

НЕЗАВИСИМЫЙ МОСКОВСКИЙ УНИВЕРСИТЕТ

globus ГЛОБУС

Общематематический семинар. Выпуск 5

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

Москва
Издательство МЦНМО
2011

УДК 51(06)
ББК 22.1я5
Г54

Глобус. Общематематический семинар / Под ред. М. А. Цфасмана и В. В. Прасолова. — М.: МЦНМО, 2004— . — ISBN 978-5-94057-064-6.

Вып. 5. — 2011. — 176 с. — ISBN 978-5-94057-847-5.

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

Пятый выпуск включает доклады В. В. Батырева, О. Я. Виро, А. А. Глуцкока, В. Ю. Калошина, Г. Кошевого, Ю. И. Манина, А. Н. Скоробогатова, А. Тоома.

УДК 51(06), ББК 22.1я5

ISBN 978-5-94057-064-6
ISBN 978-5-94057-847-5 (Вып. 5)

© НМУ, 2011
© МЦНМО, 2011.

Ю. И. М а н и н

ГЕОРГ КАНТОР И XX ВЕК

Введение

За тему доклада я должен извиниться. Я раньше всегда рассказывал на лекциях какую-нибудь математику, и решил представить более неформальный доклад не потому, что мне нечего доказывать, а потому что на лекции никогда не успеваешь доказать теорему, которую хочешь. И в какой-то момент мне это надоело, и я решил рассказать что-нибудь, что заведомо можно успеть, потому что можно начать в любом месте, прекратить в любом месте и т. д.

История лекций такого жанра начинается для меня с того, что периодически просят рассказать что-нибудь for general audience, как это называется по-английски, т. е. для людей, которые не являются профессиональными математиками; и почему они приходят на эту лекцию — неизвестно. For general audience говорить чрезвычайно трудно, и всегда начинается мучительный выбор темы.

У меня есть излюбленная тема, которую я придумал, и даже знаю, как про нее прочитать лекцию, но никогда этого не делал. Называется она «Почему дважды два — всегда четыре». В этом сюжете очень важное слово — это «всегда». Почему дважды два — четыре, можно продемонстрировать хоть на столе: взять два кусочка мела и еще два. Но вот почему всегда четыре?

Развиваться эта тема должна так. Раз «всегда», значит, речь идет о том, что тут запрятан какой-то закон сохранения. Закон сохранения — это идея, конечно, не математическая; это идея физическая. Простейший вид закона сохранения — это нечто, что аддитивно по областям пространства-времени. Значит, надо уже говорить не о пространстве, а о пространстве-времени, потому что слово «всегда» присутствует. Значит, у вас имеется 4-мерное (а может, какое-нибудь еще) многообразие — пространство-время; вы воображаете, что его можно разбивать на области, в каждой из этих областей что-то такое содержится; и вот, когда вы прибавляете

одну область к другой, то, что там содержится, должно прибавляться. Вы узнаете лагранжиан? Если есть аддитивность, значит, есть лагранжиан, действие и т. д.

Таким образом от сюжета «а почему дважды два всегда четыре» вы подымаетесь довольно быстро наверх и попадаете в область гораздо более сложную, чем первоначально рассчитывали. Обычно, дойдя до лагранжиана, я в воображении сам себя останавливал и решал, что, пожалуй, для general audience дальше будет трудно продолжать.

Но это свойство всех сюжетов, когда вы пытаетесь рассказать о чем-нибудь элементарном. Потому что такое «элементарное» по определению? Грубо говоря, вы что-то такое разлагаете на самые простые вещи. И когда вы доходите до нижнего уровня, когда вы решили, что вот это — самая простая вещь, и вы о ней что-то хотите рассказать, вы просто обнаружите, что вам некуда идти от элементарного, кроме как дальше вверх; т. е. элементарное объясняется посредством сложного. В этом смысле слова проще объяснять сложное, потому что тогда вы создаете искусственное впечатление, что вы какие-то вещи проясняете; объяснять элементарные вещи гораздо труднее.

Дойдя до этой глубокой мысли, я вернулся к сюжету «почему дважды два всегда четыре» и спросил себя: а что такое — два? Когда я спросил себя, что такое два, тут я вспомнил, что, конечно, Георг Кантор дал определение. И вот, развивая этот сюжет, я попытался представить себе, что можно сказать о Канторе сейчас, отдать ему некоторый *hommage*. В частности, потому что в моем личном развитии как математика Кантор и теория множеств сыграли совершенно особую роль.

По странной прихоти судьбы теория множеств, аксиома Цермело, континуум-проблема и всякие такие вещи были первым, что я всерьез выучил или постарался выучить, когда был в седьмом классе средней школы. Точнее сказать, так: первым был учебник Выгодского по анализу, где я научился дифференцировать, совершенно не понимая, что я делаю, но как-то научился. А потом, когда об этом узнали знакомые моей мамы, меня познакомили с преподавателем Крымского пединститута Яковом Лазаревичем Крейниным, которому я очень благодарен. А он был учеником, в свою очередь, Петра Сергеевича Новикова, и поэтому единственная математика, которую он хорошо знал, была теоретико-множественная математика. И он меня сразу научил абстрактной теории множеств, в частности, аксиоме Цермело и т. д. Воспоминание, которое у меня от этого осталось, — полной непонятности. Когда я пытался себе вообразить разницу между интервалом $[0, 1]$ с двумя концами и интервалом $(0, 1)$ без этих концов, у меня начинала кружиться голова. Я не знаю, можете ли

вы воспроизвести это состояние; я, странным образом, могу. Когда я об этом не думаю, то, конечно, никаких проблем нет, но когда я начинаю думать, у меня начинает кружиться голова. И это было очень сильным эмоциональным впечатлением.

Конечно, в университете теория множеств воспринималась как некий общематематический фон. Никакого возвращения к изначальной канторовской аксиоматике не было. В это время я не пытался представить себе, как мог думать Кантор, как его идеи могли восприниматься в математике, вообще, какое влияние всё это оказало на последующую математику, — в университетские, в молодые годы я об этом не задумывался совершенно.

Второй раз я к этому вернулся, когда писал свой курс математической логики. Тогда я пытался сформулировать для себя просто, вообще, свое отношение к какой бы то ни было проблематике оснований математики. Тогда решение, которое я для себя сформулировал, было резюмировано в первых строчках предисловия.

Я написал, что основания математики изучают математику, а не предписывают ей что бы то ни было. В той же мере, как биология изучает жизнь, а не предписывает жизни что бы то ни было.

Это был такой второй этап, после чего я опять надолго забыл о логике, основаниях математики и теории множеств.

В третий раз я вернулся сейчас к этому и написал некоторый этюд, который желающие могут найти в Интернете. А лекцию с разными вариациями на тему этого этюда я вам сейчас прочту.

Так что это была, как говорили Стругацкие, прамбула, а сейчас будет амбула.

Что такое «два»

Вернемся к вопросу: что такое два? Как известно, канторовское определение начинается с идеи множества, которая у всех у вас в головах. Поэтому вместо идеи, я думаю, лучше привести его определение, но по-немецки — ни на каком другом языке нельзя воспроизвести музыкальную структуру фразы этого определения, совершенно замечательную. Разрешите мне прочесть на моем плохом немецком языке:

«unter eine Menge verstehen wir jede Zusammenfassung M (под множеством мы понимаем любое собрание M) von bestimmten wohlunterschiedenen Objekten in unserer Anschauung oder unseres Denken (welche die „Elementen“ von M genannt werden) (вполне определенных и различных объектов нашего воображения или нашей мысли, которые мы называем элементами M) zu einem Ganzen (в единое целое)».

Вот это слово *Zusammenfassung*, которое, как скобка открывающая, стоит в начале определения, и *zu einem Ganzen* (единое целое), которое, как закрывающая скобка, стоит в конце определения, а посередине стоит, что именно объединяют, совершенно замечательно имитирует структуру мысли, структуру образования этого единого целого. И ни на одном языке такой синтаксис не позволителен — он просто не читается — кроме немецкого.

Давайте, кроме множества, напишем на современном языке морфизм множеств — отображение. Произвольное отображение — значит, мы объединяем некоторые пары элементов из X и Y снова в одно множество, так что для каждого x существует единственное y , которое является его значением; так что морфизм множеств также определяется в терминах множества и объединения.

Затем есть изоморфизм в виде биекции. Я применяю современный категорный язык, которым Кантор не пользовался.

Затем имеется определение конечного и бесконечного множества. Давайте начнем с бесконечного.

Бесконечное множество X : $\exists Y \subsetneq X$, для которого существует изоморфизм $Y \xrightarrow{(1,1)} X$.

Значит, множество, которое биективно своей собственной части, — это бесконечное. Конечное — всё, что остается; конечное — для которого нет такой биекции. В частности, пустое множество \emptyset конечно — потому что у него нет собственного подмножества.

И наконец, последнее: целое или натуральное число (в частности, двойка) — это класс изоморфизма в категории конечных множеств.

Теперь я могу дать определение двойки: 2 — это класс изоморфизма $\{\emptyset, \{\emptyset\}\}$.

Лучше всего, как это сочинил однажды фон Нейман, всё делать из пустых множеств. Тогда мы получаем пустое множество, одноэлементное множество, элементом которого является пустое множество, двухэлементное множество, элементами которого являются пустое множество и то одноэлементное множество, элементом которого является пустое множество.

Вот определение двойки по Кантору. Это, конечно, образец того, что я вам сказал: когда вы пытаетесь объяснить элементарные вещи, вам надо очень сильно усложнять их; иначе их просто некуда объяснять — мы дошли до низу, и деваться дальше некуда.

В о п р о с и з з а л а: «Юрий Иванович, а в выражении «дважды два» две двойки — это одна и та же двойка?» Очень хороший вопрос. Нет, совершенно не обязательно, потому что, если это — класс изомор-

физма двух разных, но биективных множеств, то, конечно, не обязательно. Как всегда, вы вычисления делаете на представителях, вы можете взять представителей двухэлементных множеств, которые далеко не совпадают; но как классы изоморфизма они совпадают.

На этом проблемы не кончились. Если еще немножечко подумать об этом и выйти за пределы математики, выяснится, что в естественном языке, в его жизни как бытового языка, как и, скажем, философского языка, а также и в математическом языке, эта первоначальная идея Кантора, что 2 — это 2 чего угодно, присутствует, но не всегда отдают люди себе в этом отчет. Поэтому давайте я проиллюстрирую, как это основное канторовское определение двойки на самом деле семантически представлено в выражениях разной бытовой и философской речи. Я начну с простейшей вещи:

у меня 2 уха

— это элементарная идея двойки: два предмета, которые можно показать пальцем и пересчитать.

Вот для математика уже немножко хуже:

седьмая значащая цифра в десятичной записи числа π есть двойка

— попробуйте быстро или даже не очень быстро сказать, что мы считаем, чего именно два фигурирует в таком определении? Совсем не легко, и, если кто-нибудь из вас со школьниками возится, попробуйте это выяснить.

Еще хуже:

поверхность этой доски 2-мерна

— ну, разумеется, при разумной физической идеализации. Если хотите, пространство 3-мерно — это еще будет хуже — с тройкой.

И то, что мне больше всего нравится (при моей тоже с детства любви к всяким лингвистическим, языковым явлениям):

I doubt it

— английская фраза; вот это «doubt» — это двойка, это индоевропейская двойка. И она повторяется в немецком «zweifeln», также и во французском «douter»; в немецком эта двойка видна, во французском «douter» так же видна, как в английском. Это идея сомнения. Почему «два» фигурирует, когда вы обдумываете идею сомнения? Ну, конечно, вы воображаете себе как минимум две альтернативы — так или эдак. Но все-таки, как это сказать, чего там два? Философский дуализм — это то же самое, но в более

рафинированном и терминологизированном контексте, а вообще — то же самое.

Таким образом просто факты языка позволяют очень мило проиллюстрировать первоначальное канторовское определение двойки, пропустив все эти сложные промежуточные стадии. Но они нужны, потому что вы хотите определить любое целое число, не только двойку.

Язык очень быстро кончается. Вы все знаете легендарные утверждения о языках, скажем, аборигенов Австралии, о том, что счет идет «один, два, три, много». В американском языке существует слово «zillion». Оно не означает ничего, кроме того, что это где-то за пределами того, что мы можем или хотим сказать словами. В конце концов, после двух-трех это идет или после 1984 — не так существенно. Естественный язык не может выразить много чисел; начиная с некоторого момента все выражения естественного языка для чисел, которые мы имеем, являются названиями чисел, записанных в десятичной, скажем, системе счисления, а не именами чисел как таковых.

Теория множеств

Кантор сам произвел то, что я бы назвал минималистской математикой в том смысле слова, в каком этот термин используется в искусствоведении современном — минимализм музыкальный, минимализм в живописи. Он обнаружил настолько ограниченный список фундаментальных идей, пользуясь которым можно делать необыкновенно глубокую математику, что это поразило и его, и его современников.

Позвольте мне напомнить математику, которую он развил исходя из этого списка основных идей. Опять я буду пользоваться модернизированным современным категорным языком вместо того, чтобы как-то воспроизводить его ход мыслей.

У нас имеются две вещи: категория множеств и объекты этой категории с точностью до изоморфизма. Начальный отрезок этих объектов с точностью до изоморфизма — это целые числа. Следующая канторовская идея была тогда, естественно, — а что дальше? Может быть дальше просто ничего нет? Это бы означало, что конечные множества имеют 1, 2, 3, 4, 5 и т. д. элементов, а бесконечное множество одно, и у него элементов столько, сколько целых чисел.

Конечно, мы знаем массу других бесконечных множеств. В то время самое главное да и сейчас, вероятно, самое главное было множество всех вещественных чисел. И что одних гораздо больше, чем других, не было известно до Рождества 1873 года.

29 ноября Кантор написал Дедекинду письмо, в котором в явном виде спросил, может ли быть биекция между \mathbb{N} и \mathbb{R} — между натуральными числами и вещественными. До этого, по-видимому, не было таких слов. Кажется, он тоже не пользовался словом «биекция»; я не знаю, как он называл взаимно однозначное соответствие по-немецки. До этого невозможно было задать такой вопрос, по-видимому. И это чрезвычайно интересно: мы знаем точную дату, когда впервые оказалось возможным задать такой вопрос. Этот вопрос был задан Кантором в письме Дедекинду 29 ноября 1873 г.

А после Рождества (через несколько дней после Рождества, у меня точной даты, к сожалению, нет) Кантор нашел доказательство того, что это нельзя сделать. И опять написал Дедекинду: у меня есть доказательство, но я не могу ему поверить.

Доказательство Кантора состояло, конечно, в его знаменитом диагональном процессе. Диагональный процесс существует в разных вариантах: можно воображать себе вещественные числа, записанные, скажем, в бинарных разложениях бесконечным списком, а потом в диагональном разложении изменить каждый бит на противоположный, как это обычно делается. Но это быстро обобщается и принимает гораздо более приятную форму в виде общего неравенства $2^x > x$ (строго больше!), которое для целых чисел, конечно, вполне очевидно, хотя, если подумать, тут тоже не одно доказательство.

В общем, смысл этого такой: если у вас имеется множество X , то $\text{card } 2^X > \text{card } X$. И доказывается это обычным диагональным процессом: мы воображаем себе, что мы можем установить биекцию $x \xrightarrow{(1,1)} U_x$, поставив в соответствие каждому $x \in X$ подмножество $E_x \subseteq X$. А потом доказываем, что это предположение приводит к противоречию, потому что есть очень специальное подмножество V , которое по определению состоит из тех x , которые не содержатся в подмножестве $u(x)$, которое соответствует x . Следующее множество V не имеет форму $u(z)$ ни для какого z : $V = \{y \mid y \notin U_y\} \neq U_z$.

Это обычная форма современной диагональной процедуры родственная, конечно, старинному парадоксу лжеца. Это тоже совершенно очаровательный сюжет о том, как странные словесные конструкции, очень часто восходящие к грекам, могут превратиться в точные глубокие математические теоремы, которые после Гёделя мы все однозначно интерпретируем как ограниченные возможности языка или, вообще, любого финитарного образования передать свойства бесконечности.

Итак, $2^X > X$. И из математики Кантора (если не говорить о, так сказать, предматематике, к которой я вернусь позже) самым знаменитым

стал, грубо говоря, вопрос о том, на сколько 2^X больше, чем X . Есть ли между 2^X и X , когда X счетно, что-нибудь еще? Когда X — это счетное множество, 2^X — это множество всех подмножеств счетного множества (оно имеет мощность континуума — мощность множества \mathbb{R} всех действительных чисел). И вопрос был такой: совпадает ли это с наименьшей несчетной мощностью \aleph_1 . Вот гипотеза континуума: Кантор полагал, что мощность континуума равна \aleph_1 . Сейчас точка зрения большинства людей, которые глубоко и профессионально думали об этом, состоит в том, что на самом деле мощность континуума гораздо больше.

Это вопрос является математическим или долго казался математическим, и в наших воспоминаниях, в нашем образе математики Георга Кантора в XX веке, вероятно, эта проблема — континуум-гипотеза, ее судьба, доказательство ее независимости, доказательство ее непротиворечивости, сомнения в том, что, вообще, этот вопрос имеет смысл, занимает, вероятно, основное место.

Я хочу поговорить не о судьбе этого вопроса, а о судьбе минималистской математики, которая воплощена скорее в канторовских определениях, в канторовском способе мысли, чем в канторовской задаче. Но, пожалуй, я этому посвящу вторую половину лекции, а сейчас немножко поговорю о континуум-гипотезе.

***P/NP*-проблема**

Я хочу коснуться двух сюжетов. Один состоит в том, что этот вопрос всплыл для конечных множеств довольно неожиданным образом для математиков — как *P/NP*-проблема.

P/NP-проблема совсем родственна не континуум-гипотезе, а проблеме, которая связана с аксиомой Цермело. Аксиома Цермело, как известно, постулирует возможность некоторой общей конструкции теории множеств. А именно: у вас есть множество X , есть множество всех его подмножеств 2^X , и аксиома Цермело постулирует, что вы можете выбрать по элементу в каждом подмножестве, непустом, разумеется. Больше ничего — просто выбрать по элементу в каждом подмножестве.

Предположим, что X конечно. Есть ли в этом какая-нибудь математика? Есть, если вы начинаете эти конечные объекты кодировать и спрашивать, можете ли вы все эти операции делать за полиномиальное время. Я утверждаю, что проблема *P/NP* — это в точности проблема аксиомы Цермело в финитном мире.

Давайте я чуть-чуть поговорю об этом, потому что это — сюжет занятный и потому что, предвкушая разные другие события, я хочу сразу

сказать, что он имеет отношение, по-видимому, к психологии или нейро-биологии. То обстоятельство, что теория множеств затрагивает какие-то глубокие психологические архетипы, было понятно с самого начала. В обширной переписке разных людей, которые одобряли или не одобряли теорию множеств, — Лебег, Э. Борель — многие люди высказывались в таком роде: вот я применяю аксиому Цермело, я должен думать о разных множествах, а откуда я знаю, что я в каждый момент думаю об одном и том же? То есть возникало смутное беспокойство о том, способны ли мы сохранить во время теоретико-множественных рассуждений ту ясность ума, которая постоянно связывалась с занятиями математикой и которая для Декарта, например, была просто основным критерием обоснованности научного или математического рассуждения — ясность ума в процессе этого рассуждения.

Так же, как, скажем (только в обратном направлении), для Поллока (известного американского абстрактного экспрессиониста), когда он делал свои холсты, выдавливая на них из тубиков краски или капая на них кистью, окунутой в ведро краски, — для него основным критерием того, сделал он вещь или нет, было — сумел он себя ввести в транс во время этого процесса или он сохранял ясность сознания; если он сохранял ясность сознания, значит, вещь не удалась. Для математика это прямо наоборот.

Давайте немного поговорим о проблеме, на сколько 2^X больше, чем X , на примере P/NP . Это обычная вещь: мы выбираем U_m — это будет наше первоначальное множество; там сколько-то позиций — m позиций. Потом, конечно, 2^m представлено тем, что вы на этих позициях можете ставить нули и единицы: множество всех последовательностей нулей и единиц, стоящих на этих позициях, это 2^m . Во всех разумных смыслах это гораздо больше, чем m , экспоненциально больше.

Мы хотим рассматривать подмножества U_m . И элементарная теорема состоит в том, что подмножества U_m находятся в биекции с элементами соответствующей булевой алгебры:

$$U_m \supset X \xrightarrow{(1,1)} \{f \mid f \in \mathbb{Z}_2[x_1, \dots, x_m]/(x_i^2 + x_i)\}.$$

Булева алгебра для меня (я буду думать в терминах коммутативной алгебры — во-первых, я к этому привык, а во-вторых, это гораздо удобнее) — это кольцо полиномов от m переменных над полем из двух элементов, отфакторизованное по идеалу, порожденному $x_i^2 + x_i$. Люди с алгебро-геометрическими взглядами знают, что я просто беру точку m -мерного аффинного пространства инвариантную относительно эндоморфизма Фробениуса Φ (который, конечно, действует на множестве таких последова-

тельности). А соответствие такое: $f \mapsto \{x \mid f(x) = 0\}$ — функции отвечает множество ее нулей. Есть такая элементарная и красивая теорема: вы получаете таким образом биекцию; теорема, которую я тоже рекомендую для школьников и студентов — очень приятное упражнение.

И тогда вопрос Цермело в этом контексте выглядит так: по данному f найти элемент x_0 , для которого $f(x_0) = 0$, или доказать, что его нет, а это значит — доказать, что $f \equiv 1$.

Решить эту задачу нетрудно перебором. Считайте значения f , пока не доберетесь до нуля; ни разу не доберетесь — значит, функция тождественно равна 1. Но, конечно, в P/NP -проблеме мы хотим сделать это за полиномиальное время. Тогда вопрос: за полиномиальное от чего? И тогда задача состоит в том, чтобы выбрать какое-то кодирование f (например, мы можем f задавать просто списком его коэффициентов).

Мы должны выбрать кодирование f какими-то комбинаторными объектами, у которых есть естественная величина в битах. И потом поставить вопрос о существовании алгоритма нахождения x_0 , для которого $f(x_0) = 0$, за время, полиномиальное от битовой величины кода f .

Очень красивая математика, связанная с этой задачей, содержит два утверждения, которые являются частью некоторой теории — главы вычислительной математики. Я скажу об этом только две вещи.

Ответ на первый вопрос выбирается такой: коды f — это по определению такие объекты: $u = \{N; (S_1, T_1), \dots, (S_N, T_N)\}$, где $S_i, T_i \subset \{1, \dots, m\}$ — выбирается целое число N , и потом выбирается N пар подмножеств в множестве индексов $\{1, \dots, m\}$. Каждый такой выбор числа N и N пар является кодом одного из f : если такое u задано, то

$$f_u = 1 + \prod_{i=1}^N \left(1 + \prod_{k \in S_i} (1 + x_k) \prod_{j \in T_i} x_j \right).$$

Битовая величина этого выражения равна mN .

Приятное упражнение состоит в том, что любое f можно записать в такой форме, хотя и не однозначно — это цена за то, что я сейчас скажу (в третьем пункте).

Третий пункт состоит в том, что, когда вы выбираете такое кодирование, ставите задачу таким образом, вы получаете универсальную P/NP -задачу. То есть, если вы сможете доказать, что для этой задачи есть решение за полиномиальное время, то для всех P/NP -задач оно есть; а если для этой нет, то, стало быть, P не совпадает с NP .

Таково было замечательное открытие, сделанное в вычислительной математике XX века. И прелесть его состоит для меня в том, что это

в некотором роде возвращение к минималистской математике Кантора — именно математике, а не способу ее организации, о котором я буду говорить позже. Эта финитарная версия P/NP -проблемы, как я написал у себя в этом самом тексте о Канторе — это континуум-гипотеза для бедных. Не намекая, что за ее решение назначен приз миллион долларов.

Вот это такая замечательная трансформация, травестирование гипотезы континуума в конце нашего века. До этого я говорил про аксиому выбора, но гипотеза континуума (есть ли что-нибудь между m и 2^m) в финитарном мире имеет очевидное решение: конечно, да, и много чего. Поэтому я переформулировал гипотезу: насколько 2^m больше m ? И ближе всего та переформулировка, которая связана с гипотезой Цермело. Гипотеза Цермело очень близка к тому способу, каким Кантор думал.

Вообще, эта дистанция, пройденная в математике абстрактных множеств XIX и XX века очень похожа на дистанцию, пройденную в политических системах от, скажем, гибели аристократических режимов до рождения демократических режимов. В некотором роде P/NP — это очень демократическая версия канторовской задачи теории множеств.

Кантор сам думал о своем открытии, о том, что количество разных бесконечностей бесконечно, как непосредственно боговнушенном ему. Он был мистиком или, во всяком случае, был очень близок к мистицизму. Он общался с католической церковью на предмет какой-то легитимизации своих изысканий. Он прямо и непосредственно считал, что его работа является боговдохновенной; и легитимизация требовалась для этого ощущения: является ли это действительно боговдохновенным, как он полагал.

И травестия бесконечности бесконечностей в компьютерной технике мне кажется очень занятным и символическим обстоятельством.

Биография Кантора

Я хочу начать второй час с очень краткого биографического очерка жизни Кантора.

Он родился 3 марта 1845 года в Санкт-Петербурге, в России, в немецком семействе, которое переехало в Висбаден в 56 году, когда ему было 11 лет. С этим обстоятельством связаны его попытки в начале уже XX века переехать в Россию и получить в России дипломатическую службу, когда он чувствовал себя жестоко обиженным немецким математическим истеблишментом того времени (попытки эти не оказались удачными).

Затем он до 1867 года, т. е. до 22 лет учится, в частности, в Цюрихе, Берлине, Гёттингене и затем снова в Берлине. Затем у него идут первые

публикации: теория чисел, квадратичные формы. В университете Галле в 1869 году, стало быть, в 24 года у него происходит габилизация. Точно не знаю, но по нынешним меркам в Германии это примерно как наша докторская диссертация; если это так, то он очень рано это сделал. Затем идет серия 3—4 года: работы по сходимости тригонометрических рядов. Так что он начал заниматься абстрактными множествами, воображая себе разные изысканные подмножества вещественных чисел, связанные с теорией тригонометрических рядов, т. е. с анализом, по существу.

И затем он начинает размышлять об абстрактных множествах. Обнаруживает в конце следующего периода, в 1879 году, с одной стороны, существование разных порядков бесконечности: неравномощность целых чисел и вещественных чисел. И напротив того, совершенно контринтуитивный результат: существование биекции между евклидовыми пространствами любых размерностей — в то же самое время. Сейчас мы к этому привыкли, но это было очень странно. В 1874 году происходит первая публикация теории множеств, ему, стало быть, 29 лет. Затем следующие 5 лет — серия статей: бесконечные линейные множества чисел. 1883 год, ему 38 лет — появление книги «Основы общей теории множеств».

И май 1884 года — первый приступ нервной болезни после успешного и очень приятного путешествия в Париж. У него произошло то, что бы сейчас назвали приступом депрессии — с весны по осень, с мая 1884 года.

После чего в 1885 году он ведет переписку с католическими теологами. Но в математическом мире как будто бы остается всё более и более изолированным. В 1885 году был такой конфликт с Миттаг-Леффлером, что он чувствовал, что его математические идеи в математическом мире просто не принимают, причем не принимает группа самых крупных математиков этого времени.

Он делает замечательную вещь (считаю это очень в духе демократического XX века), а именно, он основывает немецкое математическое общество — 18 сентября 1890 года. Чтобы получить, как сказали бы мы в наш политический век, какую-то такую более массовую базу и иметь возможность влияния на аморфную массу математиков более низкого уровня, чем те, которые занимали лучшие кафедры в европейских городах; чтобы иметь вес. Это было замечательное решение.

В частности, как известно, была очень сильная оппозиция Кронекера, сказавшего: «целые числа создал Бог» — в противоречие с мнением Кантора. Но Кронекер также назвал (знаменитая его фраза) Кантора «совратителем юности». И эта фраза содержит тонкое самоотрицание: как

известно, Сократ был осужден как совратитель юности. Поэтому неявное сравнение с Сократом — это довольно тонкий комплимент, но спрятанный таким вот выражением.

Затем умирает Кронекер — его крупнейший научный враг. И, напротив того, Немецкое Математическое Общество цветет. И Кантор же приложил массу усилий, теперь уже как глава этого общества, к основанию первых международных математических конгрессов.

В 1897 году происходит I Международный Математический Конгресс, на котором теория множеств действительно очень видна, и в каком-то смысле получает штамп общественного интереса — вместо того, чтобы быть отодвинутой на периферию как сочинение нездорового ума. До такой степени этот процесс происходит быстро, что на следующем математическом съезде в знаменитом списке гильбертовских проблем континуум-проблема становится задачей номер 1. Это высший момент, просто триумф Кантора, которого он добился странным образом — такой вот общественной деятельностью, повторяю, чрезвычайно созвучной XX веку, а не тому аристократическому веку, когда математикой управляли крупнейшие умы, а что такое математическое общество, было делом неизвестным и неинтересным.

Конечно, происходят и противоположные процессы, и с началом XX века в каком-то смысле дела идут под горку. Даже Гильберт, который сделал так много для того, чтобы теория множеств, в частности, континуум-гипотеза заняли почетное место в умах математиков, когда он позже высказался в том стиле, что никто не может выгнать нас из рая, созданного для нас Кантором, — этот комплимент, в противоположность кронекеровской фразе, содержал тонкую насмешку: это был намек на то, что Кантор полагает, что он боговдохновен, что его математика боговдохновенная, что, значит, из его рая нас никто не выгонит. Так что всё было очень неоднозначно.

В 1899 году у Кантора умирает сын Рудольф, и он дважды попадает в нервную клинику в Галле. 1902/03 год — опять зимой госпитализация; 1907/08 год — то же самое; 1911/12 год — то же самое.

В 1915 году в Галле собирались праздновать 70-летие Кантора на международном уровне. Но в момент подготовки началась I Мировая война. На международном уровне празднование было уже невозможно; его кое-как свернули на национальном уровне.

И после очередного периода госпитализации 6 января 1918 года, в момент очень глубокого тяжелого положения и Германии, и всей Европы, и всего западного мира, как мы знаем, Кантор умирает.

Такова примерно его биография.

Я не помню, в каком точно году, он подал заявление о принятии его в российское гражданство и на дипломатическую службу в России — до такой степени он чувствовал себя замученным неприятием теории множеств.

Кантор и XX век

Между тем, неприятие во многих отношениях было совершенно искренним. Можно, конечно, понять, что особо суровые выражения французских математиков были связаны с обычной германо-французской враждой и т. д.

Занятно, что даже эмоциональные чувства по поводу гипотезы континуума имеют тенденцию возрождаться. Я с некоторым изумлением прочел в очень замечательной статье Мамфорда в его сборнике «Математика XX века» описание малоизвестной статьи Фрелинга, которая приводит некоторые аргументы в пользу того, что гипотеза континуума очевидным образом ложна.

Аргумент занятный. Он сопровождается картинкой о бросании стрелок в цель. Нужно вообразить себе континуум, реализованный как множество точек доски; и вообразить себе континуум-гипотезу в цермеловской форме: что континуум можно вполне упорядочить таким образом, чтобы любой начальный отрезок был счетным (для каждого p то, что меньше p при этом упорядочении, чтоб было счетным). Дальше он предлагает себе вообразить исход этого эксперимента: вот сначала бросает один игрок, потом другой; скажем, один попадет в точку p_1 , другой — в точку p_2 . Посчитаем вероятность того, что $p_2 < p_1$. Вероятность равна 0, потому что только счетное число точек меньше, чем p_1 . Но, поскольку всё равно, кто первый бросает, кто — второй, то в таком же рассуждении вероятность того, что $p_1 < p_2$ тоже равна 0. А так как это исчерпывающие события, то говорить не о чем.

И Мамфорд дальше пишет очень сильную фразу о том, что он надеется, что такое рассуждение избавит, наконец, нас от всех этих глупых размышлений над бессмысленной конструкцией теории множеств. Что меня удивило в этой фразе — это искренность чувства. Это показывает, что психологические проблемы, связанные с этим канторовским открытием, склонны самовоспроизводиться. Не потому что так учили или почему-нибудь еще; Мамфорд — абсолютно замечательный математик, что нужно знать, он, конечно, знает. Но как-то нет успокоения, тем не менее. Вот это очень занятно.

Статья его называется «The Dawning of the Age of Stochasticity» (т. е. «Заря века случайности» или что-то в этом роде). И математическая

пропаганда, которая в ней содержится, состоит в том, что все основные понятия математики имеют вероятностный характер, и поэтому с самого начала, с оснований ее так и надо перестраивать, и дальше — тоже. И вся наша привычка к детерминированным вещам, к точным и т. д. — неправильная, и что, вообще, нужно иначе обо всем этом думать.

Я решил, что название статьи связано с известной историософской схемой Джамбаттиста Вико, которую я 30 лет назад вообще не принимал всерьез, а сейчас, наоборот, начинаю всерьез в нее верить. Очень грубо говоря, его историософская схема такая. В истории человечества сменяются периоды: был период варварства; потом был теократический период (грубо говоря, начала единобожия в разных вариантах и соответствующая схема общественного устройства), сменившийся аристократическим периодом (сильная монархия, герои и прочее), который сменяется демократическим периодом — предположительно мы жили в нем в XX веке; после чего, после некоторого периода хаоса, мы перейдем к очередному теократическому периоду. Почесавши себе голову после 11 сентября 2001 года, я решил, что очень похоже, что Джамбаттиста был прав, что он что-то угадал. Он писал в 1725 году. Я думаю, что он несомненно имел в виду в то время уже и Китай, и Индию, но вряд ли, скажем, Южную Америку и Африку. Это очень любопытно. Вообще, нужно прочесть. Я, к сожалению, не добрался до первичного источника и пересказываю его теорию со слов Блюма — книга «Западный канон». Я всё собираюсь добраться до настоящего текста Вико и даже хотел бы прочитать его по-итальянски, но что-то никак его не добуду.

И я написал Дэвиду, не намекает ли он на эту теорию Вико, и он мне ответил, что нет, он, собственно говоря, имел в виду название мюзикла на Бродвее, в котором его сын как-то участвовал. Я немножко огорчился, но и посмеялся тоже.

Во всяком случае, искренность чувства по отношению к континуум-проблеме меня поразила.

Кантор и Бурбаки

Я говорил до сих пор в основном о двух вещах: о математике, которую оставил нам Кантор, и о социологических и психологических обстоятельствах того, как он работал и как его идеи воспринимались — рецепции его идей при его жизни и после. Теперь я хочу сказать немного об этом круге идей, которые, грубо говоря, суммированы в списке основных канторовских определений, как мы его читаем сейчас, как его можно кратко переформулировать.

Как мне кажется, главным событием, которое является прямым наследником канторовской мысли, хотя это, вероятно, не всеми историками математики будет признано, является деятельность группы Бурбаки. Бурбаки были довольно глубоко безразличны математические проблемы бесконечности, безразличны до степени *negligeance*, до пренебрежения. Так что, когда нужно было написать том об основаниях математики, он не получился совершенно. Видно было, что это им совершенно неинтересно. А вот что им было интересно — это потенциал теории множеств как исходного минималистского материала для воображения; потенциал теории множеств для создания потенциально единого языка для всей математики, что бы это ни было — топология, теория вероятностей, алгебра, геометрия — что угодно. Потенциально единый язык. И с моей точки зрения высокая заслуга Бурбаки состояла в том, что они этого добились, по крайней мере, на некоторое время.

Они сформулировали язык структур, согласно которому любой математический объект есть, грубо говоря, множество и дополнительные данные, которые, опять же, в терминах теории множеств формулируются: там, какие-то элементы в лестнице (*echelle*), построенной из данного исходного множества. Например, группа — это множество плюс закон композиции, плюс такие-то свойства. И всё, что нужно сказать в определении математического объекта, всё формулируется в терминах множеств и элементарных операций над ними.

Свой вклад в эту программу внесли разные математики, далеко не одни Бурбаки: так, скажем, достижение Колмогорова — это чисто бурбакистский проект: показать, что теорию вероятностей можно сформулировать как теорию меры. Опять же, Синай очень сильно возражал мне на эту тему. Я не хочу сказать, что формулировка теории вероятностей как главы теории меры исчерпывает ее содержание, и тем более, не исчерпывает интуиции позади вероятностных и статистических рассуждений. Но несомненно, что сформулировать, не потеряв почти ничего в объекте, так можно.

Так вот, мне кажется, что прямым наследником этой части канторовских идей несомненно были Бурбаки, даже если они сами с этим, возможно, и не согласились бы. И что результатом этого было наблюдавшаяся нами, скажем, во второй половине XX века унификация общематематического языка, которая в принципе позволяла кому угодно — топологу читать и понимать определения и доказательства в статье по чему хотите. Но далеко не всегда так было. Нужно отдать себе отчет, что во времена, когда, предположим, Кантор работал, читать и понимать статьи по алгебраической геометрии итальянской школы мог только алгебраический геометр,

предпочтительно, итальянской школы, ну, может быть, немецкой кое-как, но уже больше никто. Сейчас это не так: в принципе, кто угодно может читать статьи на какие угодно темы. И это несомненно заслуга Бурбаки.

Опять же, мы присутствуем сейчас при разрушении этого порядка вещей, которое произошло из-за возобновившихся взаимодействий математиков с физиками. Это тоже чрезвычайно интересный процесс. И может быть, стоит немножечко сказать о том, опять же, как это виделось во времена Кантора.

Кантор отнюдь не был равнодушен к физике; но физика, которую он знал и которую он понимал, была чрезвычайно наивна. В результате, в его знаменитом определении множества зафиксировано, заморожено такое наивное представление о внешнем мире как о мире, состоящем из различных вещей. Конечно, необходимая доля абстракции там присутствует.

У Кантора было, по крайней мере, два известных мне рассуждения, где он в явном виде привлекает физический мир. Одно было замечательное: он полагал, что в мире должно быть счетное количество лейбницевских монад-атомов — материальных точек, по-видимому, массивных, и континуум безмассовых квантов — он их называл эфирными монадами; т. е. он полагал, что все эти вещи прямо есть и их можно считать. Второе его физическое рассуждение было такое: он считал, что одним из оснований для механики, для классической физики является наше представление о непрерывности, которое, в свою очередь, есть абстракция представления о непрерывном движении; и вот поэтому мы постулируем, что наше 3-мерное пространство, его точки параметризуются, скажем, тройками вещественных чисел или чем-то в этом роде. Он доказал красивую теорему, которая очень неожиданна для своего времени: если выкинуть из \mathbb{R}^3 счетное множество точек, то в оставшемся подмножестве любые две точки можно соединить непрерывной кривой. И стало быть, наше пространство вполне может пропускать какое-то всюду плотное множество, и при этом свойства непрерывности как-то сохраняются. Поэтому он ставил вопрос о том, не надо ли учесть эти новые математические факты при продумывании истинных основ механики.

Но дальше этого он не пошел. Он не дошел даже до такой простой и очевидной, скажем, для Лагранжа или для кого-то еще физической идеи о том, что математические множества наиболее существенные для физики никогда не суть множества вещей, они в лучшем случае суть множества возможностей. У нас есть конфигурационное пространство, (мы описываем систему конфигурационным пространством); одна точка конфигурационного пространства — это возможное положение или возможное состояние системы, но вовсе не то, что существует или даже реализуется: может

быть система никогда в этот угол своего конфигурационного пространства не залезет. Тем более, фазовое пространство (это значит: точка — это точка конфигурационного пространства плюс инфинитезимальный вектор, так сказать, импульса, касательный вектор), тем более ясно, что это только возможность, это не есть вещь. Когда при развитии классической физики она перешла в моду квантового описания, т. е. мы перешли от, скажем, фазового пространства, грубо говоря, к L^2 на конфигурационном пространстве, каждая точка которого, — ψ -функция — тем более очевидным образом является некоей возможностью квантового состояния системы, но отнюдь не вещью. С самого начала и во времена Кантора физикам было понятно, что если множества как-нибудь существенны для описания внешнего мира, то отнюдь не как множества вещей.

Чисто педагогически, кстати, у меня всегда вызывает возражение эта иллюстрация множеств на вещах, когда мы учим детей. Конечно, хорошо говорить о множествах стульев физически вот в этой аудитории. Но это ведет в неправильном направлении, мы не должны этого делать, мы должны очень рано уйти от этого. Множества — это то, что себе воображаешь, а не то, что есть и то, что можно пересчитывать.

Если бурбакистский проект, на мой взгляд, никак канторовскую программу не менял и не дезавуировал, а наоборот, развивал в очень позитивном смысле, то, напротив того, взаимодействия с физикой всегда уводили мышление математиков до такой степени в сторону от того, что они унаследовали от предыдущего этапа и как бы им хотелось думать. Это повторялось много раз. И для какого-то из докладов я составил такую параллельную таблицу основных вещей, которыми занимались физики, и основных вещей, которыми занимались математики в течение XX века, и получилось, что они почти не пересекались. Вот начало, скажем с 1880 г.: в математике — Кантор строит теорию множеств, Стилтес, Лебег, Борель исследуют, что такое интеграл; в это же время у физиков Максвелл занимается электричеством, магнетизмом и термодинамикой. Было видно, что, невзирая на то, что физическая интуиция оплодотворяла какую-то математику, а физики, наоборот, пользовались каким-то развитым аппаратом, было видно, что центральные вещи, которые их волнуют, почти никогда не совпадают; не только иногда. Удивительный результат такого исторического размышления. Против этого обычно возражают, когда я объясняю такие вещи на лекциях; но, во всяком случае, меня никто не смог переубедить. И мне кажется, что этот уход, потеря интереса ко всей канторовской проблематике, включающей теперь и бурбакистский подход к математике — он именно связан с этим возобновившимся периодом взаимодействия с физиками, когда идеи, которые приходят, являются очень нематематиче-

скими по существу, хотя невозможными для иного выражения, кроме как на математическом языке — такой вот замечательный парадокс. В конце концов, сейчас сотни математических работ так или иначе используют понятие фейнмановского интеграла. Между тем понятие фейнмановского интеграла сейчас определено не лучше, чем понятие интеграла Римана (я уж не говорю, Лебега) до того, как появилась статья Кеплера о стереометрии винных бочек — это было первое разумное математизированное определение интеграла. Тем не менее. Всем бурбакистским нормам это совершенно не удовлетворяет.

И я сейчас просто с каким-то расстройством недавно наткнулся на цитату из Андре Вейля, который в 1953, кажется, году на очередном математическом съезде с гордостью сказал: «Мы уже не спрашиваем, строго ли доказана теорема, мы спрашиваем, доказана ли она». То есть ему казалось, что к 1953 году вопрос о приемлемом уровне математических рассуждений уже наконец решен раз и навсегда. И конечно, подразумевалось, что благодаря деятельности Бурбаки. К счастью, мы видим сейчас, что этот уровень приемлемости опять вернулся в некоторое серое смутное состояние, в сумеречную область; и те, кто этим могут творчески пользоваться — они очень от этого счастливы. И я чувствую тоже некоторое освобождение от тяжелой обязанности находиться всё время в рамках бурбакистского объединения; тем не менее, которое я чрезвычайно ценю. И я считаю, что нужно сохранять разумный баланс между тем и другим.

Теория множеств и теория категорий

В оставшиеся 10 минут я хочу сказать еще несколько слов о том, что произошло после бурбакистской программы. Это связано с деятельностью человека, личность которого в некоторых отношениях похожа на личность Кантора, — Александра Гротендика.

Гротендик, как известно, заменил бурбакистский язык структурированных множеств на язык категорий. До сих пор достаточно строго и четко характер этой замены не был проанализирован, как-то нигде никаких текстов не написано. Поэтому давайте, очень кратенько скажу, что произошло. Были введены категории — объекты, морфизмы, как вы все хорошо знаете. Объекты перестали быть множествами, морфизмы остались множествами. И первые категории, которые рассматривались, были просто классами структурированных множеств, структурированными множествами, поэтому это не было уходом от Бурбаки.

Гомотопическая топология сделала некоторый прорыв. Основной объект гомотопической топологии — это не топологическое пространство,

а топологическое пространство с точностью до гомотопии. Топологические пространства гомотопически эквивалентные друг другу, представлены: точкой, вещественной прямой \mathbb{R} , плоскостью \mathbb{R}^2 , \mathbb{R}^3 , бесконечномерными клетками, и т. д. Поэтому уже нельзя сказать, что основной объект гомотопической топологии является структурированным множеством — это просто неверно. Вы не поймете даже толком, как строить классифицирующее пространство группы, если вы поверите, что вы должны работать с топологическими пространствами.

Второе: сама категория как математический объект. Сама категория первоначально выглядит как традиционная математическая структура. У меня имеется множество объектов, между ними множество морфизмов с частичным законом композиции. Но мы уже хорошо знаем, что категории нельзя рассматривать с точностью до изоморфизма; это бесполезное понятие. Например, есть категория, конечномерных линейных пространств над полем (каким хотите, над \mathbb{C} , например). Там есть важнейшая инволюция — звездочка (переход от пространства к сопряженному пространству — пространству линейных функционалов). Это должно было бы быть изоморфизмом категорий; но это изоморфизмом не является. Это является эквивалентностью категорий. Есть очень существенная разница между изоморфизмом и эквивалентностью категорий. И таким образом основной постулат Бурбаки о том, что объект есть структурированное множество с точностью до изоморфизма, тоже не верен; когда объектом становится категория, место изоморфизмов занимают эквивалентности. Это чрезвычайно радикальная инновация, которая продолжается в нынешнем распространении таких вещей, как, скажем, A_∞ -категории и т. д.

Мы начинаем понимать, что объекты являются математическими объектами, морфизмы между объектами являются математическими объектами, функторы являются математическими объектами. Но среди математических объектов нет равенств (за исключением тождественного — каждый объект равен самому себе), есть только эквивалентности. Эквивалентность тоже является математическим объектом. Поэтому между эквивалентностями нет равенств, кроме тривиального — каждая эквивалентность равна самой себе. Всё остальное является эквивалентностями между эквивалентностями. Вместо привычного для бурбакиста мира структурированных множеств (конкретных объектов) возникает бесконечно ступенчатый иерархический мир объектов с эквивалентностями, с эквивалентностями между эквивалентностями и т. д., и т. д. Этот мир чрезвычайно трудно зааксиоматизировать.

Толстенная рукопись Гротендика «В поисках стэков» (или даже лучше «В погоне за стэками») посвящена фиксации его каждодневных попыток

(в его уже более поздней стадии, когда Гротендик уже был психически нездоров), как-то математически записать это новое видение математического мира, которое заменяет исходный канторовский мир. Это ему не удалось. Не вполне понятно, может ли такой проект быть действительно хоть на время завершенным или он представляет собой только движение к тому, как мы станем думать о математике в ближайшие годы. Но если канторовские идеи были как-то превзойдены в XX веке, то, как мне кажется, именно этим новым категорным видением того, как математика структурирована или должна быть структурирована в послеканторовское время. И как ни странно, именно эти вещи используются в текущем периоде взаимодействия с математической физикой, чего, в общем, никто кроме самого Гротендика не ожидал; а Гротендик ожидал исключительно из гордыни, потому что физики он тоже не знал и не понимал. Но он был уверен, что его способ видеть, скажем, математическое пространство, то, что раньше себе мы воображали, как многообразие или что-нибудь в этом роде, видеть его, как сайт или как топос — он был уверен, что это доставляет новые способы формулирования нашего понимания физического мира; и он оказался прав.

Я бы хотел закончить тем, что и Кантор оказался прав после 150 лет развития его теории.

28 августа 2003 г.

В. В. Б а т ы р е в

ТОРИЧЕСКИЕ ВЫЧЕТЫ В ЗЕРКАЛЬНОЙ СИММЕТРИИ

Я начну свой доклад немножко с истории, с которой связана зеркальная симметрия, чтоб потом уже перейти к более специальным вещам. Речь идет о том, что есть так называемая теория струн. И один из вариантов понимания теории струн приводит к пространству-времени размерности 10. Это пространство с метрикой неевклидовой сигнатуры, скажем, 9 раз плюс, 1 раз минус: $(\underbrace{++ \dots +}_9 -)$. Причем размерность 10 на самом деле

раскидывается следующим образом: одна размерность относится к обычному времени, а другие, пространственные размерности (их осталось 9) делятся на $3 + 6$, причем 3 — это естественно наблюдаемые размерности в нашей природе, а 6 — так называемые скрытые размерности, которые содержат в себе некоторую дополнительную геометрию. И по некоторым соображениям эти скрытые размерности связаны с геометрическим объектом, который называется многообразием Калаби—Яу. В данном случае это будет некоторое 3-мерное комплексное проективное многообразие Калаби—Яу.

Так вот, зеркальная симметрия как раз относится к геометрии этого многообразия. Физиками было обнаружено, что она касается непосредственно геометрии этих многообразий, которая отражается в их когомологиях. Как математик, я должен хотя бы формально определить, что такое многообразие Калаби—Яу.

Меня интересуют многообразия Калаби—Яу, в принципе, любой размерности — лучше давать определение в любой размерности. Определение такое: X — проективное алгебраическое многообразие над \mathbb{C} называется многообразием Калаби—Яу, если выполнены два условия. Первое условие состоит в том, что $K_X = 0$, т. е. канонический класс многообразия X должен быть тривиальным. Скажем так: канонический дивизор многообразия X равен 0 в самом сильном смысле, т. е. относительно рациональной эквивалентности. Для второго условия есть разные возможности. Иногда добавляют условие, что фундаментальная группа тривиальна: $\pi_1(X) = 0$; иногда добавляют некоторое гомологическое условие, например, рассмат-

ривают когомологии структурного пучка этого многообразия и требуют обнуления этих когомологий для всех i между 0 и размерностью (комплексной размерностью): $h^i(\mathcal{O}_X) = 0$ для $0 < i < \dim_{\mathbb{C}} X$.

Даже если здесь присутствуют те, кто не знаком с продвинутой алгебраической геометрией, я просто на всякий случай приведу пару примеров, чтобы независимо от этого формального определения была хоть какая-то геометрическая зацепка к понятиям, о которых идет речь. Приведенное выше определение работает в любой размерности: $d = \dim_{\mathbb{C}} X$. И в малых размерностях мы видим некоторую картину устройства многообразий Калаби—Яу. В любой размерности d примером многообразия Калаби—Яу является гладкая гиперповерхность степени $d + 2$ в проективном пространстве \mathbb{P}^{d+1} , которая задается однородным многочленом степени $d + 2$. Так получается пример многообразия Калаби—Яу в любой размерности, поскольку d можно взять любое. В частности, при $d = 1$ получаются обычные кубики на проективной плоскости \mathbb{P}^2 . Кубика — это эллиптическая кривая \mathbb{C}/Λ , топологически это просто тор. В размерности 2 это кватрика, так называемая поверхность типа КЗ. К сожалению, уже здесь, поскольку мы работаем с вещественной размерностью 4, я не могу нарисовать никакой картинку. А самый первый случай, где становится всё гораздо более интересно — это квинтика в \mathbb{P}^4 (3-мерная квинтика); это простейший пример 3-мерного многообразия Калаби—Яу. В принципе, это одно из многообразий, которое с точки зрения физики является довольно интересным.

