарт в этом месте запутался, перепутал где-то числитель со знаменателем, и получил неправильную формулу.

Ещё полезная задача для тех, кто действительно хочет как-то овладеть этим, такая.

Задача 5. На столе стоит стеклянная кастрюлька, наполненная водой. Мы на неё смотрим сбоку. Эксперимент показывает, что дно кастрюльки мы увидим не так глубоко, как оно на самом деле. На сколько мельче нам кажется вода, чем она есть на самом деле (если смотреть вертикально сверху)?

Ответ в этой задаче 2/3 (вблизи вертикали). От угла, под которым смотрим, ответ зависит, но не сильно.

2.

Связь между особенностями каустик и волновых фронтов и теорией групп и алгебр Ли

Определения объектов, которые я сейчас буду рассматривать, будут даны дальше, когда я буду рассказывать об аксиоматике. Однако теоремы верны, и эти теоремы можно понимать в слабом виде, без всех общих определений. А именно, можно рассматривать волновые фронты, например, в евклидовой геометрии. Рассмотрим евклидово пространство или риманово многообразие, на нём рассмотрим гиперповерхность, потом возьмём эквидистанты и посмотрим, какие возникают особенности. Это будем называть волновыми фронтами. Тогда будут верны те теоремы, которые я сейчас сформулирую. Но они верны и в гораздо более общей ситуации контактной геометрии. Однако эту гораздо более общую ситуацию я пока не определяю. Доказательства в обоих случаях тесно связаны, но это разные теоремы.

Теперь я должен определить объекты, которые являются многомерными обобщениями полукубической параболы и ласточкиного хвоста. Теорема состоит в том, что при некоторых условиях эти объекты как раз и будут особенностями многомерных волновых фронтов.

Само определение странным образом начинается с математики, не имеющей к делу никакого отношения. Для меня это было поразительным. Я случайно наткнулся на это обстоятельство в 71-м или 72-м году, занимаясь ради денег прикладной работой в Институте электронного машиностроения. Я рассчитывал, как греются какие-то электронные схемы. При расчёте того, как они греются, я обнаружил, что это определяется асимптотиками каких-то интегралов, эти асимптотики выражаются через какие-то рациональные числа, а эти рациональные числа выражаются

через числа Коксетера каких-то групп Коксетера. В то же время они описываются и каустиками. Так оказалось, что каустики связаны с группами Коксетера. И это было совершенно поразительным фактом, который за прошедшие почти 30 лет совершенно изменил всё лицо этой науки, теории распространения волн, потому что всё связалось с другой наукой, связи с которой были до этого открытия не видны.

Итак, я переключаюсь на совершенно другую науку. Рассмотрим евклидово пространство $\mathbb{R}^k$. *Отражением* называется ортогональное преобразование, у которого множество неподвижных точек является гиперплоскостью $\mathbb{R}^{k-1}$, проходящей через точку 0. Точки этой гиперплоскости остаются неподвижными, а каждый вектор, ортогональный ей, переходит в противоположный вектор. Одно собственное число оператора равно -1, а все остальные собственные числа равны $+1$. Это — ортогональное преобразование, меняющее ориентацию. Неподвижная гиперплоскость называется *зеркалом*.

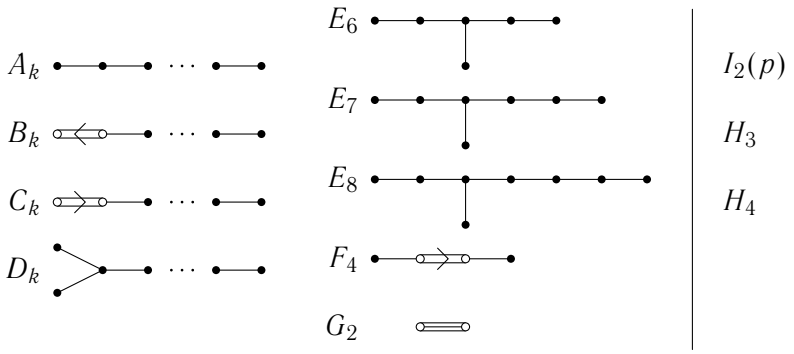
Если задано несколько зеркал, то отражения в этих зеркалах порождают группу. Например, на плоскости можно провести несколько прямых и рассмотреть отражения в этих прямых. Получится группа. Если взять две прямые на плоскости, то группа будет бесконечной тогда и только тогда, когда угол между прямыми несоизмерим с 2π. Действительно, группа содержит поворот на удвоенный угол между прямыми.

О п р е д е л е н и е. *Группой Коксетера*, или *группой отражений*, называется конечная группа, порождённая отражениями.

Нельзя говорить *группа Коксетера*, потому что алгебраисты заняли хорошее слово и дают этому термину совершенно неправильное определение. Поэтому нужно говорить иначе, будем говорить *группа отражений в евклидовом пространстве*.

Одна из главных классификационных теорем в математике состоит в том, что все группы отражений найдены. Имеется полный список этих групп. И этот список, оказывается, содержит огромное число классификаций самых разных объектов в самых разных областях математики. Например, сегодня я буду рассказывать, как эти объекты классифицируют особенности каустик и волновых фронтов. Они также связаны с простыми алгебрами Ли над комплексным полем и с многими другими разными объектами; не буду всё перечислять.

Сейчас я приведу этот список. Юрий Иванович Манин мне сказал, что все классификационные теоремы в математике приводят к этому списку по той причине, что он находится в hardware нашего мозга, поэтому мы ничего другого придумать не можем и всё подгоняем под эту схему. Я думаю, что это неправда. Как и в других вопросах, в этом Манин ошибается.



Р и с. 19. Классификация групп Коксетера

Юрий Иванович недавно опубликовал, что основное значение математики состоит в том, что она мешает прогрессу человечества.

Список групп Коксетера приведён на рис. 19.

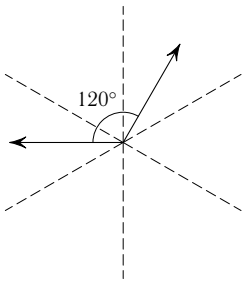
Чтобы группы были разными, нужно добавить такие условия: $A_{k\geqslant 1}$, $B_{k\geqslant 2}$, $C_{k\geqslant 3}$, $D_{k\geqslant 4}$.

Для меня сначала наибольшее значение будет иметь часть этого списка, а именно, группы A_k, D_k, E_6, E_7 и E_8.

Все группы из этого списка, кроме групп $I_2(p)$, H_3 и H_4, называются *кристаллографическими группами Коксетера*. Они дают также простые алгебры Ли. Группы E_6, E_7, E_8, F_4 и G_2 называются *экзотическими группами Коксетера*. Все группы Коксетера я сейчас опишу. Они описываются с помощью диаграмм, которые называются *диаграммами Дынкина* по общему принципу, который состоит в том, что никакой предмет не называется по имени изобретателя. Эти диаграммы Дынкина за десятки лет до Дынкина использовались Виттом и Коксетером, поэтому называются диаграммами Дынкина во всём мире. Диаграммы Дынкина кристаллографических групп Коксетера тоже приведены на рис. 19. Каждая точка диаграммы Дынкина — это вектор в евклидовом пространстве, который перпендикулярен зеркалу. Если два вектора перпендикулярны (или два зеркала перпендикулярны), то эти точки не соединяются. Если точки соединены отрезком, то угол между векторами равен 120°. Если точки соединены двумя отрезками, то угол между векторами равен 135°. Если точки соединены тремя отрезками, то угол между векторами равен 150°. Таким образом, каждая диаграмма определяет набор векторов, поэтому она определяет набор зеркал, а значит, и группу, порождённую отражениями. Что касается знаков «больше» и «меньше» (направлений стрелок), то они связны с тем, что группы B_k и C_k одинаковые, но есть две разные

решётки, которые сохраняются этими группами. «Кристаллографическая» означает, что группа сохраняет некоторую решётку. Так вот, B_k и C_k — разные решётки, которые сохраняются одними и теми же отражениями. Решётки мне не нужны, поэтому я про них больше говорить не буду.

Самый простой случай — группа A_2. Индекс k — это всегда размерность евклидова пространства. На 2-мерной плоскости есть два вектора, угол между которыми равен 120°. Рассматриваются отражения в нормалях (рис. 20). Угол между зеркалами тоже равен 120°, поэтому отражения порождают группу симметрий треугольника. Есть ещё третье зеркало; углы между зеркалами равны 60°. Группа отражений — это группа $S(3)$ (группа перестановок из трёх элементов). Эту группу лучше всего реализовывать следующим образом. Возьмём трёхмерное пространство с координатами (x_0, x_1, x_2) и возьмём группу перестановок координат. Эта группа порождена отражениями: например, транспозиция x_0 и x_1 — это отражение. Но эта группа приводима: в пространстве $\mathbb{R}^3$ есть диагональ $\Delta \subset \mathbb{R}^3$, заданная уравнениями $x_0 = x_1 = x_2$; эта диагональ инвариантна относительно всех перестановок. А так как все перестановки ортогональные, то инвариантно и ортогональное дополнение к диагонали $\Delta^\perp = \{x_0 + x_1 + x_2 = 0\}$. Это и есть двумерная плоскость, которая с зеркалами перестановок пересекается как раз по трём прямым, углы между которыми равны 60°. Геометрию равностороннего треугольника гораздо удобнее изучать не с помощью двух координат, а с помощью трёх. Если рисовать три координаты, то получается гораздо красивее.



Р и с. 20. Группа A_2

Так же определяется группа A_k. Нужно брать $x_0, \ldots, x_k$, рассматривать гиперплоскость $x_0 + \ldots + x_k = 0$ и действие на ней перестановок координат. Эта группа есть $S(k+1)$.

Группа, порождённая отражениями, называется *неприводимой*, если нет подпространства, инвариантного относительно действия всех преобразований группы. Есть теорема о том, что всякая группа, порождённая отражениями, разбивается в прямую сумму неприводимых, т. е. пространство разбивается в прямую сумму инвариантных ортогональных подпространств, на каждом из которых действие неприводимо. Тем самым, достаточно расклассифицировать неприводимые группы, порождённые отражениями.

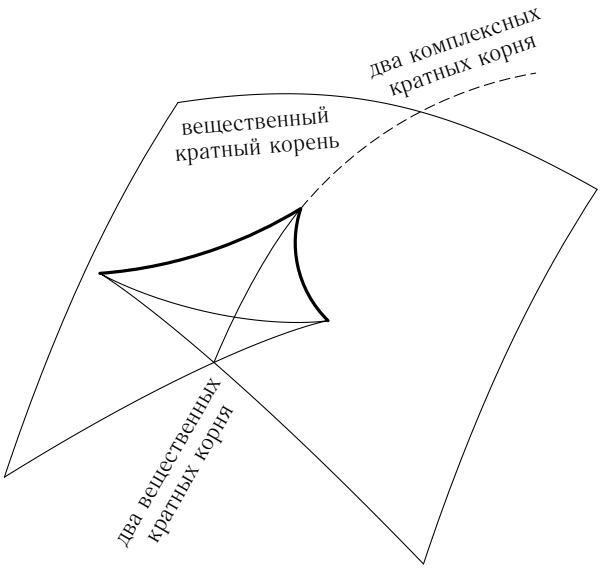
Можно взять, например, $A_1 + A_1$. Это — группа отражений относительно осей координат на плоскости. Эта группа приводима. Её диаграмма Дынкина состоит из двух изолированных точек.

Теперь мы приступаем к работе. Имея какую-то группу Коксетера, мы можем построить гиперповерхность, которая называется волновым фронтом, соответствующим этой группе. Эта конструкция такова. Легче это сделать в комплексной области, поэтому поместим вещественное пространство в комплексное пространство: $\mathbb{R}^k \subset \mathbb{C}^k$. Группа, которая действовала ортогональными преобразованиями в вещественном пространстве, действует и в комплексном. (Вещественная часть и мнимая преобразуются одинаковым образом.) Рассмотрим фактор $\mathbb{C}^k/G$.

Теорема 1. *Этот фактор гладкий, а именно,* $\mathbb{C}^k/G \approx \mathbb{C}^k$.

Для случая $G = A_k$ эта теорема называется основной теоремой теории симметрических функций. Мы берём на факторе функции. Функция на факторе — это функция от корней, которая инвариантна относительно перестановок корней многочлена. Значит, это функция от коэффициентов многочлена. Таким образом, для A_k в исходном пространстве $\mathbb{C}^k$ координатами были корни многочлена $x_0, \ldots, x_k$, где $x_0 + \ldots + x_k = 0$. Теперь мы берём инвариантные функции $\sigma_1, \sigma_2, \ldots, \sigma_{k+1}$ (σ_s обозначает сумму всех произведений s различных корней: $\sigma_1 = x_0 + \ldots + x_k$, $\sigma_2 = x_0x_1 + \ldots + x_{k-1}x_k$, $\ldots$, $\sigma_{k+1} = x_0 \ldots x_k$). Но $\sigma_1 = 0$, поэтому остаются функции $\sigma_2, \ldots, \sigma_{k+1}$. Основная теорема теории симметрических функций утверждает, что каждый многочлен от координат $x_0, \ldots, x_k$, не меняющийся при всех перестановках координат (т. е. симметрический) может быть записан в виде многочлена от переменных $\sigma_1, \ldots, \sigma_{k+1}$ («основных симметрических функций») Но аналогичная теорема есть и для гладких функций, и в вещественном пространстве.

Возникает отображение $\sigma\colon \mathbb{C}^k \to \mathbb{C}^k$ — факторизация по действию группы. Соответственно, в вещественном случае тоже возникает отображение $\mathbb{R}^k \to \mathbb{R}^k$ (корни отображаются в коэффициенты; аналогичная вещь есть и для всех случаев). В исходном пространстве $\mathbb{R}^k$ были зеркала. Образ зеркал в факторпространстве называется *дискриминантом*. В комплексном случае σ — это отображение на, поэтому комплексный дискриминант определён очень хорошо. Комплексный дискриминант — это гиперповерхность, которая определяется как образ одного зеркала (и всех, кстати, тоже — для A это всё равно). В вещественном случае есть три понятия дискриминанта, которые, вообще говоря, различны. Например, можно взять комплексный дискриминант, который лежит в комплексном факторе $\mathbb{C}^k$, и пересечь с вещественной частью $\mathbb{R}^k$ этого фактора. Другой вариант — можно взять вещественные зеркала и взять их образ. И так далее. В разных задачах оказываются полезными все варианты, которые тут можно придумать. Я не хотел бы уточнять, какой именно вариант этой конструкции мы выберем, потому что бывают полезны все варианты этой конструкции. Самый боль-



Р и с. 21. Продолжение ласточкиного хвоста

шой вещественный дискриминант — это комплексный дискриминант, пересечённый с вещественным фактором. Для случая многочленов (группа A) это будут вещественные многочлены, имеющие кратные корни — всё равно, вещественные или комплексные. Для A_2 все варианты вещественного дискриминанта дают одно и то же — полукубическую параболу. Для A_3 уже есть разные варианты. Для ласточкиного хвоста, оказывается, наибольший дискриминант будет другим. Линия самопересечения продлевается с другой стороны (рис. 21). Соответствующие многочлены имеют кратные корни в комплексной области. Хотя многочлены четвёртой степени имеют вещественные коэффициенты, у них может быть две пары комплексно сопряжённых кратных корней. Но кратных вещественных корней на этой части у них нет. Настоящий вещественный дискриминант поэтому не является вещественным алгебраическим многообразием, он является только полуалгебраическим многообразием. А если мы пересечём комплексный дискриминант с вещественной частью факторпространства, то мы получим алгебраическое многообразие, но другое. Полезно оба многообразия иметь в виду. В разных теоремах они работают. Оба они и взаимодействие между ними — всё это подробно изучено.

Т е о р е м а 2. *Для волновых фронтов общего положения в пространствах до размерности* 6 *никаких других особенностей, кроме дискриминантов* A, D *и* E, *не встречается.*

Например, в трёхмерном пространстве есть только A_3 — ласточкин хвост. В трёхмерном пространстве у волновых фронтов общего положения нет других особенностей. В четырёхмерном пространстве помимо A_4 появляется D_4. Эта поверхность хорошо изучена, но нарисовать её трудно, потому что рисовать нужно в четырёхмерном пространстве. В энциклопедии «Итоги науки» есть чертежи сечений этой поверхности, описание стратов, топологии. В пятимерном и шестимерном пространстве тоже есть конечное число особенностей. Начиная с размерности 7 — континуум; есть модули. Возникает классификация, которая уже не дискретна.

Как включить в эту схему остальные кристаллографические группы? Оказывается, что нужно построить теорию, аналогичную той, о которой я рассказывал, но для многообразий с краем. Я не буду подробно рассказывать, что такое распространение волн на многообразии с краем. Но понятно, что волны могут распространяться не только на многообразии без края, но и на многообразии с краем. Тогда ещё появятся B, C и F_4. Что же касается G_2, то оно появляется, если рассматривать многообразия с краем, который сам имеет полукубическую особенность.

Мне нужно ещё описать группы $I_2(p)$, H_3 и H_4. Индексы, как всегда, обозначают размерность пространства, в котором действуют группы.

Группа $I_2(p)$ — это группа симметрий правильного p-угольника на плоскости. Эта группа порождена отражениями. Нужно брать $p \neq 2$, 3, 4, 6. Случай $p = 2$ вырожденный; при $p = 3$ получается A_2, при $p = 4$ получается B_2, а при $p = 6$ получается G_2. Эти случаи уже есть в списке; они кристаллографические. Остальные случаи новые, например $I_2(5)$; они не кристаллографические.

Группа H_3 — это группа симметрий икосаэдра. Группа H_4 — это группа симметрий гиперикосаэдра. В четырёхмерном пространстве есть замечательный правильный многогранник, у которого 120 вершин. Этот многогранник настолько замечателен, что даже Бурбаки, которые посвятили теории групп, порождённых отражениями, 4 тома, его не описали. Хотя он был известен; у Коксетера он описан в 1928 г. У него есть очень красивая теория. Однако он не содержится в книгах Бурбаки, потому что его теория геометрическая, а не алгебраическая. Бурбаки не поместили описания этого многогранника. Алгебраисты держат в секрете описание этого замечательно красивого многогранника. Поэтому я сейчас его дам. Рассмотрим группу SO(3) вращений трёхмерного пространства. Как известно, эта группа неодносвязна. Её универсальная накрывающая накрывает её с кратностью 2. Это накрытие знают все физики, но не знают математики. В физике это явление называется *спином*. Спин имеет два значения, поэтому накрытие двулистное. Накрывающая группа в физике

называется Spin(3), а в математике она называется SU(2). Мне эти обозначения не нравятся. Я предпочёл бы сказать, что топологически SO(3) — это $\mathbb{R}P^3$. Значит, двулистное накрытие — это сфера S^3 (кватернионы с модулем 1). Теория спиноров — красивая теория. Для кватернионов есть аналог тригонометрической формы комплексного числа; кватернионам с модулем 1 можно сопоставлять вращения трёхмерного пространства. Или наоборот, вращению трёхмерного пространства можно сопоставить кватернион. Это, собственно, идея, из которой исходили Гамильтон, Родригес и многие другие. Правда, Гамильтону с трудом удалось придумать кватернионы, он смог это сделать только при помощи алкогольных паров. Когда Гамильтон, как следует воспользовавшись алкогольными парами, возвращался в туманный день и переходил в Дублине через мостик, ему пришла в голову формула $ij = k$.

Рассмотрим группу вращений икосаэдра. Эта группа состоит из 60 элементов в SO(3). Это школьный результат. Чтобы перевести икосаэдр в себя, нужно сначала перевести одну вершину в другую. Вершин 12. После этого нужно перевести одно из рёбер в другое ребро, выходящее из той же самой вершины. Там 5 вариантов. Всего получается $12 \cdot 5 = 60$ вращений. Прообраз этой группы из 60 элементов при двулистном накрытии $S^3 \to \mathrm{SO}(3)$ называется *бинарной группой икосаэдра* и обозначается Γ_{120}. Эта группа состоит из 120 элементов и лежит на S^3. А S^3, что для меня чрезвычайно существенно, лежит в $\mathbb{R}^4$ (кватернионы образуют четырёхмерное пространство). Мы получили 120 точек в четырёхмерном пространстве. Это и есть вершины многогранника, который я называю гиперикосаэдром. У него 600 граней, которые являются тетраэдрами. Это замечательный многогранник изумительной красоты, настолько красивый, что его не проходят в школе. Не знаю, почему. Его следовало бы изучать — какие у него грани разных размерностей, какова эйлерова характеристика. Для школьников — прекрасный объект изучения.

Группа симметрий гиперикосаэдра состоит из 120^2 элементов. Она является прямым произведением $\Gamma_{120} \times \Gamma_{120}$ (действие группы Γ_{120} на себе левыми и правыми сдвигами). Это и есть последняя группа H_4; она тоже не кристаллографическая.

Теорема 3 (Гивенталь). *Весь список групп, порождённых отражениями (включая кристаллографические и не кристаллографические), находится во взаимно однозначном соответствии с классификацией простых (т. е. без модулей) особенностей волновых фронтов.*

Детали используемого здесь определения простоты приведены на с. 190.

3.
Аксиоматика

Я хочу коротко рассказать об аксиоматике того, что же такое волновые фронты, что же такое каустики в симплектической и контактной геометрии; каково их общее описание. Я рассматривал случай римановой геометрии, которая является только частным случаем. А есть ещё симплектическая и контактная геометрии, и есть общее описание этих явлений. Мне потребуется некоторое описание этой новой науки. Надо сказать, что эти новые науки имеют забавную историю. В середине прошлого века была дискуссия по поводу того, что такое геометрия. Была большая дискуссия, и было много разных специалистов, которые выступали с противоположными высказываниями. Были специалисты, которые говорили, что геометрия должна быть проективной. (Как я, например, в начале лекции.) А были и другие специалисты, которые говорили, что геометрия должна быть евклидовой. Где нет длины, там нет геометрии. В конце концов этот вопрос был решён, когда появился тезис того же Кэли, на которого я уже ссылался. Кэли заявил следующее, и это было в конце концов всеми признано: «*Проективная геометрия есть вся геометрия*.» Потому что евклидова геометрия включается в проективную как часть проективной геометрии, в которой фиксирована какая-то дополнительная структура — квадратичная форма или что-то ещё, например, какая-то сфера. И если мы рассмотрим совместные инварианты, когда есть ещё какая-то дополнительная информация, например, рассмотрим проективную задачу, в которой есть эта сфера, то тогда это и будет евклидова геометрия. Значит, проективная геометрия всё содержит.

Я хочу закончить лекцию другим тезисом, который опровергает тезис Кэли. А именно, мой тезис состоит в том, что *контактная геометрия есть вся геометрия*. Контактная геометрия по отношению к проективной более общая. Есть ещё две геометрии — симплектическая и контактная. Но симплектическая геометрия играет скорее роль линейной алгебры, а контактная — роль проективной геометрии. Но эта пара относится как анализ к алгебре к паре, состоящей из проективной геометрии и линейной алгебры. Симплектическая и контактная геометрии — уже объекты сугубо бесконечномерные, где вместо операторов будут диффеоморфизмы и т. д.

Теперь я дошёл до последнего раздела — до определений. До сих пор я формулировал теоремы, но я не давал определения объектов, которые я классифицировал. Я говорил «волновые фронты», но что это такое, я не определял. Теперь я готов дать определение волновых фронтов и каустик, которое является общим. Для этого нужны основы симплектической и

контактной геометрии, и тогда будет дано общее определение. А когда будет дано общее определение, получатся и теоремы. Если начинать с этого общего определения, то получится теорема, которая такова. «При общем определении волновых фронтов список простых особенностей волновых фронтов находится в биективном соответствии со списком Коксетера.» Простые особенности — это те, у которых размерность пространства модулей равна нулю. (Подробно нужные объекты обсуждаются на с. 190.)

4.
Основы контактной геометрии

Основным объектом контактной геометрии является многообразие E^{2k-1} нечётной размерности. Основным примером, от которого происходит и название, является многообразие контактных элементов на некоторой базе. База B^k — это произвольное гладкое многообразие (размерности k). Контактным элементом на базе называется гиперплоскость *) в касательном пространстве в какой-то точке. Размерность пространства контактных элементов равна сумме размерности базы k и размерности слоя. Слой — это проективное пространство, точками которого являются все гиперплоскости в касательном пространстве; касательное пространство имеет размерность k, проективное пространство имеет размерность $k-1$. Поэтому пространство контактных элементов имеет размерность $2k-1$.

Пространство контактных элементов — нечётномерный аналог фазового пространства механики. Фазовое пространство имеет чётную размерность, а контактное пространство имеет нечётную размерность. Физики называют его *пространством конька*. Здесь $k=2$, B^2 — каток, а контактный элемент — это конёк. Положение конька определяется точкой приложения и направлением. В этом случае пространство E имеет размерность 3, и топологически E — это прямое произведение катка на окружность. В пространстве E есть замечательная структура, которая называется *условием конька*. Эта структура есть одно ограничение на скорость движения конька. А именно, конёк имеет право вращаться вокруг своей точки, которая касается катка, и имеет право ехать вперёд по своему направлению по любой кривой, но не имеет права сдвигаться поперёк. Это одно ограничение на скорость движения точки на E. Это означает, что в

*) Гиперподпространством (гиперподмногообразием) векторного пространства (многообразия) называется подпространство (подмногообразие) коразмерности один: прямая (кривая) на поверхности, плоскость (поверхность) в трёхмерном пространстве (многообразии).

точке на E есть гиперплоскость в касательном пространстве уже к E. Эта гиперплоскость в математике называется *тавтологической*. Дело в том, что есть расслоение $E^{2k-1} \xrightarrow{\mathbb{P}^{k-1}} B^k$, когда мы проектируем пару (точка, гиперплоскость) на базу и получаем точку, которая снабжена гиперплоскостью. Берём прообраз этой гиперплоскости наверху и получаем гиперплоскость наверху. Эта гиперплоскость (тавтологическая гиперплоскость, связанная с расслоением), называется *стандартной контактной структурой* в E.

Общее определение такое. В многообразии E^{2k-1} должно быть задано поле касательных гиперплоскостей. Локально поле гиперплоскостей задаётся 1-формой: $\alpha = 0$. Эта форма ещё не есть контактная структура, потому что её можно, не меняя структуры, умножить на функцию, отличную от нуля, и к тому же не требуется, чтобы 1-форма существовала глобально (ориентации не требуется). Есть единственное условие: 1-форма α должна быть максимально невырожденной. Условие максимальной невырожденности состоит в том, что есть одна самая большая орбита действия группы диффеоморфизмов на поля касательных гиперплоскостей, и в каждой точке форма должна быть именно в этой орбите. Можно написать необходимое и достаточное условие алгебраически. Это условие таково: $\alpha \wedge (d\alpha)^{k-1} \neq 0$. В трёхмерном случае, когда $k = 2$, получаем условие $\alpha \wedge d\alpha \neq 0$. Нетрудно проверить, что в примере с пространством конька условие конька удовлетворяет этой аксиоме.

Конец контактной геометрии.

