НЕЗАВИСИМЫЙ МОСКОВСКИЙ УНИВЕРСИТЕТ

globus ГЛОБУС

Общематематический семинар. Выпуск 4

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

Москва
Издательство МЦНМО
2009

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. 4. — 2009. — 224 с. — ISBN 978-5-94057-508-5.

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

Четвертый выпуск включает доклады С. Н. Артемова, А. М. Бородина, С. Г. Влэдуца, В. И. Данилова, Е. Б. Дынкина, Г. Л. Литвинова, Р. А. Минлоса, А. Н. Рыбко, В. В. Сергановой, М. В. Финкельберга, О. В. Шварцмана, В. В. Шехтмана, М. А. Шубина и Д. Б. Фукса.

УДК 51(06), ББК 22.1я5

ISBN 978-5-94057-064-6
ISBN 978-5-94057-508-5 (Вып. 4)

© НМУ, 2009.
© МЦНМО, 2009.

Предисловие

Перед Вами четвертый сборник докладов на семинаре «Глобус» — общематематическом семинаре Независимого Московского университета. Авторы, как обычно, рассказывают математикам других специальностей, как они видят свою область и что в ней нового.

Доклад С. Г. Влэдуца посвящен теории бесконечных глобальных полей: какова арифметика бесконечных расширений поля рациональных чисел и алгебраических кривых «бесконечного рода» над конечным полем. Р. А. Минлос рассказывает, как придать математический смысл столь популярному в физике методу квантования при помощи интегрирования по континуальному множеству путей. Квантование есть усложнение классических конструкций; теперь предположим, что классические конструкции — это уже квантование чего-то, и зададимся вопросом, что сверхпростое должно получиться в «классическом пределе»; это объясняет Г. Л. Литвинов. Доклад М. В. Финкельберга — об инстантонах и компактификации пространства модулей G-расслоений. М. А. Шубин пересказывает работу Нэша о равновесии, за которую тот получил Нобелевскую премию по экономике. В. В. Серганова говорит о супералгебрах Ли и методе орбит. В. В. Шехтман рассказывает о вертексных алгебрах, понятии, возникшем в физике из теории струн и тесно связанном с многообразиями Калаби—Яо. Доклад А. Н. Рыбко посвящен случайным процессам в больших коммуникационных сетях. С. Н. Артемов говорит об интуиционистской логике, понимаемой не с узко-логической, а с общематематической точки зрения. В. И. Данилов объясняет, что можно сказать о спектре сумме двух матриц, если мы знаем спектр каждой из них, и о связи этой задачи с дискретно-вогнутыми функциями. Доклад А. М. Бородина повествует о линейных разностных уравнениях и о случайных перестановках; как и все остальное на свете (см. упомянутый доклад Влэдуца) это связано с распределением нулей дзета-функции. О. В. Шварцман говорит о кристаллографических группах и фактор-пространствах по ним. Д. Б. Фукс рассказывает о лежандровых кривых и узлах. Завершается сборник докладом классика московской математической школы Е. Б. Дынкина; он объясняет, как комбинировать теорию вероятностей и анализ при изучении броуновского движения.

Как видите, спектр сюжетов весьма широк. Читайте, разбирайтесь, — Бог в помощь.

Огромное спасибо В. В. Прасолову, И. Миклашевскому, А. С. Протопопову и сотрудникам издательства МЦНМО.

М. А. Цфасман

С. Г. Влэдуц

ОТ ОСНОВНОЙ ТЕОРЕМЫ АРИФМЕТИКИ ДО БЕСКОНЕЧНЫХ ГЛОБАЛЬНЫХ ПОЛЕЙ

Я начну с основной теоремы арифметики. Она говорит, что каждое число $m \in \mathbb{Z}$ раскладывается в произведение простых чисел: $m = p_1^{\alpha_1} \dots p_r^{\alpha_r}$. Это разложение единственно в разумном смысле: простые множители можно переставлять; нужно собрать одинаковые множители в степени, чтобы $p_i \neq p_j$.

Это не единственный случай кольца алгебраических чисел, для которого имеет место такое свойство. Например, можно рассмотреть кольцо гауссовых целых $\mathbb{Z}[i]$, где $i^2 = -1$. Можно определить, что такое простое гауссово целое число. Для любого гауссова целого числа $a + bi = m$ имеется однозначное разложение на простые множители: $m = p_1^{\alpha_1} \dots p_r^{\alpha_r}$. С точностью до мелочей. Здесь мелочь такая. Для обычных целых чисел есть знак, и разложение единственно с точностью до умножения на ± 1. А здесь разложение единственно с точностью до умножения на ± 1 и на $\pm i$. Если эту операцию позволить, то разложение, по сути, единственно.

Таких колец немало, более того, гипотетически их бесконечно много. Но это не доказано. Видимо, это довольно трудная проблема. Можно, например, взять кольцо $\mathbb{Z}[\sqrt{2}]$, элементами которого являются выражения $a + b\sqrt{2}$, где $a, b \in \mathbb{Z}$. В этом кольце можно определить простые числа, и будет верно то же самое: разложение на простые множители единственно. Таких колец очень много. Но есть и кольца, в которых разложение не единственно. Например, в кольце $\mathbb{Z}[\sqrt{-5}]$ число 6 разлагается двумя способами: $6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$. Легко проверить, что все эти четыре элемента простые: дальше они ни на что не разлагаются.

Для того чтобы изучать арифметику этих колец, придумано множество разных инвариантов. Один из этих инвариантов, может быть самый важный, как раз измеряет насколько отклоняется кольцо от основной теоремы арифметики. Кольцо $\mathbb{Z}$ не отклоняется, $\mathbb{Z}[i]$ не отклоняется, $\mathbb{Z}[\sqrt{2}]$ не отклоняется, а кольцо $\mathbb{Z}[\sqrt{-5}]$ уже отклоняется.

Можно ожидать, что чем сложнее кольцо, тем сильнее оно отклоняется от основной теоремы арифметики. В каком-то смысле это верно, но не совсем буквально.

Теперь я расскажу более формально, какие имеются точные определения в этой области для полей алгебраических чисел и глобальных полей конечной характеристики (я сейчас скажу, что это такое). Потом я расскажу, как можно попытаться это перенести на случай полей бесконечной степени.

Глобальные поля конечной степени

Глобальные поля конечной степени бывают двух типов. Их можно определить абстрактно, некоторым требованием локальной компактности. Но я не буду давать это определение. Важно то, что это очень естественный класс полей, для которых, собственно, и строится арифметика. Сейчас я опишу, какие бывают поля и какие для них существуют ценности.

Числовые поля. Числовое поле K — это просто конечное расширение поля $\mathbb{Q}$: $[K:\mathbb{Q}]=n<\infty$. Когда говорят *поле алгебраических чисел*, всегда имеют в виду, что это — поле алгебраических чисел конечной степени. Это не добавляется, потому что до нашей работы с М. А. Цфасманом, можно сказать, никакой теории полей алгебраических чисел бесконечной степени, в общем-то, не существовало. Мы претендуем на то, что какую-то теорию мы начинаем все-таки развивать. Хотя, конечно, это только самое начало теории. Но есть более или менее содержательные определения и результаты, которые показывают, что это возможно.

Функциональные поля. На функциональные поля есть два эквивалентных взгляда. Согласно одному взгляду функциональное поле K — это поле, которое является конечным расширением поля $\mathbb{F}_r(T)$: $[K:\mathbb{F}_r(T)]=n$. Здесь $\mathbb{F}_r$ — конечное поле из $r=p^m$ элементов, T — переменная. Важное отличие от числовых полей состоит в том, что здесь число n неинвариантно. Для числового поля мы всегда можем сказать, какая у него степень: рассматриваем K как векторное пространство над $\mathbb{Q}$ и смотрим, какая у него размерность. А здесь нет никакого выделенного n, потому что мы можем, например, просто взять $\mathbb{F}_r(T^2)$ вместо $\mathbb{F}_r(T)$. Поле $\mathbb{F}_r(T^2)$ содержится в $\mathbb{F}_r(T)$, содержится в K, но теперь $[K:\mathbb{F}_r(T^2)]=2n$. Можно рассматривать минимальное n, но это неудобно, потому что такое крайне трудно считаемо. За исключением случая, когда минимальное n равно 2, эта задача почти что безнадежная.

Это — алгебраическая точка зрения на функциональные поля. Очень часто более удобна (и при этом эквивалентна) точка зрения, при которой поле K — это поле функций на кривой X над конечным полем $\mathbb{F}_r$: $K=\mathbb{F}_r(X)$. Кривая $X/\mathbb{F}_r$ — это решение системы алгебраических уравнений, которое имеет размерность 1. Грубо говоря, уравнений на 1 меньше, чем

неизвестных. Нужно, конечно, потребовать, чтобы множество рациональных функций (отношение одного полинома к другому) было полем. Прежде всего нужно потребовать, чтобы кривая X была неприводима (приводимая кривая задается уравнением $FG = 0$). Должно даже выполняться более сильное условие: кривая должна быть абсолютно неприводимой (т. е. неприводимой над замыканием основного поля). Кроме того, чтобы это представление было единственным, нужно потребовать, чтобы кривая X была гладкой и проективной. Гладкая означает, что соответствующая матрица частных производных имеет максимальный ранг на этой кривой. Проективность означает, что уравнения кривой на самом деле заданы в проективном пространстве. Если потребовать, чтобы кривая X была абсолютно неприводимая, гладкая и проективная, то тогда такая кривая X для каждого K существует ровно одна. Так что это все равно: говорить о поле K или о кривой X.

Свойства глобальных полей

Эти два объекта, поле K и кривая X, имеют параллельные свойства. Я уже говорил, что у них есть общее абстрактное определение. Кроме того, у них много общих свойств. Но при этом перевод одних свойств в другие — это не совсем очевидное дело. Тем не менее, оно очень полезно. Например, очень многие работы Андре Вейля состоят как раз в том, что он брал какую-то вещь, которая доказана в функциональном случае (а это по разным причинам часто оказывается проще), и искал, чему она соответствует в случае числовых полей. Как правило, это дело трудное, но зато полезное.

Какие у глобальных полей бывают инварианты? Для числовых полей первый инвариант — это степень n. Этот инвариант корректно определен.

Как я уже говорил, степень функционального поля не инварианта. Можно определять минимальную степень, но тогда это очень сложно считается. Но есть не совсем очевидная замена степени n в этом случае. А именно, $N = |X(\mathbb{F}_r)|$ — число точек кривой с координатами в конечном поле $\mathbb{F}_r$. Это множество конечно, потому что оно лежит в $\mathbb{P}^b(\mathbb{F}_r)$, а $|\mathbb{P}^b(\mathbb{F}_r)| = \dfrac{r^{b+1}-1}{r-1}$.

Число $N = |X(\mathbb{F}_r)|$ является очень хорошей заменой числа $n_0 = \min n$, где минимум берется по тем n, для которых существует такая переменная T, что $[K : \mathbb{F}_r(T)] = n$. Более того, число N некоторым образом связано с числом n_0. А именно, есть односторонняя оценка $n_0 \geqslant \dfrac{N}{r+1}$. Эта оценка существенно односторонняя; ничего с другой стороны сказать нельзя. Я сейчас в двух словах скажу, как эта оценка доказывается.

Это совершенно очевидная вещь. Если есть расширение, для которого $[K:\mathbb{F}_r(T)]=n$, то геометрически это означает, что есть отображение $X\to\mathbb{P}^1$ степени n. Действительно, поле рациональных функций на $\mathbb{P}^1$ — это как раз $\mathbb{F}_r(T)$. Мы можем поднять функции на $\mathbb{P}^1$ на X; это дает вложение поля $\mathbb{F}_r(T)$ в поле K, причем $[K:\mathbb{F}_r(T)]=n$. Расширение $[K:\mathbb{F}_r(T)]$ степени n — это то же самое, что отображение $X\to\mathbb{P}^1$ степени n. Степень отображения $X\to\mathbb{P}^1$ равна n. Это означает, что над каждой точкой $x\in\mathbb{P}^1$ лежит не более n точек кривой X. В некотором смысле, их, как правило, ровно n. Но, во-первых, бывают точки ветвления, когда точек меньше. А главное, никто не гарантировал, что если точка $x\in\mathbb{P}^1$ рациональна, т. е. ее координаты лежат в $\mathbb{F}_r$, то лежащие над ней точки тоже рациональны. Вообще говоря, это не так. Например, если отображение задано формулой $y^2=T$, то $K(y)=K(\sqrt{T})$. Мы извлекаем корень из какого-то числа в $\mathbb{F}_r$; он не обязан лежать в $\mathbb{F}_r$. Поэтому некоторые точки, лежащие над точкой x, могут быть не рациональными. Всего на $\mathbb{P}^1$ рациональных точек $r+1$, потому что точки на $\mathbb{P}^1$ — это обычные элементы $\mathbb{F}_r$ и еще ∞. Поэтому на X рациональных точек не может быть больше чем $(r+1)n$. Это число обязательно больше или равно N. Это верно для любого отображения; в частности, для минимального тоже. Тем самым неравенство $n_0\geqslant\frac{N}{r+1}$ доказано.

Это один инвариант, но он в двух ипостасях: степень для числовых полей и число точек для функциональных полей. Есть еще один важный инвариант. Он называется по-разному для этих двух случаев.

Для числовых полей этот инвариант называется *дискриминант*; мы будем рассматривать *абсолютную величину дискриминанта*:

$$D_K=|D_{K/\mathbb{Q}}|.$$

В поле K можно рассмотреть кольцо целых

$$\mathcal{O}_K=\{\xi\in K:\xi\in\overline{\mathbb{Z}}\}.$$

То, что $\xi\in\overline{\mathbb{Z}}$ (элемент ξ цел над $\mathbb{Z}$) означает следующее. Каждый элемент K удовлетворяет какому-то уравнению $a_n\xi^n+a_{n-1}\xi^{n-1}+\ldots+a_0=0$ с целыми коэффициентами. Если $a_n=1$, то элемент ξ целый. Не совсем очевидно, но можно доказать, что $\mathcal{O}_K$ — подкольцо. Более того, из общих соображений видно, что это кольцо свободно как абелева группа, причем оно является свободной абелевой группой ранга n, т. е. имеет вид $\mathbb{Z}\omega_1\oplus\mathbb{Z}\omega_2\oplus\ldots\oplus\mathbb{Z}\omega_n$. Эта абелева группа свободна потому, что там нет кручения, а если у нас есть конечно порожденный модуль над $\mathbb{Z}$ без кручения, то он свободный. По определению $D_{K/\mathbb{Q}}=\det(\omega_i\cdot\omega_j)$ (от выбора базиса это число не зависит). Число $D_{K/\mathbb{Q}}$ целое; оно имеет знак $(-1)^{r_2}$, где r_2 это вот что. Число n расщепляется на два слагаемых: $n=r_1+2r_2$, где

r_1 — число вещественных вложений поля K, а r_2 — число пар комплексных вложений. Это проще всего определить так. Рассмотрим $K \otimes \mathbb{R}$. Это — полупростая абелева алгебра. Она имеет вид $\mathbb{R}^{r_1} \oplus \mathbb{C}^{r_2}$. Знак $(-1)^{r_2}$ мы будем игнорировать, т. е. будем брать абсолютную величину.

Дискриминант — инвариант поля, хотя он и определяется через кольцо. Естественно, $D_{\mathbb{Q}} = 1$. Здесь нечего доказывать, потому что базисный элемент только один: $\omega_1 = 1$. Более того, теорема Эрмита утверждает, что $\mathbb{Q}$ характеризуется этим свойством: если у нас есть числовое поле с дискриминантом 1, то тогда это поле равно $\mathbb{Q}$. И вообще, дискриминант довольно сильно характеризует поле. Например, квадратичные поля (т. е. поля степени 2), если брать дискриминант со знаком, полностью определяются дискриминантом. Дискриминанты бывают положительные (тогда квадратичное поле вещественное) и отрицательные (тогда поле мнимое квадратичное). Квадратичные поля полностью определяются дискриминантом. Самый маленький пример, когда существуют разные поля с одним и тем же дискриминантом, довольно большой: дискриминант содержит десяток цифр (там появляются сразу три разных кубических поля). Конечно, когда дискриминант растет, таких полей становится все больше и больше. А вообще, это инвариант, который довольно хорошо характеризует поле.

Замена дискриминанта в функциональном случае — совершенно естественная вещь. Это — род кривой (род поля). Род кривой X — это целое неотрицательное число $g(X)$, которое определяется многими разными способами. Например, можно взять регулярные формы на X, которые всюду регулярны. Это — конечномерное пространство $\Omega^1[X]$ над основным полем $\mathbb{F}_r$, и $\dim_{\mathbb{F}_r} \Omega^1[X] = g$ — это по определению и есть род кривой. Можно определять род и разными другими способами. Опять же $g(\mathbb{P}^1) = 0$, и это свойство характеризует проективную прямую $\mathbb{P}^1$, т. е. если $g(X) = 0$ для некоторой абсолютно неприводимой гладкой проективной кривой X, то $X \simeq \mathbb{P}^1$ (числовые поля можно брать не с точностью до изоморфизма: они просто где-то лежат, а кривые нужно брать с точностью до изоморфизма). У кривой, изоморфной $\mathbb{P}^1$, поле функций — это $\mathbb{F}_r(T)$.

Здесь, конечно, нет никакой однозначности. Здесь есть однозначность для рода 0: кривая рода 0 изоморфна $\mathbb{P}^1$. А если $g(X) = 1$, то таких кривых (они называются *эллиптическими*) примерно $2r$ штук. Для каждого g число кривых конечно. Для рода 1 их $2r + O(1)$ штук, где $O(1)$ — целое число между 0 и 12. А если $g(X) = g \geqslant 2$, то число кривых примерно равно Cr^{3g-3}, где C — некоторая константа. Тут, конечно, род гораздо менее жесткий инвариант (по крайней мере на первый взгляд), чем дискриминант для числовых полей.

Эти две вещи, род и дискриминант, параллельны. Но на самом деле правильным аналогом рода является не дискриминант D_K, а $g_K = \log\sqrt{D_K}$. Действительно, род начинается с нуля, а дискриминант начинается с 1. И поведение в башнях, про которое я сейчас скажу, разное.

При такой нормализации равенство $g_K = 0$ характеризует $\mathbb{Q}$; равенство $g(X) = 0$ характеризует $\mathbb{P}^1$. Если мы рассмотрим все поля, для которых $g_K \leqslant M$, то их конечное число. Кривых, для которых $g \leqslant M$, тоже конечное число. Для функционального случая примерно понятно, сколько их. А сколько их для числовых полей — это трудная задача, если допускать любое n. Если $n = 2$, то это число сразу считается: оно линейно по M. А вот уже для кубических полей точного результата нет. А если допускать любые n, то абсолютно непонятно, как это число себя ведет. Точнее говоря, есть гипотезы и для асимптотики степеней и для коэффициентов, но доказано очень мало.

Еще один аргумент. Если у нас есть расширение полей $L \supset K$ степени m, то $g_L = mg_K + \ldots$; здесь mg_K — основной член, а многоточием обозначен поправочный член, который зависит не только от степени (этот поправочный член, вообще говоря, может быть и больше, чем основной; но формула имеет такой вид). Когда есть отображение кривых $X \to Y$ (это то же самое, что вложение полей) степени m, то $(g_X - 1) = m(g_Y - 1) + \ldots$ Здесь есть небольшая несогласованность. Ее можно устранить, считая, что род числового поля — это $g_K + 1$. Тогда формула для поведения в башне будет точно такая же, но зато утратится то свойство, что равенство рода 0 соответствует минимальному полю. Это вопрос вкуса, потому что это — сдвиг на 1, за которым можно проследить. Мы предпочитаем, чтобы род 0 соответствовал минимальному полю.

Определение рода кривой стандартно, а определение рода числового поля, которое здесь приведено, нестандартное. Для этого определения есть некоторые аргументы, но это не есть общепринятое определение.

Между n и g_K в случае числовых полей и между N и g в случае функциональных полей есть связь. Связь в числовом случае называется по-разному; есть много результатов, дающих эту связь. Самый первый результат называется *константой Минковского*, а самый новый результат называется *неравенством Одлыжко—Серра*. Это неравенство говорит, что дискриминант не может быть очень маленьким; он ограничен снизу следующей функцией от степени:

$$D \geqslant C_1^{r_1} C_2^{2r_2} e^{o(1)}, \text{ где } e^{o(1)} \to 1, \text{ когда степень растет.} \qquad (1)$$

Самые лучшие значения констант здесь такие: $C_1 = 8\pi e^{\gamma+\pi/2}$ и $C_2 = 8\pi e^{\gamma}$ (γ — константа Эйлера). Этот результат получен, правда, только

в предположении обобщенной гипотезы Римана. Есть вариант, когда C_1 и C_2 поменьше; этот результат безусловный, без гипотезы Римана. Факт тот, что дискриминант ограничен экспоненциальной функцией от r_1 и r_2 (можно сказать, от степени n), когда n растет.

В функциональном случае над полем констант $\mathbb{F}_r$ тоже есть соотношение между числом точек N и родом g, которое называется *теорема Вейля*. Она говорит, что

$$r+1-2\sqrt{r}\,g \leqslant N \leqslant r+1+2\sqrt{r}\,g.$$

Заметьте, что это — некое выражение, линейное по N и g. На самом деле, выражение $r+1-2\sqrt{r}\,g$, когда g растет, становится отрицательным; это не очень интересно. Интересна, конечно, верхняя оценка. Эта вещь такого же типа, как неравенство (1), только неравенство (1) нужно прологарифмировать. Если его прологарифмировать, то получится неравенство

$$2g_K \geqslant r_1 \log C_1 + r_2 \log C_2 + o(1).$$

Там линейная форма от N меньше линейной формы от g, здесь тоже линейная форма от r_1 и r_2 (при этом $n = r_1 + 2r_2$) меньше линейной формы от g. Это — еще одно объяснение, почему нужно рассматривать логарифм дискриминанта. Чуть позже мы увидим, что эти результаты не просто параллельные, но их можно и единообразно сформулировать: можно придумать такую формулировку, что эти неравенства будут выглядеть как одна и та же формула.

Это первые два инварианта: степень и дискриминант. Есть еще такой инвариант, как число классов, который как раз и отражает отклонение поля от единственности разложения на простые множители. Группа классов в числовом случае определяется следующим образом. Мы рассматриваем кольцо целых $\mathcal{O}_K$ (все целые элементы поля K), рассматриваем все его ненулевые идеалы $I \subset \mathcal{O}_K$. Они определяют полугруппу по умножению: если мы перемножим два идеала, то получим идеал. Эту полугруппу можно погрузить в группу, обратив элементы группы формально. Это означает, что можно рассматривать дробные идеалы $I \cdot J^{-1}$. Здесь J^{-1} не идеал, а модуль над $\mathcal{O}_K$, который, вообще говоря, в $\mathcal{O}_K$ не лежит. Это уже будет группа $\mathrm{Ideal}(\mathcal{O}_K)$ — группа дробных идеалов в $\mathcal{O}_K$. Она бесконечная. Например, для кольца $\mathbb{Z}$ группа дробных идеалов — это свободная абелева группа, натянутая на простые числа.

Среди идеалов в $\mathcal{O}_K$ есть главные идеалы. Главный идеал — это идеал вида (α), где $\alpha \in K$. Элемент α можно записать в виде $\alpha = a/b$, где $a, b \in \mathcal{O}_K$. Тогда $(\alpha) = (a)(b)^{-1}$. Главные идеалы образуют подгруппу P в группе $I = \mathrm{Ideal}(\mathcal{O}_K)$. Группой классов называется факторгруппа I/P.

Эта группа уже конечна. Ее мощность называется числом классов данного поля и обозначается $h(K)$. Поле удовлетворяет основной теореме арифметики тогда и только тогда, когда $h(K)=1$.

Про $h(K)$ известно удивительным образом очень мало. Например, не известно, конечно или бесконечно число полей, для которых $h(K)=1$. И это при том, что для вещественных квадратичных полей есть эвристика, что таких полей 3/4 (точнее, 73,2%). То есть для подавляющего числа вещественных квадратичных полей должно выполняться равенство $h(K)=1$, но неизвестно даже, бесконечно это число или нет. Единственный по-настоящему значительный результат в этом направлении носит скорее отрицательный характер. Он говорит, что если рассмотреть мнимые квадратичные поля $\mathbb{Q}(\sqrt{-D})$, где $D>0$, то таких полей с числом классов 1 ровно 9. Наименьший по абсолютной величине дискриминант $D=-4$, а наибольший $D=-163$. То, что их конечное число, было известно еще Гауссу, а то, что их ровно 9, доказали лет 35 назад, причем это очень трудная теорема. Соответственно, для всех остальных мнимых квадратичных полей нет единственности разложения. Более того, для мнимых квадратичных полей известно, что $h(K)\to\infty$ при $D\to\infty$ (это было известно еще Гауссу). Это более или менее единственный случай, когда ответ есть. С 30-х годов было известно, что таких дискриминантов не более 10. Поэтому это долго называлось *проблемой десятого дискриминанта*. Существует десятый дискриминант, для которого число классов равно 1, или нет? Потребовалось лет 40 больших усилий, чтобы доказать, что этого десятого дискриминанта нет. Доказательство технически не очень сложное, но идейно оно очень сложное. Идея была очень нетривиальная. Потом были найдены другие доказательства, но все они очень нетривиальные.

Существует очень похожее (можно даже сказать точно такое же) определение группы классов в функциональном случае. Я не буду его приводить. Оно дословно то же самое, только там обозначения аддитивные, а не мультипликативные. Есть еще одна маленькая разница: нужно взять некую подгруппу в группе идеалов, которая там называется группой дивизоров; нужно взять группу дивизоров степени 0. А так все то же самое. В геометрическом случае число классов — это число точек на якобиане кривой. Якобиан кривой — это g-мерное абелево многообразие над конечным полем. Число рациональных точек на нем конечно. Число классов — это и есть в точности число точек на якобиане. Поэтому с ним работать гораздо легче, чем с $h(K)$ в числовом случае. Про него гораздо больше известно, потому что это — число точек на очень хорошем многообразии. Все кривые с $h=1$ классифицированы, и это очень легко сделать. И некоторые довольно глубокие результаты теории алгебраических чисел произошли

из того, как интерпретировать некоторые свойства якобиана для групп классов. Это — теория Ивасавы. Так что это очень плодотворная аналогия. Обычно в геометрической ситуации результаты легче доказывать. Их потом переносят на числовой случай: пытаются доказать или частично доказывают.

У числового поля есть еще один инвариант, который тривиален по определению в функциональном случае, и он как раз очень сильно затрудняет работу с числом классов. Это — регулятор R. Я не буду его определять, только скажу два слова, откуда он происходит. Нужно рассмотреть мультипликативную группу $\mathcal{O}_K^*$. Теорема Дирихле о единицах говорит, что по модулю кручения эта группа изоморфна $\mathbb{Z}^{r_1+r_2-1}$, т. е. $\mathcal{O}_K^*/\operatorname{tors} \cong \mathbb{Z}^{r_1+r_2-1}$. Заметьте, что если $r_1 = 0$ и $r_2 = 1$ (т. е. поле мнимое квадратичное), то эта группа тривиальна. Регулятор — это некий детерминант, связанный с этой группой. Нужно взять ее образующие, нужно взять их логарифмы, составить из них матрицу (она не квадратная) и взять минор максимального порядка.

Равенство $R = 1$ выполняется, если поле мнимое квадратичное. Это — основная причина, почему так много известно про число классов в случае, когда поле мнимое квадратичное. Посчитать регулятор исключительно трудно. Про него есть кое-какая информация, но уже в случае вещественного квадратичного поля, когда $r_1 = 2$ и $r_2 = 0$, по определению $R = \log \varepsilon$, где ε — основная единица. (Для вещественного квадратичного поля группа $\mathcal{O}_K^*$ циклическая, у нее есть образующая; положительная образующая — это и есть основная единица.) Число $R = \log \varepsilon$ ведет себя крайне нерегулярно. Бывает так, что для двух соседних дискриминантов регулятор одного очень маленький, а другого очень большой. И никакого рационального объяснения поведения регулятора, даже в случае вещественного квадратичного поля, не известно. Для вещественных квадратичных полей есть по крайней мере алгоритм (тоже не совсем тривиальный), который вычисляет регулятор. Для каждого конкретного поля посчитать регулятор можно, но предсказать заранее, большой он или маленький, совершенно невозможно. Нетривиальность R — это основное препятствие к работе с h.

Известен следующий результат, очень трудный: $R > 0{,}1$. По определению число R вещественное положительное, но оно может быть и маленьким. И есть реально поля, у которых оно близко к 0,1. Известно поле с минимальным R.

Почему h и R так тесно связаны? Дело в том, что h и R вместе входят в вычет дзета-функции. Сейчас я напомню, что такое дзета-функция и ее вычет, а потом скажу, что такое теорема Брауэра—Зигеля, которая связывает h и R.

Дедекиндова дзета-функция определяется следующим образом:

$$\zeta_K(s) = \sum_{\mathfrak{a} \in \mathrm{Ideal}} N(\mathfrak{a})^{-s}. \tag{2}$$

Суммирование ведется по целым идеалам $\mathfrak{a}$ в $\mathcal{O}_K$. Как абелева группа $\mathfrak{a}$ имеет тот же ранг, что $\mathcal{O}_K$. Поэтому индекс $[\mathcal{O}_K : \mathfrak{a}]$ конечен. По определению $N(\mathfrak{a}) = [\mathcal{O}_K : \mathfrak{a}]$. Если $K = \mathbb{Q}$, то получается обыкновенная дзета-функция Римана $\zeta(s)$. Ряд (2) абсолютно сходится в полуплоскости $\mathrm{Re}\, s > 1$. Он сходится равномерно на компактах. В полуплоскости $\mathrm{Re}\, s > 1$ этот ряд определяет аналитическую функцию, которая продолжается до мероморфной функции с единственным полюсом в точке 1. В точке 1 полюс потому, что гармонический ряд расходится. Вычет дзета-функции в точке $s = 1$ следующий:

$$\operatorname*{Res}_{s=1} \zeta_K(s) = \frac{2^{r_1}(2\pi)^{r_2} hR}{w_K \sqrt{D_K}},$$

где w_K — порядок группы кручения в группе единиц, т. е. количество корней из единицы в $\mathcal{O}_K$. Это величина небольшая; ее легко оценить, в этой формуле она не мешает. Хорошо работать с произведением hR. Когда $R = 1$, вычет дзета-функции дает формулу прямо для h. Это — причина, по которой h в этом случае относительно хорошо известно. А в общем случае, для произвольных полей, отделить в этой формуле h от R исключительно трудно. Поэтому многие результаты относятся к произведению hR. В частности, есть результат, который называется *теорема Брауэра—Зигеля*. Он заключается в следующем. Рассмотрим последовательность полей K с растущим дискриминантом, для которой $n/\log D_K \to 0$ (в дальнейшем мы увидим, существенно это условие или нет). Тогда $\dfrac{\log hR}{\log \sqrt{D_K}} \to 1$; в наших обозначениях $\log \sqrt{D_K} = g_K$. В частности, если $R = 1$ (т. е. в случае мнимых квадратичных полей), то $(\log h)/g_K \to 1$. Для доказательства теоремы Брауэра—Зигеля нужно предположить еще какие-нибудь дополнительные технические условия: либо гипотезу Римана, либо нормальность поля, либо гипотезу Артина. Для мнимых квадратичных полей выполнено условие нормальности, поэтому для мнимых квадратичных полей $\log h$ ведет себя как g_K.