Я уже сказал, что топологию многообразия Калаби—Яу очень трудно себе представить, потому что мы работаем в вещественных размерностях > 3 , даже уже в случае КЗ-поверхностей это не просто, а тем более, в случае квинтики — это 6-мерное вещественное многообразие. Но некоторую информацию об их топологии можно все-таки получить, рассматривая их когомологические инварианты, например, числа Бетти и числа Ходжа. Я напому, что для любого комплексного многообразия можно определить его когомологии де Рама $H_{\text{DR}}^*(X; \mathbb{C})$ через комплекс де Рама комплексных дифференциальных форм. Более того, каждая группа когомологий де Рама еще дальше может быть разложена по типам (p, q) : $H_{\text{DR}}^i(X; \mathbb{C}) = \bigoplus_{p+q=i} H_{\text{DR}}^{p,q}(X; \mathbb{C})$; где пространства $H^{p,q}$ которые определя-

ются типами соответствующих дифференциальных форм. И размерности этих пространств (размерность всего пространства $H_{\text{DR}}^i(X; \mathbb{C})$ — i -е число Бетти, или их составляющих, которые называются числа Ходжа) являются топологическими инвариантами, которые при деформации комплексной структуры не меняются. Число Бетти — это $b_i(X) = \dim_{\mathbb{C}} H_{\text{DR}}^i(X; \mathbb{C})$. А чис-

ло Ходжа это $h^{p,q}(X) = \dim_{\mathbb{C}} H_{DR}^{p,q}(X; \mathbb{C})$ — здесь мы имеем 2 параметра p и q , поскольку мы работаем с комплексными многообразиями.

Моя цель состоит в том, чтобы использовать эти топологические инварианты для изучения многообразий Калаби—Яу. Я хочу посмотреть, что можно сказать об этих инвариантах в случае многообразий Калаби—Яу. Некоторую интересную информацию о топологии многообразий Калаби—Яу можно из них получить уже в размерности 2 и 3. Давайте я просто нарисую табличку. Обычно информация о числах Ходжа и числах Бетти записывается в виде таблички или диаграммы, которую также называют «кристаллом Ходжа». Я буду его рисовать в виде квадрата в координатах p, q . В случае размерности $d = 1$ картинка будет выглядеть таким образом:

1	1
1	1

— здесь все размерности по единичке; вот такой получается кристалл Ходжа. При $d = 2$ у меня есть квартика в \mathbb{P}^3 :

1	0	1
0	20	0
1	0	1

— на границе таблички стоят единички и нули, а в центре будет 20. Здесь имеет место один важный топологический факт. Я уже сказал, что примерами КЗ-поверхностей являются двумерные квартики, но таковыми являются не только квартики, есть много других семейств поверхностей КЗ. И важный топологический факт состоит в том, что топология любой КЗ-поверхности на самом деле будет всегда одна и та же. Поэтому на диаграмме Ходжа все числа Ходжа будут всегда одни и те же; здесь мы не имеем никакой топологической свободы. Она появляется начиная лишь с комплексной размерности 3. В размерности 3 мы имеем квадрат еще большего размера:

1	0	0	1
0	b	a	0
0	a	b	0
1	0	0	1

Одна часть этого квадрата (граница) однозначно определена в силу двойственности Пуанкаре, комплексного сопряжения и обнуления когомологий структурного пучка. Но остается неопределенной центральная часть

квадрата, причем, снова в силу двойственности Пуанкаре и комплексного сопряжения, размерности, которые стоят на двух диагоналях, должны равняться друг другу. То есть вся табличка, которая содержит в себе информацию о числах Ходжа, заключается в задании лишь двух чисел — a и b . Два числа a и b полностью задают диаграмму, и важно то, что на самом деле эти числа уже ни из каких соображений априори однозначно не определены. Можно показать, что существуют тысячи различных пар (a, b) , которые реализуются на 3-мерных многообразиях Калаби—Яу. В случае 3-мерной квинтики мы получаем $a = 1$ и $b = 101$.

Как я сказал, существуют тысячи различных возможностей для пар a, b . Однако пока никто не знает, существует ли для этих пар лишь конечное число возможностей. Есть даже разные мнения на этот счет. Известно только конечное множество пар (a, b) , полученное компьютером, но никто не знает, есть какая-либо оценка сверху на эти числа в случае многообразий Калаби—Яу размерности 3. Гипотезу о конечности 3-мерных многообразий Калаби—Яу совсем непонятно, как доказывать. И построить контрпример к такой гипотезе тоже непросто, потому что мы должны придумать какой-то способ конструировать бесконечно много топологически различных многообразий Калаби—Яу, и это тоже непонятно, как делать. По этому поводу есть некоторые идеи, но они лишь в стадии разработки.

Теперь мы готовы подойти к зеркальной симметрии с простейшей экспериментальной точки зрения. Есть экспериментальный материал (компьютерные вычисления), который был получен большей частью физиками, которые построили массу примеров 3-мерных гладких многообразий Калаби—Яу, и вычислили для них пары чисел (a, b) . Физики обратили внимание, что с реализацией каждой пары (a, b) почти всегда реализуется другая пара — (b, a) . Я сказал «почти», потому что в этом экспериментальном материале возникли некоторые исключения и симметрия, переставляющая местами числа a и b работала приблизительно в 96 (или 97) процентах случаев.

Если отметить на плоскости все точки с координатами (a, b) , которые реализуются числами Ходжа 3-мерных многообразий Калаби—Яу, то через расположение этих точек наглядно видна симметрия относительно прямой $x = y$. И эта симметрия у физиков не вызывала большого удивления. Они, в общем-то, считали, что это, так сказать, с их физической интуицией согласуется очень хорошо и отвечает явлению «зеркальной симметрии». Теперь я должен хоть что-то сказать относительно «обоснования» зеркальной симметрии в той форме, как оно рассматривается в физике возникает, но вместо углубления в терминологию теоретической физики, я хочу предложить «аналогию из элементарной геометрии». Дело

в том, что зеркальная симметрия в физике возникла на очень нестрогих основаниях, и они довольно непонятны для математиков. Физики используют так называемые суперконформные теории поля, которые только сейчас как-то приобретают какие-то более или менее нормальные математические основы. В момент открытия зеркальной симметрии (это было 15 лет назад) этого не было. Поэтому идти по этому пути я бы не стал. Я просто попытаюсь объяснить сам принцип рассуждений, который использовался физиками, когда они эту зеркальную симметрию предсказали.

Сам принцип рассуждений физиков довольно прост, но его легче всего объяснить с помощью некой «математической аналогии». Физики, скажем так, «умеют» геометрическому объекту (многообразию) сопоставить «другой объект», скажем, представление бесконечномерной алгебры, которое называется «суперконформная теория поля». И может так случиться, что в результате этого сопоставления разным многообразиям будет сопоставлено одно и то же (т. е. изоморфные теории). Если так происходит, то у двух многообразий X и Y может быть совершенно разная топологическая структура, но они могут привести к одному и тому же результату. В применении к комплексным многообразиям Калаби—Яу мы приходим таким образом к зеркальной симметрии, т. е. мы говорим, что два многообразия Калаби—Яу X и Y зеркально симметричны, если соответствующие теории поля изоморфны. То, что я сказал, выглядит математически пока очень плохо, поэтому я сейчас приведу ту математическую аналогию, которую обещал.

Давайте рассмотрим следующую чисто геометрическую конструкцию. Я возьму ромб и разобью его диагональю двумя способами на 2 конгруэнтных равнобедренных треугольника (рис. 1). Теперь обратите внимание:

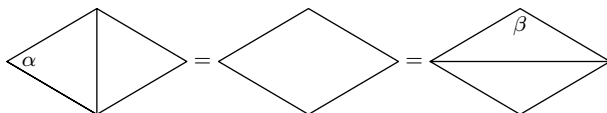


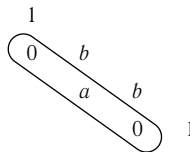
Рис. 1. Два разбиения ромба

я могу сказать, что я этот ромб получил, с одной стороны, из удвоения вот такого равнобедренного треугольника, а с другой стороны, я получил его из удвоения другого равнобедренного треугольника. Операция получения ромба одна и та же. Но треугольники, из которых мы получили этот ромб, разные. Конечно, есть связь между этими треугольниками, например, если угол при вершине одного треугольника α , а угол при вершине другого треугольника β , то есть связь: $\alpha + \beta = \pi$. Но сами равнобедренные треугольники разные — это существенный момент.

Что-то похожее происходит в зеркальной симметрии в физике. Есть два многообразия Калаби—Яу X и Y , по ним (по одному и по другому) строится некий объект, называемый суперконформная теория поля, она получается одна и та же. И процедура построения такая же; но вот исходный геометрический материал у них разный. Выше я привел некоторое условие того, что 2 треугольника при одном и том же способе построения привели к одному и тому же (т. е. конгруэнтному) ромбу: $\alpha + \beta = \pi$. Аналогичное условие есть и для многообразий, рассматриваемых физиками; в данном случае это будут d -мерные многообразия Калаби—Яу. Оказывается, что изоморфизм суперконформных теорий поля налагает некоторое необходимое условие на пару многообразий Калаби—Яу. Это необходимое условие выражается равенством чисел Ходжа: $h^{p,q}(X) = h^{d-p,q}(Y)$ для всех p, q .

Давайте посмотрим, что это условие означает в самом простейшем случае, который интересен для рассмотрения — 3-мерные Калаби—Яу. В этом случае есть лишь два числа, которые не определены в диаграмме Ходжа: $h^{1,1}(X)$ — это a , $h^{2,1}(X)$ — это b . Число a по этому необходимому условию должно быть равно $h^{2,1}(Y)$, а b должно быть равно $h^{1,1}(Y)$. То есть эти a и b — они меняются между собой. Но на самом деле то, что я здесь сформулировал, верно в любой размерности d — есть зеркальная симметрия для Калаби—Яу в любой размерности. Нас больше всего, конечно, интересует размерность 3: там больше получается замечательных свойств. Для диаграмм Ходжа зеркальная симметрия соответствует преобразованию осевой симметрии относительно вертикальной прямой, проходящей через центр квадрата.

Обратите внимание: я могу еще раз нарисовать диаграмму, чтобы проиллюстрировать эту осевую симметрию (рис. 2). Здесь было a , здесь



Р и с. 2. Диаграмма

было a , здесь b , здесь b . В одной диаграмме a — это просто второе число Бетти. Но в другой диаграмме число b — это не третье число Бетти; даже удвоенное b — это не третье число Бетти; тут еще возникает два раза число 1. Получается какая-то совершенно загадочная геометрическая связь между 3-мерными многообразиями Калаби—Яу X и Y . Так, например, если бы мы взяли квинтику X в \mathbb{P}^4 и рассмотрели зеркало, то зеркало для

квинтики (скажем, Y) должно было бы обладать следующим свойством: группа когомологий $H^{1,1}$ от зеркала должно иметь ранг 101, а $H^{2,1}$ (это размерность пространства деформаций) должно быть 1-мерным (т. е. мы получаем однопараметрическое семейство). Вот очень вкратце и очень наивно некое представление о зеркальной симметрии. Всё интересное так и осталось не высказано, я только немного пытался показать, как хотя бы приблизительно она работает.

Теперь я хочу сказать следующее: все фантазии физиков очень часто остаются фантазией, если не получается какой-то интересный математический результат. И самое удивительное, что такой результат получился. С этого момента началось действительно серьезное изучение зеркальной симметрии математиками. Эта работа была опубликована почти 15 лет назад, а сделана, конечно, она немножко раньше: P. Candelas, X. de la Ossa, P. Green, L. Parkes, Nucl. Phys. B 359 (1991), 27—74. Это работа четырех физиков: Филиппа Канделаса, его жены Ксении и двух аспирантов. Эта работа была опубликована в журнале «Nuclear physics»; работа была на 50 страниц — большая работа. Но если опустить много интересных вещей, которые были в этой работе, и вытащить интересный математический факт, то это будет одна формула. В этой работе физики получили одну замечательную формулу, которая вызвала интерес у математиков. Я сейчас хочу объяснить, как получается эта формула. Потому что она, в отличие от привычных математических формул, требует несколько больше слов для объяснения. Сейчас моя первая цель — написать эту формулу, чтобы видеть, как она получается, а потом попытаться понять её смысл.

Рассматривается некая дифференциальная форма $\Omega(x)$ от комплексного переменного x , точнее сказать, рассматривается тензор типа $(3, 0)$ в проколотой окрестности нуля с координатой x — это комплексная координата в окрестности 0. Дифференциальная форма $\Omega(x)$ пишется в следующем виде:

$$\Omega(x) = \frac{5}{(1 - 5^5 x) y_0^2(x)} \left(\frac{dx}{x} \right)^{\otimes 3}.$$

Отметим присутствие числа 5 в этой формуле (это число связано с 3-мерными квинтиками). Различие с кососимметрическим дифференциалом, т. е. с сечением внешних степеней кокасательного расслоения, состоит в том, что это сечение симметрической степени кокасательного расслоения. Этот дифференциал логарифмический, т. е. с логарифмическими особенностями в нуле. Мы рассматриваем такое сечение в окрестности нуля (в проколотой окрестности нуля, потому что здесь полюс). Осталось только определить аналитическую функцию $y_0(x)$. Функция $y_0(x)$ — это

следующий степенной ряд с явно выписываемыми коэффициентами:

$$y_0(x) = \sum_{n=0}^{\infty} \frac{(5n)!}{(n!)^5} x^n = 1 + 120x + \dots$$

Опять мы видим, что в этом выражении число 5 появляется в особой роли.

Потом с этим дифференциалом производится некая операция, которая состоит в том, что мы меняем локальную координату в проколотой этой окрестности точки 0: от координаты x переходим к координате q , которая в первом приближении совпадает с x . Она определяется следующим образом:

$q = e^{\left(\frac{y_1(x)}{y_0(x)}\right)} \sim x$. Здесь q — это экспонента отношения двух степенных рядов. Один ряд я уже определил: это $y_0(x)$, который стоит в знаменателе; а в числителе я еще должен написать ряд $y_1(x)$, он немножко громоздко выглядит. Ряд $y_1(x)$ выписывается следующим образом:

$$y_1(x) = y_0(x) \ln x + 5 \sum_{n=0}^{\infty} \frac{(5n)!}{(n!)^5} \left[\sum_{j=0}^n \frac{1}{j} \right] x^n.$$

Здесь y_0 умножается на логарифм (у y_0 в нуле никакой особенности нет, но она умножается на логарифм, тем самым y_1 приобретает особенность), и плюс в формуле дальше идет некоторый неособый добавок в виде ряда.

Зачем всё это делается, пока не объяснялось; моя цель — сперва написать формулу, чтобы появилось на доске выражение, которое мы будем далее осуждать. Можно легко заметить, что когда я разделю y_1 на y_0 , то за счет того, что у меня y_0 в формуле для y_1 умножается на логарифм, у нас здесь как бы y_0 сократится и возникнет такой главный член — $\ln x$. Но за счет взятия экспоненты мы получаем снова x . Таким образом, q — это некая другая, «чуть подправленная» координата x . Для тех, кто знаком с теорией модулярных функций, q — это то q , которое получается в теории модулярных функций. Если мы сделали бы всю эту процедуру, заменив пятерку на тройку, тогда мы бы работали с эллиптическими кривыми (кубиками), это q было бы то, что возникает в теории эллиптических кривых.

Дальше мы этот дифференциал, который был в координате x , переписываем в новую координату q : $\Omega(q) = K(q) \left(\frac{dq}{q}\right)^{\otimes 3}$, где

$$K(q) = 5 + \sum_{d=1}^{\infty} n_d \cdot d^3 \left(\frac{q^d}{1 - q^d} \right),$$

причем загадочным образом получаем: $n_d \in \mathbb{N}$. Мы получаем некоторую голоморфную функцию $K(q)$, умноженную на логарифмический диффе-

ренциал в 3-й степени в координате q . И то, что здесь получается, оказывается, несет в себе очень интересную информацию. Интересный объект — коэффициенты этой функции, ее разложение в ряд $K(q)$. Легко видеть, что 5 — это константа, с которой всё начинается, она видна с самого начала. А потом возникает суммирование уже по положительным степеням q , всё выражение лучше записать следующей форме: число $n_d d^3$, умноженное на дробь $\frac{q^d}{1 - q^d}$, которую можно рассматривать как свернутую геометрическую прогрессию.

Оказывается, что, если мы вычислим коэффициенты n_d , которые приведены в формуле выше, т. е. если мы преобразуем Ω в новую координату, разложим функцию $K(q)$ в формальный степенной ряд по степеням q , то вычисление первых коэффициентов n_d экспериментально показывает, что они все — натуральные числа. Первый коэффициент $n_1 = 2875$. Следующий коэффициент $n_2 = 609250$. Потом следующий: $n_3 = 317206375$. Ну а дальше получаются очень большие натуральные числа.

Интересно то, что число n_1 — это не что иное, как число прямых на общей квинтике в \mathbb{P}^4 , т. е. на 3-мерном многообразии Калаби—Яу. Дальше то же самое: n_2 — число коник, а n_3 — число рациональных кубик.

Этот результат поразил математиков до глубины души. Потому что в статье физиков присутствуют значения не только первых трех коэффициентов n_d , а сразу выписаны коэффициенты n_d до $d = 10$. И это было уже выше возможностей математиков. Было просто невероятно, как вообще можно такое посчитать. Более того, первые два числа было относительно несложно посчитать; а вот с третьим числом n_3 уже была довольно долгая возня, и даже была написана специально для этого компьютерная программа, которая проводила вычисление. Станным образом программа давала другой ответ. Когда математики узнали о работе физиков, они были, конечно, удивлены, почему они получили другой ответ. Но после некоторых усилий была найдена ошибка в компьютерной программе. Ответ у физиков был правильный. Тем самым физики даже в чем-то, так сказать, обогнали математиков. Нужно было каким-то образом это всё объяснить. И вся эта история началась с 1991 или 1992 года. Первые попытки найти понимание были связаны с тем, чтобы отдельный пример вычисления физиков не остался единственным и изолированным. Требовалось иметь побольше подобных примеров, побольше предсказаний. А потом первый значительный прорыв был сделан Максимом Концевичем, который предложил общий принцип подсчета кривых и математически доказал верность коэффициента при $d = 4$ (только конечно, здесь d не нужно путать с размерностью: в данном случае d — это степень рацио-

нальной кривой). С этого момента зеркальная симметрия очень увлекла математиков.

В принципе, с точки зрения математики сейчас это вычисление физиков уже не является загадочным. Физики все-таки не считают реальное количество рациональных кривых степени d на трехмерной квинтике. Они считают некоторый «индекс пересечения». Например, мы не можем заранее предсказать, что две кривые на поверхности имеют в точности столько-то общих точек: надо всегда учитывать кратность этих точек и другие подобные эффекты. То есть на самом деле, эти числа n_d — это некоторые индексы пересечения на некоторых специальных многообразиях, так называемых многообразиях стабильных отображений — то, что было предложено Максимом Концевичем. И для того, чтобы всё построить, объяснить, чтобы это всё как-то состыковалось, потребовалось достаточно долгое время. А решающий момент во всем этом, как я считаю, — в том, что для того, чтобы вычислить индексы пересечения на некоторых многообразиях, использовалось действие на нем алгебраического тора, что позволяло «локализовать» вычисление по действию тора.

Число n_d — это не число кривых степени d . Это некоторый индекс пересечения на некоторых многообразиях. Вообще говоря, сказать, что это — число кривых, в общем случае даже неверно. Можно привести какой-то другой пример, когда мы видим, что даже общее многообразие Калаби—Яу имеет бесконечное число рациональных кривых, а тем не менее, ответом будет конечное число. Для квинтики, по-моему, есть какие-то примеры, показывающие бесконечные семейства рациональных кривых для большой степени d . Мы видим, что в этих примерах интуиция нас подводит. И все же явный подсчет кривых на многообразиях Калаби—Яу представляет интерес. Это интересно хотя бы с точки зрения истории, потому что если бы не было предварительной ошибки в вычислении математиков, если бы не было бы известных чисел для количества прямых и коник, то это не произвело бы такого впечатления. Правда теперь мы знаем, что это просто некоторые индексы пересечения. Иногда эти индексы пересечения в точности совпадают с числом общих точек пересекаемых дивизоров.

Теперь я обращусь к тому, что относится к названию моего доклада. То, что я последнее время делал, было связано с работой Моррисона и Плессера: D. Morrison, R. Plesser, Nucl. Phys. B 440 (1995), 279—354. Это тоже была работа достаточно длинная — где-то 75 страниц. Но я пока опять хочу сделать из этой работы некоторый экстракт и объяснить, как это связано с формулой, предложенной Канделасом и его соавторами, это было сделано лет на 5 раньше. То есть эта работа появилась еще тогда,

когда не появилась работа Гивенталя, где было очень убедительное объяснение, почему так. А это была работа, где было дано некое приближение, по крайней мере, хоть какая-то зацепка, почему то, что там писалось, хоть как-то связано с квинтикой.

В этой работе был предложен наивный подсчет числа рациональных кривых на квинтике. На самом деле подсчет не рациональных кривых, а отображений рациональных кривых в квинтику. Я еще раз обращаю внимание: формула, которую я выписывал, содержала некоторые члены, типа вот такого: $\frac{q^d}{1-q^d}$. Они возникают из-за того, что формула считает на самом-то деле не число рациональных кривых, а число отображений. Вот я хочу как раз коснуться этого вопроса. Мы предпримем сейчас наивный подсчет числа отображений \mathbb{P}^1 в квинтику. Отображение рациональной кривой в 4-мерное проективное пространство задается однородными многочленами $f_0(u, v), \dots, f_4(u, v)$ степени d , т. е. мы должны просто взять 5 однородных многочленов от переменных u и v , где u и v — это однородные координаты на \mathbb{P}^1 . Если мы возьмем 5 однородных многочленов степени d , то они как раз дадут отображение $\mathbb{P}^1 \rightarrow \mathbb{P}^4$. Отображение фактически осуществляется так: $(u, v) \mapsto (f_0, f_1, \dots, f_4)$. Теперь можно посмотреть, а что является параметрами этих отображений; т. е. мы можем построить некоторое многообразие, параметризующее отображения при фиксированном d . Но при фиксированном d естественными параметрами являются коэффициенты многочленов, которые здесь возникают. Поскольку каждый многочлен имеет степень d , коэффициентов на единицу больше — $d + 1$. Мы должны, конечно, это всё профакторизовать по умножению на общий скаляр, и получится в результате $5(d + 1) - 1 = 5d + 4$ параметров (с точностью до общего множителя). И как наивное пространство, параметризующее эти отображения, можно просто взять проективное пространство размерности $5d + 4$, где однородными координатами являются все эти коэффициенты. А нас интересуют, конечно, не просто отображения $\mathbb{P}^1 \rightarrow \mathbb{P}^4$. Нас интересуют те отображения, которые попадают в квинтику (в общую квинтику). То есть мы должны взять на \mathbb{P}^4 координаты x_0, \dots, x_4 , потом возьмем некий многочлен 5 степени $Q(z_0, \dots, z_4)$; для того чтобы результатом отображения была кривая на квинтике, надо следующее: нам нужно просто сделать подстановку, подставить вместо переменных эти 5 однородных многочленов, и если мы получим тождественный 0 (т. е. многочлен), то это будет означать, что у нас вся кривая (образ \mathbb{P}^1) будет лежать на квинтике. Вот и надо посчитать, и сколько же таких отображений существует. Если мы просто возьмем результат подстановки в Q всех этих однородных многочленов, то это будет многочлен, вообще говоря, просто

от переменных u и v : $P = Q(f_0, \dots, f_4) = P(u, v)$. Какая будет степень P ? Легко видеть: квинтика — был многочлен степени 5, каждый из них степени d . Значит, степень будет $5d$. Но это многочлен от двух переменных. Следовательно, получается $5d + 1$ коэффициентов. И значит, если эти все $5d + 1$ коэффициентов приравнять нулю, то это и будет условие, что образ \mathbb{P}^1 лежит в квинтике.

Давайте теперь посмотрим внимательно, что означает обращение в нуль одного коэффициента. Каждый коэффициент P_i — это на самом деле будет многочлен от коэффициентов исходных (f_i) , однородный многочлен степени 5 от u и v . То есть обращение в нуль означает, что нас интересует пересечение на этом многообразии модулей, пересечение дивизоров (гиперповерхностей), заданных обращением в нуль каждого P_i , $i = 0, \dots, 5d$. У нас получается $5d + 1$ дивизоров или гиперповерхностей, и мы хотим найти их пересечение, индекс пересечения. Нас интересует то, что лежит в пересечении. Будем считать, что всё общего положения, и тогда, наивно, у нас пересекаются $5d + 1$ гиперповерхностей степени 5. У нас получится, вообще говоря, не 0-мерное многообразие, потому что здесь была размерность $5d + 4$. Но надо не забыть, что отображение нужно еще профакторизовать по автоморфизмам \mathbb{P}^1 . Дело в том, что \mathbb{P}^1 имеет 3-мерную группу автоморфизмов $\mathrm{PGL}(2)$. И если мы при фиксации, скажем, трех точек на \mathbb{P}^1 уже фиксируем как бы дополнительную жесткость, она уже дает конечное число. То есть нам как бы не хватает еще некоторых, так сказать, зафиксированных трех параметров, тогда уже получится число.

Теперь я напишу тот индекс пересечения, который меня будет интересовать. Это будет индекс пересечения, который связан с обращением в нуль всех $5d + 1$ коэффициентов, и плюс еще фиксация с точностью до автоморфизма, но это нелинейное условие. А каждое обращение в нуль — это, на самом деле, поверхность степени 5. Я сейчас посчитаю некоторый индекс пересечения на таком наивном многообразии модулей. А именно, n_d . Можно написать: $P = \sum_{i=0}^{5d} P_i u^i v^{5d-i}$. И каждый коэффициент P_i — это многочлен от коэффициентов того, что мы здесь подставим. И это будет многочлен — однородный многочлен степени 5 от коэффициентов. Именно поэтому обращение в нуль коэффициентов будет давать дивизориальное условие на этом многообразии. Но не просто дивизориальное: это будет дивизор степени 5. То есть скажем так: P_i определяет здесь дивизор D_i , и как элемент группы Пикара это будет эквивалентно пяти гиперплоским сечениям: $D \sim 5H$ для всех $i = 0, \dots, 5d$. Давайте посмотрим, какой индекс пересечения я хочу посчитать. Я беру следующий индекс

пересечения:

$$N_a = H^3 \prod_{i=0}^{5d} D_i = H^3 (5H)^{5d+1} = 5^{5d+1} (H^{5d+4})$$

— H^3 , а дальше индекс пересечения, я напишу его в виде произведения — все условия зануления всех коэффициентов. Условие H^3 возникает как раз из-за того, что мы должны убить 3 размерности с помощью учета группы автоморфизмов \mathbb{P}^1 . А остальное естественным образом суммируется. Это индекс пересечения на таком многообразии. Его посчитать не представляет никакого труда. Теперь мы знаем, что здесь у нас каждый дивизор такой. Тогда у нас получится просто $(5H)^{5d+1}$. И мы знаем, что гиперплоское сечение в максимальной степени будет давать единичку: $H^{5d+4} = 1$. Получается $5^{5d+1} H^{5d+4}$. И теперь уже получится просто число $N_d = 5^{5d+1}$.

А теперь я просто возьму степенной ряд, имеющий эти коэффициенты, т. е. я рассмотрю степенной ряд

$$\sum_{d \geq 0} N_d t^d = \frac{5}{1 - 5^5 t}.$$

В данном случае, поскольку мы это явно вычислили, тут никакого труда не представляет посчитать: это будет геометрическая прогрессия, и ответ будет $\frac{5}{1 - 5^5 t}$. Этот ответ замечателен тем, что он все-таки имеет отношение к той формуле, которая была предложена физиками. Я еще раз напомним, что физики в своих вычислениях стартовали с некоего дифференциала, который имел следующий вид:

$$\Omega(x) = \frac{5}{(1 - 5^5 x) y_0^2(x)} \left(\frac{dx}{x} \right)^{\otimes 3}.$$

Потом делалась замена переменных. Эта рациональная функция здесь так возникает. Конечно, то, что мы здесь получили — это ответ неверный: это не есть число отображений \mathbb{P}^1 . Но тем не менее, есть некие поправки. Корень этой неверности состоит в том, что мы неправильно взяли многообразии модулей, на котором мы вычисляем индекс пересечения. Поэтому мы получаем не те ответы. Но тем не менее, это все-таки является неким приближением. С чем это можно сравнить? Например, кажется, в книге Гриффитса и Харриса разбирается задача исчислительной геометрии, где касаются коники. Там есть некий такой наивный подсчет, который дает неправильный ответ, а есть более правильные подсчеты. Что-то в этом духе происходит и здесь. На правильном многообразии нужно считать

индекс пересечения, тогда получится правильный ответ. Мы сейчас делали это не на правильном многообразии, но тем не менее получили некое приближение.

В последнее время я пытался сформулировать в виде некоего утверждения, сначала гипотетического, следующее: почему совпадение этой рациональной функции и того, что дает наивное вычисление, — это не случайность? И оказалось возможным сформулировать некое более общее утверждение, сначала в виде гипотезы, которое достаточно быстро было доказано несколькими независимо математиками. Эти доказательства достаточно разные. Теперь я вынужден уточнить, о чем идет речь, и что значит обобщить или сформулировать гипотезу. Дело в том, что до сих пор я рассматривал очень простой пример многообразия Калаби—Яу, который возникает как некоторая серия в любой размерности: берем гиперповерхность степени $d + 2$ в \mathbb{P}^{d+1} . Конечно, это не все примеры; это просто один представитель, топологический, в каждой размерности. А я уже сказал, что в размерности 3 существуют тысячи различных примеров. Мне сейчас нужно объяснить следующее: как вообще получить как можно больше примеров многообразий Калаби—Яу. И притом так, чтобы зеркальная симметрия при этом приобрела некоторую естественную интерпретацию. И есть класс, для которого я могу сформулировать утверждение, которое являлось бы в каком-то смысле утверждением, показывающим, что то, что сейчас мы заметили для случая квинтик, это не случайность, это общий факт, который требует математического доказательства.

Выше я использовал параметризацию рациональных кривых с помощью однородных многочленов f_i степени d . Если эти многочлены все взаимно просты, то дают, действительно, кривую степени d . А если они не взаимно просты, то уже у кривой падает степень. И учет того, что происходит с не взаимно простыми многочленами, должен быть аккуратным. Если брать наивную компактификацию, то этого учета не происходит. Мы можем сказать, что мы делаем какое-то подразрешение, т. е. мы должны для такой модели (это тоже рациональное многообразие) кое-что добавить и несколько раздуть, и уже на этом новом многообразии рассматривать индекс пересечения.

Конечно, построение этой другой модели уже более интересно. Для случая проективных пространств, по-моему, фактически уже достаточно того, что было написано в работе Концевича и Манина. Но в общем случае нужно действовать более аккуратно. Сейчас я объясню, на какой класс примеров я хочу обобщить только что сформулированное наблюдение, и провести эвристически наивный подсчет. Здесь возникает элементарная геометрия, выпуклая геометрия, связанная с торической геометрией.

Я начну с чисто формального определения неких многогранников, которые называются рефлексивными. Речь идет вот о чем. Мы рассматриваем выпуклый многогранник $\Delta \subset \mathbb{R}^d$, $\dim \Delta = d$, $0 \in \text{int } \Delta$, т. е. выпуклую оболочку конечного числа точек. Предполагаем, что размерность совпадает с d — максимальна; и еще заодно, что точка 0 лежит строго внутри многогранника. Такой многогранник мы назовем рефлексивным, если выполнены следующие условия:

(1) все вершины Δ лежат в \mathbb{Z}^d ;

(2) $\Delta^* = \{y \in \mathbb{R}^d \mid \langle x, y \rangle \geq -1 \ \forall x \in \Delta\}$, $\langle x, y \rangle = \sum x_i y_i$; все вершины Δ^* лежат в \mathbb{Z}^d .

Первое условие: все вершины многогранника Δ лежат в стандартной решетке. А второе: мы теперь к каждому многограннику Δ можем построить двойственный, или полярный многогранник, который обозначим через Δ^* . Чтобы это формально определить, мне понадобится еще скалярное произведение. Я предполагаю, что $\langle x, y \rangle$ — стандартное скалярное произведение. Я определю другой многогранник, двойственный, или полярный, как множество таких точек в \mathbb{R}^n , что $\langle x, y \rangle \geq -1$ для любого $x \in \Delta$. Это получается некий другой многогранник, полярный. И для него я тоже потребую, что все вершины тоже лежат в \mathbb{Z}^n . Это очень сильное, на самом деле, условие. Не очевидно, но можно доказать, что в любой фиксированной размерности с точностью до автоморфизмов решетки существует только конечное число рефлексивных многогранников.

Самый простейший пример $d = 1$. Если бы мы использовали только первое условие (все вершины лежат в решетке), то это будет отрезок $[-a, b]$. А теперь мы построим двойственный $[-1/b, 1/a]$. Для того чтобы двойственный $[-a, b]^*$ тоже имел целочисленные вершины, нужно, чтобы a и b были равны 1. В общем, чтобы была целочисленность, сразу же мы приходим к единственной возможности: отрезок $[-1, 1]$ — это единственный рефлексивный многогранник в размерности 1. В размерности 2 рефлексивные целочисленные многогранники можно все перечислить (это относительно простая задача). Их будет 16 штук, 16 многоугольников. Давайте я приведу простейший пример в размерности 2. Возьмем треугольник Δ с вершинами $(-1, -1)$, $(2, -1)$, $(-1, 2)$. Для него двойственным многоугольником Δ^* будет следующее: мы должны просто брать внутреннюю нормаль к каждой грани; т. е. у нас получится треугольник, который есть выпуклая оболочка векторов $(1, 0)$, $(0, 1)$, $(-1, -1)$. В размерности 3 можно взять октаэдр, т. е. выпуклую оболочку базисных векторов со знаком плюс и минус. И двойственным для него будет куб. Тут есть двойственность между гранями и вершинами.

Я должен объяснить, какая связь между рефлексивными многогранниками и теми вещами, о которых я до этого говорил. Во-первых, общее замечание, что если есть рефлексивный многогранник, то тогда двойственный тоже рефлексивный, и два раза применяя эту операцию, мы просто вернемся обратно: $(\Delta^*)^* = \Delta$. В частности, рефлексивные многогранники всегда возникают парами.

Второе: можно с каждым рефлексивным многогранником связать семейство многообразий Калаби—Яу. Я сейчас объясню, как устроено это семейство. Пусть у нас есть рефлексивный многогранник. Тогда мы рассмотрим семейство многочленов Лорана, которое определено таким образом: $f(t) = \sum_{m \in \mathbb{Z}^d \cap \Delta} a_m t^m$. Берем суммирование по всем точкам решетки,

которые лежат внутри этого многогранника; а дальше рассматриваем просто мономы Лорана с общими коэффициентами. Если такой многочлен приравнять нулю, то это задает некую гиперповерхность $f=0$ в торе: $Z_f \subset (\mathbb{C}^*)^d$. Тор размерности d . И оказывается, что эта гиперповерхность всегда (если коэффициенты общие) может быть компактифицирована до многообразия Калаби—Яу. Это можно сделать для Δ , это можно сделать также для Δ^* ; у нас получится два семейства многообразий Калаби—Яу. И вот оказывается, что эти два семейства обладают в точности теми свойствами, которые требует зеркальная симметрия. Если мы посчитаем числа Ходжа, то они будут связаны так, как предсказывает зеркальная симметрия.

Теперь я делаю очень приблизительно набросок утверждений. Сначала у нас аффинная гиперповерхность $f=0$, затем мы должны ее компактифицировать, потом, может быть, даже разрешить особенности, но я опускаю все эти детали и говорю просто, что если есть рефлексивный многогранник, то вот это семейство гиперповерхностей дает семейство гладких многообразий Калаби—Яу (по крайней мере, если мы возьмем многогранник размерности 4, а гиперповерхность будет иметь размерность 3).

Мы получаем другое семейство, если рассмотрим двойственный рефлексивный многогранник, и эти два семейства — это и есть примеры пар зеркально симметричных многообразий Калаби—Яу.

Для размерности 2 рефлексивных многогранников 16, в размерности 3 их уже очень много, а в размерности 4, может быть, даже миллионы — очень много существует рефлексивных многогранников, и мы получаем массу примеров. В каждой размерности мы получаем конечное число рефлексивных многогранников, поэтому получаем конечное число семейств; это общий факт. То есть на этом пути мы не можем построить бесконечно много топологических типов. Нас интересует сейчас не это, нас интере-

сует зеркальная симметрия. Мы получаем большой материал, на котором можем тестировать зеркальную симметрию. То, что предсказывается зеркальной симметрией, мы можем просто проверить: а как это работает для построенных пар зеркально симметричных многообразий.

Теория зеркальной симметрии сначала на самом деле не существовала, была только некоторая догадка физиков, как и что делать. Это было очень интересно. Еще до этого были работы Хованского и Данилова по вычислению чисел Ходжа—Делиня гиперповерхностей в торах. Очень хорошие работы. Эти работы позволяли, в принципе (там такие задачи не ставились), также вычислять и числа Ходжа для компактификаций. Если взглянуть на эти числа Ходжа, то все числа Ходжа вычисляются в терминах комбинаторики, подсчета числа точек в целочисленных многогранниках. А когда я был студентом, мне в 1980 году мой научный руководитель В. А. Исковских поручил классификацию торических многообразий Фано. В принципе как раз с каждым таким рефлексивным многогранником ассоциируется многообразие не Калаби—Яу, а торическое многообразие Фано. И гладкие торические многообразия Фано размерности 3 классифицируются относительно легко. А вот в размерности 4 — эта классификация более сложна. Для меня было полезно, что я довольно долго возился с 4-мерными торическими многообразиями Фано. А если взять гиперплоское сечение 4-мерного многообразия Фано, получится 3-мерное многообразие Калаби—Яу. Уже из моего предварительного практического опыта работы с торическими многообразиями Фано эта двойственность для рефлексивных многогранников была уже как-то естественно видна. Но тогда было совершенно непонятно, как ее можно было с чем-то связать. И вот, наконец, физики что-то такое предложили, и всё сработало замечательно.

Откуда берется понятие «рефлексивного многогранника»? Мы хотим показать, что есть многогранники, которые мы называем рефлексивными, и это тот класс многогранников, из которых мы можем строить много примеров многообразий Калаби—Яу. Если он не рефлексивный, то мы тоже получаем гиперповерхности, но совершенно не обязательно они будут бирациональными к многообразиям Калаби—Яу. Я рассмотрю для примера просто 2-мерные целочисленные многогранники, т. е. многоугольники на плоскости. И хочу рассмотреть нули гиперповерхности, уравнение которой содержит мономы, соответствующие его целым точкам. В данном случае это будет аффинная кривая. Спрашивается: какой ее род? Оказывается, что ее род — число точек строго внутри многоугольника. А если нас интересуют многообразия Калаби—Яу, то род кривой должен быть равен 1. Значит, получаем уже сразу условие: внутри многоуголь-

ника должна быть ровно одна целая точка. То же самое верно в высших размерностях. Еще до открытия зеркальной симметрии был известен факт (доказанный, по-моему, Хованским), что геометрический род (т. е. число Ходжа $h^{d-1,0}$) — это число точек строго внутри многогранника. Значит, сразу мы уже знаем количество целых точек внутри многогранника: 1. Но то, что это число Ходжа равно 1 — это еще не означает, что канонический класс зануляется, это только необходимое условие. А если мы уже хотим сказать это в точности, то добавляется еще кое-что. В принципе, эта рефлексивность равносильна тому, что компактификация тора относительно этого многогранника будет горенштейново многообразии Фано. Это можно так переформулировать. И тогда мы просто берем антиканоническую гиперповерхность — Калаби—Яу. С горенштейновыми особенностями надо отдельно дополнительно разбираться.

Вот для такого класса многообразий Калаби—Яу удастся обобщить то наблюдение, о котором я говорил в случае квинтики. Даже можно сказать, какой рефлексивный многогранник соответствует квинтике. Его трудно, конечно, нарисовать, потому что это 4-мерный многогранник. Но это как бы обобщение того, что мы уже нарисовали. Это — многогранник Ньютона общего однородного многочлена степени 3 (все 10 точек, которые здесь). Если мы возьмем стандартный 3-мерный симплекс, умножим на 5 и поставим точку, то это будет как раз рефлексивный многогранник, который отвечает квинтике. Скажем так: если мы компактифицируем тор (относительно этого многогранника), то тогда получится проективное пространство, причем не просто проективное пространство, а с фиксированным проективным вложением, т. е. по Веронезе, с помощью 5-кратного вложения Веронезе, т. е. так, что общее гиперплоское сечение будет квинтика.

Теперь я должен сказать хоть что-нибудь о торических вычетах: какую роль они играют, потому что я лишь объяснил класс примеров многообразий Калаби—Яу, для которых мы должны обобщить наблюдение Моррисона—Плессера. Главным в этом обобщении является понятие торического вычета. Это, я думаю, стоит объяснить.

Если мы говорим, что мы берем двойственный многогранник, что мы компактифицируем тор относительно двойственного многогранника, и одновременно компактифицируем гиперповерхность, поскольку в общем случае она будет сначала негладкая. Но мы можем разрешить особенности, по крайней мере в размерности 3, так, что всё будет гладкое. То, что я сейчас рассказываю про рефлексивные многогранники, этого не было в работе Канделаса, там была предложена другая конструкция, но она эквивалентна абсолютно тому, что я говорю. То есть нет никакой разницы,

но на языке многогранников это гораздо более элегантно. Канделас тоже тут же перенял этот язык, когда мы встречались, сразу же перешел на язык рефлексивных многогранников, потому что это более естественно, более элегантно.

Я хотел сказать о торических вычетах. Торический вычет — это некое обобщение классического вычета Гротендика. Что такое вычет Гротендика, объясняется в книжке Гриффитса и Харриса. Речь идет о следующем. Мы берем, скажем, кольцо многочленов от n переменных (даже лучше локализовать его в нуле). И дальше рассматривается некоторое подмножество $\{f_1, \dots, f_n\} \subset \mathbb{C}[x_1, \dots, x_n]$ — регулярная последовательность n однородных многочленов — каждый своей степени, скажем, d_1, \dots, d_n . Регулярность последовательности означает, что, если мы возьмем это кольцо и профакторизуем по идеалу, порожденному этой последовательностью, то мы получим 0-мерное (артиново) кольцо; как векторное пространство это будет конечномерное пространство, его размерность c как векторного пространства будет конечна. Более того, это кольцо — очень хорошее, это горенштейново кольцо, в нем есть некая хорошая двойственность. Это кольцо, которое я обозначу $A = R/(f_1, \dots, f_n)$, устроено следующим образом: $A = \bigoplus A^i$ — градуированное кольцо, с некоторого момента все компоненты уже равны 0, потому что общая размерность как векторного пространства конечна, но не только: существует некоторая последняя ненулевая компонента, т. е. я могу сказать: $A^N \neq 0$, $A^{N+1} = 0$. И возникает некая естественная двойственность: $A^i \times A^{N-i} \rightarrow A^N = \mathbb{C}$ — это просто 1-мерное пространство. Более того, в этом 1-мерном пространстве можно даже выбрать некоторый более или менее канонический представитель: он берется просто из матрицы частных производных, мы берем определитель матрицы частных производных $J = \det\left(\frac{\partial f_i}{\partial x_j}\right)$, класс этого определителя. Тут сразу видно, чему равняется N , потому что если мы знаем, какие d_i , то в этом случае $N = \left(\sum_{i=1}^n d_i\right) - n$. И в результате у нас возникает некое отображение, и каждый многочлен этой, так сказать, критической размерности (равной степени компоненты) можно спроецировать — существует отображение вычета $\text{Res} : R^N \rightarrow \mathbb{C}$, которое отображает однородную компоненту этого кольца в \mathbb{C} . При этой проекции канонический представитель, который я обозначил Y (якобиан), переходит в произведение $d_1 \dots d_n$. Полученное отображение называется вычетом Гротендика. Оно зависит от выбора многочленов.

Оказывается, существует некое обобщение этого отображения для случая не только кольца многочленов, но и неких моноидальных ко-

лец (конструкция торического вычета принадлежит Коксу), т. е. колец, связанных с полугруппой целых точек. Мы можем сказать, что кольцо многочленов — это не что иное, как полугрупповое кольцо, связанное с моноидом целых точек в стандартном октанте — конусом над стандартным симплексом. Но можно рассматривать конуса не только над симплексами, можно рассматривать конуса, скажем, над рефлексивными многогранниками — то, что меня больше интересовало. И тоже существует некоторый способ, которым мы можем это определить. Удивительным образом рефлексивность многогранника связана с тем, что у нас получится горенштейново кольцо, и в этом случае тоже возникает некоторое отображение вычета. Это отображение используется для того, чтобы определить рациональные функции от коэффициентов уравнения гиперповерхности, их разложение в ряд меня будет интересовать.

Я напому, что в самом начале у нас возникала некоторая рациональная функция $\frac{5}{1-5^5t}$, и ее коэффициенты разложения в ряд интерпретируются как индексы пересечения. Цель такая: с одной стороны, определить некоторые рациональные функции, которые в общем случае для класса гиперповерхностей Калаби—Яу в торических многообразиях связаны с рефлексивными многогранниками, а с другой стороны, потом показать, что коэффициенты разложения этих рациональных функций в многомерные степенные ряды являются индексами пересечения на некоторых многообразиях модулей. Вот это, грубо говоря, была цель моей гипотезы. Для того чтобы это реализовать, нужно было ввести некую терминологию, вспомнить некие результаты, которые были известны еще в конце 80-х годов. Ими были результаты по вторичным многогранникам Гельфанда—Зелевинского—Капранова, связанные с триангуляциями целочисленных многогранников — всё это оказалось к месту.

Именно таким образом моя гипотеза была сформулирована, когда появилась первая моя статья с Евгением Матеровым, где мы рассматривали ее для некоторых частных случаев гиперповерхностей Калаби—Яу. На самом деле, то, что я сказал о построении примеров зеркально симметричных Калаби—Яу для гиперповерхностей в торических многообразиях, обобщается и для полных пересечений, и там тоже всё работает. Можно даже, оказывается, сформулировать аналогичную гипотезу или утверждение о совпадении разложений в степенные ряды в случае полных пересечений.

Еще раз хочу подчеркнуть: есть как бы два объекта: один объект — это рациональная функция многих переменных, а с другой стороны, есть некоторые степенные ряды. Так вот, эта рациональная функция раскладывается в ряд. Я скажу, какого плана разложение в ряд. Потому что

можно по-разному раскладывать в ряд, всё зависит от того, в какой точке. И в данном случае используется разложение рациональной функции «относительно вершины многогранника Ньютона» знаменателя этой функции. Дело в том (и это тоже уже часто использовалось, например, в последних работах Хованского), что если есть рациональная функция и мы что-то можем сказать о многограннике Ньютона знаменателя, то тогда каждая вершина многогранника Ньютона знаменателя позволяет разложить в ряд — получится некоторое суммирование по целым точкам в касательном конусе этой вершины. Так вот, эти ряды, которые получаются (коэффициенты разложения), можно интерпретировать как индексы пересечения на некоторых многообразиях модулей, которые на самом деле являются торическими многообразиями. Они являются обобщениями того самого проективного пространства размерности $5d + 4$, которое я рассматривал в примере, и некий подсчет отдельных примеров показывает, что всё должно сходиться.

Когда была сформулирована гипотеза, еще не было понятно, каким образом это доказывать. Но вскоре появились по крайней мере два доказательства. Я сейчас просто скажу имена людей, которые это сделали: Лев Борисов, Андраш Сенеш (весной этого года, ближе к лету); и Мишель Вернь и Калле Кару. Они доказали эту сформулированную мной гипотезу независимо. Первое доказательство более алгебраическое, второе — более аналитическое.

Почему меня заинтересовала эта гипотеза? В последнее время появилось много утверждений (есть даже гипотеза Виттена), когда мы, с одной стороны, рассматриваем некоторую функцию, а с другой стороны, нас интересует интерпретация разложения в ряд коэффициентов этой функции с точки зрения теории пересечений на неких многообразиях модулей. И во всех до сих пор известных примерах всё это делалось тяжело. В работе Гивенталья рассматривается многообразие, которое плохо понимаемо, но зато локализация по действию тора дает возможность вычислять. В случае гипотезы Виттена нужно рассматривать многообразие модулей стабильных кривых, и это тоже многообразие, которое очень непростое. А вот в этой ситуации все многообразия модулей получаются торическими, и все они очень хорошо понимаются. А принцип снова такой же: в одной интерпретации есть коэффициенты разложения и индекс пересечения; а другая интерпретация возникает совсем из других областей. И я рассматриваю эту ситуацию как практический пример, где можно тестировать подходящий язык для других подобных примеров.

В. Ю. К а л о ш и н

СЛУЧАЙНЫЕ БЛУЖДЕНИЯ ВДОЛЬ ОРБИТ ХАОТИЧЕСКИХ ОТОБРАЖЕНИЙ

В своем докладе я хотел бы рассказать о задаче, которой некоторое время занимались мы с Я. Г. Синаем. И в этой задаче есть несколько интересных аспектов. Один из них связан с тем, что эта задача лежит на пересечении теории вероятностей и динамических систем. Второй аспект: явление, о котором я хочу рассказать, на самом деле достаточно неожиданное; по крайней мере, если не знать технологию доказательства и если не знать некие результаты из теории вероятностей, то явление будет неожиданным. Давайте я начну с постановки задачи.

Постановка задачи

Для того чтобы увидеть интересные явления, достаточно рассмотреть отображение d -мерного тора в себя. И на самом деле интересные явления уже начинаются в размерности 2. Пусть у нас есть гладкое обратимое отображение $T: \mathbb{T}^d \rightarrow \mathbb{T}^d$, где $d \geq 1$. Его можно взять аналитическим и даже линейным. Пусть обратное тоже будет гладким. (То есть это отображение — диффеоморфизм.) И давайте предположим, что T сохраняет меру Лебега. А именно, когда мы берем d -мерный тор и берем какой-то его кусочек, то образ этого кусочка под действием T имеет ту же самую меру. Таким образом, отображение само по себе никакого перераспределения меры не делает. Теперь давайте рассмотрим функцию $P: \mathbb{T}^d \rightarrow (0, 1)$, которую можно считать случайной средой. Пока что это может звучать немножко таинственно, но сейчас я определю случайное блуждание, и тогда оба объекта, P и T , будут взаимосвязаны.

Случайное блуждание следующее. Допустим, что у нас имеется точка x , и пусть с вероятностью $P(x)$ эта точка прыгает в образ, а с вероятностью $1 - P(x)$ наша точка прыгает в прообраз. И всё это происходит на самом деле на d -мерном торе.