Но прежде чем кончить контактную геометрию, нужно сформулировать, как всегда, основной принцип. Основной принцип контактной геометрии называется принципом Аллана Вайнстейна. По простой причине: потому что Аллан Вайнстейн никогда его не формулировал. Этот принцип такой: «*В контактной геометрии все интересные объекты являются лежандровыми многообразиями*.» Чтобы это стало понятно, нужно определить, что такое лежандрово многообразие. Я это сейчас сделаю.

На самом деле, Вайнстейн сформулировал аналогичный принцип в симплектической геометрии. Его принцип, действительно им сформулированный, и всё-таки называющийся по его имени вследствие какой-то ошибки, был такой: «*В симплектической геометрии все интересные объекты являются лагранжевыми многообразиями*.» А в контактной — лежандровыми.

Лежандровы многообразия — это интегральные многообразия контактных структур наибольшей возможной размерности. Нетрудно проверить, что наибольшая возможная размерность — это $k - 1$. Существуют $(k-1)$-мерные подмногообразия в E^{2k-1}, которые в каждой свой точке касаются

контактной структуры в E^{2k-1}. А большей размерности не бывают вследствие условия невырожденности.

Пример лежандрова многообразия такой. Можно взять гиперповерхность в B^k и взять все её касательные плоскости. Это будет лежандрово многообразие в E^{2k-1}. Это очень важное лежандрово многообразие, но не самое важное, потому что можно сделать по-другому. Можно вместо гиперповерхности взять одну точку в B^k. А чтобы получить $(k-1)$-мерное многообразие, возьмём все контактные элементы — все положения конька в этом месте. Тогда условие $\alpha = 0$ будет снова выполнено. Поэтому так тоже получаем лежандрово многообразие в E^{2k-1}. У нас есть два полярные примера, но эти примеры не всё исчерпывают. Имеется ещё такой важный пример. Возьмём в B^k любое подмногообразие и возьмём все контактные элементы (касательные к B^k гиперплоскости), которые касаются этого подмногообразия. Такое многообразие всегда $(k-1)$-мерное и всегда интегральное.

Когда в B идёт какое-то распространение волн, правильное описание этого процесса состоит в том, что надо для начального фронта взять все его контактные элементы. Они будут образовывать лежандрово многообразие в E. Затем в процессе распространения через время t образуется новый волновой фронт. Его контактные элементы образуют новое лежандрово многообразие. И это лежандрово многообразие гладкое, хотя фронт может быть и негладким. У фронта могут возникать особенности; мы это видели в случае эквидистанты эллипса. Однако лежандровы многообразия, которые лежат уже в трёхмерном пространстве, — это лежандровы узлы в трёхмерном контактном многообразии. У них нет особенностей. Лежандров узел не меняется — это основной принцип топологии лежандровых узлов, которая связана с распространением волн. Все особенности у фронтов происходят от проектирования.

Теперь мы можем наконец дать определение, что такое *лежандрова особенность* — основной объект классификации. Рассмотрим расслоение $E^{2k-1} \to B^k$. Предположим, что слои этого расслоения лежандровы (в нашем случае многообразия контактных элементов на B^k это так). Такое расслоение называют *лежандровым*. Рассмотрим лежандрово подмногообразие $\Lambda^{k-1}_{\text{леж}} \subset E^{2k-1} \to B^k$. Например, если $k=2$, то у нас есть плоскость, трёхмерное многообразие и кривая в трёхмерном многообразии. На плоскости получается кривая, вообще говоря, с особенностями. Эта кривая с особенностями (образ лежандрова узла) называется *фронтом*, а вся эта диаграмма называется *лежандровым отображением*. Можно взять росток в точке из $\Lambda^{k-1}_{\text{леж}}$. Этот росток называется *лежандровой особенностью*.

Классификация производится с точностью до некоторой эквивалентности. Эквивалентность — это коммутативная диаграмма

$$\begin{array}{ccccc} \Lambda^{k-1}_{\text{леж}} & \hookrightarrow & E^{2k-1} & \longrightarrow & B^k \\ \downarrow & & \downarrow & & \downarrow \\ \Lambda^{k-1}_{\text{леж}} & \hookrightarrow & E^{2k-1} & \longrightarrow & B^k. \end{array}$$

В этой коммутативной диаграмме средняя вертикальная стрелка сохраняет контактную структуру.

Эквивалентность определена, теперь можно сформулировать теорему.

Теорема 4. *Если размерность $k \leqslant 7$, то для лежандровых отображений общего положения все особенности эквивалентны особенностям нашего списка* (*особенностям A, D, E*).

Что же касается более широкого списка, то там теорема такая. Чтобы получить общий список, нужно расширить задачу следующим образом. Не нужно предполагать, что многообразие $\Lambda^{k-1}_{\text{леж}}$ является гладким. Нужно разрешить ему быть особым. Единственное требование состоит в том, чтобы у этой классификации пространство модулей было нульмерным, чтобы не было модулей. Тогда имеется биекция между особенностями и списком.

Полное доказательство этой теоремы довольно длинное, потому что нужно исследовать орбиты группы контактных преобразований. Это довольно долгие вычисления. Я приведу только некоторый пример, чтобы вы поняли, о чём тут идёт речь. Этот пример на самом деле очень явно работает, и он объясняет, в частности, почему в нашем списке есть диаграмма из трёх отрезков, выходящих из одной точки, и нет диаграммы из четырёх отрезков, выходящих из одной точки. Это можно объяснить, но это объяснение немножко мошенническое, потому что те вычисления, которые я сейчас приведу полностью, относятся к очень простому случаю. А те вычисления, которые надо проводить для нашей задачи, требуют много часов. Само вычисление требует многих часов, а в действительности оно появилось через несколько лет после того, как всё это было придумано. Совершенно неочевидная вещь. Вычисления — алгебра, как таблицу умножения считать или таблицу интегралов Ньютона. Не сразу всё это появилось. Архимед понял, что такое интеграл, а только Ньютон составил таблицу интегралов.

Рассмотрим следующий объект: три прямые на плоскости, проходящие через нуль. Контактную структуру мы уберём, т. е. будем рассматривать более простую задачу, но аналогичную. Эквивалентность в этой задаче —

линейные преобразования плоскости, которые одну тройку прямых переводят в другую тройку прямых.

Утверждение 1. *Размерность пространства модулей равна нулю.*

Доказательство. На плоскости можно выбрать систему координат так, что одна прямая будет осью координат x, другая прямая будет осью координат y, а третью прямую растяжениями можно привести к $x+y=0$.

Бывают вырожденные случаи, когда какие-то две прямые совпали. Тогда можно сделать двукратные оси. Число таких случае конечно, поэтому модулей нет. □

Теперь рассмотрим диаграмму из четырёх отрезков. Она соответствует четырём прямым, проходящим через нуль.

Утверждение 2. *Это объект не простой.*

Доказательство. Двойное отношение касательных к четырём кривым, проходящим через одну точку, сохраняется при диффеоморфизмах плоскости. Двойное отношение принимает значение в соответствующем поле, вещественном или комплексном. Это и есть модуль. □

Для тех, кого уже не учат двойному отношению, можно ограничиться подсчётом размерностей. Многообразие четвёрок точек проективной прямой (т. е. четвёрок прямых на плоскости, содержащих начало координат) четырёхмерно. Группа проективных преобразований проективной прямой трёхмерна. Следовательно, её орбиты не более, чем трёхмерны, и их конечное число не может всё четырёхмерное многообразие четвёрок покрыть. Значит, существует хотя бы один непрерывно меняющийся инвариант четвёрок (модуль). Явная формула для этого инварианта — двойного отношения четырёх точек с аффинными координатами (x_1, x_2, x_3, x_4) — объясняет, почему он называется двойным отношением: это

$$I = \frac{x_1 - x_3}{x_1 - x_4} : \frac{x_2 - x_3}{x_2 - x_4}.$$

В остальных случаях ситуация более или менее аналогичная. Надо сказать, что первые модули во всех этих задачах почему-то, по неизвестной мне причине (хотя я это открыл, но не могу объяснить, почему; я могу только это доказать), оказываются модулями эллиптических кривых. Двойное отношение тоже ведь связано с эллиптическими кривыми. Нужно взять четыре точки на проективной прямой и рассмотреть накрытие, ветвящееся в этих точках. Тогда получим эллиптическую кривую. Почему-то первый модуль всегда один, и он всегда есть тот самый модуль, который принимает значение в модулярной группе эллиптической кривой. А дальше классификация более сложная.

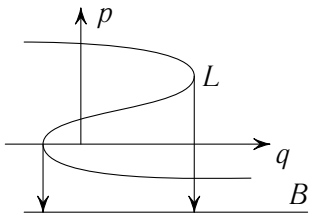
5.
Каустики

В заключение я хочу рассказать про каустики. До сих пор были фронты, а каустик не было. Каустики — это аналогичная ситуация, но в чётномерном пространстве; симплектическая геометрия. В основе симплектической геометрии лежит симплектическое пространство M^{2n} чётной размерности. Симплектическая структура — это 2-форма ω^2, которая замкнута, т. е. $d\omega^2 = 0$, и невырожденна, т. е. $(\omega^2)^{\wedge n} \neq 0$ ни в одной точке.

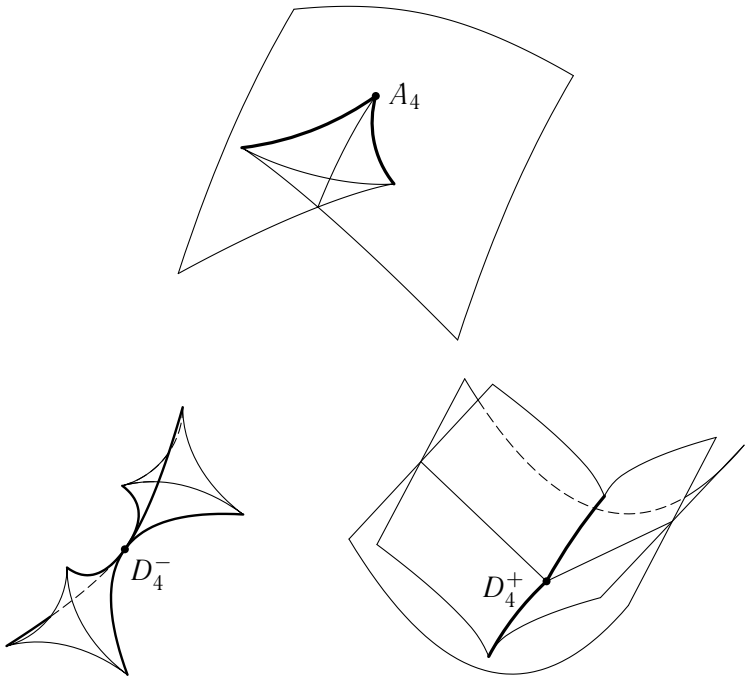
Лагранжево подмногообразие — это подмногообразие максимальной размерности, на котором $\omega^2 = 0$. Такая наибольшая размерность равна n. Требуется, чтобы ограничение ω^2 на пары касательных векторов к $L^n \subset M^{2n}$ было равно нулю.

Лагранжево расслоение — это расслоение $M^{2n} \to B^n$, слои которого лагранжевы. Классическим примером является кокасательное расслоение. Пространство кокасательного расслоения T^*B — это фазовое пространство с конфигурационным пространством B. На нём имеется каноническая (не зависящая от системы координат) 1-форма $\alpha = p\,dq$, называемая в механике *действием* и не известная математикам на первом курсе по причине плохого образования. Здесь q — координаты на B, а p — соответствующие им импульсы, т. е. координаты в кокасательном пространстве. Форма $\omega = d\alpha = dp \wedge dq$ называется ещё *интегральным инвариантом Пуанкаре*, или её интеграл называйте *инвариантным интегралом Гильберта*, в зависимости от того, геометр вы или алгебраист.

Рассмотрим теперь лагранжево подмногообразие $L^n \subset M^{2n}$ и его проекцию на базу B^n (на рис. 22 изображён случай $n = 1$). Для этой проекции имеются критические точки и их образы в B^n. В общем положении критические точки образуют $(n-1)$-мерную гиперповерхность в L^n. Их образы образуют $(n-1)$-мерную гиперповерхность в B^n, которая называется *каустикой* этого отображения. Эквивалентность каустик определяется через коммутативную диаграмму 3×2, которая уважает симплектическую структуру. Вопрос состоит в том, чтобы расклассифицировать лагранжевы особенности с точностью до эквивалентности. Ответ — тот же самый список. Но я не объяснил, как по этому списку строятся лагранжевы особенности. Сейчас я это скажу. То, что я сейчас построю, это универсальные модели, которые дают нормальные формы до размерности 6.



Р и с. 22. Проекция лагранжева подмногообразия



Р и с. 23. Лагранжевы особенности в размерности 3

Прежде чем приводить конструкцию, я приведу ответ для $n=1$, 2 и 3. Для $n=1$ всё нарисовано на картинке; другого не бывает. Если $n=2$, то особенность каустики общего положения полукубическая. Если $n=3$, то возможны три случая. При $n=3$ возможны A_4 и D_4. Вообще в симплектической задаче размерность сдвигается на 1 по сравнению с контактной: $n=k-1$. Группы A_4 и D_4 — это два случая, а ответов три. Причина состоит в следующем. Ответов два в голоморфной задаче, а в вещественной задаче ответов три, потому что D_4 имеет две вещественных формы: D_4^+ и D_4^-. Они эквивалентны в комплексной области, но не эквивалентны в вещественной: там есть некая сигнатура, которая их различает. Получаются разные картинки (рис. 23).

Все эти картинки хорошо изучены в лазерной оптике. В этой теории были некие общие топологические теоремы, которые впервые были угаданы именно лазерными физиками за счёт того, что они в эксперименте не наблюдали некоторых явлений, а потом уже математики доказали, что так и должно быть.

Картинки D_4^+ и D_4^- выглядят по-разному, но в комплексной области — это одна и та же картинка, как эллипс и гипербола.

Теперь общая теорема о том, как в терминах групп, порождённых отражениями, получать эти нормальные формы. Чтобы получить лагранжеву особенность A_k в $\mathbb{R}^{k-1}$, рассмотрим в $\mathbb{R}^k$ волновой фронт с особенностью A_k и рассмотрим расслоение на общие (трансверсальные касательной гиперплоскости хвоста) кривые.

Теорема 5. *Все такие расслоения локально эквивалентны в классе диффеоморфизмов, сохраняющих ласточкин хвост.*

В случае ласточкиного хвоста у нас были координаты (a, b, c). Это расслоение задаётся формулой $(a, b, c) \mapsto (a, b)$. Семейство кривых задаётся уравнениями $a = \text{const}$, $b = \text{const}$ $(da = 0,\ db = 0)$. К такому виду преобразованием, сохраняющим ласточкин хвост, можно привести любое семейство гладких кривых общего положения в окрестности вершины.

Аналогичный результат верен не только для группы A_3, но и для любой другой группы отражений.

Теперь спроектируем фронт вдоль кривых расслоения. На фронте есть ребро возврата. Образ ребра возврата при проектировании (он определён канонически, потому что проектирование определено канонически) — это и есть каустика. На самом деле, это отображение — просто дифференцирование многочленов. Здесь у нас был многочлен $x^4 + ax^2 + bx + c$. Забыть про c — это, в сущности, продифференцировать. Многочлен $4x^3 + 2ax + b$ приводится к виду $x^3 + ax + b$ перенормировкой коэффициентов, которая является диффеоморфизмом. Таким образом получается полукубическая парабола. Так же получены остальные картинки. Такова локальная теория.

Теперь я хочу сформулировать две нерешённые задачи. Я довольно много рассказал про то, что известно. Теперь расскажу про то, что неизвестно. До сих пор я говорил о локальных задачах. Теперь речь пойдёт о глобальных задачах. Глобальные задачи — это задачи о сосуществовании особенностей. В каком-то смысле все эти глобальные вопросы являются очень далёким обобщением теории Морса. У меня нет времени объяснять, при чём тут теория Морса и какое обобщение. Я это только сформулирую, а подробное описание этой науки можно найти в литературе. Я это придумал в 65-м году, но в действительности всё это восходит к так называемой *последней геометрической теореме Пуанкаре*. Это была первая теорема из этой глобальной теории — начало симплектической топологии. Однако впоследствии я обнаружил, что ещё до Пуанкаре был симплектический тополог, который уже применял топологию в задачах вариационного исчисления в целом и формулировал и даже доказывал теоремы вполне в стиле Пуанкаре, формулировал задачи и формулировал гипотезы. Странным образом, это был Якоби. Например, у Якоби есть работа, в которой сформулирована такая задача: «Построить теорию

Морса.» Там прямо сказано, что, рассматривая карту земной поверхности, на которой имеются горизонтали, мы убеждается, что там имеется столько-то максимумов, минимумов, сёдел. Они как-то соединяются. Что там может быть? Как это зависит от того, что Земля — шар, а не какая-то другая поверхность? У Якоби есть две странички, где прямо написано, что теория Морса является достойным предметом исследования, причём не только на сфере, но и на других поверхностях. Есть и многомерный случай, он и про это упоминает.

Сейчас я говорю про другую задачу. Вот задача, которую сформулировал Якоби, и которая на самом деле имеет к тому, о чём я рассказывал, непосредственное отношение. Но надо проходить через обобщение того, о чём я рассказывал, чтобы это объяснить. А саму задачу можно понять.

Рассмотрим в трёхмерном пространстве компактную поверхность, например эллипсоид. Рассмотрим точку на этой поверхности (например, северный полюс на сфере) и рассмотрим выходящую из этой точки геодезическую. Поверхность вложена в евклидово пространство, поэтому на ней есть метрика и есть геодезические. Выпустим из той же точки соседнюю бесконечно близкую геодезическую. В прошлом веке знали, что такое «бесконечно близкая геодезическая». Сейчас, кажется, уже не знают, поэтому нужно говорить «связность Леви—Чивиты» *), что-то такое нужно произносить. Бесконечно близкая геодезическая — я буду считать, что это известно. Эти две геодезические где-то снова пересекутся. Например, на сфере получим меридианы, которые снова пересекутся в южном полюсе. Точка пересечения двух бесконечно близких геодезических называется *сопряжённой точкой* исходной точки. Она же называется *фокальной точкой* — она связана с каустиками. Если мы возьмём первую геодезическую и начнём менять её произвольно, выходя из той же исходной точки по всем направлениям, то получим много сопряжённых точек, потому что на первой геодезической есть своя сопряжённая точка, на какой-нибудь второй геодезической есть своя сопряжённая точка и т. д. Все сопряжённые точки образуют какую-то кривую. Направления геодезических — окружность, и каждая точка этой окружности порождает первую сопряжённую точку, поэтому полученная кривая — образ окружности. Эта кривая называется *каустикой* исходной точки. (Якоби не использовал такого названия.)

Якоби доказывает следующую теорему. Это одна из первых топологических теорем вариационного исчисления в целом.

*) Физическое определение параллельного перенесения в этой связности таково: это предельно медленное («адиабатическое») перенесение направления колебаний маятника, колеблющегося около принудительно движущейся вдоль поверхности точки.

Теорема 6 (Якоби). *Каустика не может быть гладкой кривой. Каустика обязательно имеет особенности. Более того, в общем положении каустика имеет точки возврата и их чётное число.*

На сфере каустика вырожденная — она состоит из одной точки. Но если мы сферу слегка пошевелим, то получим уже не точку, а кривую.

Дальше Якоби формулирует гипотезу, которую он называет теоремой.

Гипотеза (Якоби). *На эллипсоиде на каустике любой точки имеется ровно четыре точки возврата.*

С одной стороны, это задача алгебраической геометрии, потому что Якоби проинтегрировал уравнение геодезических на эллипсоиде в тэта-функциях от двух переменных. Поэтому можно написать уравнение геодезических, продифференцировать, получить сопряжённые точки и уравнение каустик. Всё это, в принципе, можно написать явными формулами. Вопрос только в том, сколько там вещественных решений. Но эта геометрия хотя и алгебраическая, но вещественная. Поэтому алгебраические геометры совершенно неспособны что-либо сделать. А Якоби как раз умер. Самая последняя его работа — это попытка продвинуться в доказательстве этого утверждения, которое в его посмертно изданных лекциях по динамике сформулировано как теорема. Но это записывали его ученики; он сам не писал этого текста. Там это записано как теорема, но он никакого доказательства не рассказывал. Я предпочитаю называть это утверждение *гипотезой Якоби*. Но эта гипотеза может относиться к алгебраической геометрии, а может относиться и к Computer Science. Потому что все эти объекты вполне поддаются вычислению на компьютере. И если на компьютере будет обнаружен какой-то пример, в котором точек возврата две, тогда эта гипотеза будет опровергнута, потому что две точки возврата — это устойчивая вещь, это будет и в некоторой окрестности. Но те люди, которые занимаются Computer Science, они ещё хуже алгебраических геометров. Они даже понять не могут, что есть такая задача. Я пытался уже с многими это обсуждать, но они совершенно неспособны понять, что такое геодезическая, сопряжённые точки.

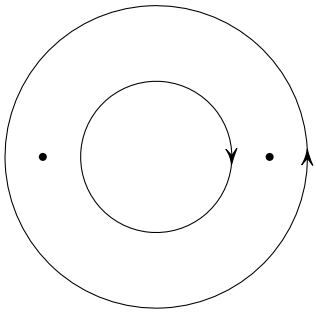
Каустик много. Мы брали первую сопряжённую точку. То, что получилось, это первая каустика. А можно взять вторые сопряжённые точки. Тогда тоже получится замкнутая кривая, которую можно назвать *второй каустикой*. Аналогичный вопрос можно поставить про все каустики, не только про первую. Доказано, что на любой выпуклой поверхности на первой каустике не меньше четырёх точек возврата. Их может быть шесть и действительно бывает на других поверхностях — не на эллипсоиде. Якоби утверждал, что на эллипсоиде (не сказано, на какой каустике; может быть, он думал про первую каустику, а может быть и нет) число точек возврата —

ровно четыре. Но, может быть, это сказал не он, а его ученики, которые плохо записали его лекции.

Второй пример, который я хочу сформулировать, — это так называемая *гипотеза Арнольда*. Она опубликована в Comptes Rendus в 1965 г. Эта гипотеза является далёким обобщением последней геометрической теоремы Пуанкаре. Геометрическая теорема Пуанкаре так называется потому, что её доказал Биркгоф после смерти Пуанкаре. Эта теорема такова.

Т е о р е м а 7. *Пусть на плоскости имеется кольцо и имеется диффеоморфизм кольца на себя, который граничные окружности сдвигает в разные стороны* (рис. 24) *и, кроме того, сохраняет площади. Тогда у него не меньше двух неподвижных точек.*

В гипотезе Арнольда вместо кольца рассматривается произвольное компактное симплектическое многообразие (M^{2n}, ω). Первый содержательный пример (в котором, правда, эта гипотеза доказана) — это двумерный тор $(T^2, dp \wedge dq)$. Рассмотрим теперь отображение, которое является симплектоморфизмом. Симплектоморфизм задаётся однозначным гамильтонианом, зависящим от времени. Это делается так. Нужно построить векторное поле $\dot{x} = v(x)$, $x \in M^{2n}$. А векторное поле строится так. Берём функцию $H(x, t)$, $t \in S^1$. Эта функция называется *функцией Гамильтона*. Возьмём дифференциал этой функции $d_x H$ по x (t считается константой). Симплектическая структура задаёт изоморфизм касательного и кокасательного пространства в каждой точке, потому что она невырожденная. Обозначим его $I_x\colon T_x^* M^{2n} \to T_x M^{2n}$. Это то же самое, что ω; билинейная форма она же является изоморфизмом между самим пространством и двойственным пространством. Возьмём dH и применим к нему I: $I_x d_x H = v(x)$; это поле v называют *гамильтоновым* с функцией Гамильтона H. Фазовый поток этого поля сохраняет симплектическую структуру. Но если там есть t, то система неавтономная, зависит от времени, но это ничего. Получается 1-параметрическое семейство диффеоморфизмов, которые являются симплектоморфизмами. В условии геометрической теоремы Пуанкаре это соответствует сохранению площадей. В простейшем примере, для тора, это будет так. Запишем на универсальной накрывающей отображение в виде $x \mapsto x + f(x)$. Условия на отображение будут такие: якобиан отображения равен 1 и среднее значение вектор-функции f равно 0 (функция f периодическая, а среднее



Р и с. 24. Геометрическая теорема Пуанкаре

значение периодической функции — это интеграл по периоду). Условие, что окружности сдвигаются в противоположные стороны, превращается в условие гомологичности нулю, или точности. Гомологичность нулю означает, что dH — это дифференциал функции, а не замкнутая 1-форма.

Гипотеза (Арнольд, 1965). *Число неподвижных точек не меньше числа Морса для* M^{2n}.

Имеется уже порядка 100 работ, где это доказывается при разных условиях. Первым, по-видимому, это доказал Элиашберг. Однако его доказательство я отверг 4 раза, потому что я находил контрпримеры к его леммам. Но он утверждает, что после этого он опубликовал правильное доказательство в 78-м году в Сыктывкаре. Но это доказательство никто не прочёл. И это пятый вариант после четырёх неверных. Поэтому есть такая задача — выяснить, верно это доказательство или нет, потому что никто его не понимает. В 83-м году Конли и Цендер создали новую теорию на основании идей Рабиновица, которые, в сущности говоря, есть перенесение на бесконечномерную ситуацию теории систем Аносова, и доказали гипотезу Арнольда для тора и для некоторых других случаев. Затем была построена теория Флоера, потом Громова, потом при разных дополнительных условиях многие другие — Корнфельд, Гивенталь — это доказывали. В многочисленных работах многих людей была построена большая наука — симплектическая топология. В этой науке доказывается, в частности эта гипотеза. Я разбирал доказательство Конли и Цендера. У них есть очень красивая работа, в которой это очень красивым образом сделано для тора размерности $2n$ со стандартной симплектической структурой. Для кэлеровых многообразий (в частности, для всех двумерных ориентируемых поверхностей) это сделал потом Флоер. В последнее время три группы людей (фамилии которых я не буду называть, потому что в каждой группе по три-четыре человека, которые публикуют по три-четыре работы, и у разных работ разные авторы, и после этого уже не разберёшься, кто же участвует в окончательном доказательстве) претендуют на доказательство этой гипотезы и даже их работы так и называются. Я в этих доказательствах ничего не понимаю. Это с одной стороны. А с другой стороны, когда я смотрю на формулировки, я вижу, что они доказывают что-то другое. Может быть, они умеют это доказывать, а может быть и нет. Это мне непонятно. Все эти доказательства основаны на одной лемме Концевича. Концевич на моём парижском семинаре пытался рассказать эти доказательства, и запутался. Он говорит, что его лемма верная, но как они её применяют, он не понимает. Ситуация сложная. Доказана гипотеза или нет, я не знаю. Основная трудность состоит в том, что во всех опубликованных текстах доказывается, что число неподвижных точек не

меньше, чем сумма чисел Бетти. А я утверждал, что оно не меньше числа Морса. А это не одно и то же. Для тора это, конечно, одно и то же.

Конечно, об оценке числа неподвижных точек снизу через число Морса речь идёт о симплектоморфизме общего положения, когда эти точки невырожденные (или же они считаются с кратностями). Если же считать только геометрически различные неподвижные точки, то и число точек надо заменить минимальным числом геометрически различных критических точек функции.