Для глобальных полей конечной степени мы перечислили 5 ценностей: 1) степень n; 2) дискриминант D (род g); 3) число классов h; 4) регулятор R; 5) дзета-функция. Между ними есть многочисленные соотношения, часть из которых я написал. Например, неравенство Одлыжко—Серра связывает n и D, но его доказательство проходит через работу с дзета-функцией, причем довольно хитрую. Теорема Брауэра—Зигеля, которая

связывает h и R с g, — это тоже довольно хитрая работа с дзета-функцией плюс еще некоторая работа с теорией представлений групп, чтобы обойти некоторые гипотезы.

Глобальные поля бесконечной степени

Теперь зададимся вопросом: «Что можно сказать про глобальные поля бесконечной степени?» Сначала нужно сказать, что такое глобальное поле бесконечной степени. Я буду говорить только про числовой случай. Числовое поле бесконечной степени — это очень простая вещь. Это — поле $\mathcal{K}$, промежуточное между $\mathbb{Q}$ и $\overline{\mathbb{Q}}$ и такое, что $\deg_{\mathbb{Q}} \mathcal{K} = \infty$. Что можно перенести из конечномерной ситуации в бесконечномерную? На прямую, очевидно, ничего. Степень бесконечная, дискриминант бесконечный, h тоже, вообще говоря, бесконечно; как определять R вообще непонятно, дзета-функцию тем более. Но мы с М. А. Цфасманом обнаружили, что можно немножко модифицировать эти определения так, чтобы для любого глобального поля бесконечной степени можно было построить некие аналоги почти всего этого. Сейчас я в двух словах расскажу, как это делается.

Пусть $\mathcal{K}$ — поле алгебраических чисел бесконечной степени. Его можно представить в виде объединения полей конечной степени: $\mathcal{K} = \bigcup\limits_{i=1}^{\infty} K_i$. Действительно, берем один элемент и присоединяем его, потом еще и еще. Каждое K_i — обычное поле алгебраических чисел конечной степени. Для него определены все перечисленные выше инварианты. Но они для разных полей K_i разные, и представления поля $\mathcal{K}$ в виде объединения полей K_i тоже есть разные. Тем не менее, оказывается, что для поля $\mathcal{K}$ можно построить инварианты, которые более или менее соответствуют инвариантам для конечного поля.

Сначала введем некоторые обозначения. Рассмотрим индекс α, принимающий значения в $\{\mathbb{R}, \mathbb{C}, 2, 3, 4, 5, 7, 8, 9, 11, \ldots\}$ ($\mathbb{R}$, $\mathbb{C}$ и степени простых чисел). Для числового поля K конечной степени положим

$$N_\alpha(K) = \begin{cases} r_1 & \text{при } \alpha = \mathbb{R}; \\ r_2 & \text{при } \alpha = \mathbb{C}; \\ \text{число } \alpha, \text{ таких что } |\mathcal{O}_K : \mathfrak{a}| = \alpha & \text{при } \alpha = q = p^n. \end{cases}$$

Теперь для поля $\mathcal{K}$ определим

$$\varphi_\alpha(\mathcal{K}) = \lim_{i \to \infty} \frac{N_\alpha(K_i)}{g(K_i)}, \tag{3}$$

где $g(K_i) = \log \sqrt{D_{K_i}}$.

Основное предложение заключается в том, что для любого α и для любого представления $\mathcal{K} = \bigcup_{i=1}^{\infty} K_i$ этот предел существует; более того, он не зависит от представления. Числа $\varphi_{\mathbb{R}}(\mathcal{K})$, $\varphi_{\mathbb{C}}(\mathcal{K})$, $\varphi_2(\mathcal{K})$, …, $\varphi_\alpha(\mathcal{K})$, … неотрицательные вещественные. С каждым полем $\mathcal{K}$ можно связать такой набор инвариантов. Естественный вопрос: «Какими бывают эти наборы инвариантов?» Конечно, они бывают все нулями. Примеры, для которых все эти инварианты равны нулю следующие.

1) Поле $\mathcal{K} = \overline{\mathbb{Q}}$. Это поле столь велико, что если посчитать эти инварианты, то окажется, что все пределы равны нулю: $\varphi(\mathcal{K}) = 0$.

2) Поле $\mathcal{K} = \mathbb{Q}^{ab} = \bigcup_{m=1}^{\infty} \mathbb{Q}(e^{2\pi i/m})$ — максимальное абелево расширение поля $\mathbb{Q}$. Для этого поля тоже $\varphi(\mathcal{K}) = 0$.

3) Поле $\mathcal{K} = \mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{4}, \sqrt{5}, \ldots)$, которое получается присоединением всех квадратных корней. Для этого поля тоже $\varphi(\mathcal{K}) = 0$.

Но бывают и поля, для которых $\varphi(\mathcal{K}) \neq 0$, только их довольно сложно построить. Эти поля существуют по теореме Голода—Шафаревича. Для любого поля Голода—Шафаревича одно из чисел $\varphi_{\mathbb{R}}(\mathcal{K})$ и $\varphi_{\mathbb{C}}(\mathcal{K})$ положительно. Я не буду подробно говорить, что такое поле Голода—Шафаревича; скажу только, каким свойством они обладают. Поле Голода—Шафаревича — это объединение бесконечной башни вложенных полей $K_0 \subset K_1 \subset K_2 \subset \ldots$ Например, $K_0 = \mathbb{Q}(\sqrt{D})$. Свойство полей такое: поле K_i над K_{i-1} не разветвлено, т. е.

$$\frac{g_{K_0}}{n_{K_0}} = \frac{g_{K_1}}{n_{K_1}} = \frac{g_{K_2}}{n_{K_2}} = \ldots$$

По-другому это называется так: эта башня не разветвлена. Она еще обладает многими другими свойствами: на каждом шагу расширение абелево и т. д.

Если поле K_0 мнимое квадратичное, то все остальные поля тоже мнимые, т. е. $n = r_2$. Поэтому $\varphi_{\mathbb{C}}(\mathcal{K}) = \frac{g_{K_0}}{n_{K_0}}$. Если поле K_0 вполне вещественное, то $\varphi_{\mathbb{R}}(\mathcal{K}) \neq 0$. Обобщая эту конструкцию, можно добиться того, чтобы любой конечный набор $\varphi_\alpha(\mathcal{K})$ был положительным. Но, например, можно ли добиться, чтобы положительным был бесконечный набор — открытая проблема.

Есть гипотеза, что если поле в некотором смысле маленькое среди бесконечных полей, то для него набор ненулевых чисел бесконечный. Дело в том, что здесь имеется некоторая монотонность: если $\mathcal{K} \subset \mathcal{L}$, то $\varphi_\alpha(\mathcal{K}) \leqslant \varphi_\alpha(\mathcal{L})$. Поэтому нужно рассмотреть минимальные поля, но только, конечно, неабелевы. Если поле абелево, то получится нуль. А если

рассмотреть так называемые почти конечные поля (это такие расширения, у которых любое подрасширение уже конечно), то для них гипотетически всегда будет бесконечно много ненулей. В пользу этого есть некоторые очень серьезные соображения, но ясно, что это очень трудная задача.

В определении (3) производится деление на род, а не на степень, потому что из-за ветвления бывает так, что степень маленькая, а род большой. Поэтому если делить на степень, то предел может оказаться бесконечным. Ихара делил на степень, он смог разобраться только с неразветвленным случаем, когда все равно на что делить: на род или на степень.

Числа $\varphi_\alpha(\mathcal{K})$ удовлетворяют некоторым неравенствам. Я напишу только одно из них — линейное:

$$\sum_q \frac{\varphi_q \log q}{\sqrt{q}-1} + \varphi_{\mathbb{R}} \log C_1 + \varphi_{\mathbb{C}} \log C_2 \leqslant 1.$$

Здесь C_1 и C_2 — константы, фигурирующие в неравенстве Одлыжко—Серра, так что $\varphi_{\mathbb{R}} \log C_1 + \varphi_{\mathbb{C}} \log C_2$ — в точности выражение Одлыжко—Серра. Таким образом, это неравенство является усилением неравенства Одлыжко—Серра. При некотором правильном определении дзета-функции это неравенство переписывается так: $\widetilde{\xi}(1/2) \geqslant 0$, где $\widetilde{\zeta} = (\log \tilde{\zeta}\zeta_K)'$. А если это же неравенство написать в функциональном случае, то получится некоторое усиление формулы Вейля, причем выглядит оно ровно так же: формула та же.

Теперь я расскажу, как в этом случае определяется дзета-функция и как выглядит теорема Брауэра—Зигеля. Дзета-функция определяется исключительно простой формулой:

$$\zeta(s) = \prod_q (1 - q^{-s})^{-\varphi_q}.$$

Пополненная дзета-функция определяется так:

$$\widetilde{\zeta}(s) = e^s 2^{-\varphi_{\mathbb{R}}} \pi^{-s\varphi_{\mathbb{R}}/2} (2\pi)^{-s\varphi_{\mathbb{C}}} \Gamma(s/2)^{\varphi_{\mathbb{R}}} \Gamma(s)^{\varphi_{\mathbb{C}}} \zeta(s).$$

Для всех полей из примеров 1—3 дзета-функция — это просто экспонента e^s.

Теорема Брауэра—Зигеля в этом случае записывается следующим образом:

$$\lim \frac{\log h_i R_i}{g_i} = 1 + \sum \varphi_q \log \frac{q}{q-1} - \varphi_{\mathbb{R}} \log 2 - \varphi_{\mathbb{C}} \log 2\pi = \widetilde{\zeta}_{\mathcal{K}}(1).$$

Мы извлекли корень степени g, поэтому в единице полюс устранился. Но из-за этого, к сожалению, функциональное уравнение пропадает.

Эта теорема является обобщением теоремы Брауэра—Зигеля. Действительно, теорема Брауэра—Зигеля соответствует тому, что $\varphi_{\mathbb{R}} = 0$ и $\varphi_{\mathbb{C}} = 0$. Но очень легко доказать, что если $\varphi_{\mathbb{R}} = 0$ и $\varphi_{\mathbb{C}} = 0$, то все φ_q тоже равны 0. А теорема Брауэра—Зигеля как раз утверждает, что $\lim \frac{\log h_i R_i}{g_i} = 1$ при условии, что $n/g \to 0$. Кстати сказать, долгое время (собственно говоря, до нашей работы) вообще не было известно, важно это условие или нет. Для доказательства оно было важно, а для результата было не известно, важно оно или нет. Оказалось, что важно. Если не требовать, чтобы $n/g \to 0$, то есть много разных примеров, когда предел отличен от 1. Он заключен между 0,5 и 1,1.

Эти формулы дают некоторое чувство, что начало этой теории в правильном направлении, хотя, конечно, это самые первые шаги. Никаких результатов по теории бесконечных глобальных полей я не видел. Я пытался это сравнивать с теорией комплексных кривых бесконечного рода, но связь непонятна. Она должна быть, но пока не ясно, в чем она.

29 августа 2002 г.

Р. А. Минлос

КВАНТОВАНИЕ ПО ФЕЙНМАНУ

Сначала я напомню общую схему квантования, которая принята в квантовой механике. Я думаю, что многие, если не все, это знают; тем не менее, я это напомню. Обычная схема состоит в следующем.

1) Выбирается некоторое комплексное гильбертово пространство $\mathcal{H}$. Элементы этого гильбертова пространства описывают состояния системы. Более точно, состояния системы описываются элементами $\psi \in \mathcal{H}$, для которых $\|\psi\| = 1$. Два элемента ψ_1 и ψ_2, которые отличаются числовым множителем (т. е. $\psi_1 = \alpha\psi_2$; при этом, естественно, $|\alpha| = 1$), описывают одно и то же состояние. Строго говоря, состояние описывается лучом в гильбертовом пространстве. Но так не говорят, потому что это — только усложнение языка. А говорят просто о векторах гильбертова пространства.

2) Каждой физической величине A (координате, импульсу и т. д.) сопоставляется свой самосопряженный оператор $\widehat{A}$ так, что если система находится в некотором состоянии ψ, то $(\widehat{A}\psi, \psi)$ — это среднее значение величины A в этом состоянии. Вообще в квантовой механике только изредка бывает так, что можно измерить величину точно; обычно измеряется среднее значение величины. Самый главный оператор, который описывает энергию системы, — это так называемый *гамильтониан* $\widehat{H}$. Самый главный он потому, что с его помощью вводится динамика системы. Пока то, что я говорил, это то, что физики обычно называют кинематикой — описание системы. Теперь нужно задать динамику системы, т. е. нужно задать закон эволюции системы во времени.

3) Эволюция описывается унитарной группой $U_t : \mathcal{H} \to \mathcal{H}$, которая по определению имеет такой вид: $U_t = e^{i\widehat{H}t}$. (Всякая унитарная группа имеет вид $U_t = e^{iSt}$, где S — самосопряженный оператор; так вот в качестве этого самосопряженного оператора выбирается гамильтониан $\widehat{H}$.) Отсюда легко выводится уравнение движения. Если вы имеете в момент времени $t = 0$ начальное состояние ψ_0, то состояние ψ_t в момент времени t таково: $\psi_t = e^{i\widehat{H}t}\psi_0$. Для ψ_t получаем уравнение

$$-i\frac{d\psi_t}{dt} = \widehat{H}\psi_t \quad \text{(уравнение Шрёдингера)}.$$

Если следить за размерностями входящих сюда величин, то в это уравнение нужно еще ввести множитель h, т. е. правильно уравнение записывается так: $-ih\frac{d\psi_t}{dt} = \widehat{H}\psi_t$. Величина ψ всегда считается безразмерной, $\widehat{H}$ имеет размерность энергии, поэтому величина h должна иметь размерность эрг · сек (это размерность величины, которая в механике обычно называется действием). В системе единиц «грамм, сантиметр, секунда» величина h измерена. Она равна приблизительно 10^{-27}; в такой системе единиц эта величина очень маленькая. Но в системе единиц, где квантовая механика действительно начинает действовать, в масштабах атома, там это уже действительно значительная величина. В каждой системе единиц измерения она имеет свое значение. В дальнейшем, поскольку это не будет для меня важным, я буду считать, что мы находимся в такой системе единиц, что $h = 1$.

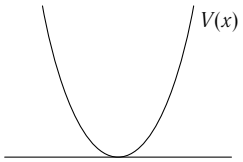
Вот я и изложил обычные основы квантовой механики. Все остальное состоит в том, что для каждой конкретной системы придумывается свое гильбертово пространство и свой оператор $\widehat{H}$. И дальше уже это изучается. Что же обычно изучается? Кроме того, что изучаются свойства этой эволюции, т. е. свойства группы $e^{it\widehat{H}}$, изучается спектр оператора $\widehat{H}$. Почему спектр так интересен? Легко показать, что если ψ_0 — собственная функция оператора $\widehat{H}$, т. е. $\widehat{H}\psi_0 = \lambda_0\psi_0$, то состояние ψ_0 является стационарным, т. е. оно не меняется во времени. Поэтому задача об отыскании всех стационарных состояний есть задача о нахождении собственных векторов и собственных значений оператора $\widehat{H}$, т. е. об изучении его спектра. Правда, оператор $\widehat{H}$ может иметь еще непрерывный спектр. Но собственные функции непрерывного спектра тоже определенным образом интерпретируются (я не буду сейчас объяснять, как именно).

Вот, собственно, и вся схема обычной квантовой механики, которая так или иначе реализуется в разных случаях. Квантование — это процедура введения гильбертова пространства и операторов в нем, и изучение свойств оператора $\widehat{H}$.

Моя цель — рассказать, как происходит квантование другим способом, который сейчас тоже весьма распространен в физике. Он, конечно, не должен противоречить первому способу квантования. Но иногда оказывается, что он приводит к другим результатам.

Я начну с квантования одной очень простой системы. Сначала я проквантую ее так, как объяснял, а потом объясню, как она квантуется по Фейнману. Эта система — частица во внешнем поле. В классическом случае ее энергия является суммой кинетической энергии и потенциальной:

$$H = \frac{mv^2}{2} + V(x), \quad x \in \mathbb{R}^1.$$



Р и с. 1. График функции $V(x)$

Потенциальная энергия задается функцией $V(x)$. В дальнейшем мы будем предполагать, что функция V имеет растущий на бесконечности график (рис. 1).

Какой оператор $\widehat{H}$ сопоставляется такому классическому случаю? Я буду сейчас объяснять, но не буду приводить никаких доказательств. (Их просто нет; весь вопрос в выборе.) Как проще всего найти оператор, соответствующий $V(x)$? Мы начнем с того, что найдем оператор, соответствующий координате: $(\widehat{x})(\psi) = x\psi(x)$. Оператор $V(x)$ — это функция координаты, поэтому оператор $\widehat{V}$ задается умножением на V: $(\widehat{V}\psi)(x) = V(x)\psi(x)$. Вас не должно смущать, что это — неограниченный оператор. В квантовой механике свободно используют неограниченные операторы. Таким образом, мы уже проквантовали потенциальную энергию. Как проквантовать кинетическую энергию? Заметим, что

$$H = \frac{mv^2}{2} + V(x) = \frac{p^2}{2m} + V(x).$$

В классической механике импульс p всегда связан с группой трансляций. Из трансляционной инвариантности следует сохранение импульса. Даже когда импульс не сохраняется, в классической механике он вводится с помощью группы трансляций. В квантовой механике поступают так же. Но каким образом? Прежде всего вводится гильбертово пространство $\mathcal{H} = L_2(\mathbb{R}^1, dx)$ (пространство функций на прямой). Если мы возьмем группу трансляций, т. е. рассмотрим оператор $U_s\psi = \psi(x - s)$, то это есть унитарная группа в этом гильбертовом пространстве. Следовательно, она представима в виде $U_s\psi = e^{i\widehat{p}s}\psi$, где $\widehat{p}$ — некоторый оператор. Его мы и назовем оператором импульса. Дифференцируя по s, легко посчитать, что $\widehat{p} = i\frac{d}{dx}$. Так вводится оператор импульса. И вообще, если какая-то классическая величина связана с какой-то группой, то соответствующий квантовый оператор всегда определяется как инфинитезимальный оператор соответствующей группы, действующей в гильбертовом пространстве. Так определяется момент количества движения и разные другие операторы.

Если $\widehat{p} = i\frac{d}{dx}$, то $\widehat{p}^2 = -\frac{d^2}{dx^2}$. Следовательно, окончательно получаем оператор Шрёдингера

$$\widehat{H} = -\frac{1}{2m}\frac{d^2}{dx^2} + V(x).$$

Для сохранения размерности тут надо было бы писать h^2, но я это не буду писать. Обычно я буду также полагать, что $m = 1$, т. е. $\widehat{H} = -\frac{1}{2}\frac{d^2}{dx^2} + V$.

Мы обычным образом проквантовали классическую систему такого вида, введя гильбертово пространство и соответствующие операторы. В нашем случае, когда мы считаем, что V растет на бесконечности, оператор $\widehat{H}$ имеет дискретный спектр $\lambda_0 < \lambda_1 < \dots$ Все собственные значения однократные. Соответствующие собственные функции $\psi_0, \psi_1, \dots$ образуют базис в нашем гильбертовом пространстве. Вот что известно про спектр оператора $\widehat{H}$. В частности, нас будут больше всего интересовать нулевое собственное значение λ_0 и нулевая собственная функция ψ_0, которая обычно называется *основным состоянием*. Основное состояние — это состояние системы с наименьшей энергией. Как правило, система больше всего времени проводит именно в этом состоянии.

Теперь я перейду к тому, как этот оператор, эту задачу, трактовал Фейнман. Он написал замечательную формулу, которая сама по себе не имеет смысла. Но это не страшно: вся наука так и развивается, что из бессмыслицы рождается некий смысл. Фейнман написал формулу для ядра оператора $e^{i\widehat{H}t}$:

$$e^{i\widehat{H}t}(q,\, q') = \text{«}C\text{»} \int\limits_{q(0)=q',\ q(t)=q} e^{i\frac{1}{2}\int\limits_0^t (\dot{q}(\tau))^2 d\tau - i\int\limits_0^t V(q(\tau))d\tau} \text{«} \prod_\tau dq(\tau) \text{»}$$

Константа C в действительности не существует, поэтому она поставлена в кавычках. Экспонента интегрируется по континуальному произведению дифференциалов; оно тоже бессмысленно, поэтому заключено в кавычки. Интеграл берется по всем траекториям, идущим от точки q' в точку q.

Тем не менее, этой формуле можно придать смысл. Сейчас я объясню, как. Эта формула и порождает всю ту науку, о которой я сейчас буду говорить. Самый простой способ объяснить, что происходит, следующий. Давайте разобьем интервал от 0 до t на маленькие кусочки длиной Δ и заменим траекторию конечнозвенной ломаной, т. е. будем говорить о значениях траектории только в конечном числе точек. Тогда это интегральное выражение естественно переписать так:

$$\int e^{\frac{i}{2}\sum\left(\frac{\Delta_i q}{\Delta}\right)^2 \Delta - i\sum V(q_i)\Delta} \prod dq_i.$$

Получается конечнократный интеграл. Правда, он, вообще говоря, абсолютно не сходится, потому что тут стоит число, по модулю равное 1. Но в силу того, что функция V растет, из-за больших осцилляций он условно сходится. И можно доказать, что для достаточно хороших (гладких) функций V этот интеграл имеет предел, который и надо понимать как континуальный интеграл, который написал Фейнман.

Чуть-чуть иначе это можно объяснить так. Забудем на время о втором слагаемом в экспоненте. Тогда оставшуюся экспоненту можно рассматривать как комплексную плотность некоторого заряда в конечномерном пространстве. И оказывается, что этот заряд, уже в бесконечномерном пространстве, сходится (при $\delta \to 0$) к некоторой комплекснозначной мере $\mathcal{F}$, которая называется *мерой Фейнмана*. И весь интеграл, уже совершенно точно, можно записать так:

$$\int\limits_{q(0)=q',\ q(t)=q} e^{-i\int\limits_0^t V(q(\tau))d\tau}\, d\mathcal{F}.$$

Эта мера, к сожалению, не является настоящей мерой. Она определена только на алгебре множеств, а не на сигма-алгебре, и не является счетно аддитивной мерой (она не имеет даже ограниченной вариации). Поэтому по ней можно интегрировать только достаточно хорошие функции.

Такая запись уже вполне законна. Правда, эта запись не очень удачна, потому что сама мера Фейнмана не очень хорошая.

Как обобщается эта конструкция для произвольной физической системы? Давайте заметим, что в экспоненте стоит следующая физическая величина:

$$\int\limits_0^t \Big(\frac{1}{2}\dot{q}^2 - V(q(\tau))\Big)d\tau.$$

В классической механике эта величина называется действием (то, что стоит под интегралом, — это функция Лагранжа, а ее интеграл по времени называется действием). Действие обозначают $S_t(q(\tau))$. Поэтому рассматриваемый интеграл можно записать так:

$$\int e^{iS_t(q(\tau))} \prod dq.$$

Именно таким образом физики сейчас этим и пользуются. Для довольно сложных систем, для которых действие имеет не такой простой вид, а другой, рассматривается обычно такой интеграл.

Работа Фейнмана относится к 1947 г. Что же произошло после этой работы? Физики, надо сказать, вначале все это не оценили. Работа Фейнмана и его подход были оценены только в середине 50-х годов. В начале 50-х годов Марк Кац, глядя на формулу Фейнмана, догадался, что так можно представить не только группу $e^{it\widehat{H}}$, но и полугруппу $e^{-t\widehat{H}}$. И это можно сделать более обоснованно. Рассмотрим тот же самый оператор $\widehat{H}$, но теперь рассмотрим $e^{-\widehat{H}t}$. Оператор $\widehat{H}$ полуограничен снизу, поэтому $e^{-\widehat{H}t}$ — ограниченный оператор. Для ядра полугруппы $e^{-\widehat{H}t}$ Марк Кац

предложил следующую формулу:

$$e^{-\widehat{H}t}(q,\,q') = C' \int e^{-\int\limits_0^t \left(\frac{1}{2}\dot{q}+V(q(\tau))\right)d\tau} \prod dq(t) =$$

$$= \int\limits_{q(0)=q',\,q(t)=q} e^{-\int\limits_0^t V(q(\tau))d\tau}\, dW. \quad (1)$$

Первое выражение снова бессмысленное. Второе выражение снова получается, когда мы трактуем выражение $\int e^{-\int\limits_0^t \left(\frac{1}{2}\dot{q}\right)d\tau}$ вместе с произведением дифференциалов как дифференциал некоторой меры. Эта мера W оказывается хорошо известной мерой Винера. Последняя запись не вызывает уже никаких нареканий; она вполне законная.

Это было предложено Марком Кацем. С тех пор математики обычно изучают континуальные интегралы именно в таком виде. Не в виде интегралов Фейнмана, которые берутся по не очень хорошей мере Фейнмана, а именно в том виде, где фигурирует мера Винера.

Опять-таки, это выражение, чтобы обобщать его на более сложные модели, условно можно записать в следующем виде:

$$\int e^{-\int\limits_0^t \left(\frac{1}{2}\dot{q}+V(q(\tau))\right)d\tau} \prod dq(t).$$

Интеграл в экспоненте — это так называемое евклидово действие $S_t^{\text{евкл}}$. Это тоже функционал от траекторий, но только знак перед $V(q(\tau))$ теперь не минус, а плюс. Евклид, конечно, ничего про это не знал, и спрашивается, почему это «евклидово»? Я в двух словах поясню. Словосочетание «евклидова теория» появилось в квантовой теории поля в 60-х годах. Квантовая теория поля обычно развертывается в четырехмерном пространстве Минковского, где метрика задается следующим образом: $x_1^2 + x_2^2 + x_3^2 - x_0^2$, где x_0 — это время. Но Швингер в свое время догадался, что можно построить некий вспомогательный объект, похожий на квантовую теорию поля, если перейти от пространства Минковского к евклидовому пространству с метрикой $x_1^2 + x_2^2 + x_3^2 + x_0^2$. Он назвал этот объект *евклидовой теорией поля*. Евклидовость, собственно, состоит в том, что мы заменили x_0 на ix_0. Оказывается, что здесь то же самое: из обычного действия в классической механике формальной заменой времени t на it мы перейдем к тому действию, которое я выписал. А теперь всякий такой переход называется переходом к евклидову объекту, по аналогии с тем, что Швингер сделал в свое время в квантовой теории поля.

Я уже объяснил, как в этом простом примере происходит квантование. Квантование состоит в том, что мы умеем записывать функцию Грина уравнения Шрёдингера с оператором $\widehat{H}$ как интеграл по винеровской мере. Теперь моя задача — рассмотреть немножко более сложную модель и поступить с ней аналогичным образом. Но прежде чем перейти к этой задаче, я хочу пойти немножко дальше в этом представлении. Винеровская мера — это мера на всех траекториях. Ее можно рассматривать поразному. Можно рассматривать ее как условную меру: фиксировать точку q' и выпускать из нее разные траектории. Тогда на всех этих траекториях эта условная винеровская мера является вероятностной мерой. Но если теперь забыть об этом условии, а рассматривать винеровскую меру на всех траекториях, то это уже не вероятностная мера, но сигма-конечная мера. Просто потому, что ее инвариантная мера в смысле распределения значений траекторий в какой-нибудь одной точке является лебеговской мерой.

Глядя на формулу (1), рассмотрим меру dM_T, для которой плотность (производная Радона—Никодима) относительно винеровской меры dW имеет следующий вид:

$$\frac{dM_T}{dW} = \frac{1}{Z_T} e^{-\int_{-T}^{T} V(q(\tau))d\tau}, \quad \text{где } Z_T = \int e^{-\int_{-T}^{T} V(q(\tau))d\tau} dW.$$

Такой интеграл при наших предположениях относительно потенциала действительно существует при любом конечном T; это можно показать. Оказывается (это тоже можно показать), что при $T \to \infty$ существует предельная мера M ($M_T \to M$), определенная на множестве всех траекторий, т. е. на пространстве $C(\mathbb{R}^1, \mathbb{R}^1)$. Для этой предельной меры M верно следующее.

1) Мера M — распределение марковского процесса. Что это значит? Я напомню, что такое марковский процесс. У нас есть мера в пространстве траекторий — в пространстве функций $q(\tau)$, заданных на всей числовой оси. Оказывается, что если я в какой-то момент времени τ_0 фиксирую всю траекторию до этого момента времени и потом рассмотрю условное распределение для всех траекторий, которые идут дальше, то распределение зависит не от всего этого фиксированного куска, а только от его последнего значения $q(\tau_0)$. Это и есть определение марковости процесса.

2) Мера M стационарная и обратимая. Стационарность означает, что если вы сдвинете все траектории по времени, то мера от этого не изменится. Обратимость означает, что если вы отразите время симметрично относительно нуля (или любого другого t), то тоже ничего не изменится.