Мы хотим понять, какое вообще поведение такого случайного блуждания. А именно, пусть мы взяли типичную точку, и начались какие-то случайные блуждания этой точки вдоль своей орбиты. Нам интересно,

какое будет поведение этой точки. Например, может ли так случиться, что она начнет застревать в каких-то местах; т. е. попадать в какие-то места на торе, и это будет как болото: она туда попадает, и у нее занимает очень много времени, для того чтобы оттуда выйти.

Формализация вопроса такая: что происходит с распределением типичной точки при больших временах? А именно, вы берете наугад точку на торе, берете ее орбиту. Наше случайное блуждание определено следующим образом: точка попадает в свой образ с вероятностью, равной значению нашей функции в этой точке, и, соответственно, она попадает в прообраз с вероятностью 1 минус значение в этой точке:

$$\xi_n(x) = \begin{cases} T\xi_{n-1}(x) & \text{с вероятностью } P(\xi_{n-1}(x)), \\ T^{-1}\xi_{n-1}(x) & \text{с вероятностью } 1 - P(\xi_{n-1}(x)). \end{cases}$$

Это случайное блуждание можно обозначить $\xi_n(x) \subset \{T_x^n\}$, $n \in \mathbb{Z}$.

Нам хотелось бы понять распределение $\xi_n(x)$, и по определению $\xi_n(x)$ должна лежать в орбите точки x .

Другой, родственный вопрос состоит в том, что вообще происходит: есть ли у процесса, который определен таким образом, хорошие инвариантные меры? Под словом «хороший» я на самом деле имею в виду меры, абсолютно непрерывные по отношению к мере Лебега. И причина, по которой эти два вопроса связаны, состоит в следующем. Пространство мер на d -мерном торе компактно, поэтому, когда время будет увеличиваться, это распределение поневоле может начать сходить к куда-то, т. е. у него могут начать возникать какие-то пределы на d -мерном торе. И мы хотим, например, понять для типичной точки, сходятся ли распределения, которые будут возникать на d -мерном торе, к этим хорошим инвариантным мерам?

Инвариантная мера для блуждания — это вот что такое. Если мы прыгнули один раз по этому правилу, то мера осталась той же. А что формально означает прыгнуть по этому правилу? У нас есть мера с гладкой плотностью $\pi(x)dx$; и правило такое: сколько из точки вышло, столько туда пришло. У нас есть точка, и в начальный момент в этой точке есть $\pi(x)$; в следующий момент из этой точки всё уйдет, потому что вы либо прыгнете в образ, либо прыгнете в прообраз. А что же в эту точку придет? В эту точку придет из прообраза, и в эту точку еще придет из образа. В результате получаем, что мера инвариантна, если она удовлетворяет такому условию:

$$\pi(x) = P(T^{-1}x)\pi(T^{-1}x) + (1 - P(Tx))\pi(Tx).$$

Слева стоит плотность, которая у нас была в начальный момент времени, а справа стоит то, что в эту точку пришло в следующий момент времени; они должны уравниваться.

Оказывается, что в зависимости от того, какая динамика T , ответы могут различаться, и достаточно сильно. И перед тем, как я начну рассказывать про динамический аспект этой задачи, я сначала остановлюсь на достаточно интересном явлении в теории вероятностей, а именно, я расскажу немного про случайные блуждания в так называемых случайных средах.

Сначала я напомним, что такое случайное блуждание, а потом расскажу, что такое случайное блуждание в случайной среде. То есть план доклада будет такой: мы начнем со случайных блужданий на целочисленной решетке, и случайные блуждания будут сначала обычные, и потом будут случайные блуждания в случайной среде. Решетка будет одномерной, потому что в многомерном случае возникает очень много дополнительных трудностей. После того как мы обсудим случайные блуждания, мы обсудим отображения тора. И первый вариант — это просто повороты. И я напишу ответы на поставленные здесь вопросы в случае поворотов. А после этого обсудим случай хаотических отображений, или, если употребить более научный термин, отображений Аносова. Затем я попытаюсь рассказать про формулировки результатов и идеи доказательств. Это наша совместная работа с Я. Г. Синаем. Таков план доклада.

Случайные блуждания

Я начну с обычного случайного блуждания. Представьте себе, что у вас есть целочисленная решетка \mathbb{Z} . В каждый целый момент времени k вы подбрасываете монетку, и если монетка падает орлом, то вы тогда двигаетесь вперед, если монетка падает решкой, вы тогда двигаетесь назад:

$$\eta_k = \begin{cases} 1 & \text{с вероятностью } 1/2, \\ -1 & \text{с вероятностью } 1/2. \end{cases}$$

Случайное блуждание задается так: $\xi_n = \xi_{n-1} + \eta_k$, $\xi_0 = 0$. И вас интересует, что же произойдет с этим случайным блужданием на больших временах.

Прежде всего я напомним, что значит распределение случайной величины. Например, у случайной величины ξ_1 следующее распределение: она с вероятностью $1/2$ равна -1 и с вероятностью $1/2$ равна 1 : $P\{\xi_1 = -1\} = 1/2$, $P\{\xi_1 = 1\} = 1/2$. То есть распределение случайной ве-

личины ξ_1 просто сконцентрировано в двух точках, потому что за один прыжок вы можете попасть либо в единицу, либо в минус единицу.

В общем случае распределением функции называется следующая величина: если у вас есть случайная величина ξ , то ее распределение — это функция $F_\xi(x)$ с таким свойством: $P(a \leq \xi < b) = F_\xi(b) - F_\xi(a)$. Все случайные величины, которые я буду рассматривать, будут иметь целочисленные значения. И распределение случайной величины — это просто функция, по которой можно восстановить все возможные вероятности. Например, самый простой способ записать определение функции распределения такой. Нас интересует, какова вероятность того, что случайная величина попала в интервал между a и b ; или, например, на целочисленной прямой вы берете какой-то интервал, и тогда ваша функция распределения должна быть вероятностью попадания в интервал $[a, b)$.

По мере того как число испытаний n растет, частичка начинает проникать всё дальше и дальше на числовую прямую. И смотреть на это распределение дискретно не очень интересно. И если вы начнете делать соответствующие перенормировки, а именно, делать перемасштабирование, то оказывается, что распределение симметричного случайного блуждания сходится к нормальному распределению. Это так называемая центральная предельная теорема. Она говорит, что вероятность того, что $\frac{\xi_n}{\sqrt{n}}$ меньше a , сходится к нормальному распределению:

$$P\left\{\frac{\xi_n}{\sqrt{n}} < a\right\} \rightarrow \frac{1}{2\pi} \int_{-\infty}^a e^{-\frac{1}{2}x^2} dx.$$

Для нас важно, что при симметричных случайных блужданиях происходит следующее: частица на временах порядка n совершает колебания порядка \sqrt{n} . То есть, если $a \gg 1$, то $P\left\{\frac{\xi_n}{\sqrt{n}} < a\right\} > 0,99$; с вероятностью 99 процентов ваша частичка заключена в интервале порядка \sqrt{n} .

Оказывается, что ситуация очень сильно меняется, если внести очень небольшие изменения в эту модель. Я перехожу к описанию случайных блужданий в случайной среде. Я хочу немного пошевелить вероятности. И способ, которым я хочу немного пошевелить вероятности, такой: я хочу добавить ε_s , где $\xi_{k-1} = s$. То есть я сейчас хочу немного изменить вероятности переходов. И тогда получится так называемое случайное блуждание в случайной среде. И изменения следующие.

До сих пор мы перед каждым движением нашей случайной частички подбрасывали монетку. И если монетка упала орлом, то мы шли вперед, а если она упала решкой, то мы шли назад. И мы считали, что у нас идеальная монетка, а именно, вероятности равны $1/2$. А когда рассмат-

риваются случайные блуждания в случайной среде, то перед тем как начать случайное блуждание, мы в каждой точке на решетке подбрасываем монетку. А именно, давайте рассмотрим нашу случайную среду. Эта случайная среда — набор $\{\varepsilon_k\}$, где $k \in \mathbb{Z}$. Давайте допустим для простоты, что ε_k равно $+\varepsilon$ с вероятностью $1/2$, и оно равно $-\varepsilon$ тоже с вероятностью $1/2$. А именно, теперь мы взяли решетку, и на этой решетке подбрасываем монетку в каждой вершине этой решетки. И теперь распределение вероятностей становится случайным.

Например, если $\xi_{k-1} = s$, то

$$\xi_k = \begin{cases} \xi_{k-1} + 1 & \text{с вероятностью } \frac{1}{2} + \varepsilon_s, \\ \xi_{k-1} - 1 & \text{с вероятностью } \frac{1}{2} - \varepsilon_s. \end{cases}$$

Это достаточно известная модель в теории вероятностей. И нам хотелось бы понять поведение ξ_k при больших k .

Мы уже практически подошли к первому сюрпризу. Оказывается, что если совсем немного пошевелить эти вероятности, то частичка начинает двигаться значительно медленнее. Если раньше она двигалась со скоростью \sqrt{n} , то после того как мы немного пошевелили случайную среду (ε может быть сколь угодно малым), частичка начинает уходить на расстояние только порядка $(\ln n)^2$. Эта теорема была доказана Синаем в 1982 г. Давайте я сформулирую теорему так: если $a \gg 1$, то

$$P \left\{ \left| \frac{\xi_n}{(\ln n)^2} \right| \leq a \right\} > 0,99.$$

Оказывается, что если внести небольшое шевеление в наше распределение вероятностей так, что в каждой отдельной точке распределение вероятностей изменилось не больше чем на ε , то полученное в результате изменение, вообще говоря, очень значительное. Как вы видите, $(\ln n)^2$ значительно меньше, чем \sqrt{n} , когда n становится большим. И в результате получается, что частички начинают гораздо медленнее двигаться, когда возникает случайная среда.

Я хочу еще сказать, что верно и существенно более странное утверждение, доказанное в работе Голосова, которая явилась продолжением работы Синая. Оказывается, что случайное блуждание в случайной среде в некотором смысле не случайно. То есть если мы при каждом шаге чуть-чуть случайно меняем вероятность, то всё меняется. Картина становится совершенно неузнаваема. И если здесь мы видим, что только скорость движения частички замедлилась, то в теореме Голосова ситуация вообще разительно отличается.

Теорема 1 (Голосов, 1984). *Для почти любой реализации случайной среды $\zeta = \{\varepsilon_k\}_{k \in \mathbb{Z}}$ существует такая функция $W_\zeta: \mathbb{Z}_+ \rightarrow \mathbb{Z}$, что при $a \gg 1$ для всех n имеет место неравенство*

$$P\{|\xi_n - W_\zeta(n)| \leq a\} \geq 0,99.$$

Эта теорема говорит следующее. Давайте зафиксируем случайную среду. А именно, перед тем как проводить случайные блуждания, в каждой вершине подбросим монетки и зафиксируем случайную среду. Реализации случайной среды — это набор $\{\varepsilon_k\}$. Если задана реализацию этих ε_n , то тогда можно построить функцию, которая зависит от случайной среды. Эта функция в каждый целый положительный момент времени будет давать целое число. И на самом деле эта функция будет говорить с очень большой точностью в глобальном масштабе, где именно находится случайная величина. А именно, существует такая функция, что при $a \gg 1$ вероятность того, что $|\xi_n - W(n)| \leq a$, не меньше 0,99.

Эта функция говорит с очень большой точностью, где находится случайная величина. Фиксируем a и будем смотреть всё в масштабах $a^{1000000}$. Пусть n порядка $a^{1000000}$. Тогда то, что вы видите, вы просто предсказываете значения случайного блуждания для любого n . Вне зависимости от того, какое время, вы не можете отклониться от детерминистического положения больше, чем на a . Конечно, как всегда в теории вероятностей, есть очень маленькая вероятность того, что это случится, но эта вероятность очень маленькая. На самом деле поведение случайного блуждания имеет такой вид на больших временах. Ваше случайное блуждание прыгает в некоторую точку, а именно, ω , в момент времени 1. После этого она какое-то время колеблется вблизи нее. После этого она перескакивает достаточно быстро в какую-то другую точку, и происходят длительные колебания в этой точке, после этого она переходит куда-то еще и после этого проводит там большое время.

Я еще отмечу, что по теореме Синяя $W_\zeta(n)$ должно быть порядка $(\ln n)^2$. Это означает, что функция будет очень медленно меняться. У вас есть целочисленная функция, у которой абсолютные значения порядка $(\ln n)^2$. И когда вы рисуете график этой функции, то на самом деле первое какое-то количество шагов она будет меняться, но после этого у нее будут такие промежутки, где она будет практически константой, потом она снова будет практически константой, потом она опять будет практически константой. Траектория здесь блуждает, после этого она попадает в болото и в этом болоте она проводит много времени, и это — центр болота. Вы проводите много времени в болоте M_1 , после этого у вас хватает, наконец-то сил, для того чтоб из него выскочить, и вы попадаете в M_2 .

Вы проводите еще больше времени в M_2 , долго там блуждаете, и в конце концов у вас находятся силы для того, чтобы из него выскочить, и вы попадаете в M_3 .

Функция W_ζ вычисляется по реализации. Более того, Кетсен посчитал распределение этой величины. Эта величина очень тесно связана с некоторым функционалом от броуновского движения.

Я сейчас хочу описать механизм застревания. Он не очень сложный. Он в некотором смысле сродни всевозможным локализациям, которые возникают в математической физике. А здесь написана на самом деле локализация. А именно, кажущиеся случайными положения случайного блуждания очень сильно локализованы в определенных местах.

Механизм застревания состоит в том, что случайное блуждание создает некоторые потенциальные колодцы, и эти потенциальные колодцы достаточно долго держат случайную величину внутри. Я сейчас попробую описать эти колодцы. Давайте введем вспомогательную случайную величину

$$\pi_k = \ln \frac{1 + \varepsilon_k}{1 - \varepsilon_k} = \begin{cases} +\rho & \text{с вероятностью } \frac{1}{2}, \\ -\rho & \text{с вероятностью } \frac{1}{2}. \end{cases}$$

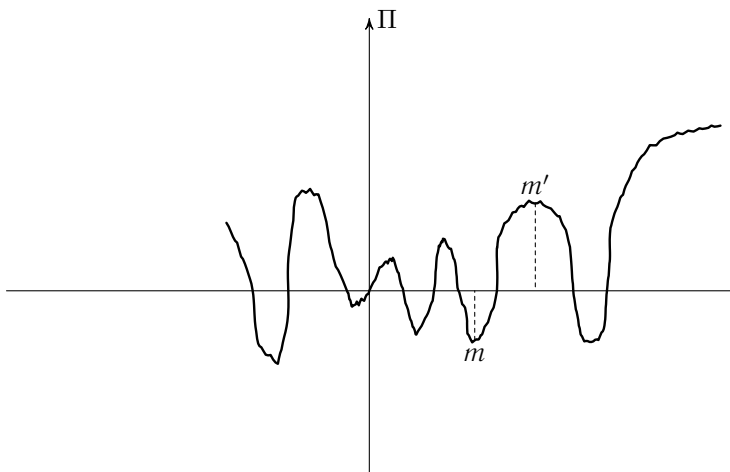
Как вы видите, π_k говорит, какая тенденция в точке k у случайного блуждания: если эта величина больше 0, то частицу сносит вправо, а именно, вероятность уйти вправо больше; если же эта величина меньше 0, то частицу сносит влево. То есть π_k — это в некотором смысле снос. Я напомним, что в точке k с вероятностью $1/2 + \varepsilon_k$ вы попадаете в $k + 1$, а с вероятностью $1/2 - \varepsilon_k$ вы попадаете в $k - 1$. И давайте сейчас введем следующую функцию. У π_k распределение точно такое же, как у ε_k , а именно, оно равно какому-то $+\rho$ с вероятностью $1/2$, и равно $-\rho$ с точно такой же вероятностью. Здесь ρ равно $\ln \frac{1 + \varepsilon}{1 - \varepsilon}$. Сейчас я хочу нарисовать график случайной среды, кумулятивный график. Это значит, что вы набираете историю, начиная от 0 до момента времени m ; т. е. вы кумулятивно собираете все π_k .

Определение следующее:

$$\Pi_m = \begin{cases} - \sum_{k=0}^{m-1} \pi_k, & \text{если } m > 0; \\ 0, & \text{если } m = 0; \\ - \sum_{k=m}^{-1} \pi_k, & \text{если } m < 0. \end{cases}$$

Для положительного m вы просто смотрите на суммарный снос. Напомню, что π_k — это степень сноса в точке траектории. И суммируется кумулятивная степень сноса, где вы набираете снос начиная от 0 до $m - 1$.

График этой функции — не что иное, как случайное блуждание, потому что с вероятностью $1/2$ функция равна $+\rho$, а с вероятностью $1/2$ она равна $-\rho$. Поэтому здесь я в действительности просто подбрасываю монетку и смотрю на сумму. То есть это то же самое случайное блуждание. Так как график этой функции — это то же самое случайное блуждание, то он должен начинаться в нуле, после этого у него есть какие-то такие колебания (рис. 1).



Р и с. 1. График функции

Оказывается, что причина застревания траекторий состоит в том, что из локальных минимумов этого графика очень трудно выходить. Грубо говоря, $W_\zeta(n)$ — это локальный минимум функции Π с характеристиками, зависящими от n . Когда мы находимся на дне колодца, из него очень трудно выходить, потому что, чтобы подняться на соседнюю вершину колодца, нам нужно преодолеть следующую вещь. От точки m до точки m' функция Π меняется, скажем, на величину Δ . Давайте для простоты предположим, что здесь был линейный график. Это означает, что в течение $m' - m$ шагов вероятность пойти вправо была строго меньше $1/2$, а именно, вероятность пойти вправо была равна $1/2 - \varepsilon$. И то, что вам нужно сделать — это возвести $1/2 - \varepsilon$ в достаточно большую степень: $\left(\frac{1}{2} - \varepsilon\right)^{m' - m}$. И причина застревания состоит в том, что вероятность того, что у вас будет много последовательных сносов в другую сторону, значительно больше:

она убывает значительно медленнее, чем растет эта степень. И выбираться из этих колодцев оказывается непреодолимо трудно.

Не очень трудоемкие вычисления показывают, что определение $W(n)$ следующее. Колодец случайной среды ξ высоты N определяется следующим образом. Мы берем первый момент времени, когда максимум Π_k будет больше N , и берем, соответственно, минимальный момент времени, когда минимум Π_k тоже не меньше N :

$$\tau^+ = \min\{k > 0: \Pi_k \geq N\} \sim N^2,$$

$$\tau^- = \min\{k < 0: \Pi_k \geq N\} \sim N^2.$$

Тогда, следуя стандартным оценкам теории вероятностей, это должно быть порядка N^2 , и это должно быть порядка N^2 . Пусть это — наш колодец, и пусть это в точности N . То есть мы останавливаем наш процесс в момент времени τ_- и в момент времени τ_+ , и после этого мы ищем m , которое равно $\min\{\Pi_k: \tau^- < k < \tau^+\}$. Значит, мы смотрим на этом графике на минимум, и вот этот наш минимум. И тогда то, что у нас получилось, это колодец.

То, что здесь нарисовано — не что иное, как график типичной реализации броуновского движения. И мы просто выделяем кусок, где высота нашего колодца над осью x равна в точности N . Мы берем это расстояние и берем из него корень. Поэтому если высота N , то длина порядка N^2 . И нужно определить $W_\zeta(n)$ как минимум колодца глубины больше $\ln n + (\ln n)^{1/2}$.

Последовательность Π_k стремится к бесконечности в обе стороны, потому что броуновское движение начинает совершать осцилляции порядка \sqrt{n} , где n — число шагов. Поэтому будут колодцы, сколь угодно глубокие.

То, что существует τ -функция, следует из того, что броуновское движение должно отклониться в какой-то момент на фиксированное число. Броуновское движение должно это сделать: пьяница в какой-то момент отойдет на 100 шагов от дома.

И оказывается, что эти колодцы захватывают траектории, а причина, по которой они захватывают траектории, состоит в том, что формулы, которые здесь написаны, говорят, что, когда случайное блуждание хочет подняться вверх, оно испытывает очень большой снос в обратную сторону. Поэтому подниматься вверх вдоль этого графика оказывается очень-очень трудно. Соответственно, спускаться вниз достаточно легко. Поэтому на этом графике случайное блуждание делает следующее: оно сначала перейдет сюда, потом у него наконец найдутся силы и оно перескочит сюда, после этого у него наконец найдутся силы, и оно перескочит сюда; после этого оно перескочит, сюда, и после этого оно перескочит туда, и здесь

существенно лучше. Поэтому случайное блуждание просто будет бегать от колодца к колодцу, и причина состоит в том, что стенки настолько крутые, что держаться на стенках очень трудно, вы сползаете всё время в точку минимума.

Тут нарисован один колодец, но он состоит из меньших подколодцев. У колодца должна быть какая-то глубина. Но глубина — это фиксированное число; соответственно, для большого n у вас будет очень большой колодец. Но если вы начнете уменьшать n , то в этом большом колодце будут колодцы поменьше.

В первой половине доклада мы рассматривали вероятностный аспект задачи, а именно, случайные блуждания в случайных средах. Я рассказал о явлении локализации для случайных блужданий в случайных средах на прямой, которое было открыто Я. Г. Синаем и потом доведено до практически полного понимания Голосовым и Кетсенем. Я еще хочу сказать, что распределение того минимума, который я определил, посчитал Кетсен, и известны все распределения.

Случайные блуждания на торе

Теперь я хочу снова вернуться к изначальной постановке задачи. Изначальная постановка задачи следующая. Мы рассматриваем диффеоморфизм (гладкое обратимое отображение) $T: \mathbb{T}^d \rightarrow \mathbb{T}^d$. Допустим, что T сохраняет меру Лебега. Для специалистов хочу сказать, что достаточно, чтобы он просто сохранял гладкую меру. И у нас есть некоторая функция, которая как раз и будет служить нам случайной средой: $P: \mathbb{T}^d \rightarrow (0, 1)$.

Мы решаем задачу про случайные блуждания на d -мерном торе. Точка x с вероятностью $P(x)$ прыгает в образ и, соответственно, с вероятностью $1 - P(x)$ прыгает в прообраз. И нас интересует следующее: 1) распределение типичной начальной точки при больших временах; 2) инвариантные меры этого случайного блуждания. Давайте я напишу формальное определение:

$$\xi_n = \begin{cases} T_{\xi_{n-1}} & \text{с вероятностью } P(\xi_{n-1}(x)), \\ T_{\xi_{n-1}}^{-1} & \text{с вероятностью } 1 - P(\xi_{n-1}(x)). \end{cases}$$

Эта задача отличается от задачи случайных блужданий в случайной среде по одной простой причине: прямая не компактна. Когда мы уходим далеко вправо, то нет никакой связи с тем, что далеко влево, и поэтому, когда у нас происходят случайные блуждания на прямой, у нас нет зависимости с прошлым. Здесь же ситуация другая: вся бесконечная траектория упакована в d -мерный тор, и поэтому убегать нам некуда. Поэтому, мы

куда-то пытаемся убежать, но из-за компактности этого тора мы всё время продолжаем возвращаться.

Я введу еще одно ограничение на этот диффеоморфизм: я хотел бы, чтобы этот диффеоморфизм был эргодический. Давайте я напомним определение. Множество $A \subset \mathbb{T}^d$ инвариантно, если $TA = A$ (образ A совпадает с самим A). И соответственно, отображение T эргодическое, если из того, что $TA = A$ (из инвариантности A) следует, что мера самого A или его дополнения равна 0: $\text{Leb}(A) = 0$ или $\text{Leb}(\bar{A}) = 0$. Требование эргодичности более или менее говорит, что типичная траектория всюду плотно заполняет тор. Если бы так получилось, что тор распадался, скажем, на две части положительной меры, причем одни траектории гуляли бы в одной части, и другие траектории гуляли бы в другой части, то тогда тор можно было бы «разрезать» на эти две части и сосредоточиться на той части, где траектории бегают всюду плотно.

Сейчас я проведу некоторые очень простые вычисления, от которых возникает иллюзия, что инвариантная мера нашего случайного блуждания существует всегда. У нас есть некоторое случайное блуждание, которое перераспределяет меру. На секунду предположим, что есть какая-то точка, которая выталкивает точки с большой силой. Тогда вероятность уйти положительна, а вероятность прийти равна 0. Это означает, что точки выходят из нашей точки, но в нее не возвращаются. Тогда в этом месте у нас не будет меры. Если инвариантная мера имеет плотность $\pi(x)dx$, то тогда свойство инвариантности записывается следующим образом:

$$\pi(x) = (1 - p(Tx))\pi(Tx) + p(T^{-1}x)\pi(T^{-1}x).$$

В момент времени 0 у нас было $\pi(x)$, и в момент времени 1 в точку x приходит из прообраза и приходит из образа. Это то, что я написал в первой части доклада, только там было $Q(T^{-1}(x))\pi(T^{-1}(x))$; T и T^{-1} поменялись ролями. В точку x мы можем попасть из прообраза, прыгая вперед, и мы прыгаем вперед из прообраза с вероятностью $P(T^{-1}(x))$. А еще в точку x мы можем попасть из образа, с вероятностью $1 - P(T(x))$. Это здесь и написано.

То есть инвариантная мера — это такая мера, что если в каждой точке вы разрешили частичке прыгнуть, то после того как она прыгнула, вы вернулись к тому же самому распределению. Нас интересуют хорошие инвариантные меры, например, имеющие плотность. А еще было бы хорошо, чтобы плотность была интегрируемой.

Давайте я проведу некие достаточно простые, но на мой взгляд интересные манипуляции. Давайте разобьем единицу на две части, а именно, на $1 - p(x)$ и $p(x)$: $(1 - p(x))\pi(x) + p(x)\pi(x)$. И после этого произведем

следующее преобразование членов:

$$(1 - p(Tx))\pi(Tx) - p(x)\pi(x) = (1 - p(x))\pi(x) - p(T^{-1}x)\pi(T^{-1}x).$$

Я просто преобразовал уравнение для плотности, разбив единицу на две части и перегруппировав члены.

Давайте обозначим функцию, написанную в правой части, через $F(x)$. Я утверждаю, что в левой части на самом деле написана $F(Tx)$. Давайте заменим x на Tx . Теперь x стало Tx , $T^{-1}x$ стало x . В результате у нас получилась функция, которая инвариантна под действием эргодического отображения. А из-за того, что наше эргодическое отображение всё очень сильно перемешивает, это означает, что эта функция — константа почти всюду. Просто потому что вдоль орбит наша функция должна быть константой. Если орбиты всюду плотны, то это по эргодической теореме дает $F(x) \equiv \text{const}$.

Введем функцию $\nu(x) = p(x)\pi(x)$. Тогда получается следующее:

$$\frac{1 - p(Tx)}{p(Tx)}\nu(Tx) - \nu(x) = \text{const} = C'.$$

Я просто ввел новые обозначения, потому что в этих обозначениях значительно проще всё делать. И мы предполагаем, что наша функция $p(x)$ гладкая и отделена от 0 и от 1, чтобы не было особенностей. А если она отделена от 0 и от 1, то здесь мы просто изменяем плотность на величину порядка константы. Давайте эту константу обозначим c . И теперь можно написать явную формулу для $\nu(x)$. Эта явная формула задается в виде некоторого рекуррентного соотношения. Мы просто суммируем вдоль орбиты. И для того чтобы просуммировать вдоль орбиты, оказывается полезным определение симметричного и несимметричного случайного блуждания. Потому что, если случайное блуждание симметричное, то в среднем частички идут вперед вдоль траекторий, и у них как бы есть скорость течения. И поэтому им значительно труднее застревать; они проскакивают все эти места. Интуитивно это можно представить так: если у вас есть некий снос, то тогда существование хорошей инвариантной меры (а именно, той, у которой есть хорошая плотность) должно быть значительно проще.

Формальное определение следующее. Это то же самое соотношение, которое я писал для π_k :

$$\int \ln \frac{1 - p(x)}{p(x)} dx = 0.$$

Если это условие выполнено, то случайное блуждание называется симметричным. Иначе оно называется несимметричным.

Так вот, в случае несимметричного блуждания создается впечатление, что следующая формула просто даст ответ:

$$\begin{aligned} \nu(x) &= C'' + \frac{1-p(Tx)}{p(Tx)}\nu(Tx) = \\ &= C'' \left(1 + \frac{1-p(Tx)}{p(Tx)} + \frac{1-p(Tx)}{p(Tx)} \cdot \frac{1-p(T^2x)}{p(T^2x)} + \dots \right) = \\ &= C'' (1 + \exp(g(Tx)) + \exp(g(Tx) + g(T^2(x)) + \dots), \end{aligned}$$

где $g(x) = \ln(1 - p(x))/p(x)$. Мы делаем подстановку и получается такое выражение.

Какие есть причины полагать, что тот ряд, который мы здесь пишем, сойдется? Причины следующие. Давайте назовем эту функцию $g(x)$. Здесь написана экспонента от $g(Tx)$, затем у нас получается экспонента от $g(Tx) + g(T^2x)$ и т. д. Поэтому по эргодической теореме сумма $\sum g(T^k x)$ должна сходиться к среднему, а именно, должна сходиться к интегралу от $g(x)dx$, который, как мы предполагаем, равен C , и можно без ограничения общности предположить, что если он не равен 0, то он, например, отрицателен:

$$\frac{1}{m} \sum_{k=1}^m g(T^k x) \rightarrow \int g(x) dx = -C.$$

В результате у нас получается, что m -й член суммы ведет себя приблизительно как $-Cm$, и этот ряд должен экспоненциально сходиться.

Казалось бы, что в произвольном несимметричном случае мы, тем самым, получаем, что плотность существует в почти каждой точке. Но, к сожалению, легко построить пример, когда она является неинтегрируемой. Сейчас я приведу этот пример. Но сначала сформулирую две теоремы.

Теорема 2 (Калошин—Синай). Если T — вращение тора $T: x \rightarrow x + \omega$, и число вращения ω иррационально, то в несимметричном случае гладкая инвариантная мера существует всегда. В симметричном случае гладкая инвариантная мера существует при условии, что ω диофантово.

Число диофантово, если оно не очень быстро аппроксимируется рациональными. Если оно быстро аппроксимируется рациональными, то в рассматриваемом случайном блуждании происходит очень быстрое возвращение к себе. Если вы слишком быстро вернулись, то вы как бы находитесь на периодической орбите. Как бы на всем многообразии вас несет вперед по траекториям, а в каком-то месте вы создали овражек и пустили течение в другую сторону. И вот в этих местах траектории начинают застревать.

В случае, когда ω не диофантово, мы практически доказали с Васамом Фаятом, что гладкой инвариантной меры не существует. Это дает ответ на вопрос о существовании инвариантных мер.

Теорема 3. Для любой точки x на торе \mathbb{T}^d с диофантовым сдвигом ω распределения на больших временах сходятся к гладкой инвариантной мере.

Рассмотрим для простоты следующее отображение Аносова: $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = A$. Эта матрица задает линейное отображение тора $T_A: \mathbb{T}^2 \rightarrow \mathbb{T}^2$. Многие видели картинку отображения кота в книжках Арнольда. Это отображение, которое растягивает тор, а после этого вы его назад складываете. Так появляется отображение тора в себя. Это отображение линейное. Более того, если вы рассмотрите касательное пространство к d -мерному тору (в данном случае 2-мерному тору), то касательное пространство разбивается на две части: $T\mathbb{T}^2 = W^u \oplus W^s$. И если вы посчитаете собственные значения и собственные направления этой матрицы, то, поскольку определитель матрицы равен 1, одно собственное значение будет больше 1, другое собственное значение будет меньше 1. В результате получаем, что W_u — собственное направление с $\lambda_u > 1$, а W_s — собственное направление с $\lambda_s < 1$. И кота растягивают в одном направлении, а в каком-то другом направлении его сжимают.

Оказывается, что если то отображение, которое вы рассматриваете, хаотическое (а именно отображение Аносова), то траектории будут застревать. И они будут застревать очень сильно. Причина для этого застревания лежит как раз в той же самой локализации Синяя—Колосова. Эту причину я попытаюсь объяснить на следующем примере.

Рассмотрим точку 0 на торе. Из-за того, что отображение линейное, точка 0 неподвижная. У нас есть неустойчивое направление (то, в котором идет растяжение) и есть направление, вдоль которого происходит сжатие. Давайте на секунду предположим, что интеграл от сноса положителен:

$$\int \ln \frac{1 - p(x)}{p(x)} dx > 0.$$

Это означает, что движение в обратном направлении (против течения траекторий) доминирует. Снос в сторону T^{-1} . Теперь давайте в этой точке снос направим в обратную сторону. Мы хотим, чтобы нас несло в обратную сторону. Пусть, скажем, $p(0) = 3/4$. Тогда соответственно $1 - p(0) = 1/4$. Что тогда происходит? Наше отображение эргодическое. Это означает, что если мы возьмем точку наугад, то эта точка будет где-то гулять, и в конце концов она подойдет очень близко, потому что она должна гулять везде. Но если она подошла очень близко, то тогда ей нужно двигаться вдоль T^{-1} ,

т. е. нужно всё время в среднем идти в прообразы. Но когда она попадает в окрестность нуля, то эта вероятность ей говорит, что она должна идти в обратном направлении. Будет происходить следующее. Траектории будут подползать к нулю. А если вы попали в ε -окрестность нуля, то вам нужно сделать порядка $\ln 1/\varepsilon$ шагов, для того чтобы выйти из этой окрестности. Действительно, в этой окрестности отображение линейно, и для того чтобы выйти из окрестности порядка ε , нужно порядка $\ln 1/\varepsilon$ шагов. А вероятность того, чтобы сделать $\ln 1/\varepsilon$ шагов в обратном направлении, равна $(1/4)^{\ln 1/\varepsilon}$.

Здесь нужно просто с той формулой, которую я написал для инвариантной меры, провести некоторые несложные вычисления. Механизм состоит в том, что есть какие-то места, где у нас есть контрречение, и это контрречение начинает останавливать траектории, и они начинают здесь застревать всё больше и больше.

Переходя на более научный язык, можно сказать так. Когда у вас есть отображение Аносова, то на больших временах траектории отображения Аносова ведут себя практически как случайное блуждание. Из-за того что они ведут себя практически как случайное блуждание, у нас получается случайное блуждание в случайной среде. И благодаря этому происходит локализация.

Более формальное утверждение говорит, что есть множество сколь угодно малой меры, например меры ε , такое что множество траекторий меры $1 - \varepsilon$ с огромной вероятностью пакуется в множество меры ε . Происходит очень сильная локализация. А именно, почти все траектории начинают застревать с подавляющей вероятностью в множестве малой меры. Они застревают именно в колодцах. Скорость стекания сюда траекторий значительно больше, чем скорость вытекания. И поэтому со временем они начинают всё больше и больше там конденсироваться.

Это соотношение скоростей можно отслеживать. В данном случае это просто тривиально. Вы находитесь в окрестности неподвижной точки линейного отображения; вы берете ε -окрестность этого отображения. И допустим на секунду, что вы пришли на границу этой ε -окрестности. На самом деле нужно оценить не только ε -окрестность, но и оценить окрестность, где контрречение доминирует. Поэтому на самом деле здесь получится не ε -окрестность, а окрестность, скажем, порядка $1/10$. Но дальше вы видите, что когда вы входите в окрестность $1/10$, вы должны в нее входить, потому что есть снос по T^{-1} . Вам нужно сделать, в зависимости от того, насколько близко вы к этой прямой, огромное число шагов: чем ближе, тем дольше. А дальше вы смотрите, сколько шагов нужно сделать с сопротивлением, и смотрите, сколько на это потребуется

времени. И несложные вычисления показывают, что происходит застревание.

Это достаточно красивое явление. Вы блуждаете на двумерном торе и, казалось бы, всё гладкое, но тем не менее в каких-то местах траектории начинают застревать, и происходит локализация.

Сначала мы смотрим глобально на поток. Это случайное блуждание. Есть ли у какая-то хорошая инвариантная мера? И после того как мы ее нашли, мы смотрим, что же происходит с индивидуальными траекториями: уважают они эту инвариантную меру или нет? И ответ состоит в том, что они ее уважают. А в гиперболическом случае инвариантная мера просто не существует.

11 марта 2004 г.

А. А. Г л у ц ю к

УНИФОРМИЗАЦИЯ. РАССЛОЕННЫЕ ВЕРСИИ И НОВОЕ ДОКАЗАТЕЛЬСТВО

Вначале я сформулирую одну из основных теорем, о которых буду говорить: теорему Пуанкаре—Кёбе об униформизации.

Т е о р е м а 1 (Пуанкаре—Кёбе). *Всякая некомпактная односвязная риманова поверхность конформно эквивалентна либо комплексной прямой \mathbb{C} , либо единичному диску.*

Эта глубокая и в высшей степени нетривиальная теорема содержит (в качестве частного случая) теорему Римана об отображении, доказательство которой элементарно и содержится почти во всех учебниках по комплексному анализу.

Т е о р е м а 2 (Риман). *Всякая односвязная область, строго содержащаяся в \mathbb{C} , конформно эквивалентна единичному диску.*

Конформная эквивалентность односвязной римановой поверхности и комплексной прямой или диска (или, вообще, некоторой односвязной области в \mathbb{C}) называется *униформизацией*.

Основная цель доклада — рассказать о различных версиях теоремы Пуанкаре—Кёбе, включая униформизацию почти комплексных структур и некоторых слоений на римановы поверхности (расслоенные версии теоремы Пуанкаре—Кёбе). Одна из них (теорема 5) говорит, что всякий двумерный тор, снабжённый произвольной гладкой почти комплексной структурой, эквивалентен некоторому стандартному комплексному тору. Эта теорема влечёт теорему Пуанкаре—Кёбе с помощью элементарных классических рассуждений, использующих компактность пространства нормированных однолистных голоморфных функций.

В докладе будет дано новое, элементарное доказательство *) теоремы 5 об униформизуемости почти комплексного тора, полученное докладчиком ([7, 8]), и будет выведена теорема Пуанкаре—Кёбе. Оно использует лишь элементарный анализ рядов Фурье, сводящий нахождение униформизации

*) Это доказательство было получено в ходе исследования одной из расслоенных версий (вопроса Э. Жиса, см. [7]; о ней будет рассказано в четвёртой части доклада). Мы приведём его с незначительными упрощениями.

к решению ограниченного линейного уравнения в соболевских пространствах функций на торе.

Доклад будет состоять из пяти частей. Вначале я расскажу о версиях теоремы Пуанкаре—Кёбе для почти комплексных структур (включая теоремы Лаврентьева, Морри, Альфорса и Берса), которые широко используются в разных областях математики, в первую очередь, в голоморфной динамике и теории клейновых групп. Для этого я введу понятие почти комплексной структуры, которое будет использоваться в дальнейшем.

Во второй и четвёртой частях я расскажу о расслоенных версиях теоремы Пуанкаре—Кёбе. Они относятся к некоторым классам слоений на римановы поверхности и утверждают, что униформизация слоёв может быть выбрана «хорошо» зависящей от трансверсального параметра.

Доказательство теоремы 5 об униформизуемости почти комплексного тора будет дано в пятой части. Теорема Пуанкаре—Кёбе будет доказана в третьей части.

1.

Почти комплексные структуры и их униформизация

Чтобы определить, что такое почти комплексная структура, напомним вначале, что задать линейную комплексную структуру на 2-мерной плоскости — это означает задать структуру 1-мерного комплексного линейного пространства. Это можно сделать не только стандартным способом, но и многими другими способами: достаточно задать «оператор умножения на i ». В качестве него можно выбрать произвольный оператор, сопряженный с поворотом на 90° . Сопрягающий оператор, приводящий «оператор умножения на i » к повороту, называется *оператором, задающим линейную комплексную структуру*. Другими словами, это — \mathbb{R} -линейный оператор $A: \mathbb{R}^2 \rightarrow \mathbb{C}$, являющийся \mathbb{C} -линейным в смысле заданной комплексной структуры на плоскости \mathbb{R}^2 . Он единствен с точностью до композиции слева с умножением на комплексное число. Если на плоскости задана ориентация, и сопрягающий оператор её сохраняет, это, в точности, означает, что рассматриваемая комплексная структура задаёт ту же ориентацию.

Почти комплексная структура на ориентированной двумерной поверхности — это семейство линейных комплексных структур (согласованных с ориентацией) в касательных плоскостях.

На римановых поверхностях имеется стандартная (почти) комплексная структура. Сейчас мы увидим, как задать произвольную нестандартную

почти комплексную структуру на римановой поверхности в голоморфной карте относительно стандартной структуры. Для этого вначале я расскажу, как задаётся линейная (нестандартная) комплексная структура в (стандартной) комплексной координате.

Рассмотрим комплексную прямую (со стандартной комплексной структурой и координатой z). Введем на ней нестандартную линейную комплексную структуру, согласованную с ориентацией. Соответствующий оператор $A: \mathbb{C} \rightarrow \mathbb{C}$ имеет вид $z \mapsto az + b\bar{z}$ (это — общий вид \mathbb{R} -линейного оператора $\mathbb{C} \rightarrow \mathbb{C}$). Оператор A является автоморфизмом, сохраняющим ориентацию, и определён с точностью до комплексной гомотетии. Следовательно, $a \neq 0$, и без ограничения общности можно считать, что $a = 1$, т. е. $Az = z + \mu\bar{z}$. Из сохранения ориентации следует, что $|\mu| < 1$.

Пример. Всякая почти комплексная структура на области $U \subset \mathbb{C}$ задается линейной 1-формой $dz + \mu(z)\bar{z}$, где $\mu(z)$ — это функция, принимающая значения в единичном диске.

Функция на почти комплексной поверхности называется *голоморфной*, если её дифференциал \mathbb{C} -линеен в каждой точке (в смысле соответствующей линейной комплексной структуры на касательной плоскости). В условии предыдущего примера это означает, что в каждой точке дифференциал должен быть пропорционален линейной форме $dz + \mu\bar{dz}$. И сразу же находится коэффициент пропорциональности: он равен $\frac{df}{dz}$. Итак, функция голоморфна, если и только если она удовлетворяет уравнению Бельтрами:

$$\frac{\partial f}{\partial \bar{z}} = \mu \frac{\partial f}{\partial z}. \quad (1)$$

Таким образом, задача проинтегрировать почти комплексную структуру, т. е. найти голоморфную карту, сводится к решению уравнения Бельтрами. Оказывается, эта задача всегда решается в достаточно слабых предположениях регулярности почти комплексной структуры (по определению, регулярность почти комплексной структуры совпадает с регулярностью соответствующей функции μ).

Теорема 3. *Всякая почти комплексная структура класса C^∞ на двумерной поверхности интегрируется локально, точнее, всякая точка поверхности имеет окрестность, где существует комплексная координата класса C^∞ , голоморфная в смысле почти комплексной структуры.*

Эта локальная теорема была доказана в более слабых предположениях немецкими математиками Корном [12] и Лихтенштейном [14] в начале XX века. Другое, более простое доказательство было дано Черном [6]. Мы дадим её новое, элементарное доказательство в конце доклада.

Более полный обзор соответствующих результатов см. в [2], [8] и соответствующих библиографиях.

З а м е ч а н и е 1. Всякая риманова метрика на поверхности задаёт некоторую почти комплексную структуру и обратно: всякая почти комплексная структура на поверхности задаётся некоторой римановой метрикой (определённой однозначно с точностью до умножения на функцию). Теорема об изотермических координатах говорит, что всякая риманова метрика в подходящих координатах имеет конформный вид. Это утверждение в точности эквивалентно предыдущей теореме о локальной интегрируемости почти комплексной структуры.

В предыдущей теореме есть два случая. Первый случай: почти комплексная структура является аналитической. В этом случае имеется элементарное доказательство, принадлежащее Гауссу. Оно состоит в комплексификации поверхности и выходе в 2-мерное комплексное пространство. Случай гладкой почти комплексной структуры сильно отличается от аналитического и гораздо более нетривиален, даже в локальной формулировке.

Имеется теорема о глобальной интегрируемости. Прежде чем ее сформулировать, я напомним определение дилатации.

Вначале вернемся к линейным комплексным структурам. С каждой линейной комплексной структурой связан эллипс, который задается уравнением $\|Az\| = 1$; это — прообраз окружности под действием оператора, задающего почти комплексную структуру. По соответствующему эллипсу линейная комплексная структура восстанавливается однозначно.

О п р е д е л е н и е 1. Рассмотрим \mathbb{C} со стандартной комплексной структурой и координатой z . Рассмотрим произвольную другую линейную комплексную структуру σ на $\mathbb{R}^2 = \mathbb{C}$. *Дилатация* структуры σ — это эксцентриситет соответствующего эллипса $\|Az\| = 1$: отношение длин большой полуоси к малой в метрике $|dz|$.

По определению, дилатация стандартной комплексной структуры равна 1.

Пусть теперь имеется риманова поверхность со стандартной комплексной структурой. Выберем на ней другую, нестандартную почти комплексную структуру (пусть μ — соответствующая функция). Дилатация нестандартной структуры определяется в каждой точке как дилатация соответствующей линейной комплексной структуры (в смысле предыдущего определения). Она равна $\frac{1+|\mu|}{1-|\mu|}$. Дилатация ограничена тогда и только тогда, когда $|\mu| < \delta < 1$.

Напомним, что диффеоморфизмы действуют на почти комплексных структурах естественным образом: линейные комплексные структуры в касательных плоскостях преобразуются с помощью производных. По определению, голоморфная функция на почти комплексной поверхности преобразует почти комплексную структуру в стандартную.

Имеет место следующая теорема.

Теорема 4. *Всякая почти комплексная структура на \mathbb{C} класса C^∞ с ограниченной дилатацией преобразуется в стандартную подходящим диффеоморфизмом $\mathbb{C} \rightarrow \mathbb{C}$ класса C^∞ .*

Эта теорема означает в точности, что уравнение Бельтрами (1) на \mathbb{C} с функцией $\mu \in C^\infty(\mathbb{C})$, $|\mu| < \delta < 1$, имеет решение, задающее диффеоморфизм $\mathbb{C} \rightarrow \mathbb{C}$. Она была доказана Лаврентьевым [13] в 1935 г. в более сильной формулировке, когда функция μ является непрерывной.

Следующая теорема является частным случаем предыдущей для двоякопериодических почти комплексных структур. Напомним, что это — одна из основных теорем, которые будут доказаны в докладе.

Теорема 5 ([1]). *Всякая C^∞ -гладкая почти комплексная структура на двумерном торе преобразуется в стандартную структуру на подходящем комплексном торе с помощью подходящего диффеоморфизма класса C^∞ .*

З а м е ч а н и е 2. Классические доказательства теоремы 4 о почти комплексных структурах на \mathbb{C} используют довольно тонкий анализ. Доказательство теоремы 5 для тора, о котором я расскажу, использует только элементарный анализ рядов Фурье и теорему Соболева о вложении. (На самом деле, это доказательство использует теорему 3 о локальной интегрируемости, а точнее, обе теоремы 3 и 5 доказываются одновременно.) С другой стороны, теорема на \mathbb{C} выводится из теоремы на торе с помощью классических рассуждений, см. [8].

Теперь сформулируем более общую теорему Морри, которая имеет очень важные приложения. Для этого мне понадобится ввести некоторые дополнительные определения.

О п р е д е л е н и е 2. Дилатация \mathbb{R} -линейного преобразования комплексной прямой — это эксцентриситет прообраза окружности, или, другими словами, дилатация прообраза стандартной комплексной структуры. *Дилатация* диффеоморфизма римановой поверхности в точке — это дилатация его производной в рассматриваемой точке.

О п р е д е л е н и е 3. Диффеоморфизм римановых поверхностей называется (K -) *квазиконформным*, если его дилатация ограничена (соответственно, не превосходит K). Гомеоморфизм, дифференцируемый почти всюду и имеющий ограниченную дилатацию (константой K) называется

(K -) квазиконформным, если, дополнительно, все его первые частные производные интегрируемы с квадратом (как обобщённые функции).

Теорема 6 (Морри, [15]). *Для любой измеримой почти комплексной структуры на \mathbb{C} с ограниченной (почти всюду) дилатацией существует единственный квазиконформный гомеоморфизм $h: \mathbb{C} \rightarrow \mathbb{C}$, переводящий её в стандартную комплексную структуру и сохраняющий точки 0 и 1. Другими словами, для любой измеримой функции $\mu: \mathbb{C} \rightarrow \mathbb{C}$, $|\mu| < \delta < 1$, существует единственный квазиконформный гомеоморфизм $h_\mu: \mathbb{C} \rightarrow \mathbb{C}$, являющийся решением уравнения Бельтрами (1) в смысле обобщенных функций и удовлетворяющий граничным условиям $h(0) = 0$, $h(1) = 1$.*

Доказательство Морри и последующие классические доказательства его теоремы используют довольно тонкий анализ. (На самом деле, теорема Морри об измеримых почти комплексных структурах может быть выведена из теоремы 4 о гладких почти комплексных структурах с помощью классических рассуждений, использующих равностепенную непрерывность нормированных квазиконформных гомеоморфизмов.) Недавно было опубликовано более простое доказательство, принадлежащее А. Дуади и Кс. Буфу [5], использующее L_2 -анализ и преобразование Фурье на \mathbb{C} .

В определенных кругах математиков теорема Морри называется «теоремой Альфорса—Берса». На самом деле, теорема Альфорса—Берса говорит, что предыдущий квазиконформный гомеоморфизм голоморфно зависит от определяющей его почти комплексной структуры.

Теорема 7 (Альфорс, Берс, [3]). *Пусть ω — это координата на \mathbb{C} , D — (поли)диск, $\mu: D \times \mathbb{C} \rightarrow \mathbb{C}$ — голоморфное по $z \in D$ семейство измеримых функций, $|\mu| < \delta < 1$. Тогда $h_{\mu(z)}$ (соответствующие квазиконформные автоморфизмы) голоморфно зависят от параметра z .*

Обе эти теоремы используются в разных ветвях математики, в первую очередь в голоморфной динамике и в теории клейновых групп. В частности, с помощью теоремы Альфорса—Берса доказывается теорема об одновременной униформизации, которую я сейчас сформулирую.

2.

Теорема Берса об одновременной униформизации

Пусть D — односвязное комплексное многообразие, W — другое комплексное многообразие на 1 большей размерности с голоморфной проекцией $\pi: W \rightarrow D$, так что тройка (W, π, D) является тривиальным C^∞ -рас-

слоением. Будем предполагать, что слои (являющиеся голоморфными кривыми) компактны и имеют род $g \geq 2$. Слой над точкой $z \in D$ будет обозначаться через S_z . Род g — один и тот же для всех слоев, но комплексный тип слоя, как правило, будет меняться. В частности, предыдущее C^∞ -тривиальное расслоение, как правило, не будет биголоморфно тривиальным.