У всей теории, которую я сегодня рассказывал, есть ещё один вариант, который физиками называется *квантовой теорией катастроф*. Этот вариант, хотя он после некоторых преобразований сводится к тому же самому и хотя он был основой моих исходных построений, в которых все эти науки возникли в 70-е годы, выглядит иначе. Если вы рассмотрите симплектическую геометрию в квантовом варианте, то возникнет такая задача. Рассмотрим интеграл

$$I_\hbar(\lambda) = \int\limits_{x\in\mathbb{R}^n} e^{i\frac{S(x,\lambda)}{\hbar}} a(x,\,\lambda)\,dx.$$

Параметр $\hbar$ называется *постоянной Планка* и считается очень маленьким: $\hbar \to 0$. Он играет роль длины волны. Функция S вещественна; она называется *фазой*. Функция a может быть как вещественной, так и комплексной; она называется *амплитудой*. Распространение волн в оптике и в квантовой механике описывается интегралами такого рода. Принцип стационарной фазы говорит, что когда $\hbar$ очень мало, то для того чтобы найти основной вклад в этот интеграл, нужно поступить следующим образом. Функцию S нужно рассматривать как семейство функций от x, зависящее от параметра λ. Нужно взять критические точки: $\dfrac{\partial S}{\partial x} = 0$. При λ общего положения критические точки будут морсовские. Простая выкладка с интегралом Френеля показывает, что около морсовской критической точки вклад в интеграл равен $\hbar^{n/2}$ (с коэффициентом, который зависит от амплитуды). Но если при каких-то каустических значениях λ критические точки начинают сливаться, то там асимптотика будет другая. На каустике асимптотика будет $\hbar^{n/2-\beta}$. Число β называется *показателем особости*. Оно отвечает за интеграл Эйри и за интеграл Пирси, насколько ярче радуга — всё это описывается этим показателем. Между прочим, первым человеком, который начал считать эти показатели, был Иван Матвеевич Виноградов. Он это делал для сумм, но суммы оценивал при помощи интегралов. Для серии A у него были такие оценки.

Если мы знаем классификацию особенностей, то мы знаем эти числа β, а значит, знаем асимптотики. Оказывается, что эти асимптотики вы-

ражаются через числа Коксетера соответствующих особенностей. В моих публикациях можно найти все списки — это хорошо известные ответы.

Для особенностей $I_2(p)$, H_2 и H_3 эта теория не построена. У меня эта теория построена для A, D, E. У моих последователей она построена для B, C, F и даже для G_2. Я думаю, что у Гивенталя это всё уже написано. Но ещё до него это делали Щербак, Варченко, Ляшко. Есть целая серия работ, в которой всё это исследовалось, и там тоже появляются числа Коксетера для соответствующих особенностей. А для некристаллографических особенностей теория осциллирующих интегралов не построена. Но я думаю, что она должна существовать. До сих пор всё, что открывалось сначала для A, D, E, потом переносилось на кристаллографические групп, а потом продолжалось и на некристаллографические группы.

Квантовая теория катастроф для некристаллографических групп должна быть как-то связана не с кристаллами, а с квазикристаллами. Но как раз в этом месте наука отстаёт, потому что люди, которые занимаются квазикристаллами, редко бывают математиками.

Вот одно из последних достижений топологической теории особенностей волновых фронтов. Мы видели, что в семействе эквидистант эллипса встречаются фронты с четырьмя точками возврата, а при дальнейшем распространении сжимающегося фронта он выворачивается наизнанку и потом становится гладко расширяющимся (рис. 2 и рис. 3). Аналогичное выворачивание происходит с любой замкнутой исходной кривой вместо эллипса и при любом законе распространения волн (важно лишь, что лежандров узел в процессе распространения не самопересекается, что происходит, например, для любой скорости распространения, зависящей и от точки, и от направления, и даже от момента времени). Достижение состоит в том, что при любом таком выворачивании обязательно приходится пройти через фронт с не менее чем четырьмя точками возврата. Это было недавно доказано Ю. В. Чекановым и П. Е. Пушкарём, а в качестве гипотезы было высказано мною в 1993 г., когда я это доказал при некоторых дополнительных предположениях. Вся эта теория является далёким обобщением теоремы Штурма, согласно которой вещественный ряд Фурье имеет не меньше нулей, чем гармоника наинизшего порядка из входящих в него с ненулевыми коэффициентами. В свою очередь, эта теорема Штурма (доказанная Гурвицем) является обобщением на высшие производные теоремы Морса, согласно которой гладкая функция на окружности имеет не менее двух критических точек, где $df = 0$. Оператор d в теореме Штурма заменён на $L = d(d^2+1)(d^2+4)\ldots(d^2+n^2)$, а число нулей Lf оказывается не меньшим чем $2n+2$. Теорема о выворачивании доставляет удивительное нелинейное обобщение случая $n=1$ в теории

Штурма, в котором число обобщённых критических точек, где $Lf = 0$, оказывается не меньшим четырёх. Доказательство Чеканова и Пушкаря использует некоторые идеи из предшествовавшей работы другого моего ученика, С. Баранникова (о теории Морса продолжения на многообразие функции, заданной вблизи его края).

28 сентября 2000 г.

А. А. Болибрух

ПРОБЛЕМА РИМАНА—ГИЛЬБЕРТА

Я хочу рассказать об этой старой известной проблеме, о том, как она решалась, какова её история. Вторая моя цель — показать, как эта проблема может быть сформулирована на геометрическом языке с использованием простейших понятий алгебраической геометрии, таких как расслоение, связность, стабильность. Такая формулировка позволяет правильно понять всю суть этой проблемы и даже получить новые результаты в этой области. Я надеюсь рассказать в конце об этих новых результатах, о том, в каком состоянии эта проблема находится сейчас, и какие обобщения на случай компактной римановой поверхности и на случай комплексного многообразия произвольной размерности здесь имеются.

Сначала я расскажу о постановке проблемы и коротко — о её довольно запутанной истории. В программе семинара «Глобус» я прочитал, что Ю. С. Ильяшенко рассказывал тут про 16-ю проблему Гильберта, которая тоже имела исключительно сложную историю. Там, как вы помните, были достижения типа теоремы Дюлака, были ошибки, были снова достижения и интересные теоремы типа теоремы Ландиса, снова были ошибки. Чем-то эта история напоминает историю 21-й проблемы Гильберта. Здесь тоже было много ошибок, много интересных продвижений. Эти проблемы очень похожи по своей истории и по тому, как развивалось исследование этих проблем. И по существу эти проблемы довольно близкие.

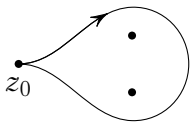
Давайте я сформулирую саму задачу. Она связана с таким простым объектом, как система линейных дифференциальных уравнений на расширенной комплексной плоскости, или на сфере Римана:

$$\frac{dy}{dz} = B(z)y, \quad z \in \bar{\mathbb{C}}.$$

Здесь $B(z)$ — известная матрица коэффициентов, а $y = (y^1, \ldots, y^p)^T$ — неизвестная вектор-функция из p компонент. Матрица $B(z)$ голоморфна вне множества особых точек $\Sigma = \{a_1, \ldots, a_n\}$. В особых точках матрица мероморфна.

Если есть такая система и есть особые точки, то у этой системы есть *монодромия*. Рассмотрим некую неособую точку z_0 и выходящую из

неё петлю γ, не проходящую через особые точки (рис. 1). Рассмотрим в некоторой окрестности точки z_0 фундаментальную матрицу решений $Y(z)$. Я напомню, что у такой системы решения образуют линейное пространство размерности p над полем комплексных чисел. Поэтому в пространстве решений можно выбрать базис. Если записать координаты каждого вектора этого базиса в столбик, то получится матрица $Y(z)$. Матрицу решений $Y(z)$ можно аналитически продолжить вдоль пути γ — это известная теорема из теории дифференциальных уравнений о том, что решение аналитически продолжается вдоль любого пути, не задевающего особых точек. После такого продолжения мы вернёмся в ту же самую окрестность точки z_0. Оказывается, что при этом получится некоторая другая фундаментальная матрица решений, т. е. базис перейдёт в другой базис. Получим отображение $Y(z) \mapsto \hat{Y}(z)$. При этом $Y(z) = \hat{Y}(z)\, G_\gamma$, где G_γ — некоторая постоянная матрица. Оказывается, что соответствие $\gamma \mapsto G_\gamma^{-1}$ зависит только от гомотопического класса пути γ. Путь можно деформировать, на задевая особых точек; это не влияет на результат продолжения, поэтому матрица получается та же самая. Это простая теорема из комплексного анализа (теорема о монодромии). Оказывается, что это соответствие порождает *гомоморфизм* фундаментальной группы

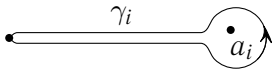


Р и с. 1. Монодромия

$$\chi\colon \pi_1\left(\bar{\mathbb{C}} \setminus \Sigma,\, z_0\right) \to \mathrm{GL}(p,\, \mathbb{C})$$

— *представление монодромии*. То, что это гомоморфизм, легко проверяется. Образ этого гомоморфизма называется *группой монодромии* уравнения.

Фундаментальная группа проколотой сферы Римана устроена очень просто. Её образующие — простейшие петли, которые обходят каждую особую точку в отдельности (рис. 2). Гомотопические классы g_i таких петель и задают образующие фундаментальной группы. Их ровно n штук. Между ними есть ровно одно соотношение $g_1 \ldots g_n = e$ (результат последовательного обхода всех петель подряд стягивается вне особых точек, с «обратной стороны» сферы Римана, в точку). Я часто буду говорить «петля», подразумевая при этом «гомотопический класс петли».



Р и с. 2. Образующие фундаментальной группы

Сама монодромия полностью определяется образами образующих фундаментальной группы, т. е. определяется аналитическим продолжением фундаментальной матрицы решений вдоль каждой образующей по отдельности. Всё определяется матрицами $G_1, \ldots, G_n$, которые задают

изменение фундаментальной матрицы решений при аналитическом продолжении вдоль образующей. Эти матрицы априори удовлетворяют только одному соотношению $G_1 \dots G_n = I$. Так описывается монодромия. Конечно, можно задать вопрос, что будет, если мы поменяем начальную точку или поменяем фундаментальную матрицу решений. Оказывается, что тогда все матрицы монодромии одновременно поменяются на сопряжённые. Я не буду это проверять; это несложно сделать. Если вы стартуете с другой фундаментальной матрицы, то она отличается от исходной на постоянную матрицу. Монодромия будет отличаться только тем, что все матрицы сопрягутся одновременно. Так что монодромия определяется с точностью до *) сопряжения.

Среди таких систем дифференциальных уравнений можно выделить наиболее простые системы. Какие самые простые особенности может иметь матрица $B(z)$ в особых точках? Полюсы первого порядка. Это означает, что в окрестности особой точки a_i матрица $B(z)$ представляет собой голоморфную функцию, делённую на $z - a_i$. Класс таких систем называется *фуксовыми системами*. Общий вид произвольной фуксовой системы очень прост:

$$\frac{dy}{dz} = \left(\sum_{i=1}^{n} \frac{B_i}{z - a_i} \right) y.$$

Здесь B_i — постоянные матрицы; они от z не зависят. Всякая система, имеющая только простые полюса (полюса первого порядка) в особых точках, имеет такой вид. Это очень лёгкое упражнение.

Именно с такими системами связана знаменитая 21-я проблема Гильберта, которая была им сформулирована в 1900 г. Она чётко и ясно сформулирована следующим образом: «Показать, что всегда существует линейное дифференциальное уравнение фуксова типа с заданными особыми точками и с заданной монодромией.» Другими словами, 21-я проблема Гильберта — это обратная задача в теории фуксовых систем дифференциальных уравнений. Гильберт именно так жёстко сформулировал свою проблему: доказать, что для любого набора особых точек $a_1, \dots, a_n$ и любого набора матриц $G_1, \dots, G_n$ найдётся фуксова система уравнений с заданными особыми точками и с заданной монодромией.

Эта проблема рассматривалась в разных аспектах. Кроме фуксовых систем рассматривались системы с регулярными особыми точками. Там эта задача тоже имеет смысл, и там тоже есть интересные результаты. Но оказалось, что самый сложный вариант проблемы связан именно с

*) Общего. — *Прим. ред.*

фуксовыми системами. И именно история задачи применительно к этим системам оказалась наиболее интересной и запутанной и наиболее важной для приложений.

Теперь я перехожу к истории изучения этой проблемы и немножко скажу о том, какие приложения она имеет. Прежде всего, когда Гильберт формулировал свои проблемы, он обычно давал некоторую мотивацию изучения этих проблем. Например, если бы можно было решить такую-то задачу, то это позволило бы развивать такую-то теорию и решить такие-то проблемы. Здесь у него мотивация довольно странная, внутриматематическая. Он пишет, что если бы эта проблема была решена, то аналитическая теория дифференциальных уравнений приобрела бы законченный вид. Это, конечно, тоже важный момент. Но по сравнению с остальными проблемами, где мотивация была более общематематической, это кажется несколько узковатым. Тем не менее, эта проблема была сформулирована, и история её довольно любопытная, чем-то, повторяю, напоминающая ситуацию с 16-й проблемой Гильберта, потому что ещё раньше, чем в случае 16-й проблемы, в 1908 году, югославский математик Племе́ль *) предложил полное положительное решение **) этой проблемы. Сам Гильберт доказал, что проблема решается положительно в случае трёх особых точек и системы размером 2×2. А Племель доказал в общем случае. Результат Племеля опубликован в его книге, которая была издана уже в 60-е годы. В этой книге есть много интересного. Там появляется формула Сохоцкого—Племеля и многое другое. Часто используются фредгольмовы операторы; техника интегральная.

Проблема была полностью решена, тем не менее, она продолжалась исследоваться вот с какой точки зрения. Теорема Племеля была чистой теоремой существования. Возникал вопрос, а как конкретно, если у вас есть точки и матрицы, восстановить систему. Для приложений это довольно интересный вопрос, и он изучался весьма подробно. В этом направлении было опубликовано много работ. Я отмечу только две наиболее знаменитые. Первая — работа нашего соотечественника Лаппо-Данилевского, который в 1928 г. развил метод матричных рядов (рядов с матричными коэффициентами), оказавшийся полезным при решении многих задач. В частности, с помощью этого метода он ещё раз доказал (передоказал, если хотите), что если для всех i выполняется неравенство $\|G_i - I\| \leqslant \varepsilon$, ***) то проблема имеет положительное решение. Это было совсем другое доказательство —

*) Он работал в университете в Черновцах, так что это тоже наш соотечественник, в каком-то смысле.

**) Впоследствии в доказательстве обнаружен существенный пробел. — *Прим. ред.*

***) Для достаточно малого ε. — *Прим. ред.*

с техникой матричных рядов, с обращением этих рядов. Это было как бы лишним подтверждением того, что Племель был прав. Ещё я хочу отметить результат голландского математика Деккерса, который в 1972 г. доказал, что при $p = 2$ (для матриц размером 2×2) проблема всегда имеет положительное решение. Он привёл алгоритм, позволяющий полюс высокого порядка свести к полюсам первого порядка, используя один результат Делиня.

Были и другие работы, но нельзя сказать, что была активная деятельность. Скорее она была на периферии развития математики, как вдруг в 70-х годах произошло одно замечательное событие в истории математической физики, да отчасти и алгебраической геометрии. Появился метод изомонодромных деформаций. Этот метод появился в трудах японских математиков Сато, Джимбо и Мивы, а также американских математиков Флашки и Ньюэла. Выявился замечательный факт: некоторые очень важные нелинейные уравнения математической физики являются условиями изомонодромной деформации систем линейных уравнений. Я сейчас поясню, что здесь имеется в виду. Пусть есть линейное уравнение, например, фуксово, — простейший случай. У него есть монодромия. Теперь давайте будем менять расположение особых точек, и поставим такой вопрос: «Как должны меняться матрицы коэффициентов, чтобы монодромия сохранялась, оставалась одной и той же?» Это — условие *изомонодромности*. Мы меняем особые точки, получается другая система уравнений на другом пространстве, потому что комплексная структура меняется, если особых точек больше трёх. (Если вы выкалываете четыре особые точки, то комплексная структура зависит от их расположения.) Многое меняется, меняется и монодромия. Если оставить те же матрицы коэффициентов, то монодромия будет другая — это легко видеть. Как матрицы коэффициентов должны меняться, чтобы монодромия не менялась? Это некое условие на матрицы коэффициентов. Теперь считается, что они зависят от расположения особых точек: $B_i(a_1, \ldots, a_n)$. Оказывается, что соответствующие уравнения, которым матрицы коэффициентов должны удовлетворять в случае изомонодромности, имеют важное значения, потому что очень многие нелинейные уравнения математической физики (скажем, все уравнения Пенлеве, некоторые частные случаи уравнения Кортевега—де Фриза) могут интерпретироваться как условия изомонодромности, как соответствующие уравнения на $B_i(a_1, \ldots, a_n)$. Вот такой удивительный факт, о котором рассказывать подробно я не могу; это тема отдельного разговора. В частности, если вы хотите исследовать какое-то уравнение математической физики, скажем, хотите исследовать решения, то что означает изомонодромность? Она означает, что данная монодромия

является первым интегралом этого уравнения. Если вы умеете вычислять монодромию, то вы можете найти первые интегралы уравнения.

Это метод исследования таких уравнений. Чтобы исследовать уравнение, его реализуют как изомонодромное условие для некоторой системы линейных уравнений. Эту систему изучают, изучают её асимптотику, и получают интересные сведения о решениях нелилейного уравнения. Такова схема. Но чтобы её реализовать, нужно сначала по данной монодромии построить саму систему, а потом её изомонодромно продеформировать. Нужно решить обратную задачу, задачу Римана—Гильберта.

Есть и другие связи с этой проблемой, о которых я скажу позднее. Они связаны с тем, когда проблему можно решить положительно, а когда нельзя.

Другими словами, оказалось, что задача, которая первоначально была сформулирована только для линейных систем, имеет очень большое значение в математической физики. Она снова оказалась в центре исследований, и после 70-х годов появилось много работ на эту тему. Я отмечу только две книги, связанные с этой деятельностью. Первая книга вышла в 1983 г. в серии Progress in Mathematics, V. 37. В этой книге напечатаны труды семинара в Эколь Нормаль с участием Дуади, Мальгранжа, Вердье и Кона, где они в течение нескольких лет изучали проблему изомонодромных деформаций и проблему Римана—Гильберта. Эта книга до сих пор является очень хорошим обзором состояния дел в этих двух задачах, и в той и в другой. Я очень рекомендую её прочитать. Там есть, скажем, замечательная статья Мальгранжа про изомонодромные деформации, актуальная и сейчас. И вот они, подробно разобрав, в частности, проблему Римана—Гильберта, в трёх статьях объяснили, что же на самом деле было сделано к тому моменту. И оказалось, что в доказательстве Племеля имеется пробел. Выяснилось, что доказательство Племеля не вызывает никаких возражений в той части, где Племель по данному представлению монодромии строит систему линейных дифференциальных уравнений, но не фуксову, а с так называемыми регулярными особыми точками (я сейчас объясню, что это такое). В этой части никаких проблем нет. А потом Племель эту систему с регулярными особыми точками, которая имеет полюса более высокого порядка, чем первый, специальными хитрыми преобразованиями доводит до фуксовой. Вот здесь-то и содержался пробел. На самом деле, он действовал во всех точках, кроме последней, аккуратно, а потом написал: «Действуя в последней точке так же, как и раньше, получаем фуксову систему.» Здесь, вроде бы, и нет особой ошибки, но и доказательства тоже нет. Пробел. Действуя в последней точке так же, как и раньше, ничего получить нельзя.

Примерно в то же самое время Ильяшенко на своём семинаре также обнаружил, что доказательство Племеля неполное. Позже этот пробел был отмечен им в книге «Динамические системы — 1» *). Был обнаружен именно пробел в доказательстве. Тем не менее, было распространено мнение, основанное на результате Деккерса, что сама теорема верна, нужно только исправить доказательство.

Пожалуй, особым этапом в изучении этой проблемы была работа Гельмута Рорля, опубликованная в 1957 г. Рорль впервые применил теорию расслоений к изучению этой задачи. Он свёл её к некой задаче из теории расслоений. На этом пути он передоказал результат Племеля, но только для случая систем с регулярными особыми точками. Когда много лет назад мы с Чернавским, Голубевой и Лексиным стали заниматься этой проблематикой, нас насторожило, что из Рорля не получается Племель. Мы, правда, не поняли, что у Племеля есть пробел, как-то руки не дошли. Но насторожило, что получается много всего хорошего, а Племель не получается.

Ещё нужно, конечно, отметить знаменитую книгу Делиня в Lecture Notes **), где он не рассматривает проблему Римана—Гильберта, но где содержится много полезного по поводу того, как интерпретировать такого рода задачи на языке алгебраической геометрии.

Вот, пожалуй, и всё, что можно сказать об этих задачах.

Я начал заниматься этой задачей в конце 80-х годов, потому что мы стали заниматься её обобщением на многомерный случай, на случай произвольных комплексных многообразий. Пусть на комплексном многообразии есть особый дивизор. Тогда тоже есть представление фундаментальной группы. Эта задача тоже была интересна, потому что многие задачи математической физики (гипергеометрические функции, уравнения Книжника—Замолодчикова) приводят к таким задачам. Когда обнаружилось, что в основном доказательстве есть дыра, пришлось подумать на эту тему. И так получилось, что к концу 1989 года удалось обнаружить контрпример. Первый контрпример у меня возник для $n = 4$ и $p = 3$ (четыре особые точки и матрицы порядка 3). Это вообще первый контрпример, который возникает.

Я напомню, что у Лаппо-Данилевского дано доказательство для произвольных n и p, но для матриц монодромии, близких к единичной матрице. С точки зрения расслоений этот результат очевиден: малая деформация

*) Динамические системы — 1 / Под ред. Д. В. Аносова и В. И. Арнольда. — М: ВИНИТИ, 1985. (Современные проблемы математики. Фундаментальные направления; Т. 1.)

**) P. Deligne. Équations différentielles à points singuliers réguliers. — Berlin etc.: Springer Verlag, 1970. — (Lecture Notes in Mathematics; V. 163.)

тривиального расслоения остаётся тривиальным расслоением. Сегодня это нам понятно.

После этого, конечно, всё не завершилось, потому что остаётся важный вопрос, когда же всё-таки можно набор данных (особые точки + представление фундаментальной группы) реализовать. (Сама проблема имеет отрицательное решение, что тоже очень интересно и оказалось совершенно неожиданным.) Интерес к этому вопросу связан с изомонодромными преобразованиями и другими задачами. Остался ещё вопрос о том, как описать те представления, которые реализуются фуксовыми системами. Общий ответ отрицательный, но вдруг есть хорошие подклассы, которые можно реализовать фуксовыми системами. На этом пути было получено много достаточных условий.

Постепенно обнаружилось, что в вопросе о достаточных условиях подход, связанный с алгебраической геометрией, очень полезен. Дело в том, что, как всегда, когда вы имеете дело со сферой Римана, всё может быть выражено в терминах комплексного анализа и дифференциальных уравнений. Но для того чтобы понимать, что делается и как, необходимо иметь некую единую правильную точку зрения на то, что происходит. И для нас такая единая точка зрения требует применения простейших методов алгебраической геометрии и понятий расслоения, связности, стабильности и полустабильности. Эти четыре понятия помогают правильно взглянуть на проблему и даже получить новые интересные результаты.

Сейчас я перехожу к основной части своего доклада, а именно, я постараюсь рассказать об этом подходе, о том, как он работает в этой задаче, и о том, какие результаты получаются с применением этого подхода. (Он, кстати сказать, работает и во многих других задачах довольно эффективно.) Два слова о контрпримере. В первом случае, когда он возникает *), он по-прежнему остаётся довольно тонким и сложным. Он связан с комплексной структурой проколотой сферы Римана. Но в размерности 4 появляются очень простые серии контрпримеров, которые могут быть поняты легко. А все контрпримеры для $n=4$ и $p=3$ нетривиальны. Они, в частности, нестабильные. Если немножко сдвинуть особые точки, сохраняя матрицы монодромии, то задача становится разрешимой.

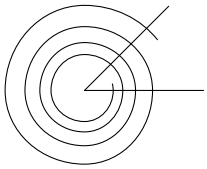
Контрпримеры имеют высокую коразмерность **), а именно, $n(p-1)$.

Особая точка a_i называется *регулярной*, если все решения данной системы имеют степенной рост в этой точке. Точнее, не более чем степенной: нет ни экспонент, ни чего-либо более сложного. Например, давайте

*) $p=3$, $n=4$. — *Прим. ред.*

**) В пространстве начальных данных. — *Прим. ред.*

рассмотрим такую простую систему: $\frac{dy}{dx} = \frac{E}{z}y$, где E — постоянная матрица. Это фуксова система; в нуле и в бесконечности у неё полюсы первого порядка. Чтобы доказать это для бесконечности, нужно сделать замену $z = 1/t$ и перейти на бесконечности к локальным координатам t. Решения этой системы моментально выписываются. Фундаментальная матрица решений равна $z^E = \exp(E \ln z)$. Посмотрим, как устроена монодромия этой системы. При обходе вокруг нуля против часовой стрелки решение заменится на $\exp(E(\ln z + 2\pi i)) = z^E G$, где $G = e^{2\pi i E}$. Любое решение — конечная логарифмическая сумма $\sum_{i=0}^{n} z^{\rho_i} \ln^{b_i} z$. При стремлении z к нулю у этого выражения не более чем степенной рост, если правильно понимать стремление к нулю. Из-за того, что здесь есть логарифм, если вы будете стремится к нулю по спирали с очень маленьким шагом (рис. 3), то после каждого оборота к логарифму будет добавляться 2π, поэтому можно получить сколь угодно быстрый рост. Поэтому здесь нужно быть аккуратным. Имеется в виду, что каждое решение имеет степенной рост, если z стремится к особой точке, оставаясь внутри некоторого сектора, т. е. без обходов вокруг особой точки. Если так понимать предел, то получается, что все решения имеют степенной рост. Логарифм растёт медленно по сравнению с z. Таков самый типичный пример уравнения с регулярной особой точкой.



Р и с. 3. Стремление к нулю по спирали

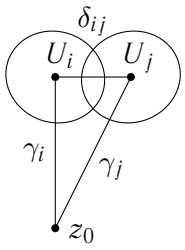
Замечательный факт состоит в том, что если система фуксова, то все её особые точки регулярны, но обратное, вообще говоря, неверно: $\mathcal{F} \subsetneq \mathcal{R}$. (Здесь $\mathcal{F}$ — фуксовы системы, $\mathcal{R}$ — системы с регулярными особыми точками.) Бывают нефуксовы системы, системы с полюсами высокого порядка, у которых тем не менее все особые точки регулярны — у решений нет существенно особых точек и скорость роста в особых точках полиномиальна.

Племель получил свой результат, абсолютно верный и правильный в несколько более широком классе — в классе систем с регулярными особыми точками. Это тоже очень интересно, но для фуксовых систем результат не был получен.

В чём же состоит метод решений, о котором я хочу рассказать? Этот метод состоит в том, чтобы по заданному представлению фундаментальной группы построить расслоение со связностью, имеющей заданную монодромию. Потом продолжим его на особые точки так, чтобы получилось расслоение с логарифмической связностью, имеющей заданную монодромию, на всей сфере Римана. Вопрос сведётся к вопросу о тривиальности

этого расслоения. Если расслоение тривиально, то задача имеет положительное решение. Если нет — то не имеет.