3) Для всякого стационарного процесса можно определить стохастический оператор

$$(T_t f)(\bar{q}) = \langle f(q_t) \mid q_0 = \bar{q} \rangle.$$

Этот оператор действует на функции от q, т. е. на функции от значений траекторий в какой-то точке. Нужно взять какую-то точку t, взять значение процесса в точке t, и проинтегрировать (условно усреднить) нашу функцию $f(q)$ при условии, что в нулевой момент времени траектория равна $\bar{q}$. В случае, когда процесс марковский, стохастические операторы $\{T_t\colon t \geqslant 0\}$ образуют полугруппу. При этом, если процесс обратимый, то эта полугруппа является самосопряженной полугруппой в пространстве $L_2(\mathbb{R}^1, \mu)$, где μ — стационарная мера. Если у вас есть стационарный процесс, то можно рассмотреть распределение значений процесса в каждой точке; это уже будут меры на прямой, а в силу стационарности процесса одни и те же. Это распределение мы и обозначили μ. Оказывается, что мера μ очень просто связана с собственной функцией ψ_0, а именно, $d\mu = \psi_0^2\, dq$, т. е. плотность меры μ относительно лебеговской меры равна квадрату основного состояния. Это — важный факт, который связывает нас с исходным оператором. Второй важный факт, который связывает нас с исходным оператором, состоит в следующем. Поскольку эта полугруппа самосопряженная, она имеет самосопряженный генератор: $T_t = e^{-Lt}$, где L — самосопряженный оператор. Оператор L унитарно эквивалентен оператору $\widehat{(H - \lambda_0 E)}$, где λ_0 — наименьшее собственное значение. Изменение энергии на константу физически не имеет значения. Дело в том, что физики всегда измеряют не сами значения уровня энергии, а разности между значениями. Поэтому вычитание константы означает, что мы измеряем энергию не от какого-то выбранного нами условно значения, а от значения в нулевом состоянии; мы просто чуть-чуть меняем шкалу энергии. Так что наш оператор L — это почти то же самое, что исходный оператор $\widehat{H}$. В этом, в каком-то смысле, и сказывается преимущество и смысл нашего подхода: мы, строя некоторый процесс, можем получить оператор, который столь же хорош, как исходный оператор $\widehat{H}$. Мы получаем новый способ описания оператора $\widehat{H}$.

То же самое верно для трехмерного и для четырехмерного пространства. Все основные формулы сохраняются.

Теперь я хочу рассмотреть одну модель (модель Нельсона). Она возникла из квантовой теории поля. Эта модель описывает взаимодействие частицы со скалярным полем. Я сразу напишу классическое евклидово действие (функционал от траектории частицы q_t и траектории поля φ_t):

$$S_T^{\text{евкл}}(q_t,\, \varphi_t) = \int\limits_{-T}^{T} \left(\frac{1}{2}\dot{q}_\tau^2 + V(q_\tau)\right) d\tau + \\ + \int\limits_{-T}^{T} \int\limits_{\mathbb{R}^d} \left[\dot{\varphi}_\tau^2 + (\nabla\varphi_\tau)^2\right] dx\, d\tau + e \int\limits_{-T}^{T} \int\limits_{\mathbb{R}^d} \rho(x - q_\tau)\varphi_\tau(x) dx\, d\tau.$$

Здесь e — заряд, ρ — распределение заряда (положительная гладкая функция, быстро убывающая на бесконечности, интеграл от которой равен 1); первое, второе и третье слагаемое — это $S^{\text{евкл}}_{\text{частицы}}$, $S^{\text{евкл}}_{\text{поля}}$ и $S^{\text{евкл}}_{\text{взаимодействия}}$.

Важно подчеркнуть, что здесь рассматривается безмассовое поле. Обычно второе слагаемое записывают в виде

$$\int_0^t \int_{\mathbb{R}^d} [\dot{\varphi}_\tau^2 + (\nabla \varphi_\tau)^2 + m\varphi_\tau^2]\, dx\, dz,$$

с дополнительным членом $m^2\varphi_\tau^2$. Здесь m обычно называют массой кванта поля. При такой добавке все значительно упрощается. Сложность состоит как раз в том, что тут этой добавки нет.

Мы должны это проквантовать по-нашему. Хочу сразу заметить, что существует вполне определенный оператор энергии, который задается с помощью операторов вторичного квантования в фоковском пространстве и который физики обычно и связывают с этой моделью. Я не буду говорить о нем ничего, потому что это завело бы нас очень далеко. В дальнейшем его форма будет представлена немножко в другом виде.

Следуя нашему рецепту, нужно сначала написать бессмысленную формулу в стиле Фейнмана для некоторой меры P:

$$dP = \frac{1}{Z} e^{-S_{\text{част}}(q_\tau)} e^{-S_{\text{поля}}(\varphi_\tau)} e^{-S_{\text{взаим}}(\varphi_\tau, q_\tau)} \prod_\tau dq_\tau \prod_{\tau, x} d\varphi_\tau(x).$$

У нас уже есть способ придавать этому смысл. Во-первых, мы уже знаем, что значит

$$e^{-S_{\text{част}}(q_\tau)} \prod_\tau dq_\tau.$$

Мы уже построили такую меру для частиц. Правда, мы ее строили в одномерном случае, а тут d-мерный случай. Но это делается аналогично. Так мы построим меру M в $C(\mathbb{R}^1, \mathbb{R}^d)$ — пространстве кривых в $\mathbb{R}^d$. Это есть марковский процесс, и все его свойства, о которых я говорил, остаются верными.

Это первый шаг. Теперь мы хотим придать смысл выражению

$$e^{-S_{\text{поля}}(\varphi_\tau)} \prod_{\tau, x} d\varphi_\tau(x).$$

Что такое евклидово действие? Евклидово действие — это некая квадратичная форма. Поэтому e в степени квадратичная форма подсказывает, что эта мера должна быть гауссовой мерой. Давайте я сейчас напомню, что такое гауссова мера.

Сначала я определю гауссову меру в конечномерном пространстве $\mathbb{R}^N$ с точками $X = (x_1, \ldots, x_N)$. *Гауссова мера* — это мера с плотностью

$$g(X) = Ce^{-\frac{1}{2}\sum a_{ij}(x_i - b_i)(x_j - b_j)}$$

относительно лебеговской меры. Вектор $b = (b_1, \ldots, b_n)$ — это среднее значение X. Действительно, $\int x_i\, g(X)dX = b_i$. Матрица $\mathcal{A} = (a_{ij})$ положительно определенная. Она определяется следующим образом. Если рассмотреть обратную матрицу $\mathcal{D} = \mathcal{A}^{-1} = (d_{ij})$, то ее элементы d_{ij} получаются как интегралы

$$d_{ij} = \int (x_i - b_i)(x_j - b_j)\, g(X)dX.$$

Обычно величины d_{ij} называют *ковариациями*, а матрицу $\mathcal{D}$ называют *матрицей ковариаций*. Гауссова мера полностью задается указанием среднего и ковариаций: если вы знаете матрицу $\mathcal{D}$, то вы знаете и матрицу $\mathcal{A}$.

Это все, что нужно знать о гауссовой мере в конечномерном пространстве. Что такое гауссова мера в произвольном линейном топологическом пространстве? Пусть B — линейное топологическое пространство, а G — мера в нем. Что значит, что мера G гауссова? Это значит следующее. Если вы возьмете любой конечный набор непрерывных линейных функционалов $F_1, \ldots, F_k$ над этим пространством, то, поскольку они заданы на пространстве с мерой, у них есть совместное распределение значений. Это уже будет мера в конечномерном пространстве. Условие такое: эта мера должна быть гауссовой для любого набора функционалов. В этом случае среднее b определяется тем, что для любого линейного функционала F имеет место равенство $\int\limits_B F(\xi)dG = F(b)$ (среднее значение любого функционала по G равно его значению на элементе b). Матрица ковариаций определяется аналогичным образом:

$$\int\limits_B F_1(\xi - b)F_2(\xi - b)dG = \operatorname{cov}(F_1, F_2).$$

Здесь cov — билинейная форма относительно этих линейных функционалов; она называется ковариацией. Оказывается, что ковариация (билинейная форма в сопряженном пространстве) и вектор b в исходном пространстве полностью определяют гауссову меру, т. е. это все, что нужно знать про гауссову меру.

Теперь вернемся снова к нашему конкретному случаю. Что в нашем случае является матрицей $\mathcal{A}$? Нужно записать квадратичную форму чуть

иначе. Я запишу ее в таком виде (после интегрирования по частям):

$$\iint \Big(-\frac{d^2}{d\tau^2} - \delta\Big)\varphi \cdot \varphi \, d\tau \, dx.$$

Роль матрицы $\mathcal{A}$ играет этот оператор. Но матрица ковариаций $\mathcal{D}$ — это обратная матрица.

Меру мы будем задавать в пространстве обобщенных функций от $d+1$ переменных (к d пространственным переменной добавляется еще одна переменная — время). В качестве B мы выберем $S'(\mathbb{R}^{d+1})$. Сопряженным пространством $B' = S(\mathbb{R}^{d+1})$ будет служить просто пространство Шварца гладких функций. Ковариация двух элементов пространства Шварца φ_1 и φ_2 это есть не что иное, как квадратичная форма:

$$S(\varphi_1, \varphi_2) = \Big(\Big(-\frac{d^2}{d\tau^2} - \nabla\Big)\varphi_1, \varphi_2\Big) = \int \frac{\widetilde{\varphi}_1(k_0, k)\widetilde{\varphi}_2(k_0, k)}{k_0^2 + k^2} \, dk_0 \, dk.$$

Второе равенство получается применением преобразования Фурье ($\widetilde{\varphi}$ — это преобразование Фурье функции φ, а k — это вектор в d-мерном пространстве). Если $d \geqslant 3$, то этот интеграл сходится; он имеет особенность в нуле, но эта особенность интегрируемая. Мы получаем непрерывную квадратичную форму в пространстве Шварца. Отсюда следует, что такая гауссова мера в пространстве обобщенных функций $S'(\mathbb{R}^{d+1})$ действительно существует, а среднее для нее равно нулю: $b = 0$. Я здесь уже перешел к пределу: я рассматриваю интеграл $\int\limits_{-\infty}^{\infty}$, а не интеграл $\int\limits_{-T}^{T}$. Это уже мера на всех обобщенных функциях. Но оказывается, что про эту меру можно сказать даже больше. Эта мера сосредоточена на более узком пространстве, чем все обобщенные функции. Она сосредоточена на пространстве непрерывных кривых $C(\mathbb{R}^1, S'(\mathbb{R}^d)) \subset S'(\mathbb{R}^{d+1})$. Эти обобщенные функции от времени зависят непрерывно.

Мы уже справились с этой задачей и ввели вторую меру. Теперь я готов объяснить, каков же тот физический оператор энергии, о котором я упомянул, и как его в этих терминах задать. Мы имеем пространство кривых со значениями в пространстве обобщенных функций и гауссову меру G. Оказывается, что возникающее распределение задает марковский процесс. Он снова обратимый и стационарный, и для него есть стационарная мера γ (мы будем обозначать большими латинскими буквами меру в пространстве траекторий, а соответствующими маленькими греческими буквами будем обозначать соответствующую стационарную меру). До этого у нас уже был другой процесс, связанный с траекториями $C(\mathbb{R}^1, \mathbb{R}^d)$.

Тут на пространстве траекторий была своя мера M и соответствующая стационарная мера μ. Мы можем построить гильбертово пространство

$$L_2(\mathbb{R}^d \times S'(\mathbb{R}^d),\, \mu \times \gamma) = L_2(\mathbb{R}^d,\, \mu) \otimes L_2(S'(\mathbb{R}^d),\, \gamma).$$

Поскольку стационарные процессы есть в одном и в другом пространстве, то можно говорить о едином стационарном процессе в пространстве траекторий частицы и поля. Их распределением будет $M \times G$, а стационарной мерой будет как раз мера $\mu \times \gamma$.

Помимо того, поскольку мы знаем, что тут есть стационарный обратимый процесс, здесь имеется генератор этого процесса. В пространстве $L_2(\mathbb{R}^d,\, \mu)$ действует оператор $L_{\text{част}}$, а в пространстве $L_2(S'(\mathbb{R}^d),\, \gamma)$ действует оператор $L_{\text{поля}}$ (как генератор соответствующего марковского гауссова процесса). Оказывается, что тот физический оператор энергии, о котором я говорил, имеет такой вид:

$$H_{\text{физ}} = L_{\text{част}} \otimes E_{\text{поля}} + E_{\text{част}} \otimes L_{\text{поля}} + \widehat{A}(q,\, \varphi),$$

где $E_{\text{поля}}$ и $E_{\text{част}}$ — единичные операторы (по полю или по частице), а $\widehat{A}(q,\, \varphi)$ — оператор умножения на функцию

$$A(q,\, \varphi) = e \int\limits_{-\infty}^{\infty} \int\limits_{\mathbb{R}^d} \rho(x - q_\tau)\varphi_\tau(x)dx\, d\tau.$$

Мы еще не дошли до основной своей цели. Обозначим уже построенную нами меру $M \times G$ через P_0. Это для нас как бы свободная мера; она описывает движение частицы и поля независимо друг от друга. А основная наша цель — построить меру P, которая уже связана со взаимодействием поля и частицы. Как это нужно сделать? Нужно опять задать, как и полагалось, меру P_T, для которой

$$\frac{dP_T}{dP_0} = \frac{1}{Z_T} e^{-e \int\limits_{-T}^{T} \int\limits_{\mathbb{R}^d} \rho(x - q_\tau)\varphi_\tau(x)dx\, d\tau}.$$

В предыдущем примере роль меры P_0 играла мера Винера. В данном случае она отличается от меры Винера. Это — вероятностная мера; не сигма-конечная, а конечная.

Теперь, по аналогии с тем, что делали раньше, построим такую меру для конечного T (она существует, как нетрудно показать). Оказывается, что верна следующая теорема, доказательство которой непросто.

Т е о р е м а. *Мера P_T сходится к некоторой мере P при $T \to \infty$.*

Мера P — это как раз та мера, которая нам нужна. Эта мера снова определяет марковский процесс. Соответствующую стационарную меру мы обозначим π.

По поводу размерности d я требую, чтобы она была не меньше 3; для 1-мерного и 2-мерного случая эта конструкция не проходит.

У нас есть первоначальное гильбертово пространство

$$L_2(\mathbb{R}^d \times S'(\mathbb{R}^d)) \text{ с мерой } \mu \times \gamma = \pi_0$$

и новое гильбертово пространство на том же самом пространстве, но с другой мерой:

$$L_2(\mathbb{R}^d \times S'(\mathbb{R}^d)) \text{ с мерой } \pi.$$

В первом пространстве есть оператор $H_{\text{физ}}$. Во втором пространстве тоже есть оператор: поскольку это марковский процесс, то у него есть генератор, и возникает новый оператор, который я обозначу $H_{\text{евкл}}$. Этот оператор ничем не хуже оператора $H_{\text{физ}}$. Возникли два оператора, которые претендуют на описание энергии этой системы. Спрашивается: когда эти операторы фактически эквивалентны, а когда не эквивалентны? Ответ такой: при $d \geqslant 4$ эти операторы унитарно эквивалентны (с точностью до сдвига на константу), а при $d = 3$ — нет. Это основной результат работы, о которой я рассказываю.

Я немножко поясню, в чем тут дело, хотя доказательство, конечно, очень сложное. У оператора $H_{\text{евкл}} = L$, поскольку он совпадает с генератором процесса, всегда есть основное состояние. У стохастической полугруппы единица сохраняется: $T_t \mathbf{1} = \mathbf{1}$. А раз единица сохраняется, то генератор превращает единицу в нуль: $L\mathbf{1} = 0$. То есть $\mathbf{1}$ — это всегда собственный вектор для генератора L с собственным значением 0. А так как оператор L положителен, то это есть основное состояние. Можно доказать, что в данном случае это собственное значение однократно. Итак, у оператора $H_{\text{евкл}}$ основное состояние всегда есть. Оказывается, что у оператора $H_{\text{физ}}$ при $d = 3$ основного состояния нет. Его спектр начинается с какой-то точки. Для оператора $H_{\text{евкл}}$ спектр начинается с точки 0 и не имеет щели. Но это уже более тонкий результат. В случае же оператора $H_{\text{физ}}$ дело обстоит так, что есть какая-то точка, с которой начинается спектр, но сама эта точка не является собственным значением. Для оператора $H_{\text{евкл}}$ точка, с которой начинается спектр, является собственным значением, а для оператора $H_{\text{физ}}$ — нет.

Как это доказывается? Дело в том, что если бы у оператора $H_{\text{физ}}$ существовало основное состояние ψ, то можно было бы показать, что $\psi^2 = \frac{d\pi}{d\pi_0}$ (производная Радона—Никодима меры π по мере π_0). В случае, когда основное состояние есть, это всегда так. И наоборот. Таким образом, если эти меры абсолютно непрерывны относительно друг друга, то основное состояние всегда есть. Если же основного состояния нет, то они друг

относительно друга сингулярны. И здесь прямо в лоб доказывается, что эти две меры (в одном и том же пространстве) сингулярны друг относительно друга в случае размерности 3. Следовательно, основного состояния быть не может, а поэтому эти операторы не унитарно эквивалентны.

Результат неожиданный, потому что всегда физики считали, что оба способа описания системы (способ непосредственного задания оператора и способ построения марковского процесса) приводят с точностью до сдвига к унитарно эквивалентным гамильтонианам. Этот пример показывает, что это не всегда так.

12 сентября 2002 г.

Г. Л. Литвинов

ДЕКВАНТОВАНИЕ МАТЕМАТИКИ И ВВЕДЕНИЕ В ИДЕМПОТЕНТНЫЙ АНАЛИЗ

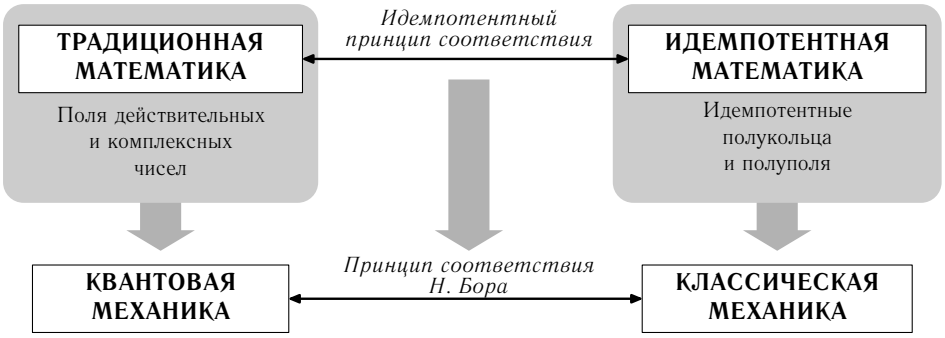
Предлагаемый текст основан на слегка отредактированной и обновленной стенограмме доклада, представленного 10 октября 2002 г. В частности, обновлен список литературы. Неформальный стиль изложения по возможности сохранен. Автор выражает горячую благодарность А. Н. Соболевскому за помощь и компьютерные версии рисунков.

Введение

Я хочу описать некоторый новый взгляд на несколько разделов математической науки. С нашей *) точки зрения обычная традиционная математика над числовыми полями — это наука квантовая. И у этой квантовой науки есть классический аналог. Этот классический аналог называется сейчас *идемпотентная математика* или *тропическая математика*. Обычно в последнее время в математической науке слово *квантование* понимается в пиквикском смысле. Если есть какая-то деформация, есть какие-то формальные аналогии, то тогда это называется квантованием. Квантование, о котором я хочу рассказать, связано с самым настоящим квантованием (из квантовой физики). Общая схема примерно такая. Традиционная математика над числовыми полями соответствует вовсе не классической механике, а квантовой механике. Есть еще одна математика, где роль полей играют идемпотентные полуполя и полукольца, о которых я вскоре буду говорить. Эта математика соответствует классической механике. Как известно, есть принцип соответствия Нильса Бора о том, что квантовая механика должна в пределе давать классическую механику. Оказывается, что аналогичный принцип соответствия действует между традиционной математикой и идемпотентной математикой (см. рис. 1).

Если в традиционной математике что-то особо интересно, то есть очень хорошие шансы, что в параллельной математике это тоже будет очень интересно [7]—[9].

*) Я имею в виду некоторую неформальную группу в Москве, которая концентрируется вокруг Виктора Павловича Маслова. Я буду рассказывать в основном о том, что возникло в работах участников этой группы.



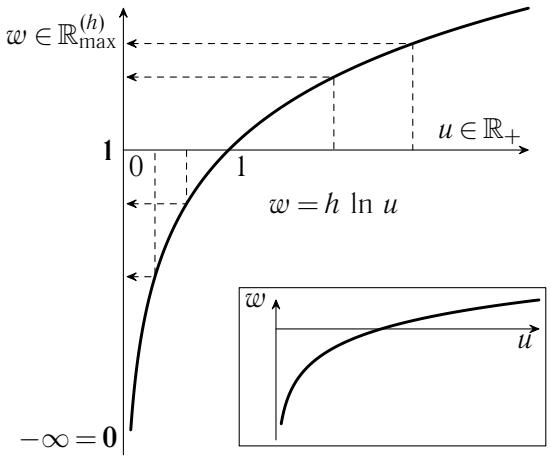
Р и с. 1. Принципы соответствия в идемпотентной математике и физике

Вот пример, который нужно постоянно держать в голове. Мы используем обычные числа, но базовые операции мы выбираем по-другому. Самый важный пример — это так называемая *алгебра Max Plus*. Эта алгебра состоит из обычных действительных чисел и символа $-\infty$, который играет в ней роль нуля, в то время как обычный нуль играет роль единицы: $\mathbf{0} = -\infty$ и $\mathbf{1} = 0$. Операции в этой алгебре следующие: $a \oplus b = \max\{a, b\}$ (новое сложение) и $a \odot b = a + b$ (новое умножение). Идемпотентных полуполей и алгебр — огромное количество, но этот пример самый важный. Именно он связан с физикой.

Аналогично есть и алгебра Min Plus, когда максимум заменяется на минимум. В этой алгебре $\mathbf{0} = +\infty$ и $\mathbf{1} = 0$. Операции в ней такие: $a \oplus b = \min\{a, b\}$ и $a \odot b = a + b$.

Условие идемпотентности состоит в том, что $x \oplus x = x$. Я не могу сказать, что название — идемпотентная математика (или тропическая математика) — удачное; так сложилось исторически. Более правильное название придумал один из предтеч этой науки, а именно ленинградский математик Н. Н. Воробьев, специалист по теории игр и математической экономике. Он предвидел появление такой науки, и назвал ее экстремальной математикой. Но его идеи, как и идеи многих других предтеч, не были замечены.

Посмотрим, как можно получить алгебру Max Plus (часто обозначаемую через $\mathbb{R}_{\max}$) в результате естественной деформации. Давайте вместо поля всех вещественных чисел (или даже поля всех комплексных чисел) рассмотрим множество $\mathbb{R}_+$ неотрицательных чисел и проделаем следующую замену переменных: $u \mapsto w = h \ln u$ (рис. 2). При этом полуось $\mathbb{R}_+$ отображается на ось всех вещественных чисел, 0 переходит в $-\infty$. Полуось $\mathbb{R}_+$ — полуполе (это означает, что там есть все свойства поля за одним единственным исключением: отсутствует вычитание). Перенесем операции умножения и сложения из $\mathbb{R}_+$ на всю вещественную ось с помощью этого логарифмического преобразования. Ясно, что новое



Р и с. 2. Замена переменных: на вставке — случай малого значения h

умножение будет сложением: $u_1 \cdot u_2 \mapsto w_1 \odot w_2 = w_1 + w_2$, а новое сложение будет выражаться следующей формулой:

$$u_1 + u_2 \mapsto w_1 \underset{h}{\oplus} w_2 = h \ln(e^{w_1/h} + e^{w_2/h}).$$

Ясно, что при различных h все эти полуполя изоморфны. У нас есть семейство полуполей, зависящее от h. Теперь, как обычно, устремим h к нулю справа. Тогда сложение перейдет в операцию взятия максимума из двух чисел: $w_1 \underset{h}{\oplus} w_2 \to \max\{w_1, w_2\}$ при $h \to +0$. Этот параметр $h > 0$ и есть наша постоянная Планка. Впервые логарифмическое преобразование в интересующем нас контексте появилось в 20-е годы в работах Эрвина Шрёдингера, который изучал переход от квантовой механики к классической. У него это преобразование было, но при этом вместо h он использовал выражение $i\hbar$, где $\hbar$ — обычная постоянная Планка. Если постоянная Планка вещественная, то никакого перехода к алгебре Max Plus получится не может, никаких пределов не будет. Это — упражнение для студентов первого курса. А если заставить постоянную Планка принимать чисто мнимые значение, то тогда оказывается, что неотрицательные числа перейдут в довольно любопытную алгебру, которая и является нашим основным объектом.

Какой главный рецепт нашего деквантования? Надо постоянную Планка сделать чисто мнимой, а затем устремить ее к нулю. Давайте эту общую идею запомним для будущего. Указанную процедуру принято называть *деквантованием Маслова*. Занимался такими деформациями и Эбергард Хопф в начале 50-х годов. У него был метод исчезающей вязкости, и эта исчезающая вязкость как раз и была нашей постоянной Планка, однако алгебра Max Plus в явном виде не обсуждалась.

Понятно, что соответствующий переход от поля вещественных или комплексных чисел к алгебре Max Plus осуществляется при помощи деквантования Маслова и отображения $x \mapsto |x|$. Такой переход мы также назовем деквантованием Маслова.

Любопытно, что здесь можно сделать еще и вторичное деквантование. Если процедуру деквантования Маслова проделать второй раз (опять выделить положительную полуось и опять прологарифмировать), то мы получим алгебру Min-Max, когда сложение — это минимум, а умножение — это максимум (или наоборот). То, что получится в результате, уже не будет полуполем, а будет только полукольцом. Оно не такое важное, но в приложениях оно тоже выплывает.

Идемпотентные полукольца и полуполя

Что же такое полукольца и полуполя? Это — алгебры с операциями сложения и умножения, которые удовлетворяют стандартным аксиомам. Общая идея такова: есть все, кроме вычитания. Обе операции ассоциативны, есть две дистрибутивности (левая и правая). Сложение у нас коммутативно всегда. Если умножение коммутативно, то объект называется коммутативным. Самое главное — условие идемпотентности. Если это условие выполнено, то полукольцо (или полуполе) называется *идемпотентным*. Идемпотентные полукольца с нулем и единицей (а именно такие объекты мы и будем обсуждать) иногда называют *тропическими алгебрами*. В последнее время тропической алгеброй чаще называют алгебру Max Plus или изоморфную ей алгебру Min Plus.

Полукольцо является *полуполем*, если оно коммутативно и любой ненулевой элемент обратим по отношению к умножению. Но поскольку, например, в алгебре Max Plus умножение — это сложение, а у сложения есть вычитание, то ясно, что алгебра Max Plus будет полуполем. Вообще есть такой экспериментальный факт. Ассоциативная операция является особенно интересной и хорошей в двух случаях: либо когда она обратима, либо когда она идемпотентна. По крайней мере, других интересных случаев я не знаю.

Что будет, если обе операции обратимы? Тогда будет обычное поле; это будет традиционная математика. Если обе операции идемпотентны, то эта область тоже очень хорошо изучена: это булевы алгебры, теория решеток. Словом, читайте книгу Г. Биркгофа «Теория решеток». А вот случай, когда одна операция является идемпотентной, а другая обратимой, т. е. случай полуполя, был пропущен. А он оказывается очень интересным. В этом месте в науке была большая дырка.

По поводу классификации идемпотентных полуполей известны некоторые результаты. Идемпотентных полуполей очень много. Если вы сложите обычные поля (прямая сумма), то поля не получится. А если правильно складывать идемпотентные полуполя, то будет снова идемпотентное полуполе. По поводу классификации полуколец есть разные результаты, но, например, уже классификация конечных идемпотентных полуколец — дело совершенно безнадежное. Хотя это очень интересно с точки зрения математической логики. Но зато есть только одно конечное идемпотентное полуполе: это алгебра Буля (состоящая из двух элементов). Михаил Александрович Шубин описал все конечные коммутативные полукольца, у которых 3 или 4 элемента, см. [21].

По поводу истории вопроса и терминологии см., например, [7].

Канонический порядок и алгебраические конструкции

Отметим, что наша теория исключительно положительная (или хотя бы неотрицательная). Дело в том, что если у нас есть идемпотентное полукольцо или даже идемпотентная полугруппа, то возникает *канонический частичный порядок*. По определению $a \preccurlyeq b$ тогда и только тогда, когда $a \oplus b = b$. Опять же, в голове нужно держать алгебру Max Plus; в этом случае канонический порядок совпадает с обычным порядком на вещественной оси. В общем случае $a \oplus b$ — точная верхняя грань множества из двух элементов. Естественно, эта точная верхняя грань может не совпадать ни с a, ни с b. Ясно, что $a \succcurlyeq \mathbf{0}$ для любого элемента a, так как $\mathbf{0} \oplus a = a$. Сложение фактически порождается порядком: $a \oplus b = \sup\{a, b\}$. Так что пример с максимумом в этом смысле очень типичный.

Есть еще одна чрезвычайно важная процедура. Если у нас есть произвольное частично упорядоченное множество, то его можно пополнить. Это делается с помощью дедекиндовых сечений, в точности так же, как строится пополнение множества рациональных чисел. Подробности можно найти в книге Г. Биркгофа. Поэтому возникают *полные полукольца*. Правда, от полного полукольца требуется, чтобы его операции были совместимы с этим пополнением. Но об этом — чуть позже.

И есть еще одна чрезвычайно важная операция — операция «звездочка», или вычисление *квазиобратного элемента* (называемая также *операцией замыкания*). Это — сумма геометрической прогрессии:

$$a \mapsto a^* = \mathbf{1} \oplus a \oplus a^2 \oplus \ldots = \sup\{\mathbf{1}, a, a^2, \ldots\}.$$

Ясно, что если наше полукольцо полное, то результат всегда найдется. Это что-то вроде $(1-a)^{-1}$, т. е. что-то вроде резольвенты. Вообще-то,

резольвента важнее, чем обратный элемент, и в идемпотентной науке это ясно видно.

Кроме того, полезно еще рассматривать *ограниченное пополнение* и *ограниченно полные полукольца*. В таких полукольцах существуют суммы любых ограниченных множеств.