Рассмотрим теперь многообразие \tilde{W} , составленное из универсальных накрывающих \tilde{S}_z слоев S_z и снабжённое естественной комплексной структурой. Многообразие \tilde{W} является универсальной накрывающей над W (односвязность базы) и допускает естественную голоморфную проекцию $\tilde{\pi}: \tilde{W} \rightarrow D$. Её слои \tilde{S}_z — односвязные голоморфные кривые, являющиеся универсальными накрывающими над слоями S_z многообразия W . Тройка $(\tilde{W}, \tilde{\pi}, D)$ также является C^∞ -тривиальным расслоением. Кривые S_z гиперболичны, поэтому, их универсальные накрывающие \tilde{S}_z конформно эквивалентны диску. Более того, можно показать (используя теорему Альфорса—Берса), что тривиализация нового расслоения может быть выбрана голоморфной вдоль слоёв. Но, как правило, её нельзя сделать биголоморфной. В то же время, следующая теорема утверждает существование одновременной биголоморфной униформизации слоёв, но не дисками, а некоторым подходящим семейством односвязных областей в \mathbb{C} .

Теорема 8 (Берс, [4]). *В предыдущих предположениях существует биголоморфный изоморфизм $h: \tilde{W} \rightarrow U \subset \bar{\mathbb{C}} \times D$, образующий коммутативную диаграмму с проекциями на D (U — некоторая односвязная область в $\bar{\mathbb{C}} \times D$).*

Итак, предыдущая многомерная область U слонится над D на одномерные односвязные области $U_z \subset \bar{\mathbb{C}}$, $h(\tilde{S}_z) = U_z \times z$. Оказывается, что можно построить h так, чтобы униформизирующие области U_z были инвариантными относительно квазифуксовых групп, действующих дробно-линейными преобразованиями и голоморфно зависящих от параметра.

Дополнение к теореме Берса. Пусть в предыдущих предположениях g — род кривых S_z . Тогда биголоморфизм из предыдущей теоремы может быть выбран так, чтобы существовали голоморфные функции $a_1, \dots, a_g, b_1, \dots, b_g: D \rightarrow \text{PSL}_2(\mathbb{C})$, такие что для каждого $z \in D$ область U_z инвариантна относительно преобразований $a_i(z), b_i(z)$, $i = 1, \dots, g$. Более того, порождённая ими группа действует свободно и дискретно на U_z , и фактор области U_z по её действию изоморфен кривой S_z .

З а м е ч а н и е 3. Как правило, предыдущие области U_z будут довольно дикими. Точнее, можно добиться того, чтобы область U_z была круглым диском над произвольно выбранной индивидуальной точкой z . Но в типичном случае, области $U_{z'}$, отвечающие подходящим соседним

точкам, будут квазидисками, у которых граница не имеет касательной ни в одной точке. И более того, хаусдорфова размерность границы будет больше 1.

Доказательство теоремы Берса. Мы докажем вначале локальную версию теоремы Берса, утверждающую, что каждая точка $z_0 \in D$ имеет окрестность, над которой существует биголоморфизм h из теоремы Берса. Затем мы докажем теорему Берса в общем случае.

Мы построим послойную униформизацию

$$H_z: \tilde{S}_z \rightarrow U_z \subset \bar{\mathbb{C}}$$

следующим образом. Фиксируем точку $z_0 \in D$. Универсальная накрывающая \tilde{S}_{z_0} конформно эквивалентна единичному диску D_1 (фиксируем конформную эквивалентность). Для каждого слоя S_z мы выберем подходящий диффеоморфизм $\phi_z: S_{z_0} \rightarrow S_z$. Рассмотрим соответствующий прообраз стандартной комплексной структуры на S_z : это — почти комплексная структура на S_{z_0} (обозначим её σ_{ϕ_z}). Поднимем её на универсальную накрывающую $\tilde{S}_{z_0} = D_1$. Продолжим поднятую структуру на всю сферу Римана, содержащую D_1 , полагая её стандартной во внешности единичного диска. Продолженную структуру на $\bar{\mathbb{C}}$ обозначим $\bar{\sigma}_{\phi_z}$.

Фиксируем три различные точки $a, b, c \in \partial D_1$ (раз и навсегда, независимо от z и ϕ_z). По теореме Морри, существует (единственный) квазиконформный гомеоморфизм $h_{\phi_z}: \bar{\mathbb{C}} \rightarrow \bar{\mathbb{C}}$, переводящий структуру $\bar{\sigma}_{\phi_z}$ в стандартную, а точки a, b, c — в $0, 1$ и ∞ , соответственно.

Отметим точку $w_0 \in S_{z_0}$ (независимо от z и ϕ_z). Для выбранного диффеоморфизма ϕ_z отметим точку $\phi_z(w_0) \in S_z$. Рассмотрим универсальные накрывающие соответствующих слоёв как построенные по выбранным отмеченным точкам: при этом на каждой универсальной накрывающей отмечается точка, отождествляемая с соответствующей отмеченной точкой на слое. Рассмотрим поднятие $\tilde{\phi}_z: \tilde{S}_{z_0} \rightarrow \tilde{S}_z$ диффеоморфизма ϕ_z , переводящее одну отмеченную точку в другую. По построению, диффеоморфизм $\tilde{\phi}_z$ переводит структуру $\bar{\sigma}_{\phi_z}$ в стандартную. Тем самым, композиция

$$H_z = h_{\phi_z} \circ \tilde{\phi}_z^{-1}: \tilde{S}_z \mapsto U_z \subset \bar{\mathbb{C}}, \quad U_z = h_{\phi_z}(D_1),$$

является квазиконформным гомеоморфизмом, сохраняющим стандартную комплексную структуру. Следовательно, $H_z: \tilde{S}_z \rightarrow U_z$ — конформное отображение (локальная теорема о единственности). \square

О п р е д е л е н и е 4. Предыдущее отображение $H_z: \tilde{S}_z \rightarrow U_z \subset \bar{\mathbb{C}}$ называется *униформизацией* слоя \tilde{S}_z , ассоциированной с диффеоморфизмом ϕ_z .

3.

Локальная одновременная униформизация

Фиксируем произвольную точку $z_0 \in D$. Покажем, что существует такая её окрестность $V \subset D$, над которой теорема Берса справедлива: существует биголоморфизм области $\bar{\pi}^{-1}(V)$ на область в $\bar{\mathbb{C}} \times V$, согласованный с проекцией. Для этого проведём локальную голоморфную трансверсаль к отмеченному слою S_{z_0} через каждую его точку, так чтобы трансверсаль бесконечно гладко зависела от соответствующей точки. Уменьшая трансверсали, можно добиться того, чтобы они не пересекались и проектировались на одну и ту же область $V \subset D$ (компактность слоя). Голономия вдоль трансверсалей задаёт семейство C^∞ -диффеоморфизмов $\phi_z: S_{z_0} \rightarrow S_z$, голоморфно зависящее от $z \in V$. Пусть $H_z: \tilde{S}_z \rightarrow U_z$ — униформизации слоёв, ассоциированные с диффеоморфизмами ϕ_z . Искомый биголоморфизм из теоремы Берса есть отображение

$$H: x \in \pi^{-1}(V) \mapsto (H_{\pi(x)}(x), \pi(x)).$$

Его инъективность и согласованность с проекциями очевидны. Докажем его биголоморфность. По построению, отображение H голоморфно вдоль слоёв. Покажем, что оно голоморфно вдоль трансверсалей. Это вместе с теоремой Хартогса влечёт глобальную голоморфность (впрочем, здесь можно и не использовать теорему Хартогса, а напрямую доказать гладкость отображения H : отсюда и из предыдущих утверждений будет следовать голоморфность).

Прообраз $\pi^{-1}(V)$ диффеоморфно (но, вообще говоря, не голоморфно) параметризуется прямым произведением:

$$\psi: D_1 \times V \rightarrow \pi^{-1}(V), \quad \psi: (\omega, z) \mapsto (\phi_z(\omega), z).$$

Трансверсали к S_{z_0} параметризуются горизонталями $\omega \times V$ с сохранением голоморфной координаты z . Отображение H записывается в новых координатах в виде

$$H \circ \psi: (\omega, z) \mapsto (h_{\phi_z}(\omega), z).$$

Значение $h_{\phi_z}(\omega)$ голоморфно по z (теорема Альфорса—Берса и голоморфность по z почти комплексной структуры $\tilde{\sigma}_z$). Голоморфность семейства структур $\tilde{\sigma}_z$ следует из голоморфности по z определяющего его семейства диффеоморфизмов ϕ_z . Следовательно, отображение $H \circ \psi$ голоморфно по z вдоль горизонталей. Тем самым, отображение H голоморфно вдоль трансверсалей. Биголоморфность отображения H и локальная теорема Берса доказаны.

Глобальная униформизация. Доказательство теоремы Берса в общем случае основано на следующей лемме о единственности униформизации.

Лемма 1. Пусть S_{z_0} и S_z — римановы поверхности одинакового рода $g \geq 2$. Униформизация поверхности S_z , ассоциированная с данным диффеоморфизмом $S_{z_0} \rightarrow S_z$, зависит только от его класса изотопии.

Определим теперь глобальную униформизацию из теоремы Берса на каждом слое. Фиксируем произвольную точку $z_0 \in D$ и отображим конформно универсальную накрывающую \tilde{S}_{z_0} на единичный диск. Выберем произвольную другую точку $z \in D$ и соединим её с z_0 некоторым путём

$$\gamma: [0, 1] \rightarrow D, \quad \gamma(0) = z_0, \quad \gamma(1) = z.$$

Поднимем путь γ до непрерывного семейства диффеоморфизмов $\phi_t: S_{z_0} \rightarrow S_{\gamma(t)}$ (произвольным образом), $\phi_0 = Id$. Пусть $H_z: \tilde{S}_z \rightarrow U_z \subset \mathbb{C}$ — униформизация слоя \tilde{S}_z , ассоциированная с диффеоморфизмом ϕ_1 . Она не зависит от выбора пути и семейства диффеоморфизмов. Это следует из предыдущей леммы о единственности и из того, что класс изотопии диффеоморфизма ϕ_1 также не зависит от их выбора (односвязность базы и свойство накрывающей гомотопии).

Итак, на каждом слое построена своя униформизация. Покажем, что она голоморфна не только вдоль слоёв, но и вдоль подходящих трансверселей к слоям. Это вместе с теоремой Хартогса докажет её глобальную голоморфность на \tilde{M} и, тем самым, теорему Берса. Рассмотрим произвольную точку $z \in D$ и проведём ровно одну голоморфную трансверсаль к слою S_z через каждую его точку, так чтобы трансверсали C^∞ -гладко зависели от параметра слоя и проектировались на одну и ту же окрестность V точки z (как при доказательстве локальной теоремы Берса). Голономия вдоль трансверселей индуцирует семейство диффеоморфизмов $\psi_{z'}: S_z \rightarrow S_{z'}$, голоморфно зависящее от параметра $z' \in V$. Фиксируем диффеоморфизм $\phi = \phi_1: S_{z_0} \rightarrow S_z$, построенный выше. Композиция $\psi_{z'} \circ \phi: S_{z_0} \rightarrow S_{z'}$ также голоморфно зависит от z' . По построению и лемме о единственности, ассоциированная с ней униформизация слоя $\tilde{S}_{z'}$ совпадает с его униформизацией $H_{z'}$, определённой выше (по, вообще говоря, другому диффеоморфизму, но изотопному предыдущей композиции). Униформизация $H_{z'}$ голоморфна вдоль предыдущих трансверселей, как и в доказательстве локальной теоремы Берса. Голоморфность и теорема Берса доказаны по модулю леммы о единственности.

Доказательство леммы о единственности. Основным его шагом является следующая лемма о квазиконформности.

Лемма 2. Пусть S — компактная риманова поверхность рода $g \geq 2$. Отождествим её универсальную накрывающую \tilde{S} с единичным диском. Всякое поднятие на \tilde{S} любого квазиконформного гомеоморфизма $\psi: S \rightarrow S$ (например, диффеоморфизма) продолжается до квазиконформного отображения замкнутого диска.

Доказательство леммы опущено для простоты изложения.

Пусть $\phi, \phi': S_{z_0} \rightarrow S_z$ — два изотопных диффеоморфизма. Пусть σ, σ' — соответствующие прообразы стандартной комплексной структуры кривой S_z . Это — почти комплексные структуры на S_{z_0} . Поднимем их с S_{z_0} на $D_1 = \tilde{S}_{z_0}$ и обозначим поднятые почти комплексные структуры символами $\bar{\sigma}$ и $\bar{\sigma}'$, соответственно. Продолжим их на сферу $\bar{\mathbb{C}}$, полагая их стандартными вне диска D_1 . Пусть $\bar{\phi}, \bar{\phi}': D_1 \rightarrow \tilde{S}_z$ поднятия диффеоморфизмов ϕ и ϕ' , соответственно, построенные как в начале доказательства теоремы Берса, $i: D_1 \rightarrow \bar{\mathbb{C}}$ — стандартное включение. Для доказательства леммы достаточно показать, что

$$h_\phi \circ i \circ \bar{\phi}^{-1} = h_{\phi'} \circ i \circ (\bar{\phi}')^{-1},$$

или, другими словами, коммутативность диаграммы

$$\begin{array}{ccccc} (\bar{\mathbb{C}}, \bar{\sigma}) & \xrightarrow{h_\phi} & \bar{\mathbb{C}} & \xleftarrow{h_{\phi'}} & (\bar{\mathbb{C}}, \bar{\sigma}') \\ \uparrow i & & & & \uparrow i \\ (D_1, \bar{\sigma}) & \xrightarrow{\bar{\phi}} & \tilde{S}_z & \xleftarrow{\bar{\phi}'} & (D_1, \bar{\sigma}') \\ \downarrow & & \downarrow & & \downarrow \\ (S_{z_0}, \sigma) & \xrightarrow{\phi} & S_z & \xleftarrow{\phi'} & (S_{z_0}, \sigma'). \end{array}$$

Итак, достаточно показать, что $h_{\phi'}^{-1} \circ h_\phi = (\bar{\phi}')^{-1} \circ \bar{\phi}$. Заметим, что левая часть предыдущего равенства — квазиконформный гомеоморфизм сферы, фиксирующий три заданные точки на единичной окружности. Это следует из построения и группового свойства квазиконформных гомеоморфизмов. Покажем, что правая часть продолжается до квазиконформного гомеоморфизма сферы, для которого целая единичная окружность состоит из неподвижных точек. Это вместе с утверждением теоремы Морри о единственности докажет предыдущее равенство.

Предыдущая правая часть — это диффеоморфизм единичного диска, непрерывно продолжающийся тождественным гомеоморфизмом на границу. Это следует из построения и из того, что диффеоморфизм $(\phi')^{-1} \circ \phi: S_{z_0} \rightarrow S_{z_0}$ изотопен тождественному (так как ϕ и ϕ' изотопны).

Предыдущее продолжение диффеоморфизма — квазиконформный гомеоморфизм замкнутого диска (лемма о квазиконформности), тождественный на границе. Следовательно, его продолжение на всю сферу Римана (определённое как тождественное вне диска) также квазиконформно и тождественно на единичной окружности. Лемма о единственности доказана. Теорема Берса полностью доказана. \square

Доказательство дополнения к теореме Берса (о голоморфном семействе квазифуксовых групп). Напомним определение униформизации слоёв \tilde{S}_z . Мы отождествили каждый слой с диском $D_1 = \tilde{S}_{z_0}$, снабжённым подходящей почти комплексной структурой, и продолжили последнюю на всю сферу Римана как стандартную в дополнении к диску. Униформизация слоя \tilde{S}_z совпала с ограничением на D_1 квазиконформного гомеоморфизма сферы, переводящего продолженную структуру в стандартную. Рассмотрим действие фундаментальной группы отмеченного слоя S_{z_0} на накрывающем слое $\tilde{S}_{z_0} = D_1$. Это действие продолжается до действия конформными автоморфизмами сферы Римана (по принципу симметрии). Предыдущие почти комплексные структуры на сфере, с помощью которых строились униформизации слоёв, инвариантны относительно этого действия. Это следует из построения. Поэтому предыдущий квазиконформный гомеоморфизм сопрягает предыдущее действие на почти комплексных сферах с действием той же группы квазиконформными гомеоморфизмами стандартной сферы, сохраняющими стандартную комплексную структуру. Тем самым, последнее (сопряжённое) действие является конформным (утверждение о единственности из теоремы Морри). Из построения следует его голоморфность по параметру z . Для каждого $z \in D$ область U_z инвариантна относительно соответствующего действия, и ограничение последнего на U_z конформно сопряжено с действием фундаментальной группы слоя S_z . Это тоже следует из построения и доказывает дополнение к теореме Берса. \square

Еще я хочу сказать, что имеется обобщение теоремы об одновременной униформизации, принадлежащее Юлию Сергеевичу Ильяшенко [11], для случая, когда имеется один особый слой с простыми двойными особыми точками. В этом случае соответствующее голоморфное семейство групп дробно-линейных преобразований состоит не из квазифуксовых групп, как в теореме Берса (с двумя инвариантными односвязными областями), а из так называемых невырожденных Б-групп, каждая из которых имеет ровно одну односвязную инвариантную область. В более общем случае, когда имеется не один а несколько особых слоев, биголоморфной униформизации, вообще говоря, не существует [9].

Схема доказательства теоремы Пуанкаре—Кёбе (по статье [8]).

Пусть S — односвязная некомпактная риманова поверхность. Для доказательства теоремы Пуанкаре—Кёбе достаточно установить существование однолистной голоморфной функции $S \rightarrow \mathbb{C}$. Это сведёт её к теореме Римана об отображении, применённой к образу функции.

Первый (и основной) шаг доказательства состоит в существовании однолистной голоморфной функции на произвольном компакте.

Лемма 3. Всякий компакт в S допускает однолистную голоморфную функцию.

Второй шаг — исчерпывание поверхности S компактами $D_1 \subset D_2 \subset \dots \subset S$ (гомеоморфными диску) и рассмотрение соответствующих однолистных функций $h_n: D_n \rightarrow \mathbb{C}$ (существующих по предыдущей лемме). Доказывается сходимость подпоследовательности подходящим образом нормированных функций h_n к искомой однолистной функции на S .

Первый шаг: доказательство леммы 3. Оно основано на теореме 5 об униформизуемости почти комплексного тора (которая, напомним, будет доказана в конце доклада). Для этого представим рассматриваемую риманову поверхность S (которая диффеоморфна плоскости) как плоскость \mathbb{C} , снабжённую нестандартной бесконечно гладкой почти комплексной структурой σ : фиксируем произвольный бесконечно гладкий диффеоморфизм $S \rightarrow \mathbb{C}$ и рассмотрим образ σ стандартной комплексной структуры. (Почти комплексная структура σ , вообще говоря, не обязана иметь ограниченную дилатацию.)

Достаточно доказать, что на любом (замкнутом) диске $K \subset \mathbb{C}$ существует однолистная функция, голоморфная в смысле структуры σ . В случае, если последняя двоякопериодична относительно некоторой двумерной решётки, утверждается большее: вся рассматриваемая риманова поверхность $S = (\mathbb{C}, \sigma)$ конформно эквивалентна \mathbb{C} . Действительно, фактор по решётке периодов структуры есть почти комплексный тор. По теореме 5, он конформно эквивалентен некоторому стандартному комплексному тору. Искомая конформная эквивалентность $S \rightarrow \mathbb{C}$ получается поднятием предыдущей на универсальные накрывающие.

Пусть теперь структура σ не двоякопериодична. Фиксируем ограниченную окрестность $V \supset K$ диска K . Выберем произвольную (например, ортогональную) решётку периодов на \mathbb{C} так, чтобы некоторый прямоугольник периодов (фундаментальный прямоугольник) содержал \bar{V} . Теперь фиксируем ограничение на V почти комплексной структуры σ и продолжим его до гладкой двоякопериодической структуры σ' на \mathbb{C} с той же решёткой периодов. Например, можно вначале продолжить $\sigma|_V$ вплоть до границы прямоугольника периодов, так чтобы новая, продолженная структура

была стандартной в некоторой окрестности границы, а затем продолжить периодически на всю плоскость, замощённую копиями рассматриваемого прямоугольника. По построению, новая структура σ' на \mathbb{C} двоякопериодична и совпадает со старой на K . Применяя предыдущие рассуждения к новой структуре σ' , получаем конформное отображение $(\mathbb{C}, \sigma') \rightarrow \mathbb{C}$. Его ограничение на компакт K , где обе структуры совпадают, даёт искомую однолиственную голоморфную функцию $(K, \sigma) \rightarrow \mathbb{C}$. Лемма 3 доказана. \square

Доказательство второго шага. Пусть $D_n, h_n: D_n \rightarrow \mathbb{C}$ — определённые выше исчерпывающие компакты, гомеоморфные диску, и однолистные голоморфные функции. Покажем, что из подходящим образом нормированных функций h_n можно выбрать подпоследовательность, равномерно сходящуюся на каждом компакте в S к непостоянной функции. Тогда, по построению, предел — искомая однолистная голоморфная функция $S \rightarrow \mathbb{C}$ (так как он не постоянен и сами функции h_n однолисны). Доказательство возможности выбора сходящейся подпоследовательности использует только однолистность функций h_n и нормальность семейства нормированных однолистных функций (здесь *нормальность* — это равномерная ограниченность производных на каждом компакте). Напомним, что из всякой нормальной последовательности голоморфных функций всегда можно выделить подпоследовательность, равномерно сходящуюся на компактах (это следует из теоремы Арцела—Асколи). \square

Предложение 9. Пространство однолистных функций

$$f: D_1 \rightarrow \mathbb{C}$$

на единичном диске, таких что $f(0) = 0$ и $f'(0) = 1$, — нормально.

Доказательство предложения будет опущено для простоты: оно может быть найдено в учебниках по комплексному анализу.

Теперь для нормировки предыдущих функций h_n фиксируем произвольную точку внутри наименьшего компакта D_1 : обозначим ее через 0 и фиксируем локальную голоморфную координату z в её окрестности, $z(0) = 0$. Нормируем функции h_n (применяя к ним подходящие аффинные преобразования) так, чтобы $h_n(0) = 0$ и $\frac{dh_n}{dz}(0) = 1$. Утверждается, что полученная последовательность нормальна на S (а отсюда, как отмечено выше, вытекает существование сходящейся подпоследовательности; её предел не постоянен, что следует из условий нормировки).

Действительно, последовательность h_n нормальна внутри каждого фиксированного компакта D_j . Это следует из предыдущего предложения, нормировки и из того, что D_j конформно эквивалентен диску: эквивалентность осуществляется функцией $h_j|_{D_j}$ и последующим применением отображения Римана к её образу. (Напомним, что по построению, каждый

компакт D_j гомеоморфен диску.) Компакты D_j исчерпывают S , поэтому, последовательность h_n нормальна на всей поверхности S . Второй шаг доказан. Теорема Пуанкаре—Кёбе доказана по модулю теоремы 5 об униформизуемости почти комплексного тора.

4.

Метрическая одновременная униформизация слоений (по статье [7])

Соответствующие результаты относятся к слоениям на римановы поверхности. Более подробно, рассмотрим вещественное компактное многообразие M , на котором задано слоение на двумерные поверхности. Это означает, что задано интегрируемое двумерное распределение: каждая точка имеет окрестность, расслаивающуюся на двумерные диски и послойно диффеоморфную прямому произведению двумерного диска на диск дополнительной размерности; склейки между пересекающимися картами прямых произведений уважают разбиения на диски. Это задаёт разбиение всего многообразия на несамопересекающиеся и попарно непересекающиеся поверхности, называемые *листами* (которые, вообще говоря, не являются замкнутыми подмногообразиями и, например, могут быть всюду плотными, как, например, двумерная иррациональная обмотка тора). Будем считать, что каждый лист снабжён структурой римановой поверхности, которую можно задать, например, следующим образом. Пусть g — гладкая метрика на M . Её ограничение на каждый лист задаёт почти комплексную структуру на нём. Эта почти комплексная структура конформно эквивалентна некоторой римановой поверхности (в силу теоремы об интегрируемости из первой части доклада). Нас будет интересовать случай, когда все листы — параболические римановы поверхности, т. е. их универсальные накрывающие, снабжённые поднятой комплексной структурой, конформно эквивалентны \mathbb{C} .

З а м е ч а н и е 4. Конформный тип универсальной накрывающей листа не зависит от выбора метрики. Действительно, фиксируем метрику g_0 (и соответствующую комплексную структуру, которую мы назовём стандартной). Любая другая метрика g задаёт (нестандартную) структуру, имеющую равномерно ограниченную дилатацию (компактность объемлющего многообразия). Отсюда и из теоремы 4 следует утверждение замечания.

На каждой параболической поверхности имеется единственная (с точностью до константы) плоская полная конформная метрика (плоская = локально евклидова = гауссова кривизна равна нулю). На \mathbb{C} — это стандартная евклидова метрика $|dz|^2$. Поэтому на каждом индивидуальном

листе L существует функция $\phi: L \rightarrow \mathbb{R}_+$, такая что метрика ϕg листа является плоской и полной.

Э. Жис (É. Ghys) предложил следующий вопрос:

В о п р о с. Верно ли, что в предыдущих предположениях существует глобальная бесконечно гладкая функция $\phi: M \rightarrow \mathbb{R}_+$, такая что ограничение на каждый лист метрики ϕg является плоским? (Полнота следует из компактности.) Следует исключить слоения (типа рибовского), где на объемлющем многообразии не существует ни одной гладкой метрики, плоской вдоль листов (не обязательно конформной).

Заметим, что предыдущая функция ϕ на каждом индивидуальном листе единственна с точностью до константы. Вопрос может быть переформулирован следующим образом: можно ли на каждом листе выбрать свою нормирующую константу так, чтобы соответствующие функции ϕ были гладки не только вдоль листов, но и по трансверсальному направлению?

В статье [7] доказаны положительные и отрицательные результаты и дан более детальный исторический обзор. В настоящем докладе мы сформулируем основной положительный результат о линейных слоениях на торе с произвольной гладкой метрикой.

Рассмотрим n -мерный тор $\mathbb{T}^n = \mathbb{R}^n / 2\pi\mathbb{Z}^n$ и рассмотрим на \mathbb{R}^n какое-нибудь слоение на 2-мерные параллельные плоскости. То есть, другими словами, возьмем какую-нибудь линейную вектор-функцию коранга 2 и рассмотрим слоение на её плоскости уровня. Это индуцирует некоторое двумерное слоение на торе, которое мы будем называть *линейным слоением*. В типичном случае, это — 2-мерная иррациональная обмотка. Пусть g — риманова метрика класса C^∞ на торе. Она индуцирует семейство почти комплексных структур на листах. Все листы параболичны в силу предыдущего замечания и поскольку стандартная евклидова метрика индуцирует на слоях-плоскостях стандартную комплексную структуру.

Теорема 10 (см. [7]). *Для любого двумерного линейного слоения на торе \mathbb{T}^n в любой размерности и для любой C^∞ -гладкой метрики g на торе существует C^∞ -гладкая функция $\phi: \mathbb{T}^n \rightarrow \mathbb{R}_+$, такая что ограничение на каждый лист метрики ϕg плоско.*

Ранее Жис [10] доказал частный случай теоремы 10 в размерности 3 и при дополнительном предположении, что линейное слоение удовлетворяет диофантову условию. Оказывается, теорема верна в общем случае, даже если не налагать диофантова условия.

Наметим доказательство теоремы 10. Для этого выберем разложение тора в прямое произведение $\mathbb{T}^n = \mathbb{T}^2 \times \mathbb{T}^{n-2}$ так, чтобы проекция листов

на первый 2-мерный тор была локально взаимно однозначной. Пусть z — это комплексная координата на 2-мерном торе.

Метрика g задаёт семейство почти комплексных структур на листах, которое, в свою очередь, задается дифференциальной 1-формой

$$\omega_\mu = dz + \mu(x)d\bar{z}, \quad \mu: \mathbb{T}^n \rightarrow \mathbb{C}, \quad \mu \in C^\infty, \quad |\mu| < 1,$$

где z — координата на листе, поднятая с двумерного тора. Без ограничения общности будем считать, что $g = |\omega_\mu|^2$: этого можно добиться, умножая метрику на константу. Теперь нахождение функционального множителя ϕ сводится к решению уравнения Бельтрами на каждом листе. Точнее, мы ищем дифференциал униформизирующего отображения в виде $f\omega_\mu$.

Л е м м а 4. Для любого линейного слоения на торе и любой функции μ вышеуказанного типа существует функция $f: \mathbb{T}^n \rightarrow \mathbb{C} \setminus 0$, $f \in C^\infty$, такая что ограничение 1-формы $f\omega_\mu$ на каждый лист слоения замкнуто.

Искомая конформная метрика, плоская на листах, есть $|f|^2|\omega_\mu|^2$, т. е. $\phi = |f|^2$.

Для простоты изложения, мы опустим доказательство самой леммы, а докажем следующую её упрощённую версию, которая, как мы увидим, влечёт униформизуемость почти комплексного тора. Доказательства обеих лемм дословно повторяют друг друга с очевидными изменениями.

5.

Униформизуемость почти комплексного тора: доказательство теоремы 5 (по статьям [7] и [8])

Обозначим $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$.

Л е м м а 5. Для любой функции $\mu: \mathbb{T}^2 \rightarrow \mathbb{C}$, $\mu \in C^\infty$, $|\mu| < 1$ (обозначим $\omega_\mu = dz + \mu d\bar{z}$), существует другая C^∞ -функция $f: \mathbb{T}^2 \rightarrow \mathbb{C} \setminus 0$, такая что 1-форма $f\omega_\mu$ замкнута.

Вначале мы выведем теорему 5 из предыдущей леммы, а затем докажем последнюю. Пусть $\mu: \mathbb{T}^2 \rightarrow \mathbb{C}$ — функция, задающая почти комплексную структуру на \mathbb{T}^2 , f — соответствующая функция из предыдущей леммы. Поднимем форму $f\omega_\mu$ с тора на его универсальную накрывающую \mathbb{C} (поднятую форму обозначим тем же символом $f\omega_\mu$). Рассмотрим отображение $F: \mathbb{C} \rightarrow \mathbb{C}$, заданное её интегралом:

$$z \rightarrow F(z) = \int_0^z f\omega_\mu.$$

Это — диффеоморфизм, что следует из односвязности плоскости \mathbb{C} и из того, что F — неразветвлённое накрытие (утверждение $f \neq 0$ и компактность тора).

Рассмотрим почти комплексную структуру на \mathbb{C} , заданную поднятой функцией μ (обозначим эту структуру через σ). Диффеоморфизм F переводит её в стандартную комплексную структуру. По построению исходная структура σ инвариантна относительно действия группы \mathbb{Z}^2 сдвигами по решётке периодов тора; ни один сдвиг не имеет неподвижных точек. Поэтому диффеоморфизм F сопрягает эти сдвиги с конформными автоморфизмами плоскости без неподвижных точек, которые, следовательно, сами являются сдвигами. Итак, F сопрягает предыдущее действие группы \mathbb{Z}^2 сдвигами почти комплексной плоскости с её действием сдвигами на \mathbb{C} . Опуская F на соответствующие факторы, получаем искомый диффеоморфизм рассматриваемого почти комплексного тора на некоторый стандартный комплексный тор.

Доказательство леммы 5. Первый шаг:

Предложение 11. Для любой функции $\mu: \mathbb{T}^2 \rightarrow \mathbb{C}$, удовлетворяющей условиям предыдущей леммы, существует функция $f: \mathbb{T}^2 \rightarrow \mathbb{C}$ класса C^∞ , не обращающаяся тождественно в нуль (но, возможно, не обязательно нигде не обращающаяся в нуль), такая, что форма $f\omega_\mu$ замкнута.

Затем, на втором шаге, мы покажем, что, на самом деле f нигде не обращается в нуль. Доказательство второго шага основано на теореме 3 о локальной интегрируемости гладкой почти комплексной структуры и использует элементарные топологические рассуждения. Теорема о локальной интегрируемости будет следовать из первого шага и элементарных оценок.

Доказательство предложения 11. Замкнутость формы $f\omega_\mu$ эквивалентна следующему уравнению в частных производных:

$$\frac{\partial f}{\partial \bar{z}} = \frac{\partial(\mu f)}{\partial z}. \quad (1)$$

Чтобы его решить, мы сведём его к ограниченному линейному уравнению в соболевских пространствах. Для этого применим к обеим частям «обратный оператор» $\left(\frac{\partial}{\partial \bar{z}}\right)^{-1}$. Получим уравнение

$$f = \left(\frac{\partial}{\partial \bar{z}}\right)^{-1} \circ \frac{\partial}{\partial z}(\mu f).$$

Оператор $\left(\frac{\partial}{\partial \bar{z}}\right)^{-1}$ определён не на всех гладких функциях (только на функциях с нулевым средним), а его значение на данной функции опреде-

лено не однозначно, а с точностью до аддитивной константы. Тем не менее, предыдущей операции можно придать смысл. Для этого мы используем следующее замечание (на котором основано всё доказательство леммы 5 и теоремы 5).

З а м е ч а н и е 5. Существует (единственный) оператор

$$U = \left(\frac{\partial}{\partial \bar{z}} \right)^{-1} \circ \frac{\partial}{\partial z}, \quad \text{т. е.} \quad U \circ \frac{\partial}{\partial \bar{z}} = \frac{\partial}{\partial \bar{z}} U = \frac{\partial}{\partial z},$$

корректно определённый на всех гладких функциях на \mathbb{T}^2 , унитарный во всех соболевских гильбертовых пространствах функций на торе и сохраняющий средние. Действительно, операторы $\frac{\partial}{\partial z}$ и $\frac{\partial}{\partial \bar{z}}$ имеют общие собственные функции $e_n(x) = e^{2\pi i(n,x)}$, $n \in \mathbb{Z}^2$, образующие ортогональный базис в каждом гильбертовом соболевском пространстве. Для каждой из этих функций соответствующие собственные значения операторов $\frac{\partial}{\partial z}$ и $\frac{\partial}{\partial \bar{z}}$ равны по модулю и, более точно, получаются друг из друга последовательным применением замены знака и комплексного сопряжения. Это следует из того, что первый оператор сопряжён второму, взятому со знаком «минус» (формула интегрирования по частям). Теперь оператор U можно определить как оператор с тем же собственными функциями: соответствующие собственные значения находятся как частные соответствующих предыдущих собственных значений (за исключением нулевой гармоники (константы), на которой они оба равны нулю: на ней мы доопределяем U как тождественный оператор). Унитарность оператора U в соболевских пространствах следует из построения: все его собственные значения имеют модуль 1.

Перепишем теперь предыдущее уравнение на f в виде $f = (U \circ \mu) f$ или

$$f = (U \circ \mu) f + c, \quad c = \text{const.}$$

Всякое решение последнего уравнения (с произвольной константой c) удовлетворяет уравнению (1) по построению. Решим его при $c = 1$:

$$f = (Id - U \circ \mu)^{-1}(1) = \left(1 + \sum_{k=1}^{\infty} (U \circ \mu)^k \right)(1). \quad (2)$$

Предыдущий обратный оператор существует в $L_2(\mathbb{T}^2)$: $\|U \circ \mu\|_{L_2} < 1$, так как U унитарен (в частности, $\|U\|_{L_2} = 1$), а $\|\mu\|_{L_2} = \max |\mu| < 1$. Этот же оператор корректно определён и ограничен в каждом пространстве Соболева $H_s(\mathbb{T}^2)$: соответствующая операторная норма членов предыдущего

ряда убывает экспоненциально. Это доказывается с помощью элементарных оценок и с использованием коммутирования оператора U с дифференцированиями по аффинным координатам тора. Следовательно, функция f , заданная формулой (2), корректно определена и принадлежит всем соболевским пространствам. Пересечение последних состоит в точности из бесконечно гладких функций (теорема Соболева о вложении). Следовательно, $f \in C^\infty(\mathbb{T}^2)$. По построению, функция f является решением дифференциального уравнения замкнутости (1). Она не обращается тождественно в нуль, поскольку оператор $(Id - U \circ \mu)^{-1}$ обратим. Предложение 11 доказано. \square

Сформулируем свойство непрерывности формулы (2), которое будет использоваться при доказательстве второго шага (необращение в нуль) и теоремы 3.

Предложение 12. Функционал $\mu \mapsto f$, заданный формулой (2), непрерывен в C^3 -топологии на функциях μ и соболевской H_3 -топологии (а следовательно, и C^0 -топологии) на функциях f .

Предложение вытекает из элементарных оценок на члены ряда из формулы (2). Его последнее утверждение вытекает из теоремы Соболева о вложении: $H_3 \subset C^0$.

Замечание 6. По построению, $f \equiv 1$ при $\mu \equiv 0$. Тем самым, если норма $\|\mu\|_{C^3}$ — достаточно мала, то соответствующая функция f — C^0 -близка к 1 (и, в частности, нигде не обращается в нуль), по предыдущему предложению.

Локальная интегрируемость: доказательство теоремы 3. Фиксируем произвольную точку 0 рассматриваемой почти комплексной поверхности и некоторую (неголоморфную) комплексную координату в её окрестности. Можно считать, что рассматриваемая почти комплексная структура стандартна в точке 0 во введенной координате: этого можно добиться вещественно-линейным преобразованием координаты, переводящим соответствующий эллипс в окружность. Можно добиться того, чтобы производные почти комплексной структуры вплоть до порядка 3 были сколь угодно малы в подходящей окрестности точки 0, применяя замену масштаба. Можно выбрать малый диск в предыдущей окрестности, поместить его на тор \mathbb{T}^2 и продолжить почти комплексную структуру с диска на тор, сохраняя близость к стандартной структуре и малость производных. Пусть $\mu: \mathbb{T}^2 \rightarrow \mathbb{C}$ — функция, задающая продолженную структуру, $f(x)$ — соответствующая функция (2). Предыдущее построение можно сделать так, чтобы функция μ была сколь угодно мала в C^3 -норме. Тем самым, можно добиться того, чтобы $f \neq 0$, в силу предыдущего замечания. Искомая локальная голоморфная координата теперь задаётся

интегралом $x \mapsto \int_0^x f(dz + \mu d\bar{z})$. Она локально однолистка в окрестности точки 0, так как $f \neq 0$. Теорема 3 доказана. \square

Второй шаг: обращение в нуль. На предыдущем шаге по функции $\mu: \mathbb{T}^2 \rightarrow \mathbb{C}$ мы построили решение f уравнения (1), не обращающееся в нуль тождественно на \mathbb{T}^2 . Покажем теперь, что функция f нигде не обращается в нуль. Для этого будем использовать уже доказанную теорему 3, согласно которой рассматриваемый почти комплексный тор имеет гладкий атлас из локальных голоморфных карт (тем самым, является римановой поверхностью). По построению, 1-форма $f\omega_\mu$ — голоморфна (замкнутость и \mathbb{C} -линейность).

Будем доказывать второй шаг от противного. Предположим, что функция f имеет нули. Это суть, в точности, нули голоморфной формы $f\omega_\mu$ (гладкость голоморфных карт). Поэтому они изолированы и их число конечно (компактность). Рассмотрим один из нулей x_0 и голоморфную координату w в его окрестности, $w(x_0) = 0$. Имеем $f\omega_\mu = cw^k(1 + o(1))dw$ в координате w , $c \neq 0$, $k \in \mathbb{N}$ (голоморфность). В точке x_0 формы dw и ω_μ пропорциональны (считаем, что $c dw = \omega_\mu$, домножая w на константу). Тогда имеем

$$f(w) = w^k(1 + o(1)).$$

Аргумент правой части (а следовательно, и функции f) имеет положительное приращение $2\pi k$ вдоль положительно ориентированной малой окружности вокруг точки x_0 . Оно совпадает с интегралом от логарифмической формы $\frac{df}{f}$ по той же окружности.

Рассмотрим область $E \subset \mathbb{T}^2$, полученную из тора вырезанием малых дисков, каждый из которых содержит ровно один нуль функции f . Предыдущая логарифмическая форма, с одной стороны, голоморфна на области E , а с другой стороны, имеет ненулевой интеграл по её границе (предыдущее утверждение). Это противоречит формуле Стокса и доказывает второй шаг. Теорема 5 полностью доказана. \square

Список литературы

- [1] *Abikoff W.* Real analytic theory of Teichmüller space. Springer-Verlag, 1980. (Lect. Notes in Math.; V. 820).
- [2] *Ahlfors L.* Lectures on quasiconformal mappings. Wadsworth, 1987.
- [3] *Ahlfors L., Bers L.* Riemann's mapping theorem for variable metrics // Ann. of Math. 1996. V. (2)72. P. 385—404.
- [4] *Bers L.* Simultaneous uniformization // Bull. Amer. Math. Society. 1960. V. 66. P. 94—97.

- [5] *Douady A., Buff X.* Le théorème d'intégrabilité des structures presque complexes // The Mandelbrot set, theme and variations. Cambridge: Cambridge Univ. Press, 2000. (London Math. Society Lecture Note Ser.; V. 274). P. 307—324.
- [6] *Chern S.-S.* An elementary proof of the existence of isothermal parameters on a surface // Proc. AMS. 1955. V. 6. P. 771—782.
- [7] *Glutsyuk A.* Simultaneous metric uniformization of foliations by Riemann surfaces // Comm. Math. Helv. 2004. V. 79, Issue 4. P. 704—752.
- [8] *Glutsyuk A.* Simple proofs of uniformization theorems. <http://arxiv.org/abs/math/0510071>.
- [9] *Glutsyuk A.* Nonuniformizable skew cylinders. A counterexample to simultaneous uniformization conjecture // C. R. Acad. Sci. Paris. Série I Math. 2001. V. 332. P. 209—214.
- [10] *Ghys É.* Sur l'uniformisation des laminations paraboliques // Integrable systems and foliations / Ed. C. Albert, R. Brouzet, J.-P. Dufour (Montpellier, 1995). 1996. (Progress in Math.; V. 145). P. 73—91.
- [11] *Ильяшенко Ю. С.* Невырожденные Б-группы // ДАН СССР. 1973. Т. 208. С. 1020—1022.
- [12] *Korn A.* Zwei Anwendungen der Methode der sukzessiven Annäherungen. Berlin: Schwarz Festschrift, 1919. P. 215—229.
- [13] *Лаврентьев М. А.* Об одном классе непрерывных отображений // Мат. сб. 1935. Т. 42, № 4. С. 407—434.
- [14] *Lichtenstein L.* Zur Theorie der konformen Abbildungen; Konforme Abbildungen nicht-analytischer singularitätenfreier Fläschenstücke auf ebene Gebiete // Bull. Acad. Sci. Cracovie. 1916. P. 192—217.
- [15] *Morrey C. B. (Jr.)* On the solutions of quasi-linear elliptic partial differential equations // Trans. Amer. Math. Soc. 1938. V. 43, № 1. P. 126—166.

15 апреля 2004 г.

Ю. И. М а н и н

ДРОБНЫЕ РАЗМЕРНОСТИ В ГЕОМЕТРИИ И АЛГЕБРЕ

Сюжет, который я сегодня хочу рассказать, не произошел из какой-нибудь отдельной моей или чужой работы, а выкристаллизовался постепенно из осознания того, что размерности, которые встречаются в математике, образуют довольно странный спектр разнородных понятий, у которых, тем не менее, много общего. И было бы любопытно понять какой-то более единый контекст, в котором разного типа размерности, возникающие в геометрии и в алгебре, включая и, конечно, наиболее интересные нам размерности, которые не являются обыкновенными целыми числами, — так вот, такой контекст, в котором все бы они, их родство было бы лучше проявлено, чем это мы понимаем сейчас. Поэтому количество разных сюжетов, которые могли бы войти в этот доклад, очень велико. И те из них, которые я выбрал, образуют хотя и сравнительно маленькое, но все еще тоже слишком большое для одного доклада подмножество. Я воображаю себе уже примерно полугодовой курс на тему о том, какие бывают размерности.

Я этим летом стал размышлять над этой темой и впервые здесь попробую прокатать, что из этого получилось. Вкратце: там будет немножко истории, потом плотности и размерность Хаусдорфа, потом фон-неймановские размерности, связанные с C^* -алгебрами и факторами, переход к некоммутативной геометрии в стиле Конна, потом скачок в другую сторону — нецелые размерности, которые возникают из теории модулярных форм, p -адические веса Серра и функции периода Льюиса—Цагира, и наконец, если останется время, дробные размерности в гомологической алгебре и некоммутативной геометрии — это сравнительно недавний сюжет, придуманный в основном физиками, но уже осмысляющийся математиками, так называемый поток градуировок в производных категориях. То, что физики сейчас придумывают нетривиальные конструкции уже в контексте триангулированных и производных категорий, это замечательно и очень мне нравится.

История

Теперь немного истории. Я не знаю, как сейчас преподают геометрию в средней школе. Когда мы учились, геометрия основывалась на Киселеве, а Киселев, в свою очередь, пользовался традицией евклидовых «Элементов», и люди моего поколения, безусловно, помнят определение Евклида: точка это нечто, не имеющее частей, линия — это протяженность без ширины, граница линии состоит из точек. Это его формулировки, взятые из первой книги, о плоской геометрии. А вот из 11-й книги, о пространственной геометрии, определения несколько более современные. Там сказано приблизительно так: твердое тело — это нечто, что имеет длину, ширину и высоту (или глубину). И второе — очень важный принцип: границей твердого тела является поверхность. Это начало индуктивного определения, которое апеллирует к интуитивному понятию границы. У твердого тела граница — поверхность, у поверхности границей, по-видимому, должна быть линия, у линии границей должны быть точки. Ясно, что в этом месте Евклид что-то пропустил. Пропустил он чрезвычайно важный принцип: что граница границы пуста. Он не мог сформулировать этого, потому что «пустота» и «нуль» очень долго не были в обиходе математики. Было бы очень интересно услышать историков математики с тщательным филологическим анализом и с привлечением последующей истории возникновения нуля в арабской нотации. Пропущено в этой линии мысли фундаментальное уравнение гомологической алгебры: $\partial^2 = 0$. Если бы Евклид или кто-нибудь написал это словами: граница не имеет границы, то можно было бы считать этого человека тем, кто заложил первый камень в основу гомологической алгебры. Но Евклид этого не сделал.

Во всяком случае, у Евклида сделано примерно следующее. Имеется нечто, у чего есть граница, у этого нечто есть размерность. Обозначим это нечто через X^d , обозначим размерность этого нечто через d . Это d подчиняется такому принципу: граница этого чего-то является чем-то другим, причем размерность на единицу меньше: $\partial(X^d) = Y^{d-1}$. У Евклида d принимает значения 1, 2, 3. И это приблизительно то, что можно сказать на этом языке.

Теперь я процитирую второй фрагмент истории. Евклид — это примерно 300 лет до н. э. Дальше я перескочу сразу к Лейбницу (Leibniz). Это его письмо к Лопиталю (de l'Hôpital). Письмо к Лопиталю датировано 30 сентября 1695 г. Я сначала прочту буквально, а потом скажу, что, по-видимому, он имеет в виду. Он пишет: «Кажется, Иоганн Бернулли сообщил вам о том, что я ему рассказал о замечательной аналогии, которая позволяет объяснить тот факт, что последовательные дифференциалы функции об-

разуют геометрическую прогрессию». Смысл этого выражения, если я его правильно реконструирую, на нашем языке такой: нужно написать ряд Тейлора для значения функции в точке плюс приращение, но считая, что этот дифференциал является малым, но конечным приращением:

$$f(x + dx) = \sum \frac{1}{n!} f^{(n)}(x)(dx)^n.$$

В этом ряде $(dx)^n$ есть последовательный n -й дифференциал, они образуют геометрическую прогрессию. Ряд Тейлора к этому времени, по-видимому, еще не был открыт, но все-таки что-то в этом роде Лейбниц имел в виду.

Дальше идет вопрос, который имеет прямое отношение к теме этой лекции. «Можно спросить, что означал бы дифференциал в дробной степени?» Замечательно, что, воображая такую формулу или что-то в этом роде, он хочет вообразить себе это: $(dx)^\alpha$, где α дробное. Дальше Лейбниц замечает: «Хотя кажется, что этот вопрос отдален от геометрии, которая еще не знает такие дробные степени, я думаю, что когда-нибудь эти парадоксы доставят полезные следствия, потому что почти не существует парадоксов, которые были бы бесполезны».

Я хочу сразу сказать, что то, что я написал на доске, это просто сжатое напоминание о том, что было сказано словами и в том, и в другом случае. Но за словами Лейбница явно стоят формулы, тогда как за словами Евклида формулы определенно не стоят. За словами Евклида стоят картинки. В частности, возможно, что проблема пустоты и пустого множества или нуля именно потому так трудна, что трудно представить себе нечто отсутствующее. Символ для этого отсутствующего — более понятная вещь, чем само это отсутствие.

И это, конечно, иллюстрация вещей стандартных, которые на языке нейрофизиологии связаны с право- и левополушарным мышлением, условно говоря, материал для правого мозга — геометрия, музыка, пространственные отношения, а материал для левого мозга — язык, символика. Сейчас, в современной культурологии даже общества делятся на общества словесные, вербальные и общества образа. И, скажем, засилье телевидения среди средств массовой информации воспринимается как знак того, что нынешнее состояние общества является скорее правополушарным, чем левополушарным, а предшествующее книжное, языковое общество было вербальным. Те, кто знают феминистскую терминологию, слышали, наверное, слово «фаллологотризм», ассоциирующее мужской доминант с левым мозгом, и т. д.

Это два фрагмента истории — просто указание на то, что хорошо бы на эту тему подумать с такой точки зрения, потому что из этого могло бы

произойти что-нибудь любопытное. Я буду обращаться к геометрическим и алгебраическим языковым образам в равной мере, поскольку мы все как математики привыкли соединять то и другое. Но будет интересно подчеркнуть, какие фрагменты являются чисто алгебраическими, условно говоря, для которых не хватает геометрии. Алгебра всегда доставляет в каком-то смысле больший объем информации, чем геометрия. Ряд вещей, о которых я буду говорить, можно сформулировать как «алгебраический объект в поисках своего геометрического прообраза».

Плотности

Давайте начнем теперь с двух простейших ситуаций, где появляются дробные размерности. Первая — чисто алгебраическая, где можно придать смысл выражению $(dx)^\alpha$ по Лейбницу в вполне скучном бурбакистском контексте. Давайте мы ее построим, чтобы, по крайней мере, у нас какие-то вещи были точно определены.

Пусть M — дифференцируемое многообразие. Сейчас увидите, почему я хочу, чтобы оно было дифференцируемым. Я определю обратимый пучок ранга 1 или, если хотите, вещественное 1-мерное векторное расслоение V_s . Я определю его в картах. Если у меня есть локальные координаты (x_α) , то связанное с V_s и (x_α) каноническое сечение V_s в этих локальных координатах — это формальное выражение $|dx_1 \wedge \dots \wedge dx_n|^s$, где s — произвольное вещественное число. А функции перехода такие: если я меняю систему координат, то у меня возникает модуль якобиана в s -й степени (положительное число можно возвести в любую вещественную степень), который умножается на такой элемент объема:

$$|dx_1 \wedge \dots \wedge dx_n|^s = |\det(\partial x_i / \partial y_j)|^s \cdot |dy_1 \wedge \dots \wedge dy_n|^s.$$

А обычная бурбакистская машинерия позволяет определить что-нибудь, где такие вещи приобретают смысл. Это что-то становится сечением вполне канонически определенного 1-мерного пучка на расслоении. Конечно, с этим можно интересно играть, там есть разные дополнительные структуры, например, D -модули и т. д., это хорошая игра и полезные понятия.