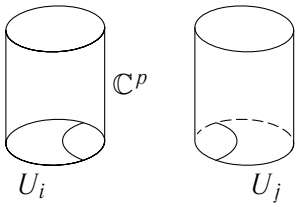
Я сейчас сказал несколько слов, которые, может быть, не все студенты знают. Поэтому я теперь всё аккуратно объясню. Эта конструкция впервые появилась в работе Рорля, потом она была у Атьи. Как по представлению фундаментальной группы построить расслоение $F \xrightarrow{\pi} \bar{\mathbb{C}} \setminus \Sigma$ над проколотой сферой Римана? Это делается очень просто. Нужно взять покрытие базы достаточно мелкими открытыми множествами U_i. Я буду брать такие открытые множества, что и сами они связны и односвязны, и пересечение любых двух из них связно и односвязно. Проколотую сферу Римана всегда можно покрыть конечным набором таких множеств. Теперь я фиксирую начальную точку z_0 и в каждой окрестности U_i выберу некоторую точку и соединю её раз и навсегда некоторым путём γ_i с этой начальной точкой. Если $U_i \cap U_j \neq \varnothing$, то я рассмотрю путь δ_{ij}, который лежит в $U_i \cup U_j$ и соединяет концы путей γ_i и γ_j (рис. 4).



Р и с. 4. Петля

Если $U_i \cap U_j \neq \varnothing$, то можно рассмотреть петлю $\gamma_i \delta_{ij} \gamma_j^{-1}$. У каждой петли есть гомотопический класс (я не буду вводить новое обозначение для гомотопического класса, а буду обозначать его так же, как и саму петлю). К нему можно применить гомоморфизм χ. Получится матрица $\chi(\gamma_i \delta_{ij} \gamma_j^{-1})$, которую я обозначу g_{ij}. Эта матрица постоянная. Она не зависит от выбора пути δ_{ij}, потому что как сами области, так и их пересечение, связны и односвязны.

Итак, что же я сделал? Каждому непустому пересечению $U_i \cap U_j$ я сопоставил матричную функцию g_{ij}, которая, к счастью, оказалась постоянной; такое бывает. Получаем отображение $U_i \cap U_j \xrightarrow{g_{ij}} \mathrm{GL}(p, \mathbb{C})$. По g_{ij} можно склеить расслоение. Очевидно, что $g_{ij} = g_{ji}^{-1}$ (всё обходится в обратную сторону, поэтому получается обратная матрица) и $g_{ij} g_{jk} g_{ki} = I$ (тут-то и нужно соображение о том, что петли можно гомотопировать, потому что если вы возьмёте три петли и рассмотрите их произведение, то получится петля, равная произведению путей δ_{ij}, а это произведение гомотопно нулю — стягивается в точку). Все свойства коцикла выполнены, поэтому я могу построить расслоение хорошо всем известным способом. Умножу каждое U_i на комплексное пространство $\mathbb{C}^p$. Получим столбики (рис. 5). Эти столбики склеиваются следующим образом: если есть два столбика, соответствующих U_i и U_j с общей



Р и с. 5. Столбики

частью $U_i \cap U_j$, то точка $(z, \vec{v})$ из столбика $U_i \times \mathbb{C}^p$ склеивается с точкой $(z, g_{ji}\vec{v})$ из столбика $U_j \times \mathbb{C}^p$. Если всё так склеивать, то условие коцикла означает, что такая склейка будет корректно определена и получится некое многообразие, которое я обозначу F. Для этого многообразия есть естественная проекция $\pi\colon F \to \bar{\mathbb{C}} \setminus \Sigma$. Тройка $(\pi, F, \bar{\mathbb{C}} \setminus \Sigma)$ называется *векторным расслоением* ранга p.

Эта конструкция — обобщение прямого произведения. Если бы я склеивал по единичному коциклу, то получил бы прямое произведение базы на $\mathbb{C}^p$. А так получилось что-то более хитрое. Как многообразие строится из шаров в евклидовом пространстве, так и расслоение *). Очень простая конструкция.

Замечательный факт состоит в том, что реализация рассматриваемого расслоения задана постоянным коциклом. Что из этого следует? Напомню, что отображение $s\colon \bar{\mathbb{C}} \setminus \Sigma \to F$ называют *сечением*, если $\pi \circ s = \mathrm{id}$. Понятие сечения является обобщением понятия векторной функции в случае расслоения. Действительно, если вы всё делаете над окрестностью U_i, то сечение — это просто отображение базы в $\mathbb{C}^p$; все такие отображения склеиваются. Если мы рассмотрим стандартный базис в $U_i \times \mathbb{C}^p$, который получается из стандартного базиса в $\mathbb{C}^p$, то любое такое отображение можно по такому базису разложить. Поэтому любое сечение s, рассмотренное над окрестностью U_i, в стандартном базисе сечений имеет некоторые координаты $({}^i y^1, \ldots, {}^i y^p)^T = {}^i\vec{y}$.

ыло: $\begin{pmatrix} {}^i y^1 \\ \vdots \\ {}^i y^p \end{pmatrix}$

Вопрос о том, можно ли продифференцировать сечение расслоения, нетривиален. Но в нашем случае ввести операцию дифференцирования сечения расслоения очень просто. Возьмём произвольное сечение, возьмём его координаты в стандартном базисе, и скажем, что дифференциал сечения s — это дифференциал вектора ${}^i\vec{y}$, т. е. $\nabla s \leftrightarrow d\,{}^i\vec{y}$. Единственный нетривиальный вопрос такой. Если я хочу получить после такого взятия производной настоящее сечение над всей базой, нужно, чтобы результаты дифференцирования (которые есть некие другие сечения с коэффициентами в дифференциальных формах) склеились при склейке цилиндриков на всём многообразии. Но у нас склейка происходит с подкруткой. Поэтому если есть какой-то график, то для того чтобы он склеился с другим графиком, нужно, чтобы на общей части соответствующие значения функций, задающих эти графики, были связаны тем же самым соотношением. Поэтому нужно, чтобы дифференциалы сечений ∇s, которые пока заданы в локальных координатах над каждой окрестностью и которые тоже представляют собой некие другие сечения, были расположены так, что на

*) Из цилиндров. — *Прим. ред.*

общей части они связаны такой же функцией. Нужно, чтобы выполнялось условие $d^i\vec{y} = g_{ij}\, d^j\vec{y}$. Если это условие будет выполнено, то тогда из этих кусочков сечений склеится общее сечение, которое будет хорошо определённой производной. Вот единственное, что нужно, если я хочу научиться дифференцировать сечения. Нужно сначала продифференцировать локально, а потом склеить результаты и проверить, удовлетворяют ли они условию корректности склейки в глобальное сечение. Но когда g_{ij} постоянны, тут всё очевидно. К счастью, у нас g_{ij} постоянны. Поэтому $g_{ij}\, d^j\vec{y} = d\, g_{ij}{}^j\vec{y}$, а по определению ${}^i\vec{y} = g_{ij}{}^j\vec{y}$. Такое простое введение дифференцирования сечений стало возможным только потому, что матрицы g_{ij} постоянные, т. е. коцикл постоянный. Если бы коцикл был не постоянный, то появился бы дополнительный член. Тогда было бы непонятно, как с ним бороться.

Итак, это расслоение замечательно тем, что я могу дифференцировать сечения. Дифференциал сечения я буду обозначать ∇s. Каждое сечение может быть продифференцировано по такому правилу: я должен взять координаты сечения в окрестности U_i и взять обычный дифференциал векторной функции; это и будут координаты ∇s по определению. В силу того, что коцикл постоянный, это совершенно корректное определение.

Рассмотрим уравнение $\nabla s = 0$. Сечение s, которое удовлетворяет такому уравнению, называют *горизонтальным*. Для нашего расслоения условие, что сечение горизонтально, означает просто, что $d^i\vec{y} = 0$. Таким образом, задав связность ∇, т. е. задав правило дифференцирования сечений, мы в каждой координатной окрестности получаем простенькое дифференциальное уравнение $d^i\vec{y} = 0$. В каждой координатной окрестности уравнения разные, потому что координаты разные; но вид уравнение имеет всегда один и тот же. Задание связности эквивалентно заданию в каждой координатной окрестности таких простейших дифференциальных уравнений.

Важнейшим понятием является *монодромия* связности. Рассмотрим произвольный замкнутый путь γ и покроем его координатными окрестностями U_i (это покрытие конечное, потому что координатных окрестностей конечное число). Пусть начальная (и одновременно конечная) точка пути покрыта окрестностью U_1. В этой окрестности уравнение имеет вид $d^1\vec{y} = 0$. Я возьму его решение. В качестве решения можно выбрать единичную матрицу I. Затем я посмотрю, какой вид будет иметь решение в координатной окрестности U_2, пересекающей U_1. При переходе к этой окрестности координаты сечений умножаются на матрицу g_{12}, поэтому в координатной окрестности U_2 это решение будет иметь вид Ig_{12}. В следующей окрестности это решение будет иметь вид $Ig_{12}g_{23}$ и т. д. Когда

я после полного обхода вернусь в исходную точку, я получу матрицу $g_{12}g_{23}\dots g_{m1}$. Итак, после полного обхода, матрица решений умножается на матрицу $g_{12}g_{23}\dots g_{m1}=G^{-1}$. Это по определению и есть матрица монодромии связности. Мне не важно, что есть глобальное уравнение. Мне достаточно, что в каждой окрестности есть по уравнению. Монодромия этой системы всё равно определяется.

В нашем случае матрица монодромии имеет такой замечательно простой вид: $G_\gamma=(g_{12}g_{23}\dots g_{m1})^{-1}$.

У п р а ж н е н и е. $G_\gamma=\chi(\gamma)$, где γ — петля, а χ — исходное представление монодромии фундаментальной группы.

В этом смысле оказывается, что так введённая связность (которая, повторяю, может рассматриваться просто как совокупность большого числа согласованных систем уравнений над каждой окрестностью) имеет заданную монодромию — ту, которую мы хотели получить. Это — простейшая проверка. Нужно просто нарисовать картинку и убедиться, что если подставить вместо g_{ij} произведение $\gamma_i\delta_{ij}\gamma_j^{-1}$, то все γ_i сокращаются и остаётся произведение путей δ_{ij}, которое гомотопно петле γ.

По исходному представлению мы построили расслоение $F\to\bar{\mathbb{C}}\setminus\Sigma$, на котором задали связность ∇, имеющую заданную монодромию. Осталось сделать ещё один шаг. Подумаем вот над чем. Мы научились дифференцировать сечения, используя данную специальную тривиализацию расслоения, т. е. данный коцикл g_{ij}. Но ведь можно рассмотреть другую тривиализацию того же расслоения, ей эквивалентную, т. е. можно выбрать другой склеивающий коцикл, уже не постоянный, но эквивалентный исходному. Я сейчас не буду объяснять, что такое эквивалентные коциклы. Все, кто знает расслоения, конечно, знает, что это такое. Как там будет выглядеть связность?

Если вместо стандартного базиса $e_1,\dots,e_p$ вы возьмёте другой базис (другую тривиализацию) $\xi_1,\dots,\xi_p$, то эти базисы связаны посредством матрицы $\Gamma^{-1}(z)$, т. е. $(\xi_1,\dots,\xi_p)=(e_1,\dots,e_p)\Gamma^{-1}(z)$. Матрица $\Gamma^{-1}(z)$ аналитически зависит от z, потому что сечения аналитические. Тогда сечение s, которое в первом базисе имело координаты ${}^i\vec{y}$, будет иметь координаты ${}^i\vec{f}=\Gamma^i\vec{y}$ во втором базисе. Поэтому если вы посмотрите, как же теперь выглядит производная s по отношению к новым координатам, вы получите более сложную формулу. Нужно просто честно продифференцировать это выражение. После несложных вычислений вы получите, что $d{}^i\vec{f}=-(d\Gamma\cdot\Gamma^{-1})\,{}^i\vec{f}$. Вот, оказывается, как будет выглядеть применение связности для дифференцирования сечения в локальных координатах, если вы выбрали другую тривиализацию расслоения. Нужно продифференцировать ${}^i\vec{f}$, а потом вычесть матрицу дифференциальных

форм $\omega = d\Gamma \cdot \Gamma^{-1}$, умноженную на ${}^i\vec{f}$. Только в тривиализации, заданной постоянным коциклом, связность выглядит так просто. В другой тривиализации она приобретает более сложный вид. Но заметьте, что поправка линейна по ${}^i\vec{f}$. В итоге мы получаем, что для того чтобы найти горизонтальное сечение, нужно решить уравнение $d{}^i\vec{f} = \omega^i\vec{f}$. Это — другая форма записи нашей системы линейных дифференциальных уравнений. Над каждой окрестностью возникает система линейных дифференциальных уравнений.

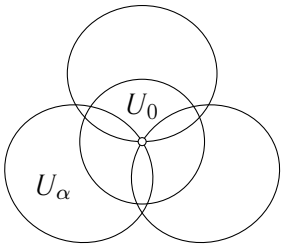
К чему я всё это говорю? Представьте себе на минуту, что построенное нами расслоение $F \to \bar{\mathbb{C}} \setminus \Sigma$ оказалось бы тривиальным, т. е. если бы был глобальный базис голоморфных сечений. В этом базисе уравнение, задающее горизонтальное сечение, имело бы вид $d\vec{f} = \omega\vec{f}$ без всяких i, потому что в качестве локальной тривиализации можно взять всю базу сразу. Если бы был глобальный базис сечений, то тогда уравнение $\nabla s = 0$ не разбивалось бы в совокупность уравнений для каждого покрытия, над которым расслоение тривиально, а давало бы глобальное дифференциальное уравнение на всей проколотой сфере Римана, и это уравнение имело бы заданную монодромию по той простой причине, что такую монодромию имеет связность (это уравнение — просто вид связности в заданном базисе). Вот в чём прелесть этого метода. Всё, что я сделал, — я построил расслоение со связностью, имеющей заданную монодромию. В данном случае задание связности эквивалентно заданию большого числа дифференциальных уравнений над каждой окрестностью, над которой расслоение тривиально. И связность имеет заданную монодромию. Если расслоение глобально тривиально, то в базисе глобальных сечений мы получаем только одну систему уравнений, имеющую заданную монодромию.

Связность может быть любой, потому что мы получаем систему линейных дифференциальных уравнений $d\vec{f} = \omega\vec{f}$ на проколотой сфере Римана с заданной монодромией, но мы не имеем никакой информации о том, как ведёт себя матрица коэффициентов ω в особых точках. Фуксова она или не фуксова — ничего мы пока сказать не можем. Но если расслоение тривиально, то это — решение такой интересной задачи. В базисе глобальных сечений наша связность задаёт систему дифференциальных уравнений, которая имеет заданную монодромию и заданные особые точки. Исходная задача почти решена: по заданному представлению построена система с заданной монодромией и с заданными особыми точками. Но я совершенно не контролирую поведение системы в особых точках.

Замечательный факт состоит в том, что расслоение всегда тривиально. Этот факт можно доказывать разными способами. Я скажу некоторые слова, но можно и без этих учёных слов всё доказать. Дело в

том, что любое комплексное расслоение над проколотой сферой Римана топологически тривиально, потому что проколотая сфера Римана — это топологически букет окружностей, а любое комплексное расслоение над окружностью тривиально. Поэтому расслоение топологически тривиально всегда, а аналитически оно тривиально потому, что проколотая сфера Римана — многообразие Штейна *). Я сказал умные слова, но на самом деле тут можно всё доказать вручную. Ничего кроме теоремы Рунге из комплексного анализа для доказательства этого факта не нужно. Такое доказательство приведено в книге Форстера «Римановы поверхности».

Для того чтобы не просто решить такую симпатичную задачу и получить систему, в которой никак не контролируются особенности, нужно продолжить расслоение на всю сферу Римана, продолжить в особые точки. Если мы это сделаем, то тогда сможем решить исходную задачу. Что значит «продолжить расслоение в особые точки»? Рассмотрим некоторую особую точку. Пусть это будет точка 0. Эта точка особая, поэтому в покрытие она не входит. Я нарисую покрытие тремя окрестностями (рис. 6). Нужно взять некую окрестность U_0, содержащую точку 0, и продолжить расслоение на U_0 — подклеить соответствующий столбик $U_0 \times \mathbb{C}$ к нашему расслоению. Получится продолжение расслоения.



Р и с. 6. Покрытие окрестности особой точки

Как это сделать? Рассмотрим некую окрестность U_α из покрытия. Над этой окрестностью есть специальный базис горизонтальных сечений $(e_1, \ldots, e_p)$, тот самый, по которому строилось расслоение. Что получится, если я продолжу этот базис вдоль петли, обходящей особую точку? По определению он умножится на матрицу монодромии, отвечающую особой точке: $(e_1, \ldots, e_p) \mapsto \mapsto (e_1, \ldots, e_p)G_0$. Рассмотрим другой базис $(\xi_1^0, \ldots, \xi_p^0) = (e_1, \ldots, e_p)z^{-E_0}$, где $E_0 = \frac{1}{2\pi i} \ln G_0$. Логарифм матрицы определяется неоднозначно; чтобы он был определён однозначно, нужно зафиксировать собственные значения. Я буду считать, что действительные части собственных значений лежат между 0 и 1, т. е. $0 \leqslant \operatorname{Re} \rho_0^j < 1$. Эти условия однозначно определяют матрицу.

*) Многообразие Штейна — это, грубо говоря, комплексное многообразие, на котором достаточно много голоморфных функций. Например, для любых двух различных точек есть голоморфная функция, которая их разделяет. Если вы из сферы Римана выкинули хотя бы одну точку, то любые две точки разделяются функцией z.

Замечательный факт состоит в том, что базис $(\xi_1^0, \ldots, \xi_p^0)$ при обходе особой точки переходит сам в себя. Действительно, z^{-E_0} умножается на G_0^{-1}. Поэтому при обходе вокруг особой точки базис $(\xi_1^0, \ldots, \xi_p^0)$ умножается на $G_0 G_0^{-1} = I$. Я подправил базис на матричную функцию z^{-E_0} и снова получил базис (над каждой точкой $z \neq 0$ это постоянная невырожденная матрица, т. е. снова получаем базис сечений, даже аналитических). Но так уже получается базис сечений над всей окрестностью $U_0 \setminus \{0\}$. Причина простая: во-первых, будучи продолжен на любую окрестность, он является там базисом, а во-вторых, будучи продолжен в исходную точку он переходит сам в себя. Я построил тривиализацию расслоения над окрестностью $U_0 \setminus \{0\}$, т. е. я построил базис сечений в цилиндре $U_0 \times \mathbb{C}^p$, из которого выброшена ось $\{0\} \times \mathbb{C}^p$. Как этот базис сечений ведёт себя около нуля, я не знаю. Он там может вести себя плохо. Но вне точки 0, вне оси цилиндра, — это глобальные сечения. Действительно, переходя из одной окрестности в другую, мы совершим обход вокруг точки 0, и при этом при возвращении в исходную точку базис останется тем же самым.

Дальше ясно, как продолжать расслоение. Выберем в цилиндре $U_0 \times \mathbb{C}^p$ стандартный базис сечений, который получается из прямого произведения, и отождествим эти сечения с $(\xi_1^0, \ldots, \xi_p^0)$ на общей части двух этих цилиндров, т. е. при $z \neq 0$. Я приклеил цилиндр $U_0 \times \mathbb{C}^p$ и тем самым продолжил расслоение на особую точку. Вклеить так цилиндрик — это то же самое, что принудительно объявить сечения голоморфными в точке 0.

Как будет выглядеть матрица связности, т. е. матрица дифференцирования сечений в этом базисе? Первоначально дифференцирование было простое: $d\vec{f} = 0$. Я уже писал такую формулу, поэтому сейчас напишу без объяснений: $d\vec{f} = \frac{E_0}{z} dz\, \vec{f}$. Так выглядит действие связности для произвольного сечения в этом базисе. Правая часть — это в точности $(-d\Gamma \cdot \Gamma^{-1}) f$ — можете проверить. Заметьте, что получилось фуксово уравнение $d\vec{f} = \frac{E_0}{z} dz\, \vec{f}$. Итак, если я продолжу расслоение таким способом в особую точку, то моя связность будет иметь, как говорят, *логарифмическую* особенность. Или, говоря по-другому, соответствующая система будет иметь полюс первого порядка. Замечательное продолжение — оно контролирует поведение связности в особой точке. Я напомню, что форма связности — это $-d\Gamma \cdot \Gamma^{-1}$, а $\Gamma = z^{-E_0}$.

Продолжим так расслоение в каждой особой точке. Получим расслоение, которое называется *каноническим* продолжением исходного расслоения над всей сферой Римана. Если это расслоение вдруг случайно окажется голоморфно тривиальным, то в глобальном базисе сечений

моя связность задаст уравнение $d\vec{f} = \omega \vec{f}$, которое локально отличается от уравнения $d\vec{f} = \frac{E_0}{z} dz\, \vec{f}$ только сопряжением на голоморфную обратимую матрицу, потому что два голоморфных базиса отличаются на голоморфную обратимую матрицу. Значит, это уравнение имеет только простые полюса в особых точках. И я получу решение задачи Римана—Гильберта.

Ещё раз повторю, что я сделал. Я продолжил расслоение $F \to \bar{\mathbb{C}} \setminus \Sigma$ в особые точки специальным образом, выбрав базис сечений над проколотой окрестностью особой точки и вклеив соответствующий цилиндрик. Это продолжение замечательно тем, что наша связность будет теперь иметь особенности в точках a_i, но эти особенности будут простыми полюсами — связность будет логарифмическая. Получается расслоение $F^0 \to \bar{\mathbb{C}}$ над всей сферой Римана с логарифмической связностью ∇^0, имеющей заданную монодромию. Если это расслоение голоморфно тривиально, то тогда уравнения на горизонтальные сечения связности принимают вид системы линейных дифференциальных уравнений на всей сфере Римана с заданной монодромией. И что замечательно, мы можем предсказать, какие особенности имеет матрица коэффициентов. Она в каждой особой точке a_i отличается от матрицы $\frac{E_0}{z - a_i}$ на голоморфную обратимую матрицу, т. е. тоже имеет полюс первого порядка. Всё свелось к вопросу о том, тривиально расслоение $F^0 \to \bar{\mathbb{C}}$ или нет. Это расслоение специальное. Как я уже говорил, оно называется каноническим продолжением исходного расслоения. Оно описано в книге Делиня *).

На самом деле есть и другие продолжения, которые приводят к логарифмическим связностям. Мы нуждаемся во всех этих продолжениях. Нам хотелось бы описать все продолжения, у которых связности имеют только логарифмические особенности. Оказывается, что описать все такие продолжения очень несложно. Нужно вместо базиса $(\xi_1^0, \ldots, \xi_p^0)$ выбрать базис $(\hat{\xi}_1^0, \ldots, \hat{\xi}_p^0) = (e_1, \ldots, e_p) z^{-E_0} z^{\Lambda}$, где я буду считать, что матрица E_0 верхняя треугольная (всегда можно выбрать базис так, чтобы матрица была верхней треугольной), а

$$\Lambda = \begin{pmatrix} \lambda_0^1 & 0 & \ldots & 0 \\ 0 & \lambda_0^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_0^p \end{pmatrix}$$

— произвольная диагональная матрица, у которой диагональные элементы образуют невозрастающую последовательность целых чисел: $\lambda_0^1 \geqslant \ldots \ldots \geqslant \lambda_0^p \in \mathbb{Z}$. Множитель z^{Λ} ветвления не имеет; он однозначный. Поэтому

*) См. сноску на с. 208. — *Прим. ред.*

у нас снова есть базис сечений, и я могу продолжить расслоение с помощью этого базиса. Тогда соответствующее уравнение запишется так:

$$d\vec{f} = \frac{1}{z}(\Lambda + z^{\Lambda}E_0 z^{-\Lambda})\,dz\,\vec{f}.$$

Вот так изменится моё уравнение. Замечательный факт состоит в том, что эта матрица голоморфна: произведение матриц $z^{\Lambda}E_0z^{-\Lambda}$ с условием $\lambda_0^1 \geqslant \ldots \geqslant \lambda_0^p \in \mathbb{Z}$ и с условием верхней треугольности матрицы E_0 является голоморфной матрицей. Снова моё уравнение имеет фуксовы особенности. Так продолженное расслоение снова имеет связность с логарифмическими особенностями. Это ещё одно такое расслоение. Замечательная теорема (которая получается, например, из книги Гантмахера) состоит в следующем. Как мы выяснили, расслоения $\bar{\mathcal{F}} = \{F^{\Lambda}, \nabla^{\Lambda}\}$, $\Lambda = \{\Lambda_1, \ldots, \Lambda_n\}$ имеют логарифмические особенности.

Теорема 1. *Так описываются все расслоения с логарифмическими связностями, имеющие заданную монодромию.*

Таких расслоений бесконечно много; это бесконечное семейство параметризуется матрицами Λ. Если хотя бы одно из этих расслоений тривиально, то проблема Римана—Гильберта имеет положительное решение. Это мы будем сокращённо обозначать PRH$\oplus$. Итак, PRH$\oplus$ тогда и только тогда, когда существует матрица Λ, для которой расслоение F^{Λ} голоморфно тривиально.

Первая часть программы выполнена. Ответ на вопрос Гильберта сводится к исследованию следующего вопроса. Приведена совершенно чёткая конструкция бесконечного семейства расслоений с логарифмическими связностями, имеющими заданную монодромию. Вопрос в том, есть в этом семействе расслоений хотя бы одно тривиальное или нет. Если есть хотя бы одно, то ответ на вопрос Гильберта положителен, если нет, то отрицателен.

Каждое конкретное комплексное расслоение над сферой Римана имеет очень простое устройство. Сейчас я сформулирую теорему, которая принадлежит Гротендику, а потом объясню, что она означает на простом координатном языке. Каждое комплексное расслоение над сферой Римана есть прямая сумма одномерных расслоений:

$$F^{\Lambda} \cong \mathcal{O}(k_1^{\Lambda}) \oplus \ldots \oplus \mathcal{O}(k_p^{\Lambda}),$$

где расслоение $\mathcal{O}(k)$ имеет такое координатное описание:

$$\mathcal{O}(k) = \{\mathbb{C},\ \bar{\mathbb{C}} \setminus \{0\},\ g_{0\infty} = z^k\}.$$

Здесь имеется в виду, что сфера Римана покрыта двумя окрестностями $\mathbb{C}$ и $\bar{\mathbb{C}} \setminus \{0\}$, а в качестве функций склейки берутся функции $g_{0\infty} = z^k$. Стало

быть,

$$F^\Lambda = \{\mathbb{C},\ \bar{\mathbb{C}} \setminus \{0\},\ g_{0\infty} = z^{\begin{pmatrix} k_1^\Lambda & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & k_p^\Lambda \end{pmatrix}}\}.$$

Числа $k_1^\Lambda, \dots, k_p^\Lambda$ называют *типом расщепления* расслоения. Они полностью определяют его как голоморфное рассбоение на $\bar{\mathbb{C}}$.