Имеются стандартные конструкции прямого произведения (суммы) полуколец. Это — обычное прямое произведение и *правильное* (или *тропическое*) прямое произведение. Обычное прямое произведение K_1 и K_2 — это декартово произведение $K_1 \times K_2$ с поточечными операциями. Получается полукольцо. А если вы сложите два полуполя, то полуполя не получите. Правильное прямое произведение идемпотентных полуполей K_1 и K_2 определяется так:

$$(K_1 \setminus \{\mathbf{0}\}) \times (K_2 \setminus \{\mathbf{0}\}) \cup \{\mathbf{0}\}.$$

Тогда получается, что тропическое прямое произведение любых двух полуполей снова есть полуполе.

Можно брать непрерывные прямые произведения. Например, если у вас есть более или менее произвольное локально выпуклое пространство функций над $\mathbb{R}$, то это — полуполе. На функциях есть частичный порядок (одна функция меньше или равна другой) и есть сложение (обычное сложение функций). Это сложение будет новым умножением, а новое сложение порождено порядком.

Все стандартные банаховы решетки — это тоже примеры непрерывных сумм алгебры Max Plus. Например, пространства L^p. Это — интересные функциональные пространства (и одновременно полуполя) в идемпотентной науке.

Кроме того, если есть коммутативное полукольцо K, то всегда можно обычным образом построить полукольцо $\mathrm{Mat}_n(K)$ матриц размера $n \times n$ с элементами из K. Это будет хороший пример некоммутативного идемпотентного полукольца.

Новые примеры

Теперь есть смысл рассказать несколько примеров.

Пример 1. $K = [a, b]$ с операциями $\oplus = \max$ и $\odot = \min$. Тогда $\mathbf{0} = a$ и $\mathbf{1} = b$. Это — полное полукольцо с линейным порядком $\leqslant$.

Пример 2. $K = \mathbb{R}_+$ с операциями $\oplus = \max$ и $\odot = \cdot$ (обычное умножение). Тогда $\mathbf{0} = 0$ и $\mathbf{1} = 1$. Получаем идемпотентное полукольцо. Но если его прологарифмировать, то мы увидим, что ничего особенно нового не получим.

Надо сказать, что любая комбинация максимума (или минимума) с арифметическими операциями порождает числовую идемпотентную алгебру.

Наши операции являются аналогами булевых операций. Именно так они у многих предтеч и возникали. И связано это опять же с квантовой теорией и квантовой логикой. Ясно, что любая дистрибутивная решетка — это некоторое идемпотентное полукольцо относительно операций взятия точных нижних и верхних граней подмножеств, состоящих из двух элементов.

Пример 3. Булева алгебра: $\vee = \oplus = \max$, $\odot = \min = \wedge$ и $K = \{\mathbf{0}, \mathbf{1}\}$.

Пример 4. $K = \{\mathbf{0}, \mathbf{1}, a\}$ — идемпотентное кольцо из трех элементов. Здесь элемент a играет роль «бесконечности», так что $a \oplus a = a$, $a \odot a = a$, $\mathbf{0} \odot a = \mathbf{0}$, $\mathbf{1} \odot a = a$, $\mathbf{0} \oplus a = \mathbf{1} \oplus a = a$.

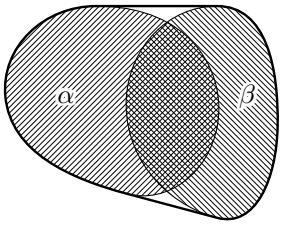


Рис. 3. Полукольцо выпуклых компактов. Сплошной линией показан результат вычисления суммы $\alpha \oplus \beta$

Это — нетрадиционная логика (многозначная логика). Вообще надо сказать, что логики внесли достойный вклад в предмет.

Очень интересен еще такой пример (он у нас в разных видах еще всплывет). Обычно он используется в математической экономике, в теории эволюции множеств Парето.

Пример 5. Объект этого полукольца — компактное выпуклое подмножество (обычно в арифметическом пространстве). Нуль — это пустое множество. Единица — множество, состоящее из нуля. Операции такие: $\alpha \oplus \beta$ — это выпуклая оболочка α и β (см. рис. 3), а умножение — это сложение по Минковскому:

$$\alpha \odot \beta = \{a + b : a \in \alpha,\ b \in \beta\}.$$

Это очень важное полукольцо. Оно работает и в алгебраической геометрии. Здесь, конечно, порядок не линейный.

Пример 6. Если R — коммутативное кольцо, то пространство его двусторонних идеалов — идемпотентное полукольцо. Сумма идеалов состоит из попарных сумм их элементов, а произведение идеалов состоит из сумм всевозможных произведений.

Еще один пример, который мы все знаем из средней школы, но мы не знали, что говорим прозой. Оказывается, что натуральные числа — это тоже замечательное полукольцо.

Пример 7. Натуральные числа 1, 2, 3, ... со следующими операциями: $a \oplus b$ — наименьшее общее кратное a и b, а $a \odot b$ — наибольший

общий делитель a и b (или наоборот). Тогда все аксиомы полукольца выполняются: можно раскрывать скобки и т. п. Здесь даже больше правил, потому что не одна дистрибутивность, а две. Чтобы это полукольцо имело нуль и единицу, нужно добавить еще один элемент.

Пожалуй, первый интересный пример идемпотентного полукольца возник в 1956 г. в работе знаменитого логика Стивена Клини. Он же заложил основы идемпотентной линейной алгебры. Там полукольцо такое.

Пример 8. Элемент полукольца — язык над фиксированным конечным алфавитом. Язык — это просто словарь. Операции следующие:

$$A \odot B = \{x : x = yz,\ y \in A,\ z \in B\}, \quad A \oplus B = A \cup B.$$

Оказывается, что это полукольцо имеет довольно важное значение в компьютерных языках; оно активно применяется и сейчас.

Еще один важный пример возник в квантовой теории измерения. Причем работы на эту тему крайне малодоступны (во всех смыслах). Но по крайней мере определения известны.

Пример 9. Если есть C^*-алгебра, то оказывается, что множество замкнутых правых идеалов в ней образует важную решетку. И эта решетка является идемпотентным полукольцом относительно операций $I \odot J = \overline{I \cdot J}$ и $I \oplus J = \overline{I + J}$ (замыкание берется в топологии C^*-алгебры, по норме). Оказывается, что это полукольцо коммутативно тогда и только тогда, когда исходная C^*-алгебра коммутативна. Двусторонние идеалы тоже образуют полукольцо, но оно не такой интересный объект.

Объекты такого рода называются *кванталями* благодаря своему происхождению. Но на самом деле кванталь — это просто полное идемпотентное полукольцо.

Полные идемпотентные полукольца

Пусть K — идемпотентное полукольцо с каноническим частичным порядком. Если X — подмножество в K, то через $\bigoplus X$ обозначим точную верхнюю грань этого подмножества, если она существует, т. е. $\bigoplus X = \sup X = \sup_{x \in X}\{x\}$. Полукольцо K назовем *a-полным* (или *алгебраически полным*), если K полно как частично упорядоченное множество (то есть любое подмножество в K имеет точные верхнюю и нижнюю грани) и справедливы «бесконечные» дистрибутивные законы:

$$k \odot (\bigoplus X) = \bigoplus (k \odot X), \qquad (\bigoplus X) \odot k = \bigoplus (x \odot k)$$

для любых элементов $k \in K$ и подмножеств $X \subset K$.

Как уже было сказано, кванталы — это в точности a-полные идемпотентные полукольца.

Полукольцо K является *b-полным* (или *ограниченно полным*), если любое ограниченное подмножество X в K имеет точную верхнюю грань $\bigoplus X$ и «бесконечные» дистрибутивные законы справедливы для любых элементов $k \in K$ и ограниченных подмножеств $X \subset K$.

Разумеется, любое a-полное идемпотентное полукольцо является и b-полным. Полукольца, описанные в примерах 1, 3, 4, 7—9, являются a-полными. Алгебра Max Plus, обозначаемая также через $\mathbb{R}_{\max}$, является b-полной. Если к $\mathbb{R}_{\max}$ добавить элемент $+\infty$ и доопределить операции очевидным образом, то возникает a-полное идемпотентное полукольцо $\widehat{\mathbb{R}}_{\max} = \mathbb{R}_{\max} \cup \{+\infty\}$. Однако $\widehat{\mathbb{R}}_{\max}$ не является полуполем. Более того, любое a-полное идемпотентное полуполе совпадает с алгеброй Буля (пример 3). С другой стороны, идемпотентное полуполе будет b-полным (как полукольцо), если оно ограниченно полно как частично упорядоченное множество.

Принцип соответствия и идемпотентный анализ

Теперь давайте посмотрим на идемпотентный принцип соответствия [8], [9]. Принцип этот такой. Если в классической теории есть какие-то интересные конструкции, понятия, теоремы и т. д., то есть очень хорошие шансы, что им в идемпотентной математике соответствует тоже что-то очень интересное. А переход такой. Вместо поля мы пишем какое-нибудь идемпотентное полукольцо и смотрим, что получится.

Первый шаг в этом направлении сделал Виктор Павлович Маслов, когда попробовал построить теорию идемпотентного интегрирования, см., например [6], [16]—[18]. Обычный интеграл $\int\limits_a^b f(x)\,dx$ — это предел интегральных сумм $\sum f(x_i)\sigma_i$. По принципу соответствия давайте посмотрим, что получится, если обычные операции заменить на операции из алгебры Max Plus. Тогда вместо суммы появится максимум:

$$\max\{f(x_i) \odot \sigma_i\} = \max\{f(x_i) + \sigma_i\},$$

причем $\sigma_i \to 0$. В результате получим, что идемпотентная версия интеграла — это точная верхняя грань функции:

$$\int\limits_X^{\oplus} f(x)\,dx = \sup_{x \in X}\{f(x)\}.$$

Нужно, чтобы эта точная верхняя грань существовала. А для этого полукольцо должно быть a-полным или, если функция ограниченная, то b-полным.

Замечательно, что это определение работает для любых идемпотентных полуколец (и любого пространства или множества X), потому что в любом полукольце есть порядок и можно брать точную верхнюю грань.

Если мы знаем, что такое интеграл, то можно определить еще и *меру Маслова*. Это — линейный (в смысле идемпотентной математики) функционал на множестве «интегрируемых функций», который определяется как следующее «скалярное произведение»:

$$\varphi \mapsto m_\psi(\varphi) = \int\limits_X^{\oplus} \varphi(x) \odot \psi(x)\, dx.$$

Здесь интеграл нужно понимать как идемпотентный интеграл. Впоследствии были доказаны абстрактные теоремы, что никаких других линейных функционалов с разумными свойствами не бывает. Таким образом, скалярное произведение задает типичную меру Маслова.

Эта мера, конечно, глубоко положительна. Поэтому несколько французских математиков сообразили, что есть сильная аналогия с теорией вероятностей, с вероятностными мерами. Тогда (принцип соответствия!) оказалось, что масловские меры соответствуют вероятностным мерам, уравнение Беллмана соответствует уравнению Чепмена—Колмогорова, а принцип Беллмана соответствует принципу причинности Маркова в стохастических цепях Маркова, см., например [2]. В результате идеи и методы теории вероятности удается перенести в теорию оптимизации.

Теперь посмотрим, как обстоят дела с гармоническим анализом. Если есть группа G, то можно попробовать построить групповое кольцо (или полукольцо). Давайте, например, в качестве такого полукольца возьмем пространство всех ограниченных функций $B(G, K)$. Тогда свертка определяется следующим способом:

$$(\varphi \circledast \psi)(x) = \int\limits_G^{\oplus} \varphi(y) \odot \psi(y^{-1}x)\, dx.$$

Так мы получаем групповое полукольцо.

Общий принцип гармонического анализа состоит в том, что какую групповую алгебру вы выберете, такой и возникнет вариант гармонического анализа. Здесь — одна из групповых алгебр и один из вариантов гармонического анализа. Естественно возникает вопрос, нельзя ли построить что-нибудь в духе преобразования Фурье так, чтобы эта свертка при преобразовании Фурье переходила в обычное произведение. Такое преобразование построить можно. Посмотрим, что такое идемпотентное

преобразование Фурье. Обычное преобразование Фурье—Лапласа определяется следующим образом:

$$f \mapsto \widehat{f}(\xi) = \int\limits_X e^{i\langle x,\xi\rangle} f(x)\,dx,$$

где $X = \mathbb{R}^n$, $x \in X$, а ξ — линейный функционал $x \mapsto \langle x,\, \xi\rangle$ на X. При этом самую важную роль играет экспонента, то есть характер группы X. Чтобы обобщить преобразование Фурье, скажем, на алгебру Max Plus, нам нужно понять, что такое характер. Характер — это решение функционального уравнения $f(x+y) = f(x)f(y)$. Правильный аналог этого уравнения такой: $f(x+y) = f(x) \odot f(y) = f(x) + f(y)$. Разумные решения такого функционального уравнения — линейные функции. Вообще, линейные и кусочно линейные функции играют в идемпотентном анализе очень важную роль. Стало быть, соответствующее тропическое преобразование Фурье

$$f \mapsto \widehat{f}(\xi) = \int\limits_X^{\oplus} \langle x,\, \xi\rangle \odot f(x)\,dx = \sup_x\{\xi \cdot x + f(x)\}$$

— это просто преобразование Лежандра с точностью до косметических изменений. А преобразование Лежандра очень важное, поскольку оно в классической механике отвечает за перевод лагранжевой картины в картину Гамильтона и наоборот. Оказывается, что преобразование Лежандра — это не что иное, как идемпотентная версия преобразования Фурье. В результате преобразование Лежандра становится более понятным. Это замечательное наблюдение принадлежит Виктору Павловичу Маслову.

В идемпотентной математике можно строить всевозможные пространства функций. Вообще, изначально идемпотентная математика называлась идемпотентным анализом. В таком анализе мы интересуемся пространствами функций, которые принимают значения в идемпотентных полуполях или идемпотентных полукольцах. А то пространство, на котором определены наши функции, его геометрическую структуру, мы не трогаем. Но это все-таки только часть идемпотентной математики. В последнее время Олег Виро очень удачно тронул и геометрическую структуру, см. [22].

В идемпотентном анализе возникает большое количество всевозможных функций, функциональных пространств и линейных операторов. Но функции возникают, как правило, негладкие. Это ясно, потому что операции максимум и минимум гладкость нарушают. Поэтому, вообще-то, идемпотентный анализ — это анализ негладких функций и негладких решений дифференциальных уравнений.

Примерами пространств функций являются: пространство всех функций, пространство всех ограниченных функций, пространство всех полунепрерывных (снизу или сверху) функций на топологическом пространстве, а также пространства выпуклых и вогнутых функций на линейных пространствах. Эти примеры играют важную роль в идемпотентном анализе. Вообще, в идемпотентной науке особенно хорошие пространства — это ядерные пространства. И ядерных пространств здесь гораздо больше, чем в обычном анализе. В обычном анализе ядерных пространств довольно мало, но все они обладают знаменитым свойством аппроксимации Гротендика. В идемпотентной науке все проще: свойство аппроксимации и ядерность — это в точности одно и то же. Подробности см. в [15].

Теперь нам ясно, что такое интеграл, что такое функциональное пространство (можно сформулировать и понятие абстрактного линейного пространства в идемпотентном анализе). Самое главное в этом анализе, что можно брать суммы неограниченного семейства слагаемых, потому что на самом деле это просто точная верхняя грань этого семейства: $\bigoplus x_\nu = \sup\{x_\nu\}$. Если наше пространство (полукольцо) полное, то это имеет смысл. Линейность (над идемпотентным полукольцом) в этой науке понимается как линейность относительно бесконечного числа слагаемых. Оказывается, что такая обобщенная линейность в точности соответствует полунепрерывности. Таким образом, можно строить чисто алгебраический вариант анализа, где аналогом непрерывности является следующее равенство: $f(\bigoplus x_\nu) = \bigoplus f(x_\nu)$. Здесь f может быть функцией, оператором и т. д. Операции называются a-линейными, если здесь суммы любые, и b-линейными, если суммы ограниченные.

Ясное дело, что по принципу соответствия всякий интегральный оператор A имеет следующий вид:

$$A f(x) = \int_X^{\oplus} K(x, y) \odot f(y)\, dy = \sup_{y \in X} \{K(x, y) + f(y)\}$$

(если мы переходим к алгебре Max Plus). Обычный интеграл заменяется супремумом, а произведение заменяется на сумму. Но это выражение совершенно стандартно в теории оптимизации и часто там используется. Вообще, идемпотентный анализ — это, кроме всего прочего, еще и общая, унифицирующая, форма теории оптимизации.

Оказывается, что любой хороший линейный оператор (хороший в указанном смысле) в хорошем пространстве (т. е. в ядерном) является интегральным. То есть справедливы аналоги теоремы Лорана Шварца о ядре. Более того, построен алгебраический идемпотентный анализ, который

включает основные теоремы типа Хана— Банаха, теорию топологических тензорных произведений, теорию ядерных пространств и ядерных операторов и, в частности, весьма общую теорему о ядре, которая и была центральным пунктом теории Гротендика. Подробности см. в [11], [12], [15].

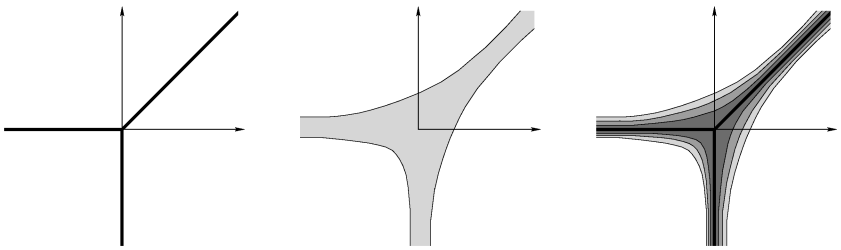
Тропические полиномы и деквантование геометрии

Прежде чем переходить к физике, давайте посмотрим, что будет тропическим аналогом полинома. Если у нас есть моном $x_1^{n_1}\dots x_k^{n_k}$, то он перейдет в функцию $n_1x_1+\dots+n_kx_k$. Это — линейная функция (в обычном смысле). Полином — это сумма мономов; здесь, стало быть, возникнет максимум. Так возникает кусочно линейная функция, которая, вдобавок, выпуклая. Это наблюдение лежит в основе метода Олега Виро, который заметил, что метод, которым он пользовался много лет назад, на самом деле связан с квантованием и деквантованием [22]. И он говорил прозой. Если деквантовать вещественную алгебраическую геометрию, то возникнут задачи из теории многогранников. Поскольку многогранники с нужными свойствами построить легко, а построить соответствующие алгебраические многообразия — трудная проблема, то Виро строил сначала многогранники, потом проводил процедуру квантования и получал то, что ему было нужно. Причем использовалось именно то квантование, о котором мы сейчас говорим.

Так из полиномов появляются выпуклые кусочно линейные функции. А если, кроме того, учесть, как выпуклая функция получается из своих касательных, то она тоже в некотором смысле кусочно линейная.

Итак, идемпотентная версия вещественной алгебраической геометрии была открыта Виро и представлена в его докладе на конгрессе в Барселоне [22]. Исходя из идемпотентного принципа соответствия, Виро построил кусочно-линейную геометрию многогранников специального вида в конечномерных евклидовых пространствах как результат деквантования Маслова обычной вещественной алгебраической геометрии. Он также указал на важные приложения (например, в рамках 16-й проблемы Гильберта о построении вещественного алгебраического многообразия с предписанными свойствами) и на связь с комплексной алгебраической геометрией и амебами в смысле И. М. Гельфанда, М. М. Капранова и А. В. Зелевинского (см. их книгу [3] и статьи Виро [22], [23]). Затем комплексная алгебраическая геометрия была деквантована Г. Михалкиным (с тем же результатом), который указал новые важные приложения, см. доклад Михалкина на семинаре «Глобус» 10 ноября 2005 г. и его доклад на конгрессе в Мадриде

[20]. Эта новая «идемпотентная» (или асимптотическая) геометрия теперь обычно называется *тропической алгебраической геометрией*.



Р и с. 4. «Деквантование» амебы

Грубо говоря, деформация алгебраического многообразия при конечных значениях «постоянной Планка» h и является амебой в смысле [3]. В пределе при $h \to 0$ амеба переходит в *тропическое алгебраическое многообразие* (скелет амебы). Например, для прямой $V = \{(x, y) \in \mathbb{C}^2 \colon x + y + 1 = 0\}$ соответствующая тропическая прямая $\mathrm{Tro}(V)$ представлена на рис. 4(a). Амеба представлена на рис. 4(b), а соответствующая деформация амебы — на рис. 4(c).

Давайте посмотрим, бывают ли какие-то связи между обычными кольцами и идемпотентными. Допустим, у нас есть полином $\mathcal{P}(x)$ от одной переменной с неотрицательными коэффициентами. Сопоставим ему его степень $\deg \mathcal{P}$. Я утверждаю, что отображение $\mathcal{P}(x) \mapsto \deg \mathcal{P}$ является гомоморфизмом в тропическое полукольцо Max Plus.

А что будет, если мы возьмем полином $\mathcal{P}(x_1, \ldots, x_p)$ от нескольких переменных? Аналогом степени в этом случае является многогранник Ньютона. Многогранник Ньютона — это выпуклый многогранник. Он является частью того примера, который я указывал (см. пример 5). Я утверждаю, что отображение полинома в свой многогранник Ньютона — гомоморфизм относительно операций выпуклой оболочки двух многогранников Ньютона и их суммы по Минковскому.

Опишем этот гомоморфизм более подробно. Для функции f, определенной на $\mathbb{R}^n_+ = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \colon x_i \geqslant 0$ для всех $i = 1, \ldots, n\}$ или на области в $\mathbb{C}^n$, включающей $\mathbb{R}^n_+$, определим функцию $\widehat{f}$ формулой

$$\widehat{f}(x) = \lim_{h \to 0} h \ln |f(\exp(x/h))|,$$

где $h > 0$, а $x \in \mathbb{R}^n$.

Назовем функцию $\widehat{f}(x)$ *деквантованием* функции $f(x)$, а преобразование $f(x) \mapsto \widehat{f}(x)$ — *преобразованием деквантования* [14]. Разумеется, это преобразование порождено деквантованием Маслова. Справедлив следующий результат.

Если функция f является полиномом, то субдифференциал $\partial\widehat{f}$ функции $\widehat{f}$ в начале координат совпадает с многогранником Ньютона полинома f. Для полукольца полиномов с неотрицательными коэффициентами преобразование $f \mapsto \partial\widehat{f}$ является гомоморфизмом этого полукольца в полукольцо выпуклых многогранников относительно операций Минковского (см. выше, пример 5).

Используя преобразование деквантования, можно обобщить этот результат на широкий класс функций и выпуклых множеств; подробности см. в [14], [7]. Разумеется, все это связано с тропической геометрией.

Связь с физикой и принцип суперпозиции

Давайте теперь разберемся с физикой. Связано ли это как-нибудь с настоящей физикой? Вспомним, что такое вообще квантование и деквантование. Если у нас есть классическая гамильтонова функция

$$H = H(p,\, x) = \sum_{i=1}^{n} \frac{p_i^2}{2m_i} + V(x)$$

(координаты здесь обобщенные, $x = (x_1, \ldots, x_N)$, $p = (p_1, \ldots, p_N)$), то, чтобы получить соответствующую квантовую систему, т. е. уравнение Шрёдингера, можно написать уравнение вида

$$-\frac{\hbar}{i}\frac{\partial\psi}{\partial t} = \widehat{H}\psi,$$

где $\widehat{H} = \widehat{H}(\widehat{p}_i,\, \widehat{x}_i)$. Оператор энергии $\widehat{H}$ получается следующим образом: вместо координаты x_i нужно поставить оператор умножения на эту координату, а вместо импульса p_i — оператор $\frac{\hbar}{i}\frac{\partial}{\partial x_i}$. Тогда мы получим уравнение Шрёдингера.

Есть важные классические функции. Во-первых, функция Лагранжа

$$L = \sum_{i=1}^{N} \frac{p_i^2}{2m} - V(x).$$

Во-вторых, знаменитая функция действия

$$S(x,\, t) = \int_{t_0}^{t} L(x,\, \dot{x},\, t)dt,$$

где интеграл берется вдоль траектории системы.

Процедура у нас такая. Мы делаем постоянную Планка чисто мнимой: $h = i\hbar$. Тогда мы приходим к уравнению

$$h\frac{\partial u}{\partial t} = H\left(-h\frac{\partial}{\partial x_i}, \widehat{x}_i\right) u,$$

которое является обобщенным уравнением теплопроводности. При этом вещественнозначная функция u соответствует волновой функции (точнее, ее модулю, что имеет очевидный физический смысл).

Как известно, в квантовой теории действует *принцип суперпозиции*: если есть два состояния системы (два решения уравнения Шрёдингера), то их сумма снова является состоянием. Это ясно, потому что уравнение Шрёдингера линейное: сумма решений снова будет решением. А уравнение теплопроводности отличается от него постоянными множителями. Оно тоже линейное. Стало быть, если u_1 и u_2 — решения уравнения, то их линейная комбинация $u = \lambda_1 u_1 + \lambda_2 u_2$ также является решением этого уравнения.

Теперь давайте сделаем нашу стандартную замену переменной, т. е. логарифмическое преобразование, которое делал еще Шрёдингер: $S = \mp h \ln u$. Только постоянная Планка будет у нас чисто мнимой; мы устремим ее к нулю. В зависимости от знака у нас появится алгебра Max Plus или Min Plus; обе эти алгебры годятся. Если мы все это аккуратно подставим, то получим уравнение

$$-\frac{\partial S}{\partial t} = V(x) + \sum_{i=1}^{N} \frac{1}{2m_i}\left(\frac{\partial S}{\partial x_i}\right)^2 \mp h \sum_{i=1}^{N} \frac{1}{2m_i}\frac{\partial^2 S}{\partial x_i^2}.$$

Это по-прежнему уравнение теплопроводности, только в новых координатах.

Таким образом, хотя наше уравнение нелинейно в обычном смысле, оно линейно с точностью до замены переменной. Если S_1 и S_2 — его решения, то функция

$$S = \lambda_1 \odot S_1 \underset{h}{\oplus} \lambda_2 \odot S_2,$$

полученная заменой переменных $S = h \ln u$, также является решением уравнения. В этом случае обобщенное умножение $\odot$ совпадает с обычным сложением, а обобщенное сложение $\underset{h}{\oplus}$ определяется выбранной заменой переменной (см. выше, Введение).

Теперь давайте устремим постоянную Планка к нулю. Тогда последний член в уравнении пропадает. В результате мы получим уравнение Гамильтона—Якоби, т. е. основное уравнение классической механики

$$\frac{\partial S}{\partial t} + V(x) + \sum_{i=1}^{N} \frac{1}{2m_i}\left(\frac{\partial S}{\partial x_i}\right)^2 = 0 \quad \text{или} \quad \frac{\partial S}{\partial t} + H\left(\frac{\partial S}{\partial x_i}, x_i\right) = 0.$$

Ясно, что эти операции (т. е. $\odot$ и $\underset{h}{\oplus}$) перейдут в операции в алгебре Max Plus или Min Plus. Поэтому естественно считать, что если S_1 и S_2 — решения уравнения Гамильтона—Якоби, то и их линейная комбинация

$$S = \lambda_1 \odot S_1 \oplus \lambda_2 \odot S_2$$

снова является решением этого уравнения. Получается, что уравнение Гамильтона—Якоби линейно над алгебрами Max Plus и Min Plus. Получается, что основное уравнение классической механики, которое в обычном смысле нелинейно, линейно в идемпотентной теории. Поэтому к этому уравнению можно применять все линейные трюки: дельта-функции, функции Грина и т. п. В результате возникает *принцип суперпозиции Маслова*, который он впервые сформулировал [16], [17]. Он, правда, начинал не с уравнения Шрёдингера, а с обычного уравнения теплопроводности.

Оказывается, кстати, что дифференциальное уравнение Беллмана очень похоже на уравнение Гамильтона—Якоби и для него также справедлив принцип суперпозиции. Это не удивительно, потому что принцип суперпозиции связан с вариационными принципами механики и экстремальными задачами.

Общий принцип суперпозиции состоит в том, что *многие важные задачи* (*особенно в теории оптимизации*), *нелинейные в обычном смысле*, *являются линейными над подходящими идемпотентными полукольцами*.

Что же такое вариационные принципы механики (принцип наименьшего действия)? Принцип наименьшего действия говорит, что действие должно быть минимальным на траекториях движения нашей системы. А минимум — это масловский интеграл в алгебре Min Plus:

$$\min S(x,\, t) = \int\limits^{\oplus}_{T \in \{\text{пространство траекторий}\}} S(x,\, t)\, dT.$$

То есть это как раз интеграл по траекториям. Более точно, если с этим поразбираться, то выяснится, что принцип наименьшего действия сводится к тому, что интеграл $\int\limits^{\oplus} e^{-S} dT$ соответствует максимальному значению величины e^{-S}. А это в точности означает, что действие S минимально. Таким образом, оказывается, что вариационные принципы механики — это всего-навсего идемпотентный вариант фейнмановского подхода к квантовой механике через интегралы по траекториям. И это тесно связано с формулой Лакса—Олейник и рядом других изысканий в этой области.

З а м е ч а н и е. Я вас, конечно, обманываю, потому что сумма решений — это уже не гладкая функция. Это же идемпотентная сумма. Если

мы будем брать максимумы, да еще в бесконечном количестве, то, конечно, функция будет негладкой. В идемпотентном анализе можно грамотно определить, что такое негладкое решение уравнения этого типа. Для этого строится идемпотентная версия теории обобщенных функций, и соответствующее обобщенное решение и будет правильно сформулированным понятием решения уравнения Гамильтона—Якоби. То, что они линейны (в идемпотентном смысле), означает следующее. Существует оператор эволюции (как бы интегральный оператор), который является линейным и непрерывным и переводит пространство обобщенных решений в себя.