Давайте посмотрим на второе, немного менее очевидное, но очень хорошо разработанное в геометрии XX века понятие, где возникают настоящие размерности. Выше шла речь, условно говоря, о размерности чего-то, что не является геометрическим объектом. Это индекс некоей системы векторных расслоений. Может быть, прежде чем переходить к новому понятию, давайте запомним пару свойств. Положим здесь $s = 1$. Тогда мы обна-

ружим, что некоторые сечения V_1 (некоторые в том смысле, что мы можем рассматривать не обязательно дифференцируемые сечения, а измеримые, непрерывные и т. д.) являются мерами, поэтому их можно интегрировать. Можно даже брать сечения, которые являются распределениями, тогда их можно интегрировать после умножения на подходящие функции и т. д. Кроме того, есть очевидное чисто алгебраическое спаривание:

$$V_s \times V_{1-s} \rightarrow V_1 \xrightarrow{\int} \mathbb{R}.$$

А так как здесь меры, мы можем после этого проинтегрировать и получить скалярное произведение (так мы и сделаем — это очень полезно). Есть разные связанные с этим обстоятельства (я не буду сейчас говорить о структуре D -модулей и прочих таких вещах), которые позволяют получить какой-то алгебраический формализм, связанный с плотностями.

Теперь вернемся к сюжету, где аналогичный индекс s выступает действительно в качестве меры, но будет видно, что там существуют какие-то действительно пространства дробной размерности. Выше дробная размерность — это, грубо говоря, вес; что-то вроде веса формального выражения. Потом мы к этой точке зрения вернемся, это совсем другое.

Размерность Хаусдорфа—Безиковича

Теперь давайте посмотрим на размерность Хаусдорфа—Безиковича. Эта размерность приписывается множествам странной структуры. Эта размерность сейчас наиболее популярна даже среди нематематиков, из-за того что Мандельброт написал длинную книгу на тему о фракталах, где он популяризировал эту самую размерность. Эту книгу интересно читать с точки зрения психолога: там поразительное количество психологии талантливой, яркой, творческого раздраженного человека. Но математическое содержание по сравнению с тем, что было известно до Мандельброта, не слишком велико.

Давайте я дам просто определение, потому что совершенно очевидно, что происходит, если просмотреть все шаги формального определения. Здесь такая сцена: некое метрическое пространство M . Картинка, по большей части — это толстый лист бумаги или что-нибудь в этом роде, евклидово пространство. Но существенно отметить, что, тогда как здесь выше важна была дифференцируемость, здесь важна метрика. Мы пытаемся определить, посчитать дробную «размерность» (я буду пока писать «размерность» в кавычках) подмножества M , априори произвольного, с компактным замыканием. Стратегия счета такая. Мы начинаем с формулы, которая имеет вполне определенный смысл и которую можно доказать:

если у вас имеется евклидов шар радиуса ρ в евклидовом пространстве размерности d , то тогда его d -мерный объем — это некая константа, связанная с Γ -функцией, умноженная на ρ^d , т. е. $\text{vol}_d(B_\rho) = c_d \rho^d$. Это факт. И мы притворяемся, что мы можем в этой формуле положить d равным любому вещественному числу. Принимаем в качестве интуитивного предварительного определения, что эта формула или приблизительно эта формула имеет смысл и верна для произвольных вещественных d . После этого мы декларируем, что мы будем считать d -мерный (для любого вещественного d) объем такого множества S с компактным замыканием таким образом. Мы покрываем S конечным числом шариков разных радиусов (мы потребовали компактности замыкания, чтобы можно было покрыть конечным числом; а исходное метрическое пространство нужно, потому что мы должны знать, каков радиус шарика). У нас получились шарики каких-то радиусов ρ_m , после чего, естественно, берем сумму объемов этих шариков (по элементам покрытия). После этого делаем предельный переход следующим образом: берем инфимум, ограничивая все радиусы данной величиной ρ . А потом можно доказать, что есть предел, когда ρ стремится к 0. И объявляем это по определению d -мерным объемом S :

$$\lim_{\rho \rightarrow 0} \inf_{\rho_m < \rho} \sum \text{vol}_d(B_{\rho_m}) =: v_d(S).$$

Тут, наконец, возникает теорема. Хаусдорф ее не формулировал, хотя это определение более или менее он дал. Сформулировал и доказал ее Безикович. Теорема совершенно замечательная, о том что существует критическое значение D , которое обладает таким свойством, что число $v_d(S)$ равно 0 при $d > D$, и ∞ при $d < D$. Я очень советую вам подумать, как это получается для простых множеств. В этой формуле d рассматривается как переменное, и тут есть просто критическое значение. Оно по определению и является размерностью Хаусдорфа—Безиковича.

Дальше, конечно, интересно, что мы не приписали никакой меры множеству S . Но это более тонкая вещь. Некоторые из таких подмножеств будут даже иметь и меру, а некоторые не будут, но размерность у них будет. Если вы посмотрите книгу Мандельброта, то там много всяких фрактальных множеств и вычислений того, какова у них конкретная размерность. И видно, что это имеет очень хороший интуитивный правополушарный смысл: множество как бы какое-то разреженное, потому что мы из него много дырок вырезаем и оно становится разреженным. И это число D , грубо говоря, — размерность того, что осталось.

Заметьте, что я не определил объем, я определил только это критическое значение. Объем может даже и вообще не существовать или как

угодно плохо себя вести. Есть отдельный, более узкий класс множеств, которые имеют меру.

Я не смог найти ничего в литературе по поводу следующего вопроса, который очень нечетко сформулирован, но тем не менее есть. При нецелом d можно ли определить какие-то специальные подмножества, которые будут заслуживать названия многообразий или, скажем, клеток? Какие S будут лучше других в том смысле слова, что они будут многообразиями или многообразиями с границей (с краем)? Так чтобы, например, граница имела на единицу меньшую размерность, возвращаясь к евклидовому определению. Я спрашивал у разных специалистов. Конечно, не обязательно в такой форме, но я думаю, что где-нибудь в литературе может быть такое определение, но я его не знаю.

Мне бы, например, очень хотелось, чтобы вдоль таких многообразий можно было бы интегрировать плотности V_d , которые были определены в предыдущем примере. Иногда можно — это не так безнадежно; но там сами многообразия не определены.

Может быть, нужна какая-нибудь однородность, типа самоподобности. Но если это даже и так, то, по-видимому, это не все, что там есть. Я не знаю в точности, что. Очень хотелось бы иметь какие-то такие квазимногообразия, границы которых бы имели такую же размерность на единицу меньше, и по которым можно было бы интегрировать плотности веса d и веса $d - 1$ и т. д. Гомологическую алгебру пока нельзя делать по той же причине, что у Евклида: у нас нет приличного определения границы, с формулой $\partial^2 = 0$.

Факторы

Я перехожу к следующему, по видимости, не связанному с этим кругу идей. Это замечательная бесконечномерная геометрия, которую я не знаю, с кого начинать: традиционно фон Нейман (von Neumann) и Мюррей (Murray), но, может быть, были какие-нибудь предшественники. Дробная размерность в их контексте такая. Тут дробные размерности возникают по интуитивно чрезвычайно простой причине: вы смотрите на размерность чего-то внутри бесконечномерного пространства, и это что-то тоже бесконечномерное, и то дробное число, которое при этом получается, это, грубо говоря, конечное отношение двух бесконечностей. Это интуитивная идея позади этого. Давайте посмотрим, как это оформляется.

Сцена — это линейное пространство M , скажем, над комплексными числами. Определяется/вычисляется размерность линейных пространств, которые являются подпространствами в M . Как возникают дробные размерности? Условно говоря, вы вычисляете $\frac{\dim L}{\dim M}$, которая есть бесконеч-

ность поделить на бесконечность, надлежащим образом регуляризованная. И это оказывается какое-то вещественное число.

Стратегия определения и счета такая. Сначала переписываете обычную теорию размерности конечномерных пространств, введя матричные алгебры, следующим образом. Предположим сначала, что у нас в обычном смысле слова пространство конечномерно. Обозначим через E_M матричную алгебру линейных эндоморфизмов M над \mathbb{C} и заменим L проектором $p_L \in E_M$ в этой алгебре. Это идемпотент ($p_L^2 = p_L$), $\text{Im } p_L = L$. Строим нормализованный функционал следа $\text{tr}: E_M \rightarrow \mathbb{C}$. Это просто след матрицы, разделенный на $\dim M$ (потому что я хочу здесь делить на размерность пространства) — так чтобы тождественный идемпотент, т. е. единичный оператор имел след 1: $\text{tr}(p_M) = 1$. И потом определяем нормализованную размерность L как след соответствующего проектора. Здесь надо еще доказывать, что от произвола выбора ничего не зависит. Это конечномерная теория.

В бесконечномерной теории предметом уже служат некоторые бесконечномерные \mathbb{C} -алгебры E , которые аналогичны алгебрам матриц. И в первоначальной работе Мюррея и фон Неймана было совершенно очевидно, как происходит этот предельный переход. Матрицы вкладывались в матрицы большего размера так, чтобы выполнялись некоторые условия когерентности, и в конце концов в пределе получался некий хороший объект. Я не буду сейчас, конечно, определять, что такое W -алгебры Мюррея и фон Неймана. Главное, что, если этот класс алгебр хорошо ввести, то ли конструктивным, то ли аксиоматическим определением, то после этого можно имитировать это описание. А именно, у алгебры будет каким-то образом нормализованный след. Вместо размерностей пространств мы рассматриваем следы соответствующих идемпотентов.

А то, что определение хорошее и правильное, оправдывается задним числом, потому что у нас получается совершенно невероятная по красоте классификационная теорема Мюррея и фон Неймана — знаменитая теорема, что алгебры делятся, грубо говоря, на 5 типов по спектру значений размерности, который таким образом получается — спектр следов самосопряженных идемпотентов. Он может быть одного из следующих типов:

$$I_n \quad \left\{ 0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1 \right\}$$

$$I_\infty \quad \{1, \dots, \infty\}$$

$$II_1 \quad [0, 1]$$

$$II_\infty \quad [0, +\infty]$$

$$III \quad \{0, +\infty\}$$

Первый тип получается из конечномерной теории. Здесь происхождение дробных размерностей связано именно с тем, что мы нормализуем, делим одно на другое; здесь можно было бы обойтись и без нормализации. В I_∞ без нормализации обойтись уже нельзя. В знаменитом типе II_1 получаются все вещественные числа между 0 и 1. Соответственно, в типе II_∞ — между 0 и плюс бесконечностью. И в типе III — только 0 и плюс бесконечность. Это довольно знаменитая классификация алгебр, которую можно рассматривать как придание смысла тому интуитивному ощущению, что в некоторых бесконечномерных линейных ситуациях у нас естественно возникают такие возможности — дробные спектры размерности.

Самые красивые приложения, вероятно, связаны с тем обстоятельством, что когда мы начинаем строить теории кохомологий разных геометрических пространств, в разных функционально аналитических контекстах, где цепи, коцепи и кохомологии все бесконечномерны, но имеют какие-то естественные гильбертовы структуры, то может оказаться, что пространства кохомологий сами по себе хотя и бесконечномерны, но у них есть размерность Мюррея—фон Неймана, и возникают, скажем, числа Бетти, которые являются такими размерностями и поэтому нецелыми образованиями: эйлерова характеристика, всякие замечательные формулы для индексов операторов и т. д. — это ситуация, где мы реально измеряем линейные бесконечномерные пространства. На этом уровне совсем не видно возможных связей между размерностью Хаусдорфа—Безиковича.

Подпространства и факторпространства; базы данных

И тут возникает очень интересная ситуация, о которой я просто сделал наблюдение и сформулировал. Я сейчас обсужу ее очень неформально, прежде чем сделать перерыв. Наблюдение такое. Я не знаю, какая за этим стоит точная математика, но как мне кажется интуитивно, возникает следующая ситуация. В размерности Хаусдорфа—Безиковича и во всех фрактальных картинках Мандельброта плохие пространства, которые мы себе воображаем, это подпространства хороших пространств. В ситуациях, где хорошо работает с большим успехом эта идеология и ее обобщения — это ситуация конновской некоммутативной геометрии, к которой я перейду. Там плохие пространства — это факторпространства хороших пространств. Все знают, что если есть слоение, скажем, иррациональная обмотка тора или что-нибудь в этом роде, то это отношение эквивалентности на вполне хорошем пространстве, но так как оно плохо согласовано с топологией, факторпространство не хаусдорфово и т. д. Конн много лет пропагандирует идею, что после того, как мы переходим

к факторпространствам, мы должны совершенно изменить язык описания нашего геометрического объекта, в том смысле, что там, где мы работали с коммутативными алгебрами функций и модулями над ними, мы должны начать работать с какими-то некоммутативными алгебрами, для которых есть эвристическое описание в каждом случае, и потом применять новую идеологию некоммутативной геометрии. Эта идея, конечно, является совершенно правильной, более того, эта же идея с другими словами обрабатывается в других разделах геометрии.

Перед перерывом я хочу сделать совершенно неформальное отступление о базах данных. Это такой сюжет, который все время обсуждается в теоретической и практической computer science, который пожирает много денег и т. д. Базы данных можно себе математически воображать как метрические пространства сложной и скверной структуры, и вычислительные задачи, которые ставятся по поводу базы данных обычно являются задачами типа поиска. Например, вы задаете термин, слово, фразу и вам нужно выдать все контексты, в которых эта фраза возникает. Значит, у вас в довольно скверном метрическом пространстве возникает еще более скверное метрическое подпространство, которое приходится задавать полным списком. И всегда считается, что это то, что нужно. Теперь давайте произведем такой мысленный эксперимент. Вот Дарвин плывет на корабле «Бигль», вот он записывает свои наблюдательные данные в своих дневниках, вот у него скапливается куча бумаги и отображение всего этого у него в голове, и вот из всего этого выходит теория эволюции. Нет сомнения, что теория эволюции есть скорее результат факторизации, сжатия базы наблюдаемых данных, чем результат отбора части наблюдений. То же можно сказать о любой теории, исходным пунктом которой являются обширные наблюдения и эксперименты, будь это данные о геноме человека или космологические данные о Вселенной, которыми мы обладаем, или еще чего-то в таком роде. Вопрос такой: как описывать факторы плохих пространств вместо подпространств. Это будет мотивировка к следующей главе моего доклада, где я буду больше говорить о конновской некоммутативной геометрии и о возникновении нецелых размерностей в ее контексте.

Некоммутативная геометрия

Я расскажу про идеологию Конна (Connes), а потом будет несколько подробнее сюжет, в котором почти все нити или фрагменты определений, интуитивных образов как-то сплетаются в одну очень интересную картину, которая не до конца продумана. Как я сказал, идеология Конна состоит в том, что главные плохие пространства — это не подмножества хороших,

а фактормножества хороших. И что думать об этих пространствах нужно как о некоммутативных пространствах, что бы это ни означало.

Перед тем, как приводить примеры и аргументы в пользу этой точки зрения, я сразу скажу, чего мы совершенно не понимаем с точки зрения этой идеологии. Конновские некоммутативные пространства почти во всех категориях, в которых они мало-мальски строились, почти во всех контекстах, в которых они мало-мальски исследованы, не образуют категории. Не только не понятно, что такое морфизм между такими пространствами, непонятно даже, что такое изоморфизм, непонятно даже, что такое канонический изоморфизм. Даже когда мы явно знаем, что вот это некоммутативное пространство просто равно, в некотором смысле слова является одним и тем же с другим, если мы подходим к его описанию с разных точек зрения, то мы совершенно не можем сказать, как математически выразить это равенство. Это сейчас одна из главных проблем некоммутативной геометрии. Мы не знаем ни одной ситуации, где некоммутативные пространства (кроме самых тривиальных случаев) образовывали бы хорошую категорию. Вполне возможно, что это принципиальная трудность и что некоммутативные пространства вообще образуют не категорию, а какой-то вариант поликатегории, ∞ -категории, может быть, A_∞ -категории, может быть, еще чего-нибудь в этом роде. Но это совершенно непонятно.

Сразу же сказав это, давайте я перейду к позитивным сторонам. Те, кто хоть немножко знает теорию Конна, понимают, о чем я говорю сейчас, а тем, кто не знает, я хочу проиллюстрировать эту идеологию на таком архетипичном примере.

Типичный плохой фактор такой: вы берете точку (одноточечное пространство), берете группу и заставляете эту группу действовать на одноточечное пространство. Группе этой совершенно невозможно действовать никак иначе, кроме как тождественно: каждый групповой элемент оставляет эту точку на месте. После этого вы спрашиваете, что такое фактор такого одноточечного пространства по действию такой группы. В любом классическом смысле слова этот фактор является самой точкой, а в смысле Конна — отнюдь нет.

Этот фактор представлен некоммутативной алгеброй, которая является неким вариантом групповой алгебры группы G . В каком смысле пространство представлено алгеброй? В алгебраической геометрии мы знаем, что, скажем, аффинные схемы представлены коммутативными кольцами — функциями этих аффинных схем. Скажем, векторные расслоения представлены проективными модулями над кольцами функций. Удержусь от того, чтобы сказать: а морфизмы представлены гомоморфизмами ко-

лец. Это последнее не переносится на некоммутативную ситуацию. Но первые два — да: мы воображаем себе, что некоторые некоммутативные алгебры представляют некоммутативные пространства, проективные модули над ними представляют векторные расслоения над этими некоммутативными пространствами. И дальше с этим надо работать, подгоняя геометрическую интуицию под алгебраический язык, с помощью которого мы эту интуицию выражаем. Тогда эта эквивалентность, что наше пространство является в каком-то смысле слова некоммутативным спектром групповой алгебры этой группы, становится интуитивно очень понятным.

Про пространство сказать многое нельзя, но скажем про категорию векторных расслоений. Что такое векторное расслоение над точкой? Это просто линейное пространство. Что такое векторное расслоение над точкой, у которой есть группа симметрий G ? Это векторное пространство, на которое G действует. Хотя она действует на точку тривиально, на векторное пространство она уже вполне может действовать нетривиально. Поэтому понятие векторного расслоения над таким пространством становится тождественным понятию представления группы G , линейного представления группы G . Очень хорошо: как раз линейные представления группы G и отвечают модулям над алгеброй $C[G]$. Так что с этим все в порядке, все хорошо. Это — архетипичный пример.

В более сложном случае вы можете себе представлять группу G , действующую на более богатое пространство, но так, что стабилизаторы точек меняются от места к месту, возникают какие-то скачки. Некоммутативность позволяет уловить эти скачки. Или вы можете заменить группу G на какой-то группоид эквивалентности, отношение эквивалентности, которое может быть не согласовано с топологией, или с тем, что еще нас интересует в этом пространстве. Все равно, имеется какой-то достаточно общий эвристический рецепт конструкции некоммутативной алгебры, которая представляет ваш фактор. Вот, условно говоря, жизнь идет таким образом в конновской некоммутативной геометрии.

Теперь вместо того, чтобы повторять много разных вещей, которые у Конна очень увлекательно написаны в массе его статей и нескольких книгах, я хочу описать такой контекст, где очень много всех этих интуитивных соображений как-то складывается в очень увлекательную, но не до конца понятую целостную картину. Это связано с модулярными формами.

(Дробные) размерности и модулярные формы. Давайте я напому некоторые классические определения. Обычная модулярная форма одной переменной — функция на верхней полуплоскости. На верхнюю полу-

плоскость действует группа дробно линейных преобразований $\mathrm{PSL}(2, \mathbb{Z})$: $z \mapsto \frac{az+b}{cz+d}$ — действие матрицы $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ на $z \in H$. И мы можем рассматривать голоморфные функции на верхней полуплоскости, удовлетворяющие функциональному уравнению вида $f\left(\frac{az+b}{cz+d}\right) = f(z)(cz+d)^{\omega+2}$. Это число ω называется весом. Впрочем, в разных контекстах весом называют немножко разные вещи. И мы хотим, чтобы это было выполнено для всех элементов некоторой подгруппы конечного индекса $\Gamma \subset \mathrm{PSL}(2, \mathbb{Z})$, иногда конгруэнц-подгруппы. Плюс к этому обычно добавляют еще разные условия поведения на бесконечности. Например, можно определять параболические модулярные формы.

Это — основное функциональное уравнение. Естественно, такие голоморфные функции образуют линейное пространство. Можно посчитать их размерности и прочее. Это, конечно, эквивалентно (и тут мы встречаемся с лейбницевской формулой) тому, что формальное выражение $f(z)(dz)^{\frac{\omega+2}{2}}$ является Γ -инвариантом. Это означает, что, если мы перейдем к модулярной кривой X_Γ , которая есть $\Gamma \backslash H$ (фактор верхней полуплоскости H по действию Γ), то такое выражение спускается до лейбницевского высшего дифференциала, голоморфного на такой модулярной кривой.

Но на самом деле это можно еще немножечко переписать. Лейбниц не знал внешних форм. Внешние формы, сколько я уж их лет знаю, все равно, остаются для меня некоторой загадкой. Лейбниц не знал внешних форм, он не знал, что по какой-то причине дифференциалы от независимых переменных должны антикоммутировать, что только в этом случае они садятся в хорошую теорию. И это функциональное уравнение можно переписать иначе: как условие того, что некоторая голоморфная форма объема на большем многообразии спускается — на что же? Ну, опять же, люди, которые занимаются модулярными формами, понимают геометрию этого дела. Верхняя полуплоскость параметризует эллиптические кривые: над каждой точкой z живет эллиптическая кривая E_z , которая является фактором \mathbb{C} по решетке, порожденной 1 и z . Дробно линейное преобразование из $\mathrm{PSL}(2, \mathbb{Z})$ переводит точку в полную орбиту таких точек, где соответствующие эллиптические кривые оказываются изоморфными. В этом, собственно, смысл этой модулярной категории. Поэтому над $\Gamma \backslash H$ у нас имеется некоторая универсальная эллиптическая кривая \mathcal{E} (с точностью до тонкостей, которые я пропускаю). Просто сажаем на орбиту ту эллиптическую кривую, которая этой орбитой параметризуется. Дифференциальная форма $f(z)dz \wedge d\zeta_1 \wedge \dots \wedge d\zeta_\omega$ на $H \times \mathbb{C}^\omega$ спускается

тогда на $\mathcal{E}^{(\omega)} := \overbrace{\mathcal{E} \times \dots \times \mathcal{E}}^\omega$ — расслоенное произведение ω -кратной этой

универсальной кривой над модулярной кривой:

$$\begin{array}{c} \mathcal{E}^{(\omega)} \\ \downarrow \\ \Gamma \backslash H \end{array}$$

В конечном счете действительно оказывается, что это ω -кратное произведение имеет голоморфную размерность $\omega + 1$, и оказывается, что это исходное функциональное уравнение описывает голоморфные формы старшего веса (с какими-то условиями на каспах, на бесконечности) на таком алгебраическом многообразии. Поэтому ω , $\omega + 1$ или $\omega + 2$ есть некая размерность — настоящая целочисленная размерность комплексного многообразия. Заметьте, что мы начали с лейбницевского варианта, где есть какой-то старший дифференциал, но после такого анализа функциональных уравнений пришли к тому, что на самом деле это ω впервые нам встретилось, так сказать, в неправильном контексте. А правильный контекст такой: у нас получается плотность, голоморфная плотность на комплексном аналитическом многообразии.

Поэтому если у нас в каком-нибудь контексте возникнут модулярные формы нецелого веса ω , они должны нам указывать на неявное присутствие каких-то многообразий нецелой размерности, пространств нецелой размерности. Это эвристический принцип, который я хочу проиллюстрировать.

Я буду называть весом ω , $\omega + 1$, $\omega + 2$ — безразлично, что в каком контексте удобнее; и заранее за это извиняюсь. Модулярные формы нецелого веса, если они в каком-нибудь контексте возникнут, должны указывать на скрытое присутствие каких-то пространств нецелой размерности. Я буду применять это как эвристический принцип к тому, чтобы как-то по-новому думать о тех ситуациях, где модулярные формы нецелого веса действительно появлялись раньше в математической литературе, возможно, по совсем другим причинам. Один контекст, в котором это произошло, это были серровские p -адические модулярные формы.

Серровские модулярные формы p -адического веса. Первый пример, когда у нас тут появляются модулярные формы, это очень интересный пример, потому что он не осмыслен ни в каком другом контексте, кроме того, в котором он первоначально возник. Это серровские модулярные формы p -адического веса. Впервые мы имеем размерность, которая является не вещественным числом, а p -адическим. Какие пространства стоят за ними, я совершенно не представляю себе, было бы очень интересно что-нибудь такое понять.

Здесь конструкция состоит из следующих шагов. Давайте будем рассматривать для простоты только случай, когда Γ — полная группа $\text{PSL}(2, \mathbb{Z})$. И давайте в качестве первого шага условимся, что мы будем задавать модулярную форму рядом Фурье. Если вы возьмете здесь матрицу $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, т. е. сдвиг на единицу, то вы обнаружите, что функция должна иметь период 1, а поэтому она разлагается в ряд Фурье по стандартной экспоненте: $f(z) = \sum_{n=0}^{\infty} a_n q^n$, $q = e^{2\pi iz}$. А что у него коэффициенты Фурье будут ненулевые только начиная с 0 — это условие поведения в каспе на бесконечности.

Определение Серра. p -адическая модулярная форма — это формальный ряд. В этом месте вы можете считать q формальной переменной. Коэффициенты здесь, конечно, были комплексные числа; тут мы хотим, чтобы они были p -адические числа. И этот формальный ряд должен быть пределом, т. е. существует последовательность настоящих модулярных форм (подчиняющихся этому уравнению) (f_i) весов каких-то (k_i) с рациональными коэффициентами, которая p -адически равномерно покоэффициентно сходится к f . Заметьте, что в определении про веса ничего не сказано. А вот сейчас про них будет сказано в теореме.

Теорема 1. *Если есть такая p -адическая модулярная форма f , то у нее есть вес $k(f)$, который является пределом (k_i) в группе $\mathbb{Z}_p \times \mathbb{Z}_{p-1}$.*

Группа здесь почти что p -адическая. На самом деле, p -адические числа здесь нужно умножить на циклическую группу порядка $p - 1$. Это мелочь, на которой я сейчас не буду останавливаться.

Есть отдельный вопрос о существовании таких модулярных форм. Но целый набор их строится конструктивно, так что это, оказывается, не проблема.

Итак, для p -адической модулярной формы имеется предел, который является почти что p -адическим числом, и который не зависит от того, какую последовательность вы выбрали, сходящуюся к f . Это очень замечательно и непонятно почему. Совершенно непонятно, какой объект, какое пространство стоит за этим. Я себе это воображаю, как расслоенное произведение, как какой-то предел таких расслоенных произведений

$$\overbrace{\mathcal{E} \times \dots \times \mathcal{E}}^{k_i},$$

где количество множителей равно k_i , но в каком-то таком p -адическом смысле, что пространства переменные и в общем случае стремящейся к бесконечности размерности, каким-то образом p -адически к чему-то

сходится. Спрашивается: к чему, и что все это означает? И вообще, есть ли теория размерности, похожая на что угодно, на Хаусдорфа—Безиковича или на что хотите, в которой естественно возникали бы p -адические, а не вещественные значения размерности? Даже, по правде говоря, нам очень не хватает Мюррея и фон Неймана, W -алгебры или C^* -алгебры, или неизвестно чего, что приводило бы к p -адическим размерностям бесконечномерных объектов, не вещественным. Все попытки строить p -адический функциональный анализ приводили к разным интересным вещам, но не к такого сорта теории, насколько я знаю.

Периодные функции Льюиса и Цагира. Теперь давайте я опишу второй контекст. Это так называемые периодные функции Льюиса и Цагира. Тут возникают вообще комплексные веса, связанные с нулями дзета-функции. Здесь мы переходим к чему-то, что должно быть областью некоммутативной геометрии конновского типа и что с предыдущей геометрической картиной связано так. Группа $\mathrm{PSL}(2, \mathbb{Z})$ действует, конечно, не только на верхнюю полуплоскость, но и на верхнюю полуплоскость с границей \mathbb{R} . В частности, она границу переводит в себя. Область алгебраической геометрии в общем случае включает в себя рациональные точки этой границы, только рациональные числа. Когда мы добавляем их, то при переходе к фактору H по подгруппам конечного индекса группы $\mathrm{PSL}(2, \mathbb{Z})$, мы получаем компактную модулярную алгебраическую кривую вместо некомпактной, мы добавляем каспы, добавляем точки на бесконечности, а все иррациональные точки абсолюта пропускаются. Алгебраическая геометрия не знает теории, где бы они возникали. Возникает поэтому такая идея, что у нас должна быть какая-то тень или какой-то новый объект теории модулярных функций, связанный с пространством $\mathrm{PSL}(2, \mathbb{Z}) \backslash \mathbb{R}$, а не $\mathrm{PSL}(2, \mathbb{Z}) \backslash H$, как это происходит в алгебраической геометрии. Какую точную математику можно навести на это? Это такая невидимая некоммутативная граница классической теории модулярных кривых, над которой могут существовать пространства такого типа, расслоения универсальных эллиптических кривых, которые будут тогда предположительно расслоениями на некоммутативные торы. Одна из первых причин моего интереса ко всему этому была вызвана такой программой теории классов для вещественных квадратичных полей вместо мнимых квадратичных полей. Потому что если у всего этого хозяйства есть интересная арифметика, она может привести к доказательству гипотез Старка о явной конструкции абелевых расширений вещественных квадратичных полей и всякого такого. Но об этом я сейчас говорить не буду, я буду говорить о более классическом куске всей этой идеологии, который связан еще с интуицией Гаусса.

У нас имеется такая интуитивная картина: есть некоммутативная граница. Здесь появляются конновские плохие пространства, которые являются плохими факторами и одновременно аналогами классических модулярных кривых. Давайте возьмем $\Gamma = \text{GL}(2, \mathbb{Z})$. Если этот определитель равен -1 , то верхняя полуплоскость переходит в нижнюю, но \mathbb{R} все равно переходит в себя, только ориентация меняется.

У $B_\Gamma := \Gamma \backslash P^1(\mathbb{R})$ есть другое описание как множества. А именно, это множество классов вещественных точек по такому отношению эквивалентности:

$$x \equiv y \Leftrightarrow \exists m, n: T^m x = T^n y.$$

Две точки эквивалентны, если существует два целых числа m и n с таким свойством. А T — это оператор, который переводит x в дробную долю $1/x$:

$$T: x \mapsto \frac{1}{x} - \left[\frac{1}{x} \right].$$

Если вы вообразите себе x записанным в виде непрерывной дроби (бесконечной в типичном для иррационального числа случае), то это будет сдвиг этой непрерывной дроби. Скажем, если x живет между 0 и 1, то таким образом:

$$x = \frac{1}{a_0 + \frac{1}{a_1 + \dots}} \mapsto \frac{1}{a_1 + \frac{1}{a_2 + \dots}}$$

Мы про a_0 забываем и берем эту дробную долю — сдвиг неполных частей. Это отношение заключается в том, что ниже чего-то цепные дроби совпадают. У одной дроби — ниже одного места, у другой — ниже другого места.

Это знаменитое гауссовское отношение эквивалентности. Можно рассмотреть оператор L на функциях, который дуален к T , и он будет выглядеть следующим образом:

$$(Lf)(x) = \sum_{k=1}^{\infty} \frac{1}{(x+k)^2} f\left(\frac{1}{x+k}\right).$$

Можно теперь рассматривать функции от вещественной переменной и такой оператор. Он поразительно похож на оператор Гекке над бесконечным простым числом. Никак не могу этому замечанию придать точный смысл, но что-то, мне кажется, в этом есть. То ли над бесконечным, то ли над нулевым. Смысл этой двойственности состоит в том, что если вы проинтегрируете по $[0, 1]$ произведение f на Lh — это то же самое, что интегрировать f , к аргументу которой применено T , умноженную на h :

$$\int_{[0,1]} f \cdot Lh \, dx = \int_{[0,1]} (f|T)h \, dx.$$

Так что этот гауссовский оператор — это такое функциональное выражение для оператора сдвигов.

И приятно сразу включить этот оператор в серию, зависящую от параметра. Тут-то и возникнут впервые формально наши дробные веса:

$$(L_s f)(x) = \sum_{k=1}^{\infty} \frac{1}{(x+k)^{2s}} f\left(\frac{1}{x+k}\right).$$

Просто вместо квадрата мы поставили здесь степень $2s$. Если посмотреть на эту формулу и записать ее не с дифференциалами, а с плотностями, то формула тогда будет выглядеть таким образом:

$$\int_{[0,1]} f |dx|^{1-s} L(h |dx|^s) = \int (f |dx|^{1-s} |T)h |dx|^s.$$

И тогда становится понятным, что оператор L_s двойствен, так сказать, оператору T , но рассматриваемому на таких плотностях, вместо того чтобы рассматривать на функциях. И интуиция тогда состоит в следующем. Оператор L_s представляет действие L на плотностях вместо функций. L -инвариантные s -плотности — это некая замена модулярной формы веса s , но только определенная не на верхней полуплоскости, с чего мы начали, а на ее некоммутативной границе. Если угодно, вы можете это формализовать и считать это формальным определением. Но я хотел объяснить, почему естественно считать ее таковой, и поэтому всю эту цепь последовательных шагов перед вами развернул.

Удачный функционально аналитический контекст для этого формализма был предложен Майером лет 15 назад.

Один классический пример гауссовской 1-плотности, которая является модулярной формой, это $1/(1+x)$. Это плотность, которую открыл Гаусс и которая у него возникла, когда он обнаружил замечательную статистику неполных частных непрерывных дробей — гипотеза Гаусса, которую доказали потом Кузьмин и Поль Леви независимо в 28—29 г. Такая модулярная форма, которая является 1-плотностью, инвариантная 1-плотность. Вообще, динамическая система, связанная с этим оператором T , изучалась много. Есть ее приложения к космологическим моделям, так называемый Mixmaster Universe, там возникает статистика непрерывных дробей при движении назад к моменту Большого Взрыва по неким довольно специальным метрикам в эйнштейновских уравнениях Общей Теории Относительности. Но это отдельный большой сюжет, я не буду сейчас об этом говорить.

Так вот, естественный вопрос такой: возникают ли естественные модулярные формы веса s в этом смысле слова в каком-нибудь контексте, где

с было бы не просто 1, а что-нибудь еще. У меня уже нет времени на то, чтобы приводить подробности, и я скажу просто ответ. Ответ качественно состоит в следующем.

Этот оператор я описал формально. Это чисто формальное выражение; я не сказал, на каком классе функций я это рассматриваю, где он сходится, какие у него функционально аналитические свойства и т. д. Первый пункт, который наводит на всем этом хозяйстве чистоту и который в высшей степени нетривиальный: есть хорошее функциональное пространство, состоящее из аналитических функций, банахово пространство, на котором L_s оказывается следовым оператором, ядерным оператором. И у этого оператора есть мероморфное продолжение на всю комплексную плоскость значений s , которое голоморфно всюду, кроме простых полюсов в $2s = 1, 0, -1, \dots$

Затем у этого оператора можно взять фредгольмовский детерминант такого вида: $\det(1 - L_s)$. И этот фредгольмовский детерминант оказывается зельберговской дзета-функцией, связанной с модулярной кривой. На модулярной кривой вы можете взять систему всех замкнутых геодезических, у них взять длины и обращаться с этими длинами так, как если бы они были простыми числами — нарисовать эйлерово произведение. Это будет зельберговская дзета-функция, и эта зельберговская дзета-функция оказывается этой замечательной вещью. У него есть известный список нулей, а именно, 1, нули на прямой $\operatorname{Re}(s) = 1/2$ и тривиальные нули типа $1 - k$, где $k = 2, 3, 4, \dots$; и наконец, еще критические нули ζ -функции Римана, деленные пополам, т. е. на $\operatorname{Re}(s) = 1/4$ — такой список нулей зельберговской ζ -функции.

Оказывается, что для этих значений s и только для них существуют модулярные формы соответствующего веса. Те, которые отвечают тривиальным нулям, порождают как раз классические модулярные формы. Все те, которые отвечают нетривиальным нулям, дают модулярные формы на некоммутативной границе совершенно загадочной природы и невероятно красивой теории. Те, кто хочет узнать подробности, могут обратиться к длинной статье Льюиса и Цагира, опубликованной недавно в «Annals of Mathematics».

Основной вопрос: что за пространства стоят за этими нулями. Я ожидаю, что есть какие-то пространства дробной и даже комплексной размерности, которые стоят за этими нулями, и что все те примеры, которые я вам сегодня рассказал, должны быть объединены в более общем контексте, чем тот, который мы понимаем сейчас.

А. Тоом

НЕЭРГОДИЧНОСТЬ В 1-D ПРОЦЕССЕ С ПЕРЕМЕННОЙ ДЛИНОЙ

Я довольно долго занимался так называемыми случайными процессами с локальным взаимодействием, о которых все вы слышали, а некоторые даже с ними много работали. Сейчас я хочу предложить вашему вниманию некоторый новый тип таких процессов, которые я называю «процессы с переменной длиной».

Предположим сначала, что у нас есть конфигурационное пространство $\{\ominus, \oplus\}^{\mathbb{Z}}$; оно состоит из бесконечных в обе стороны последовательностей, у которых каждый член — это либо плюс, либо минус: $\dots\ominus\oplus\oplus\ominus\oplus\dots$ И вопрос в том, какие мы с ними можем делать преобразования.

Некоторые преобразования вполне традиционные. Например, «превращение»: $\ominus \rightarrow \oplus$ с какой-то вероятностью. Один минус может превратиться в плюс с какой-то вероятностью. Я буду указывать только некоторые преобразования; вы должны домысливать, что симметричное преобразование тоже может произойти: если минус в плюс может переходить, то плюс в минус тоже может переходить. Интерпретация этого зависит от того, какое у меня время: дискретное или непрерывное. В непрерывном времени я случайным образом выбираю какую-то позицию, и в ней минус превращается в плюс. А в дискретном времени для каждого минуса специально бросается монетка, и в зависимости от результата этого бросания он превращается или не превращается. Я буду говорить только о дискретном времени, и в этом случае мы делаем это преобразование сразу со всеми плюсами и минусами; каждый из них изменяется или нет с какой-то вероятностью, независимо друг от друга.

А вот теперь будут операции нового типа. «Рождение»: из ничего может получиться частица: $\wedge \rightarrow \oplus$. Специальный знак \wedge означает «ничего», пустое место; его не было. Я говорю неформально: он возьмет и появится между двумя символами. Эта возможность уже обсуждалась в литературе; например, в очередном романе про Гарри Поттера было рассказано, как ряд домов стоит, и вдруг по мановению волшебной палочки возникает между ними дом, раздвигает соседей и занимает свое место между ними.

Операция «смерть» (один знак вдруг исчезает): $\oplus \rightarrow \wedge$. Например, вы можете себе представить какую-то очень тонкую, длинную биологическую формацию, типа лианы, которая состоит из каких-то единиц, скажем, из клеток; и эти клетки могут рождаться, умирать.

Может быть еще митоз, когда клетка превращается в две того же типа: $\ominus \rightarrow \ominus\ominus$.

А может быть аннигиляция, когда две соседние клетки, которые были разные (плюс и минус), обе исчезают с какой-то вероятностью: $\oplus\ominus \rightarrow \wedge$.

Всё это происходит с какими-то вероятностями.

Я буду рассматривать только системы очень конкретного вида, у которых может происходить следующее. Вначале все минусы: $\dots\ominus\ominus\ominus\dots$. Время t дискретно, и в каждый этап времени происходит две операции. Первая операция — Flip_β ; я буду называть ее флип, с параметром β (пока только один параметр). Flip_β — это значит, что каждый минус превращается в плюс с вероятностью β , независимо от всех остальных: $\ominus \xrightarrow{\beta} \oplus$.

Вторая операция — аннигиляция $\text{Ann}_\alpha: \oplus\ominus \xrightarrow{\alpha} \wedge$ (только в таком порядке). Аннигиляция может произойти только тогда, когда плюс и минус являются соседями; они уничтожают друг друга, а их соседи соответственно становятся соседями друг друга.

Это надо применять либо к конечным цепочкам, чего я делать не буду, либо, если проделать это с бесконечными, то надо считать, что мы имеем дело только с однородными мерами. Мера считается однородной, если она переходит в себя при сдвигах по этому пространству. Для конфигураций позиции нумерованы, для мер — нет. Я буду писать операторы справа от меры.

Таким образом, вначале существует мера δ_\ominus — это мера, сконцентрированная в одной конфигурации: все минусы. Потом применяется флип с параметром β ; потом применяется аннигиляция с параметром α ; и всё это повторяется много раз.

С мерой мы поступаем, как всегда: в этом пространстве мы выбираем некоторую Σ -алгебру, порожденную цилиндрическими множествами; и когда я говорю «мера», это мера на этой Σ -алгебре. Сходимость на всех цилиндрических множествах.

Основной вопрос, который меня интересует, это поведение меры

$$\delta_\ominus(\text{Flip}_\beta \text{ Ann}_\alpha)^t = \mu_t,$$

когда параметр t стремится к бесконечности.

Мера, которая получается, будет тоже трансляционно инвариантной, потому что то, что происходит, однородно по всей прямой.

Теорема 1. $\mu_t(\oplus) \leq \frac{300\beta}{\alpha^2}$.

Здесь $\mu_t(\oplus)$ — это μ_t от плюсов, т. е. частота плюсов в мере μ_t меньше или равна, чем $300\beta/\alpha^2$. Обратите внимание, что существуют такие положительные α и β , что это число очень маленькое, а именно, β должно быть много меньше, чем α^2 . И эта оценка не зависит от t ; т. е. плотность плюсов всегда останется очень маленькой.

Теорема 2. Если $2\beta > \alpha$, то $\mu_t \xrightarrow[t \rightarrow \infty]{} \delta_{\oplus}$.

Таким образом, мы имеем фазовый переход, потому что поведение этих мер различно при различных соотношениях между α и β . Действительно, выберем систему координат с координатами α и β (рис. 1). Теорема 2

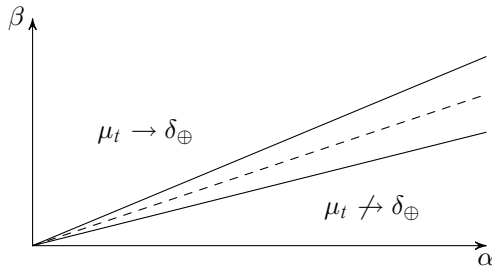


Рис. 1. Два графика

показывает, что во всей области над графиком μ_t стремится к δ_{\oplus} . Теорема 1 показывает, что если провести параболу, то под этой параболой μ_t не стремится к δ_{\oplus} . Значит, есть какая-то граница между ними, по меньшей мере одна (я еще не доказал, что только одна); выше этой границы меры стремятся к δ_{\oplus} , а ниже — не стремятся. Это еще не следует из обеих теорем, потому что пока формально мы можем предположить ситуацию, что мы идем снизу вверх, и мера сначала не стремилась, потом стремится, потом снова не стремится, потом стремится, потом снова не стремится и т. д., сколько угодно раз.

Теорема 3. Возьмем любую меру из нашего множества, т. е. нормированную меру на том пространстве, о котором сейчас всё время идет речь. Предположим, что $\beta > 0$ и $(1 - \beta)\mu(\theta) \leq \frac{1}{2}$. Тогда $\mu(\text{Flip}_{\beta} \text{Ann}_{\alpha})^t \rightarrow \delta_{\oplus}$.

То есть, если начальная мера такая, что плотность минусов в ней не слишком большая, то применение нашего оператора к ней ведет ее к δ_{\oplus} . От α предел вообще не зависит.

Теорема 4. Обозначим $S(\alpha, \beta) = \sup_t \mu_t(\oplus)$ (это супремум по всем натуральным t). Тогда для любого фиксированного α функция $S(\alpha, \beta)$ разрывна как функция от β .

Случай $\alpha = 1$ я не рассматриваю. Этот случай плох тем, что тогда в некоторых ситуациях оператор аннигиляции не всегда удастся применить. Представьте себе, что была мера, где минусы и плюсы строго чередовались (на четных местах стояли плюсы, а на нечетных — минусы). Если применить операцию аннигиляции, то непонятно, что получится. Это особый случай, к которому оператор аннигиляции с $\alpha = 1$ не применим. Если же $\alpha < 1$, то оператор аннигиляции можно применять к этой конфигурации. Получится та же самая мера. Некоторые из них аннигилируют, но не все; а оставшиеся составят достаточно, чтобы заполнить всю прямую.

Я предпочитаю иметь дело с мерами. Дело в том, что наша традиция определять операторы, как они действуют на конфигурации, и считать, что тем самым определено, как они действуют на меры, следует из того, что оператор непрерывен. Эти операторы действительно непрерывны, но это еще надо доказывать. Я предпочитаю говорить о том, как они действуют на меры.

Оператор Ann_α определяется как $\text{Duel}_\alpha \text{Clean}$. И остается определить этот оператор Duel_α , который зависит от α , и оператор Clean , который не зависит от α . Оператор Duel_α превращает каждую меру на пространстве $\{\ominus, \oplus\}^{\mathbb{Z}}$ в меру на $\{\ominus, \odot, \oplus\}^{\mathbb{Z}}$. Это третье вспомогательное состояние \odot (я буду называть его пустое состояние); и оператор Duel_α состоит в следующем: если были \oplus, \ominus , то с вероятностью α им дается команда «огонь», и с вероятностью α они оба убивают друг друга и переходят в пустое состояние. А оператор Clean потом все пустые состояния вычищает. Это делается сразу для всех пар одновременно, но каждая берется с вероятностью α ; с вероятностью α дуэль происходит, а с вероятностью $1 - \alpha$ она не происходит, и независимо для каждой пары дуэлянтов.

Теперь определение оператора Clean :

$$\mu_{\text{Clean}}(a_0, \dots, a_k) = \frac{1}{1 - \mu(\odot)} \sum_{n_1, \dots, n_k=0}^{\infty} \mu(a_0 \odot^{n_1} a_1 \dots a_{k-1} \odot^{n_k} a_k).$$

Это мера на слове; но в силу однородности, это то же самое, что мера на цилиндре. Поскольку я рассматриваю только однородные меры, если вы сдвинете цилиндр по прямой, значение меры будет то же самое. Поэтому меру можно просто определять на словах, т. е. на конечных последовательностях тех букв, которые мы употребляем.

Единственная мера, для которой μ_{Clean} не определено, это мера, сосредоточенная на конфигурации «все точки»; для всех остальных мер определено. Надо, конечно, еще доказывать, что эти μ_{Clean} от цилиндров согласованы, но это не трудно.

Как появляется множитель, вполне понятно: поскольку точки исчезают, то шансы всех остальных повышаются. Возьмем, например, $\mu_{\text{Clean}}(\oplus)$. Ясно, что исчезновение точек делает каждый плюс вносящим больший вклад в частоту, потому что мы считаем число плюсов и делим на ту длину, на которой они расположены, но длина уменьшилась. В данном случае, поскольку это слово состоит из одной буквы, пробелов нет; поэтому мы получаем

$$\mu_{\text{Clean}}(\oplus) = \frac{\mu(\oplus)}{1 - \mu(\ominus)} = \frac{\mu(\oplus)}{\mu(\oplus) + \mu(\ominus)}$$

Обозначение $\mu(\oplus)$ означает следующую вероятность. Вы тыкаете пальцем в точку и спрашиваете: «Какова вероятность того, что здесь плюс?» Например, $\mu(\ominus)$ для конфигурации, состоящей из одних плюсов, это нуль.

Очень важно, что этот оператор не линейный; с этим связаны определенные трудности. Например, я поинтересовался тем, как мы доказываем, например, теорему, считающуюся доказанной, что каждый клеточный автомат имеет по меньшей мере одну инвариантную меру. Оказывается, там очень существенно используется линейность.

Теперь я хочу прокомментировать эти теоремы. Они показывают и сходство, и различие этого процесса с хорошо известными контактными процессами. Естественно сравнивать этот процесс с контактными процессами. Среди них есть давно известный процесс Ставской: это вариант контактных процессов для систем с дискретным временем. Поэтому мне удобнее сравнивать этот процесс с процессом Ставской.

Я сформулирую определение процесса Ставской в тех обозначениях, которые я использую сегодня: $\delta_{\ominus}(\text{Flip}_{\beta}\text{Stav})$. Мы тоже начинаем с меры δ_{\ominus} , которая сосредоточена в конфигурации «все минусы», а потом применяем тот же самый флип с параметром β и еще оператор Stav, который от параметров никаких не зависит. Оператор Stav — это детерминированный оператор, который можно рассматривать как применяемый к конфигурациям: $\text{Stav}: \{\ominus, \oplus\}^{\mathbb{Z}} \rightarrow \{\ominus, \oplus\}^{\mathbb{Z}}$. А именно, $(x \text{Stav})_i = \min(x_i, x_{i+1})$, т. е. если к конфигурации x применяется Stav, и мы интересуемся i -й координатой образа, то она просто равна минимуму от x_i и x_{i+1} ; минимум в том смысле, что минус меньше плюса.

Поведение этого оператора разделяет все свойства, типичные для контактных процессов. Ясно, что оператор Ставской Stav благоприятствует минусам: он иногда превращает плюсы в минусы (а именно, он превращает плюс в минус, если его правый сосед — минус). А оператор Flip_{β} благоприятствует плюсам; он сильно благоприятствует плюсам, если β большое, и слабо благоприятствует плюсам, если β маленькое. Поэтому не удивительно, что тут возникает борьба двух тенденций, и в результате мы имеем

различное поведение при маленьких β и при больших β . Пока вроде бы похоже на то, что получается в нашем процессе. Однако есть и различия тоже. Например, было доказано для операторов такого типа, что аналог функции s непрерывен, и поэтому там фазовый переход 2-го рода; а у нас эта функция разрывна, поэтому тут фазовый переход 1-го рода.

Когда эта статья рассматривалась в журнале, там было три референта — два математика и один физик. И физик сказал, что это единственный ценный результат в этой работе, ради которого ее стоит печатать. Но этот результат — следствие основных результатов; а изолированно доказан быть не может. Самая важная тут первая теорема; вторая и третья — легкие, они нужны для контраста.