Получается такая задача, к которой мы свели исходную задачу Римана—Гильберта. У нас есть особые точки $a_1, \dots, a_n$, есть матрицы монодромии $G_1, \dots, G_n$ и есть матрицы $\Lambda_1, \dots, \Lambda_n$, по которым мы строим наше продолжение. На выходе нужно получить числа $k_1^\Lambda \geqslant \dots \geqslant k_p^\Lambda$. Если бы можно было это сделать, то мы смогли бы дать полный ответ на вопрос Гильберта. Дело в том, что расслоение голоморфно тривиально в том и только том случае, если все числа k_i^Λ равны нулю. К сожалению, этого сделать нельзя. Это очень тяжёлая задача, которая известна в теории интегральных уравнений как *задача нахождения частных индексов*. Она общего решения не имеет. Но выясняется, что иногда об этих числах можно кое-что сказать, и на основании этой информации дать ответ на вопрос, имеет проблема положительное решение или нет.

Я не буду говорить о контрпримере, скажу о положительном результате — о том, как можно решить задачу в некотором специальном случае, когда представление имеет специальный вид. Монодромия называется *неприводимой*, если у набора матриц $G_1, \dots, G_n$ нет общих инвариантных подпространств, кроме нулевого подпространства и всего пространства $\mathbb{C}^p$. Следующая теорема была доказана в 1992 г. независимо мной и Володей Костовым.

Т е о р е м а 2. *Если представление χ неприводимо, то* PRH$\oplus$.

Таким образом, любое неприводимое представление монодромии может быть реализовано фуксовой системой. Тем самым, контрпримеры нужно искать среди приводимых представлений. Так они, собственно, и были найдены.

Сейчас я коротко объясню доказательство. Оказывается, имеет место такая замечательная лемма.

Л е м м а. *Если χ неприводимо, то для любого набора Λ тип расщепления построенного расслоения обладает таким свойством*: $k_i^\Lambda - k_{i+1}^\Lambda \leqslant n-2$.

Мы берём бесконечное множество различных матриц Λ, строим по ним бесконечное множество расслоений. У них, вообще говоря, бесконечное множество типов расщепления. Казалось бы, какие угодно. Но нет. Оказывается, что если исходное представление было неприводимо, то нельзя получить какие угодно типы расщепления. Разности между соседними ин-

дексами не должны превосходить число особых точек минус два. Если вас интересуют тривиальные расслоения, то вы должны потребовать, чтобы $k_1^\Lambda + \ldots + k_p^\Lambda = 0$. Но если сумма чисел равна нулю и разности между соседними числами не превосходят $n-2$, то таких наборов конечное число. Что же получается? С одной стороны, по неприводимому представлению вы строите бесконечное число расслоений (топологически тривиальных — нулевой степени). С другой стороны, в этом бесконечном наборе в силу этой леммы лишь конечное число гомеоморфно неэквивалентных типов. Весьма правдоподобно звучит, что там должны быть тривиальные расслоения. Это, конечно, не доказательство, но это — объяснение, почему так должно быть. И действительно, доказательство не такое уж трудное. Основано оно на том, что построено бесконечное число расслоений, среди которых лишь конечное число различных, неэквивалентных. Поэтому там обязательно встречаются тривиальные расслоения.

Неравенство $k_i^\Lambda - k_{i+1}^\Lambda \leqslant n-2$ связано с понятием стабильности и полустабильности расслоений со связностями, но я не имею ни малейшей возможности за оставшееся время обо всём этом подробно рассказать.

Вся эта конструкция может быть перенесена и на компактные римановы поверхности. (Заметьте, что она существенно локальна.) Там можно ставить вопросы разумным образом, и там есть свои трудности. Похожую конструкцию можно применять и в многомерном случае, что и сделали Hain и N. Katz.

Я не так уж много успел рассказать. Но я надеюсь, что сумел объяснить, почему сам подход, связанный с такими простыми и естественными понятиями, как расслоение и связность, уже помогает правильно поставить задачу и свести её к более простой задаче, которую в некоторых случаях уже удаётся решить.

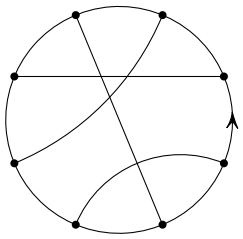
12 октября 2000 г.

С. К. Ландо

ОБ ОДНОМ КЛАССЕ ИНВАРИАНТОВ ГРАФОВ, СВЯЗАННОМ С ИНВАРИАНТАМИ ВАСИЛЬЕВА УЗЛОВ

Я сегодня буду рассказывать про один класс инвариантов графов, связанных с инвариантами Васильева узлов. Несмотря на то что за последние 6—7 лет про инварианты Васильева узлов рассказывалось неоднократно, эту тематику нельзя считать общеизвестной. Поэтому я посвящу первую половину своего доклада рассказу о конструкции Васильева в случае узлов в трёхмерном пространстве. А вторая часть доклада в основном будет ориентирована на класс инвариантов графов, которые возникают в связи с инвариантами Васильева.

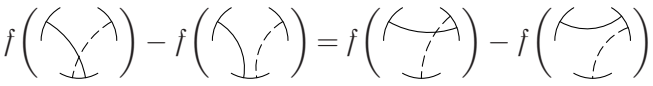
Меня в первую очередь будут интересовать комбинаторные структуры, связанные с инвариантами Васильева. Эти комбинаторные структуры называются *весовыми системами*. Весовая система — это некоторая функция на хордовых диаграммах, удовлетворяющая специальному соотношению. *Хордовая диаграмма* — это очень простой комбинаторный объект. С комбинаторной точки зрения хордовая диаграмма представляет собой окружность (эту окружность я всегда буду предполагать ориентированной против часовой стрелки); на этой ориентированной окружности выбрано несколько пар точек, и эти пары точек соединены хордами (рис. 1). Хорды могут пересекаться между собой, могут и не пересекаться. Такие картинки мы рассматриваем с точностью до изотопии окружности (т. е. с точностью до гомеоморфизма окружности, сохраняющего ориентацию). То есть это чисто комбинаторный объект; никакой дополнительной структуры на нём нет.



Р и с. 1. Хордовая диаграмма

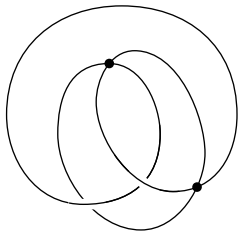
Функция на хордовых диаграммах называется *весовой системой*, если она удовлетворяет четырёхчленному соотношению, которое выглядит следующим образом. Возьмём 4 хордовые диаграммы, у которых одинаково расположены все хорды за исключением двух выделенных хорд (рис. 2). На самом деле, пунктирная хорда тоже неподвижна; она во всех четырёх диаграммах тоже одинакова. Изменяется только сплошная хорда: её

конец — соседний с концом пунктирной хорды. Между двумя концами этих двух хорд на нарисованной дуге окружности не может быть никаких других концов хорд. Конец сплошной хорды может располагаться четырьмя способами. Мы получаем 4 диаграммы. Функция на хордовых диаграммах называется весовой системой, если она удовлетворяет линейному соотношению, изображённому на рис. 2.

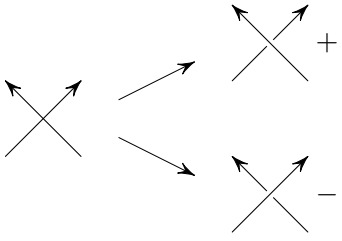


Р и с. 2. Четырёхчленное соотношение

Откуда взялись хордовые диаграммы и откуда взялись четырёхчленные соотношения? Васильев рассматривает инварианты узлов (т. е. функции на узлах), где под узлом понимается изотопический класс отображения $S^1 \to \mathbb{R}^3$. На самом деле теория Васильева гораздо более широкая; она относится к различным функциональным пространствам. Это лишь один из примеров, хотя и наиболее яркий. В этом примере вычисления в значительной мере удаётся провести до конца. Идея Васильева состоит в том, что любой инвариант, определённый на узлах, можно продолжить на особые узлы. Если под узлом мы понимаем изотопический класс отображения $S^1 \to \mathbb{R}^3$, которое является взаимно однозначным отображением на свой образ, гладко и дифференциал отображения нигде не обращается в нуль, то особым узлам разрешается иметь двойные точки (рис. 3). Никаких более сложных особенностей особому узлу иметь не полагается.



Р и с. 3. Особый узел



Р и с. 4. Разрешение двойной точки

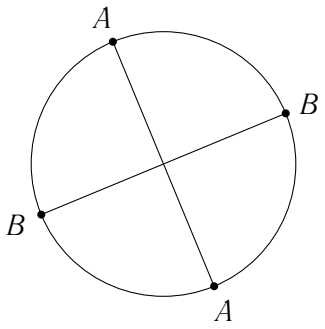
Окружность изначально ориентирована, поэтому узел тоже можно считать ориентированным. Продолжение инварианта выглядит следующим образом. Если инвариант уже определён на узлах с n двойными точками, то на узлы с $n+1$ двойной точкой инвариант продолжается следующим образом. Каждую двойную точку можно разрешить двумя способами (рис. 4). Ориентация узла позволяет одно разрешение считать положительным, а другое отрицательным. Значение инварианта на узле с дополнительной двойной точкой мы определяем так: $v(*) = v(+) - v(-)$, т. е. из значения инварианта на узле, разрешённом в

положительную сторону, вычитается значение инварианта на узле, разрешённом в отрицательную сторону. Возможность разрешать таким образом особые узлы означает, что дискриминант в пространстве узлов *допускает кориентацию*.

При таком продолжении инварианта на первый взгляд возникает неоднозначность. Мы можем разрешать двойные точки в произвольном порядке. Простое утверждение гласит, что результат не зависит от того, в каком порядке мы разрешаем особые точки. Это верно для любых инвариантов.

В пространстве всех инвариантов можно выделить подпространство $\mathcal{V}_d$, состоящее их тех инвариантов, продолжения которых обращаются в нуль на особых узлах с более чем d двойными точками. Это подпространство называется *пространством инвариантов порядка не выше* d. Мы имеем вложения $\mathcal{V}_0 \subset \mathcal{V}_1 \subset \mathcal{V}_2 \subset \dots$ Объединение $\mathcal{V} = \bigcup_i \mathcal{V}_i$ называется *пространством инвариантов узлов конечного порядка*. Обычно рассматриваются инварианты со значениями в коммутативном кольце.

С каждым особым узлом можно связать хордовую диаграмму следующим образом. Особый узел — это отображение окружности в трёхмерное пространство, рассматриваемое с точностью до изотопии. При этом отображении у каждой двойной точки есть ровно два прообраза. Каждая пара прообразов, соответствующих двойной точке, соединяется хордой. Например, особому узлу, изображённому на рис. 3, соответствует хордовая диаграмма, изображённая на рис. 5.



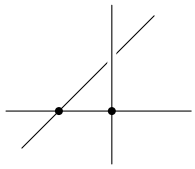
Р и с. 5. Хордовая диаграмма, соответствующая особому узлу с рис. 3

Оказывается, что всякий инвариант $v \in \mathcal{V}_d$ определяет функцию на хордовых диаграммах порядка d. (*Порядок* хордовой диаграммы — это число хорд.) Утверждение состоит в том, что если есть особый узел с d двойными точками, то значение инварианта порядка не выше d на этом узле полностью определяется хордовой диаграммой этого узла. Это утверждение почти очевидно. Что такое инвариант порядка не выше d? Это инвариант, который обращается в нуль на любом узле с $d+1$ двойной точкой. Это означает, что мы можем произвольным образом проводить друг сквозь друга ветви особого узла с d двойными точками, не меняя значение инварианта. Скачок инварианта оказывается равным нулю, потому что значение инварианта на любом узле с $d+1$ двойной точкой равно нулю. Несложно доказать, что если у двух особых

узлов одинаковые хордовые диаграммы, то их можно преобразовать один в другой, проводя ветви узла друг сквозь друга. Тем самым, инвариант порядка не выше d определяет функцию на хордовых диаграммах степени d.

Естественный вопрос: «Какие функции на хордовых диаграммах могут так получиться?» Ответ состоит в том, что эти функции должны удовлетворять четырёхчленному соотношению. Это не единственное соотношение, которому они должны удовлетворять. Я скажу об этом несколько слов. Правильный объект, который нужно рассматривать, это не просто узлы, а оснащённые узлы, т. е. узлы, в каждой точке которых задана кориентирующая прямая (кориентирущее направление). Это то же самое, что узел, которому приписано целое число. Для оснащённых узлов помимо четырёхчленного соотношения никаких других соотношений нет. А для обычных узлов есть ещё так называемое *одночленное соотношение*. Но оно играет несущественную роль, поэтому я о нём рассказывать не буду.

Четырёхчленное соотношение возникает в ситуации, когда на узле есть две близкие двойные точки (рис. 6), которые хотят пройти друг сквозь друга. У них для этого есть разные возможности. Если аккуратно проанализировать поведение инварианта для таких двух соседних точек, то окажется, что должно выполняться четырёхчленное соотношение. Это соотношение было найдено Васильевым. Естественно встал вопрос, нет ли других соотношений. Утверждение о том, что других соотношений нет в случае, когда значения инварианта лежат в поле нулевой характеристики, составляет содержание известной теоремы Концевича. Интеграл Концевича доказывает это утверждение для инвариантов со значениями в поле нулевой характеристики. Для инвариантов со значениями в кольцах или в полях положительной характеристики ответ неизвестен.

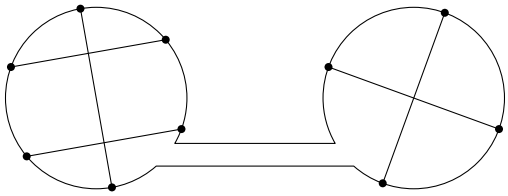


Р и с. 6. Две близкие двойные точки

Если есть функция, которая удовлетворяет четырёхчленному соотношению, т. е. если есть весовая система на хордовых диаграммах степени d, то по ней восстанавливается инвариант, из которого она получена. Есть процедура, которая позволяет восстановить какой-нибудь инвариант. Таких инвариантов достаточно много. Они образуют аффинное пространство, потому что любые два из них отличаются на инвариант порядка $d-1$.

Нас будут интересовать функции на хордовых диаграммах, удовлетворяющие четырёхчленному соотношению. Мне будет удобнее забыть про функции и рассматривать четырёхчленное соотношение на хордовых диаграммах. Я буду рассматривать его как отношение эквивалентности в векторном пространстве, порождённом хордовыми диаграммами. Мне

понадобится пространство $\mathcal{A} = \mathcal{A}_0 \oplus \mathcal{A}_1 \oplus \mathcal{A}_2 \oplus \ldots$, где $\mathcal{A}_i$ — пространство, порождённое хордовыми диаграммами степени i, профакторизованное по подпространству, порождённому всеми четырёхчленными соотношениями. На этом пространстве имеется структура алгебры Хопфа. Главными элементами структуры алгебры Хопфа являются две операции — умножение и коумножение. Умножение $m\colon \mathcal{A} \otimes \mathcal{A} \to \mathcal{A}$ берётся из операции связной суммы узлов. Это произведение узлов переносится на хордовые диаграммы (рис. 7); здесь подразумевается, что произведение уважает ориентацию



Р и с. 7. Произведение хордовых диаграмм

окружности. Утверждение состоит в том, что если мы рассматриваем хордовые диаграммы по модулю четырёхчленного соотношения, то это произведение корректно определено. Вообще говоря, мы можем разрывать хордовые диаграммы в разных точках и получать разные произведения, но по модулю четырёхчленного соотношения они оказываются одинаковыми. Всегда найдётся цепочка четырёхчленных соотношений, связывающая два разных произведения двух данных хордовых диаграмм. Это произведение коммутативно.

В двойственном пространстве к пространству хордовых диаграмм есть естественное умножение — это просто умножение инвариантов (умножение весовых функций). Здесь важно, чтобы область значений инварианта была коммутативным кольцом. Если в двойственном пространстве есть произведение, то в исходном пространстве есть копроизведение $\mu\colon \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}$. В данном случае копроизведение устроено следующим образом. Разбиваем хорды данной диаграммы всеми возможными способами на два непересекающихся подмножества (одно из них может быть пустым). Из первого множества образуем первую диаграмму, из второго вторую. Между этими диаграммами ставим знак тензорного умножения. Так мы получаем элемент тензорного квадрата. Копроизведение — это сумма всех таких элементов. Если диаграмма содержит n хорд, то сумма состоит из 2^n слагаемых. Очевидным образом это коумножение кокоммутативно: его результат инвариантен относительно перестановки множителей в тензорном квадрате.

Чтобы определить структуру биалгебры, нужно ещё сказать, какая здесь единица и какая коединица. Единица очевидная — это пустая хордовая диаграмма. Коединица тоже очевидная. Чтобы превратить биалгебру в алгебру Хопфа, нужно ещё добавить антипод. Антипод там тоже есть. Но я не буду вдаваться в эти алгебраические тонкости.

Коммутативные и кокоммутативные алгебры Хопфа устроены очень просто. На этот счёт есть теорема Милнора—Мура. Она гласит следующее. В кольце многочленов $\mathbb{C}[x_1, \ldots, x_k]$ есть естественная структура алгебры и структура коалгебры, которая устроена следующим образом. Если есть какой-нибудь моном $y_1 \ldots y_l$, то результат коумножения этого монома строится так:

$$\mu(y_1 \ldots y_l) = 1 \otimes y_1 \ldots y_l + y_1 \otimes y_2 \ldots y_l + \ldots$$

Образующие монома всеми возможными способами разбиваются на два подмножества и между ними ставится знак тензорного умножения. Это очень простая алгебра Хопфа. Она называется *полиномиальной* алгеброй Хопфа. Теорема Милнора—Мура гласит, что любая коммутативная и кокоммутативная алгебра Хопфа является полиномиальной. В частности, полиномиальной является алгебра Хопфа хордовых диаграмм. Тем самым, в ней имеются какие-то образующие (*примитивные элементы*), которые очень непохожи на хордовые диаграммы. Обратите внимание, что для хордовой диаграммы $\bigotimes$

$$\mu\Big(\bigotimes\Big) = \bigcirc \otimes \bigotimes + 2 \ominus \otimes \ominus + \bigotimes \otimes \bigcirc,$$

а в полиномиальной алгебре Хопфа для примитивного элемента $\mu(x_i) = 1 \otimes x_i + x_i \otimes 1$. Тем самым, хордовые диаграммы не являются примитивными элементами. В полиномиальной алгебре Хопфа элемент примитивен тогда и только тогда, когда является линейной комбинацией образующих.

Примитивные элементы полностью определяют структуру алгебры Хопфа. В каждом пространстве $\mathcal{A}_i$ есть подпространство $\mathcal{P}_i \subset \mathcal{A}_i$, состоящее из примитивных элементов. (В пространстве $\mathcal{A}_0$ есть только один примитивный элемент 0.) Чтобы описать алгебру Хопфа хордовых диаграмм, достаточно описать пространства примитивных элементов, которые априори должны быть устроены проще, чем сами пространства $\mathcal{A}_i$. Нас интересуют размерности пространств $\mathcal{P}_i$. Вот что известно о размерностях этих пространств (в первой строке записана степень хордовых диаграмм, во второй — размерность пространства примитивных элементов):

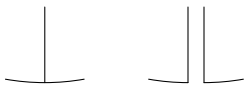
1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	2	3	5	8	12	18	27	39	55	?

Все полагают, что размерность пространства примитивных элементов растёт, но это не доказано. Скорость роста тоже неизвестна.

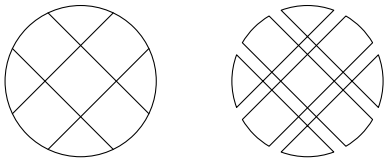
Теперь я приведу несколько примеров весовых систем, т. е. функций, удовлетворяющих четырёхчленному соотношению.

1) Число хорд.

2) Число пересечений хорд. Действительно, число пересечений хорд для всех четырёх диаграмм в четырёхчленном соотношении легко находится; эти числа для двух диаграмм одинаковые и для двух других диаграмм тоже одинаковые, но отличаются на 1 от первых двух чисел.



Р и с. 8. Удвоение хорды

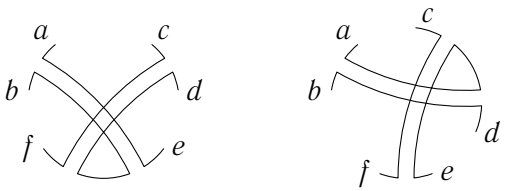


Р и с. 9. Пример удвоения хорд в диаграмме. В результате удвоения получилось 3 окружности

3) Хордовую диаграмму можно рассматривать как граф с вершинами степени 3. В окрестности каждой вершины можно произвести удвоение хорды (рис. 8). Пример операции удвоения хорды для конкретной хордовой диаграммы приведён на рис. 9. Пересечения хорд при этом вершинами не считаются. В результате такой операции хордовая диаграмма распадается на несколько окружностей.

Число этих окружностей является весовой системой.

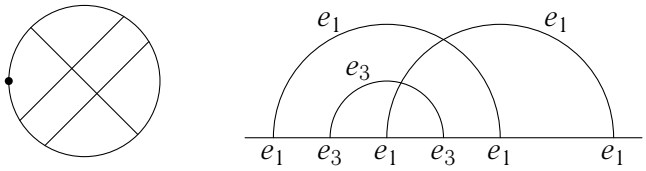
Эта весовая система удовлетворяет даже более сильному соотношению (двучленному). Действительно, для этой весовой системы легко проверить, что $(1)=(3)$ и $(2)=(4)$. Например, равенство $(1)=(3)$ проверяется на рис. 10; соединяются концы, помеченные одинаковыми буквами.



Р и с. 10. Проверка равенства $(1)=(3)$

Помимо этих весовых систем есть огромный класс весовых систем, которые происходят из алгебр Ли. Для построения такой весовой системы нужна алгебра Ли $\mathfrak{g}$ с невырожденным биинвариантным скалярным произведением. Для простоты я буду рассматривать алгебры Ли над $\mathbb{C}$, хотя

аналогичная конструкция имеется и для алгебр Ли над $\mathbb{R}$. По каждой такой алгебре Ли весовая система строится следующим образом. Возьмём в этой алгебре ортонормированный базис $e_1, \ldots, e_g$. Хордовую диаграмму нужно разорвать в какой-то точке и развернуть окружность в прямую (рис. 11).



Р и с. 11. Построение весовой системы по алгебре Ли

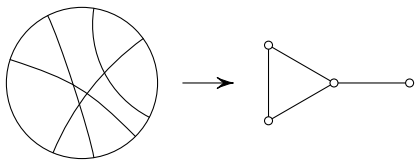
Затем на каждой дуге диаграммы нужно написать одну из букв $e_1, \ldots, e_g$; это показано на том же рисунке. Эти буквы мы перенесём на концы дуг и получим слово — элемент универсальной обёртывающей алгебры $U(\mathfrak{g})$. Просуммируем все такие слова по всем расстановкам букв. Для данного конкретного случая получается сумма $\sum_{i,j,k=1}^{g} e_i e_j e_k e_j e_i e_k$. Теорема, которая в таком виде принадлежит Концевичу, гласит, что эта сумма, во-первых, лежит в центре универсальной обёртывающей алгебры, во-вторых, этот элемент универсальной обёртывающей алгебры не зависит от того, в каком месте разрывается хордовая диаграмма, и не зависит от выбора базиса, в-третьих, и это главное, полученная функция со значениями в центре универсальной обёртывающей алгебры (коммутативном кольце) удовлетворяет четырёхчленному соотношению. Оно вытекает из тождества Якоби в алгебре Ли. Тем самым мы получили весовую систему со значениями в центре универсальной обёртывающей алгебры. Эту весовую систему можно несколько упростить, если взять представление алгебры Ли и взять след этого представления. Представление алгебры Ли заодно задаёт представление универсальной обёртывающей алгебры, и композиция со следом даёт весовую систему со значениями в поле комплексных чисел.

В частности, инвариант с числом дуг (окружностей) графа, полученного удвоением хорд, строится по стандартному представлению алгебры Ли $\mathfrak{sl}(N)$.

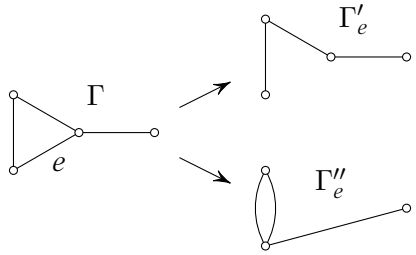
По конструкции весовых систем из алгебры Ли видно, что это — объект, с которым в таком виде работать невозможно. Для хордовой диаграммы порядка n и алгебры Ли размерности g набор слагаемых состоит из g^n элементов труднообозримой универсальной обёртывающей алгебры. Кроме того, известно, что не все весовые системы получаются таким образом.

В общем, структура этой алгебры Хопфа остаётся весьма неясной. Как мы пытаемся что-то прояснять для себя, работая с непонятным объектом? Мы пытаемся откусить какой-то кусок. Если с этим куском удаётся разобраться, то можно двигаться дальше. То, что я сейчас буду рассказывать, это попытка откусить такой кусок.

Исходным мотивом для этой попытки послужило следующее замечание Дужина. Каждой хордовой диаграмме D можно сопоставить её граф пересечений $\Gamma(D)$. Вершины этого графа находятся во взаимно однозначном соответствии с хордами диаграммы; две вершины соединены ребром в том и только том случае, если две хорды пересекаются (рис. 12). Дужин заметил, что хроматический многочлен $\chi_{\Gamma(D)}(t)$ графа пересечений $\Gamma(D)$ является весовой системой. Я напомню, что *хроматический многочлен* $\chi_\Gamma(t)$ графа Γ — это многочлен, который при любом натуральном t принимает значение, равное количеству правильных раскрасок вершин графа Γ в t цветов (раскраска называется *правильной*, если любые две соседние вершины раскрашены в разные цвета).



Р и с. 12. Граф пересечений



Р и с. 13. Две операции над графом

Хроматические многочлены замечательны тем, что они, как нетрудно видеть, удовлетворяют следующему условию. Если есть граф Γ и в этом графе есть ребро e, то удалив ребро e, мы получим граф, который будем обозначать Γ'_e, а стянув ребро e в точку, получим граф Γ''_e (рис. 13). Тогда хроматические многочлены удовлетворяют соотношению $\chi_\Gamma(t) = \chi_{\Gamma'_e}(t) - \chi_{\Gamma''_e}(t)$. Действительно, если из правильных раскрасок графа Γ'_e выбросить те раскраски, для которых концы ребра e окрашены в один цвет, то получим правильные раскраски графа Γ.

Из этого соотношения, в частности, следует, что $\chi_\Gamma(t)$ — многочлен. Действительно, это соотношение последовательно сводит вычисление хроматического многочлена к графам с меньшим числом рёбер и, тем самым, — к вычислению хроматического многочлена графа вообще без рёбер. А хроматический многочлен графа с n вершинами и без рёбер равен t^n.

Доказательство утверждения Дужина такое. Пусть четырёхчленное соотношение имеет вид $(1)-(2)=(3)-(4)$ и диаграмме (1) соответствует граф пересечений Γ. Тогда диаграмме (2) соответствует граф пересечений Γ'_e (одно ребро стирается, поскольку две пересекающиеся хорды становятся непересекающимися). Для пары диаграмм (3) и (4) в правой части графы пересечений связаны таким же образом. Утверждение, которое легко проверить, состоит в том, что графы Γ''_e для одной пары и для другой пары совпадают.