Матричные уравнения Беллмана и их приложения

Исходя из принципа суперпозиции ясно, что самые важные задачи — это линейные задачи. Давайте посмотрим, как решать системы уравнений в этой идемпотентной науке. Вообще-то, первым до этого додумался тот, кто придумал метод последовательных приближений. Давайте рассмотрим систему линейных уравнений $\mathcal{A}X = B$. Если это система над каким-то полем, то ее всегда можно преобразовать к виду $X = AX + B$, где $\mathcal{A} = I - A$, а I — единичная матрица. Давайте решать уравнение $X = AX + B$ методом последовательных приближений: $X_0 = 0$, $X_1 = B$ и т. д. В конечном итоге находим решение

$$X = \lim(I + A + A^2 + \ldots)B = A^*B.$$

Здесь A^* — сумма геометрической прогрессии. У A^* есть одно фундаментальное свойство:

$$A^* = I + AA^* = I + A^*A.$$

Если нужно сильно аксиоматизировать ситуацию, это свойство можно взять в качестве аксиомы. Получается универсальный метод, который годится для любых полуколец. Если только удастся «звездочку» определить (а в идемпотентном случае ее можно определить, потому что это — верхняя грань в полном полукольце; см. выше, раздел 3), то мы имеем совершенно универсальное решение системы линейных уравнений.

Эта система линейных уравнений, т. е.

$$X = AX + B,$$

называется *матричным уравнением Беллмана*.

Это — конечномерная версия стационарных уравнений Беллмана. Сейчас мы увидим, как она связана с оптимизацией на графах. Вы понимаете, что любая задача приближается конечномерной; нелинейная задача

приближается линейной. Поэтому самое главное — разобраться с конечномерными линейными задачами. Теперь мы знаем ответ, причем в явном виде.

Я забыл сказать, что операция «звездочка» во всех конкретных «элементарных» идемпотентных алгебрах исключительно проста. Например, если $x \prec \mathbf{1}$, то $x^* = x$, а если $x \succ \mathbf{1}$, то $x^* = \infty$ или $x^* = \mathbf{0}$. Это очень хорошее упражнение, посмотреть в каждом полукольце, что такое «звездочка». Легко убедиться, что вычислительная трудность операции «звездочка» равна нулю. Эта операция исключительно удачно работает. Можно было бы прямо указать алгоритм вычисления матрицы A^*.

Решение системы не единственно, но этот метод дает нам минимальное решение. Это именно то решение, которое нужно в приложениях.

Первым человеком, который написал формулу A^*B, был, видимо, Стивен Клини [5]. Он написал ее для своего идемпотентного полукольца, состоящего из языков над фиксированным алфавитом. Он первым научился решать эти системы линейных уравнений. И его метод оказался очень мощным. Но первым человеком, который действительно разобрался в ситуации, был Б. Карре (B. Carré, 1971), см. [1]. Он загадочный человек. Это француз, который жил в Англии и там печатался, и даже его соавторы не знают, что с ним сейчас, куда он делся. Он первый человек, который понял, что все стандартные оптимизационные задачи на графах сводятся к решению матричных уравнений Беллмана для различных идемпотентных полуколец. По дороге была работа примерно 1961 г. у Н. Н. Воробьёва, который решал немножко другие задачи, но проявил очень большую проницательность.

Карре первым заметил, что различные оптимизационные алгоритмы на самом деле соответствуют различным методам решения систем линейных уравнений. Он заметил, например, что алгоритм Беллмана для проблемы кратчайшего пути — это идемпотентный метод Якоби, а алгоритм Форда — это идемпотентный итерационный алгоритм Гаусса—Зайделя (в полном согласии с принципом соответствия).

Это был следующий чрезвычайно важный шаг, и мы сейчас попробуем это разобрать. Что здесь важно? Во-первых, как мы видели на примере решения системы уравнений методом последовательных приближений, методы абсолютно универсальные. Они годятся для любых полуколец. И это важно с точки зрения, например, программирования. Имеется специальная техника программирования, которая называется техникой *абстрактных типов данных* и позволяет одним алгоритмом решать большое количество задач, см., например, [8], [9], [13]. А во-вторых, важно, что алгоритмы — линейные (над полукольцами). А значит, все

основные методы распараллеливания линейных алгоритмов проходят. От того, что мы поняли принцип суперпозиции, и осознали, что задача является линейной над полукольцами, мы сразу получили новый алгоритм.

Наш основной принцип, принцип идемпотентного соответствия, справедлив для алгоритмов и даже для компьютерных устройств (процессоров). Потому что, собственно говоря, алгоритм, программа и компьютерный процессор — это, в сущности, одно и то же. Это просто разные стадии реализации алгоритма: на бумаге, в программе, в памяти машины или в железе. Из этого, казалось бы, вполне невинного принципа соответствия можно делать выводы большой практической силы.

Сейчас я попробую объяснить, почему некоторые важные оптимизационные задачи являются линейными над подходящими полукольцами. Эти оптимизационные задачи формулируются на так называемых взвешенных направленных графах. У нас есть точки и дуги. При этом у дуги есть направление: она идет из одной точки в другую. Термин «взвешенный» означает, что каждой дуге приписан вес; обычно это вещественное число. Оказывается, что если у нас есть какое-нибудь числовое идемпотентное полукольцо, то мы всегда сможем построить с помощью такого графа матрицу. Матрица строится так: $A = (a_{ij})$, где $a_{ij} = \mathbf{0}$, если нет дуги $i \to j$, а если дуга $i \to j$ есть, то a_{ij} — вес дуги $i \to j$. Ясно, что у нас получается взаимно однозначное соответствие между взвешенными графами и матрицами над идемпотентным полукольцом, лишь бы оно было числовым.

Давайте сначала рассмотрим задачу из динамического программирования. Это некая версия задачи о коммивояжере. У нас есть граф. Предположим, что дуги — это какие-то пути в сельской глубинке. Разъезжает коммивояжер, и от каждой поездки по маршруту из точки i в точку j он получает какой-то доход a_{ij}. Это и есть вес соответствующей дуги. Кроме того, когда он покидает эту область в точке i_n, то получает терминальный приз f_{i_n}, который может быть положительным (откупные от местных купцов) или отрицательным (налог от местной мафии). Основная проблема состоит в том, чтобы получить максимальный доход в результате этого путешествия. Это — типичная задача динамического программирования.

Общий доход M выражается формулой

$$M = a_{i_1 i_2} + a_{i_2 i_3} + \ldots + a_{i_{n-1} i_n} + f_{i_n}.$$

С другой стороны, если мы будем считать, что веса — это элементы алгебры Max Plus, то M есть следующее выражение:

$$a_{i_1 i_2} \odot a_{i_2 i_3} \odot \ldots \odot a_{i_{n-1} i_n} \odot f_{i_n}.$$

Давайте сначала посчитаем максимум дохода, если сделано ровно n шагов,

начиная с точки i. Легко доказать, что тогда $\max M = (A^n f)_i$ (i-я координата). Мы уже видим, что это — линейная задача. Это прямо следует из этой формулы. А если мы не хотим ограничивать число шагов, то нужно сложить нулевую попытку, попытку, когда он проехал один раз и на этом закончил, и т. д. А это и будут суммы геометрической прогрессии. Следовательно, окончательный ответ такой: максимум равен $(A^* f)_i$. Поэтому максимум — это решение соответствующей системы линейных уравнений над алгеброй Max Plus.

Многие другие оптимизационные задачи на графах (например, задача о минимальном пути и задача о пути наибольшей ширины) сводятся к решению матричных уравнений Беллмана над подходящими идемпотентными полукольцами. Подробности см., например, в [1], [6], [8], [9], [18].

Краткий обзор идемпотентной/тропической математики в целом и дополнительную библиографию можно найти, например, в [7]. Современное состояние предмета хорошо представлено в статьях из сборника [10].

Литература

[1] *Carré B. A.* An algebra for network routing problems // J. Inst. Appl. 1971. V. 7. P. 273—294.

[2] *Del Moral P.* A survey of Maslov optimization theory // *Kolokoltsov V. N., Maslov V. P.* Idempotent Analysis and Applications. Dordrecht: Kluwer Acad. Publ., 1997. P. 243—302 (Appendix).

[3] *Gelfand I. M., Kapranov M. M., Zelevinsky A.* Discriminants, resultants, and multidimensional determinants. Boston: Birkhäuser, 1994.

[4] Idempotency / Ed. J. Gunawardena // Publ. of the Newton Institute. V. 11. Cambridge: Cambridge University Press, 1998.

[5] *Kleene S. C.* Representation of events in nerve sets and finite automata // Automata Studies / Ed. J. McCarthy, C. Shannon. Princeton: Princeton University Press, 1956. P. 3—40.

[6] *Kolokoltsov V. N., Maslov V. P.* Idempotent Analysis and Applications. Dordrecht: Kluwer Acad. Publ., 1997.

[7] *Литвинов Г. Л.* Деквантование Маслова, идемпотентная и тропическая математика: краткое введение // Записки научных семинаров ПОМИ. 2005. Т. 326. С. 145—182. E-print arXiv:math.GM/0507014.

[8] *Litvinov G. L., Maslov V. P.* Correspondence principle for idempotent calculus and some computer applications, (IHES/M/95/33). Bures-sur-Yvette: Institut des Hautes Etudes Scientifiques, 1995. E-print arXiv:math.GM/0101021.

[9] *Litvinov G. L., Maslov V. P.* The correspondence principle for idempotent calculus and some computer applications // In [4]. P. 420—443.

[10] *Litvinov G. L., Maslov V. P.* (Eds.) Idempotent mathematics and mathematical physics. Providence, RI: AMS, 2005. (Contemporary Mathematics.; V. 377).

[11] *Литвинов Г. Л., Маслов В. П., Шпиз Г. Б.* Тензорные произведения идемпотентных полумодулей. Алгебраический подход // Матем. заметки. 1999. Т. 65, № 4. С. 572—585. E-print arXiv:math.FA/0101153.

"all4" — 2009/5/27 — 11:23 — page 53 — #53

[12] *Литвинов Г. Л.*, *Маслов В. П.*, *Шпиз Г. Б.* Идемпотентный функциональный анализ. Алгебраический подход // Матем. заметки. 2001. Т. 69, № 5. С. 758—797. E-print arXiv:math.FA/0009128.

[13] *Литвинов Г. Л.*, *Маслова Е. В.* Универсальные численные алгоритмы и их программная реализация // Программирование. 2000. № 5. С. 53—62. E-print arXiv:math.NA/0102114.

[14] *Litvinov G. L.*, *Shpiz G. B.* The dequantization transform and generalized Newton polytopes // In [10]. P. 181—186.

[15] *Litvinov G. L.*, *Shpiz G. B.* Kernel theorems and nuclearity in idempotent mathematics: an algebraic approach // Journal of Mathematical Sciences. 2006. V. 139, № 3. E-print arXiv:math.FA/0609033.

[16] *Маслов В. П.* О новом принципе суперпозиции для оптимизационных задач // Успехи матем. наук. 1987. Т. 42. № 3 С. 39—48.

[17] *Маслов В. П.* Операторные методы. М.: Наука, 1973.

[18] *Маслов В. П.*, *Колокольцов В. Н.* Идемпотентный анализ и его применение в оптимальном управлении. М: Наука, 1994.

[19] *Maslov V. P.*, *Samborskii S. N.* (*Eds.*) Idempotent Analysis. Providence, RI: AMS, 1992. (Adv. in Soviet Math.; V. 13).

[20] *Mikhalkin G.* Tropical geometry and its applications // Proceedings of the ICM. Madrid, 2006. E-print arXiv:math.AG/0601041.

[21] *Shubin M. A.* Algebraic remarks on idempotent semirings and the kernel theorem in spaces of bounded functions // In [19]. P. 151—166.

[22] *Viro O.* Dequantization of real algebraic geometry on a logarithmic paper // In: 3rd European Congress of Mathematics. Barcelona, 2000. E-print arXiv:math/0005163.

[23] *Viro O.* What is an amoeba? 2002. P. 916—917. (Notices of the Amer. Math, Soc.; V. 49).

10 октября 2002 г.

М. В. Финкельберг

КОМПАКТИФИКАЦИЯ УЛЕНБЕК И АФФИННЫЕ АЛГЕБРЫ ЛИ
(по работе А. Бравермана и Д. Гайцгори)

Само пространство модулей G-расслоений и его компактификация возникли из математической физики, поэтому я очень кратко напомню, откуда они возникли, но в дальнейшем это никак использоваться не будет. У нас есть четырехмерное пространство $\mathbb{R}^4$, а также компактная группа Ли K; почти все время у нас будет $K = \mathrm{SU}(n)$. Изучается K-связность A в K-расслоении. Для $K = \mathrm{SU}(n)$ это будет просто эрмитова связность в n-мерном комплексном расслоении с унитарной структурной группой. У этой связности A есть кривизна F_A, которая является 2-формой с коэффициентами в алгебре Ли $\mathfrak{su}(n)$ косоэрмитовых матриц. Уравнение Янга—Миллса (Yang—Mills) на эту связность гласит, что кривизна антиавтодуальна: $F_A = - * F_A$; оператор $*$ — это «звездочка» Ходжа. Кроме того, требуется, как говорят, *конечность энергии*, т. е. требуется, чтобы интеграл $\int\limits_{\mathbb{R}^4} |F_A|^2$ был конечен. В таком случае известно, что этот интеграл принимает дискретное множество значений, а именно, $\frac{1}{8\pi^2} \int\limits_{\mathbb{R}^4} |F_A|^2 = k \in \mathbb{N}$. Это число k называют *инстантонным числом* или *топологическим зарядом*, а еще можно сказать, что это — *второй класс Черна*. Нужно только сказать, второй класс Черна чего именно. Пространство $\mathbb{R}^4$ вложено в сферу S^4. Наше расслоение можно продолжить в бесконечно удаленную точку. А если есть расслоение на S^4, то у него есть топологический инвариант — второй класс Черна. Это элемент группы $H^4(S^4; \mathbb{Z}) \cong \mathbb{Z}$, т. е. целое число.

Связностей очень много (так же, как функций на четырехмерном пространстве со значениями в алгебре Ли). Решений уравнения антиавтодуальности тоже много, но зато на них действует очень большая группа калибровочных преобразований (функций со значениями в группе, которые сопрягают связности). Существенно различные решение — это те, которые не переводятся друг в друга никакой заменой калибровки. Это — пространство модулей решений $\{\mathrm{YM}\}^k/\mathbb{C}^\infty(\mathbb{R}^4, K)$. Это уже получается конечномерное многообразие, которое любят изучать в четырехмерной топологии. Его часто называют *многообразием инстантонов* на $\mathbb{R}^4$. Тут важно, что мы фиксировали топологический заряд $k = c_2$.

Дональдсон (S. Donaldson) изучал инстантоны на 4-мерных многообразиях. Он даже филдсовскую премию получил за изучение этих пространств модулей. А кроме того, он открыл, что это связано с алгебраической геометрией следующим образом. Главное его наблюдение состоит в том, что $\mathbb{R}^4 = \mathbb{C}^2$. Поэтому все 2-формы имеют комплексный тип: сколько там dz и сколько $d\bar{z}$. В частности, кривизна разлагается в следующую сумму: $F_A = F_A^{0,2} + F_A^{1,1} + F_A^{2,0}$. Тогда уравнение антиавтодуальности переписывается очень просто. Оно эквивалентно тому, что $F_A^{0,2} = F_A^{2,0} = 0$ и $F_A^{1,1} \wedge \omega = 0$, где $\omega = dz_1 \wedge d\bar{z}_1 + dz_2 \wedge d\bar{z}_2$ — стандартная кэлерова форма на $\mathbb{C}^2$. Итак, уравнение Янга—Миллса можно переписать в такой комплексной форме. Что это означает? Саму связность тоже можно разложить: $A = A^{0,1} + A^{1,0}$. Компоненту $A^{0,1}$ можно воспринимать как оператор $\bar{\partial}$ дифференцирования по $\bar{z}$. На нашем расслоении можно ввести голоморфную структуру, объявив голоморфными те сечения, которые зануляются этим оператором. Тогда уравнения $F_A^{0,2} = F_A^{2,0} = 0$ будут условиями интегрируемости: локальная комплексная структура интегрируется до настоящей комплексной структуры. Тем самым на нашем расслоении возникает голоморфная структура. Достижение Дональдсона состояло в том, что если выполнено условие $F_A^{1,1} \wedge \omega = 0$, то можно однозначно восстановить $F_A^{1,1}$, которое будет удовлетворять этому условию. Тем самым, больше ничего требовать не надо.

Т е о р е м а 1 (Дональдсон). *Имеется биекция между решениями уравнения Янга—Миллса и различными голоморфными структурами в нашем расслоении.*

Таким образом,

$$\mathrm{Inst}^k = \{G\text{-расслоения на } \mathbb{C}^2\};$$

здесь G — комплексная группа Ли с максимальной компактной подгруппой K; например, в данном случае, когда $K = \mathrm{SU}(n)$, это будет группа $G = \mathrm{SL}(n)$. Просто, когда речь идет о голоморфных расслоениях, плохо говорить о $\mathrm{SU}(n)$-голоморфных расслоениях, лучше говорить о n-мерных векторных расслоениях. О векторных расслоениях на $\mathbb{C}^2$ тоже плохо говорить. Правильнее говорить о расслоениях на каком-то компактном пространстве, например, на $\mathbb{P}^2$, но с фиксированной тривиализацией вдоль бесконечной прямой $\mathbb{P}^1_\infty \subset \mathbb{P}^2$. Для расслоения на компактном пространстве уже имеет смысл говорить про второй класс Черна c_2. Топологическое условие $c_2 = k$ остается.

Значит, пространство модулей инстантонов — это то же самое, что пространство некоторых голоморфных, или алгебро-геометрических, данных. А именно, n-мерных векторных расслоений на $\mathbb{P}^2$ с тривиализацией вдоль прямой.

Физики любят интегрировать по пространству модулей инстантонов, а оно очень некомпактно (сейчас мы увидим, до какой степени оно некомпактно). Поэтому полезно иметь какую-нибудь компактификацию. Карен Уленбек (K. Uhlenbeck) открыла, что в некотором пространстве обобщенных связностей (т. е. в пространстве некоторых распределений) последовательность связностей A_i при $i \to \infty$ может сходиться к $A + \sum\limits_{x \in \mathbb{R}^4} m_x \delta_x$, где $m_x \in \mathbb{N}$. То есть у связности могут возникать дельтаобразные особенности. Нужно сказать, что у этой связности падает топологический заряд. А именно, если $c_2(A_i) = k$, то $c_2(A) = k - \sum m_x$. Это означает, что можно снабдить топологией несвязное объединение

$$\bigcup_{b \leqslant k} \mathrm{Inst}^b \times \mathrm{Sym}^{k-b}(\mathbb{R}^4).$$

Инстантоны имеют алгебро-геометрическую интерпретацию, поэтому естественно попытаться найти алгебро-геометрическую интерпретацию этой компактификации. Этим мы и будем заниматься: попытаемся алгебраически определить

$$\mathrm{Uhl}^a \supset \mathrm{Inst}^a = \mathrm{Bun}_G^a(\mathbb{A}^2).$$

Группа G у нас пока что $\mathrm{SL}(n)$, а вообще это может быть любая полупростая группа.

Тут на самом деле тоже отличился Дональдсон, который заметил следующее. Есть старая конструкция Атьи—Дринфельда—Хитчина—Манина, которые научились строить расслоения на 4-мерной сфере в терминах данных линейной алгебры. Потом ее более красиво переписали Барт, Бейлинсон и другие. Описание $\mathrm{Bun}^a_{\mathrm{SL}(n)}(\mathbb{A}^2)$ в терминах линейной алгебры следующее (здесь важно, что группа именно $\mathrm{SL}(n)$). У нас есть пространство $V = \mathbb{C}^n$, есть пространство $W = \mathbb{C}^a$, действуют линейные операторы $V \xrightarrow{i} W$, $W \xrightarrow{j} V$ и $W \xrightarrow{B_1, B_2} W$. Сначала мы рассматриваем пространство четверок (B_1, B_2, i, j). Это — линейное пространство

$$E = \mathrm{Hom}(V, W) \oplus \mathrm{Hom}(W, V) \oplus \mathrm{End}(W) \oplus \mathrm{End}(W),$$

на котором есть самодвойственность, потому что пространства $\mathrm{Hom}(V, W)$ и $\mathrm{Hom}(W, V)$ двойственны, а эндоморфизмы $\mathrm{End}(W)$ двойственны сами себе. У нас получилось некое линейное пространство плюс двойственное, поэтому там можно ввести каноническую симплектическую форму ω. Кроме того, на W действует заменами координат группа $\mathrm{GL}(W)$; она сопрягает B_1 и B_2, и умножает i справа на себя, а j слева на обратный. Это действие сохраняет симплектическую форму ω. В этой ситуации

можно провести гамильтонову редукцию. Это как раз то, что нужно сделать. Метаутверждение такое: «$\mathrm{Bun}^a(\mathbb{A}^2)$ есть гамильтонова редукция E относительно $\mathrm{GL}(W)$». Оказывается, что в зависимости от того, какой придавать этому смысл, получится либо пространство инстантонов, либо пространство Уленбек.

Смысл этому нужно придавать следующим образом. Во-первых, есть отображение моментов $\mu\colon E \to \mathfrak{gl}(W)$ (отображение моментов отображает E в двойственную алгебру Ли, но алгебра Ли $\mathfrak{gl}$ самодвойственна). Это отображение задается формулой $(B_1, B_2, i, j) \mapsto [B_1, B_2] + ij$. Когда делают гамильтонову редукцию, в первую очередь нужно взять нулевой уровень отображения моментов: $\mu^{-1}(0) \subset E$. Мы наложили на эти четверки условие $[B_1, B_2] + ij = 0$. Дальше нужно профакторизовать по действию группы $\mathrm{GL}(W)$. Но как профакторизовать? Утверждается, что там есть особенно хорошие четверки, на которые теперь наложено условие не замкнутое, а открытое (некоторое условие общности). На этом множестве группа $\mathrm{GL}(W)$ действует свободно, и факторизовать там — одно удовольствие. А именно, есть условие стабильности и условие костабильности. Условие стабильности заключается в следующем: четверка (B_1, B_2, i, j) стабильна, если, грубо говоря, образ i все порождает. А научно выражаясь, если подпространство $W' \subset W$ таково, что $i(V) \subset W'$ и $B_1 W' \subset W' \supset B_2 W'$, то $W' = W$. Условие костабильности двойственно. Четверка (B_1, B_2, i, j) костабильна, если выполняется следующее условие: если подпространство $W' \subset \mathrm{Ker}\ j$ таково, что $i(V) \subset W'$ и $B_1 W' \subset W' \supset B_2 W'$, то $W' = 0$.

Имеет место следующее утверждение, которое восходит, вероятно, к Атье—Хитчину—Дринфельду—Манину (а в такой форме оно написано, например, в книге Дональдсона и Кронхаймера).

1. Действие группы $\mathrm{GL}(W)$ на множестве стабильных и костабильных четверок свободно, и фактор по этому действию — пространство модулей $\mathrm{Bun}^a_{\mathrm{SL}(n)}(\mathbb{A}^2)$.

2. Дональдсон заметил, что если всех этих ограничений не накладывать, а просто взять и наивно профакторизовать (взять категорный фактор — спектр инвариантных функций), то получится алгебраическое многообразие, которое имеет такую же стратификацию, как и пространство Уленбек:

$$\mu^{-1}(0) \,/\, \mathrm{GL}(W) = \mathrm{Uhl}^a_{\mathrm{SL}(n)}(\mathbb{A}^2) = \bigsqcup_{b \leqslant a} \mathrm{Bun}^b \times \mathrm{Sym}^{a-b}(\mathbb{A}^2).$$

Давайте я теперь напишу, как по данным линейной алгебры строится расслоение. Данным (B_1, B_2, i, j) сопоставляется *монада*

$$H^1\big(W \otimes \mathcal{O}_{\mathbb{P}^2}(-1)\big) \xrightarrow{e} (W \oplus W \oplus V) \otimes \mathcal{O}_{\mathbb{P}^2} \xrightarrow{f} W \otimes \mathcal{O}_{\mathbb{P}^2}(1),$$

где

$$e = (z_0 B_1 - z_1,\ z_0 B_2 - z_2,\ z_0 j), \quad f = (-z_0 B_2 + z_2,\ z_0 B_1 - z_1,\ z_0 i),$$

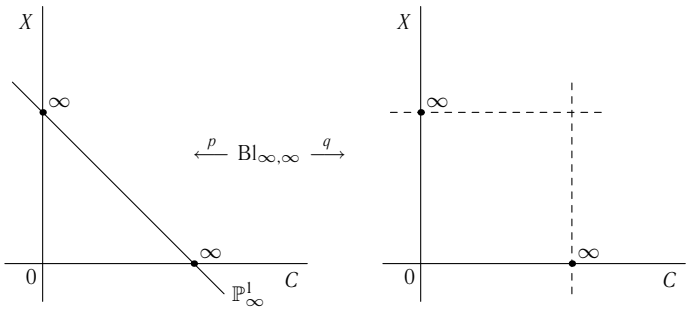
(z_0, z_1, z_2) — однородные координаты на $\mathbb{P}^2$, а $z_0 = 0$ — уравнение бесконечной прямой $\mathbb{P}^1_\infty$. На бесконечной прямой, где $z_0 = 0$, ничего не происходит и остается только $V \otimes \mathcal{O}_{\mathbb{P}^2}$. Так и должно быть: там должна быть тривиализация.

Первый член имеет когомологическую степень -2, средний член — когомологическую степень -1, последний — степень 0. Утверждается, что когда четверка стабильна и костабильна, не будет ни когомологий в первом члене (ядра), ни когомологий в последнем члене (коядра), а будут только когомологии в среднем члене, и это и есть то самое векторное расслоение. Эта конструкция восходит к Атье—Хитчину—Дринфельду—Манину.

Пространство модулей расслоений параметризует расслоения (bundles). А пространством модулей чего является $\mathrm{Uhl}^a_{\mathrm{SL}(n)}(\mathbb{A}^2)$? Дринфельд недавно сообразил, что оно является пространством модулей gundles (гос-слоений) (название в честь того, что этим занимались Гинзбург, Уленбек, Накаджима, Дринфельд, ...). Что такое gundle? Это то, что получится, если подставить произвольные (B_1, B_2, i, j). Тогда возникнут другие когомологии, получится комплекс. Но этот комплекс получится не произвольный, а то, что называется *извращенный пучок*. Для произвольных $(B_1, B_2, i, j) \in \mu^{-1}(0)$ монада — извращенный пучок $\mathcal{F}$, локально свободный в коразмерности 1. Раньше было расслоение, оно вообще всюду локально свободно. А теперь особенности допускаются только в точках; на кривых не может быть особенностей. А в точках особенности могут быть, но какие? Если есть какая-то точка $x \in \mathbb{A}^2$, то в ней можно брать схемно-теоретический слой $i^*(\mathcal{F})$. В общей точке он степени -1. А в любой точке $i^*(\mathcal{F}) \in D^{\leqslant 0}(\mathrm{Vect})$, т. е. степень должна быть не выше нуля. То же самое верно для двойственного по Серру пучка: $i^*(D\mathcal{F}) \in D^{\leqslant 0}(\mathrm{Vect})$. Обычно ровно так определяются извращенные пучки в конструктивной теории, но то же самое можно сказать и в когерентной теории. Кроме того, пучок по-прежнему тривиализован на бесконечной прямой. Значит, если мы возьмем $\mu^{-1}(0)$ и профакторизуем по действию группы $\mathrm{GL}(W)$, то это будет то, что называется *стек* модулей извращенных пучков (таких, как сказано выше), тривиализованных на $\mathbb{P}^1_\infty$ и для которых $c_2(\mathcal{F}) = a$.

Но на самом деле это еще немножко слишком тонко. А именно, оно по-прежнему только отображается в пространство Уленбек:

$$\frac{\mu^{-1}(0)}{\mathrm{GL}(W)} \to \mathrm{Uhl}^a_n.$$



Р и с. 1. Перестройка

То есть нужно дополнительно факторизовать по некоторому отношению эквивалентности. Gundle — это даже не просто пучок, а некоторый класс эквивалентности. И это дополнительное отношение эквивалентности следующее. Два пучка $\mathcal{F}_1$ и $\mathcal{F}_2$ эквивалентны, если есть третий пучок $\mathcal{F}$, который отображается в оба пучка так, что он почти им обоим равен, но поскольку это все-таки комплексы, то там есть такая мера неравенства, как конус (в гомологической алгебре). Так вот, конус сосредоточен на конечном множестве (имеет конечный носитель). А тогда у него определен *цикл* в симметрической степени $\mathbb{A}^2$. Эти два цикла должны совпадать, т. е. $\mathcal{F} \xrightarrow{f_1} \mathcal{F}_1$, $\mathcal{F} \xrightarrow{f_2} \mathcal{F}_2$ и $[\mathrm{Cone}\ f_1] = [\mathrm{Cone}\ f_2]$. Такой класс эквивалентности и есть gundle.

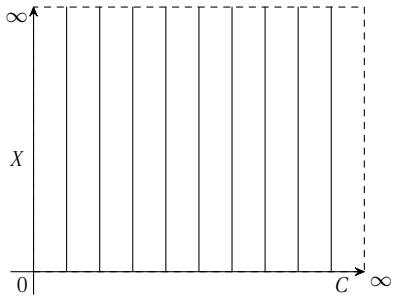
Забавно, что $\dfrac{\mu^{-1}(0)}{\mathrm{GL}(W)}$ — просто стек, это не алгебраическое многообразие. Но если чуть профакторизовать по дополнительному отношению эквивалентности, то получится уже замечательное алгебраическое многообразие.

Плохо то, что все это действует только для группы $\mathrm{SL}(n)$, т.е. для векторных расслоений. Если бы нас интересовали $\mathrm{SO}(n)$-расслоения или, тем более, E_8-расслоения, то совершенно непонятно было бы, что делать.