Теперь я расскажу про доказательство того, о чем я говорил в начале. Была сформулирована теорема 1: для всех t имеет место неравенство

$$\delta_{\ominus}(\text{Flip}_{\beta} \text{Ann}_{\alpha})^t(\oplus) \leq \frac{300\beta}{\alpha^2}.$$

(У Flip только один параметр здесь, потому что минусы превращаются в плюсы, а плюсы в минусы не превращаются.) Надо сказать, что я очень смущен этим α^2 ; я очень хотел бы тут написать просто α . Действительно, предположим, что мы захотим переходить к пределу, когда β и α стремятся к 0 в определенной пропорции. Интуитивно кажется ясно, что при этом мы переходим к оператору с непрерывным временем; и могли бы даже его так определить. Но с этой оценкой я ничего не могу сделать. Тут должно быть по всему здравому смыслу $\frac{\beta}{\alpha}$; но я не смог это получить.

Я применяю контурный метод Пайерлса, чтобы доказать этот факт. И сейчас я скажу несколько слов о том, как его применяю. Давайте вспомним, что вначале у нас все минусы. Давайте возьмем какую-то реализацию нашего процесса и построим соответствующий граф. Пусть есть частица. Во-первых, с ней может ничего не произойти, и может ничего не происходить сколь угодно долго. Может ничего не происходить до некоторого момента, а потом тут может возникнуть плюс благодаря оператору. Дело в том, что эти натуральные числа не равны t ; поскольку у меня есть два оператора, то переход от четного t к нечетному соответствует действию оператора Flip, а переход от нечетного t к четному соответствует действию оператора аннигиляции. Поэтому тут может возникнуть плюс; далее этот минус и плюс могут аннигилировать, и в этом случае они оба превращаются в точки; а уж с точками ничего не меняется, они остаются точками навсегда.

Это другое представление того же самого процесса, при котором оператор Clean не действует, но плюсы и минусы взаимодействуют так, как

будто точек не было; т. е. они через эти точки находят своих ближайших соседей живых и с ними как-то взаимодействуют.

Построим граф, у которого вершинами будут все живые точки. У этого графа мы соединяем две точки либо горизонтальным ребром, либо вертикальным. С вертикальным ребром всё очень просто: если одна точка над другой, мы их соединяем ребром. Горизонтальные точки соединяются ребром, если между ними все точки мертвые. Каждая живая точка соединяется снизу; если она не умирает в следующий момент, она со своим состоянием в следующий момент тоже соединяется; а также соединяется с левым и правым живыми соседями. Получается некий граф; это граф не ориентированный. Мертвые точки не являются вершинами графа.

Этот граф плоский; он прямо изображен на плоскости без самопересечений. Раз он изображен на плоскости без самопересечений, у него есть двойственный граф. И каждой вершине исходного графа соответствует некоторая страна двойственного графа.

С положительной вероятностью на каждой вертикали стоит минус, с которым до интересующего нас момента времени еще ничего не произошло. Поэтому тут тоже есть такой упорный минус, который ни с кем не реагировал и остался минусом, до интересующего нас момента времени. Меня интересует то, что происходит в коробке между этими двумя вертикальными стенками. Тут меня интересует момент времени t , и меня интересуют плюсы в этот момент времени t . Я их называю цветами. Но любой садовник вам скажет, что цветов не бывает без корней. И у этих цветов тоже есть корни. А именно, если тут есть цветок, то мы можем от него пойти вниз и вправо. И если там, куда мы идем (вниз или вправо), тоже плюс, то это его корень. Сам цветок тоже будет по определению корнем а также южный сосед корня — тоже корень, если он в состоянии \oplus , и восточный сосед корня — тоже корень, если он в состоянии \oplus . Таким образом у каждого цветка есть некая корневая система.

Я беру все цветы, беру все их корни (это корневая система), и можно доказать, что это некий кластер, у которого есть некоторая граница, и на этой границе определенные вещи должны происходить. Например, если эта граница здесь проходит, значит, тут плюс, а тут минус, значит, тут должен был сработать оператор Flip, что происходило с вероятностью β . А справа, если я от какого-то корня не могу пойти вправо, значит, в этом месте не произошла аннигиляция. Если тут минус, значит, этот и этот не проаннигилировали, что добавляет еще один фактор.

И каждый такой шаг — я иду по контуру. Тут получается нечто вроде случайных блужданий — я иду по этому контуру, и каждому шагу соот-

ветствует некий фактор. И вот если просуммировать произведения этих факторов по всем контурам, мы получаем эту оценку.

Сейчас я скажу, что такое контур. Не забывайте, что в исходном графе корень — это точка, а в дуальном графе это страна. И совокупность этих стран окружена границей; эта граница состоит из шагов; и вот контур — это граница, состоящая из шагов. Это как в методе Пайерлса, очень похоже; некоторая специфика есть, но основные идеи оттуда; также как дуальность плоских графов.

Я вам рассказал про работу, которая опубликована в «Journal of statistical physics» (Т. 115, № 3/4, май 2004).

16 сентября 2004 г.

ТОЧКИ НА КРИВЫХ ШИМУРЫ НАД ЧИСЛОВЫМИ ПОЛЯМИ

Я сегодня буду говорить про рациональные точки на кривых Шимуры. Начну с того, что расскажу кратко предысторию вопроса, а потом объясню, что такое кривые Шимуры.

1.

Модулярная кривая

Надо начать с того, как Барри Мазур (B. Mazur) в 1977 г. нашел все \mathbb{Q} -точки на модулярных кривых $X_1(N)$. Напомню, что это такое.

Рассмотрим группу $\Gamma_1(N)$; пусть N — это целое положительное число. Рассмотрим матрицы 2×2 с целыми коэффициентами $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, которые по модулю N сравнимы с матрицей $\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$. Потребуем также, чтобы определитель был равен 1, т. е. $ad - bc = 1$, $a, b, c, d \in \mathbb{Z}$. Эти матрицы — подмножество алгебры $\text{Mat}_2(\mathbb{Q})$ квадратных матриц порядка 2 над рациональными числами.

Такие матрицы действуют на верхней полуплоскости \mathbb{H} (верхнюю полуплоскость можно рассматривать как множество комплексных чисел с положительной мнимой частью: $\{z \in \mathbb{C} \mid \text{Im } z > 0\}$). Тогда, как известно, есть дробно-линейное действие $z \mapsto \frac{az+b}{cz+d}$. И можно рассмотреть фактор $\Gamma_1(N) \backslash \mathbb{H}$. Фактор этот будет обозначаться $Y_1(N)$. Многие знают, что этот фактор некомпактен. Чтобы сделать его компактным, нужно добавить конечное множество точек — параболические точки или каспы. Пусть $X_1(N)$ — это такая компактификация (гладкая). Это — модулярная кривая.

Модулярная кривая, как я ее определил, это риманова поверхность. Но замечательное обстоятельство состоит в том, что на самом деле $X_1(N)$ — это алгебраическая кривая над \mathbb{Q} . И классики знали, как при желании написать ее уравнение. В классической науке люди делали следующее: они брали параболическую точку; в этой точке можно написать разложение в ряд. Если есть рациональная функция, инвариантная относительно

действия $\Gamma_1(N)$, то можно разложить ее в ряд; если есть несколько функций, то можно написать уравнение, которое их связывает. Таким образом можно получить многочлен с постоянными коэффициентами, и это дает уравнение модулярной кривой над \mathbb{Q} . Но это получалось из вычислений. А глубокое объяснение этого факта состоит в том, что $Y_1(N)$ есть грубое многообразие модулей таких объектов: пара, состоящая из эллиптической кривой E и точки P на ней, где P — точка порядка ровно N . Это только грубое многообразие модулей, но всё равно, это очень важное обстоятельство. Проблема модулей над \mathbb{Q} или даже над \mathbb{Z} — это другой подход к тому факту, что $X_1(N)$ — многообразие над \mathbb{Q} .

Мазур нашел все рациональные точки этого алгебраического многообразия над \mathbb{Q} . Важность этого шага для теории чисел в том, что из этого он смог вывести точное описание всех возможных групп кручения эллиптических кривых. Теорема из этой работы говорит следующее: если род кривой $X_1(N)$ больше 0 (т. е. если это не проективная прямая), то все рациональные точки кривой $X_1(N)$ — это, в сущности, параболические точки (то, что нужно добавить к $Y_1(N)$, чтобы получить $X_1(N)$). Это замечательное описание.

Многообразием модулей эллиптических кривых является Y , а X получается компактификацией Y . В принципе, можно придать этому смысл, сказать, что X классифицирует как бы обобщенные эллиптические кривые, т. е. не только эллиптические кривые, но и их вырождения. А эллиптические кривые в собственном смысле отвечают только точкам Y . В этих обозначениях Мазур, собственно, доказал, что на Y рациональных точек как бы нет, т. е. $Y_1(N)(\mathbb{Q}) = \emptyset$.

Поэтому если у вас есть эллиптическая кривая E над \mathbb{Q} и точка $P \in E(\mathbb{Q})$, имеющая порядок в точности N , то N должно быть не равно 11 и $N \leq 12$. То есть не может быть, скажем, рациональной точки, у которой порядок ровно 20. Это замечательное обстоятельство позволило полностью классифицировать все подгруппы кручения, состоящие из рациональных точек на эллиптических кривых.

2.

Кватернионы

Это была мотивация. Теперь я хочу поговорить про кривые Шимуры. Они получаются таким образом. Конгруэнц-подгруппа — это подмножество алгебры $\text{Mat}_2(\mathbb{Q})$ матриц порядка 2 над \mathbb{Q} . А теперь вместо такой алгебры рассмотрим алгебру кватернионов B . Это как бы обычная кватернионная алгебра; все знают, как она определяется: четырехмерное

векторное пространство $B = \mathbb{Q} \oplus \mathbb{Q}i \oplus \mathbb{Q}j \oplus \mathbb{Q}k$ над \mathbb{Q} со стандартными соотношениями $k = ij = -ji$; единственное отличие от алгебры обычных гамильтоновых кватернионов в том, что $i^2 = a$, $j^2 = b$, $a, b \in \mathbb{Z}$, $a \neq 0$, $b \neq 0$.

Не при всяких числах a и b получается алгебра кватернионов, иногда может получиться алгебра, изоморфная матричной алгебре. B — тело тогда и только тогда, когда уравнение $ax^2 + by^2 = z^2$ не имеет ненулевых решений в \mathbb{Q} . Рассмотрим такие a и b ; например, можно взять $-1, -1$; а можно взять что-нибудь другое.

У этой кватернионной алгебры есть редуцированная норма — отображение $N: B \rightarrow \mathbb{Q}$. Это аналог определителя; его квадрат — определитель умножения как линейного преобразования B . Есть также редуцированный след $T: B \rightarrow \mathbb{Q}$. Одно отображение мультипликативное, другое аддитивное. А еще у нее есть редуцированный дискриминант или просто дискриминант. Он определяется таким образом: это произведение простых чисел по всем p , для которых, если помножить B тензорно на p -адические числа, то получится по-прежнему алгебра кватернионов, т. е. нечто не изоморфное алгебре матриц над \mathbb{Q}_p :

$$D = \text{discr } B = \prod_{p - \text{ простое: } B \otimes \mathbb{Q}_p \not\cong \text{Mat}_2(\mathbb{Q}_p)} p.$$

Для таких p после тензорного умножения на \mathbb{Q}_p алгебра кватернионов остается телом. Есть только две возможности: стать алгеброй матриц или остаться телом. Таких простых чисел конечное количество. И говорят, что это числа, которые разветвлены в B . А также бесконечность разветвлена в B или нет; иными словами, умножая на \mathbb{R} тензорно, мы тоже можем получить либо алгебру матриц, либо останутся кватернионы.

Кватернионные алгебры над \mathbb{Q} с точностью до изоморфизма соответствуют целым положительным числам, свободным от квадратов. При этом алгебре сопоставляется ее дискриминант $B \leftrightarrow D$.

Из глобальной теории полей классов вытекает, что число нормирований поля \mathbb{Q} , в которых алгебра остается телом (если сюда включить и вещественные числа), есть четное число, т. е. мощность множества

$$\{\text{нормирования } \mathbb{Q} \mid B \otimes \mathbb{Q}_\nu \not\cong \text{Mat}_2(\mathbb{Q}_\nu)\}$$

четна. Поэтому можно узнать, что если D состоит из произведения нечетного количества простых, то, значит, в бесконечности B остается телом, а если четное количество, то в бесконечности B — матричная алгебра. Например, если взять стандартную алгебру кватернионов ($a = -1$, $b = -1$), то она разветвлена в двойке и в бесконечности, ее дискриминант равен 2.

Для того чтобы определить аналог конгруэнц-подгрупп, мне нужен аналог кольца целых. Я напомним, что если у вас есть числовое поле k , то в нем есть кольцо целых, которое определяется как целое замыкание \mathbb{Z} в k — все элементы, которые целы над \mathbb{Z} , все корни унитарных многочленов (т. е. многочленов со старшим коэффициентом 1) с целыми коэффициентами. А соответствующий объект в кватернионной алгебре так легко нельзя определить. Надо рассматривать порядки в B — подкольца $\mathcal{O} \subset B$, содержащие единицу и порождающие над рациональными числами всё B : $\mathcal{O} \otimes \mathbb{Q} = B$. Аналогом кольца целых являются максимальные порядки; например, такие, которые действительно максимальны в смысле включения. Но проблема состоит в том, что, поскольку B не коммутативна, то максимальных порядков много; можно взять один максимальный порядок, сопрячь его, и получится другой. Здесь проявляется различие между алгебрами, которые в бесконечности расщепляются, и теми, которые не расщепляются. Оказывается, что верен такой факт: если в бесконечности B расщепляется, т. е. $B \otimes \mathbb{R} \simeq \text{Mat}_2(\mathbb{R})$, то все максимальные порядки сопряжены. Это очень хорошая ситуация. (Есть стандартная книга, в которой всё это написано: это Vignéras, *Lecture Notes Mathematics* 800. Это стандартная ссылка для всего, что нужно знать про порядки в кватернионных алгебрах.) И мы только такие кватернионные алгебры будем рассматривать; они называются неопределенными.

3.

Кривые Шимуры

Из-за того что B на бесконечности является алгеброй матриц, можно рассмотреть действие разных групп, которые в ней содержатся, на верхней полуплоскости. Более точно, зафиксируем максимальный порядок $\mathcal{O} \subset B$. Редуцированная норма $N: \mathcal{O} \rightarrow \mathbb{Z}$, отображает $\mathcal{O} \subset B$ в целые числа, так же как норма кольца целых числового поля; при этом обратимые элементы переходят в ± 1 (обратимые элементы \mathbb{Z}): $\mathcal{O}^* \rightarrow \{\pm 1\}$. Возьмем те из них, у которых норма равна 1: $\mathcal{O}^+ = \{x \in \mathcal{O} \mid N(x) = 1\}$. И теперь, поскольку я потребовал, чтобы норма была равна 1, то \mathcal{O}^+ вкладывается в $\text{SL}_2(\mathbb{R})$ — из-за отождествления $B \otimes \mathbb{R} \simeq \text{Mat}_2(\mathbb{R})$. Я могу подействовать соответствующими матрицами на верхней полуплоскости и рассмотреть фактор $\mathcal{O}^+ \backslash \mathbb{H} = X$. Как и раньше, это риманова поверхность. Это кривая Шимуры в ее простейшем варианте.

Особенность состоит в следующем: для кватернионной алгебры этот фактор компактен, если алгебра не матричная (а если алгебра матричная,

то фактор некомпактен). Это чрезвычайно важное обстоятельство: никаких параболических точек добавлять не надо. И это в каком-то смысле хорошо, а в некотором смысле плохо. Кроме этого вся остальная теория переносится в общем-то довольно хорошо.

Я упомянул раньше, что классический вариант кривой можно задать уравнением (это часть классической теории модулярных форм), которое получается из-за того, что можно написать разложение в каспе. Можно выбрать параболическую точку, в ней написать разложение в ряды функций, потом получить соотношение на эти функции — уравнение кривой. Здесь ничего такого сделать нельзя. Отсутствие параболических точек означает, что описывать уравнением кривые Шимуры — очень сложное занятие. Не то чтобы это было совсем невозможно, есть много разных примеров, но нет такого единого метода, который бы с этим справлялся.

Есть очень важный факт (это как раз работа Шимуры 60-х годов), что эти кривые определены над \mathbb{Q} , то есть кривая Шимуры — гладкая проективная алгебраическая кривая над \mathbb{Q} . Это не очень простой факт. И Шимура его доказывал, используя теорию комплексного умножения, то, как действует группа Галуа на всём — таким классическим способом. А современный подход к этому состоит в том, что надо написать проблему модулей; т. е. надо сказать, что эта кривая решает некую задачу модулей, написать объекты, которые надо параметризовать. Это маленькая статья Дринфельда в «Функциональном анализе», которая потом превратилась в большую книгу Катца и Мазура «Arithmetic moduli of elliptic curves».

Кривая Шимуры X решает следующую проблему модулей (это грубое многообразие модулей, но всё равно): она параметризует пары (A, i) , где A — абелева поверхность (значит, абелево многообразие размерности 2), а $i: \mathcal{O} \subset \text{End } A$ — вложение максимального порядка, который мы зафиксировали, в эндоморфизмы. То есть это абелевы поверхности с кватернионным умножением, как говорят. Такое описание достаточно для задания проблемы модулей над \mathbb{Q} ; в принципе, можно поднапрячься и задать проблему модулей над \mathbb{Z} . Труды Дринфельда как раз позволяют это всё сделать. Это нужно уточнить, я не буду про это говорить; там появляются дополнительные трудности, довольно тонкие.

Про эти кривые Шимуры много всего известно. В частности, из трудов Дринфельда можно вывести, как выглядит ее модель над спектром \mathbb{Z} , где редукция хорошая, где плохая. Модель над \mathbb{Z} устроена так: если p не делит дискриминант, то редукция хорошая, т. е. эту кривую можно продолжить до гладкой кривой над спектром \mathbb{Z}_p ; а если p делит дискриминант, то тут

есть абсолютно явное описание того, как устроен плохой слой: все компоненты являются проективными прямыми, известно, как они пересекаются, известно, как действует группа Галуа на компонентах. Это описание получается из p -адической униформизации Чередника—Дринфельда. Это абсолютно фундаментальная вещь, но сейчас у меня нет возможности про это детально говорить.

4.

Точки кривых Шимуры над числовыми полями

Вопрос, который я хотел бы обсудить, такой: что можно сказать про точки кривых Шимуры над числовыми полями. Почему над числовыми полями, а не над \mathbb{Q} ? Потому что над \mathbb{Q} их нет. А над \mathbb{Q} их нет, потому что их нет над \mathbb{R} . Шимура в своих работах выяснил, что над \mathbb{R} эти кривые точек не имеют: $X(\mathbb{R}) = \emptyset$.

Поэтому первое, что осмыслено рассматривать, это мнимо-квадратичные поля. Пусть k — мнимо-квадратичное поле, т. е. такое расширение \mathbb{Q} степени 2, которое получается присоединением корня из какого-нибудь отрицательного целого числа, свободного от квадратов.

Почему надо говорить именно про числовые поля, а не про локальные вначале? Дело в том, что над локальными полями получен окончательный ответ лет 15 назад. Джордан (B. Jordan) и Ливне (R. Livné) нашли необходимое и достаточное условие над локальными полями. То есть над \mathbb{R} всё ясно, над \mathbb{C} всё ясно, остаются еще p -адические поля. Они полностью разобрались с ситуацией с p -адическими полями, т. е. с конечными расширениями \mathbb{Q}_p (когда p делит D и когда не делит — для всех). Простейшая кривая Шимуры, как я ее определил, задается исключительно дискриминантом; я сказал, что надо зафиксировать максимальный порядок, но от этого, в общем, ничего не зависит. Значит, если есть дискриминант и есть конечное расширение поля p -адических чисел, то просто есть некая процедура (не очень короткая).

Интересно, каким методом они действовали. Когда p не делит D , в ситуации с хорошей редукцией, надо просто посчитать число точек над конечным полем. И потом, если у вас есть гладкая точка в замкнутом слое, то по лемме Гензеля она поднимается до точки над p -адическим полем. А как посчитать число точек над конечным полем? Для этого есть большая наука: формула следа Эйхлера—Сельберга для вычисления следа операторов Гекке — очень громоздкая аналитическая формула; это чистый анализ. Поэтому можно написать точную формулу. А когда p делит дискриминант, нужно пользоваться явным описанием p -адической

униформизации Дринфелда и явным описанием компонент вырожденного слоя, которое из нее получается. Но это всё на самом деле вполне вычислимые вещи, с которыми можно работать. Получаются результаты, которые легко использовать на практике.

Поэтому следующий этап — это как раз работа над числовыми полями. И тут нет большой ясности. Сейчас я скажу, что известно над числовыми полями. Я не буду подробно рассказывать, зачем нужны кривые Шимуры. Они сами по себе очень интересны; они используются в теореме Ферма, в гипотезах Ленглендса, они используются везде. Можно, например, спросить, можно ли параметризовать эллиптические кривые кривыми Шимуры. Это осмысленная деятельность, потому что тогда можно строить рациональные точки на эллиптических кривых.

Есть следующая относительно малоизвестная теорема Джордана. Пусть p — простое число, $p \geq 11$, $p \equiv 3 \pmod{4}$, p делит дискриминант. Пусть k — мнимо-квадратичное поле, такое что над k кватернионная алгебра расщепляется, т. е. $B \otimes k \simeq \text{Mat}_2(k)$. И пусть вдобавок p инертно, т. е. p остается простым в k . Это такие условия, которые легко понять; а еще есть условие, которое понять несколько более трудно. Пусть не существует сюръективного гомоморфизма конечных абелевых групп следующего вида:

$$Cl_k^{(p)} \rightarrow Cl_k \times \mathbb{Z}/\frac{p^2-1}{12}.$$

Тогда над k у X нет точек: $X(k) = \emptyset$.

Группа Cl_k — это группа классов поля k . Это фактор группы дробных идеалов по главным идеалам. Напомню, что дробные идеалы — это подмножества $I \subset k$, для которых существует ненулевой элемент $a \in k$, такой что если помножить a на I , то получится обычный идеал в кольце целых \mathcal{O}_k . Идеал I называется главным, если $I = b\mathcal{O}_k$ для некоторого $b \in k^*$. Это аналог группы Пикара. А $Cl_k^{(p)}$ — это дробные идеалы, взаимно простые с p ; факторизовать их нужно по главным идеалам, которые порождаются b , сравнимым с 1 по модулю p :

$$Cl_k^{(p)} = \frac{\text{дробные идеалы, взаимно простые с } p}{(b) : b \equiv 1 \pmod{p}}.$$

Эта группа иногда называется лучевой группой классов с кондуктором p . Как ясно из этого описания, есть точная последовательность:

$$1 \rightarrow (\mathcal{O}_k/p)^* / \mathcal{O}_k^* \rightarrow Cl_k^{(p)} \rightarrow Cl_k \rightarrow 1.$$

Группа классов с кондуктором p отображается в просто группу классов сюръективно, потому что любой идеал можно сдвинуть с p . Такая кон-

струкция известна из геометрии — это конструкция обобщенных якобианов. Cl_k — это группа Галуа максимального абелева расширения k , которое нигде не разветвлено, а $Cl_k^{(p)}$ — такого, которое разветвлено только в p .

Инертность простого числа p означает, что p остается простым в k . Тогда фактор кольца целых по p будет конечным полем, которое квадратично над полем из p элементов. Значит, здесь будет содержаться циклическая группа порядка $p^2 - 1$, которой изоморфна $(\mathcal{O}_k/p)^* = \mathbb{F}_{p^2}^*$. А $\mathcal{O}_k^* = \{\pm 1\}$ — единственные единицы кольца целых, за исключением того случая, когда $k = \mathbb{Q}(\sqrt{-1})$ или $k = \mathbb{Q}(\sqrt{-3})$. Хорошо известный простой результат, что кольцо целых в квадратичном поле содержит нетривиальные корни из 1 только когда оно содержит корень 4-й степени из 1 или корень 6-й степени из 1. То есть в большей части случаев это есть циклическая группа порядка $\frac{p^2 - 1}{2}$. Поэтому на самом деле вполне реально, что такой сюръективный гомоморфизм существует. И действительно, при нынешнем развитии техники, с помощью программы типа Mathematica вы просто пишете ваше мнимо-квадратичное поле и p , и она вам говорит, чему эта группа будет равна. То есть это такие вещи, которые сейчас считаются компьютером мгновенно.

Я приведу вкратце доказательство этой теоремы и объясню происхождение условия инертности. Я также хочу привести пример.

Пусть дискриминант есть произведение двух простых чисел $D = 23 \times 107$, а $k = \mathbb{Q}(\sqrt{-23})$. Число 23 здесь разветвлено, а 107, наоборот, инертно. Можно посчитать символ Лежандра и увидеть, что это достаточно, чтобы убить B переходом к k . А с другой стороны, все условия выполнены. В качестве p я предлагаю взять 107. Тогда лучевая группа классов есть произведение трех циклических групп $Cl_k^{(107)} = \mathbb{Z}/4 \times \mathbb{Z}/81 \times \mathbb{Z}/53$. А просто группа классов — группа из трех элементов: $Cl_k = \mathbb{Z}/3$. И если разложить в произведение циклических групп, там получится 4 циклических фактора. И соответственно, заключение такое, что у этой кривой нет точек над таким полем.

Я не буду выписывать теорему Джордана—Ливне, которую я процитировал. Но ее можно применить, и окажется, что у этой кривой есть точки всюду локально. То есть такая кривая является контрпримером к принципу Хассе: это многообразие, которое имеет точки над всеми пополнениями основного поля. Принцип Хассе — это такой принцип (иногда он бывает верен, а иногда нет), который говорит, что в некоторых случаях, для некоторых классов многообразий из наличия точек над всеми пополнениями следует наличие точек над глобальным полем. В нашем примере принцип Хассе не выполнен.

5.

Препятствие Манина

Манин 35 лет назад предложил подход, который дает общий способ построить препятствие к принципу Хассе. Я отвлекусь от темы о кривых Шимур, поговорю просто о препятствии Манина. Если X — многообразие над числовым полем k , такое что над всеми пополнениями оно имеет рациональные точки ($X(k_v) \neq \emptyset$ для всех v), то можно рассмотреть произведение множеств локальных точек по всем нормированиям $\prod X(k_v)$ и в нем определить некоторое подмножество $\prod X(k_v)^{\text{Br}}$: это множества локальных точек (по одной локальной точке для каждого нормирования), такие что сумма локальных инвариантов элементов группы Брауэра X равна 0:

$$\prod X(k_v)^{\text{Br}} = \left\{ (P_v), P_v \in X(k_v), \sum \text{inv}_v A(P_v) = 0 \right\}.$$

Здесь A принадлежит группе Брауэра $\text{Br } X = H^2(X, \mathbb{G}_m)$. Группу Брауэра многообразия X можно определить разными способами, например, как вторую группу этальных когомологий с коэффициентами в \mathbb{G}_m . По ее функториальности можно рассмотреть специализацию ее элемента в любой точке, например, в P_v . Тогда получится элемент группы Брауэра локального поля $A(P_v) \in \text{Br } k_v$. Локальная теория полей классов доставляет отображение inv_v , которое задает изоморфизм $\text{Br } k_v$ с \mathbb{Q}/\mathbb{Z} (для \mathbb{R} — с циклической группой из двух элементов). Если набор локальных точек происходит на самом деле из одной точки с координатами в k , то тогда значение A в этой точке будет элементом группы Брауэра поля k , и закон взаимности глобальной теории полей классов говорит, что сумма всех локальных инвариантов равна 0. Таким образом, если вложить k -точки X диагонально в произведение $\prod X(k_v)$, то они будут лежать в $\prod X(k_v)^{\text{Br}}$. Получается такой промежуточный объект, как бы верхняя граница, вместе с k -точками в этом произведении:

$$X(k) \subset \prod X(k_v)^{\text{Br}} \subset \prod X(k_v).$$

Смысл того, что Манин ввел группу Брауэра, состоит в том, что в ряде случаев она вычисляется. Для кубических поверхностей она по модулю группы Брауэра k — просто конечная группа. Была большая деятельность, которая продолжается, по исследованию всех мыслимых контрпримеров к принципу Хассе и объяснению того, что на самом деле происходит в ситуации, когда локальные точки есть всюду, а глобальных нет. Иногда это происходит за счет препятствия Манина, т. е. за счет того что уже

$\prod X(k_v)^{\text{Br}}$ пусто. Если можно доказать, что $\prod X(k_v)^{\text{Br}}$ пусто, то автоматически $X(k)$ тоже пусто. Это универсальный подход, который, когда он работает, позволяет сводить трудную задачу решения диофантова уравнения над k (над числовым полем) к более простой задаче. Если группа Брауэра конечна, то в принципе условия, которыми $\prod X(k_v)^{\text{Br}}$ задается в $\prod X(k_v)$, вычислимы.

Но группа Брауэра не всегда конечна. В частности, если X — гладкая проективная кривая (именно это мы сегодня обсуждаем), то интересная часть группы Брауэра (фактор по группе Брауэра поля, которая никакой роли здесь не играет) — это когомологии Галуа с коэффициентами в группе Пикара кривой над замыканием: $\text{Br } X / \text{Br } k = H^1(k, \text{Pic } \bar{X})$, где $\bar{X} = X \times_k \bar{k}$. И это является фактором первой группы когомологий с коэффициентами в якобиане $J = \text{Jac}(X)$.

Вопрос о том, является ли препятствие Манина на кривых единственным препятствием к принципу Хассе (иначе говоря, верно ли, что всегда, когда на данной кривой есть локальные точки, а глобальных точек нет, то это происходит из-за того, что $\prod X(k_v)^{\text{Br}}$ пусто) в очень сильной степени является открытым. В частности, известно только, что это так для кривых рода 1, что получается немедленно из стандартных глобальных теорем двойственности. Известны также отдельные очень частные случаи с разными дополнительными условиями. Вопрос этот имеет смысл не только как чисто теоретическое удовлетворение любопытства, но он еще занимает людей, которые занимаются вычислениями. Есть очень много специалистов по вычислительной теории чисел, которые любят решать такие задачи: дана явно кривая (большого рода), найти все рациональные точки. И они находят эту идеологию полезной.

Кривая X над k может быть двух типов: такая кривая, у которой есть класс дивизоров степени 1, определенный над k , и такая, у которой его нет. Если считать, что X его не имеет, то второй случай трудности не представляет, потому что в этом случае X можно представить себе как вложенную в главное однородное пространство якобиана, которое не имеет рациональных точек. Тогда вопрос просто решается опять же из глобальных теорем двойственности, при условии что мы примем гипотезу, что группа Тейта—Шафаревича якобиана конечна: $|\text{Ш}(J)| < \infty$.

А в первом случае ситуация такая. Если такой класс дивизоров существует, то можно использовать его для того, чтобы вложить X в его собственный якобиан: $X \subset J$. И всё сводится, используя опять же стандартные теоремы, которые я всё время упоминаю, к такой задаче. k -рациональные точки X можно вложить в произведение точек X над локальными полями,

внутри такого же произведения для якобианов:

$$X(k) \hookrightarrow \prod X(k_v) \cap \text{замыкание } J(k).$$

Возьмем замыкание рациональных точек якобиана в топологии произведения его локальных точек. Есть теорема Серра, опубликованная в двух статьях в «Известиях Академии Наук СССР», о том, что это замыкание — то же самое, что проконечное пополнение $J(k)$; если эта группа конечна, слово «замыкание» можно забыть. Ясно, что k -точки X лежат в вышеуказанном множестве. Задача состоит в том, верно ли, что здесь имеется равенство, что включение можно заменить на равенство? И это ровно есть то, как формулируется проблема о единственности препятствия Манина для кривых, в этом трудном случае. Равенство есть тогда и только тогда, когда препятствие Манина объясняет такие контрпримеры к принципу Хассе.

Я не буду это использовать, просто хочу привлечь внимание к этой проблеме. Очень трудно доказать, что это неверно. Есть люди, которые считают, что все контрпримеры для кривых к принципу Хассе должны объясняться препятствием Манина.

Я недавно присутствовал на докладе, где Пунен (В. Рoonen) проделал такие чисто вероятностные рассуждения. Например, можно сделать некоторое приближение к этой задаче: можно рассмотреть произведение точек над конечными полями. Пусть у нас есть X — кривая над \mathbb{Q} и на ней есть рациональный класс степени 1. Можно вложить X в якобиан и рассмотреть

$$\prod X(\mathbb{F}_q) \cap \text{замыкание образа } J(\mathbb{Q}) \subset \prod J(\mathbb{F}_p).$$

(Здесь произведение по всем p ; а можно брать некоторые p , тогда будет множество конечно, тогда замыкания нет.) И он сделал такое замечание: если считать, что на кривой находится в среднем $p + 1$ точка, на якобиане сколько-то точек, то размеры этих множеств таковы, что они не должны пересекаться случайным образом; т. е. если эти подмножества являются случайными, то пересечение должно быть пусто. Это навело на мысль о том, что гипотеза должна быть верна. Но присутствовавший при этом Серр сказал, что он не убежден. Конечно, проблема в том, когда простых бесконечное количество. По тем же соображениям Пунен считал, что это должно быть верно для абелевых многообразий.

Этот метод любят специалисты по вычислительной теории чисел, потому что он позволяет строить алгоритмы, которые для разных кривых находят точки, определенные над \mathbb{Q} , которые, в принципе, довольно сложно искать.

6.

Метод спуска

Теперь я хочу рассказать про метод спуска, который в практических случаях часто удобнее для тех же целей. Метод спуска имеет дело с конечным неразветвленным накрытием. Пусть $Y \xrightarrow{f} X$ — отображение, которое получается факторизацией по свободному действию конечной группы G (или конечной групповой схемы), $X = Y/G$. Тогда Y — это X -торсор относительно G (или G -торсор на X). Типичный пример этой ситуации — это изогении эллиптических кривых.

Пусть у нас есть 1-коцикл, т. е. элемент группы когомологий Галуа $\xi \in H^1(k, G)$, где G коммутативна. Тогда есть операция скручивания Y на ξ . (Как известно, если группа действует на каком-нибудь объекте, например, на многообразии, то имея класс из первой группы когомологий, можно скрутить то, на чем она действует, на этот класс.) Скрученный объект Y^ξ , так же как и Y , отображается на X . На нем действует G , и $f^\xi: Y^\xi \rightarrow X$ является торсором над X относительно G . И первое следствие (это верно для любого поля): множество рациональных точек $X(k)$ является непересекающимся объединением образов рациональных точек на всех скрученных формах:

$$X(k) = \bigsqcup_{\xi \in H^1(k, G)} f^\xi(Y^\xi(k)).$$

А объясняется это таким образом: мы берем точку P на X , смотрим на ее слой; P — точка определенная над k , поэтому группа Галуа поля k будет действовать, переставляя точки слоя. Иными словами, она будет действовать на слое при помощи некоторого характера. В ситуации, когда на G нет действия группы Галуа, можно думать, что это просто гомоморфизмы группы Галуа в G . И таким образом точке P , лежащей на X и определенной над k , отвечает такой гомоморфизм. И если я скручу ровно на этот самый гомоморфизм, то окажется, что в прообразе P у меня есть рациональные точки (все точки в прообразе будут рациональными). Каждая точка на X поднимается до рациональной точки ровно на одной из скрученных форм.

Это пока общая картина. Применение к арифметике состоит в следующем. Опять пусть k — числовое поле; если есть k -рациональная точка на X , то существует единственный коцикл $\xi \in H^1(k, G)$, такой что скрученная форма Y^ξ тоже имеет k -точку: $Y^\xi(k) \neq \emptyset$. Это следует из того, что я сказал. И можно утверждение ослабить: из непустоты $X(k)$ следует, что существует $\xi \in H^1(k, G)$ такое, что $Y^\xi(k_v) \neq \emptyset$ для всех v . Это абсолютно тривиальное, банальное наблюдение. Его можно перевернуть и сказать

следующее: если для любого ξ существует v такое, что $Y^\xi(k_v)$ пусто, то $X(k)$ тоже пусто — упражнение в логике. Но несмотря на абсолютно банальный характер этих замечаний, они полезны на практике, потому что позволяют получить нетривиальную информацию о наличии точек на X над числовым полем, исходя из чисто локальных вычислений на Y . Можно сделать еще замечание, что если X проективно, то достаточно рассматривать только конечное количество классов ξ и конечное количество нормирований v ; т. е. на самом деле это конечная процедура. Этот принцип используется традиционно для вычисления группы Сельмера.

Пусть теперь у нас имеется торсор $f: Y \rightarrow X$ такой, что для любого ξ существует v , для которого $Y^\xi(k_v)$ пусто. Если X имеет точки над всеми локальными полями, то, как мы видели, X — контрпример к принципу Хассе. Нетривиальная, хотя и не очень сложная теорема (Colliot-Thélène, Sansuc) утверждает, что любой такой контрпример объясняется препятствием Манина. То есть, если у вас есть набор локальных точек на X , лежащий в $\prod X(k_v)^{\text{Br}}$, то какое бы ни было неразветвленное накрытие $Y \rightarrow X$, можно подобрать скрученную форму, на которую они все разом подымутся. Соответственно, это дает некоторый рецепт на практике, как доказать тот факт, что контрпример к принципу Хассе объясняется при помощи препятствия Манина, таким достаточно простым способом, работая с одним накрытием. Понятно, что это очень маленькая грань, потому что препятствие Манина, как я говорил, связано с бесконечной группой (H^1 с коэффициентами в якобиане), а тут это как бы его маленькая тень. Тем не менее, это эффективный метод.

7.

Накрытие Шимуры

Теперь моя цель состоит в том, чтобы подойти к результату Джордана, который я выше сформулировал, при помощи этого подхода. Если X — кривая Шимуры, то какова может быть кривая Y ? Понятно, что нелепо ожидать, что будет только одно накрытие, которое годится для всех мыслимых контрпримеров к принципу Хассе на кривых Шимуры. Но есть одно такое совершенно замечательное накрытие, которое объясняет теорему Джордана. Оно называется накрытием Шимуры.

Давайте построим Y следующим образом. Я сказал (возвращаясь к предыдущей проблематике), что редуцированная норма отображает \mathcal{O} в \mathbb{Z} . Рассмотрим простые числа p , которые делят дискриминант, и рассмотрим целые числа, кратные p . Их прообраз назовем I_p . Тогда I_p — это двусторонний идеал в \mathcal{O} , и у него есть разные приятные свойства.

Например, его квадрат — это просто главный идеал, порожденный p ; фактор по нему — конечное поле $\mathcal{O}/I_p \cong \mathbb{F}_{p^2}$. Я хочу использовать аналог конгруэнц-подгруппы. Пусть $\Gamma_p = \{x \in \mathcal{O}^+ \mid x \equiv 1 \pmod{I_p}\}$. Это такие элементы \mathcal{O}^+ (я напомним, что это кватернионы из максимального порядка $\mathcal{O} \subset B$, которые имеют норму 1), что x сравним с 1 по модулю p . Это выделяет некоторую подгруппу в \mathcal{O}^+ ; и я могу ей действовать на верхней полуплоскости и рассмотреть фактор $\Gamma_p \backslash \mathbb{H}$. Это тоже компактная риманова поверхность, чуть более общего класса, чем раньше.

Но она имеет следующий дефект. Шимура вычислил поле, над которым будет определена эта алгебраическая кривая. Можно представить, что это алгебраическая кривая над числовым полем; но это поле оказывается не \mathbb{Q} . Она просто не определена над \mathbb{Q} . Определена она над круговым полем $\mathbb{Q}(\mu_p) = \mathbb{Q}(\sqrt[p]{1})$, получающимся добавлением корня p -ой степени из 1.

Иными словами, то, что мне нужно, будет приводимой кривой, уже определенной над \mathbb{Q} : X_p будет кривая над \mathbb{Q} , состоящая из $p - 1$ неприводимых компонент, изоморфных $\Gamma_p \backslash \mathbb{H}$. И есть забывающее отображение $X_p \rightarrow X$.

Кривую X_p можно также рассматривать как многообразие модулей с дополнительной структурой. Напомним, что X было многообразием модулей пар (A, i) , где A — абелева поверхность, i — это вложение $\mathcal{O} \subset \text{End } A$. У нас есть действие \mathcal{O} на A (кватернионное умножение), и надо рассмотреть подгруппу $A[I_p] \subset A$, которая аннулируется умножением на I_p . Это аналог кривой с комплексным умножением, где можно рассматривать ядро умножения на разные элементы кольца целых соответствующего мнимо-квадратичного поля. Имеем $A[I_p] = \mathcal{O}/I_p \simeq \mathbb{F}_{p^2}^*$. Тогда X_p — грубое многообразие модулей троек (A, i, P) , где P — образующая $A[I_p] \subset A$ как \mathcal{O} -модуля.

Проекция на X просто забывает про P : $(A, i, P) \mapsto (A, i)$. Это отображение, к сожалению, разветвлено. То, что называется накрытием Шимуры, является его максимальным этальным фактором. Я хочу рассмотреть промежуточное пространство Y . В дальнейшем p будет не равно 2 и 3. Тогда отображение $X_p \rightarrow X$ является накрытием Галуа с группой $\mathbb{F}_{p^2}^*/\pm 1$, так как точку P можно выбрать $p^2 - 1$ способами.

$$\begin{array}{ccc}
 X_p & \xrightarrow{\text{группа } \mathbb{Z}/\frac{p^2-1}{2}} & X \\
 & \searrow & \nearrow \\
 & & Y & \xrightarrow{\text{группа } \mathbb{Z}/\frac{p^2-1}{12}} & X
 \end{array}$$

Пусть Y есть то единственное промежуточное накрытие Галуа, для которого группа Галуа Y/X — это $\mathbb{Z}/\frac{p^2-1}{12}$. Утверждается, что отображение $Y \xrightarrow{f} X$ неразветвлено. Этот факт, как вы понимаете, чисто топологический, потому что речь идет о действии некоторых фуксовых групп на верхней полуплоскости, и там у них есть эллиптические точки, за счет которых это отображение не является неразветвленным, и надо просто посчитать стабилизаторы и по ним отфакторизовать. Это факт на уровне классификации матриц, у которых есть неподвижные точки в верхней полуплоскости.

Теперь я могу абстрактный аппарат теории спуска, о котором я говорил выше, применить в нашей конкретной ситуации. Моя цель состоит в том, чтобы доказать, что для любого характера группы Галуа $\varphi \in \text{Hom}\left(\text{Gal}(\bar{k}/k), \mathbb{Z}/\frac{p^2-1}{12}\right)$ существует v , для которого соответствующая скрученная форма не имеет локальных точек: $Y^\xi(k_v) = \emptyset$. Но я не хочу ничего вычислять про Y , я хочу действовать иначе. Эквивалентная формулировка состоит в следующем. Мы рассматриваем любые наборы локальных точек P_v на X . Как я уже говорил, если Y неразветвлено накрывает X и имеется рациональная точка на X , то на слое группа Галуа действует некоторым характером, переставляя точки слоя. Соответственно, любая локальная точка определяет локальный характер: для любой точки $P_v \in X(k_v)$ определен характер

$$\varphi \in \text{Hom}\left(\text{Gal}(\bar{k}/k), \mathbb{Z}/\frac{p^2-1}{12}\right).$$

Эквивалентная формулировка требуемого утверждения состоит в том, что никакой набор таких характеров, задаваемых локальными точками, не происходит из глобального характера. Значит, надо доказать, что для любого семейства точек (P_v) характеры φ_v не происходят из глобального характера φ . Это всё, что нужно сделать.

Если я это сделаю, то по теореме теории спуска, которую я процитировал, я смогу заключить, что контрпримеры к принципу Хассе, которые получаются на основании теоремы Джордана, объясняются препятствием Манина. Сейчас я опишу кратко план доказательства, не вдаваясь в детали. Всё сводится к такой арифметической проблеме: как здесь найти противоречие, какие нужно использовать свойства этих локальных характеров, которые делают их несовместимыми друг с другом?

Есть два утверждения. Первый факт такой. Если рассмотреть нормирование $v \neq p$, то φ_v неразветвлен; иными словами, его ограничение на группу инерции тривиально. Объяснение этому можно дать совершенно геометрическое: надо рассмотреть X и Y как схемы над \mathbb{Z} и доказать,

что отображение $Y \rightarrow X$ будет этальным накрытием вне p . Это следует из того, что Y получается из X наложением условия в p (дополнительное условие касается только p). И из неразветвленности этого морфизма схем следует, что характер, который таким образом получается, будет тоже неразветвлен.

А второй факт состоит в том, что если $v = p$, то характер φ_v , наоборот, полностью разветвлен. Что я имею в виду, когда говорю «полностью разветвлен»? Ограничение на подгруппу инерции сюръективно отображает ее туда, куда характер бьет; т. е. характер сюръективен на подгруппе инерции группы Галуа максимального абелева расширения k_v .

Эти факты я не буду доказывать. То геометрическое объяснение, которое я дал, использует интерпретацию кривых Шимурь как грубых схем модулей. Можно этого не использовать, можно действовать по-другому. Но в любом случае используется теория полей классов и другие результаты: теоремы Нерона—Огга—Шафаревича и Серра—Тейта. А второй факт основан на теореме Мишеля Райно (Michel Raynaud) про конечные плоские группы.

Мы знаем следующее обстоятельство: эллиптическая кривая определена над тем полем, над которым определен ее модулярный инвариант. Но пара (A, i) (абелева поверхность с кватернионным умножением) уже не обязательно определена над полем вычетов $k(P)$ точки P , которая в пространстве модулей ее задает, потому что это только грубое пространство модулей. Джордан доказал, что это происходит ровно тогда, когда $k(P)$ расщепляет B , т. е. B становится изоморфной матричной алгебре при подъеме поля с k на $k(P)$.

И теперь, соединяя все эти факты, мы получаем противоречие с условием теоремы. А именно, таким образом: если глобальный характер φ неразветвлен всюду вне p , то он пропускается через лучевую группу классов $Cl_k^{(p)}$. Поскольку φ полностью разветвлен в p , то его ограничение на подгруппу инерции дает сюръективное отображение. И отсюда получается противоречие с условием теоремы. Примерно такова схема доказательства. Всё это доказательство, должен сказать, на самом деле является более или менее пересказом того, что сделал Джордан. Надо просто чуть-чуть подправить его местами, и тогда оно переводится на геометрический язык.

23 декабря 2004 г.

КРИСТАЛЛИЧЕСКАЯ КОМБИНАТОРИКА

Масаки Кашивара придумал кристаллы — это цветные ориентированные графы, у которых ребра покрашены некоторым множеством цветов, без монокромных циклов и в которых выполняются некоторые условия на «взаимодействия» ребер разного цвета с общей вершиной *). В этом классе кристаллов находятся так называемые регулярные кристаллы, которые соответствуют интегрируемым представлениям соответствующих алгебр. Для каждой (классической) картановской матрицы, регулярные кристаллы образуют тензорную категорию, изоморфную категории представлений соответствующей (классической алгебры) алгебры Каца—Мули. Кристаллы — это комбинаторные «скелеты» представлений, на которых можно отвечать на многие вопросы теории представлений, используя комбинаторику.

Я расскажу про элементарную конструкцию кристаллов Кашивары типа A и как ее можно использовать для некоторых фундаментальных конструкций в комбинаторике, таких как соответствие Робинсона—Шенстеда—Кнута, правило Литтлвуда—Ричардсона, инволюции Шютценберже и многие другие. Этот новый взгляд мы развивали совместно с В. И. Даниловым **).

Основной объект исследования — это массивы. Массив $a \in \mathcal{A}(I, J)$, где $I = \{1, \dots, n\}$ и $J = \{1, \dots, m\}$ множества столбцов и строк массива, соответственно, это неотрицательная матрица в декартовых координатах. Поскольку мы не умножаем и не складываем массивы, а определим на них другие операции, мы их называем массивами, а не матрицами. Представлять их себе будем так: возьмем двумерную целочисленную решетку, \mathbb{Z}^2 ,

*) Эти условия определяются обобщенной картановской матрицей. На диагонали такой матрицы стоят двойки, а внедиагональные члены неположительны. С этими матрицами можно связать системы корней, что ведет к теории алгебр Ли, вообще говоря бесконечномерных. Среди таких матриц есть и конечномерные, про соответствующие картановские матрицы говорят, что они классические.

***) Данилов В. И., Кошевой Г. А. Массивы и комбинаторика таблиц Юнга // УМН. 2005. Т. 60, № 2(362). С. 79—142.

в ней выделяем прямоугольник до n по горизонтали, до m по вертикали, и смотрим на это как на набор единичных $m \times n$ ящиков. В эти ящики можно помещать шарики (в дискретном случае) и массы чего-либо (в непрерывном варианте). В дискретном случае (который будет основным), в ящик $(i, j) \in I \times J$ мы помещаем $a(i, j)$ шаров.

Введем операции на множестве целочисленных массивов. Эти операции перемещают шарики массива или вправо, или влево, или вверх, или вниз. Основное что мы хотим от этих операций — их обратимость при эффективном действии.

Определим операцию перемещения вниз на двухстрочном массиве $a \in \mathcal{A}(I, \{1, 2\})$. Представим, что шарики в верхней строчке — это мужчины, а в нижней — женщины. А положение шарика — зарплата, т. е. $a(i, 1)$ — это количество женщин, получающих зарплату $i \in I$, соответственно $a(i, 2)$ — количество мужчин с зарплатой i . Мужчины и женщины хотят заключить браки и имеют следующие предпочтения. Каждый мужчина хочет взять себе жену, которая зарабатывает строго меньше, чем он, но как можно больше среди такого множества женщин. Ему хочется быть главой семьи, но при этом чтобы жена зарабатывала как можно больше, не нарушая его главенства. Женщина, наоборот, хочет выбрать мужа, который зарабатывал бы больше ее, но с минимальной возможной разницей. То есть ей хочется, чтобы муж не очень сильно командовал дома. Теперь скажем, какие множества браков (паросочетания) устойчивы. Паросочетание называется *устойчивым*, если мужчины и женщины состоят в браке не более чем с одним партнером, и если не бывает таких ситуаций: а) два человека (мужчина и женщина) остались одинокими, хотя могли бы вступить в брак; б) какой-нибудь мужчина женился, но есть незамужняя женщина, которая для него более привлекательна (это нехорошо, он будет тянуться к этой незамужней), и то же самое для женщин: если какая-нибудь женщина вышла замуж, но есть свободный мужчина, который для нее более предпочтителен. Ясно, что в устойчивом паросочетании мужчина или женщина, состоящие в браке, не могут улучшить своего положения, не разбив другой семьи. Легко понять, что устойчивые паросочетания существуют.

Рассмотрим устойчивое паросочетание. Операция D (down) опускания вниз определяется по такому правилу: нужно взять самого правого (т. е. самого богатого) неженатого мужчину и объявить, что это на самом деле женщина; и перевести его в нижний ряд (опускаем шарик в соответствующем столбце из второй строки в первую). Обратная к ней операция U (up) — поднятие: нужно взять самую бедную незамужнюю женщину и, наоборот, ее объявить мужчиной, т. е. поднять соответствующий шарик

вверх. Несложно проверить что эти операции не зависят от того какое устойчивое паросочетание используется для определения операций D и U . Если есть мужчины и женщины не вступившие в брак, тогда эти операции — обратимы. То есть, если операция D переводит массив a в новый массив, $a \neq Da$, тогда операция поднятия переводит измененный массив в первоначальный, $UDa = a$. Аналогичное свойство выполняется для U .