После того как было обнаружено это свойство, мы занялись изучением графов пересечений хордовых диаграмм, чтобы понять, какой кусок алгебры Хопфа хордовых диаграмм можно так откусить.

Какие вообще соотношения между хордовой диаграммой и её графом пересечений? Конечно, граф пересечений сопоставляется хордовой диаграмме однозначно. Но если задан граф, то он, во-первых, может не соответствовать никакой хордовой диаграмме, а во-вторых, он может соответствовать сразу большому количеству хордовых диаграмм. Так что соотношение не взаимно однозначное.

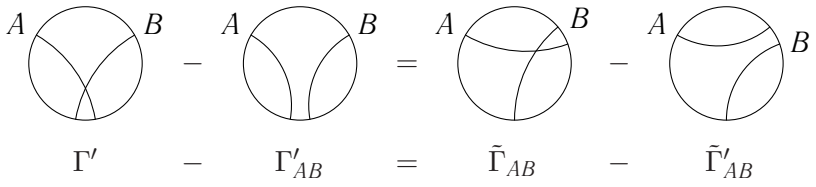
Для хордовой диаграммы есть два понятия близости хорд. Во-первых, близкими можно считать хорды, концы которых соседние. Во-вторых, близкими можно считать пересекающиеся хорды. Граф пересечений запоминает только одну из этих структур. Мы теряем информацию о близости хорд вдоль окружности. Это и даёт надежду получить что-то более обозримое, чем просто хордовые диаграммы.

Соотношение $f(\Gamma)=f(\Gamma'_e)+f(\Gamma''_e)$ называют *соотношением Татта*. Если хроматический многочлен заменить на многочлен $f(\Gamma)=(-1)^{\#V(\Gamma)}\chi_\Gamma$, где $\#V(\Gamma)$ — число вершин графа Γ, то такой многочлен будет удовлетворять соотношению Татта. Нужно только правильно понимать стягивание ребра в соотношении Татта. При таком стягивании могут появляться кратные рёбра. А если мы стягиваем одно из простых рёбер, то остальные превращаются в петли.

По поводу функций, удовлетворяющих такому соотношению, есть замечательная работа Татта 1947 г. Там они полностью расклассифицированы и дано их полное описание. Это описание состоит в следующем. Соотношение Татта можно рассматривать как рекуррентную формулу для подсчёта значений нашей функции. Мы взяли граф, взяли в нём ребро, стянули это ребро, получили два графа. В каждом из них рёбер меньше, чем в исходном. Если мы знаем значения функции на них, то мы знаем значение функции на исходном графе. Тем самым вычисление инварианта сводится к его вычислению для графа, у которого нет рёбер, отличных от петель. В зависимости от того, в каком порядке мы стягиваем рёбра,

могли бы получиться разные ответы. Теорема Татта гласит, что ответ не зависит от того, в каком порядке стягиваются рёбра.

Татт в этой работе вводит понятие кольца графов. Кольцо графов — это линейное пространство, натянутое на графы. Произведение задаётся несвязным объединением графов. Его долго ругали за эту работу все комбинаторщики, потому что непонятно, зачем нужна кольцевая структура на графах. С другой стороны, эта работа считается классической. Естественное желание, которое возникло, — проверить, не дадут ли другие инварианты Татта весовые системы. Оказалось, что нет. Единственный инвариант Татта, который приводит к весовым системам, — это хроматический многочлен. Но при этом выяснилось следующее. Давайте посмотрим на четыре хордовые диаграммы, которые участвуют в четырёхчленном соотношении. Им соответствуют четыре графа: Γ, Γ'_{AB}, $\tilde{\Gamma}_{AB}$ и $\tilde{\Gamma}'_{AB}$ (рис. 14).



Р и с. 14. Четырёхчленное соотношение для графов пересечений

Оказывается, что графы $\tilde{\Gamma}_{AB}$ и $\tilde{\Gamma}'_{AB}$ однозначно восстанавливается по графу Γ и хордам A, B. Эти хорды являются вершинами графа Γ; граф Γ'_{AB} получается из графа Γ стиранием ребра AB. Каким образом восстанавливаются графы $\tilde{\Gamma}_{AB}$ и $\tilde{\Gamma}'_{AB}$? Во первых, в хордовой диаграмме, соответствующей графу $\tilde{\Gamma}_{AB}$, все хорды, кроме хорды A, имеют между собой те же пересечения, что и в хордовой диаграмме, соответствующей графу Γ. А пересечения с хордой A меняются только для тех хорд, которые пересекают хорду B; пересечение меняется в точности на противоположное. Если хорда C пересекала хорды A и B, то в новом графе она не пересекает A и пересекает B. Если хорда C не пересекала A и пересекала B, то в новом графе она пересекает A и B.

Вот такое преобразование определяет переход от графа Γ к графу $\tilde{\Gamma}$. Мы получаем то, что называется *четырёхчленным соотношением для графов*:

$$\Gamma - \Gamma'_{AB} = \tilde{\Gamma}_{AB} - \tilde{\Gamma}'_{AB}.$$

Забудем теперь, что наши графы были графами пересечений хордовых диаграмм, и будем рассматривать четырёхчленное соотношение для произвольных графов.

В отличие от соотношения Татта, в четырёхчленном соотношении играет роль порядок вершин. В соотношении Татта роль играет только ребро. В четырёхчленном соотношении третий член зависит не только от ребра, но и от порядка вершин этого ребра.

Мы можем рассматривать пространство, порождённое графами (пока не учитывая четырёхчленное соотношение). Во-первых, на этом пространстве есть таттовское умножение, заданное несвязным объединением графом. Во-вторых на нём есть коумножение. Коумножение графов устроено следующим образом. Мы берём граф и разбиваем его вершины произвольным образом на два подмножества. Эти два подмножества определяют два графа. Между этими графами мы ставим знак тензорного умножения. Копроизведение графа — это сумма таких слагаемых по всем разбиениям вершин графа на два подмножества. Получающаяся при этом алгебра Хопфа не очень интересна, поскольку её примитивные элементы находятся во взаимно однозначном соответствии со связными графами. Каждому связному графу можно сопоставить примитивный элемент, и эти примитивные элементы будут порождать всю алгебру Хопфа.

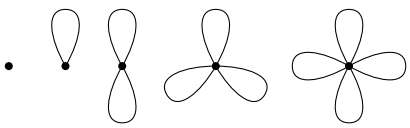
В отличие от ситуации с хордовыми диаграммами, для введения структуры алгебры Хопфа нам не нужно четырёхчленное соотношение. Структура алгебры Хопфа присутствует здесь изначально, но она слабая. Интересно то, что структура алгебры Хопфа выдерживает факторизацию по четырёхчленному соотношению. То, что сохраняется умножение, очевидно. То, что сохраняется коумножение, требует небольшого доказательства.

Мы получили коммутативную и кокоммутативную алгебру Хопфа $\mathcal{F} = \mathcal{F}_0 \oplus \mathcal{F}_1 \oplus \mathcal{F}_2 \oplus \dots$ По теореме Милнора—Мура она изоморфна полиномиальной алгебре Хопфа. Соответственно, она порождена примитивными образующими $\mathcal{P}_i \subset \mathcal{F}_i$. Известны следующие размерности пространств примитивных элементов:

1	2	3	4	5	6	7	8
1	1	1	2	3	5	7	?

У нас имеется отображение $\mathcal{A} \to \mathcal{F}$ из алгебры Хопфа хордовых диаграмм в алгебру Хопфа графов (профакторизованных по четырёхчленному соотношению). Это отображение является гомоморфизмом алгебр Хопфа. Есть гипотеза о том, что этот гомоморфизм является эпиморфизмом. Проведённый подсчёт показывает, что это отображение — не мономорфизм; в размерности 7 есть 1-мерное ядро. Что дальше, неизвестно. Вообще, про алгебру $\mathcal{F}$ почти ничего не известно.

Итак, нас интересуют инварианты графов, удовлетворяющие четырёхчленному соотношению. Коумножение вводит структуру алгебры Хопфа и в кольце Татта, но там алгебра получится уже не градуированная, а только фильтрованная, потому что соотношение Татта не сохраняет число вершин. Теорема Татта гласит, что в этой алгебре Хопфа в каждой размерности есть ровно одна образующая (рис. 15). А алгебра Хопфа $\mathcal{F}$ гораздо богаче. По крайней мере, в начале число образующих в каждой размерности растёт быстро.



Р и с. 15. Образующие кольца Татта

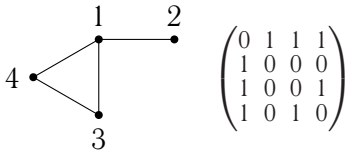
Если гипотеза о том, что отображение алгебр Хопфа — эпиморфизм, окажется верной, то надежда на то, что построенный объект проще, чем исходная алгебра Хопфа, сохраняется. Более того, приобретает самое веское основание. Гипотетически разница между этими двумя алгебрами небольшая. Первое различие наступает только в размерности 7, и различие незначительное. Видимо, это различие будет расти.

Ядро этого гомоморфизма — тоже какой-то кусок алгебры Хопфа хордовых диаграмм, который, быть может, самый интересный. Во всяком случае, есть старая проблема, возникшая сразу после того, как появилась работа Васильева. Она состоит в следующем. Возьмём хордовую диаграмму и рассмотрим её зеркальный образ (окружность ориентирована, поэтому мы, вообще говоря, получаем другую диаграмму). Есть ли весовая система, которая может различить эти две диаграммы (если они различны)? Если такой весовой системы нет, то из этого вытекает, что инварианты Васильева неполны, т. е. они могут различить не все узлы. (Это тоже одна из проблем; в этой области, пожалуй, самая главная.) Естественно, когда мы переходим от хордовых диаграмм к графам пересечений, мы забываем про ориентации. Поэтому различающий элемент, если он существует, может лежать только в ядре.

Поскольку есть гомоморфизм $\mathcal{A} \to \mathcal{F}$, есть и гомоморфизм двойственных алгебр $\mathcal{A}^* \leftarrow \mathcal{F}^*$, действующий в обратную сторону. Тем самым, любой инвариант графов, удовлетворяющий четырёхчленному соотношению, определяет весовую систему. Поэтому если у нас есть большой запас инвариантов графов, удовлетворяющих четырёхчленному соотношению, то мы автоматически получаем большой запас весовых систем, который, как можно ожидать, будет более обозримым, чем те весовые системы, которые можно было изначально получить по хордовым диаграммам.

Каждому графу можно сопоставить его матрицу инцидентности (матрицу примыканий). А именно, занумеруем вершины и положим $a_{ij} = 1$,

если вершины i и j соединены ребром, и $a_{ij}=0$, если вершины i и j не соединены ребром. (Пример такого сопоставления изображён на рис. 16.) Всегда получается симметрическая матрица, которую можно рассматривать как матрицу над $\mathbb{Z}/2\mathbb{Z}$. Рассматривая эту матрицу как матрицу над $\mathbb{Z}/2\mathbb{Z}$, возьмём её коранг. Утверждение состоит в том, что коранг является инвариантом, удовлетворяющим четырёхчленному соотношению. И этот коранг в точности совпадает с весовой системой, задаваемой числом окружностей в диаграмме с удвоенными хордами. Это теорема Жени Соболевой. В частности, коранг удовлетворяет двучленному соотношению. Доказательство этого утверждения основано на том, что описанная выше процедура перехода от графа Γ к графу $\tilde{\Gamma}$ на уровне матриц инцидентности состоит в умножении над $\mathbb{Z}/2\mathbb{Z}$ слева и справа на матрицу



$$\begin{pmatrix} 0&1&1&1\\ 1&0&0&0\\ 1&0&0&1\\ 1&0&1&0 \end{pmatrix}$$

Р и с. 16. Граф и его матрица примыканий

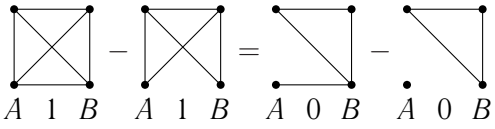
$$\begin{pmatrix} 1&0&0&0\\ 1&1&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{pmatrix}$$

и на транспонированную к ней. (Это для графов с четырьмя вершинами. В общем случае матрица устроена так же: на диагонали стоят единицы, $a_{21}=1$, на остальных местах стоят нули.) Таким образом, на матрицу инцидентности нужно смотреть как на квадратичную форму, а на преобразование — как на замену базиса. Замена базиса не меняет ранг и коранг квадратичной формы. По-видимому, у этого преобразования есть какой-то глубокий смысл.

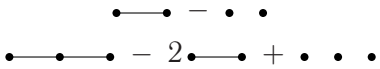
Отсюда возникает естественный вопрос. По-видимому, среди инвариантов графов, удовлетворяющих четырёхчленному соотношению, есть большой класс инвариантов, которые происходят из алгебр Ли. Что это за инварианты? Как их описывать? Дело в том, что изначальное описание через хордовые диаграммы не работает, потому что гомоморфизм действует в противоположную сторону. И, вообще говоря, я не могу надеяться, что инвариант хордовых диаграмм, построенный по алгебре Ли, удастся вернуть обратно в инварианты графов. В данном конкретном случае получается; коранг связан со стандартным представлением алгебры Ли $\mathfrak{gl}_n$.

Все те примеры, которые пока удалось построить, тоже связаны с алгебрами Ли. Для инварианта, который строится по $\mathfrak{sl}_2$, Чмутов и Варченко построили рекуррентное соотношение, которое позволяет его значения на хордовых диаграммах выразить через значения на хордовых диаграммах с меньшим числом хорд. Такое значение — многочлен от одной переменной

со старшим коэффициентом 1, потому что центр универсальной обёртывающей алгебры для $\mathfrak{sl}_2$ отождествляется с алгеброй многочленов от одной переменной. Можно попытаться интерпретировать коэффициенты этого многочлена. Старший коэффициент единица. Второй коэффициент — число пересечений хорд в хордовой диаграмме; он же, соответственно, — число рёбер в графе пересечений. Третий коэффициент выражается через число треугольников и число квадратов в графе. Из этого удаётся вывести, что число вершинных квадратов в графе удовлетворяет четырёхчленному соотношению. *Вершинный квадрат* — это четвёрка вершин, на которые можно натянуть квадрат (неважно, сколькими способами). Например, если взять полный граф с четырьмя вершинами и записать для него четырёхчленное соотношение (рис. 17), то у графов в левой части есть по одному вершинному квадрату, а у графов в правой части вершинных квадратов нет.



Р и с. 17. Четырёхчленное соотношение и вершинные квадраты



Р и с. 18. Примитивные элементы в размерностях 2 и 3

Как интерпретировать следующий коэффициент многочлена, неизвестно. Вообще, неизвестно, поднимается ли $\mathfrak{sl}_2$-инвариант на графы.

Вот ещё парочка несложных примеров. На рис. 17 видно, что число рёберных квадратов не удовлетворяет четырёхчленному соотношению. (*Рёберный квадрат* — это четвёрка рёбер, образующих квадрат.) Графы в левой части равенства содержат 3 и 1 рёберных квадрата, а в правой части рёберных квадратов нет. Однако число рёберных квадратов становится инвариантом, если его рассматривать по модулю 2. То же самое справедливо и для рёберных n-угольников при произвольном $n > 3$. Число совершенных паросочетаний в графе тоже удовлетворяет четырёхчленному соотношению. (*Совершенное паросочетание*, или 1-*фактор*, это такой набор из n рёбер в графе с $2n$ вершинами, что любая вершина принадлежит одному ребру.) Совершенные паросочетания каким-то пока таинственным способом связаны с весовой системой, построенной по супералгебре Ли $\mathfrak{gl}(1|1)$.

Теперь я расскажу, откуда в алгебре Хопфа графов берутся примитивные элементы. Сами по себе графы, как правило, не являются примитивными элементами в этой алгебре (точечный граф является, а остальные нет). Например, примитивны элементы, изображённые на рис. 18.

Пространство $\mathcal{F}_n$ можно представить в виде прямой суммы пространства $\mathcal{P}_n$ и его прямого дополнения, которое натянуто на несвязные графы (то, что получается умножением элементов меньшего порядка): $\mathcal{F}_n = \mathcal{P}_n \oplus \mathcal{P}_n^{\perp}$. Тем самым, каждому графу можно сопоставить его проекцию на пространство примитивных элементов. Проекция идёт вдоль того дополнения, которое я описал. Для этого проектора имеется явная формула

$$\pi_n(\Gamma) = \Gamma - 1! \sum \Gamma(V_1)\Gamma(V_2) + 2! \sum \Gamma(V_1)\Gamma(V_2)\Gamma(V_3) - \dots$$

Суммирование здесь производится следующим образом. Мы берём все вершины графа и разбиваем их всеми возможными способами на два непустых подмножества. На каждое из подмножеств V_i вершин натягиваем граф $\Gamma(V_i)$ (полный подграф нашего графа), и суммируем по таким графам. Затем суммируем по тройкам непустых подмножеств и т. д. Эта формула указывает на то, что этот проектор является на самом деле логарифмом. Вместо коэффициентов $1!, 2!, \dots$ должны стоять $\frac{1}{2}, \frac{1}{3}, \dots$ Так и будет, если мы будем рассматривать упорядоченные наборы вершин вместо неупорядоченных наборов вершин. Он действительно является логарифмом в следующем вполне точном смысле. Если у нас есть мультипликативный (т. е. сохраняющий мультипликативную структуру) инвариант W со значениями в кольце R, то у такого мультипликативного инварианта можно взять логарифм $w = \log W : \mathcal{F} \to R$. Этот логарифм обладает следующим свойством: $w = W \circ \pi$. В этом смысле проекция является логарифмом.

Дополнение

В заключение позволю себе пофантазировать на следующую тему. Согласно общей философии Арнольда (которую он излагал на Студенческих чтениях в НМУ) у любого вещественного объекта есть естественная комплексификация (их может быть несколько). Такая комплексификация должна быть и у инвариантов Васильева. Что это такое, сказать трудно, однако можно предположить, как выглядит комплексификация хордовой диаграммы. Это — оснащённое зацепление в трёхмерной сфере. Сфера S^3 служит комплексификацией окружности S^1, а роль хорд (= нульмерных сфер) играют одномерные компоненты зацеплений. Индекс пересечения хорд, который в вещественной ситуации может принимать значения из $\mathbb{Z}/2\mathbb{Z}$, переходит в целочисленный индекс зацепления. Для комплексифицированных хордовых диаграмм можно даже написать четырёхчленное соотношение (в нём участвует так называемое второе движение Кирби), однако его смысл пока совершенно неясен.

26 октября 2000 г.

А. Я. Хелемский

ПЛОСКИЕ МОДУЛИ И ГАРМОНИЧЕСКИЙ АНАЛИЗ НА ГРУППАХ

Мне предложили рассказать на семинаре «Глобус» о науке, которой я занимаюсь. Эта наука выспренно называется *гомологическая теория алгебр анализа*. На Западе её большей частью называют *топологическая гомология*, подразумевая, что гомологии сугубо топологические. Как такую вещь рассказывать? Наука разрослась. За полтора часа не расскажешь и её малой части. Поэтому я предлагаю пойти по такому пути. Нужно взять какой-то свежий результат (я возьму самый свежий). Сформулировать, потом объяснить ингредиенты. И ещё объяснить идею доказательства в самых общих чертах. По ходу дела, если результат синтетический, многое нужно будет рассказать, и всё это вместе, возможно, какое-то впечатление даст.

Результат, о котором я расскажу, будет опубликован в совместной работе трёх авторов: Далеса, Гхахрамани и моей *). Он таков.

Теорема 1. *Если G — непрерывная* (*т. е. не дискретная*) *локально компактная группа, то её алгебра мер $M(G)$ не аменабельна.*

Здесь есть два ключевых слова, которые потребуют объяснения. Во-первых, что такое «алгебра мер», а во-вторых, что такое «аменабельность» банаховой алгебры, каковой эта алгебра мер является. Но сразу же я продолжу. В объединении с давно известным, классическим в данной проблематике, результатом Барри Джонсона, о котором ещё будет много слов сказано ниже, отсюда получается автоматическое следствие, которое позволяет уже для любых локально компактных групп дать полное описание того, когда алгебра мер аменабельна.

Следствие. *Пусть G — произвольная локально компактная группа. Тогда её алгебра мер $M(G)$ аменабельна тогда и только*

Работа выполнена при частичной поддержке Научно-технической программы сотрудничества между Россией и Грецией, обеспеченной Генеральным секретариатом научно-технических исследований в Греции (грант 70/3/4970, Афинский университет).

*) H. G. Dales, F. Ghahramani, A. Ya. Helemskii. The amenability of measure algebras // J. London Math. Soc. II Ser. — 2002. — V. 66, № 1. — P. 213—226.

тогда, когда G дискретна и аменабельна в классическом теоретико-групповом смысле, идущем от фон Нойманна (*т. е. аменабельна как дискретная группа*).

У аменабельности в теоретико-групповом смысле есть много разных эквивалентных определений. Проще всего можно сказать так. Дискретная группа G *аменабельна*, если на ней существует конечная аддитивная мера (нетривиальная и определённая на всех подмножествах), которая инвариантна относительно сдвигов. Здесь достаточно инвариантности относительно левых сдвигов, потому что тогда будет инвариантность и относительно правых сдвигов.

Когда-то тот факт, что такой меры не существует на группе вращений (дискретной) позволило фон Нойману объяснить парадокс Банаха—Тарского. Но если говорить ещё и об этом, то я ничего не успею.

Теперь я начинаю объяснять ингредиенты.

Алгебра мер. Пусть G — локально компактная группа. В алгебре, насколько мне известно, есть только одна конструкция групповой алгебры группы. Это линейные комбинации элементов группы с коэффициентами из какого-то поля. Групповые алгебры одной и той же группы отличаются только тем, какие берутся поля. Для банаховых алгебр поле только одно — поле комплексных чисел $\mathbb{C}$. Но существует довольно много разных конструкций, которые исходя из локально компактной группы делают банахову алгебру. Они обслуживают разные вещи, и трудно сказать, какая из них лучше. Но две из них выделяются по своей важности. Это самая старая по времени алгебра $L^1(G)$, которую когда-то ввёл Ирвинг Сигал. Она состоит из функций на G, интегрируемых по мере Хаара. (Значит, её определение зависит от этого глубокого понятия — меры Хаара, поэтому оно не столь просто.) Групповая операция — обычная свёртка.

Но есть ещё и другая банахова алгебра — алгебра мер $M(G)$, которая как раз и есть предмет нашего разговора. Определяется она гораздо проще (не требует меры Хаара), но устроена она гораздо сложнее и изучена гораздо хуже. Её элементы — конечные комплексные регулярные борелевские меры на G. Что такое «регулярные», я подробно объяснять не буду; это означает, что они хорошо согласованы с топологией. Измеримые множества аппроксимируются изнутри компактными, а сверху открытыми. Важно, что мера конечная. Это — линейное пространство. Оно является банаховым, если нормой меры объявить вариацию, т. е. $\|\mu\| := \operatorname{Var}\mu$. И главное, в этом банаховом пространстве можно ввести операцию, относительно которой оно является банаховой алгеброй. Эта операция тоже называется свёрткой. Пусть есть две борелевские меры μ и ν. Что такое свёртка мер как мера? Мне нужно объяснить, что такое

$\mu * \nu$ от множества E. Свёртка определяется так:

$$\mu * \nu(E) = \iint\limits_{GG} \chi_E(st)d\mu(s)d\nu(t) = \int\limits_G \nu(s^{-1}E)d\mu(s).$$

Здесь χ_E — характеристическая функция множества E, т. е. $\chi_E(st) = 1$, если $st \in E$, и $\chi_E(st) = 0$, если $st \notin E$.

В этой алгебре есть несколько замечательных двусторонних идеалов. Один из них состоит из абсолютно непрерывных по мере Хаара функций. (Теперь включается мера Хаара.) Он отождествляется как раз с $L^1(G)$. Сейчас речь идёт о случае, когда группа непрерывна. Когда группа дискретна, всё упрощается; об этом я скажу позже.

Но для нас сейчас более интересен другой замкнутый двусторонний идеал $M_c(G)$, который состоит из так называемых непрерывных мер (т. е. мер, которые в каждой отдельной точке равны нулю):

$$L^1(G) \subset M_c(G) \subset M(G).$$

Идеал $M_c(G)$ как линейное пространство имеет банахово дополнение, более того, это банахово дополнение является банаховой алгеброй, т. е. банахова алгебра $M(G)$ представима (как линейное пространство) в виде прямой суммы банаховых алгебр: $M(G) = M_c(G) \oplus l_1(G)$. Подалгебра $l_1(G)$ состоит из так называемых дискретных мер. Дискретные меры — это пределы линейных комбинаций мер Дирака. Кстати сказать, алгебра $M(G)$, в отличие от алгебры $L^1(G)$, обладает единицей. Единица алгебры $M(G)$ — это мера Дирака, сосредоточенная в единице группы. А если взять разные меры Дирака и замкнуть их линейные комбинации, то получим банахово пространство $l_1(G)$.

Тот факт, что идеал $M_c(G)$ дополняем, — это наше чеховское ружьё, которое выстрелит в последнем акте.

Если группа G дискретна, то $M_c(G)$ пропадает, и остаётся $M(G) = l_1(G)$.

Для понятия аменабельности есть много эквивалентных определений, но основных подходов два. Я сейчас изложу более традиционный и старый по времени подход, до сих пор наиболее распространённый на Западе. Этот подход основан на так называемых дифференцированиях. Второй подход основан на плоских модулях. Как раз он и участвовал в доказательстве этой теоремы. О нём речь пойдёт позже.

Вообще, мой доклад делится на три части. Первая часть называется «На Западе». Она как раз сейчас начинается. Вторая часть будет называться «На Востоке». И третья часть будет «Синтез».

1.
На Западе

Пусть A — банахова алгебра, X — двусторонний банахов модуль над A (бимодуль, как говорят). Это означает, что X — бимодуль в чисто алгебраическом смысле, но ещё он — банахово пространство, и обе операции внешнего умножения непрерывны. *Дифференцированием* A со значениями в X называется непрерывный линейный оператор $D\colon A \to X$, который удовлетворяет тождеству Лейбница $D(ab) = a \cdot D(b) + D(a) \cdot b$. Если зафиксировать элемент $x \in X$ и рассмотреть оператор $D_x\colon a \mapsto a \cdot x - x \cdot a$, то, как легко проверить, этот оператор будет дифференцированием. Такое дифференцирование называется *внутренним*. Типичная задача в теории банаховых алгебр состоит в следующем. Заданы банахова алгебра A и бимодуль X. Верно ли, что любое дифференцирование из A в X является внутренним? По-настоящему интересными считаются как раз не внутренние дифференцирования, называемые *внешними*.

Одна из мотиваций того, почему этот вопрос интересен, состоит в следующем. Исторически всё начиналось со случая $X = A$, т. е. с дифференцирований алгебры. В анализе такие дифференцирования давали инфинитезимальный вариант автоморфизмов, потому что если D — дифференцирование, то e^D — автоморфизм алгебры. А внутреннее дифференцирование по такому правилу даёт внутренний автоморфизм, который действует сопряжением. Причём это верно более или менее в обе стороны, потому что если заданный внутренний автоморфизм близок к единичному, то он обязательно имеет вид e^D, где D — внутреннее дифференцирование. Есть такая теорема. А поэтому, зная все дифференцирования, мы можем узнать, например, все однопараметрические группы автоморфизмов банаховой алгебры. При определённой доле воображения можно считать, что это помогает знать что-то в квантовой механике, потому что однопараметрические группы автоморфизмов — это модели развития во времени квантово-механической системы. Но это пусть говорят физики. А нам важно, что знать автоморфизмы — это интересная задача.