Теперь вопрос в том, как определить аналогичную компактификацию

$$\mathrm{Bun}_G^a \subset \mathrm{Uhl}_G^a = \bigsqcup_{b \leqslant a} \mathrm{Bun}_G^b \times \mathrm{Sym}^{a-b}(\mathbb{A}^2)$$

для произвольной полупростой группы Ли G над $\mathbb{C}$. Для этого сначала произведем некоторую перестройку. У нас было $\mathbb{P}^2$, в нем была бесконечная прямая $\mathbb{P}^1_\infty$. Выберем еще две прямые C и X (рис. 1). Две точки пересечения прямой $\mathbb{P}^1_\infty$ с прямыми C и X можно раздуть. Вклеются две прямые, но зато прямую $\mathbb{P}^1_\infty$ можно будет сдуть. Получится произведение $C \times X$. Если, как раньше, было векторное расслоение, тривиализованное



Р и с. 2. Вертикальные прямые

вдоль прямой $\mathbb{P}^1_\infty$, то его можно поднять в раздутие $Bl_{\infty,\infty}$, а потом спустить в $C \times X$, потому что оно тривиализовано вдоль прямой, которую мы сдуваем. Но поскольку мы раздули две точки, в которых оно было тривиализовано, оно станет тривиализованным вдоль $\{C \times \infty\} \cup \{\infty \times X\}$. Тем самым, имеется эквивалентность между векторными расслоениями на $\mathbb{P}^2$, тривиализованными вдоль $\mathbb{P}^1_\infty$, и векторными расслоениями на квадрике $C \times X$, тривиализованными вдоль креста $\{C \times \infty\} \cup \{\infty \times X\}$.

Если теперь группа G произвольная, то что такое G-расслоение? Во-первых, по каждому представлению нашей группы можно построить ассоциированное векторное расслоение. А если два представления тензорно перемножаются, то ассоциированные расслоения тоже тензорно перемножаются. Так вот, G-расслоение — это ровно такие данные, а именно, тензорный функтор из G-модулей в векторные расслоения на поверхности. Если мыслить так и произвести эту операцию с каждым ассоциированным векторным расслоением, то мы увидим, что G-расслоение на $\mathbb{P}^2$, тривиализованное вдоль $\mathbb{P}^1_\infty$, это то же самое, что расслоение на квадрике $C \times X$, тривиализованное вдоль $C \times \infty \cup \infty \times X$. Поэтому дальше мы будем мыслить в терминах расслоений на квадрике. Это просто другая реализация Bun^a_G (второй класс Черна остается тем же).

На это нужно смотреть так. Нужно провести вертикальные прямые (рис. 2). На поверхности у нас есть G-расслоение, но если его ограничивать на прямые $c \times X$, то будет получаться семейство G-расслоений на прямых, причем все они будут тривиализованы в точках $c \times \infty$. Если мы ограничимся на прямую $\infty \times X$, то это расслоение будет тривиализовано полностью. Таким образом, ограничивая расслоение на вертикальные прямые вида $c \times X$, получаем отображение $C \to \mathrm{Bun}_G(X)$. Но что такое $\mathrm{Bun}_G(X)$, т. е. модули G-расслоений на X?

Давайте разберем пример, когда $G = \mathrm{SL}_2$, т. е. мы говорим просто про двумерные расслоения степени 0 на проективной прямой X. Таких

расслоений бывает очень немного. Бывает $\mathcal{O}\oplus\mathcal{O}$, бывает $\mathcal{O}(-1)\oplus\mathcal{O}(1)$, $\mathcal{O}(-2)\oplus\mathcal{O}(2)$ и т. д. То есть это получается как бы набор точек. Но на самом деле это опять же то, что называется стек. На самом деле, есть понятие семейства расслоений. Классов изоморфизма всего дискретное множество, но бывают семейства.

Так что $\mathrm{Bun}_G(X)$ — это стек. А стек — это фактор алгебраического многообразия по какому-то отношению эквивалентности. Алгебраическое многообразие здесь — это аффинный грассманиан $\mathcal{G}_X$ группы G, связанный с кривой X. Чтобы объяснить, что такое грассманиан, нужно, как всегда это бывает, ввести дополнительную жесткость. После того как жесткость введена, получается многообразие модулей. Потом нужно профакторизовать по отождествлениям с разными жесткостями. Жесткость такая: расслоение с тривиализацией в некоторой точке. Но этого мало. На самом деле нужно тривиализовать не просто в точке, а в формальной окрестности, т. е. тривиализовать все струи. Таким образом, аффинный грассманиан группы G — это многообразие модулей пар $(\mathcal{F},\tau)$, где $\mathcal{F}$ есть G-расслоение на прямой X, а τ — формальная тривиализация в точке ∞. Формальная тривиализация это уже много всего. Для этого нужно отождествить ограничение нашего расслоения на формальную окрестность точки с тривиальным расслоением. Поэтому получается бесконечномерное многообразие. Если на него смотреть на уровне точек, то, как всегда, нужно ввести параметр t, который в начале координат равен 0, а в точке ∞ равен ∞, нужно взять группу петель $G((t^{-1}))$ и профакторизовать ее по $G[t]$ (полиномам от t). На уровне точек — это наш грассманиан $\mathcal{G}_X$. На нем продолжает действовать проалгебраическая группа формальных рядов $G[[t^{-1}]]$ (замены тривиализации формальной окрестности). Группа формальных рядов — это хорошее образование. Оно хотя и бесконечномерное, но проалгебраическое. Если по ней профакторизовать, то получится как раз стек модулей расслоений: $\mathcal{G}_X/G[[t^{-1}]]=$ $=\mathrm{Bun}_G$. Все как обычно: просто расслоение склеивается в проколотой окрестности точки ∞. Говорится, что всякое расслоение тривиально на дополнении, тривиально в диске, но как-то склеивается в проколотом диске. Склейки в проколотом диске — это $G((t^{-1}))$. Здесь мы не интересуемся дополнением, поэтому профакторизовали по $G[t]$. Остается действие замен тривиализаций в формальном диске. Если по нему тоже профакторизовать, то получится пространство модулей всех расслоений. Точнее, не пространство, а стек.

Оно очень похоже на пространство флагов. Это — параболический вариант пространства флагов группы петель. Оно, в частности, имеет клетки Брюа; все хорошее про него верно. Там есть открытая клетка

Брюа, которая отвечает тривиальным расслоениям. Общее расслоение на $\mathbb{P}^1$ тривиально. Грамотно нужно говорить так. Если есть семейство расслоений, и в некоторой точке этого семейства расслоение тривиально, то в некоторой окрестности в топологии Зарисского оно тоже будет тривиально. В этом смысле общее расслоение на $\mathbb{P}^1$ тривиально. Но могут быть точки, в которых оно вырождается; на дивизоре так: $\mathcal{O}(-1)\oplus\mathcal{O}(1)$ (в случае SL(2)). А потом в коразмерности 3 оно становится $\mathcal{O}(-2)\oplus\mathcal{O}(2)$. Затем в коразмерности 5 (размерности скачут через 2) оно становится $\mathcal{O}(-3)\oplus\mathcal{O}(3)$.

Самый простой способ это представлять — примерно так же, как для флагов с клетками Брюа. Эти классы расслоений как раз задают разбиение на клетки Брюа, точнее, чуть более грубое разбиение. Это разбиение параметризуется доминантными ковесами группы G. Иными словами, двойным фактором аффинной группы Вейля по конечной группе Вейля.

То, что нам пригодится, — это то, что и в $\mathrm{Bun}_G(X)$ и в $\mathcal{G}_X$ есть открытые подмножества (каждое из этих подмножеств — дополнение к дивизору), которые параметризуют тривиальные G-расслоения. У нас есть отображение из C в этот стек, но в бесконечности он попадает как раз в этот открытый кусок — по определению, потому что расслоение тривиализовано на бесконечном кресте, в частности, на $\infty\times X$. Но это значит, что оно почти всюду попадает в этот кусок. Мы получаем, что наше расслоение в ограничении на $c\times X$ тривиально для почти всех c. Кроме того, сразу получается (это нам тоже будет полезно) некоторый дивизор на C, а именно, обратный образ дивизора нетривиальных расслоений. Получаем дивизор $\eta(V)$, где V — наше расслоение. Это — эффективный дивизор на кривой. Он имеет степень a. Второй класс Черна расслоения переписывается таким способом. Получаем $\eta(V)\in\mathrm{Sym}^a(C)$ — прообраз нетривиального дивизора с $\mathrm{Bun}_G(X)$. Но теперь заметим еще, что если поменять местами X и C, то получится, что ограничение нашего расслоения V на почти любую горизонтальную прямую тоже тривиально. Есть некоторый дивизор, в ограничении на соответствующие горизонтальные прямые расслоение будет нетривиально, но почти всегда оно тривиально. Следовательно, оно тривиально не просто на выделенной прямой, но и в целой ее окрестности по Зарисскому. В частности, оно тривиально и в формальной окрестности. А раз расслоение тривиально в формальной окрестности, то мы получили подъем отображения

$$\begin{array}{ccc} C & \longrightarrow & \mathrm{Bun} \\ & & \uparrow \\ & & \mathcal{G}_X \end{array}$$

т. е. отображение $C \to \mathcal{G}_X$. То есть расслоение не просто тривиализовано на этой прямой, но еще снабжено тривиализацией в ее формальной окрестности. Мы получили отображение $C \to \mathcal{G}_X$, не просто в стек модулей расслоений, а в его покрытие, в грассманиан. Это отображение обладает тем свойством, что в бесконечной точке C оно попадает в отмеченную точку грассманиана. А именно, там расслоение тривиальное и еще снабжено постоянной тривиализацией. Получаем базированное отображение $(C, \infty) \to (\mathcal{G}_X, g)$.

Нужно еще вспомнить, что за условие имеется на второй класс Черна. Оно переписывается следующим образом. Отображение $(C, \infty) \to (\mathcal{G}_X, g)$ имеет степень a. Это означает следующее. У нас есть проективная прямая (т. е. сфера), которая попадает в грассманиан. У нее есть фундаментальный цикл, а у грассманиана есть вторые гомологии $H_2(\mathcal{G}_X, \mathbb{Z}) = \mathbb{Z}$ (у грассманиана есть единственная клетка Брюа коразмерности 1). У них есть каноническая образующая. Должно выполняться условие $[C] = a$ (фундаментальный класс нашей кривой равен a). Еще можно сказать так. Образующая вторых когомологий $H^2(\mathcal{G}_X, \mathbb{Z})$ — это $c_1(\mathcal{L})$, где $\mathcal{L}$ — детерминантное расслоение на грассманиане. Детерминантное расслоение имеет много разных определений. В частности, $\mathcal{L} = \mathcal{O}(\mathrm{Div})$, где Div — дивизор нетривиальных расслоений. Поэтому не удивительно, что когда мы брали прообраз дивизора нетривиальных расслоений, он имел степень a на нашей кривой. Это и означает, что в этом смысле степень отображения равна a.

Короче говоря, из всего этого следует такое забавное утверждение.

У т в е р ж д е н и е 1. *Пространство расслоений* $\mathrm{Bun}_G^a(\mathbb{A}^2)$ — *это то же самое, что пространство отображений* $\mathrm{Map}^a(C, \infty; \mathcal{G}_X, g)$ (*базированные отображения степени a в грассманиан*).

Это довольно забавно, потому что $\mathrm{Bun}_G^a(\mathbb{A}^2)$ — это некоторое конечномерное многообразие, а грассманиан сам по себе чудовищно большой. Но пространство базированных отображений данной степени опять получается конечномерным.

Это позволяет легко определить компактификацию. Что такое отображение из кривой в грассманиан? Сам по себе грассманиан хотя и бесконечномерный, но проективное многообразие. А именно, на нем есть обильное расслоение (детерминантное). Его сечения известны:

$$\Gamma(\mathfrak{g}_x, \mathcal{L}) = V_{\omega_0}$$

— это то, что называется базисное фундаментальное представление соответствующей аффинной алгебры Ли (алгебры петель) $\hat{\mathfrak{g}}$. Оно тоже бесконечномерное. Тем самым получается обычное проективное вложение $\mathcal{G}_X \subset \mathbb{P}(V_{\omega_0}^*)$; грассманиан вкладывается в проективизацию двойственного

пространства. Из-за того, что пространство было бесконечномерное, двойственное пространство — проконечномерное векторное пространство, топологическое. Там грассманиан задается какими-то уравнениями. Это то, что называется *уравнения Плюккера*. В правильных координатах для SL_2 эти уравнения переписываются как нелинейные дифференциальные уравнения Кадомцева—Петвиашвили. Это неважно. Главное, что какими-то уравнениями грассманиан здесь задается, хотя и этих уравнений в бесконечномерном пространстве бесконечное число. Мы только этим и будем пользоваться.

Что такое отображение из кривой в такое хозяйство? Это означает, что в каждой точке кривой задана некоторая прямая из $V^*_{\omega_0}$, которая бегает вместе с точкой кривой, но все время остается внутри грассманиана, т. е. удовлетворяет уравнениям Плюккера. Иными словами, задано линейное подрасслоение в тривиальном расслоений со слоем $V^*_{\omega_0}$. Таким образом, пространство отображений Maps^a — это пространство линейных подрасслоений $L \subset V^*_{\omega_0} \otimes \mathcal{O}_C$ с поточечными условиями Плюккера (в каждой точке отображение попадает в грассманиан) и есть условие в бесконечности, что ∞ попадает именно в точку g. Теперь видно, почему это некомпактное многообразие: потому что мы ограничивались линейными подрасслоениями. Линейное подрасслоение можно задавать его сечением; нужно только потребовать, чтобы оно не обращалось в нуль. А можно разрешить сечениям обращаться в нуль, т. е. рассматривать обратимые подпучки. Значит, Maps^a лежит в замыкании $\overline{\mathrm{Maps}}^a$, которое придумал Дринфельд для несколько иных целей. Здесь $\overline{\mathrm{Maps}}^a$ — обратимые подпучки с теми же условиями. (Я забыл сказать, что a — это условие на степень: $\deg L = -a$.) На самом деле, это то же самое, что сечения, потому что мы знаем, что такое линейное расслоение на проективной прямой степени $-a$: это $\mathcal{O}(-a)$. Если, наоборот, это все подкрутить на $\mathcal{O}(a)$, то получится $\mathcal{O} \subset V^*_{\omega_0} \otimes \mathcal{O}(a)_C$ с поточечными условиями Плюккера. Но подпучок $\mathcal{O}$ — это буквально то же самое, что сечение, только с точностью до пропорциональности. Так что речь идет о пространстве сечений (не тождественно нулевых) с точностью до пропорциональности. Но раньше было, что они нигде не зануляются, ни в одной точке, а здесь мы выбрасываем это условие. Просто берем все сечения с точностью до пропорциональности. Так получается компактификация.

Про нее известно, как она устроена. Она на самом деле обладает похожей стратификацией:

$$\overline{\mathrm{Maps}}^a = \bigsqcup_{b \leqslant a} \mathrm{Maps}^b \times \mathrm{Sym}^{a-b}(C \setminus \infty).$$

Здесь только симметрическая степень не поверхности, как мы хотели, а кривой. Понятно, откуда это происходит. Если у нашего сечения разрешаются нули, то мы помним, какие были нули и с какими кратностями. Так и получается стратификация. Но еще раз повторю, это несколько меньше, чем мы хотели. Мы хотели получить симметрическую степень поверхности, а не кривой. И это проявляется в том, что такое пространство, хотя оно и склеено из кусочков вполне приличных, конечномерных, все вместе оно плохое. Оно, что называется в алгебраической геометрии *бесконечного типа*. Оно примерно такого сорта, как если взять плоскость и стянуть на ней прямую. То есть, иначе говоря, если взять $\mathrm{Spec}(1 \oplus xk[x, y])$. Это такое странное образование, которое стратифицировано плоскостью без прямой и точкой. Хотя оно стратифицировано вполне прилично, но все вместе это нечто ужасное: бесконечного типа. Так и это пространство: хотя оно и стратифицировано приличными вещами, но все вместе склеивается в нечто ужасное.

Требуется его немножко разрешить, чтобы получить компактификацию Уленбек. Это уже совсем просто. А именно, нужно все это проделать в таком относительном варианте. У нас раньше были фиксированы две прямые, а теперь рассмотрим произвольные пары прямых. Пусть R — множество пар прямых (X_r, C_r); обе прямые пересекают бесконечную прямую, но в разных точках. Это аффинное многообразие (как бы множество координат на плоскости). Имея одну такую пару прямых, можно проделать все то, что мы раньше говорили. В частности, можно по расслоению получить некоторый дивизор на одной из этих кривых. У нас есть относительные прямые $\mathcal{X} \to R$ и $\mathcal{C} \to R$. Для одного расслоения мы построили некоторое сечение. То есть для каждой пары прямых мы получили некоторую точку относительной симметрической степени $\mathrm{Sym}^a_R\, \mathcal{C}$. Таким образом, мы получили отображение

$$\eta\colon\ \mathrm{Bun}^a_G(\mathbb{A}^2) \to \mathrm{Sect}_R\, \mathcal{C}^{(a)}$$

из всех расслоений в пространство всех сечений этой симметрической степени.

Вообще, сечений тоже много, они бывают разных степеней. Но здесь можно априори оценить, куда мы попадаем; это еще вполне конечномерное пространство. Все. Теперь у нас есть вложение $\mathrm{Bun}^a_G(\mathbb{A}^2)$, с одной стороны, в компактификацию Дринфельда $\overline{\mathrm{Maps}^a}$, которая слишком мала, а с другой стороны, в пространство сечений $\mathrm{Sect}_R\, \mathcal{C}^{(a)}$. Поэтому можно взять вложение

$$\mathrm{Bun}^a_G(\mathbb{A}^2) \to \overline{\mathrm{Maps}^a} \times \mathrm{Sect}_R\, \mathcal{C}^{(a)}.$$

Теперь нужно только взять замыкание в этом произведении. Тут уже все будет хорошо. Таким образом, Uhl^a_G — это просто замыкание в произведении. Оно будет иметь такую стратификацию, как надо. Давайте я только поясню, какое отношение имеет симметрическая степень плоскости к этим сечениям. Если есть какой-то 0-цикл на плоскости и есть такое разложение плоскости, то его можно спроецировать вдоль прямой X на прямую C. Тогда получится как раз 0-цикл степени a на кривой C. Но это и есть вложение симметрической степени плоскости в эти сечения. Так они там лежат.

Теперь я хочу рассказать, что про это пространство Уленбек известно. Пространство Уленбек конечномерное, но особое. Есть стандартный способ исчисления особенностей. Это — подсчет слоев пучка Горески—Макферсона. Я не буду объяснять, что это такое. По крайней мере, ответ — это некоторый набор чисел и производящая функция для этих чисел. Например, если взять двумерный квадратичный конус в трехмерном пространстве, то у него есть особенность в вершине, и там слои Горески—Макферсона тривиальны (постоянный пучок). А если взять трехмерный конус в четырехмерном пространстве, там уже получаются слои степени 0 и степени 2. В таком случае принято писать $1+q$.

Размерность Bun^a_G равна $2a\check{h}$, где $\check{h}$ — дуальное число Коксетера для группы G. Для SL_n это будет n, а, например, для G_2 это будет 4. Чтобы написать производящие функции для этих особенностей, нужно немножко углубиться в аффинные алгебры Ли. Проще всего ответ сказать в случае, когда G типа A, D или E (то, что называется *с простыми связями*). Дело в том, что алгебра Ли, которая имеет к нам отношение, такая. Мы берем сначала аффинную алгебру $\hat{\mathfrak{g}}$ (которая уже возникала: у нее был грассманиан), а потом мы берем двойственную по Ленглендсу алгебру $\hat{\mathfrak{g}}^\vee$ (для нее матрица Картана транспонированная). Но беда в том, что если исходная алгебра была не ADE, то $\hat{\mathfrak{g}}^\vee$ будет не обычная алгебра петель, а, скрученная. Если мы начнем с группы G_2, то будет, что называется, $D_4^{(3)}$, и для нее комбинаторика еще хуже. Поэтому давайте я разберу только случай ADE. В этом случае алгебра $\hat{\mathfrak{g}}$, которая в данном случае будет изоморфна алгебре, двойственной по Ленглендсу, выглядит так: $\hat{\mathfrak{g}} = \mathbb{C}[t, t^{-1}] \otimes \mathfrak{g} \oplus CK \oplus CD$; добавляется одна образующая центральная и еще одна образующая — обращение петли, или оператор энергии. В ней есть подалгебра $\mathfrak{n} = t\mathbb{C}[t] \otimes \mathfrak{g}$ (нильпотентный радикал параболической). В самой алгебре Ли $\mathfrak{g}$ есть главный нильпотент e. Для SL_n это матрица, в которой на побочной диагонали стоят единицы, а в остальных местах — нули. Во всякой группе относительно присоединенного действия конечное число нильпотентных орбит, поэтому среди них есть одна открытая; из

нее берется главный нильпотент. Он присоединенно действует на алгебре Ли и задает на ней так называемую монодромическую фильтрацию. Его действие по теореме Джекобсона—Морозова можно продолжить до действия SL_2, поэтому все это распадется в сумму SL_2-модулей. Нас будет интересовать ядро этого действия, т. е. старшие векторы, но мы будем помнить, какие именно старшие веса там есть. Я хочу сказать, что инварианты этого главного нильпотента градуированы некоторым образом: $\mathfrak{g}^e = \bigoplus\limits_{k\geqslant 0} (\mathfrak{g}^e)_k$. Это называется градуировкой Костанта, или монодромической градуировкой. Вся алгебра, стало быть, градуирована двумя степенями:

$$(t\mathbb{C}[t]\otimes \mathfrak{g}^e) = \bigoplus_{d,k} \mathfrak{n}_k^d.$$

Одна градуировка берется по нильпотенту, а другая по оператору энергии.

Теперь можно сформулировать ответ таким образом. Рассмотрим точку

$$\Big(V\in \mathrm{Bun}^b,\ \sum m_i s_i\Big),\ \text{ где }\ \sum m_i = a-b.$$

Здесь $\sum m_i s_i$ — это несколько точек плоскости. Важно, как они сливаются. Если точки различные, без кратностей, то там будут простейшие особенности. А если точки начинают сливаться, то возникают более сложные особенности. Поэтому я написал набор точек с кратностями. Слой пучка Горески—Макферсона в этой точке — это будет

$$\bigotimes_i \Big(\bigoplus_{k\geqslant 0} \mathrm{Sym}(\mathfrak{n}^e)^k_{m_i}\Big)[2k].$$

Здесь k отвечает за сдвиг гомологической степени, а вторая градуировка отвечает за то, сколько точек слилось.

Давайте я скажу два слова, откуда возникает комбинаторика двойственной по Ленглендсу аффинной алгебры. Мы рассматривали отображения $\mathrm{Maps}^a(C,\, c;\, \mathcal{G}_X,\, g)$ из базированной кривой в грассманиан. Я уже говорил, что грассманиан — это как бы пространство флагов для группы петель. А у нее есть настоящее пространство флагов $\mathcal{B}$. Из него есть проекция на грассманиан: $\mathcal{B}\to\mathfrak{g}$; в слое стоят конечномерные флаги. У $\mathcal{B}$ вторых гомологий уже много. Они нумеруются не просто целыми числами, а целой решеткой кокорней: $H_2(\mathcal{B};\,\mathbb{Z}) = Y$ — решетка кокорней аффинной алгебры Ли $\hat{\mathfrak{g}}$. Здесь тоже есть клетки Брюа, занумерованные аффинной группой Вейля. В частности, среди них есть клетки коразмерности 1: дивизоры Брюа. Их столько штук, сколько простых корней для аффинной группы.

Для α из положительного конуса в Y можно рассмотреть пространство базированных отображений степени α: $\mathrm{Maps}^{\alpha}(C, c; \mathcal{B}, b)$. Если есть такое отображение, то мы можем брать прообраз не одного единственного дивизора, как было в грассманиане, а прообраз любого дивизора Шуберта. Это получится уже крашеный дивизор на нашей кривой. Красок будет столько, каков ранг нашей группы. Поэтому есть отображение

$$\mathrm{Maps}^{\alpha}(C, c; \mathcal{B}, b) \xrightarrow{\varphi} (C \setminus c)^{\alpha} = \mathbb{A}^{\alpha}.$$

На $\mathrm{Maps}^{\alpha}(C, c; \mathcal{B}, b)$ есть симплектическая форма, относительно которой это — интегрируемая система. То есть все эти координаты, все функции находятся в инволюции в соответствующей гамильтоновой структуре. Если взять специальный слой $\varphi^{-1}(\alpha \cdot 0)$, то это будет лагранжево многообразие (относительно симплектической формы). Для каждого кокорня α появилось некоторое множество неприводимых компонент. Их можно объединить вместе, т. е. взять $B = \bigsqcup_{\alpha} \mathrm{Irr}\, \varphi^{-1}(\alpha \cdot 0)$. Получится кристалл Кашивары для алгебры $\hat{\mathfrak{g}}^{\vee}$. Кристалл это не просто множество, а множество с операциями e_i и f_i, т. е. между множествами есть некоторые примыкания, которые я сейчас не буду определять.

31 октября 2002 г.

М. А. Шубин

РАВНОВЕСИЕ НЭША

Я должен сразу вас предупредить, что я не являюсь экспертом по теории игр, и этот доклад в основном будет носить популярно-методологический и исторический характер. Честно признаюсь, что я вплотную заинтересовался Нэшем и его работами, в частности, его работами по теории игр, после того как появилась замечательная книга о нем, которая называется «Beautiful Mind», и фильм, который по английски тоже называется «Beautiful Mind», а по-русски он называется «Игры разума». Этот фильм, конечно, гораздо хуже, но что-то в нем тоже есть. Во всяком случае, он вызвал такой ажиотаж вокруг имени Нэша, что стало неудобно ничего об этом не знать. Я решил с этим познакомиться и получил от этого большое удовольствие. Хочу этим поделиться.

Нэш сейчас вполне работоактивен. Недавно я слышал его доклад на математическом конгрессе в Пекине, где он рассказывал, между прочим, и о теории игр тоже. В частности, про какие-то вещи, которые в некотором смысле являются продолжением его старых работ. Но я хотел рассказать про его старую работу и ее некоторое развитие. Это — работа, за которую он получил Нобелевскую премию. Фактически, в этой работе одна страница. Поразительным образом эту работу можно сравнительно легко прочесть и понять, что я и сделал. Это доставило мне большое удовольствие.

Нэш хотел, чтобы эта работа была его кандидатской диссертацией (собственно, в Америке других и не бывает) PhD. Но его руководитель засомневался, я бы даже сказал испугался, и стал его уговаривать написать немного больше: добавить какие-то примеры и т. д. Кончилось дело тем, что Нэш придумал некое другое доказательство, которое имеет свои преимущества, но с эстетической точки зрения, как мне кажется, не такое интересное. Это другое доказательство было опубликовано в его работе в Annals of Mathematics. В этой работе уже 11 страниц, но там зато много

Я глубоко благодарен Виктору Васильевичу Прасолову за редактирование и подготовку этой лекции к печати, а также Сергею Игоревичу Соболеву за высококвалифицированную техническую помощь.

примеров. И эта работа уже была его диссертацией. Она переведена на русский язык и помещена в одном из сборников по теории игр.

Можно было бы просто прочитать эту работу и постараться увидеть, что же мы в ней можем понять. Но я предпочту рассказывать об этом независимо. Потом, когда я немножко подробнее расскажу, мы вернемся к этому тексту и на него посмотрим.

Первая моя цель состоит в том, чтобы объяснить формулировку и доказательство теоремы Нэша. Именно то доказательство, которое занимает одну страницу. Но поскольку мне придется рассказать некоторые предварительные сведения и еще одну теорему, на которую Нэш ссылается, ее не доказывая, то у меня это займет немножко больше времени и немножко больше места, чем одна страница.

Речь идет об играх, которые называются *бескоалиционными*. Что такое бескоалиционная игра? У нее есть n игроков. У каждого игрока есть свое конечное множество возможных стратегий S_i. Пусть $|S_i| = n_i$. У каждого игрока в каждый момент игры есть возможность выбора одной из стратегий. После того как каждый из игроков выбрал стратегию, каждый игрок получает некоторую сумму денег p_i (выигрыш или проигрыш). Таким образом, p_i — функция от $s_1 \in S_1, \ldots, s_n \in S_n$, т. е.

$$p_i : S_1 \times S_2 \times \ldots \times S_n \to \mathbb{R}.$$

Если число p_i положительное, то это выигрыш, а если отрицательное, то проигрыш. Друг с другом игроки никак не советуются, и каждый игрок заинтересован выиграть как можно больше. Пока мы говорим про то, что происходит в данный момент времени. Здесь, на самом деле, не очень удачные обозначения; через некоторое время я их поменяю.

В таком виде трудно сказать что-то содержательное об этой игре. Предположим теперь, что эту игру можно повторять. Введем смешанные стратегии. Смешанная стратегия i-го игрока — это когда он выбирает чистые стратегии, т. е. элементы множества S_i, с какими-то вероятностями. Таким образом, смешанная стратегия i-го игрока — это распределение вероятностей на множестве S_i. Давайте теперь поменяем обозначения. Пусть $\{\pi_{i\alpha}\} = S_i$ — множество всех чистых стратегий i-го игрока. Это то же самое множество, только стратегии теперь занумерованы дополнительно индексом α. Тогда смешанная стратегия задается числами $c_{i\alpha} \geqslant 0$, причем $\sum c_{i\alpha} = 1$. Такой выбор (смешанную стратегию) мы будем представлять себе еще как формальную линейную комбинацию $s_i = \sum\limits_{\alpha} c_{i\alpha}\pi_{i\alpha}$. И мы будем рассматривать ее как точку множества $\widehat{S}_i = \{s_i\}$, которое представляет собой $(n_i - 1)$-мерный симплекс с вершинами $\pi_{i\alpha}$. Мы представляем себе $\pi_{i\alpha}$ как линейно независимые векторы в пространстве размерности

n_i. Тогда условие, что сумма координат равна 1, а сами координаты неотрицательны, задает симплекс.

Выбор всеми игроками смешанных стратегий назовем *ситуацией*. Ситуация — это точка

$$s = (s_1, \ldots, s_n) \in \widehat{S}_1 \times \ldots \times \widehat{S}_n = \widehat{S}.$$

Множество ситуаций — это прямое произведение симплексов. Прямое произведение симплексов можно рассматривать как выпуклое подмножество в пространстве соответствующего числа измерений.

Теперь мы должны продолжить функцию выплаты p_i на множество смешанных стратегий; она была определена на множестве чистых стратегий. Продолжим ее следующим образом: функция $p_i : \widehat{S} \to \mathbb{R}$ — это математическое ожидание выигрыша i-го игрока.