Давайте рассмотрим какой-нибудь простой пример двухстрочного массива и действия на нем операций D и U :

$$\begin{array}{cccc} 1 & 1 & 3 & 2 \\ 2 & 0 & 3 & 1 \end{array}$$

Применим операцию D , тогда получим следующее распределение шариков по ящикам:

$$\begin{array}{cccc} 1 & 1 & 2 & 2 \\ 2 & 0 & 4 & 1 \end{array}$$

Если еще раз применим операцию D , тогда получится

$$\begin{array}{cccc} 1 & 1 & 1 & 2 \\ 2 & 0 & 5 & 1 \end{array}$$

Еще раз применив D , получим

$$\begin{array}{cccc} 0 & 1 & 1 & 2 \\ 3 & 0 & 5 & 1 \end{array}$$

Этот массив уже не будет изменяться под действием операции D , такие массивы называются **D**-плотными.

Теперь можно применить операции U и пройти снизу вверх по этим массивам и дойти до **U**-плотного массива

$$\begin{array}{cccc} 1 & 1 & 4 & 3 \\ 2 & 0 & 2 & 0 \end{array}$$

который не изменяется при действии U .

В случае массива с большим числом строк $a \in \mathcal{A}(I, J)$, $J = \{1, \dots, m\}$, операция опускания, которая действуют на двух соседних строчках j и $j+1$ обозначается D_j , а обратная к ней — U_j , $j = 1, \dots, m-1$. Какие свойства есть у этих операций? Во-первых, $D_j U_j = U_j D_j = \text{id}$, если они действуют эффективно. Поэтому эти операции образуют полугруппу. Во-вторых, если $|j - j'| > 1$, тогда они коммутируют: $D_j D_{j'} = D_{j'} D_j$, а если они соседние, $|j - j'| = 1$, тогда или они коммутируют или выполняется соотношение

Верма $D_j D_{j+1}^2 D_j = D_{j+1} D_j^2 D_{j+1}$. При этом какое соотношение выполняется, коммутирование или Верма, зависит от массива к которому они применяются *).

Определим операцию $\mathbf{D}_j = D_j^\infty$, которая применяет операцию D_j до тех пор пока действие стабилизируется. Эта операция идемпотентна $\mathbf{D}_j^2 = \mathbf{D}_j$ и удовлетворяет соотношению Янга—Бакстера или Коксетера—Мура: $\mathbf{D}_j \mathbf{D}_{j+1} \mathbf{D}_j = \mathbf{D}_{j+1} \mathbf{D}_j \mathbf{D}_{j+1}$.

Набор операций D_j и U_j , $j = 1, \dots, m-1$, задает структуру кристалла Кашивары типа A_{m-1} на массивах $A(I, J)$. Для этого покажем что эти операции «правильно» связаны с отображением «весов» (строчные суммы).

Отображение весов посылает массив a в вектор его строчных сумм, J -вес:

$$a \rightarrow wt(a) = \left(\sum_i a(i, 1), \dots, \sum_i a(i, m) \right).$$

$wt(a)$ — вектор в m -мерном пространстве (для целозначного массива это вектор в решетке весов).

Покажем как отображение весов связано с операциями опускания и поднимания D_j и U_j . Вернемся к интерпретации этих операций через бракосочетания.

После заключения браков в устойчивом паросочетании, удалим шарики, соответствующие всем мужчинам и женщинам, вступившим в брак. Тогда получится массив, в котором любой одинокий мужчина в $j+1$ -й строчке расположен левее (не строго) любой одинокой женщины в j -й строчке. Вот типичный случай

○	○		○○		
			○		○○

Обозначим через $\varepsilon_j(a)$ число одиноких мужчин, а через $\varphi_j(a)$ — число одиноких женщин, тогда $\varepsilon_j(a)$ равно числу эффективных действий D_j , а $\varphi_j(a)$ — числу эффективных действий U_j , и выполняется следующее соотношение:

$$\varepsilon_j(a) - \varphi_j(a) = wt_{j+1}(a) - wt_j(a).$$

Это соотношение следует из того, что супружеские пары дают равный вклад в координаты веса $wt_{j+1}(a)$ и $wt_j(a)$. Также из брачной интерпретации операций на массивах понятно, что если операция D_j действует

*) Подробности можно найти в нашей статье с В. И. Даниловым и А. В. Карзановым «Combinatorics of A_2 -crystals» (J. of Algebra. 2007. V. 310. P. 218—234).

эффективно, тогда вес $wt(D_j a)$ связан с весом $wt(a)$ следующим образом

$$wt(D_j a) = wt(a) + e_j - e_{j+1},$$

где e_j — это единичный вектор в \mathbb{R}^J , который имеет единичку на j -м месте: $e_j = (0 \dots 1 \dots 0)$.

Такое согласование действий операций D_j и U_j на массивах с весовым отображением wt задает на $\mathcal{A}(I, J)$ (для любого I) кристалл Кашивары типа A_J (*Kashiwara M. Bases Cristallines des Groupes Quantiques // Cours Spécialisés 9, Société Math. France, 2002*). Особенный интерес состоит в описании неразложимых (неприводимых) кристаллов, поскольку на таких кристаллах можно изучать комбинаторику неприводимых представлений GL . Сейчас мы к этому приступим.

Определим категорию кристаллов $\mathcal{M}(J)$, $J = \{1, \dots, m\}$. Объекты этой категории — инвариантные подмножества массивов с множеством строк J и конечным множеством столбцов (т. е. подмножества $\mathcal{A}(I, J)$ с некоторым I , которое не фиксируется) относительно действия операций D_j, U_j , $j = 1, \dots, m - 1$. Морфизмами являются отображения таких множеств, которые коммутируют с операциями поднятия и опускания. На этой категории можно научиться теории представлений $GL(m)$. А именно, объекты категории (инвариантные подмножества) — это аналог представлений; но, вообще говоря, приводимых. Раскладываются они в объединение минимальных инвариантных подпространств, которые называются орбитами. Орбита устроена так. Берем какой-нибудь массив a и начинаем на него действовать всевозможными словами из операций вверх и вниз, т. е. получаем множество массивов вида $m(D_1, D_2, \dots, D_{m-1}, U_1, \dots, U_{m-1})a$, где m — моном от переменных D_j, U_j , $j = 1, \dots, m - 1$.

У каждой орбиты есть выделенный **D-плотный конец** — это такой массив, который инвариантен относительно любой операции D_j , т. е. это такой массив b , что $D_j(b) = b$ для всех $j = 1, \dots, m - 1$.

Действительно, если мы берем какой-то элемент орбиты и начинаем на него действовать мономами от D_j , $j = 1, \dots, m - 1$, то поскольку каждая операция D_j опускает некоторое неотрицательное количество шариков, то через конечное число применений таких операции никакие шарики не будут опускаться, и мы получим массив, на котором операции D_j действуют идемпотентно, т. е. **D-плотный конец**.

Могло бы оказаться, что из какого-нибудь массива можно получить несколько **D-плотных концов**. Этого не случается, я позже объясню, в чем тут дело.

Итак, у каждой орбиты есть один **D-плотный конец**.

D-плотные концы или **D**-плотные массивы — это ничто иное как таблицы Юнга, объекты хорошо известные в комбинаторике и теории представлений (см., например, книжку У. Фултона «Таблицы Юнга и их применения к теории представлений и геометрии», М.: МЦНМО, 2006). **D**-плотный массив b можно превратить, как и любой массив, в таблоид.

Таблоид — это графическое изображение массива в декартовых координатах, или переход в каменный век, когда число $b(i, j)$ заменяется на $b(i, j)$ камешков на которых написан символ i и которые последовательно располагаются в j -м ряду. Поясним на примере, пусть у нас есть массив

$$\begin{array}{cccc} 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 3 \end{array}$$

Мы его превращаем в таблоид. Для этого мы его прочитаем: в 1-й строчке массива написано 1, 0, 0, 3; это преобразуется в один камешек, помеченный единицей, и три камешка, помеченных четверкой, т. е. в соответствующем таблоиде в первой строке стоит 1, 4, 4, 4; во 2-й строчке массива написано 0, 0, 1, 0 что соответствует второй строчке таблоида с одним камешком с номером 3; 3-й строчке 2, 0, 0, 1 соответствует два камешка с номером 1 и один с номером 4, т. е. 1, 1, 4, и, наконец, 4-й строчке 0, 1, 0, 0 соответствует один камешек с цифрой 2. Получается такой таблоид:

$$\begin{array}{cccc} & & & 2 \\ & & & 1 & 4 & 4 \\ & & & 3 \\ & & & 1 & 4 & 4 & 4 \end{array}$$

Характеризацией **D**-плотных массивов является то, что соответствующие таблоиды являются *полустандартными таблицами Юнга*. А именно, соответствующий **D**-плотному массиву таблоид-таблица будет устроен так: во-первых, в каждой следующей строке число камешков будет не больше, чем в предыдущей; во-вторых, вдоль каждой строчки стоящие числа (символы на камешках) не строго возрастают, а вдоль каждого столбца — строго.

Форма таблоида — последовательность, составленная из количества камешков в каждой строке. У таблицы форма является невозрастающей последовательностью, λ_1 клеток в первой строке, больше или равно λ_2 клеток во второй и т. д.

Например, массив

0	0	1	3	4
0	3	3	3	3
4	2	3	4	0

является **D**-плотным и у него такая полустандартная таблица Юнга

3	4	4	4	5	5	5	5					
2	2	2	3	3	3	4	4	4	5	5	5	
1	1	1	1	2	2	3	3	3	4	4	4	4

Напомним, что морфизмы — это такие отображения инвариантных подмножеств (объединения орбит), которые коммутируют с операциями U_j и D_j .

В категории $\mathcal{M}(J)$ выполняется лемма Шура, которая говорит, что если есть две орбиты O и O' , и морфизм $f: O \rightarrow O'$, то тогда f или изоморфизм, или O' состоит из единственной точки (из единственного массива). Таким образом есть много изоморфных орбит, и они «помечены» разными **D**-плотными массивами, т. е. таблицами Юнга. Оказывается, что формы у таблиц, помечающих изоморфные орбиты, одинаковые. По аналогии с теорией представлений, **D**-плотные концы орбит соответствуют векторам старшего веса. Форма таблицы соответствует весу представления. Кристаллический подход позволяет пометить орбиты не весом, а таблицами, и изоморфные орбиты персонифицируются своими таблицами Юнга. То есть можно различать изоморфные «представления» $GL(m)$.

Чтобы лучше понять структуру орбит, введем другие две серии операций L_i и R_i , $i = 1, \dots, n - 1$. Возьмем два соседних столбца i и $i + 1$, и определим операции L_i и R_i следующим образом: транспонируем массив (относительно диагонали в декартовых координатах), $a \rightarrow a^t$. К транспонированному массиву применим операцию D_i и транспонируем $D_i a^t$, т. е. получим $(D_i a^t)^t$. Этот массив и определяет действие L_i на a ,

$$L_i a = (D_i a^t)^t.$$

Обратная операция определяется из $R_i a = (U_i a^t)^t$.

Если определить весовую функцию ωt^t как вектор столбцовых сумм массива, I -вес, тогда операции L_i и R_i , $i = 1, \dots, n - 1$, определяют на $\mathcal{A}(I, J)$ кристалл Кашивары типа A_I .

Важный результат, который и составляет суть кристаллической комбинаторики, состоит в том, что эти две кристаллические структуры коммутируют. А именно, справедлив кристаллический аналог двойственности Хоу.

Теорема. Для любых i и j справедливы коммутирования

$$L_i D_j = D_j L_i \quad \text{и} \quad L_i U_j = U_j L_i.$$

Из этой теоремы, в частности, следует, что каждая орбита относительно пары операций D_j и U_j , $j = 1, \dots, m - 1$, имеет ровно один **D**-плотный конец.

Рассмотрим две изоморфные D, U -орбиты. Тогда две таблицы Юнга, соответствующие **D**-плотным концам в этих орбитах, имеют одну и ту же форму. Действительно, из этой теоремы следует, что если на все элементы орбиты подействовать любой из операций L_i и R_i , $i = 1, \dots, n - 1$, то получится изоморфная орбита, и соответствующие **D**-плотные концы имеют одну и ту же форму, поскольку формой **D**-плотного массива является его J -вес, т. е. вектор строчных сумм, который не изменяется при действии операции L_i, R_i , $i = 1, \dots, n - 1$.

У любой D, U -орбиты есть *эталонная орбита*, которая имеет **D**-плотный конец, являющийся одновременно и **L**-плотным, т. е. являющийся **DL**-плотным массивом. Легко видеть, что **DL**-плотный массив — это диагональный массив с диагональю $a(1, 1) = \lambda_1 \geq a(2, 2) = \lambda_2 \geq \dots$. Эталонные орбиты с разными **DL**-плотными концами не изоморфны. Из этого, в частности, следует, что J -вес **D**-плотных концов изоморфных орбит является одним и тем же.

Можно сказать и как устроена D, U -эталонная орбита $O(\mu)$ формы μ . Она состоит из **L**-плотных массивов формы μ , что то же самое, что полустандартные таблицы Юнга формы μ , заполненные из алфавита J . Аналогично L, R -эталонная орбита $O(\nu)$ формы ν состоит из **D**-плотных массивов формы ν , т. е. из полустандартных таблиц Юнга формы ν , заполненных из алфавита I .

Возьмем произвольный массив a . Действуя на него операциями L_i , $i = 1, \dots, n - 1$, мы можем уплотнить его влево и получить **L**-конец соответствующей L, R -орбиты. Действуя на массив a операциями D_j , $j = 1, \dots, m - 1$, можно перевести его в **D**-конец соответствующей D, U -орбиты.

Таким образом, каждому массиву мы сопоставляем пару полустандартных таблиц Юнга (одинаковой формы, в силу предыдущей теоремы коммутирования). Оказывается, это сопоставление является биекцией.

Теорема. Пространство массивов $A(I, J)$ изоморфно скрещенному произведению $\mathbf{DA} \times_{LDA} \mathbf{LA}$ множества **D**-плотных массивов в $A(I, J)$ и множества **L**-плотных массивов в $A(I, J)$ относительно множества **LD**-плотных массивов.

Эта биекция напоминает знаменитое соответствие RSK (Робинсона—Шенстеда—Кнута), но на самом деле является модифицирован-

ным RSK-соответствием. Традиционно RSK-соответствие определяется алгоритмическим способом, используя так называемый алгоритм игры в пятнашки, jeu-de-taquin, придуманный Ласку и Шютценберже, и согласовано только с одной из двух кристаллических структур. Наш подход избавляется от «мистики» игры в пятнашки, и, в частности, можно показать, что игра в пятнашки является серией кристаллических операции L_i , $i = 1, \dots, n - 1$, в D , U -орбите.

Тензорное произведение орбит. Массивы можно тензорно перемножать. Рассмотрим объекты категории $\mathcal{M}(J)$, т. е. инвариантные подмножества относительно операций U и D . Тогда их можно тензорно перемножить и получится инвариантное множество.

Пусть A — U , D -инвариантное подмножество в $A(I, J)$, и A' — U , D -инвариантное множество в $A(I', J)$. Их тензорное произведение $A \otimes B$ состоит из пар $(a, b) \in A(I \amalg I', J)$ таких, что $a \in A$, $b \in B$ и b располагается правее a . Последнее означает, что мы сдвигаем I' вправо на $n = |I|$, при этом массив b переходит в массив $b' = b'(n + i, j) = b(i, j)$, $i \in I'$, $j \in J$, и (a, b) обозначает конкатенацию массивов a и b' .

Вопрос о разложении на орбиты множества $A \otimes A'$ сводится к случаю, когда инвариантные множества A и A' сами являются орбитами. Более того, мы будем считать эти орбиты эталонными, т. е. что они имеют вид $O(\mu)$ и $O(\nu)$ для разбиений (или биплотных массивов) μ и ν .

Как и любое инвариантное множество, $O(\mu) \otimes O(\nu)$ является объединением орбит вида $O(a, b)$, где $a \in O(\mu)$, $b \in O(\nu)$. Кроме того, пару (a, b) можно считать \mathbf{D} -плотным массивом. Отсюда видно, что a \mathbf{D} -плотный, т. е. фактически совпадает с μ . Что касается b , то условие $b \in O(\nu)$ можно переписать как требование, что b является \mathbf{L} -плотным массивом формы ν . Такую пару (a, b) назовем *стандартной парой* типа (μ, ν) ; это пара массивов, удовлетворяющая трем требованиям:

- 1) a — \mathbf{L} -плотный массив столбцового I -веса μ ,
- 2) b — \mathbf{L} -плотный массив столбцового I' -веса ν , и
- 3) составной массив (a, b) \mathbf{D} -плотный.

Из 3) следует, что форма орбиты $O(a, b)$ равна строчному весу (a, b) .

Зафиксируем итог: *орбиты $O(\mu) \otimes O(\nu)$ находятся в биективном соответствии со стандартными парами типа (μ, ν) .*

Обозначим через $\text{SP}(\mu, \nu; \lambda)$ множество стандартных пар (a, b) типа (μ, ν) и J -веса λ . Тогда *множество орбит $O(\mu) \otimes O(\nu)$, изоморфных $O(\lambda)$, отождествляется с $\text{SP}(\mu, \nu; \lambda)$.*

Множество $\text{SP}(\mu, \nu; \lambda)$ находится в биекции с косыми таблицами Литтлвуда—Ричардсона формы $\lambda \setminus \mu$ и веса ν . Кардинальность последнего

множества определяет кратность вхождения неприводимого представления V_λ группы $GL(m)$ в тензорное произведение $V_\mu \otimes V_\nu$.

Скажем, как определить таблицу Литтлвуда—Ричардсона по стандартной паре (a, b) . Для этого надо взять полустандартную таблицу, отвечающую этой паре, удалить таблицу, отвечающую a , и в оставшейся косоугольной таблице сдвинуть алфавит $I' + |I|$ в прежний I' . Несложно проверить, что полученная косоугольная таблица будет таблицей Литтлвуда—Ричардсона.

Таким образом сложный вопрос о кратностях разложений тензорных произведений в теории представлений становится прозрачным на языке массивов.

Важный пример изоморфизма в категории массивов \mathcal{M} дает конструкция, навеянная понятием инволюции Шютценберже. Для этого определим операцию $*$ центральной симметрии массивов в $\mathcal{A}(I, J)$,

$$a \rightarrow *a, \quad *a(i, j) = a(n - i, m - j).$$

Пусть $A \subset \mathcal{A}(I, J)$ — инвариантное множество массивов формата $I \times J$. Рассмотрим множество $*A = \{*a, a \in A\}$. Очевидно, что оно тоже инвариантно. И хотя само отображение $*$ не является морфизмом, мы утверждаем, что существует канонический изоморфизм

$$\text{Sch} = \text{Sch}_I : A \rightarrow *A.$$

Он однозначно определен тем, что переводит орбиту массива $a \in A$ в орбиту массива $*a$. Форма массива a совпадает с формой массива $*a$, и, следовательно, орбиты $O = O(a)$ и $*O = O(*a)$ канонически изоморфны.

Если d — \mathbf{D} -конец орбиты O , то \mathbf{D} -конец орбиты $*O$ равен $\mathbf{D}(*d)$. Покажем, как это выглядит на конкретном примере. Пусть \mathbf{D} -плотный массив d имеет вид

			1
	1	1	2
4		1	

Перевернутый массив $*d$ имеет вид

	1		4
2	1	1	
1			

Делая уплотнение вниз, мы получаем $Sch(d)$:

			1
	2		2
3		1	1

Обратим внимание, что исходный I -вес $(4, 1, 2, 3)$ поменялся на противоположный $(3, 2, 1, 4)$.

Коммутативность. Конечно, $A \otimes A'$ не равно $A' \otimes A$. Интуитивно ясно, что они должны быть тесно связаны. Однако естественная, казалось бы, биекция, которая отправляет (a, a') в (a', a) , не является морфизмом, т. е. не согласована с действиями D и U . Тем не менее, множества $A \otimes A'$ и $A' \otimes A$ *изоморфны*.

Для этого воспользуемся введенным выше изоморфизмом Шютценберже $Sch: A \otimes A' \rightarrow *(A' \otimes A)$. Так как для любых a, a' верно равенство $*(a, a') = (*a', *a)$, мы имеем равенство $*(A' \otimes A) = *A' \otimes *A$. Тем самым имеется канонический изоморфизм $Sch: A \otimes A' \xrightarrow{\sim} *A' \otimes *A$. Теперь композиция изоморфизмов

$$A \otimes A' \xrightarrow{Sch \otimes 1} *A \otimes A' \xrightarrow{Sch} *A' \otimes A \xrightarrow{Sch \otimes 1} A' \otimes A$$

дает изоморфизм $Com: A \otimes A' \rightarrow A' \otimes A$. Назовем его *компьютером*.

В связи с этим естественно задать вопрос: *будут ли изоморфизмы Sch и/или Com удовлетворять уравнению Янга—Бакстера?* Иначе говоря, будет ли коммутативна диаграмма

$$\begin{array}{ccccccc}
 A \otimes B \otimes C & \xrightarrow{Sch_{12} \otimes 1} & *B \otimes *A \otimes C & \xrightarrow{1 \otimes Sch_{23}} & *B \otimes *C \otimes A & \xrightarrow{Sch_{12} \otimes 1} & C \otimes B \otimes A \\
 \parallel & & & & & & \parallel \\
 A \otimes B \otimes C & \xrightarrow{1 \otimes Sch_{23}} & A \otimes *C \otimes *B & \xrightarrow{Sch_{12} \otimes 1} & C \otimes *A \otimes *B & \xrightarrow{1 \otimes Sch_{23}} & C \otimes B \otimes A
 \end{array}$$

Можно показать, что это так, если алфавиты I, I' и I'' одноэлементные.

Биекция ассоциативности. Аналогично можно рассмотреть разложение на орбиты тройного произведения $O(\mu) \otimes O(\nu) \otimes O(\lambda)$. Они соответствуют стандартным тройкам (a, b, c) , т. е. тройкам массивов, которые удовлетворяют условиям:

- а) a, b, c **L**-плотные массивы формы (или горизонтального веса) μ, ν и λ соответственно;
- б) массив (a, b, c) **D**-плотный.

Множество таких троек обозначим $ST(\mu, \nu, \lambda)$.

По каждой стандартной тройке (a, b, c) можно построить две пары стандартных пар, соответствующие двум способам расстановки скобок в произведении $O(\mu) \otimes O(\nu) \otimes O(\lambda)$.

Сначала мы расставим скобки так: $(O(\mu) \otimes O(\nu)) \otimes O(\lambda)$. Это означает, что мы сначала разлагаем множество $O(\mu) \otimes O(\nu)$ на орбиты $O(a, b)$, а затем каждое произведение $O(a, b) \otimes O(\lambda)$ на орбиты. Это дает нам две стандартные пары (a, b) и $(\mathbf{L}(a, b), c)$. Ясно, что мы получаем биекцию между $\text{ST}(\mu, \nu, \lambda)$ и $\coprod_{\sigma} \text{SP}(\mu, \nu; \sigma) \times \text{SP}(\sigma, \lambda)$.

Расставим теперь скобки иначе: $O(\mu) \otimes (O(\nu) \otimes O(\lambda))$. Это дает две другие стандартные пары: $(a, \mathbf{L}(b, c))$ и (b', c') , где $(b', c') = \mathbf{D}(b, c)$. Стандартность этих пар очевидна; отметим только, что вторая пара имеет тип (ν, λ) , так как операция \mathbf{D} не меняет $I' \amalg I''$ -вес. Как и в первом случае, эти две пары связаны соотношением: формы массивов $\mathbf{L}(b, c)$ и $(b', c') = \mathbf{D}(b, c)$ совпадают. Тем самым мы построили отображение

$$\text{ST}(\mu, \nu, \lambda) \rightarrow \coprod_{\tau} \text{SP}(\mu, \tau) \times \text{SP}(\nu, \lambda; \tau).$$

Тут уже чуть менее очевидно, что это биекция. Но это действительно так. В самом деле, пусть мы имеем элемент правого множества, т. е. две стандартные пары (a, d) и (b', c') , причем форма d совпадает с формой (b', c') . Массив d \mathbf{L} -плотный, а (b', c') — \mathbf{D} -плотный. Тогда теорема о биекции (модифицирование RSK) утверждает, что существует в точности один массив (b, c) , такой что $\mathbf{D}(b, c) = (b', c')$ и $\mathbf{L}(b, c) = d$.

Собирая все вместе, мы получаем биекцию ассоциативности

$$\coprod_{\sigma} \text{SP}(\mu, \nu; \sigma) \times \text{SP}(\sigma, \lambda) \xrightarrow{\sim} \coprod_{\tau} \text{SP}(\mu, \tau) \times \text{SP}(\nu, \lambda; \tau).$$

Однако стоит предупредить, что биекция ассоциативности не линейна, а только кусочно линейна.

Ассоциатор Дринфельда. Рассмотрим тензорное произведение трех объектов категории массивов (орбит) $A \otimes B \otimes C$. Расставим сначала скобки так: $(A \otimes B) \otimes C$ и, используя компьютер, получим $C \otimes (A \otimes B)$, затем применим компьютер в скобках $\text{Com}: A \otimes B \rightarrow B \otimes A$, получим $C \otimes (B \otimes A)$. Расставим теперь скобки по-другому: $A \otimes (B \otimes C)$. Применив компьютер получаем $(B \otimes C) \otimes A$ и, применив еще раз компьютер в скобках, получим окончательно $(C \otimes B) \otimes A$. Несложно показать, что полученные результаты переводятся один в другой сменой скобок, т. е. по-

лучаем, что категория массивов имеет ассоциатор Дринфельда

$$\begin{array}{ccccccc}
 & & C \otimes (A \otimes B) & \longrightarrow & C \otimes (B \otimes A) & & \\
 & \nearrow & & & & \searrow \simeq & \\
 (A \otimes B) \otimes C & \longrightarrow & A \otimes (B \otimes C) & \longrightarrow & (B \otimes C) \otimes A & \longrightarrow & (C \otimes B) \otimes A
 \end{array}$$

С массивами можно связать некоторый класс функций и получить функциональные варианты приведенных выше конструкций.

Возьмем массив $a = (a(i, j))$ и свяжем с ним распределение масс в вершинах целочисленной решетки \mathbb{Z}^2 плоскости \mathbb{R}^2 , поместив в точку (i, j) массу $a(i, j)$. Рассмотрим функцию распределения — интеграл от массива:

$$\left(\iint a \right) (i, j) = \sum_{i' \leq i, j' \leq j} a(i', j')$$

Мы получим функцию на положительном октанте, с нулевым граничным условием.

Функция распределения масс F является супермодулярной, т. е. имеют неотрицательную вторую смешанную производную или, эквивалентно,

$$(i) \quad F(i, j) + F(i + 1, j + 1) \geq F(i, j + 1) + F(i + 1, j)$$

выполняется для любых $i, j \in \mathbb{Z}_+$.

Иными словами массивы — то же самое, что супермодулярные функции, а целочисленные массивы — то же самое что целочисленные супермодулярные функции.

Можно сказать, какие функции соответствуют **D**-плотным массивам. Это супермодулярные функции, которые удовлетворяют неравенствам

$$(ii) \quad F(i, j) + F(i + 1, j) \geq F(i, j - 1) + F(i + 1, j + 1),$$

для любых $i, j \geq 1$. Действительно, что такое **D**-плотность массива? Это условие на каждые две соседние строки массива. В терминах паросочетаний это условие отсутствия неженатых мужчин. В j -й строке сумма $a(1, j) + \dots + a(i - 1, j)$ равна числу невест с зарплатой не больше $i - 1$. Это число должно быть не меньше числа женихов в строке $j + 1$ с зарплатой не большей i , т. е. для любого $i \in I$ выполняется

$$a(1, j) + \dots + a(i - 1, j) \geq a(1, j + 1) + \dots + a(i, j + 1).$$

Для функции $F := \iint a$, последнее условие и есть в точности (ii).

Рассмотрим последовательность приращений у функции, удовлетворяющей (i) и (ii), вдоль прямых $x = 1, 2, \dots, n - 1$. Такая последовательность удовлетворяет условию перемежевания схем Гельфанда—Цейтлина.

То есть схема Гельфанда—Цейтлина — последовательность вертикальных приращений функции распределения масс **D**-плотного массива.

L-плотные массивы находятся в биекции с супермодулярными функциями, которые удовлетворяют для любых $i, j \geq 1$ неравенствам

$$(iii) \quad F(i, j) + F(i, j + 1) \geq F(i - 1, j) + F(i + 1, j + 1).$$

Функции на целочисленном ортанте, удовлетворяющие всем трем условиям (i), (ii) и (iii) называются *дискретно-выпуклыми*.

Чтобы лучше увидеть дискретную выпуклость, разрежем \mathbb{R}^2 тремя сериями прямых $x = a$, $y = b$, $x - y = c$ с целыми $a, b, c \in \mathbb{Z}$. Тогда плоскость разобьется на треугольники с вершинами в целочисленной решетке. Любую функцию, заданную на целочисленной решетке в положительном ортанте, можно продолжить по аффинности на каждый треугольник и получить функцию на всем ортанте. Так вот, дискретно-выпуклые функции продолжаютя до выпуклых функций на ортанте; функции, соответствующие **D**-плотным массивам, продолжаютя до выпуклых в каждой целой вертикальной полосе ширины 1, а функции для **L**-плотных массивов — до выпуклых в каждой целой горизонтальной полосе ширины 1.

Например, при $I = I'$, имеется биекция между стандартными парами в $SP(\mu, \nu; \lambda)$ и дискретно-выпуклыми функциями на прямоугольнике $I \times J$ с граничными приращениями μ, ν, λ и 0 вдоль сторон при обходе границы по часовой стрелке.

Таким образом, число Литтлвуда—Ричардсона $c_{\mu, \nu}^{\lambda}$, равное кратности орбиты $O(\lambda)$ в тензорном произведении $O(\mu) \otimes O(\nu)$, совпадает с количеством целозначных дискретно вогнутых функций на прямоугольнике с граничными приращениями μ, ν, λ и 0. Отметим, что такие функции являются целыми точками в многограннике всех дискретно-вогнутых функций с заданными приращениями. То, что это многогранник, следует из того, что класс дискретно-вогнутых функций на двумерной решетке замкнут относительно сложения.

Вопрос о том, является ли $c_{\mu, \nu}^{\lambda}$ ненулевым становится задачей из полиэдральной комбинаторики, о разрешимости системы от $3n^2$ переменных (в размерности n), и поэтому решается за полиномиальное времени. Отметим, что работающего такого алгоритма пока нет.

Отметим, что дискретно-выпуклые функции на двумерной решетке дуальны к так называемым ульям (hives), которые ввели Кнутсон—Тау при решении гипотезы о насыщении (Saturation conjecture), что очень близко к треугольникам Бернштейна—Зелевинского. В заключение я расскажу о биекции между целочисленными дискретно-выпуклыми функциями и картинками Зелевинского.

Дискретно-выпуклые функции и картины Зелевинского. Чтобы определить картины Зелевинского (*Zelevinsky A. V. A generalization of the Littlewood—Richardson rule and the Robinson—Schensted—Knuth correspondence // J. Algebra. 1981. V. 69. P. 82—94*) нам понадобятся следующие два порядка на $\mathbb{N} \times \mathbb{N}$ (я рассматриваю все в декартовых координатах, а Зелевинский рассматривал матричные координаты). Для (k, l) и (p, q) определим

- 1) $(k, l) \leq_P (p, q)$, если $k \leq p$ и $l \leq q$;
- 2) $(k, l) \leq_J (p, q)$, если $k < p$ или $k = p$ и $l \geq q$.

Рассмотрим два подмножества $X, Y \subset \mathbb{N} \times \mathbb{N}$. Тогда

1. Отображение $f: X \rightarrow Y$ называется *PJ-стандартным*, если для любых пар, таких что $(k, l) \leq_P (p, q)$, $(k, l), (p, q) \in X$, выполняется $f(k, l) \leq_J f(p, q)$.

2. Отображение $f: X \rightarrow Y$ называется *картиной*, если оно является биекцией и оба отображения f и f^{-1} являются PJ-стандартными.

Для пары косых диаграмм Юнга *) κ и $\kappa' \subset \mathbb{N} \times \mathbb{N}$, обозначим

$$\mathbf{P}(\kappa, \kappa') := \{f: \kappa \rightarrow \kappa' \mid f \text{ является картиной}\}.$$

Нам удобнее транспонировать картины, поэтому определим транспонированный $\kappa \leq_J$ порядок следующим образом

$$(k, l) \leq_J^t (p, q), \quad \text{если } l < q \text{ или } l = q \text{ и } k \geq p.$$

Назовем отображение $g: X \rightarrow Y$ *tPJ-стандартным*, если для любой пары, такой, что $(k, l) \leq_P (p, q)$, $(k, l), (p, q) \in X$, выполняется $g(k, l) \leq_J^t g(p, q)$.

Отображение $g: X \rightarrow Y$ называется *транспонированной картиной*, если это биекция и оба отображения g и g^{-1} являются tPJ-стандартными.

Для пары косых диаграмм Юнга $\kappa, \kappa' \subset \mathbb{N} \times \mathbb{N}$, обозначим

$$\mathbf{Pt}(\kappa, \kappa') := \{g: \kappa \rightarrow \kappa' \mid g \text{ — транспонированная картина}\}.$$

Пусть косые диаграммы Юнга $\kappa = \lambda \setminus \mu$ и $\kappa' = \lambda' \setminus \mu'$ содержат по n строк. Предположим, что множество транспонированных картин не пусто, и возьмем какую-нибудь транспонированную картину $g \in \mathbf{Pt}(\kappa, \kappa')$, и определим функцию на квадрате $\square_n := I \times I \subset \mathbb{Z}^2$, $a_g: \square_n \rightarrow \mathbb{Z}_+$ по

*) Косая диаграмма Юнга получается из диаграммы Юнга удалением некоторой ее поддиаграммы Юнга. Напомним, что диаграммы Юнга мы рассматриваем в декартовых координатах, иногда говорят о французском стиле рисования диаграмм Юнга, и это набор единичных ящиков в положительном ортанте, которые устойчивы, т. е. под каждым ящиком есть ящик или стенка октаэдра, на котором он стоит, и слева от него тоже есть соседний ящик или стенка ортанта.

следующему правилу:

$a_g(i, j)$ равняется числу ящиков в j -й строке диаграммы κ , которые g отправляет в i -ю строчку κ' .

Можно показать, что двойной интеграл функции a_g , $\iint a_g$, является дискретно-выпуклой функцией на квадратном гриде \square_n с граничными приращениями $\mu', \lambda, \lambda', \mu$ вдоль сторон квадрата.

Верно и обратное, по любой целозначной дискретно-выпуклой функции на квадратном гриде \square_n с граничными приращениями $\mu', \lambda, \lambda', \mu$, можно построить транспонированную картину из $\mathbf{Pt}(\kappa, \kappa')$.

Сначала, по дискретно-выпуклой функции F на гриде \square_n с граничными приращениями $\mu', \lambda, \lambda', \mu$, мы построим пару полустандартных косых таблиц Юнга формы $\kappa = \lambda \setminus \mu$ и $\kappa' = \lambda' \setminus \mu'$.

Для этого рассмотрим две последовательности — вертикальных и горизонтальных приращений F . Последовательность вертикальных приращений

$$F_1^v, \dots, F_n^v,$$

где F_i^v обозначает приращения F вдоль вертикальной прямой $x = i$, $F_i^v(1) = F(i, 1)$, \dots , $F_i^v(j) = F(i, j) - F(i, j - 1)$, \dots , $F_i^v(n) = F(i, n) - F(i, n - 1)$, является перемежающей, в силу условий (i) и (ii) дискретной выпуклости.

Аналогично, в силу условий (i) и (iii), последовательность горизонтальных приращений

$$F_1^h, \dots, F_n^h,$$

$F_j^h(1) = F(1, j)$, \dots , $F_j^h(i) = F(i, j) - F(i - 1, j)$, \dots , $F_j^h(n) = F(n, j) - F(n, j - 1)$, является тоже перемежающей.

Отметим, что условия на граничные приращения дают следующие равенства: $F_1^v = \mu$, $F_n^h = \lambda'$, $F_n^v = \lambda$, $F_1^h = \mu'$.

Свойство перемежевания позволяет определить две полустандартные косые таблицы Юнга T_F^v и T_F^h формы κ и κ' , соответственно. Первая получается объединением косых диаграмм $F_j^v \setminus F_{j-1}^v$ заполненных камешками только с цифрой j , $j = 1, \dots, n$, вторая получается объединением косых диаграмм $F_j^h \setminus F_{j-1}^h$, заполненных камешками с цифрой j , $j = 1, \dots, n$.

Вторая смешанная производная

$$\partial\partial F(i, j) := F(i, j) + F(i - 1, j - 1) - F(i, j - 1) - F(i - 1, j),$$

с одной стороны, равняется числу камешков (ящиков) в j -й строке таблицы T_F^v , помеченных буквой i , а, с другой стороны, равняется числу камешков в i -й строке таблицы T_F^h , на которых написана буква j .

Это позволяет определить отображение косых диаграмм $f_F: \kappa \rightarrow \kappa'$ по следующему правилу:

для каждого $(i, j) \in \square_n$ отображение f_F посылает самый правый ящик j -й строки диаграммы κ , из тех ящиков, в которых располагаются камешки с буквой i (согласно заполнению T_F^v), в самый левый ящик i -й строки диаграммы κ' , из тех в которых располагаются камешки с буквой j (согласно заполнению T_F^h).

Согласно предыдущему свойству второй смешанной производной, когда мы пройдем по всем $(i, j) \in \square_n$, получится, что f_F осуществит биекцию между ящиками косых диаграмм Юнга κ и κ' . Более того, f_F является транспонированной картиной (не очень сложная проверка). Итак, резюмируем, что мы получили, в следующей теореме.

Теорема. *а) Пусть $F: \square_n \rightarrow \mathbb{Z}_+$ является целочисленной дискретно-выпуклой функцией с граничными приращениями $F_1^v = \mu$, $F_n^h = \lambda'$, $F_n^v = \lambda$, $F_1^h = \mu'$. Тогда отображение $f_F: \kappa \rightarrow \kappa'$ является транспонированной картиной.*

б) Пусть $g \in \mathbf{Pt}(\kappa, \kappa')$, тогда функция

$$F_g := \iint a_g: \square_n \rightarrow \mathbb{Z}_+, \quad F_g(i, j) := \sum_{i' \leq i, j' \leq j} a_g(i', j'),$$

является целочисленной дискретно-выпуклой функцией на гриде \square_n с граничными приращениями μ' , λ , λ' , μ вдоль сторон грида.

Отметим, что эта биекция между картинами и дискретно-вогнутыми функциями позволяет за полиномиальное время от n решить вопрос о непустоте множества $\mathbf{Pt}(\kappa, \kappa')$.

Из этой теоремы и теоремы о биекции (модифицированное RSK) несложно получить следующую биекцию:

$$\mathbf{Pt}(\kappa, \kappa') \cong \coprod_{\nu} (\mathbf{Pt}(\nu, \kappa) \times \mathbf{Pt}(\nu, \kappa')),$$

где ν пробегает обычные (не косые) диаграммы Юнга. Так полученную биекцию можно назвать модифицированной биекцией Зелевинского из его статьи в журнале «J. Algebra» (1981. V. 69. P. 82—94).

13 января 2005 г.

О. Я. В и р о

ОЦЕНКИ СНИЗУ ДЛЯ ЧИСЛА РЕШЕНИЙ ИСЧИСЛИТЕЛЬНЫХ ПРОБЛЕМ В ТОПОЛОГИИ И ВЕЩЕСТВЕННОЙ АЛГЕБРАИЧЕСКОЙ ГЕОМЕТРИИ

1.

Введение

В моем докладе речь пойдет только о вещественной исчислительной алгебраической геометрии, хотя для начала нужно, естественно, сказать кое-что и про комплексную. Эти исследования происходят прямо сейчас; всё не оконченное, результатов будет немного, скорее будут постановки задач.

1.1. Задачи исчислительной геометрии. Позвольте сначала напомнить, что такое исчислительная геометрия, какие задачи так называются. Типичная задача выглядит так: на плоскости, в пространстве или в каком-нибудь еще алгебраическом многообразии фиксируется какое-то количество точек, какое-то количество кривых и спрашивается, сколько кривых какого-то там деформационного типа проходит через эти точки, пересекается с этими кривыми, со всеми кривыми. Конечно, задач тут много, и не обязательно задачи укладываются в эту схему. Например, можно потребовать, чтобы кривые не только пересекались с данными кривыми, но и касались их или касались их очень хорошо, с определенным порядком, или чтобы они проходили через данные точки в данном направлении или чтобы, вообще, это были не кривые, а многообразия большего числа измерений. Но во всех этих задачах, конечно, условия должны быть подобраны так, чтобы кривых было конечное число, чтоб было что считать; иначе придется вместо того чтобы считать кривые, брать какие-то численные характеристики многообразия всех решений. Это, конечно, имеет смысл, но уже выходит за рамки этого предмета.

1.2. Уход исчислителей в комплексную область. Примерами таких задач могут служить простые задачи, хорошо известные из школы. Например: дана окружность, дана точка, сколько прямых проходят через эту точку и касаются этой окружности? И здесь мы сталкиваемся с явлением, которое, собственно, определяет предмет моего доклада. Дело в том, что

в зависимости от того, где находится точка, решений либо два, либо ни одного. Это вынудило специалистов по этому предмету давным-давно, еще в XIX веке, перейти в комплексную алгебраическую геометрию, где количество решений меняется не так решительно и на открытом всюду плотном (или на открытом по Зарискому) множестве это количество постоянно. В частности, через точку, расположенную внутри окружности, тоже проходят 2 прямые, касающиеся окружности, только они комплексные, и касаются они комплексной окружности.

1.3. Конники, касающиеся пяти данных коник. Вот еще пример такой задачи. Пусть на плоскости даны 5 коник. Сколько существует коник, которые касаются всех этих пяти коник? Эту задачу предложил Штайнер [8], он и решил ее, но решил неправильно. Найденный Штайнером ответ $6^5 = 7776$ был не верен. Всё это происходило в XIX веке. Потом Жонкьер и Шаль [2] решили ее правильно и получили 3264 коники. Так что задача нетривиальная, решение нетривиальное. Но, между прочим, это решение относится к области комплексной. Сколько их в вещественной области, сказать трудно, потому что это опять же зависит от конфигурации коник. В 1997 году была опубликована работа трех известных математиков Ронги, Тоньоли и Вуста [7]. В ней было доказано, что все 3264 могут оказаться вещественными при определенном расположении исходных пяти коник.

А может ли их быть меньше, может ли их, например, быть 0, я не знаю. *) Может быть, это простой вопрос, я не думал над ним уж очень много, но во всяком случае ясно, что комплексные исчислительные задачи решать проще, а в соответствующих вещественных задачах решение комплексной задачи играет роль оценки сверху. Если комплексных решений немного, то и вещественных решений не больше.

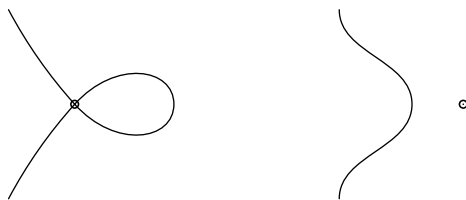
2.

Считаем вещественные рациональные кривые

2.1. Вещественные плоские рациональные кубики через 8 точек.

Я хочу разобрать более детально еще один классический пример: имеется 8 точек на плоскости, сколько вещественных рациональных кривых степени 3 проходит через эти 8 точек? Вещественные рациональные кривые степени 3 общего положения бывают двух видов (рис. 1). Рациональная кривая степени 3 имеет одну двойную точку. Двойная точка может быть

*) Примечание при корректуре. Вельшинжер доказал, что если исходные 5 коник ограничивают в проективной плоскости попарно не пересекающиеся области, то это число не меньше 32. Полный ответ по-прежнему не известен.



Р и с. 1. Рациональные кривые степени 3

либо изолированной (в вещественной области) и тогда через нее проходят 2 комплексно сопряженные ветви, либо через двойную точку проходят 2 вещественные ветви. Вообще-то через 8 точек проходит бесконечно много кубических кривых, но большая их часть — эллиптические кривые. Они составляют пучок, т. е. проективную прямую, и нумеруются одним параметром. В таком пучке над комплексными числами присутствуют 12 рациональных кривых, ровно 12. Это легко доказать, я сейчас не буду этим заниматься, это несложное упражнение. Впрочем, аргументов, которые я приведу сейчас для вещественного случая, хватит и в комплексном случае.

Посмотрим на эту задачу с вещественной точки зрения. Какое количество решений здесь может быть? Я сначала скажу ответ. Ответ такой: может быть либо 8, либо 10, либо 12 вещественных рациональных кубических кривых, проходящих через данные 8 вещественных точек в общем положении.

Вы спросите: а как насчет девяти? 9 тоже может, но тогда пучок будет вырожден, и в нём будет присутствовать кривая с точкой возврата (каспом). Кроме кривых, показанных на рис. 1, бывает ещё полукубическая парабола. Она нужна для того чтобы заполнить промежутки: в кривую с каспом сливаются две кривые с обыкновенными двойными точками перед тем как исчезнуть. А меньше 8 не может быть.

Почему это так? К этому ответу легко прийти, сделав вычисление с эйлеровой характеристикой. Рассмотрим пучок всех кубических кривых, проходящих через эти 8 точек, как отображение вещественной проективной плоскости — сначала без этих точек — в проективную кривую. Кроме этих точек придётся удалить ещё одну. На удалённые точки это отображение продолжить нельзя, нужно их раздуть, т. е. произвести σ -процесс в каждой из этих точек. Тогда вещественная проективная плоскость, где всё это живёт, будучи раздута в 9 точках, превратится в связную сумму десяти экземпляров проективной плоскости, и получится её отображение в проективную прямую $\mathbb{R}P^1$, т. е. в окружность: $\#^{10}\mathbb{R}P^2 \rightarrow \mathbb{R}P^1 = S^1$.

Это отображение строится так. Кривые нашего пучка определяются уравнениями вида

$$\alpha F(x_0, x_1, x_2) + \beta G(x_0, x_1, x_2) = 0,$$

где $F(x_0, x_1, x_2) = 0$ и $G(x_0, x_1, x_2) = 0$ — уравнения каких-нибудь двух кубических кривых, проходящих через заданные 8 точек. Коэффициенты α и β можно рассматривать как однородные координаты кривой в пучке. Все кривые пучка проходят через точки пересечения кривых $F(x_0, x_1, x_2) = 0$ и $G(x_0, x_1, x_2) = 0$. Среди этих точек — исходные 8 точек. Однако, по теореме Безу две кривые степени 3 пересекаются в 9 точках, так что помимо данных 8 точек имеется ещё одна. Все кривые нашего пучка проходят через данные 8 точек и эту девятую точку. Через любую другую точку проходит ровно одна кривая пучка. Определим образ такой точки как номер $(\alpha:\beta)$ той единственной кривой пучка, которая проходит через эту точку. Если точка совпадает с одним из узлов пучка, то этого нельзя сделать. Вместо такой точки мы вставили целую проективную прямую — прямую направлений в этой точке. Кривые нашего пучка проходят через эту точку со всевозможными направлениями, так что мы можем направлению поставить в соответствие кривую, и, значит, её номер $(\alpha:\beta)$.

Посмотрим на это отображение с точки зрения эйлеровой характеристики. Эйлерова характеристика поверхности, которая получилась, равна -8 . Мы взяли $\mathbb{R}P^2$, у $\mathbb{R}P^2$ эйлерова характеристика 1. Каждое раздутие вычитает 1: на место точки, у которой эйлерова характеристика 1, мы вставляем окружность, эйлерова характеристика которой ноль. В результате девяти раздутий получается эйлерова характеристика -8 . А у $\mathbb{R}P^1$ эйлерова характеристика 0. Далее, общий слой этого отображения тоже имеет эйлерову характеристику 0, потому что это замкнутая гладкая кривая, её компоненты связности гомеоморфны окружности. Так что -8 может образоваться только за счет особых кривых. Посмотрим, какие особые кривые бывают. Особые кривые — это как раз те, которые нас интересуют. У кривой одного вида эйлерова характеристика -1 — топологически это букет двух окружностей на проективной плоскости. А у кривой другого вида эйлерова характеристика $+1$, потому что это окружность плюс точка. И мы получаем, что -8 должно быть равно числу кривых одного вида минус число кривых другого вида. Интересующее же нас число — это общее число всех этих кривых. Значит, оно никак не может быть меньше 8.

А больше 12 оно не может быть по другой причине. Рассмотрим ту же самую ситуацию в комплексной области. Там эйлерова характеристика будет другой. Аналогичный подсчет там даст, что всегда это равно 12.

2.2. Обобщение: теорема Вельшинжера. Мы здесь столкнулись с таким явлением, что количество вещественных решений нашей исчислительной задачи можно оценить не только сверху, но и снизу. Недавно Жан-Ив Вельшинжер [11], [12] доказал обобщение этой оценки снизу. Его работы опубликованы в arXiv'e в 2003 г.

Имеются две теоремы по этому поводу, о кривых на поверхности и в трёхмерном многообразии. Я буду их формулировать только в простейшей, проективной ситуации. У Вельшинжера эти теоремы доказываются в гораздо бóльшей степени общности.

Первая теорема плоская. Рассматриваются рациональные вещественные кривые степени d , проходящие через $3d - 1$ точку в общем положении. Число $3d - 1$ здесь появляется просто потому, что для этого числа задача имеет конечное число решений: таких кривых конечное число.

Так вот, в общем положении каждая из этих кривых может быть снабжена знаком так, что их количество, подсчитанное с учетом этих знаков, не зависит от расположения точек. Какое бы мы расположили точки ни взяли, всегда будет получаться одно и то же число. Позвольте мне его обозначить через W_d .

Выглядит эта формулировка, тем самым, так: у каждой кривой есть такой вес, что количество этих кривых с весами одно и то же для всех общих конфигураций точек.

Мы в действительности уже рассмотрели частный случай этой теоремы, когда $d = 3$. В этом случае мы выяснили, что $W_d = 8$.

Как определить вес кривой, то есть знак, с которым кривую следует считать? У вещественной рациональной кривой общего положения имеются только двойные точки. Вещественные двойные точки могут быть двух типов, о которых я уже говорил — точка с двумя вещественными ветвями и точка с двумя мнимыми сопряженными ветвями. Будем считать количество точек второго сорта, т. е. точки, через которые проходят две комплексно сопряженные мнимые ветви; и возьмем минус единицу в степени количество таких точек — это и есть тот знак, с которым нужно брать кривую. Так что действительно в случае кубических кривых получается то, что нужно.