Из того, что я сказал, следует, что очень важно знать, когда дифференцирования являются внутренними. Но это в конкретном примере. А дальше было замечено, что часто бывает так, что для алгебры A дифференцирование $D\colon A \to A$ не всегда является внутренним, но если вместо A рассмотреть какую-нибудь разумную объемлющую алгебру $\mathscr{A} \supset A$, то в такой расширенной области $\mathscr{A}$ любое дифференцирование алгебры A уже будет внутренним.

Типичный пример такой: $A = \mathscr{K}(H)$ — компактные операторы в гильбертовом пространстве H (упражнение: придумать дифференцирование этой алгебры, которое не является внутренним). Если вы рассмотрите дифференцирования со значениями в алгебре $\mathscr{B}(H)$ всех ограниченных операторов, то все такие дифференцирования являются внутренними, т. е. задаются коммутаторами. Почему это так, мы ещё скажем.

Пытаясь ответить на вопрос, являются ли все дифференцирования со значениями в данном бимодуле внутренними, естественно рассмотреть факторпространство всех дифференцирований по внутренним. Традиционно оно обозначается $\mathscr{H}^1(A, X)$ и называется *первой группой когомологий* банаховой алгебры A с коэффициентами в бимодуле X. То, что в обозначениях используется рукописная буква $\mathscr{H}$, говорит о том, что рассматриваются непрерывные когомологии. Есть ещё и n-мерные когомологии для всех натуральных n, но о них сегодня разговора не будет.

Если Вы занимаетесь дифференцированиями, естественно спросить себя, как же устроены банаховы алгебры A, для которых $\mathscr{H}^1(A, \cdot) \equiv 0$, т. е. их любые дифференцирования со значениями во всех бимодулях являются внутренними? В алгебре эти вещи тоже интересны; там они называются *сепарабельными* алгебрами. В функциональном анализе нельзя это слово произнести, оно занято совсем другими объектами. Здесь они называются *стягиваемыми* алгебрами. Название унаследовано из топологии, потому что из условия $\mathscr{H}^1(A, \cdot) \equiv 0$ следует, что $\mathscr{H}^n(A, \cdot) \equiv 0$ для всех n. А в топологии, скажем, в классе полиэдров, такое условие выделяет стягиваемые полиэдры. Отсюда терминология.

Итак, каковы же стягиваемые банаховы алгебры? В чистой алгебре ответ давно известен. Классическая теорема гласит, что стягиваемые (сепарабельные) алгебры — прямые суммы конечного числа полных матричных алгебр. Для банаховых алгебр такая прямая сумма (т. е., как говорят, классически полупростая алгебра), конечно же, является стягиваемой. А вот верно ли обратное, это один из старых открытых вопросов. Что здесь известно? Известно, что если Вы рассмотрите какой-то конкретный класс с разумными ограничениями, то всегда ответ положительный. Любая стягиваемая алгебра из этого класса классически полупроста. Например, если Вы возьмёте C^*-алгебры, групповые алгебры обоих типов, коммутативные алгебры, то это верно. А в общем случае пока не получается из-за патологических свойств геометрии банаховых пространств. Между прочим, если пример есть, то он очень экзотичен. Тем не менее, я бы не рискнул делать предположение, что обратная теорема верна. Сейчас, по слухам, такой контрпример пытается построить Чарльз Рид — тот самый, который придумал оператор без инвариантных подпространств в банаховом

пространстве. А косвенное подтверждение того, что в принципе ничто не запрещает такие контрпримеры, такое. Если мы чуть-чуть расширим класс алгебр, и вместо банаховых алгебр рассмотрим более общие топологические алгебры и даже метризуемые, то алгебра $\mathbb{C}^M$ всех функций на произвольном множестве M с покоординатной сходимостью будет стягиваемой.

Мораль, тем не менее, такова. Мы ещё толком не знаем в точности, каковы стягиваемые алгебры. Но, во всяком случае, мы видим, что узок круг этих алгебр. Их очень мало. Условие стягиваемости является чрезвычайно сильным.

Барри Джонсон в 1972 г. в терминах дифференцирований предложил класс банаховых алгебр, который имеет оптимальный размер. Этот класс оказался чрезвычайно удобным и впоследствии всё более и более популярным. Это и есть аменабельные алгебры. Чтобы их определить, мне нужно понятие дуального банахова модуля. Пусть X — левый банахов модуль над банаховой алгеброй A. Тогда его дуальное (сопряжённое) банахово пространство X^* автоматически оказывается правым банаховым модулем: $\langle f \cdot a,\, x\rangle = \langle f,\, a \cdot x\rangle$. Точно так же, если был правый модуль, то сопряжённый к нему является левым модулем. А если с самого начала был бимодуль, то и сопряжённое пространство окажется бимодулем.

Сопряжённых бимодулей гораздо меньше, чем произвольных, и работать с ними гораздо приятнее. Главная причина заключается в том, что единичный шар компактен в слабой* топологии. Когда мы что-то ищем, можно брать ограниченные направленности и их предельные точки. Как правило, они дадут то, что нужно.

Теперь определение аменабельности по Джонсону (1972 г.). Банахова алгебра A называется *аменабельной* *), если для любого бимодуля X первая группа когомологий с коэффициентами не в самом X, а в его сопряжённом, равна нулю: $\mathscr{H}^1(A, X^*) = 0$. Это условие гораздо терпимее, поэтому аменабельных алгебр больше, чем стягиваемых. То, что данное определение весьма содержательно, показывает следующая теорема Джонсона; заодно она даёт и оправдание термина «аменабельная», который до этого был известен только в теории групп.

Т е о р е м а 2 (Джонсон). *Банахова алгебра $L^1(G)$ аменабельна тогда и только тогда, когда G аменабельна как локально компактная группа.*

Я ещё не объяснял, что такое аменабельность для локально компактных групп. Это проще всего сказать так. Одно (а значит, и все остальные)

*) Теперь говорят *аменабельной по Джонсону*, потому что появился целый ряд других вариантов. Второй по важности, самый известный, — Алэна Конна.

из нескольких стандартных функциональных пространств, связанных с G, имеет инвариантные средние. А именно, если мы возьмём, например, все ограниченные непрерывные функции $C_b(G)$, то на этом пространстве существует функционал $L\colon C_b(G) \to \mathbb{C}$, который непрерывен, нормализован (т. е. константу 1 переводит в число 1), а главное, левоинвариантен — он не зависит от сдвигов. Отсюда следует, что он правоинвариантен.

Времени доказывать теорему Джонсона у меня нет, а если бы было, я бы дал простое доказательство этой теоремы, но совершенно другими методами, основанными как раз на плоских модулях, о которых речь пойдёт немножко дальше.

После теоремы Джонсона популярность аменабельных алгебр стала расти. Она и сейчас продолжает расти. Сейчас все уже согласны с тем, что это один из самых важных классов банаховых алгебр в общих терминах (не обслуживающий конкретную ветвь анализа, а сформулированный в общих терминах теории).

Возник типовой вопрос, ответ на который каждый раз будет приносить приятную конкретность. Банаховы алгебры работают в разных областях анализа. Групповые алгебры — в гармоническом анализе, операторные — в теории операторов и т. д. Возьмём какой-нибудь конкретный класс, действующий в той или иной науке. И спросим, какие алгебры внутри этого класса аменабельны, а какие нет. Образец — теорема Джонсона; она даёт полный ответ в классе $L^1(G)$. Сразу же люди стали думать о C^*-алгебрах. Но это оказался трудный орешек. По крайней мере, та теорема, о которой я сейчас упомяну, это результат совместных усилий по меньшей мере шести человек. Самый большой вклад внёс, по-видимому, Алэн Конн в работе 1978 г. А завершающий удар нанёс датчанин Уффе Хаагеруп, который вообще известен тем, что он ставит точки в старых открытых проблемах. Теорема получилась такая.

Теорема 3. *C^*-алгебра аменабельна тогда и только тогда, когда она ядерная.*

Ядерные алгебры — это очень важный и определяемый во внутренних терминах класс в теории C^*-алгебр. Но у меня нет времени о нём подробно рассказывать, потому что тогда мы уйдём в сторону. Есть много разных определений. Самое короткое такое. C^*-алгебра ядерная, если её обёртывающая алгебра фон Нойманна гиперфинитна. Ядерные алгебры очень хороши для тензорных произведений.

Стягиваемых C^*-алгебр гораздо меньше, потому что есть следующая теорема.

Теорема 4 (Селиванов). *C^*-алгебра стягиваема тогда и только тогда, когда она конечномерна.*

Ещё один результат заслуживает упоминания. Он из теории функций и относится к равномерным алгебрам. Равномерная алгебра — это замкнутая подалгебра алгебры всех непрерывных функций на своём гельфандовском спектре.

Теорема 5 (М. В. Шейнберг, 1973). *Равномерная алгебра A аменабельна тогда и только тогда, когда $A = C(\Omega)$, т. е. A совпадает со всеми непрерывными функциями.* (*Стало быть, алгебра на диске не аменабельна.*)

Для сравнения отметим, что равномерная алгебра A стягиваема тогда и только тогда, когда она изоморфна $\mathbb{C}^n$ с покоординатным умножением.

Между прочим, эта теорема была доказана давно, но в своё время не была должным образом оценена. А сейчас вокруг неё творится запоздалый бум. Пизье её как-то применил и т. д.

Постепенно за эти годы все наиболее важные и наиболее популярные классы алгебр анализа получили соответствующую разгадку: какие из них аменабельны, а какие нет. Но одна алгебра сопротивлялась — та самая алгебра $M(G)$. Теорема, которую я сформулировал, была доказана в июле 1999 г. Что было известно, и почему возникла та гипотеза, которая была в конце концов подтверждена? Был хорошо известен коммутативный случай. Если G коммутативна, то и её алгебра мер тоже коммутативна. А значит, работает вся гельфандовская теория. У алгебры мер тогда есть пространство максимальных идеалов — гельфандовский спектр, и она представима в виде алгебры каких-то (не всех, конечно) непрерывных функций. Беда лишь в том, что хотя для алгебры $L^1(G)$ в случае коммутативной G спектр известен, и очень хорошо известен — это просто группа, двойственная по Понтрягину, спектр алгебры $M(G)$ — нечто невообразимо ужасное. Чем больше люди с ним работают, тем больше они всяких парадоксов обнаруживают.

В конце 60-х годов Джозеф Тэйлор сделал такое наблюдение, которое ещё раз говорит о том, какой это странный объект. На спектре алгебры $M(G)$ есть участки, которые являются аналитическими многообразиями, и даже бесконечномерными. Но для нас сейчас важно, что на этом спектре всегда есть аналитические диски, т. е. после соответствующего отождествления сужения наших функций (преобразований Гельфанда) будут аналитическими на этих дисках. А аналитичность — это враг аменабельности. Достаточно в качестве бимодуля взять комплексную плоскость *) $\mathbb{C}$ и рассмотреть функционал, который сопоставляет каждой мере производ-

*) Это, кстати сказать, в своё время тоже было одной из причин, почему нужно рассматривать по возможности больше бимодулей, а не только саму алгебру.

ную соответствующего сужения на этот диск преобразования Гельфанда. Легко видеть, что это как раз и будет дифференцирование (так называемое *точечное* дифференцирование), не являющееся внутренним. Так довольно дёшево получается, что если группа G коммутативна и непрерывна, то об аменабельности нечего и мечтать.

Я замечу, что Тэйлор известен своими выдающимися работами начала 70-х годов, где он впервые дал правильное определение совместного спектра коммутирующих операторов в банаховом пространстве. Но и до этого у него были очень интересные работы по алгебре мер, которые я цитировал.

Так вот, был известен коммутативный случай. Потом был целый ряд работ, в которых, отправляясь от этого случая, обобщался подобный результат. Например, пусть группа некоммутативная, но у неё есть достаточно большая коммутативная факторгруппа. Тогда тоже можно такую вещь доказывать. А что делать, если группа совсем некоммутативная (например, нет нормальных подгрупп), было непонятно. Забегая вперёд, я скажу, что общее рассуждение основано на совсем другом и ничего этого не использует. Оно основано на том, что в непрерывных группах есть тощие (meagre) множества.

2.
На Востоке

Всё остальное время будет посвящено разным обсуждениям, как же эту теорему доказать. Для этого нужно вернуться в тот же 72-й год и вспомнить, что тогда был «Железный занавес». На Западе есть город Ньюкасл; там сидит Барри Джонсон. Он держит плакат, на котором написано: «Аменабельность». На Востоке есть Москва. Там в эти же годы имела место некая деятельность, с виду никакого отношения к аменабельности и Джонсону не имевшая. А именно, была попытка построить нечто вроде общей гомологической теории сначала банаховых алгебр, а затем и более общих топологических алгебр. Причём цели этой деятельности поначалу тоже не имели никакого отношения к аменабельности. Они были таковы. Уже к этому времени существовали апробированные и важные (все были с этим согласны) непрерывные варианты понятий когомологий $\mathscr{H}^n(A, X)$ и гомологий $\mathscr{H}_n(A, X)$ банаховых алгебр с коэффициентами в банаховых бимодулях. Они определялись как гомологии и когомологии так называемых стандартных когомологических комплексов — прямой аналог (непрерывный) комплексов Хохшильда. Главной мотивацией этой деятельности было желание научиться вычислять эти группы, не будучи привязанными к их

исходным определениям в терминах стандартных комплексов, а пользуясь мощными методами, открытыми Картаном, Эйленбергом, Маклейном — отцами-основателями гомологической алгебры. Эти методы заключаются в том, что мы уже не привязаны к стандартным комплексам, а вычисляем наши инварианты с помощью так называемых резольвент тех или иных модулей. А резольвенты эти состоят из модулей, в гомологическом смысле наилучших. А именно:

— проективных;

— инъективных;

— плоских.

Здесь я произнёс три ключевых слова. Это три кита как чистой гомологической алгебры, так и банаховой гомологической алгебры. Но некоторые знают, что для топологических алгебр остаётся два кита, потому что инъективных модулей, как правило, нет. Это недавнее открытие Алёши Пирковского. А у нас будет три кита.

Чтобы всё это применять (а как применять — об этом разговор и пойдёт), нужно дать правильные (это значит — работающие) функционально-аналитические варианты этих трёх ключевых понятий гомологической алгебры: проективности, инъективности и плоскости. Самыми главными для нас будут плоские модули.

Пусть A — фиксированная банахова алгебра. Возникает много разных категорий модулей над A; теперь важно, что это категории. Категория левых модулей обозначается A-mod, правых — mod-A, бимодулей — A-mod-A. Объекты этих категорий — модули, а морфизмы — это морфизмы в чисто алгебраическом смысле (то, что в теории представлений называется сплетающими операторами), но только они должны быть непрерывными, как всегда всё у нас.

Важным для дальнейшего обстоятельством является то, что всё это разнообразие категорий модулей на самом деле кажущееся. Все модули сводятся к левым, но только над более сложной алгеброй. Для меня будет важно, что любой бимодуль можно рассматривать как левый модуль над алгеброй A^e — так называемой *обёртывающей* алгеброй исходной банаховой алгебры. Алгебра A^e строится следующим образом. Присоединим к алгебре A единицу, если её не было. Получим алгебру A_+. Затем рассмотрим противоположную алгебру A_+^{op}, в которой произведение ab определяется как произведение ba в алгебре A_+. Наконец, нужно построить проективное тензорное произведение (классическое произведение Гротендика в теории банаховых пространств) $A_+ \hat{\otimes} A_+^{\mathrm{op}}$. Тогда для заданного бимодуля левое внешнее умножение корректно определено по правилу $(a \otimes b) \cdot x = a \cdot x \cdot b$. Поэтому, если что-то определить для левых модулей,

то это будет определено и для всех модулей. Понятно, что всё можно свести и к правым модулям. Это тоже будет важно.

Фиксируем $X \in A\text{-mod}$. Из всех определений категорий проективных, инъективных и плоских модулей я выберу самое короткое, в терминах главных функторов, действующих в этих категориях модулей. Для фиксированного X можно определить три функтора. А именно, ковариантный функтор морфизмов ${}_Ah(X, ?)\colon A\text{-mod} \to \mathscr{B}an$, контравариантный функтор морфизмов ${}_Ah(?, X)\colon A\text{-mod} \to \mathscr{B}an$ и функтор тензорного произведения над A, который действует уже не на категории левых модулей, а на категории правых модулей: $? \,{}_A\hat{\otimes}\, X\colon \ \text{mod-}A \to \mathscr{B}an$. Здесь $\mathscr{B}an$ — категория банаховых пространств. Первые два функтора — специализации функтора морфизмов, которые есть в любой категории. Например, первый функтор на объекты действует так: $Y \mapsto {}_Ah(X, Y) \subset \mathscr{B}(X, Y)$. Между прочим, подпространство ${}_Ah(X, Y)$, состоящее из морфизмов из X в Y, замкнуто в пространстве всех непрерывных операторов X в Y. То же самое правило, которое есть на первых страницах любого учебника по теории категорий, показывает, как этот функтор действует на морфизмах. Только результат теперь будет не просто отображение множеств, а линейный непрерывный оператор. Точно так же сопоставление $Y \mapsto {}_Ah(Y, X) \subset \mathscr{B}(Y, X)$, вместе с соответствующими определениями для морфизмов, даёт контравариантный функтор. Между прочим, самым важным и общеизвестным специальным его случаем является тот случай, когда $A = X = \mathbb{C}$. Тогда получается функтор перехода к сопряжённому банахову пространству и сопряжённым операторам.

Третий функтор действует так. Каждому правому банахову модулю Y сопоставляется банахово пространство

$$Y \,{}_A\hat{\otimes}\, X := Y \,\hat{\otimes}\, X / \overline{l.\ h.}\{x \cdot a \otimes y - x \otimes a \cdot y\}.$$

Здесь $\overline{l.\ h.}$ — замыкание линейной оболочки. Факторизация производится для того, чтобы в результате получить свойство универсальности для сбалансированных операторов. В этом смысл, как в алгебре, так и в анализе, понятия тензорного произведения модулей. Только я хочу подчеркнуть, что левое и правое умножения как бы убивают друг друга, и результат будет всего лишь банаховым пространством.

Как определяются проективность и прочее? Если бы я был в чистой алгебре, то там нет никаких проблем. Я бы просто сказал, что модуль X является проективным, инъективным или плоским, если, соответственно, ковариантный функтор морфизмов, контравариантный функтор морфизмов или функтор тензорного произведения сохраняет точность последовательностей. Если дана точная последовательность модулей из катего-

рии области определения, то после применения нашего функтора должна остаться точность.

Если я теперь повторю это определение, но произнесу слова «непрерывность», «банаховость», поставлю над тензорным произведением «крышку», то получится какое-то определение, и даже довольно осмысленное. Беда в том, что будет слишком мало проективных, инъективных и, что сейчас для меня наиболее важно, плоских модулей, чтобы развить содержательную гомологическую теорию, в частности, хорошо считать гомологии и когомологии. Нужно сделать понятия более терпимыми.

В своё время это было сделано с помощью такого предварительного определения. Рассмотрим точную (в чисто алгебраическом смысле) последовательность банаховых модулей $\dots \leftarrow Y_n \xleftarrow{d_n} Y_{n+1} \leftarrow \dots$, где d_n, $n \in \mathbb{Z}$, — линейный ограниченный оператор. Эта последовательность называется *допустимой*, если она в некотором смысле идеальна как последовательность банаховых пространств и линейных непрерывных операторов после забывания дополнительной структуры — внешнего умножения. Сказать это точно можно по-разному. Алгебраический метод такой: последовательность допустима, если обладает стягивающей гомотопией, состоящей из линейных непрерывных операторов $\dots \dashrightarrow Y_n \overset{s_n}{\dashrightarrow} Y_{n+1} \dashrightarrow \dots$ Я пишу здесь пунктирные стрелки, потому что хочу подчеркнуть, что это, вообще говоря, не морфизм. Он не реагирует на внешнее умножение. Просто должна существовать последовательность линейных непрерывных операторов s_n, для которых выполняется тождество $ds + sd = 1$, где 1 — тождественный оператор в соответствующем банаховом пространстве. Это алгебраический подход. Ясно, что он унаследован из топологии. Есть и геометрический подход. То же самое получится, если в каждом члене потребовать следующее. Поскольку мы уже знаем, что последовательность точная, то можно рассмотреть банахово подпространство $\operatorname{Ker} d_n = \operatorname{Im} d_{n+1}$. Эквивалентное требование состоит в том, что это банахово подпространство обладает банаховым дополнением, т. е. существует дополнительное замкнутое подпространство.

Теперь дадим правильное определение. Банахов модуль называется *проективным*, если его ковариантный функтор морфизмов переводит любую допустимую точную последовательность в точную. Банахов модуль называется *инъективным*, если таким же свойством обладает его контравариантный функтор морфизмов. Наконец, и для сегодняшнего доклада это самое главное, банахов модуль называется *плоским*, если он сохраняет точность допустимых последовательностей после применения функтора тензорного произведения.

Сделаем некоторые наблюдения над этими понятиями.

1. Связь между инъективностью и плоскостью. Оказывается, что модуль X плоский тогда и только тогда, когда модуль X^* инъективный. Когда я говорю о плоскости правого модуля, то я говорю об инъективности левого модуля. Но я уже говорил, что так можно делать, потому что всё сводится к левым модулям.

Между прочим, это совсем нетрудная вещь. Её можно рекомендовать в качестве приятного упражнения. Главным моментом, который либо можно принять на веру, либо доказать (но тогда это уже будет долгая история), является следующее утверждение, которое вообще лежит в основе всей банаховой гомологии. *Комплекс банаховых пространств* (о модулях мы забыли) *и линейных непрерывных операторов точен тогда и только тогда, когда сопряжённый комплекс точен.*

Доказательство этого упражнения довольно приятное, потому что оно использует основные принципы функционального анализа — теорему Хана—Банаха и теорему Банаха об обратном операторе; все эти теоремы работают.

2. Проективные модули могут быть охарактеризованы внутренним образом, а именно, как прямые слагаемые так называемых свободных банаховых модулей. Можно дать общекатегорное определение, но я лучше дам явную конструкцию. Рассмотрим банахово пространство $A_+ \hat{\otimes} E$, где E — банахово пространство. Это банахово пространство является также и банаховым модулем: ясно, как определить внешнее умножение с помощью внутреннего умножения в A_+. Это и есть свободный банахов модуль.

3. Из предыдущего свойства легко выводится, что если модуль проективен, то он заведомо плоский. Теперь заявление, которое оправдывает всю эту науку: «Обратное, вообще говоря, неверно.» Без этого наука была бы тривиальна и неинтересна.

Это — общее место в чистой алгебре. В чистой алгебре этот феномен (разница между жёстким, негибким, требовательным понятием проективности и гораздо более часто встречающимся гибким понятием плоскости) — одно из важнейших наблюдений. То же самое (мне даже кажется, что ещё в большей степени) имеет место и в анализе. И в разных областях анализа это приобретает разные формы. У меня нет времени рассказывать об этих формах подробнее, но хочется об этом хотя бы упомянуть. Так мы естественно подойдём к нашей главной теме. То, что я сейчас буду писать, это на самом деле — содержательные теоремы. Запишу я их в виде пропорции:

$$\frac{\text{проективность}}{\text{плоскость}} = \frac{\text{компактные группы}}{\text{аменабельные группы}} = \frac{\text{факторы типа I}}{\text{произвольные гиперфинитные факторы}} =$$

$$= \frac{\text{спектр имеет окрестность, содержащую аналитический диск}}{\text{спектр имеет слой, содержащий аналитический диск}} =$$
$$= \frac{\text{паракомпактное локально компактное пространство}}{\text{произвольное локально компактное пространство}} =$$
$$= \frac{\text{банаховы алгебры с единицей}}{\text{банаховы алгебры с ограниченной аппроксимативной единицей}} =$$
$$= \frac{\text{стягиваемая алгебра}}{\text{аменабельная алгебра}}.$$

Что такое ограниченная аппроксимативная единица? В отличие от алгебры, обладать единицей в алгебре функционального анализа — большая роскошь, иногда непозволительная. Такие почтенные алгебры, как алгебра компактных операторов, уже упоминавшаяся алгебра $L^1(G)$, единицами не обладают. А присоединять единицы, как это делают в алгебре, — это совершать насилие. Появляется чуждый элемент, например, некомпактный оператор (тождественный). Но часто так бывает, что единицы нет, но есть её следующая замена. Правая ограниченная аппроксимативная единица в банаховой алгебре A — это такая ограниченная направленность e_λ, что для любого элемента $a \in A$ направленность ae_λ стремится к a. Понятно, как определяется левая аппроксимативная единица; понятно, как определяется двусторонняя.

Ясно, что компактные операторы в гильбертовом пространстве обладают аппроксимативной единицей: нужно взять диагональные матрицы, у которых на первых n местах стоят единицы, а все остальные элементы равны нулю. Самый классический пример — дельтаобразная последовательность в алгебре $L^1(G)$.

Теперь теорема, которая является основной леммой. Она носит общий характер. Для формулировки этой теоремы нужны понятия дополняемого и слабо дополняемого банахова модуля. Левый модуль $I \subset A$ называется *дополняемым*, если он имеет дополнительное банахово подпространство, т. е. естественное вложение in: $I \to A$ является коретракцией в $\mathscr{B}an$ (имеет левый обратный ограниченный линейный оператор). Бывает так, что вложение не имеет левого обратного оператора, но зато сопряжённый оператор in^*: $A^* \to I^*$ является ретракцией, т. е. имеет правый обратный. В такой ситуации мы говорим о *слабой дополняемости*. Классический пример: все сходящиеся к нулю последовательности образуют идеал $c_0 \subset l_\infty$ в пространстве всех ограниченных последовательностей. Этот идеал не дополняем, но слабо дополняем.

Теорема 6 (Шейнберг, Хелемский). *Пусть A — банахова алгебра, $I \subset A$ — замкнутый левый идеал. Предположим, что алгебра A обладает единицей или хотя бы ограниченной аппроксимативной единицей. Рассмотрим фактормодуль $X := A/I$. Предположим, что*

«I хорошо в A расположен», т. е. является дополняемым или хотя бы слабо дополняемым. В таком случае модуль X — плоский тогда и только тогда, когда идеал I обладает правой ограниченной аппроксимативной единицей.