Как это нужно себе представлять? Предположим, что у i-го игрока есть ровно две стратегии π_{i1} и π_{i2}, и он выбирает стратегию π_{i1} с вероятностью c_1, а стратегию π_{i2} с вероятностью c_2, где $c_1 + c_2 = 1$. Предположим, что при выборе стратегии π_{i1} он получает p_1, а при выборе стратегии π_{i2} он получает p_2. Тогда, если стратегии остальных игроков фиксированы, средний выигрыш будет $c_1 p_1 + c_2 p_2$. То есть это будет линейная функция на отрезке. А в общем случае это будет линейная функция на соответствующем симплексе. Мы продолжаем функцию таким образом, чтобы она была линейной по каждой группе переменных. В результате получится многочлен степени n, где n — число игроков.

В итоге мы получаем набор функций $p_i : \widehat{S} \to \mathbb{R}$, где $i = 1, \ldots, n$. Каждая из этих функций линейна по каждой группе переменных, входящих в $\widehat{S}$. Каждый игрок заинтересован в максимизации своего выигрыша; p_1 задает его выигрыш или проигрыш при выборе стратегии каждым игроком.

В статье Нэша мы продвинулись на один абзац, что уже немало. Нэш пишет, что можно определить понятие игры n игроков, в которой каждый игрок обладает конечным числом чистых стратегий, и при этом каждому выбору каждым игроком одной из этих чистых стратегий соответствует определенная выплата каждому игроку. Потом он описывает смешанные стратегии: «Смешанные стратегии являются вероятностными распределениями на множестве чистых стратегий. Функция выплаты является математическим ожиданием этих игроков. Тем самым она становится полилинейной формой от вероятностей, с которыми различные игроки выбирают свои чистые стратегии».

Во втором абзаце Нэш пишет, что каждый набор n стратегий (по одной для каждого игрока) можно рассматривать как точку в произведении n пространств, представляющих собой стратегии каждого игрока. Мы пока дошли до этого места в статье Нэша. Я буду продолжать. Дальше

Нэш пишет, что набору n стратегий (т. е. тому, что я назвал ситуацией) противостоит второй набор n стратегий, если стратегия каждого игрока во втором наборе дает наибольшее возможное ожидание выигрыша для этого игрока при условии, что стратегии остальных $n-1$ игроков фиксированы.

Это я должен объяснить чуть более подробно. Давайте выберем две ситуации $s=(s_1, \ldots, s_n)\in\widehat{S}$ и $s'=(s'_1, \ldots, s'_n)\in\widehat{S}$. Мы будем говорить, что s' *противостоит* s, если

$$p_1(s'_1, s_2, \ldots, s_n)=\max_{\widetilde{s}_1} p_1(\widetilde{s}_1, s_2, \ldots, s_n)$$

и то же самое верно для остальных игроков. То есть если мы заменим стратегию одного игрока в ситуации s на стратегию того же игрока в противостоящей ситуации s', то это даст максимум того, что он может заработать, если стратегии всех остальных игроков в ситуации s фиксированы.

Равновесием называется ситуация, которая противостоит сама себе. Если мы находимся в равновесии, то каждый игрок получает максимум того, что он может получить при фиксированных стратегиях остальных игроков. Но не исключено, что оптимальный выбор не единствен.

Теорема, за которую Нэш получил Нобелевскую премию, состоит в следующем.

Т е о р е м а 1. *Равновесие всегда существует.*

Эта теорема, разумеется, довольно элементарна. Чтобы вы не питали слишком больших надежд, я могу сказать, что сейчас вряд ли можно получить Нобелевскую премию за что-нибудь в этом роде. За прошедшие 50 лет эта наука ушла очень далеко вперед. В частности, ей занимались такие выдающиеся математики, как Смейл, который написал около 1000 страниц работ на эту тему. Там добавлена динамика (теорема Нэша относится к статической ситуации). Можно также добавить критерии, которые определяются не одним числом, а многими числами. Наука ушла далеко вперед, тем не менее красота этой теоремы остается. Я немножко отвлекусь на историю и скажу, что теория игр была впервые изобретена фон Нойманом и основательно описана в его книге с Моргенстерном, которая называется «Теория игр и экономическое поведение». Это классическое сочинение, на которое опирался Нэш. Но в основном там рассматривались игры с нулевой суммой: если кто-то что-то проиграл, то другой выиграл. У Нэша это совершенно не обязательно. Это была одна из первых работ, где рассматривались игры с ненулевой суммой. Но вообще фон Нойман, который сам несомненно получил бы Нобелевскую премию за создание теории игр, если бы дожил до нужного момента, отнесся к работе Нэша очень прохладно. Он сказал, что это тривиальное следствие теоремы о неподвижной точке.

Давайте я теперь объясню, почему это — теорема о неподвижной точке, и как это доказывать. А потом мы поговорим о разных других аспектах теоремы Нэша.

Приступим теперь к доказательству теоремы Нэша, которое изложено в оставшемся абзаце его статьи.

Доказательство. Сопоставим каждой ситуации $s \in \widehat{S}$ множество $\Phi(s) \subset \widehat{S}$, состоящее из всех ситуаций, противостоящих s. Множество $\Phi(s)$ замкнутое, поскольку в пределе все сохранится, но кроме того оно еще и выпуклое ввиду линейности каждой из этих функций по каждому переменному. Мы можем даже сразу сказать больше. Дело в том, что это очень удачное определение. В ситуации, противостоящей s, нам не надо заменять все стратегии. Нужно заменять только одну. Мы максимизируем по одной переменной. По сути дела, мы максимизируем линейную функцию. Где линейная функция, заданная на симплексе, может достигать максимума? Множество, где она достигает максимума, — это подсимплекс. Это очевидно, поскольку линейная функция достигает максимума обязательно в вершине, но может достигать максимума на нескольких вершинах сразу. Соответственно, максимум достигается на выпуклой оболочке этих вершин, которая представляет собой симплекс. Отсюда вытекает, что $\Phi(s)$ — произведение подсимплексов симплексов $\widehat{S}_i$. То есть это не просто выпуклое множество, а более конкретное выпуклое множество, устроенное таким образом.

Поскольку функция линейная, давайте рассмотрим график этой многозначной функции:

$$G = \{(s,\, s') \colon s' \in \Phi(s)\} \subset \widehat{S}^2.$$

Утверждается, что множество G замкнуто. Это тривиальное утверждение, но хорошо иметь его в виду. Его доказательств я оставляю в качестве тривиального упражнения. Замкнутость графика означает, что если $\sigma_k \to s$ при $k \to \infty$ и $\sigma'_k \in \Phi(\sigma_k)$, причем $\sigma'_k \to s'$, то тогда $s' \in \Phi(s)$ (предел точек графика тоже принадлежит этому графику).

Точка $s \in \widehat{S}$ называется *неподвижной точкой* многозначной функции Φ, если $s \in \Phi(s)$. Очевидное утверждение состоит в том, что равновесие и неподвижная точка — это одно и то же. Действительно, $\Phi(s)$ — это все стратегии, которые противостоят s. Поэтому $s \in \Phi(s)$ в точности означает, что стратегия s противостоит самой себе.

Получается многозначное отображение с замкнутым графиком, и даже выпуклозначным (значение в каждой точке — выпуклое множество). Равновесие — это в точности неподвижная точка. Была известна следующая теорема.

Теорема 2 (Какутани, 1941). *Пусть S — ограниченное замкнутое выпуклое подмножество в $\mathbb{R}^d$, $\Phi\colon S \to S$ — многозначная выпуклозначная функция на S со значениями в S с замкнутым графиком. Тогда Φ имеет неподвижную точку.*

Исторически, Нэш не знал этой теоремы; он ее сам доказал. Теперь, я думаю, можно более или менее любому студенту Независимого университета поручить это как домашнюю работу. Ее доказательство занимает полстранички. Если к работе Нэша добавить это доказательство, все равно она не будет превосходить двух страниц.

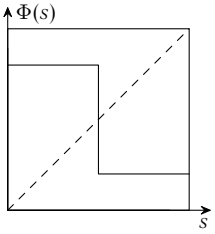


Рис. 1. График многозначной функции: 1

Давайте я приведу несколько примеров таких многозначных функций, когда $S = [0, 1]$ — отрезок. Я начну с такого замечания. Если функция Φ однозначная, то замкнутость графика и непрерывность — это одно и то же (при условии, что область определения и множество значений компактны). В этом случае теорема Какутани превращается в теорему Брауэра. Таким образом, теорема Брауэра есть частный случай теоремы Какутани. Теорему Брауэра я не буду доказывать. А теорема Какутани выводится из теоремы Брауэра.

Рассмотрим график многозначной функции, изображенный на рис. 1. Сначала график идет горизонтально, потом вертикально, потом снова горизонтально. Точка, над которой висит вертикальный отрезок, будет неподвижной. Из этого же примера вы видите такую вещь. Такая примитивная идея, как взять какую-нибудь однозначную непрерывную ветвь отображения, не проходит.

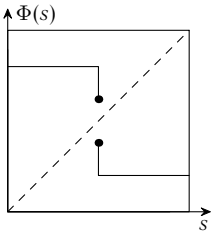


Рис. 2. График многозначной функции: 2

Давайте теперь рассмотрим график, который сначала идет горизонтально, потом вертикально вниз, затем он обрывается и где-то возобновляется (рис. 2). Это пример функции, которая имеет замкнутый график, но не выпуклозначна (значение одной точки состоит из двух отрезков). Неподвижной точки в этом случае нет. Так что в теореме Какутани что-то такое должно предполагаться.

Мы дочитали работу Нэша до конца. Он здесь объясняет, что для каждой стратегии множество всех противостоящих стратегий выпукло, т. е. отображение, которое каждой стратегии ставит в соответствие множество всех противостоящих стратегий, выпуклозначное. Используя непрерывность функции выплаты, он говорит, что график этого отображения замкнут, и он подробно объясняет, что значит замкнутость этого графика. Поскольку график

замкнут и образ каждой точки при этом отображении является выпуклым, мы заключаем из теоремы Какутани, что это отображение имеет неподвижную точку, т. е. точку, содержащуюся в ее образе. Следовательно, имеется положение равновесия. Дальше он говорит, что в теории игр двух игроков с нулевой суммой то, что называется главной теоремой в книге фон Ноймана и Моргенстерна в главе 3 (существование седловой точки) и существование положения равновесия эквивалентны. В этом случае любые два положения равновесия приводят к одному и тому же ожиданию выигрыша для игроков, но в общем случае это уже не так.

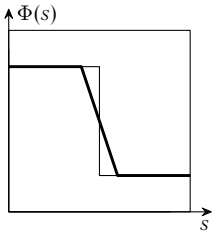
Давайте я теперь расскажу доказательство теоремы Какутани, потому что оно красивое и поучительное. Потом я вернусь к теории игр, чтобы рассказать о дилемме заключенного.

Первый шаг в доказательстве теоремы Какутани — сведéние к случаю, когда S — симплекс. Если S не симплекс, то нужно поступить следующим образом. Нужно взять симплекс S', который включает S. После этого нужно взять ретракцию $\psi\colon S' \to S$ ($\psi|_S = \mathrm{Id}_S$). Для построения ретракции можно взять внутреннюю точку множества S и рассмотреть центральную проекцию из нее, но проецировать только до S. У нас есть многозначное отображение Φ. Рассмотрим теперь многозначное отображение $\widetilde{\Phi} = \Phi \circ \psi$. Очевидно, что у $\widetilde{\Phi}$ неподвижные точки будут те же самые, что и у Φ. Свойство выпуклозначности и замкнутости графика при переходе от Φ к $\widetilde{\Phi}$ сохраняются. Кроме того, ясно, что неподвижные точки $\widetilde{\Phi}$ лежат внутри S, поскольку образ отображения $\widetilde{\Phi}$ лежит внутри S; остальные свойства сохраняются. Это простое замечание сводит ситуацию к случаю, когда S является симплексом.

Пусть теперь S является симплексом. Возьмем его подразделение $S^{(n)}$ на более мелкие симплексы. Можно, например, взять n-е барицентрическое подразделение. Важно лишь, чтобы максимальный диаметр симплексов $S^{(n)}$ стремился к 0 при $n \to \infty$. Выберем непрерывное однозначное отображение $\phi_n\colon S \to S$, которое будет аппроксимацией многозначного отображения Φ и которое будет ассоциировано с подразделением $S^{(n)}$, следующим образом. Если $s_i^{(n)}$ — какая-то вершина подразделения $S^{(n)}$, то в качестве $\phi_n(s_i^{(n)})$ мы выбираем любую точку $\Phi(s_i^{(n)})$. Множество $\Phi(s_i^{(n)})$ — это некое выпуклое множество; мы выбираем произвольную точку из этого множества. Тем самым отображение определено на вершинах симплекса. На все симплексы в $S^{(n)}$ это отображение продолжается по линейности. Получается уже однозначное отображение $\phi_n\colon S \to S$. Выбираем точку x_n, которая является неподвижной точкой этого отображения, т. е. $\phi_n(x_n) = x_n$. Это возможно по теореме Брауэра о неподвижной точке. Теперь, выбрав какую-нибудь подпоследовательность, считаем, что

$x_n \to x_0$ при $n \to \infty$. Утверждается, что x_0 будет наша неподвижная точка. Это очень естественная конструкция. Единственное, что нужно понять, это как здесь используется замкнутость графика и как здесь используется выпуклость значений в каждой точке. Это требует некоторой аккуратности. Давайте я проведу эти рассуждения. Пусть Δ_n — симплекс подразделения $S^{(n)}$, содержащий неподвижную точку x_n. Пусть, далее, x_0^n, x_1^n, …, x_d^n — его вершины. Диаметр симплекса Δ_n стремится к 0 при $n \to \infty$, поэтому $x_i^n \to x_0$ для всех i. Представим x_n в виде $x_n = \sum\limits_{i=0}^{d} \lambda_i^n x_i^n$, где $\sum\limits_{i=0}^{d} \lambda_i^n = 1$, $\lambda_i^n \geqslant 0$. Теперь, еще раз переходя к подпоследовательности, будем считать, что $\lambda_i^n \to \lambda_i^0$ при $n \to \infty$. Более того, мы можем еще считать, что $y_i^n \to y_i^0$, где $y_i^n = \varphi_n(x_i^n)$ — образы вершин. Вернемся к формуле $x_n = \varphi_n(x_n)$. Здесь $x_n = \sum\limits_{i=0}^{d} \lambda_i^n x_i^n$, а отображение φ_n линейно на каждом симплексе. Поэтому $x_n = \varphi_n(x_n) = \sum\limits_{i=0}^{d} \lambda_i^n y_i^n$. Эта сумма сходится: $\sum\limits_{i=0}^{d} \lambda_i^n y_i^n \to \sum\limits_{i=0}^{d} \lambda_i^0 y_i^0$. С другой стороны, $x_n \to x_0$. Из этого следует, что $x_0 = \sum\limits_{i=0}^{d} \lambda_i^0 y_i^0$. При этом $y_i^0 \in \Phi(x_0)$; это вытекает из замкнутости графика (заметьте, что $x_i^n \to x_0$). Теперь мы пользуемся тем, что образ точки x_0 — это выпуклое множество $\Phi(x_0)$. Значит, $x_0 = \sum\limits_{i=0}^{d} \lambda_i^0 y_i^0 \in \Phi(x_0)$. Я доказал, что x_0 — неподвижная точка. На последних двух шагах я объяснил, как использовалась выпуклость и как использовалась замкнутость графика. □

На примере, который я уже рисовал, это можно проиллюстрировать следующим образом. Представьте себе график, который идет горизонтально, потом падает вертикально вниз, а потом снова идет горизонтально (рис. 1). В соответствии с этой конструкцией предлагается действовать следующим образом. Разбить симплекс, о котором идет речь, на мелкие подсимплексы. Затем построить однозначное отображение так. Выбрать его каким угодно на мелких подсимплексах, чтобы вершины отображались в свои образы, т. е. выбрать однозначную ветвь, но только на этом конечном множестве точек. А дальше провести линейную интерполяцию. Это даст нам отображение, график которого изображен на рис. 3. Одно звено графика идет круто, но все-таки не вертикально. Тогда получается неподвижная точка. Предельная точка таких неподвижных точек — это и есть наша неподвижная точка.



Р и с. 3. График аппроксимации

Я хочу рассмотреть два примера игр. Один из них — стандартная игра с нулевой суммой, которая разбирается согласно фон Нойману. Вторая игра называется *дилемма заключенного*. Это игра с ненулевой суммой. Для нее не применима ни теория фон Ноймана, ни теория Нэша. Она показывает некую сложность того, что происходит в реальной жизни.

Рассмотрим игру со следующей таблицей выплат.

	Орел	Решка
Орел	3	−2
Решка	−2	1

Игра состоит в следующем. Есть два игрока. У каждого из них есть монетка, и на каждом шаге игры они показывают друг другу орла или решку. В таблице указан выигрыш первого игрока, которого мы назовем A. Выигрыш второго игрока равен выигрышу первого со знаком минус, поэтому в таблице достаточно указать выигрыш одного игрока. Если оба показали орла, то первый игрок выигрывает 3 у. е. Если оба игрока показали разные стороны монет, то второй игрок выигрывает 2. А если оба показали решку, то первый игрок выигрывает 1.

Давайте проанализируем, кому эта игра более выгодна. Априори это неясно. Есть ли, например, в этой игре равновесие? По крайней мере, среди чистых стратегий равновесия не видно. Скажем, первому игроку выгодно все время выигрывать 3. Но второй игрок на это, конечно, не пойдет. Если он увидит, что первый игрок всегда выбирает орла, то он будет показывать решку, и выигрывать 2. То же самое и с другими клетками таблицы; ни в одной из клеток покоя не будет.

Что будет, если каждый игрок будет показывать орла или решку с вероятностью 1/2? Как легко сообразить, вероятность каждой клетки будет 1/4. Средний выигрыш получается равным нулю. Является ли это оптимальным? Оказывается, что нет. Давайте попробуем в этом разобраться. Пусть первый игрок выбирает орла с вероятностью p, а решку с вероятностью $1-p$. Пусть второй игрок выбирает орла с вероятностью q, а решку с вероятностью $1-q$. Тогда средний выигрыш первого игрока равен

$$G(p,\,q) = 3pq + (1-p)(1-q) - 2[p(1-q) + (1-p)q] = \\ = 1 + 8pq - 3(p+q) = 8\Big(p - \frac{3}{8}\Big)\Big(q - \frac{3}{8}\Big) - \frac{1}{8}.$$

Посмотрев на эту формулу, мы видим, что у второго игрока есть следующая стратегия: взять $q = \frac{3}{8}$, т. е. выбирать орла с вероятностью 3/8 и решку с вероятностью 5/8. Тогда, что бы ни делал первый игрок, он

всегда в среднем будет проигрывать 1/8. И ничего лучшего он сделать не может. Но если второй игрок попытается это изменить в какую-то сторону, возьмет $q \neq \frac{3}{8}$, то первый игрок может изменить ситуацию в свою пользу, выбрав $p = 0$ или $p = 1$ в зависимости от знака множителя $q - \frac{3}{8}$. Равно, и первый игрок, полагая $q = \frac{3}{8}$, ограничивает свой проигрыш $\frac{1}{8}$ независимо от действий второго. Здесь точка $\left(\frac{3}{8}, \frac{3}{8}\right)$ — положение равновесия. Если вы нарисуете график функции $G(p, q)$, то это будет седловая точка.

Теперь давайте рассмотрим пример игры, которая называется *дилемма заключенного*. Если игра имеет ненулевую сумму, то рисовать табличку более сложно, даже если игроков всего двое, потому что в каждой клетке должны стоять два числа: выигрыш первого и выигрыш второго. Их часто пишут через запятую, но я напишу это следующим образом:

$-7 \setminus -7$	$-1 \setminus -9$
$-9 \setminus -1$	$-3 \setminus -3$

Есть два игрока, A и B. В классической литературе это два грабителя банка. В каждом квадрате слева записан выигрыш того, кто выбирает строку, а справа — того, кто выбирает столбец. Простейшая интерпретация здесь выглядит следующим образом. Двое неудачливых грабителей банка были пойманы при подходе к банку и арестованы. Их посадили в тюрьме в отдельные камеры. Перед ними стоит следующий выбор. Должны ли они сознаться в том, что они вместе шли грабить банк, или молчать, не сознаваться. Если они не сознаются, то они получат 3 года тюрьмы за незаконное обладание оружием. Если они оба сознаются, то каждый получит по 7 лет. Если один сознается, а другой нет, то тот, кто сознался, за помощь следствию получит всего 1 год тюрьмы, а тот, кто не сознался, получит 9 лет тюрьмы.

Другая интерпретация этого такая. Есть две страны, у которых есть атомное оружие. Они решают вопрос о том, нужно ли нанести превентивный атомный удар по своему противнику или нет. Эти цифры (если к ним добавить нуль) означают затраты (в процентах) ресурсов данной страны, который повлечет та или иная ситуация. Если обе стороны начнут атаку одновременно, то они потеряют по 70% своих ресурсов. Если они не начинают войну, то у них уходит 30% ресурсов на гонку вооружений. Страна, которая начала войну, теряет 10% ресурсов, а другая страна теряет 90% своих ресурсов.

Посмотрим на эту таблицу с точки зрения чисел. Я утверждаю, что клетка $\boxed{-7 \setminus -7}$, и только она, представляет собой равновесие. Эта клет-

ка соответствует тому, что оба игрока признались в намерении ограбить банк (или оба решили начать атомную войну). Предположим, что первый игрок отклонился от этой стратегии. Тогда второй игрок может сознаться, и это уменьшит выигрыш первого игрока. Первому игроку это невыгодно. Наоборот, положение, когда ни один из них не будет сознаваться, которое, казалось бы, им выгодно, если они между собой как-то договорились сотрудничать, не является положением равновесия. В этом положении каждый игрок может увеличить свой выигрыш, если другой игрок не меняет свою стратегию.

Положением равновесия является очень невыгодная стратегия взаимного разрушения (если рассматривать это с точки зрения атомной войны). Надо сказать, что фон Нойман, учитывая разные выводы (может быть, связанные с теорией игр, но точно это не известно), уговаривал президента Трумена начать превентивную атомную бомбардировку Советского Союза, считая, что это единственный способ спасти западную цивилизацию. Среди прочих замечательных людей его поддерживал Бертран Рассел, который потом это всячески отрицал, когда об этом говорил, но это хорошо документировано.

Как же быть с дилеммой узника? Это не так просто. Окончательного понимания ситуации с этой игрой до сих пор нет. Если эту игру можно повторять много раз (с атомной войной это не так), то выясняется, что после многократного повторения игроки учатся каким-то образом друг с другом сотрудничать. Постепенно в 60% случаев они выбирают квадратик $\boxed{-3\,\backslash\,-3}$, который, в конце концов, им выгоднее, чем то положение равновесия, о котором здесь идет речь. В Америке в Rand Corporation был такой эксперимент еще в 50-м году, который действительно показал такой результат.

Если можно выбирать смешанные стратегии, то это ничего не меняет. Для смешанных стратегий положение равновесия остается то же самое. Эта игра не очень хорошо вписывается в жесткие рамки.

Я закончу тем, что сознаюсь, какие книги и статьи я читал по этому поводу.

Литература

[1] *Nash J.* Equilibrium points in n-person games // Proc. Nat. Acad. Sci. 1950. V. 36. P. 48—49. (Русский перевод см. на следующей странице.)

[2] *Nash J.* Non-cooperative games // Annals of Mathematics. 1951. V. 54. P. 286—295.

[3] *Milnor J.* A Nobel prize for John Nash // Math. Intelligencer. 1995. V. 17, №. 3. P. 11—17.

"all4" — 2009/5/27 — 11:23 — page 80 — #80

[4] *von Neumann J., Morgenstern O.* Theory of Games and Economic Behavior. New York: Princeton Univ. Press, 1944.

[5] *Nasar S.* A Beautiful Mind. Simon & Schuster, 1998.

[6] *Poundstone W.* Prisoner's Dilemma. Doubleday, 1992.

[7] *Kakutani S.* A generalization of Brouwer's fixed point theorem // Duke Math. J. 1941. V. 8. P. 457—459.

[8] *Mehlman A.* The Gane's Afoot! Game Theory in Myth and Paradox. Providence, RI: AMS, 2000.

5 декабря 2002 г.

Приложение: перевод статьи [1]

Точки равновесия в играх с n участниками

Сообщено С. Лефшецем, 16 ноября 1949

Можно определить понятие игры с n участниками, в которой у каждого игрока есть конечное множество чистых стратегий и в которой определенное множество выплат n игрокам соответствует каждому набору из n чистых стратегий, по одной стратегии для каждого игрока. Для смешанных стратегий, которые задаются распределением вероятностей чистых стратегий, функции выплаты являются математическими ожиданиями выплат игрокам, и таким образом становятся полилинейными формами от вероятностей, с которыми разные игроки выбирают свои чистые стратегии.

Любой набор n стратегий, по одной для каждого игрока, можно рассматривать как точку в произведении пространств, полученном при умножении n пространств стратегий игроков. Один такой набор n стратегий противостоит другому, если стратегия каждого игрока в противостоящем наборе n стратегий приводит к наибольшему математическому ожиданию для этого игрока по отношению к $n-1$ стратегиям других игроков в наборе n стратегий, которому он противостоит. Набор n стратегий, противостоящий самому себе, называют точкой равновесия.

Сопоставляя каждому набору n стратегий множество противостоящих ему наборов n стратегий, получаем многозначное отображение произведения пространств в себя. Из определения противостояния мы видим, что множество точек, противостоящих одной точке, *выпукло*. Воспользовавшись непрерывностью функций выплат, мы получаем, что график этого отображения замкнут. Замкнутость эквивалентна следующему утверждению: если $P_1, P_2, \ldots$ и $Q_1, Q_2, \ldots$ — последовательности точек в произведении пространств, причем $Q_n \to Q$ и $P_n \to P$, и если Q_n противостоит P_n, то тогда Q противостоит P.

Так как график замкнут и образ каждой точки при рассматриваемом отображении выпуклый, из теоремы Какутани *) следует, что отображение

*) *Kakutani S.* Duke Math. J. 1941. V. 8. P. 457—459.

имеет неподвижную точку (т. е. точку, которая содержится в своем образе). Следовательно, существует точка равновесия.

В случае игры с двумя участниками и с нулевой суммой «основная теорема» *) и существование точки равновесия эквивалентны. В этом случае любые две точки равновесия приводят к одним и тем же математическим ожиданиям для игроков, но в общем случае это неверно.

*) *Von Neumann J., Morgenstern O.* Theory of Games and Economic Behavior. Ch. 3. Princeton: Princeton University Press, 1947. (Русский перевод: *фон Нейман Дж., Моргенштерн О.* Теория игр и экономическое поведение. М.: Наука, 1970.)

Автор благодарен доктору Дэвиду Гэйлу за предложение применить теорему Какутани, чтобы упростить доказательство, и А. Е. С. за финансовую поддержку.

В. В. Серганова

ТЕОРЕМА ЛОКАЛИЗАЦИИ И МЕТОД ОРБИТ ДЛЯ СУПЕРАЛГЕБР ЛИ

Гомоморфизм Хариш-Чандры

Сначала я сделаю краткое отступление про обычные алгебры Ли: что я, собственно, собираюсь обобщать. Пусть $\mathfrak{g}$ — полупростая алгебра Ли над $\mathbb{C}$, $U = U(\mathfrak{g})$ — ее универсальная обертывающая. У нас есть треугольное разложение $\mathfrak{g} = \mathfrak{n}^- + \mathfrak{h} + \mathfrak{n}^+$ (верхние треугольные матрицы, диагональные матрицы, нижние треугольные матрицы). Первое, что я хочу напомнить, — это описание центра Z универсальной обертывающей. Центр можно отобразить в симметрическую алгебру от $\mathfrak{h}$, т. е. существует отображение $h : Z \to S(\mathfrak{h})$ (проекция). Проекция Хариш-Чандры — это отображение $U \to S(\mathfrak{h}) = U(\mathfrak{h})$. Равенство $S(\mathfrak{h}) = U(\mathfrak{h})$ возникает из-за того, что алгебра $\mathfrak{h}$ коммутативна. Ядром этой проекции является $\mathfrak{n}^- U + U\mathfrak{n}^+$.

Теорема 1 (Хариш-Чандра). $h(Z) = S(\mathfrak{h})^W = \mathrm{Pol}(\mathfrak{h}^*)^W$, *где* W — *группа Вейля*.

Нужно учесть, что когда я говорю о действии группы W, то это — сдвинутое действие. Есть специальный элемент $\rho = \frac{1}{2} \sum_{\alpha \in R^+} \alpha$ (полусумма всех положительных корней), и действие задается такой формулой: $\lambda^w = w(\lambda + \rho) - \rho$.

Пучок скрученных дифференциальных операторов

Мы можем определить центральный характер $\chi_\lambda : Z \to \mathbb{C}$. Он определяется весом $\lambda \in \mathfrak{h}^*$. Центральный характер — это просто вычисление (evaluation). В теории представлений известно, что если вы хотите описывать неприводимые представления, то центр по лемме Шура действует центральным характером. Поэтому имеет смысл рассматривать фактор $U^\lambda = U/\big(z - \chi_\lambda(z)\big)$. Допустим для начала, что вес λ целочисленный. Это означает, что одномерное представление $\mathfrak{h}$ с весом λ интегрируется как представление соответствующего тора. Формальное определение целочисленного веса такое. В $\mathfrak{h}$ есть базис Шевалле $H_1, \ldots, H_n$. Вес λ целочисленный, если $\lambda(H_i) \in \mathbb{Z}$ для всех i.

Рассмотрим многообразие флагов $X = G/B$ (группа G односвязная). Целочисленный вес λ определяет обратимый пучок $\mathcal{O}(\lambda)$ на многообразии флагов. Мы рассматриваем пучок D^λ дифференциальных операторов на X с коэффициентами в пучке $\mathcal{O}(\lambda)$. Бейлинсон и Бернштейн называют этот пучок *пучком скрученных дифференциальных операторов*.

Этот же объект можно определить для любого λ, не только целочисленного. Есть абстрактное определение скрученных дифференциальных операторов. Это такой пучок, который локально изоморфен пучку дифференциальных операторов. Можно показать, что такие пучки нумеруются в точности весами. Но для целочисленных λ эти пучки имеют геометрический смысл.