Доказательство Вельшинжера состоит в анализе того, что может происходить при деформации начальных данных, при деформации этой конфигурации точек. Действительно, две такие рациональные кривые могут убежать в мнимую область, и при этом у них должна быть разная чётность числа изолированных двойных точек. Вот и всё. Ничего больше пока про это не известно. И все вычисления чисел W_d , которые здесь проводились,

проводились просто предъявлением в конкретной ситуации конкретных примеров.

Двойные точки бывают мнимые; они могут убежать в мнимую область. Они тогда вообще не нужны. Две изолированные двойные точки могут убежать, это на четности их числа не отразится. Это тоже учтено.

2.3. Теорема Вельшинжера о кривых в пространстве. Вторая теорема Вельшинжера относится к пространственным кривым. Формулировка та же самая, с тем только изменением, что нужно проводить эти кривые через $2d$ точек. Каждую кривую тоже можно снабдить весом. Кривые рациональные, степени d . Степени d — это значит, что их комплексификация пересекает плоскость в d точках. И опять-таки есть такое универсальное число для каждой степени, что сумма таких весов всех этих кривых, всех решений этой задачи, равна этому числу.

Весы определяются, конечно, не так, как в плоской ситуации, потому что пространственные кривые все неособы в общем положении, но зато у них имеется специальное оснащение. У каждой такой кривой в 3-мерном пространстве есть специальное оснащение. Оно получается из-за того, что нормальное расслоение такой кривой расщепляется в прямую сумму двух расслоений.

Увидеть это оснащение можно так. Возьмем эту кривую. Она определяется нашими точками жестко. Потянем за одну из точек. Кривая начнет деформироваться. Она будет крутиться вокруг тех точек, которые фиксированы. И поле скоростей определит оснащение. Оно, конечно, в каком-то смысле рвется, но в каком-то смысле и нет. Потому что, когда вы переходите через эту неподвижную точку, поле скоростей изменяет знак; но полоска, составленная из скоростей, остается. Так что кривая оказывается оснащенной. Поле прямых остается.

И дальше мы включаем спинорную структуру проективного пространства. Спинорная структура считает, сколько раз это поле прямых обернется вокруг нашей кривой. Поскольку это поле прямых, и действительно бывает, что повернется полуцелое число раз, нужно проявлять некоторую осторожность. В общем, четность/нечетность этого количества оборотов (по модулю некоторой подправки) это и есть этот самый вес — я не буду сейчас вдаваться в детали.

В действительности и в плоской ситуации можно сформулировать нечто в таком духе. Потому что, когда завиток превращается в изолированную двойную точку, количество оборотов касательного вектора в плоскости меняется ровно на единичку. Так что спинорная структура на самом деле явление общее, она присутствует не только в пространственной, но и в плоской теореме.

Попытки применить те же самые веса, которые работают в рациональной ситуации, в других классах кривых, вели к контрпримерам. Как их можно подправить — не известно. Поскольку теоремы доказываются именно через теорию особенностей, т. е. мы смотрим, что не меняется при деформации, очень важно угадать, что же не должно меняться. Хотя есть всякие соображения по этому поводу, пока никаких теорем нет. По-видимому, эта теорема распространяется и, возможно, уже распространена (я видел работы на эту тему) на рациональные кривые в других пространствах.

3.

Назад в топологию

3.1. Коэффициент зацепления и подсчёт прямых. Но я хочу оставить рациональные кривые и обратить ваше внимание на то, что эти все теоремы очень похожи на теоремы, которые давно известны из топологии. В топологии выбор большой, таких теорем много. Я буду говорить о теоремах, которые сколько-нибудь похожи на только что рассмотренные алгебро-геометрические теоремы.

В теоремах Вельшинжера речь шла только о проведении кривых через точки; кривые, с которыми должны были бы пересекаться подсчитываемые кривые, пока не появлялись. Мне хочется двигаться в этом направлении.

Рассмотрим такой пример. Возьмем в 3-мерном пространстве две кривые (ориентированные, хотя их ориентация понадобится не сразу) и возьмем ещё точку. И зададим вопрос: сколько есть прямых, которые проходят через эту точку и пересекают обе этих кривых?

Конечно, оценки сверху здесь никакой нет. Можно так неудачно пододвинуть эту точку или подвинуть наше зацепление, что таких прямых будет бесконечно много.

Меня здесь интересует оценка снизу. Если эти окружности не зацеплены, мы можем их как-то расположить свободно в пространстве, совсем в стороне от этой точки. Тогда таких прямых и вовсе не окажется. Однако если кривые зацеплены, то обязательно такие прямые есть. И больше того, в общем положении их количество оценивается числом: $|2 \text{lk}(C_1, C_2)|$ (абсолютной величиной удвоенного коэффициента зацепления этих двух кривых). Причина совершенно ясна: дело в том, что удвоенный коэффициент зацепления можно подсчитать следующим образом: мы встаем в фиксированную точку, проектируем из нее наше зацепление на плоскость, тогда каждая из наших прямых (решение нашей задачи) даёт точку скрещивания; в точке скрещивания можно по известному правилу поста-

вить знак (плюс или минус единица), и если вы теперь просуммируете все эти ± 1 , то у вас получится удвоенный коэффициент зацепления.

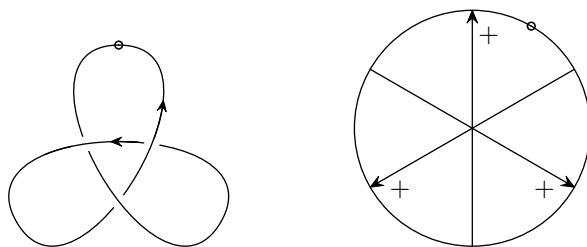
Более того, если вы просуммируете эти числа только по тем точкам, где первая кривая встречается со второй, то у вас получится просто коэффициент зацепления. Так что есть оценка и на следующую исчислительную задачу: сколько существует прямых, которые проходят через точку и пересекают эти две кривые в заданном порядке (сначала первая кривая, потом вторая)? Количество таких прямых оценивается просто абсолютной величиной коэффициента зацепления.

3.2. Исчислительный смысл формул Поляка—Виро. Вот еще несколько результатов того же сорта. Известно, что все инварианты Васильева можно вычислять при помощи комбинаторных формул, см. [6] и [3]. Каждое такое вычисление представляет собой подсчет по диаграмме некоторого количества решений исчислительной задачи такого же сорта. Только вместо отдельных прямых, проходящих через заданную точку, мы подсчитываем определенные конфигурации прямых. Я сформулирую это лишь в простейшем случае, чтобы не вводить всей этой сложной терминологии.

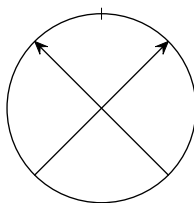
Самый простой инвариант Васильева для узлов — это так называемый инвариант Кассона. Это инвариант степени 2. Он может быть характеризован тем, что для трилистника он равен 1, для тривиального узла он равен 0, а для восьмерки он равен -1 . Этого уже вполне достаточно, может быть, даже более чем достаточно. Это числовой инвариант, и его можно вычислять следующим образом. Нарисуем диаграмму узла, нарисуем гауссову диаграмму этого узла, выберем какую-нибудь точку на этом узле.

Что такое гауссова диаграмма? Я просто рисую окружность, параметризующую узел, и для каждой пары точек, которые отображаются при проектировании на плоскость в одну, я рисую хорду, соединяющую соответствующие две точки на параметризующей окружности, соединяю их стрелочкой — от верхней ветви к нижней; снабжаю эту стрелочку знаком — в каждой точке я вычисляю тот же самый знак, который я вычислял для коэффициентов зацепления (рис. 2). На рис. 2 все эти знаки будут плюсы. А теперь нужно выбрать на узле точку, соответствующую выбранной точке на гауссовой окружности, и найти все подконфигурации конфигурации этих стрелок, которые выглядят так, как показано на рис. 3. В данном случае у нас есть всего одна такая пара стрелок. Нужно перемножить знаки стрелок у каждой такой пары и сложить по всем таким парам. То, что получится, это инвариант.

Такие формулы есть для любого инварианта Васильева, см. [3]. А что же эта формула означает? Что мы в действительности решаем следующую



Р и с. 2. Узел и его гауссова диаграмма



Р и с. 3. Конфигурация стрелок

задачу исчислительной геометрии: мы берем точку и через нее проводим пары прямых, каждая из которых проходит через эту точку и пересекает наш узел дважды. Вот две хорды; нужно, чтобы эта пара хорд пересекала узел в соответствии с этой схемой. У каждой из этих хорд есть знак, мы перемножаем знаки и складываем. Тем самым, количество пар хорд никак не может быть меньше, чем инвариант Кассона.

Как видите, здесь чуть более деликатная формулировка: мы рассматриваем не просто прямые, а какие-то конфигурации прямых, прямые должны располагаться определённым образом. Это связано, конечно, с тем, что инварианты Васильева — более деликатный предмет, чем коэффициент зацепления. Коэффициент зацепления — это самый простой инвариант Васильева.

Сегодня, собственно, ничего кроме примеров таких формулировок больше и не будет.

3.3. Прямые, пересекающие четверет кривых. Формулировка теоремы про две окружности и точку допускает многомерные обобщения. И следующая теорема которую мы рассмотрим — тоже многомерная, но я её сформулирую сначала в трёхмерном пространстве. Это теорема Юлии Виро (Дроботухиной) [9].

Пусть имеются 4 кривые, скажем, в проективном пространстве:

$$C_1, C_2, C_3, C_4 \subset \mathbb{R}P^3.$$

Эти кривые замкнуты, ориентированы, дизъюнкты. Каждая из них не обязательно связна — может состоять из нескольких окружностей. Кривые в общем положении, не обязательно алгебраические, не обязательно гладкие — какие угодно топологические кривые. Тогда (если кривые действительно в общем положении) существует лишь конечное число прямых, которые пересекают все эти 4 кривые. Такие прямые называются квадрисекантами. Утверждается, что каждую такую кривую можно снабдить знаком так, чтобы сумма всех этих знаков была равна следующему выражению:

$$\sum \omega(l) = 2 \text{lk}(C_1, C_2) \text{lk}(C_3, C_4) - \text{lk}(C_2, C_3) \text{lk}(C_4, C_1)$$

(разность произведений коэффициентов зацепления). Здесь $\omega(l)$ — это те знаки, которые связываются с каждой прямой.

Есть некая конструкция, которая даёт знак каждой прямой, пересекающей эти кривые в данном циклическом порядке. У нас ситуация проективная, так что каждая прямая — это окружность, и на ней точки пересечения с кривыми C_i расположены в определенном циклическом порядке. Мы требуем, чтобы циклический порядок был именно такой: C_1, C_2, C_3, C_4 . Мы считаем те прямые, которые пересекают все эти кривые, и получается такое выражение. Тем самым, количество этих прямых никак не меньше, чем абсолютная величина этого выражения.

Можно обсуждать вопрос, насколько точна эта теорема. Она точна для некоторых зацеплений, а для некоторых, конечно, нет. Оценки сверху здесь никакой нет. Между прочим, если бы кривые были алгебраическими, то была бы оценка сверху, и эта оценка сверху была бы удвоенным произведением степеней всех этих кривых.

Эта теорема обобщается на случай, когда в n -мерном проективном пространстве имеются многообразия размерностей p, q, p, q , причем $p + q = n - 1$, $C_1, C_2, C_3, C_4 \subset \mathbb{R}P^n$. Эти ограничения на размерности нужны для того, чтобы были определены коэффициенты зацепления.

Для кривых в R^3 или подмногообразий в R^n имеется аналогичная теорема, только выражение будет попроще, потому что мы можем наложить еще более жесткие условия на прямые. Пусть теперь прямые пересекают их в данном линейном порядке, а не циклическом. Тогда тоже имеются знаки у прямых, и сумма всех этих знаков равна следующему выражению:

$$\sum \omega(l) = \text{lk}(C_1, C_2) \text{lk}(C_3, C_4).$$

Эта теорема была независимо получена Поляком.

Она обобщается на n -мерную ситуацию похожим образом, только теперь меньше ограничений на размерности, потому что меньше коэффициентов зацепления надо определить.

3.4. Перспективы и трудности перехода от прямых к кривым. Конечно, это выглядит как шаг в сторону обобщения теоремы Вельшинжера. Вельшинжер рассматривал рациональные кривые степени d , проходящие через $2d$ точек. В формулировке каждую точку мы можем разменять на две кривых. И в случае, когда у нас степень равна 1, теорема Вельшинжера превращается в такой тривиальный факт, что через любые 2 точки проходит единственная прямая. Тут говорить не о чем; но если вы одну из этих точек разменяете на окружности, то получится та формулировка о коэффициенте зацепления, с которой я начал. А если вы обе точки разменяете на окружности, то получатся 4 кривые, и нужно считать прямые, которые проходят через них всех, и получится теорема Юлии Виро. Можно ли найти общую теорему, которая бы обобщала и эти топологические формулировки, и теорему Вельшинжера? К сожалению, пока я ничего не знаю о такой общей формулировке. И если просто проанализировать доказательство теоремы Юлии Виро, то видны трудности, с которыми мы тут встречаемся при переходе к старшим степеням.

Я должен обратить ваше внимание на то, что в теореме Вельшинжера порядок прохождения рациональной кривой через точки не фиксирован, он произволен. И, по-видимому, это неспроста: такой формулировки, в которой бы порядок был фиксирован, может быть, просто и нет. Во всяком случае, если мы хотим оставить эту теорему в каких-то простых терминах.

Вот почему. Что может случиться с рациональной кривой, с решением этой задачи, вельшинжеровской задачи, если мы двигаем наши точки? Речь идет уже о доказательстве теоремы Вельшинжера. Про одну неприятность я уже рассказывал: две кривые могут слиться и убежать в комплексную область. Но есть еще одна неприятность: кривая может стать приводимой. У нас была рациональная кривая, и на ней может появиться особая точка, как это бывает, скажем, с коникой. У нас была гипербола, мы ее подвигали, и она превратилась в пару прямых. При движении тех точек, через которые мы проводим нашу кривую, кривая может выродиться.

Что же в этом плохого? С точки зрения формулировки теоремы Вельшинжера — ничего. Здесь как раз ничего неприятного для нее не происходит, и вклад кривой в сумму не меняется. Однако меняется порядок, в котором точки расположены на кривой. Дело в том, что после того как мы прошли через это вырождение, через пару прямых, у нас появляется гипербола другого сорта. У нас были какие-то точки, через которые мы эту

гиперболу (или кривую более высокой степени рациональную) проводили. В тот момент, когда кривая выродилась, какие-то из них оказались на одной половинке, а какие-то — на другой. В следующий момент кривая перестроилась, две ветви слились, и наши точки опять оказались на одной кривой. Но дело в том, что эта перестройка поменяла ориентацию части кривой. Сначала у нас кривая проходила через какие-то точки в одном порядке, а потом она будет проходить вторую половинку в обратном порядке. Поэтому порядок фиксировать в теореме Вельшинжера не удастся. Разве что мы научимся следить за тем, что здесь происходит.

Почему с порядком удастся справиться в более элементарной ситуации? Потому что прямая не умеет распадаться. Прямая в трёхмерном пространстве не может распасться ни на что, потому что уже это прямая, её степень равна единице. Конечно, что-то в этом духе может происходить и с другими классами кривых. И тогда возникают соответствующие формулировки.

3.5. Окружности, пересекающие секстет кривых. Ещё один такой замечательный класс кривых — окружности в трёхмерном пространстве. Окружность тоже не умеет распадаться. Правда, она может стянуться в точку и исчезнуть, и в момент стягивания в точку она превращается в пару мнимых пересекающихся прямых. Но за этим легко проследить. А еще окружность может превратиться в прямую плюс бесконечно удаленная прямая. Но в конечной области ничего страшного не происходит: просто окружность распрямилась, а в следующий момент она опять окружность. Если мы следим только за тем, что происходит в конечной области евклидова пространства, то ничего не происходит. А если хотите, то можно посредством стереографической проекции переехать на трёхмерную сферу, а уж там окружность ни на что распасться не может, потому что она является пересечением сферы с плоскостью. Она неприводима и не может быть приведена. Разве что стянется в точку. Поэтому соответствующая теорема есть и для окружностей. Формулировка ее выглядит так.

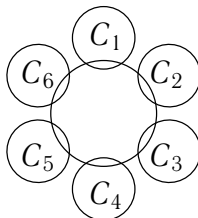
Теперь у нас всё происходит не в проективном пространстве, а там, где живут окружности, т. е. либо в аффинном пространстве, либо в сфере. Зато у нас есть 6 кривых или 6 подмногообразий в \mathbb{R}^n , и мы считаем окружности, проходящие через них. Каждую из этих окружностей можно снабдить знаком так, чтобы общее количество сосчитанных с этими знаками окружностей было равно

$$\text{lk}(C_1, C_2) \text{lk}(C_3, C_4) \text{lk}(C_5, C_6) - \text{lk}(C_2, C_3) \text{lk}(C_4, C_5) \text{lk}(C_6, C_1).$$

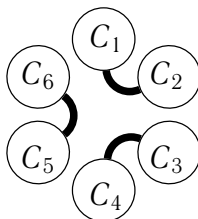
Эта формула очень похожа на предыдущую: произведение коэффициентов зацепления первой кривой со второй, третьей с четвертой, пятой с шестой

минус произведение коэффициентов зацепления второй с третьей, четвертой с пятой и шестой с первой.

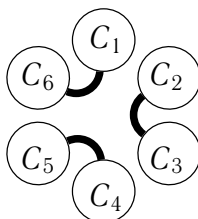
Это естественная формула. Если хотите представить её схематически, то давайте расположим кривые C_i так, как они должны пересекаться с окружностями:



Из произведения следующих трёх коэффициентов зацепления:



вычитаем произведение следующих трёх:



Я не знаю никакой аналогичной теоремы про кривые старших степеней или про общие коники. По причине, о которой я уже говорил, не удаётся различать кривые, пересекающие данные кривые в заданном порядке.

В простейшем случае эта теорема даёт оценку снизу числа окружностей, которые пересекают каждую компоненту 6-компонентного зацепления, составленного из гладких замкнутых ориентированных кривых $C_1, C_2, C_3, C_4, C_5, C_6$. Как и в применениях предыдущей теоремы, связность кривых C_i не существенна. Мы считаем те настоящие окружности, которые пересекают сначала кривые C_i в заданном циклическом порядке. Количество таких окружностей оказывается не меньше, чем абсолютная величина этой разности произведений коэффициентов зацепления.

В 3-мерном пространстве все C_i имеют размерность 1. А в обобщении этой теоремы всё происходит в n -мерном пространстве, C_i являются

попарно дизъюнктными ориентированными гладкими замкнутыми подмногообразиями этого пространства. Они должны иметь размерности p и q , причем размерности должны чередоваться, и сумма $p + q$ должна быть $n - 1$ для того, чтобы были определены коэффициенты зацепления.

3.6. Набросок доказательства. Теперь я хочу сказать несколько слов о доказательстве теорем Юлии Виро, которые мы обсуждали. Обе эти теоремы — это практически одна и та же работа, там просто одно и то же доказательство. В доказательстве второй теоремы используется первая. Я не буду, конечно, вдаваться в детали. Я хочу сказать только про принципиальные моменты.

Сначала напомним формулировку. Формулировка следующая. Скажем, самая первая формулировка, когда имеется 4 ориентированные замкнутые кривые, дизъюнктные, в проективном пространстве $C_1, C_2, C_3, C_4 \subset \mathbb{R}P^3$, и мы снабжаем таким знаком каждую прямую, пересекающую эти кривые в данном циклическом порядке, что в сумме получается число, зависящее только от попарных коэффициентов зацепления этих кривых, и, значит, не меняющееся при изотопиях. Получается всегда такая комбинация коэффициентов зацепления, независимо от того, какие кривые мы брали — удвоенный определитель:

$$2 \begin{vmatrix} \text{lk}(C_1, C_2) & \text{lk}(C_2, C_3) \\ \text{lk}(C_4, C_1) & \text{lk}(C_3, C_4) \end{vmatrix}$$

Число прямых с весами равняется этому числу.

Возьмем первые 3 кривые и рассмотрим все прямые, которые пересекают каждую из них. Они составляют поверхность. Представьте себе, что все C_i — прямые. Тогда, в общем случае, эта поверхность является гиперboloидом. А если берутся не прямые, а кривые, то это какая-то поверхность, у которой имеются разнообразные особенности. Эти особенности можно разрешить некоторым естественным образом, как это обычно делается в алгебраической геометрии.

Как это делается в теории особенностей? Просто эти прямые нужно рассматривать сразу вместе с точками пересечения с кривыми. Каждую прямую — столько раз, сколько на ней троек точек. И нужно рассмотреть пересечение этой поверхности с четвертой кривой. Эта схема работает и в других случаях, но здесь интересна сама по себе эта поверхность. Оказывается, что особенности такой поверхности очень легко рассортировать. Это несложно. Максимум, что там бывает в общем случае — двойные кривые. Ну и самопересечения, от которых избавиться совсем легко. Прямые рассматриваются только в нужном циклическом порядке. Поэтому в действительности мы имеем здесь дело не с индексом пересечения чет-

вертой кривой со всей этой поверхностью, а с индексом пересечения этой четвертой кривой с некоторым куском этой поверхности. Замечательно, что поверхность даёт пленку между третьей и первой кривой, которая лежит на этой поверхности. Это — область на поверхности между C_3 и C_1 , и она примыкает к C_3 с кратностью, равной удвоенному коэффициенту зацепления C_1 с C_2 , а к C_1 — с кратностью равной удвоенному коэффициенту зацепления C_2 с C_3 . Кратность означает, что к кривой подходит несколько листов. Так здесь возникает эта формула.

3.7. Квадрисеканты как необходимый признак нетривиальности.

Эти прямые — квадрисеканты, в том смысле, что они пересекают кривую $C_1 \cup C_2 \cup C_3 \cup C_4$ четырежды. Квадрисеканты появлялись в топологии много раз. Самый первый раз, насколько я знаю, они появились в 1933 г., когда была опубликована работа [5] Эрики Панвиц (E. Pannwitz), где она доказала следующую замечательную теорему: если узел нетривиален, то у него есть хотя бы одна квадрисеканта. То есть для любого нетривиального узла существует прямая, которая пересекает его в четырех точках. У стандартного тривиального узла никаких квадрисекант нет, есть только бисеканты — секущие, которые пересекают его в двух точках.

Потом эту теорему несколько раз переоткрывали и передоказывали. Последний раз это сделал Куперберг (G. Kupferberg) [4], который ее еще и обобщил: он доказал, что то же самое верно и для зацеплений в евклидовом пространстве. В проективном пространстве всё не так. Там бывают нетривиальные узлы, у которых нет квадрисекант. См. рис. 8.

Доказательства теоремы Панвиц красивы, но они не содержат инвариантов. В отличие от того, о чём я сейчас рассказывал, это доказательство существования, которое основано не на инвариантах, а на том, что если квадрисекант нет, то на узел натягивается диск. Такое прямое доказательство.

3.8. Инвариант Кассона и число квадрисекант узла.

Квадрисеканты ещё появились недавно в препринте [1] четырех авторов — Бадни, Коннанта, Скэнелла и Синха. Этот препринт есть в arXiv. И там квадрисеканты снабжаются знаками так, что сумма этих знаков оказывается равной инварианту Кассона — всё тому же первому нетривиальному инварианту Васильева, о котором я сегодня рассказывал. Только суммирование надо производить не по всем квадрисекантам, а лишь по тем, которые пересекают узел в точках, расположенных на узле в квадрисеканте в порядках, определённым образом связанных друг с другом. Тогда такие прямые можно снабдить знаками так, чтобы в сумме получился инвариант Кассона. Знак можно считать нулём, когда порядки пересечения не правильны. Когда порядки правильны, знак равен ± 1 . Знак определя-

ется довольно сложной формулой, я её сейчас, пожалуй, воспроизводить не буду.

Затем Поляк нашел замечательное доказательство, гораздо более элементарное, чем у них. Формулу Бадни, Коннанта, Скэннелла и Синха для инварианта Кассона, Поляк свёл посредством некой непрерывной деформации к формуле для инварианта Кассона в терминах гауссовой диаграммы (я уже рассказывал об этой формуле в самом начале этой лекции). В $S^2 \times S^2 \times S^2$ строятся два четырёхмерных целочисленных цикла. Они строятся по узлу. Их связывает непрерывная деформация, так что они реализуют один и тот же гомологический класс. Индекс пересечения одного из них с диагональю $\{(x, x, x) \mid x \in S^2\} \subset S^2 \times S^2 \times S^2$ оказывается равен сумме знаков квадрисекант, появившейся у Бадни, Коннанта, Скэннелла и Синха, а индекс пересечения другого с той же диагональю — правой части формулы Поляка—Виро для инварианта Кассона этого узла, то есть выражению инварианта Кассона в терминах гауссовой диаграммы.

Так или иначе, подсчёт квадрисекант узла с подходящими знаками даёт инвариант Кассона этого узла, то есть простейший нетривиальный инвариант Васильева. Интересно, конечно, что будет, если рассматривать пересечения узла с окружностями, подсчитывать шестикратные секанты из окружностей. Какие инварианты Васильева это даст? Так должно получаться много разных инвариантов, но это пока не сделано.

4.

Обратимся снова к алгебраическим кривым

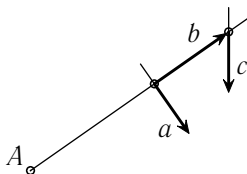
Во всех формулировках до сих пор кривые были произвольными, они не были алгебраическими, мы не предполагали никакой алгебраичности. Действительно, алгебраичность здесь не по существу. Но это происходит по одной из двух причин. Одна причина: в теоремах Юлии Виро прямая с каждым многообразием пересекается всего один раз. А в теоремах Эрики Панвиц и Куперберга прямая пересекается много раз, но схема пересечения такая, что между любыми двумя точками пересечения есть еще другая точка пересечения; они как бы зацеплены. Первая точка пересечения не может слиться со второй точкой и исчезнуть до того, как мы пройдем через третью точку, что, в общем-то, для пересечения с прямой трудновато. Здесь просто комбинаторика пересечения, которая вмонтирована в формулировку, не позволяет произойти следующему: концы хорды на кривой скользят навстречу друг другу, сливаются, и вот хорды уже нет, она как бы соскочила.

С этим явлением можно справиться по-другому и написать формулу такого же сорта, если кривая (или многообразие большей размерности) алгебраическая. Я хочу еще сказать несколько слов об этом.

4.1. Неинвариантность коэффициента самозацепления и его реабилитация. Начнем с самой простой формулировки, которая по этому поводу могла бы быть. Представьте себе, что у нас есть не две кривых, как в самом начале лекции, а одна (узел). И еще есть точка. Я хочу сосчитать количество прямых, проходящих через эту точку и пересекающих мой узел дважды. Можно ли такие прямые снабдить знаками так, чтобы их количество, подсчитанное с этими знаками, было бы инвариантом?

Прежде всего, сразу совершенно ясно, какими знаками можно было бы снабжать прямые. Это те же самые знаки, которые и раньше были. Давайте я скажу об этих знаках подробнее.

Пусть у нас есть две ветви кривой, есть точка, и прямая, проходящая через них. Как в такой ситуации выбрать знак? Надо ориентировать ветви кривой, взять касательные векторы к ним и взять вектор, соединяющий рассматриваемые точки пересечения (рис. 4), взять смешанное произведение этих трёх векторов и взять его знак. Этот знак ± 1 естественно



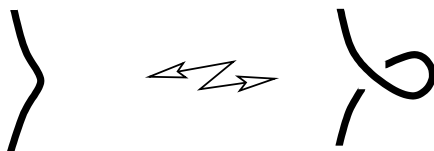
Р и с. 4. Выбор знака



Р и с. 5. Отрицательная точка скрещивания

ассоциируется с этой прямой. Он уже участвовал в определении коэффициента зацепления. Что же это за знак? Если мы смотрим из точки A , то видим такую картинку, как на рис. 5. И знак определяется так: если верхнюю ветвь тащить к нижней, кратчайшим образом, так чтобы ориентации уважались, то если это делается по часовой стрелке, то мы ставим -1 , а если против часовой стрелки, то ставим $+1$. Это обычное соглашение, которое участвует в определении коэффициентов зацепления. Если мы посмотрим на наш узел в целом, то из точки A мы увидим картинку — диаграмму этого узла. Предлагается следующий кандидат для инварианта: просто сумма этих чисел по всем точкам скрещивания.

Хорошо известно, что эта сумма не является инвариантом: она не инвариантна относительно первого движения Рейдемейстера. Если вы возьмете



Р и с. 6. Неинвариантность суммы знаков

завиток (рис. 6) и разгладите его, то эта точка скрещивания исчезает вместе со своим знаком.

Я уже упоминал, что если кривая алгебраическая, то двойная точка не исчезает. Двойная точка остаётся, как в самом начале этой лекции было, она просто становится изолированной. Что это означает на языке прямых? Просто есть прямая, вещественная прямая, которая по-прежнему проходит через нашу точку, и пересекается с кривой дважды, только она пересекается теперь не в вещественной области, а в мнимой, в сопряжённых мнимых точках. Вот как раз здесь-то и нужно, чтобы кривая была алгебраическая. Если она не алгебраическая, то ничего не поделаешь. Если ее нельзя продолжить в комплексную область, то инварианта нет.

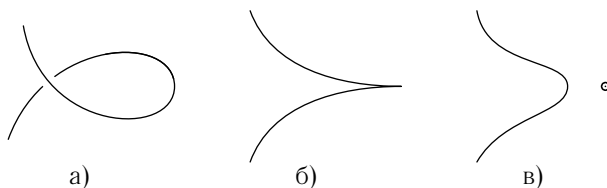
Сопоставление всех этих соображений показывает, что никакого инварианта, который бы оценивал количество решений этой исчислительной задачи и вычислялся бы как сумма каких-то весов прямых, просто не существует. Однако если кривая алгебраическая, то нужно просто поставить исчислительную задачу по-другому: мы считаем со знаками вещественные прямые, которые проходят через данную точку и пересекают нашу кривую в двух точках, не упоминая, какие это точки — вещественные или мнимые. Надо всё считать вместе. Просто дело вот в чем: вещественной прямой, пересекающей кривую в мнимых сопряжённых точках, тоже можно присвоить знак, так чтобы сумма всех знаков не менялась.

4.2. Знак изолированной вещественной точки скрещивания. Сначала я опишу геометрический способ определения этого знака. Что это за прямая? Это прямая, которая пересекается в своей комплексификации с нашей кривой в двух точках, которые сопряжены друг другу. Выберем одну из них. Потом я буду доказывать, что всё, что делается, не зависит от этого выбора. Я выбрал эту точку пересечения, тем самым я выбрал ветвь алгебраической кривой, которая пересекает мою прямую. Как только я сделал этот выбор, на вещественной части этой прямой появляется ориентация, потому что половинка прямой обладает комплексной ориентацией: это часть комплексной прямой, на комплексной прямой есть ориентация, определяемая комплексной структурой. Эта ориентация по-

ловинки прямой индуцирует ориентацию на своем крае — вещественной части прямой. Если бы мы начали с другой половинки, то получилась бы противоположная ориентация. Итак, наша прямая, вдоль которой мы проектируем кривую, оказывается ориентированной.

А теперь давайте посмотрим на плоскость проекции. Локально, это, конечно, просто \mathbb{R}^2 ; но у этого \mathbb{R}^2 есть комплексификация \mathbb{C}^2 . В \mathbb{C}^2 картинка чуть богаче. Во-первых, там лежит \mathbb{R}^2 . А во-вторых, там лежат две мнимые сопряженные ветви кривой, которые тоже двумерны. И одна из них выбрана (та, которая является проекцией выбранной ветви). Я смотрю на неё, она ориентирована, на ней есть её родная комплексная ориентация. И теперь я могу выбрать ориентацию плоскости \mathbb{R}^2 . Я выбираю её так, чтобы индекс пересечения плоскости \mathbb{R}^2 и проекции выбранной ветви вместе с её родной комплексной ориентацией оказался равен $+1$. Так что здесь у меня появляется ориентация на \mathbb{R}^2 . Просто из комплексной картинки в этой точке и из выбора одной из двух ветвей. Что же у нас теперь есть? У нас есть ориентация прямой и ориентация плоскости. И всё это происходит в пространстве $\mathbb{R}P^3$, которое тоже ориентируемо, и, значит, мы можем фиксировать там ориентацию; давайте её фиксируем. Так что у нас здесь возникает знак плюс или минус — знак индекса пересечения этой прямой с этой плоскостью. Этот знак не зависит от выбора ветви, потому что если мы поменяем выбор ветви, ориентация прямой поменяется; но и ориентация плоскости поменяется, потому что мы от этой картинки переходим к комплексно сопряженной.

А еще есть другое замечательное более алгебраическое описание того же самого знака, о котором я должен рассказать, для того чтобы стало понятным, что этот знак связан с тем при деформации рисунка 7. Для этого сначала позвольте мне переписать определение этого знака в вещественном случае.



Р и с. 7. Первое движение Рейдемейстера вещественного алгебраического узла

Переход от зацепления к кривой с изолированной точкой на плоскости — это то, что происходит при первом движении Рейдемейстера. Представьте себе, что у вас есть пространственная кривая, вы берете касательную к ней, встаете на эту касательную и смотрите вдоль неё на

кривую. То, что вы видите — это касп, полукубическая парабола в общем положении (рис. 7, б)). Теперь, если вы немножко сдвинетесь в одну сторону и посмотрите снова, то вы увидите то, что изображено на рис. 7, а), а если вы сдвинетесь в другую сторону, то увидите то, что изображено на рис. 7, в). И вместе эта деформация проекции называется первым движением Рейдемейстера.

В алгебраической ситуации все эти кривые алгебраические. У проекции алгебраической кривой рисунка 7, а) имеется двойная точка. Мы рассматриваем проекцию нашей пространственной кривой, из движущейся точки (да и кривая тоже могла бы двигаться). Мы рассматриваем кривую из этой точки — делаем проекцию. После проекции, конечно, снова получаются алгебраические кривые. Локально это может быть представлено именно как кубические рациональные кривые. И одна-единственная двойная точка не может убежать в мнимую область. Она просто обязана остаться вещественной области. Для нее нет пары, чтобы в таком семействе стать мнимой. Поэтому она остается известно где, для нее можно написать уравнение.

Пусть есть двойная точка. Я уже говорил об этой двойной точке в более общей ситуации, когда она не только что родилась, а просто где-то живет. У нее всегда, во все времена есть знак. А теперь я хочу объяснить, почему знак в точке на рис. 7, а) равен знаку в точке на рис. 7, в). Как получается, что эти знаки совпадают? Для этого мне хочется сначала проанализировать, какой знак на рис. 7, а). А это я уже написал. По-другому можно сказать так: что мы берем и выписываем координаты векторов a , b , c , потом вычисляем детерминант полученной 3×3 -матрицы, а потом берем его знак:

$$\text{sign det} \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

Что такое векторы a , b , c ? Вектор a — это вектор скорости для параметризации, вдоль первой ветви. Пусть эта ветвь параметризуется вектор-функцией r_1 . Тогда $a = \dot{r}_1$. Пусть вторая ветвь параметризуется вектор-функцией r_2 . Тогда вектор b — это просто $r_2 - r_1$, а c — это \dot{r}_2 . В этих обозначениях

$$\text{sign det} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \text{sign det} \begin{pmatrix} \dot{r}_1 \\ r_2 - r_1 \\ \dot{r}_2 \end{pmatrix}.$$

Теперь представьте себе, что всё это мы повторяем для ситуации, когда ветви — мнимые сопряжённые. Вектор-функция r_1 у нас комплексная, r_2 тоже комплексная, но зато $r_2(t)$ является комплексно сопряженным числом к $r_1(t)$, для любого значения параметра t . Далее, \dot{r}_1 — это ка-

кой-то комплексный вектор, а \hat{r}_2 — это сопряженный комплексный вектор. Тогда $r_2 - r_1$ это чисто мнимый вектор (его вещественная часть равна 0), поскольку $r_2 = \bar{r}_1$. Я вычел из вектора сопряженный.

Теперь давайте вычислим этот определитель. Во-первых, можем прибавить первую строку к последней, у нас получится вещественная строка, определитель не изменится. Затем мы можем вычесть половину последней строки из первой. Наконец, мы можем вынести число i за знак определителя дважды:

$$\det \begin{pmatrix} A+iB \\ iC \\ A-iB \end{pmatrix} = \det \begin{pmatrix} A+iB \\ iC \\ 2A \end{pmatrix} = \det \begin{pmatrix} iB \\ iC \\ 2A \end{pmatrix} = (-1) \det \begin{pmatrix} B \\ C \\ 2A \end{pmatrix}.$$

Ясно, что получилось нечто вещественное. У этого вещественного есть знак. Конечно, этот знак связан с тем знаком, потому что в предельном случае, в переходном, когда там был касп, знак этого определителя можно сосчитать по правилу Лопиталья, и получится знак кручения фрагмента. Мораль такая: у вещественного алгебраического узла имеется инвариант — сумма всех таких знаков по двойным точкам, которые возникают на плоскости, как изолированным, так и не изолированным. Эта сумма не меняется при алгебраической изотопии этого узла.

4.3. Жёсткие изотопии вещественных алгебраических узлов. Что я понимаю под алгебраической изотопией? Её ещё называют жёсткой изотопией. Это деформация, при которой узел остаётся вещественным алгебраическим и у него ни в какой момент нет самопересечений.

При деформации узла могут возникать двойные точки. Нужно запретить им возникать: если возникает двойная точка, это уже никакая не изотопия. Двойные точки могут возникать двумя способами. Один способ — банальный, который мы знаем из теории узлов. А второй способ вот какой. Две мнимые сопряженные ветви в пространстве двигаются навстречу друг другу, и на один момент они пересекаются в вещественной точке; а потом они снова расходятся в мнимую область. Значит, вещественная двойная точка мелькнет где-то в стороне от узла (в вещественной области) и исчезнет.

Но это она в 3-мерном пространстве будет мелькать. А если вы смотрите на этот узел со стороны, то вы ее с самого начала видите, потому что у проекции узла жила-была изолированная двойная точка, мы в ней ставили знак, и просто в некоторый момент этот знак пропадёт, а потом он сменится на противоположный. Это тоже явление, которое, коль скоро мы говорим об изотопии, надо запретить.

4.4. Окомплексненная скрученность. Вещественно-алгебраические узлы должны рассматриваться с точностью до алгебраической жесткой

изотопии, и тогда сумма знаков двойных точек его проекции — это инвариант степени 1 по Васильеву.

С точки зрения пространств вещественных алгебраических узлов он объясняется следующим образом. У дискриминантного многообразия есть часть, которая состоит из неизолированных точек скрещивания, и часть, которая состоит из изолированных точек. Каждая из них снабжается ко-ориентацией. Инвариант имеет степень 1, потому что он меняется на одно и то же число, когда вы пересекаете это дискриминантное многообразие, где бы вы его ни пересекали; первая производная постоянна. Во всяком случае, у вещественных кривых в 3-мерном аффинном пространстве инвариантов степени 1 нет. А у вещественных алгебраических кривых, как мы видим, есть. Но тут стоит, конечно, пространство замкнуть, чтоб была сфера. А в проективном пространстве все наши рассуждения и так применимы.

Этот инвариант был определён в моей статье [10]. Он там был назван *encomplexed writhe*. На русский это название следует перевести как «окомплексненная скрученность».

Это явление, конечно, не специфически одномерное. Оно наблюдается всегда, когда есть подмногообразие размерности n в пространстве размерности $2n + 1$. Когда есть коэффициенты зацепления, тогда есть и такой алгебраический коэффициент зацепления, или, лучше сказать, самозацепления. В этом случае мы тоже можем рассматривать исчислительные задачи, но приходится уже (если мы хотим иметь дело с хордами и если не хотим зацеплять эти хорды сами за себя и рассматривать сложные условия прохождения — последовательности прохождения хорды через точки узла) принимать во внимание алгебраичность. То есть надо потребовать с самого начала, чтобы объёмлющее многообразие было алгебраическим. Для вещественных алгебраических узлов и зацеплений можно делать всё, что можно делать для обычных узлов и зацеплений. В частности, для них есть инварианты конечной степени.

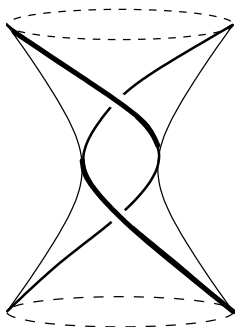
4.5. Теории вещественных алгебраических узлов. В отличие от топологической ситуации, существует много теорий вещественных алгебраических узлов. Ровно столько, сколько есть деформационных семейств кривых в 3-мерном пространстве, а это — большое хозяйство. Они все-таки немножко взаимодействуют между собой. Дело в том, что бывают операции: по кривым из разных семейств вы можете строить кривую нового семейства. Например, если у вас есть две рациональные кривые, которые пересекаются в одной точке, вы можете взять их объединение и продеформировать. Но это только если мы хотим продеформировать объединение в одной точке пересечения. Более сложные ситуации, когда

пересечение происходит в большем количестве точек и мы хотим продеформировать объединение сразу в нескольких из них, не деформируясь. Так что нерациональную кривую из двух рациональных так не получишь.

По-видимому, наиболее интересная и наиболее простая с алгебраической точки зрения вещественная алгебраическая теория узлов возникает, когда мы имеем дело с рациональными кривыми. Если род кривой положителен, то это уносит куда-то в мнимую область топологическую сложность из вещественной области.

Как я уже говорил, объединение пространственных рациональных кривых можно возмущать, но только деформации, сглаживающие всего одну точку пересечения кривых, строятся легко. Независимые деформации мы можем устраивать, если всё происходит на плоскости; в пространстве, к сожалению, этого нет. Поэтому свобода в построении здесь весьма ограничена.

4.6. Проективный нетривиальный узел без квадрисекант. Известно, что в проективном пространстве самый простой нетривиальный узел возникает степени 4, но там он один-единственный. Это как раз тот узел, у которого нет квадрисекант. Это замечательный узел — однокомпонентная кривая (рис. 8). Он дважды пересекает бесконечно удаленную плоскость;



Р и с. 8. Нетривиальный узел без квадрисекант в $\mathbb{R}P^3$

он гомологичен нулю. Это такая гипербола, в которой ветви зацеплены друг за друга. Этот узел лежит на гиперboloиде, и на гиперboloиде это — кривая бистепени $(3, 1)$. Поскольку он лежит на гиперboloиде и имеет бистепень $(3, 1)$, у него нет квадрисекант. Прямая, которая пересекала бы его в четырех точках, должна была бы лежать на этом гиперboloиде. Действительно, прямая, пересекающая гиперboloид в четырех точках, обязана на нем лежать. Следовательно, это одна из образующих гиперboloида. Но если она образующая, то она пересекает наш узел либо в одной точке, либо

в трех, поскольку бистепень такая. Если вы хотите эту кривую построить прямо руками — возьмите 3 прямолинейных образующих и еще одну из другого семейства — и возмутите объединение. Получится какая-то картинка, не такая красивая как на рис. 8, но легко проверить, что при надлежащем подборе точек, которые вы возмущаете, и возмущения у неё тот же изотопический тип.

Это — простейший нетривиальный вещественно-алгебраический узел в трёхмерном проективном пространстве; он имеет степень 4. Кроме него и его зеркального образа в степени 4 есть только тривиальные узлы. В степени 5 есть трилистник, который завязан на проективной прямой. Топологически он получается, если объединить этот узел с некой прямой и взять связную сумму.

Какие есть алгебраические узлы степени 6, я не знаю. Их, наверное, много; есть некоторые примеры, но полный их список неизвестен.

4.7. Пространственные кубики как узлы. В общем, есть такая наука — вещественная алгебраическая теория узлов. Самый простой не плоский вещественный алгебраический узел, топологически тривиальный, это, конечно, просто пространственная кубика. Пространственные кубики бывают двух жестких изотопических типов. И различаются они при помощи этого инварианта, потому что если вы спроектируете такую кривую на плоскость, то вы обязательно увидите у нее двойную точку. Она всегда одна и ее знак при отражении плоскости меняется на противоположный. Так что это узел, как говорят, не зеркален, то есть не эквивалентен своему зеркальному образу. В жестком изотопическом смысле. А в топологическом смысле это, конечно, просто проективная прямая.

Дискриминант в пространстве пространственных кубик состоит из плоских рациональных кривых. Если вы эту кривую будете деформировать в свой зеркальный образ, то по дороге она сплющится, и эта двойная точка проекции станет настоящей двойной точкой, изолированной или неизолированной, но там всё просто устроено.

Но уже в следующих степенях всё устроено не так просто. Это целая наука, которая пока не то что не завершена, но едва появилась. *)

4.8. Encomplexify! Обогащение картинки посредством учета того, что происходит в мнимой области, это явление общего характера, оно никак

*) Примечание при корректуре. К настоящему времени (лето 2011 года) проективные узлы степени ≤ 5 расклассифицированы. Бьёрклунд расклассифицировал рациональные узлы с точностью до жёстких изотопий, а Оревкин и Михалкин расклассифицировали не рациональные узлы, но лишь с точностью до топологических изотопий. Любопытно, что, как показал Бьёрклунд, рациональный узел степени ≤ 5 характеризуется с точностью до жёсткой изотопии своей окомплексненной скрученностью.

не связано с этой размерностью, это происходит и в других размерностях, и это, конечно, имеет смысл рассматривать.

5.

Заключение. Ориентации и сигнатуры нульмерных многообразий

Я хотел бы в конце немножко поговорить о том, что это вообще за предмет. Я в этом докладе рассказал о нескольких ситуациях, когда возникают знаки, приписанные вещественным алгебраическим объектам, и эти знаки позволяют оценивать снизу количество решений исчислительных задач. Эти знаки могут приписываться вещественным алгебраическим объектам или топологическим. Конечно, каждый раз это связано с тем, что соответствующее пространство этих алгебраических объектов ориентируется. И то пространство, с которым мы имеем дело, нульмерно. Скажем, мы хотим считать прямые; это нульмерное пространство. Мы хотим оценить число его точек. А вместо этого мы вычисляем (это всегда легче сделать) сигнатуру этого ориентированного нульмерного многообразия. Раз уж многообразие ориентировано, то у него есть сигнатура.

Сигнатура в нульмерном случае это и есть разница между количеством положительных точек и отрицательных точек. Сигнатура очень инвариантна, она гораздо более инвариантна, чем количество точек само по себе: она инвариантна относительно ориентированных кобордизмов.

Ориентируемое нульмерное многообразие получается, если рассмотреть пересечение ориентированных многообразий бóльшей размерности. Так что если хотите, всё, что я рассказывал сегодня, можно свести к банальному призыву: давайте будем рассматривать ориентации у вещественных алгебраических многообразий, когда они появляются более или менее естественно. Это хороший инструмент для оценки снизу количества решений исчислительных задач или каких-нибудь других задач.

Конечно, это не новый призыв, такое уже много раз было. Например, комплексные ориентации вещественных алгебраических кривых, разбивающих свой комплексификацию, которые ввел В. А. Рохлин, сыграли большую роль в изучении топологии вещественных алгебраических кривых на плоскости. Есть и другие примеры, о некоторых шла речь сегодня, на другие времени не хватило. Но на этом позвольте мне закончить.

Список литературы

[1] R. Budney, J. Conant., K. Scannell, and D. Sinha. *New perspectives on self-linking*, *Advances in Mathematics*. 191 (2005) 78-113; arXiv: math.GT/0303034.

- [2] M. Chasles, *Construction des coniques qui satisfont à cinq conditions*, C. R. Acad. Sci. Paris 58 (1864), 297—308.
- [3] M. N. Goussarov, M. Polyak, and O. Viro, *Finite type invariants of classical and virtual knots*, Topology, vol. 39, no. 5, (2000) 1045—1068; arXiv:math.GT/9810073.
- [4] G. Kuperberg, *Quadriseccants of knots and links*, J. Knot Theory Ramifications, **3:1** (1994), 41—50.
- [5] E. Pannwitz, *Eine elementargeometrische Eigenschaft von Verschlingungen und Knoten*, Math. Ann. 108 (1933), 629—672.
- [6] Michael Polyak and Oleg Viro, *Gauss diagram formulas for Vassiliev invariants*, International Mathematics Research Notes 1994:11.
- [7] F. Ronga, A. Tognoli, and T. Vust, *The number of conics tangent to 5 given conics: The real case*, Rev. Mat. Univ. Complut. Madrid 10 (1997) 391—421.
- [8] J. Steiner, *Elementare Lösung einer geometrischen Aufgabe, und über einige damit in Beziehung stehende Eigenschaften der Kegelschnitte*, J. Reine Angew. Math. 37 (1848), 161—192.
- [9] Julia Viro, *Lines joining components of a link*, Journal of Knot Theory and Its Ramifications, **18:6** (2009) 865—868; *Lines and circles joining components of a link*, arXiv: math.GT/0511527.
- [10] Oleg Viro, *Encomplexing the writhe*, «Topology, Ergodic Theory, Real Algebraic Geometry. Rokhlin's Memorial» Amer. Math. Soc. Transl. ser 2, vol 202, ed. V. Turaev and A. Vershik; arXiv: math.GT/0005162.
- [11] J.-Y. Welschinger, *Invariants of real symplectic 4-manifolds and lower bounds in real enumerative geometry*, Invent. Math., 162(1) (2005) 195—234; arXiv: math.AG/0303145.
- [12] J.-Y. Welschinger, *Spinor states of real rational curves in real algebraic convex 3-manifolds and enumerative invariants*, Duke Math. J., 127(1) (2005) 89—121, 2005; arXiv: math.AG/0311466.

24 февраля 2005 г.

Оглавление

Ю. И. М а н и н. Георг Кантор и XX век	3
В. В. Б а т ы р е в. Торические вычеты в зеркальной симметрии . . .	25
В. Ю. К а л о ш и н. Случайные блуждания вдоль орбит хаотических отображений	47
А. А. Г л у ц ю к. Униформизация. Расслоенные версии и новое доказательство	63
Ю. И. М а н и н. Дробные размерности в геометрии и алгебре	85
А. Т о о м. Неэргодичность в 1-D процессе с переменной длиной . .	105
А. Н. С к о р о б о г а т о в. Точки на кривых Шимур над числовыми полями	113
Г. К о ш е в о й. Кристаллическая комбинаторика	129
О. Я. В и р о. Оценки снизу для числа решений исчислительных проблем в топологии и вещественной алгебраической геометрии	147

ГЛОБУС

Общематематический семинар. Выпуск 5

Научный редактор *М. А. Цфасман*

Редактор *В. В. Прасолов*

Подписано в печать 23.08.2011 г. Формат $70 \times 100 \frac{1}{16}$. Бумага офсетная.
Печать офсетная. Печ. л. 11. Тираж 400 экз. Заказ №

Издательство Московского центра непрерывного математического образования.
119002, Москва, Большой Власьевский пер., 11. Тел. (499) 241–74–83.

Отпечатано с готовых диапозитивов в ППП «Типография „Наука“».
121099, Москва, Шубинский пер., 6.