Идеалы были левыми, а ограниченная аппроксимативная единица должна быть правой. Такой критерий. Как он будет работать — об этом дальше. Пока что я сначала дам интерпретацию наглядную, но менее важную для того, что будет потом, а затем уже дам более существенную. Что лежит на поверхности, если уже знать эту теорему? Обратимся к теории операторных алгебр. Пусть A — C^*-алгебра. Предположим, что дано её алгебраически циклическое представление, т. е. инволютивное представление с циклическим вектором. Очень просто доказать, что тогда гильбертово пространство этого представления является фактормодулем $H = A_+/I$, где идеал I — аннулятор циклического вектора. Есть хорошо известная теорема Сигала о том, что всякий левый замкнутый идеал в C^*-алгебре обладает правой ограниченной аппроксимативной единицей. Отсюда мы немедленно получаем, что наш модуль плоский. А дальше общие методы гомологической алгебры дают, в частности, что $\mathscr{H}^n(A, \mathscr{B}(H)) = 0$ для всех $n > 0$. Это может представлять интерес в связи с классической в теории операторных алгебр проблемой, стоящей с 1957 года — так называемой проблемой подобия Кадисона, которая такова: всякое ли (вообще говоря, неинволютивное) представление C^*-алгебры в гильбертовом пространстве подобно инволютивному. Это довольно важный вопрос. Недавно выяснилось, что такого рода вопросы — это вопросы о когомологиях. А именно, этот вопрос эквивалентен тому, что $\mathscr{H}^1(A, \mathscr{B}(H)) = 0$ для всех без исключения представлений (инволютивных). Наш результат имеет с желаемым «симметрическую разность»: тривиальность когомологий доказана для весьма специальных представлений, но зато для всех n.

Теперь более важная для нас вещь. Теорема 6 позволила в своё время обнаружить бесконечномерные банаховы алгебры, полупростые, у которых все без исключения банаховы модули плоские. Для этого достаточно (это переносится из обычной алгебры без особых изменений) знать плоскость бимодуля A_+. Из этого уже следует, что все модули — односторонние или двусторонние — являются плоскими. Но если на бимодуль A_+ внимательно посмотреть, то выяснится, что он как раз является частным случаем тех самых фактормодулей. А именно, мы уже знаем, что бимодуль — это то же самое, что левый модуль над обёртывающей алгеброй A^e. Оказывается, что $A_+ = A^e/I^\Delta$, где I^Δ — так называемый *диагональный идеал*. Он представляет собой ядро морфизма произведения $\pi_+ \colon A_+ \hat{\otimes} A_+ \to A_+$, который

переводит $a \otimes b$ в ab. Морфизм произведения корректно определён и является морфизмом бимодулей. Идеал называется диагональным по следующей причине. Если бы алгебра A была коммутативна, то у неё был бы гельфандовский спектр Ω. Тогда у обёртывающей алгебры, как тензорного произведения, гельфандовский спектр — декартов квадрат $\Omega \times \Omega$. Тогда ядро (диагональный идеал) состоит из функций, обращающихся в нуль на диагонали, т. е. $(s, s) \mapsto 0$ для всех $s \in \Omega$. Между прочим, этот идеал весьма популярен в некоммутативной геометрии, особенно после работ Конна. Элементы его называются *некоммутативными дифференциальными формами первого порядка*.

Что это даёт в данном случае? Рассмотрим самую популярную коммутативную банахову алгебру — C на любом компакте; если угодно, возьмите отрезок. Кстати сказать, когда сама алгебра A обладает единицей, «плюсик» писать не нужно: $A_+ = A$. Пусть $A = C(\Omega)$. Её обёртывающая алгебра — $C(\Omega) \hat{\otimes} C(\Omega)$. Это, между прочим, не то же самое, что $C(\Omega \times \Omega)$; этих функций меньше. Но алгебра $C(\Omega) \hat{\otimes} C(\Omega)$ хорошо известна. Она называется *алгеброй Варопулоса*. Эта алгебра очень важна в гармоническом анализе. С помощью этой алгебры Варопулос объяснил решение Майявена проблемы спектрального синтеза. Алгебра Варопулоса довольно сложная, но её идеалы, тем не менее, хорошо изучены. В частности, хорошо известно, что диагональный идеал, состоящий из функций, обращающихся в нуль на диагонали, заведомо обладает ограниченной аппроксимативной единицей. Значит, бимодуль $A_+ = A$ плоский.

3.
Синтез

В своё время мы получили эту информацию и, может быть, так бы с ней и сидели. И этот класс банаховых алгебр, плоских как бимодули, может быть, так бы и остался в качестве интересного примера — не более того. Но в этот момент сквозь железный занавес пробился мемуар Джонсона, любезно посланный автором. Мы его читаем и видим определение аменабельных алгебр. Класс банаховых алгебр явно очень интересный. И всё время по ходу обсуждения возникают ограниченные аппроксимативные единицы. Например, если алгебра аменабельная, то она заведомо должна обладать ограниченной аппроксимативной единицей. Идеалы при определённых свойствах тоже должны обладать такой единицей. Тут запахло жареным — вроде бы, это очень похоже на то, чем мы занимались. И действительно, после некоторого раздумья была доказана такая теорема.

Т е о р е м а 7 (Шейнберг—Хелемский). *Для любой банаховой алгебры A следующие условия эквивалентны*:

1) *A аменабельна по Джонсону*;

2) *все модули над A плоские*;

3) *бимодуль $A_+ \in A$-mod-A плоский*;

4) *гомоморфизм $\pi_+^* \colon A_+^* \to (A_+ \hat{\otimes} A_+)^* = \mathrm{Bil}(A_+)$, сопряжённый гомоморфизму произведения π_+, обладает левым обратным в категории бимодулей* (*т. е. таким ζ, что $\zeta\pi_+^* = 1$*).

Теперь мы видим альтернативный подход к понятию аменабельности. Можно на аменабельность смотреть двумя глазами. В частности, какие из этого можно сделать практические выводы? Главный, к которому я подбираюсь, — это про алгебру мер. Но можно рассказать ещё про одно приложение, в своё время оказавшееся для нас неожиданным. Прежде чем об этом говорить, нужно сказать, что теорема в действительности тройная. Теорема 7 — это только её треть. Вторая её часть, значительно более простая и доказанная значительно раньше, говорит о стягиваемости. А именно, алгебра A стягиваема тогда и только тогда, когда любой модуль над A проективный (а не плоский), бимодуль $A_+ \in A$-mod-A проективный (а не плоский), наконец, гомоморфизм π_+ (а не сопряжённый к нему) обладает правым обратным. Кроме того, есть ещё третья колонка, которая касается аменабельности по Конну. Это, наоборот, более поздний и более трудный результат.

Теперь иллюстрация — новые примеры не дополняемых (и даже не слабо дополняемых) подпространств, т. е. приложение к геометрии банаховых пространств — старой науке. Предположим, что нам сказали, что алгебра A аменабельна. И вдруг мы нашли у неё левый идеал $I \subset A$, который по каким-то причинам заведомо не обладает правой ограниченной аппроксимативной единицей. Когда это может произойти? Если бы идеал I был слабо дополняемым, то тогда фактор по нему не мог бы быть плоским. А над аменабельной алгеброй все модули плоские. Противоречие. Это всё выглядит схоластически. Но такие примеры, оказывается, есть. Их даёт классический гармонический анализ. Рассмотрим алгебру $L^1(G)$, где G — коммутативная некомпактная группа. Про такие группы классический результат Майявена 1956 года (он и называется *отрицательным решением проблемы спектрального синтеза*) состоит в том, что для них в алгебре $L^1(G)$ существует левый идеал, который не только не обладает правой ограниченной аппроксимативной единицей, но даже не совпадает со своим топологическим квадратом: $\overline{I^2} \neq I$. Следствие такое: идеалы Майявена (а они есть) не являются слабо дополняемыми (как подпространства). Ещё раз повторю, почему. Предполо-

жим, что идеал Майявена I слабо дополняем. Рассмотрим фактормодуль $X = A/I$. С одной стороны, он плоский, потому что алгебра аменабельна: все модули над ней плоские. А раз он слабо дополняем, то из критерия следует, что он обладает правой ограниченной аппроксимативной единицей.

Теперь второй пример, значительно более важный для основной темы доклада. Предположим, что мы занимаемся какой-то алгеброй, и хотим знать, аменабельна она или нет. Пусть мы догадались, что она не аменабельна. Как это доказать? Для этого нужно просто предъявить хотя бы один неплоский модуль. И это в точности то, что было сделано с алгеброй мер. В силу теоремы 6, чтобы предъявить неплоский модуль, достаточно предъявить левый замкнутый идеал, который был бы (хотя бы слабо) дополняемым, но по каким-то причинам не обладал бы правой ограниченной аппроксимативной единицей. Нашли такой идеал — всё, алгебра не аменабельна. Где нашли такой идеал в алгебре мер $M(G)$?

Теорема 8. *Пусть G — произвольная непрерывная локально компактная группа. Тогда в ней существует замкнутый* (*и даже компактный*) *нормальный делитель $N \subset G$, обладающий следующим свойством*: *левый замкнутый идеал $I = \{\mu\colon |\mu|(xN) = 0,\ x \in G\}$ дополняем в алгебре мер $M(G)$ и $\overline{I^2} \neq I$.*

Из того, что было сказано ранее, автоматически следует, что алгебра $M(G)$ не аменабельна. Дополнительно можно сказать, что если группа G метризуема (а это главный случай, к которому сводятся общие разговоры), то тогда в качестве N можно взять единичную подгруппу: $N = \{e\}$. Тогда $I = M_c(G)$ — идеал непрерывных мер. Этот идеал не совпадает со своим топологическим квадратом. По-видимому это верно не только для метризуемых групп, но и для всех остальных. *)

Вот что конкретно было сделано в гармоническом анализе.

За оставшееся время я попробую рассказать, как доказывается теорема 8. Я сосредоточусь на метризуемом случае. Чтобы доказать, что $\overline{I^2} \neq I$, достаточно предъявить линейный непрерывный функционал $\Lambda\colon M_c(G) \to \mathbb{C}$, который обладает следующими свойствами: $\Lambda \neq 0$ и $\Lambda(\mu * \nu) = 0$, т. е. сам функционал не равен нулю, но на любой свёртке он равен нулю. Ядро этого функционала меньше $M_c(G)$, но оно содержит его топологический квадрат. Этот функционал связан с тощими множествами, о которых я упоминал. Оказывается (и это главное), если группа G непрерывна, то существует компактное множество $K \subset G$ (канторообразное), которое обладает следующими свойствами:

*) Недавно (осенью 2000) это было доказано.

1) $x_1 x_2^{-1} x_3 x_4^{-1} \neq e$ для любых четырёх различных элементов K. Из этого следует, что $|sK \cap tK| \leqslant 3$ при $s \neq t$, т. е. компакт в каком-то смысле достаточно маленький;

2) существует непрерывная мера $\mu > 0$, сосредоточенная на K, т. е. компакт в каком-то смысле достаточно большой.

Если такой компакт K найден, то мы полагаем $\Lambda(\mu) = \mu(K)$. Формула для свёртки, написанная в начале лекции, и свойство 1 показывают, что $\mu * \nu(K)$ — это интеграл от функции, отличной от нуля не более чем на счётном множестве точек. Такой интеграл равен нулю.

16 ноября 2000 г.

Приложение

Й. Меннике

ЛИНЕЙНЫЕ ГРУППЫ НАД $\mathbb{Z}$

1.
Конгруэнц-подгруппы $\mathrm{SL}_2(\mathbb{Z})$

Тема лекции — линейные группы над $\mathbb{Z}$. Примером линейной группы над $\mathbb{Z}$ служит группа $G = \mathrm{SL}_2(\mathbb{Z})$, состоящая из матриц $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, где a, b, c, $d \in \mathbb{Z}$ и $\det = ad - bc = 1$. Группа G порождена образующими $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ и $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Чтобы это доказать, рассмотрим произведения

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+tc & b+td \\ c & d \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ sa+c & sb+d \end{pmatrix}.$$

Используя алгоритм Евклида, от матрицы $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ можно перейти к матрице $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Другой набор образующих группы G состоит из матриц $A = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$ и $B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Доказательство этого утверждения — несложное упражнение. Эти образующие удовлетворяют соотношению

$$A^3 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = B^2,$$

Поэтому $A^3 = B^2$ и $B^4 = 1$. Можно доказать, что группа G задаётся образующими A и B и соотношениями $A^3 = B^2$ и $B^4 = 1$.

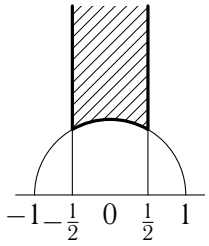
Группа G действует на гиперболической плоскости

$$\mathbb{H}^2 = \{z \in \mathbb{C}\colon\ \mathrm{Im}\ z > 0\}$$

следующим образом. Если $X = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, то $X(z) = \dfrac{az+b}{cz+d}$.

Доклад был прочитан по-русски и начинался словами: «Лекцию я буду читать по-русски. Я люблю русский язык, но русский язык не очень любит меня, поэтому иногда возникают трудности». — *Прим. ред.*

Поскольку наша группа действует на плоскости, можно поинтересоваться, есть ли у неё фундаментальная область. Ответ такой: да, есть. Пример фундаментальной области изображён на рис. 1.



Р и с. 1. Фундаментальная область $\mathrm{SL}_2(\mathbb{Z})$

Теперь мы определим в $G = \mathrm{SL}_2(\mathbb{Z})$ конгруэнц-подгруппы. Пусть $m \in \mathbb{N}$. Конгруэнц-подгруппа N_m состоит из матриц $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$, для которых $a \equiv d \equiv 1 \pmod m$ и $b \equiv c \equiv 0 \pmod m$. Подгруппа N_m нормальна в G. Это мы будем обозначать так: $N_m \lhd G$. Факторгруппа G/N_m изоморфна $\mathrm{SL}_2(\mathbb{Z}/m\mathbb{Z})$. Если $m = p_1^{e_1} \dots p_t^{e_t}$ — разложение на простые множители, то

$$G/N_m \cong \mathrm{SL}_2(\mathbb{Z}/p_1^{e_1}\mathbb{Z}) \times \dots \times \mathrm{SL}_2(\mathbb{Z}/p_t^{e_t}\mathbb{Z}).$$

Это — китайская теорема об остатках.

Структура группы $\mathrm{SL}_2(\mathbb{Z}/p^e\mathbb{Z})$ такова. Имеется такая последовательность подгрупп

$$\mathrm{SL}_2(\mathbb{Z}/p^e\mathbb{Z}) \supset H_1 \supset H_2 \supset \dots \supset H_n = \{1\},$$

что $\mathrm{SL}_2(\mathbb{Z}/p^e\mathbb{Z})/H_1 \cong \mathrm{SL}_2(\mathbb{F}_p)$ и

$$H_i/H_{i+1} \cong C_p \times C_p \times C_p,$$

где C_p — циклическая группа порядка p.

Проблема конгруэнц-подгрупп формулируется следующим образом. Пусть $N \lhd G$ и $|G/N| < \infty$, т. е. N — нормальная подгруппа в G конечного индекса. Верно ли, что найдётся $m \in \mathbb{N}$, для которого $N_m \subset N$?

Т е о р е м а 1. *Для группы $G = \mathrm{SL}_2(\mathbb{Z})$ решение проблемы конгруэнц-подгрупп отрицательно.*

Причина этого следующая. Группа чётных подстановок A изоморфна факторгруппе G/U, где $U \lhd G$. Этого уже достаточно, чтобы прийти к противоречию.

Дело в том, что согласно теореме Жордана—Гёльдера (см. [2]) для любой конечной группы H существует такая последовательность подгрупп

$$H = H_0 \supset H_1 \supset H_2 \supset \dots \supset H_n = \{1\},$$

что H_i/H_{i+1} — простая группа (фактор Гёльдера). При этом факторы Гёльдера определены однозначно с точностью до изоморфизма.

Группа U не может содержать конгруэнц-подгрупп. Действительно, последовательность подгрупп

$$G \supset U \supset \{1\}$$

обладает тем свойством, что $G/U \cong A_n$ — простая группа. Поэтому A_n — фактор Гёльдера. А если бы группа U содержала конгруэнц-подгруппу, то нашёлся бы фактор Гёльдера $\mathrm{PSL}_2(\mathbb{F}_p) \ncong A_n$.

На самом деле ситуация даже ещё хуже. Справедливо следующее утверждение.

Т е о р е м а 2. *Для любого $r \geqslant 2$ существует нормальная подгруппа $U \lhd G$ конечного индекса, изоморфная свободной группе F_r ранга r.*

Это, в частности, означает, что для любой конечной группы K существует подгруппа $V \lhd U$, для которой $U/V \cong K$.

2.
Конгруэнц-подгруппы $\mathbf{SL}_2(\mathbb{Z})$, $n \geqslant 3$

Обратимся теперь к группам $G = \mathrm{SL}_n(\mathbb{Z})$, $n \geqslant 3$. Для этих групп проблема конгруэнц-подгрупп имеет положительное решение.

Т е о р е м а 3. *Для любой нормальной подгруппы N в G конечного индекса существует $m \in \mathbb{N}$, для которого $N_m \subset N$.*

Д о к а з а т е л ь с т в о. Рассмотрим группу $Q_m = \mathrm{NCl}_G(I + me_{12})$, где

$$e_{12} = \begin{pmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

и I — единичная матрица, т. е.

$$I + me_{12} = \begin{pmatrix} 1 & m & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Здесь NCl_G — нормальное замыкание (нормализатор) относительно группы G. Чтобы получить нормальное замыкание элемента, берутся все элементы, с ним сопряжённые, и рассматривается порождённая ими группа.

Непосредственно из определения видно, что $Q_m \lhd N_m$.

Л е м м а 1. *Для любой нормальной подгруппы N в G конечного индекса существует $m \in \mathbb{N}$, для которого $Q_m \subset N$.*

Действительно, индекс N в G конечен, а значит найдётся такое t, что

$$\begin{pmatrix} 1 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}^t = \begin{pmatrix} 1 & t & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} \in N.$$

Тогда при $m = t$ группа Q_m лежит в N.

Теперь нам достаточно доказать, что $N_m = Q_m$, т. е. $N_m/Q_m = 1$.

Лемма 2. *Пусть $X \in N_m$. Тогда найдётся*

$$X' = \begin{pmatrix} a & b & \dots & 0 \\ c & d & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

сравнимый с X по модулю Q_m.

Приведём доказательство (см. [1]). Оно напоминает алгоритм Евклида.

Наша цель — изготовить единицу на диагонали. Если мы это сделаем, то, умножая на элементы из Q_m слева и справа, можно обнулить все внедиагональные элементы соответствующих строки и столбца. После этого можно применить индукцию (при $n = 2$ утверждение очевидно).

Если все внедиагональные элементы первого столбца нулевые, то доказывать нечего. Пусть НОД внедиагональных элементов равен km. С помощью алгоритма Евклида можно сделать так, чтобы у нашей матрицы (обозначим её $H = (h_{ij})$) $h_{21} = km$, $h_{i1} = 0$ (при $i > 2$). При этом, очевидно, $(h_{11}, k) = 1$. Найдём такое c, что $(ch_{11} + k, h_{32}) = 1$ (почему это можно сделать?). Умножим H слева на $I + cme_{21}$. Получим матрицу L, сравнимую с H по модулю Q_m, у которой $l_{32} = h_{32}$, $l_{21} = (ch_{11} + k)m$. Теперь подберём такие d и d', чтобы у матрицы $(I + de_{12})(I + d'e_{23})L(I - de_{12})(I - d'e_{23})$ на пересечении второй строки и второго столбца стояла единица.

Лемма 3. *Группа N_m/Q_m содержится в центре группы G/Q_m.*

Доказательство. Пусть $X \in N_m$,

$$X = \begin{pmatrix} a & b & \dots & 0 \\ c & d & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Поскольку $\mathrm{SL}_n(\mathbb{Z})$ порождается элементами вида $I + e_{ij}$, нам надо проверить, что коммутатор X с $I + e_{ij}$ лежит в Q_m для любых $i \neq j$.

Если $i, j > 2$, то коммутатор равен единице. При $i = 1, 2$, $j > 2$ (или наоборот) несложная проверка показывает, что коммутатор лежит в Q_m. Остались только $I + e_{12}$ и $I + e_{21}$. Но их можно выразить через остальные $I + e_{ij}$. Например, $I + e_{12} = (I - e_{13})(I - e_{32})(I + e_{13})(I + e_{32})$. □

Пусть X, $X' \in N_m$,

$$X = \begin{pmatrix} a & b & \dots & 0 \\ c & d & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} \quad \text{и} \quad X' = \begin{pmatrix} a & b & \dots & 0 \\ c' & d' & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Коэффициенты c, d, c', d' удовлетворяют уравениям $ad - bc = 1$, $ad' - bc' = 1$. Следовательно, найдётся такое целое число t, что $c' = c + ta$, $d' = d - tb$, т. е. $X \equiv X' \pmod{Q_m}$.

Таким образом, любой элемент из N_m/Q_m мы можем записать как

$$X = \begin{pmatrix} a & b & \dots & 0 \\ * & * & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Более того, умножая X справа на

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ tm & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} 1 & tm & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

мы видим, что нас интересуют не a и b, а $b \pmod a$ и $a \pmod b$.

Окончание доказательства вытекает из следующей леммы:

Л е м м а 4. *Пусть* $X, X' \in N_m$,

$$X = \begin{pmatrix} a & b & \dots & 0 \\ * & * & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} \quad \textit{и} \quad X' = \begin{pmatrix} a & b' & \dots & 0 \\ * & * & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Тогда

$$XX' \equiv \begin{pmatrix} a & bb' & \dots & 0 \\ * & * & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} \pmod{Q_m}.$$

Детали доказательства можно прочитать в [1].

Приведём набросок аккуратного доказательства теоремы.

Л е м м а 5. *Пусть* $X \in N_m$,

$$X = \begin{pmatrix} a & b & \dots & 0 \\ c & d & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Тогда для любого n *найдётся*

$$X'_n = \begin{pmatrix} a & b^n & \dots & 0 \\ (-1)^{n+1}c^n & d_n & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

сравнимый с X^n *по модулю* Q_m.

Доказательство — несложное упражнение.

Л е м м а 6. *Пусть* $X \in N_m$,

$$X = \begin{pmatrix} a & b & \dots & 0 \\ c & d & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

причём $b^n \equiv \varepsilon \pmod a$, *где* $\varepsilon = \pm 1$. *Тогда* $X^n \in Q_m$.

Доказательство. Пусть

$$X_n'' = X_n'(I + xme_{12}) = \begin{pmatrix} a & axm + b^n & \dots & 0 \\ c' & d' & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Так как $I + xme_{12} \in Q_m$ для любого целого x, а $X_n' \equiv X^n \pmod{Q_m}$, нам достаточно доказать, что $X_n'' \in Q_m$ для некоторого x.

Поскольку $a \equiv 1 \pmod{m}$, мы можем представить a в виде $a = 1 + km$. Определим x из уравнения $axm + b^n = -\varepsilon km$ (почему оно разрешимо в $\mathbb{Z}$?).

Несложная проверка показывает, что $(I - \varepsilon e_{21})X_n''(I + \varepsilon e_{21}) \in Q_m$, т. е. X_n'' лежит в Q_m. □

Пусть $X \in N_m$,

$$X = \begin{pmatrix} a & b & \dots & 0 \\ c & d & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Положим

$$X' = (I - te_{21})X(I + te_{21}) = \begin{pmatrix} a + tb & b & \dots & 0 \\ c' & d' & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Поскольку группа G/N_m — абелева, $X' \equiv X \pmod{Q_m}$.

Предположим, что a' и a'' являются членами арифметической прогрессии $\{a + tb\}$, $t \in \mathbb{Z}$. Пусть n' и n'' выбраны так, что $b^{n'} \equiv \varepsilon' \pmod{a'}$, а $b^{n''} \equiv \varepsilon'' \pmod{a''}$, где ε', $\varepsilon'' = \pm 1$. Тогда, по лемме 6, $X^{n'} \in Q_m$, $X^{n''} \in Q_m$, а следовательно, и $X^{(n', n'')} \in Q_m$ (алгоритм Евклида). Наша цель — найти такие a' и a'', чтобы $(n', n'') = 1$.

Выберем простое $p = a + tb$ (это можно сделать по теореме Дирихле). Пусть n — наименьшее число, для которого верно $b^n \equiv 1 \pmod{p}$. Тогда, очевидно, $n \mid p - 1 = 2^l q_1^{l_1} q_2^{l_2} \dots q_i^{l_i}$. Теперь выберем такие простые r и r', чтобы $r \equiv -p \pmod{bq_1q_2\dots q_i}$, $r' \equiv -1 \pmod{bq_1q_2\dots q_i}$ (опять же, это можно сделать по теореме Дирихле). Определим a': $a' = rr'$. Из того, как мы выбирали r и r' следует, что $a' \equiv p \pmod{b}$, а следовательно, a', как и p, является членом прогрессии $\{a + tb\}$. Пусть n' — наименьшее число, для которого выполняется $b^{n'} \equiv 1 \pmod{a'}$. Тогда $n' \mid (r - 1, r' - 1)$. Нетрудно проверить, что $r - 1 \equiv -2 \pmod{q_j}$, $r' - 1 \equiv -2 \pmod{q_j}$. Таким образом, у n и n' нет общих нечётных делителей.

Если либо n, либо n' нечётно, то положив $a'' = p$, $n'' = n$ мы получаем, что $X \in Q_m$. Осталось разобрать случай, когда n и n' чётные.

Пусть $b \not\equiv 0 \pmod{4}$ или $a \equiv -1 \pmod{4}$. Тогда возьмём $p = a + tb$, $p \equiv -1 \pmod{4}$. Если $2 \mid n$, то $b^{n/2} \equiv -1 \pmod{p}$. Следовательно, $b^{n/2} \equiv$

$\equiv -2 \pmod 4$, т. е. $n/2$ нечётно. В этом случае мы можем взять $a'' = p$, $n'' = n/2$.

Пусть $b \equiv 0 \pmod 4$ и $a \not\equiv -1 \pmod 4$. Тогда $a \equiv 1 \pmod 4$, так как $(a, b) = 1$. Возьмём $-p = a + tb$. Тогда $p \equiv -1 \pmod 4$, и дальнейшие рассуждения повторяют предыдущий случай. □

Литература

[1] J. M e n n i c k e. Finite factor groups of the unimodular group // Ann. Math. — 1965. — V. 81, № 1. — P. 31—37.

[2] М. Х о л л. Теория групп. — М.: ИЛ, 1962.

12 ноября 1999 г.

Оглавление

Предисловие 3
Ю. С. Ильяшенко. Столетняя история 16-й проблемы Гильберта 8
В. А. Васильев. Ветвящиеся интегралы и теории Пикара—Лефшеца 22
Б. Л. Фейгин. Конформные теории поля 59
М. А. Цфасман. Алгебраическая геометрия, теория чисел и их приложения к плотным упаковкам 66
В. Я. Иврий. Всё началось с Вейля 78
Ю. И. Манин. Некоммутативная геометрия и квантовые тэта-функции 91
Я. Г. Синай. Динамика адиабатического поршня (нарушение второго начала термодинамики) 109
С. И. Гельфанд. О числе решений квадратного уравнения 124
Ю. И. Манин. Проблема Морделла—Вейля для кубических поверхностей 134
А. В. Зелевинский. Обобщённые коэффициенты Литтлвуда—Ричардсона, канонические базисы и полная положительность 147
В. И. Арнольд. Теория распространения волн 161
А. А. Болибрух. Проблема Римана—Гильберта 202
С. К. Ландо. Об одном классе инвариантов графов, связанном с инвариантами Васильева узлов 222
А. Я. Хелемский. Плоские модули и гармонический анализ на группах 238
Й. Меннике. Линейные группы над $\mathbb{Z}$ 257