Алгебра Ли $\mathfrak{g}$ действует векторными полями на $\mathcal{O}(\lambda)$, поэтому есть отображение $\mathfrak{g} \to \Gamma(D^\lambda)$ (в глобальные сечения). Этот гомоморфизм естественно продолжается до гомоморфизма универсальной обертывающей $\gamma_\lambda \colon U \to \Gamma(D^\lambda)$.

Т е о р е м а 2. $\operatorname{Ker} \gamma_\lambda = (\{z - \chi_\lambda(z) \colon z \in Z\})$.

Соответственно, индуцированный гомоморфизм факторов $\bar{\gamma}_\lambda \colon U^\lambda = U / \operatorname{Ker} \gamma_\lambda \to \Gamma(D^\lambda)$ — изоморфизм. То есть факторалгебра по центральному характеру — это в точности глобальные сечения пучка дифференциальных операторов.

Это первая часть теоремы. У нее есть и вторая часть (она самая важная).

Т е о р е м а 3. *Допустим, что вес $\lambda + \rho$ регулярный и доминантный (это означает, что $(\lambda + \rho)(H_i) \notin \mathbb{Z}_{\leqslant 0}$ для всех i). Мы можем рассмотреть две категории: категорию U^λ-модулей и категорию пучков D^λ-модулей. Эти категории эквивалентны.*

Функтор эквивалентности — это просто взятие глобальных сечений.

Эта теорема очень важна для теории представлений. Если оставить только условие регулярности, то будет эквивалентность производных категорий.

П р и м е р 1. Пусть $G = \mathrm{SL}(2)$ и $\mathbb{P}^1$ — флаговое многообразие. Предположим, что $\lambda = 0$. Какие мы знаем там модули? Во-первых, мы знаем тривиальный модуль. Какому D-модулю он соответствует? Он соответствует структурному пучку $\mathcal{O}$. Другой пример — модуль Верма $U(\mathfrak{g}) \otimes_{U(\mathfrak{b})} \mathbb{C}$, где $\mathfrak{b} = \mathfrak{h} + \mathfrak{n}^+$. Он соответствует D-модулю, порожденному дельта-функцией.

Для $\mathrm{SL}(2)$ есть два модуля Верма. Один модуль Верма — это $M(0)$. Его веса относительно $\mathfrak{h}$ — это струна $0, -2, -4, \ldots$ А есть другой модуль Верма, $M(-2)$. Он является подмодулем $M(0)$. Тривиальный модуль — это фактор одного модуля Верма по другому. Модуль $M(-2)$ — это и будет модуль дельта-функции.

А в общей теории модули Верма связаны с клетками Шуберта многообразия.

Чтобы получить второй модуль Верма мы выкалываем точку: рассматриваем все функции, голоморфные вне этой точки, и берем прямой образ как пучка. У него очень много сечений, в отличие от всех голоморфных функций.

Связь с орбитами коприсоединенных представлений

В нашем случае присоединенное и коприсоединенное представления изоморфны: $\mathfrak{g}^* = \mathfrak{g}$. Рассмотрим $\mathfrak{g}^*$. На нем есть стандартная скобка Пуассона — то, что называется скобкой Кириллова. А именно, $\mathcal{O}(\mathfrak{g}^*) = S(\mathfrak{g})$, а дальше коммутатор [,] задает скобку Пуассона. То есть это — пуассоново многообразие, и в нем есть симплектические листы. Симплектические листы — это то же самое, что орбиты. Рассмотрим орбиту общего положения в $\mathfrak{g}^*$ относительно действия группы G. Тогда ограничение на нее пуассоновой формы дает симплектическую структуру. Скажем, если $G = \mathrm{SL}_n$, то это будет орбита некоторой диагональной матрицы, у которой все собственные значения разные. Орбиты общего положения нумеруются тем же самым параметром; они нумеруются весами $\lambda \in \mathfrak{h}^*$.

На орбите O возникает симплектическая структура. С другой стороны, у нас есть алгебра U^λ. Получается, что U^λ — это квантование симплектической структуры на O.

Обычно, если у вас есть дифференциальные операторы, то это — квантование стандартной симплектической структуры на кокасательном расслоении. Тут структура скрученная.

Супералгебры Ли

Теперь я хочу обсудить все это для супералгебр Ли. Супералгебра Ли — это $\mathbb{Z}_2$-градуированное векторное пространство $\mathfrak{g} = \mathfrak{g}_0 + \mathfrak{g}_1$, в котором есть скобка [,], имеющая степень 0. Эта скобка обладает двумя свойствами:

$$[x, y] = -(-1)^{\bar{x}\,\bar{y}}[y, x],$$
$$[x, [y, z]] = [[x, y], z] + (-1)^{\bar{x}\,\bar{y}}[y, [x, z]].$$

Здесь $\bar{x}$ и $\bar{y}$ — степень (градуировка) элементов x и y.

Простая супералгебра Ли — это супералгебра Ли, не имеющая идеалов. Простые супералгебры описаны в статье Каца. Я дам два примера простых супералгебр, для которых я буду объяснять всю ситуацию.

Пример 2. Алгебра $\mathfrak{gl}(m|n)$. Это — алгебра автоморфизмов суперпространства блочных матриц $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$, где A — квадратная матрица порядка m (четные переменные), а D — квадратная матрица порядка n (нечетные переменные). Коммутатор определяется следующим образом:

$$[X, Y] = XY - (-1)^{\bar{X}\bar{Y}} YX.$$

Как и в обычном случае, в $\mathfrak{gl}(m|n)$ есть подалгебра $\mathfrak{sl}(m|n)$. Она состоит из матриц, у которых суперслед равен 0, т. е. $\operatorname{tr} A - \operatorname{tr} D = 0$.

Эта алгебра проста всегда, за исключением случая, когда $m = n$. Тогда у этой алгебры есть центр; по нему можно профакторизовать.

В классификации Каца есть три типа алгебр. Первый тип — это алгебры, которые можно задавать матрицей Картана, как и обычные алгебры Ли. Алгебра $\mathfrak{gl}(m|n)$ относится к этому типу. Кроме того, есть странные алгебры. А еще есть сорт алгебр, про который я сегодня не буду говорить. Это супералгебры векторных полей. Они берутся из очень понятной вещи. Что такое полиномиальные векторные поля? Это просто дифференцирования полиномиальной алгебры. Что такое супервекторные поля? Это дифференцирования суперполиномиальной алгебры, которая уже одна. Если у вас только нечетные переменные, она становится алгеброй Грассмана. Поэтому получается алгебра дифференцирований алгебры Грассмана, которая конечномерна. Это дает другие примеры алгебр, которые задаются не матрицами Картана, а строятся как алгебры векторных полей.

И есть еще интересный пример, который я, собственно, хочу обсудить. Эта алгебра обычно называется $P(n)$. Мы можем смотреть на $\mathfrak{gl}(m|n)$ как на алгебру $\operatorname{End} V$ эндоморфизмов суперпространства $V = V_0 \oplus V_1$. Рассмотрим случай, когда $\dim V_0 = \dim V_1 = n$. Тогда можно рассмотреть нечетную симметрическую форму ω (это форма, которая спаривает четные и нечетные векторы). Тогда $P(n)$ — алгебра Ли, которая сохраняет форму ω.

Вы можете построить аналог ортогональной или симплектической алгебры Ли, если будете рассматривать ортогональную или симплектическую форму. Обычную, которая сохраняет четность. Но возникает еще новая структура, которая сохраняет нечетную форму.

В матричном виде эта алгебра имеет очень интересный вид, а именно $\begin{pmatrix} A & B \\ C & -A^t \end{pmatrix}$, где $B^t = B$ и $C^t = -C$. Эта алгебра Ли не имеет инвариантной билинейной формы. Она состоит из матриц в $\mathfrak{gl}(m|n)$, удовлетворяющих таким свойствам.

Я хочу попробовать обобщить то, что я говорила сначала, скажем, для этих двух алгебр. Я начну с алгебры $P(n)$, потому что для нее вычисления

легче. Что мы можем сказать про эту алгебру? Сергеев доказал, что центр ее универсальной обертывающей алгебры U тривиален. Казалось бы, вся теория рассыпается уже в этом месте. Но тут возникает очень интересная вещь, которую я только недавно заметила. Рассмотрим, тем не менее, универсальную обертывающую U. Оказывается, что у нее радикал Джекобсона нетривиален. (Радикал Джекобсона — это аннулятор всех неприводимых представлений: все, что аннулируется любым неприводимым представлением. В данном случае это просто максимальный нильпотентный идеал.) У нас есть разложение $\mathfrak{g} = \mathfrak{n}^- \oplus \mathfrak{h} \oplus \mathfrak{n}^+$. Алгебра $\mathfrak{b}^+$ состоит из матриц вида $\begin{pmatrix} A & * \\ 0 & -A^t \end{pmatrix}$, где матрица A верхняя треугольная. Мы можем рассмотреть модуль Верма $M(\lambda) = U(\mathfrak{g}) \otimes_{U(\mathfrak{b})} C_\lambda$, где $\lambda \in \mathfrak{h}^*$.

У т в е р ж д е н и е 1. *Радикал Джекобсона J является аннулятором всех модулей Верма*:

$$J = \bigcap_{\lambda \in \mathfrak{h}^*} \operatorname{Ann} M(\lambda).$$

Доказательство этого предложения несложно.

Дальше мы делаем вот что. Изначально задача, которую я хотела решать, — описание неприводимых представлений, не только конечномерных, а любых. По определению я могу профакторизовать по радикалу Джекобсона. При этом ничего не изменится. Рассмотрим $\bar{U} = U/J$. Интересно, что у $\bar{U}$ есть центр, причем достаточно большой. Мы будем его описывать.

Пусть Z — центр $\bar{U}$. Я могу построить гомоморфизм Хариш-Чандры $h\colon Z \to S(\mathfrak{h})$ так же, как я это делала раньше. Пусть W — группа Вейля четной части $\mathfrak{g}_0$. В данном случае четная часть — это просто $\mathfrak{gl}_n$, поэтому W — симметрическая группа S_n. Тогда

$$h(Z) \cong \mathbb{C} + S^W(\mathfrak{h})\Omega,$$

где Ω — многочлен степени $n(n-1)$. В каком-то смысле это аналог многочлена Вандермонда. Алгебра $\mathfrak{h}$ состоит из диагональных матриц $\operatorname{diag}(\lambda_1, \ldots, \lambda_1)$. Многочлен Ω определяется следующим образом:

$$\Omega(\lambda) = \prod_{i \neq j} (\lambda_i - \lambda_j + j - i).$$

Он инвариантен относительно сдвинутого действия группы Вейля.

Назовем вес λ *типичным*, если $\Omega(\lambda) \neq 0$. Для любого типичного доминантного веса теорема Бернштейна—Бейлинсона про локализацию верна дословно после замены U на $\bar{U}$. В нашем случае

$$\bar{U}^\lambda = \bar{U}/(z - \chi_\lambda(z))_{z \in Z}.$$

В частности, вы можете писать все формулы инфинитезимальных характеров для любых представлений. Допустим, вы рассматриваете любое типичное весовое неприводимое представление. Тогда очень легко написать его инфинитезимальный характер.

Одно из следствий этой теоремы заключается в том, что алгебра $\bar{U}^\lambda$ является матричной алгеброй над $U(\mathfrak{g}_0)^\lambda$.

Мы в принципе понимали, что есть типичные веса, но только для конечномерных представлений, потому что никогда раньше не было центра.

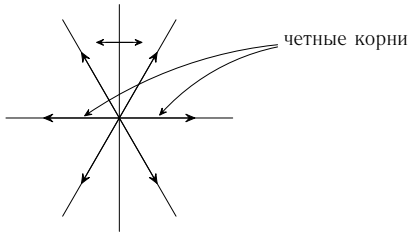
Теперь я хочу объяснить, откуда взялся этот радикал и каков его геометрический смысл. Рассмотрим $\mathfrak{g}^*$ как супермногообразие с действием супергруппы G. Первое замечание заключается в том, что $\mathfrak{g}^*$ не изоморфно $\mathfrak{g}$ как $\mathfrak{g}$-модуль. А чтобы описывать орбиты, нужно описать $\mathfrak{g}^*$. Оно состоит из матриц $\begin{pmatrix} A & B \\ C & A^t \end{pmatrix}$, где $B^t = -B$ и $C^t = C$. А действие будет тем же суперкоммутатором. Заметим, что это — супермногообразие, и в нем есть подстилающее многообразие $\mathfrak{g}_0^* \subset \mathfrak{g}^*$. Супермногообразие — это некоторая схема с нильпотентами на обычном многообразии. Обычное здесь — это просто $\mathfrak{g}_0^*$. Что такое обрита? Тут начинается такой интересный феномен супергеометрии, что супермногообразие, вообще говоря, не обязано быть объединением орбит. Если вы рассмотрите точку на подстилающем многообразии, то орбита под действием группы будет супермногообразием. Однако это вовсе не означает, что само все супермногообразие (если вы перейдете к функтору точек) будет объединением орбит. То есть неверно, что любая матрица такого вида получается из четной матрицы под действием группы G.

Например, если мы рассмотрим суперпространство $\mathbb{R}^{(0|1)}$ и рассмотрим действие $\mathbb{R}^*$ (умножение), то орбитой будет одна точка 0. Ровно такая ситуация встречается здесь. Если мы возьмем $Q = \overline{G \cdot \mathfrak{g}_0^*}$ (замыкание всех орбит всех четных матриц), то это будет супермногообразие не полной размерности; оно не совпадает с $\mathfrak{g}^*$. У этого супермногообразия есть идеал I_Q (это — аффинное супермногообразие). С другой стороны, имеется другой идеал, а именно, радикал J и можно взять $\mathrm{Gr}\, J$. Есть гипотеза, что $I_Q = \mathrm{Gr}\, J$. Я пока не могу ее доказать. Пока доказано только следующее утверждение.

Т е о р е м а 4. *Q — неприводимая компонента многообразия, ассоциированного с* $\mathrm{Gr}\, J$.

Если бы я доказала, что это многообразие неприводимо, то гипотеза была бы доказана.

Теперь я расскажу, что делать с другим примером. Мне кажется, что так можно еще многое получить, хотя это еще явно не доделано.



Р и с. 1. Корни алгебры $\mathfrak{sl}(1|2)$

Рассмотрим $\mathfrak{g} = \mathfrak{sl}(m|n)$. Прежде всего мы хотим понять, что такое центр U. Есть теорема Александра Сергеева, которая дает описание, аналогичное описанию Хариш-Чандры. Опять-таки, диагональная подалгебра — это подалгебра Картана. Однако условие, описывающее образ гомоморфизма $h\colon Z \to S(\mathfrak{h})$, другое. Прежде всего, есть группа $W = S_m \times S_n$ — группа Вейля алгебры $\mathfrak{g}_0 = \mathfrak{sl}(m) \oplus \mathfrak{sl}(n) \oplus \mathbb{C}$ (для простоты мы рассматриваем случай $n \neq m$). Легко показать, что образ обязательно должен быть инвариантен относительно этой группы. Но еще есть дополнительное условие: $h(Z)$ — это те полиномы $p \in S(\mathfrak{h})^W$, которые удовлетворяют следующему условию. У этой алгебры есть корневое разложение (разложение по собственным подпространствам относительно действия $\mathfrak{h}$). И там есть нечетные корни (они еще называются изотропными), для которых $(\alpha, \alpha) = 0$. Условие такое: если $(\lambda + \rho, \alpha) = 0$, то $p(\lambda + t\alpha) = p(\lambda)$ для любого изотропного корня α. Для алгебры $\mathfrak{sl}(1|2)$ см. рис. 1. Эта алгебра имеет ранг 2, поэтому мы можем нарисовать картинку на плоскости. Система корней фактически не отличается от системы корней для $\mathfrak{sl}(3)$. Но на эту систему корней надо смотреть в суперслучае. Квадратичная форма, которая здесь возникает, имеет сигнатуру $(+, -)$. Есть два четных корня, а все остальные корни изотропные.

В этом случае условие такое. Центр изоморфен кольцу многочленов, которые, во-первых, симметричны относительно вертикали, а кроме того, они постоянны на кресте.

Отсюда возникает определение типичного веса. Любой вес $\lambda \in \mathfrak{h}^*$ задает характер $\chi_\lambda\colon Z \to \mathbb{C}$. Вес λ типичен, если он не лежит на этом кресте, а в общем случае — если $(\lambda + \rho, \alpha) \neq 0$ для любого изотропного корня.

Некоторые точки всегда задают один и тот же центральный характер. Вес λ типичен, если только орбита группы Вейля задает один и тот же центральный характер. Для типичного λ теорема Бернштейна—Бейлинсона была доказана (в общем виде, не только для $\mathfrak{sl}(n)$) Пенковым, довольно давно. Интересно было бы понять, как устроены нетипичные веса. Очень многие интересные неприводимые представления имеют нетипичные веса.

Один из феноменов теории супералгебр Ли такой. Имеются конечномерные не вполне приводимые представления. Так же, как, например, в характеристике p. Это связано с тем, что в обычном случае на каждой орбите группы Вейля есть только один старший вес, отвечающий неприводимому конечномерному представлению, таким образом, любое конечномерное представление расщепляется в сумму неприводимых согласно действию центра универсальной обертывающей. А здесь есть много конечномерных представлений, между которыми существуют нетривиальные расширения: в нашем примере — все целочисленные веса, сидящие на одной прямой, задают конечномерные представления. До недавнего времени даже не было известно, как написать характер неприводимого конечномерного атипичного представления.

Будем говорить, что два веса эквивалентны, если они задают один и тот же центральный характер. То есть если $\chi\colon Z \to \mathbb{C}$ — центральный характер и

$$S_\chi = \{\lambda \in \mathfrak{h}^* \colon \chi_\lambda = \chi\}.$$

В нашем примере весь крест отвечает одному центральному характеру. Используя теорему Сергеева, легко описать S_χ в общем виде. Если вес λ регулярен, т. е. $(\lambda+\rho, \alpha) \neq 0$ для некоторого α, то существует ровно k ортогональных положительных изотропных корней $\alpha_1, \ldots, \alpha_k$, для которых $(\lambda+\rho, \alpha_i) = 0$. Пусть $\chi = \chi_\lambda$. Тогда S_χ — это некоторая плоскость, разнесенная действием группы Вейля: $S_\chi = \bigcup\limits_{w \in W} l_\chi$, где $l_\chi = \lambda + \mathbb{C}\alpha_1 + \ldots + \mathbb{C}\alpha_k$.

Из комбинаторики корней вытекает, что можно выбрать борелевскую систему простых корней, иначе говоря, борелевскую подалгебру $\mathfrak{b}$, для которой корни $\alpha_1, \ldots, \alpha_k$ простые. Тогда можно построить параболическую подалгебру

$$\mathfrak{p} = \mathfrak{b} \bigoplus_{i=1}^{k} \mathfrak{g}_{-\alpha_i}.$$

Эта параболическая подалгебра будет играть важную роль.

Соответственно, можно рассмотреть аналог модуля Верма:

$$\mathcal{F}_\mu = U(\mathfrak{g}) \otimes_{U(\mathfrak{p})} C_\mu.$$

А именно, модули, которые индуцированы уже не борелевской подалгеброй, а параболической. Одномерный модуль C_μ должен быть модулем над параболической подалгеброй. Вообще говоря, он существует не всегда. Но он существует для любого $\mu \in l_\chi$. Дальше я рассмотрю $U^\chi = U/(z - \chi(z))$.

Оказывается, что если мы начали с нетипичного веса χ, то U^χ имеет нетривиальный радикал Джекобсона J^χ. А именно, $J^\chi = \bigcap\limits_{\mu \in l_\chi} \operatorname{Ann} \mathcal{F}_\mu$.

Дальше естественно рассмотреть $\bar{U}^\chi = U^\chi/J^\chi$ и предположить, что у него есть центр, и центр довольно хороший. В общем виде описывать центр я еще не умею. Есть гипотеза, что центр $Z(\bar{U}^\chi)$ достаточно велик. Я могу доказать некоторые общие факты, например, что центр бесконечномерен.

Разобран случай $\mathfrak{sl}(1|n)$. Там центр описан. В этом случае $k=1$, т. е. все аффинные плоскости, о которых я говорила, — это прямые. Описание центра в этом случае получилось довольно красивое. Центр мы описываем тем же самым способом. Мы отображаем центр в полиномы, но не на всем пространстве $\mathfrak{h}$, а только на l_χ. У нас есть гомоморфизм Хариш-Чандры $h\colon Z(\bar{U}^\chi) \to \mathrm{Pol}(l_\chi) = \mathbb{C}[t]$. Для каждого характера χ на прямой l_χ есть выделенные точки $x_1, x_2, \ldots, x_{n-1}$. Рассматриваемое отображение инъективно. Полиномы в образе описываются так:

$$h(Z(\bar{U}^\chi)) = \{p(t)\colon\ p(x_i) = p(x_i+1)\}.$$

Правильнее сказать так: $p(\lambda) = p(\lambda+\alpha)$; сдвиг на 1 — это сдвиг на α.

Выделенные точки $x_1, x_2, \ldots, x_{n-1}$ зависят от λ, с которого мы начинали. Это, примерно говоря, точки, в которых нарушается регулярность веса (со сдвигом). Интересно, что иногда, например, $x_2 = x_1 + 1$. Тогда будут полиномы, имеющие одно и то же значение в трех точках. Что такое спектр? Я беру прямую и отождествляю какие-то точки. Например, для $\mathfrak{sl}(1|2)$ ответ совсем простой. У нас есть две прямые. Я выбираю одну из них. Условие на полиномы такое: $p(0) = p(1)$.

Есть плохие точки. Вне этих точек есть теорема Бернштейна—Бейлинсона. Каждая точка индуцирует новый характер. У меня был центр, я по нему факторизовала, появился радикал, я по нему профакторизовала. Появился новый центр. Теперь я беру новый центральный характер

$$\Phi_\mu\colon Z(\bar{U}^\chi) \to \mathbb{C}.$$

Если $\mu \neq \chi_i$, то имеется эквивалентность категорий $\bar{U}^\chi/\operatorname{Ker}\Phi_\mu$ и $D^\mu_{G/P}$-модулей.

Это помогает, например, написать формулу характера любого неприводимого модуля в категории O.

Вообще говоря, идея такая. Когда степень атипичности больше, этот процесс, вероятно, нужно повторять несколько раз. Сначала степень атипичности 1, и вы поступаете так. Потом будут получаться не точки, а какие-то кривые. Но пока непонятно, как в общем виде посчитать новый центр. Тут он подсчитан с помощью некоторого трюка, который нет смысла рассказывать в этом докладе.

26 декабря 2002 г.

В. В. Шехтман

ВЕРТЕКСНЫЕ АЛГЕБРЫ, СВЯЗАННЫЕ С АЛГЕБРАИЧЕСКИМИ МНОГООБРАЗИЯМИ

Мотивация того, о чем я буду рассказывать, пришла из теории струн (квантовая физика). Речь идет о некоторых моделях так называемой конформной теории поля, связанных с многообразиями. Многообразие — это то, что называется target space. По этому, так сказать, пространству-времени бегают в обычной теории частицы, а в теории струн — струны. Это — квантовая теория, поэтому с точки зрения математической мы хотели бы описать некоторое линейное пространство с операторами, которые на нем действуют. Кроме того, это — то, что называется двумерной конформной теорией поля, так что это пространство — представление алгебры Вирасоро, замечательной алгебры Ли. И эти операторы сами по себе образуют некий сорт алгебры. Эта алгебра была открыта физиками. Математики ее иногда называют вертексной алгеброй, операбельной алгеброй. Физики называют эти модели струн, бегающих по многообразию, σ-моделями.

У нас есть гладкое многообразие (у физиков оно часто с метрикой, но я буду интересоваться алгебраическим случаем). Мы хотим ассоциировать с ним такой интересный математический объект — вертексную алгебру. Она аналогична когомологиям многообразия, но эта аналогия не полная.

В первой части я бы хотел обсудить вещи конечномерные, очень классические. Речь идет вот о чем: дело в том, что явное описание σ-моделей не известно, хотя физики массу вещей о них знают и, собственно, они явились источником того, что называется квантовыми когомологиями. Однако для некоторых многообразий физики предложили ответ. Во-первых, те многообразия, о которых идет речь, в большинстве случаев суть многообразия Калаби—Яо, с тривиальным каноническим классом. И те модели, о которых пойдет речь, связаны с так называемыми гиперповерхностями Ферма.

Квинтика Ферма

Я хочу немножко поговорить о гиперповерхности Ферма, и даже совершенно конкретной гиперповерхности Ферма — квинтике Ферма. У нас есть уравнение

$$x_0^5 + x_1^5 + x_2^5 + x_3^5 + x_4^5 = 0.$$

Это гиперповерхность в $\mathbb{P}^4$, обозначим ее $\mathcal{H}$. Это будет 3-мерное многообразие. Гиперповерхности Ферма точно решаемы не только в той науке, о которой я говорил. На самом деле и для математиков они точно решаемы. А именно, еще Андре Вейль пришел к своим гипотезам о числе точек, о числе решений уравнений над конечными полями, вычисляя число решений именно таких уравнений над конечными полями. Диагонали были чуть-чуть более общего вида, но, в общем, гиперповерхности были именно такие. В свою очередь, он обобщал вычисления Гаусса. Эта гиперповерхность точно решаема в том смысле, что Андре Вейль явно вычислил ее ζ-функцию как алгебраического многообразия. Он установил, что ζ-функция имеет такой вид, как если бы на ее комплексных когомологиях действовал некий оператор с характеристическими многочленами. В частности, мы можем из вычисления Гаусса—Вейля узнать числа Бетти этой гиперповерхности. Давайте напишем числа Бетти; они будут таковы: $b_0=1,\ b_1=0,\ b_2=1,\ b_3=204,\ b_4=1,\ b_5=0,\ b_6=1$. Здесь b_3 — самое интересное. Единички тут возникают по теореме Лефшеца. А средняя группа когомологий — самая нетривиальная. Спрашивается, как получить число 204? Очень просто. Рассмотрим группу $G=(\mu_5)^5$, где μ_5 — группа корней пятой степени из 1. Эта группа действует на рассматриваемой гиперповерхности, умножая каждую координату на соответствующий корень из 1. И поэтому она действует на всем, что связано с гиперповерхностью и, в частности, на ее когомологиях. И поэтому ее когомологии, например, третьи когомологии, раскладываются по характерам нашей группы. В частности, $H^3(\mathcal{H})$ выглядит следующим образом: это — произведение собственных подпространств:

$$H^3(\mathcal{H})=\bigoplus_{\chi\in X} H^3(\mathcal{H})_\chi.$$

Во-первых, каковы характеры этой группы? Двойственная группа — это, очевидно, просто $(\mathbb{Z}/5\mathbb{Z})^5=(\chi_1,\ \ldots,\ \chi_5)$. А X — это некоторое подмножество в характерах. Согласно вычислению Вейля X состоит из таких пятерок $(\chi_1,\ \ldots,\ \chi_5)$, что $\chi_i\neq 1$ и $\chi_1\ldots\chi_5=1$. Таким образом, H^3_χ — это подпространство в H^3 таких элементов в H^3, что для любого $g\in G$ действие элемента g на элемент x есть умножение $\chi(g)$ на x, т. е. $gx=\chi(g)x$.

Нетривиальная теорема состоит в том, что имеет место такое разложение на собственные подпространства, и каждое собственное подпространство одномерно (нет никаких кратностей, все собственные подпространства одномерны). Можно взять любую статью Вейля или любую книжку по теории чисел, и вы увидите такое разложение.

Таким образом, размерность этого пространства — это число элементов этого множества; мы его можем отождествить с множеством таких пятерок чисел $(a_1, \ldots, a_5)$, что все a_i расположены от 1 до 4, а их сумма сравнима с 0 по модулю 5:

$$X = \Big\{(a_1, \ldots, a_5)\colon 1 \leqslant a_i \leqslant 4,\ \sum a_i \equiv 0 \pmod 5\Big\}.$$

Приятное комбинаторное упражнение — посчитать, что мощность этого множества равна 204.

Сейчас мы сформулируем еще более тонкое комбинаторное утверждение. Давайте поинтересуемся не просто числами Бетти, а числами Ходжа нашего многообразия. Будем считать, что все это происходит над $\mathbb{C}$. Давайте напишем числа Ходжа нашей квинтики. Через h^{pq} я обозначаю размерность пространства $H^p(\mathcal{H}, \Omega^q_{\mathcal{H}})$, где $\Omega^q_{\mathcal{H}}$ — пучок голоморфных форм. Это определение годится в алгебраической ситуации. Если хотите, h^{pq} — это размерность пространства гармонических форм типа p, q. Числа Ходжа нашей гиперповерхности выглядят следующим образом:

$$\begin{array}{ccccccc} & & & h^{3,0}=1 & & & \\ & & & h^{2,1}=101 & & & \\ h^{0,0}=1 & & h^{1,1}=1 & & h^{2,2}=1 & & h^{3,3}=1 \\ & & & h^{1,2}=101 & & & \\ & & & h^{0,3}=1 & & & \end{array}$$

Сумма $h^{3,0}+h^{2,1}+h^{1,2}+h^{0,3}$ равна 204 — сумма по столбцу. В моих обозначениях $b_i = \sum\limits_{p+q=i} h^{pq}$ — это i-е число Бетти. Это следует из теории Ходжа, и не зависит от конкретного вида нашего многообразия.

Прежде чем обсудить эти числа, я хочу сказать о том, что можно было бы назвать coté miroir — зеркальная сторона проблематики, так сказать. Для этого мы рассмотрим другой интересный алгебраический объект, тоже связанный с нашим уравнением. А именно, рассмотрим кольцо $A = \mathbb{C}[x_1, x_2, x_3, x_4, x_5]$ (кольцо многочленов от 5 переменных), и еще рассмотрим так называемое кольцо Милнора M, а именно, фактор кольца A по идеалу, порожденному частными производными f. В нашем случае это будет просто фактор A по идеалу, порожденному мономами x_i^4. Нам очень повезло, что f диагональна. Поэтому частные производные будут просто мономы. Это кольцо называется кольцом Милнора особенности $f(x) = 0$. Это конечномерный объект (конечномерное кольцо). Легко выписать его базис. Базисом может служить множество Y, состоящее из произведений $x_1^{n_1}x_2^{n_2}x_3^{n_3}x_4^{n_4}x_5^{n_5}$, где n_i будут от 0 до 3 включительно.

Заметим, что на этом кольце тоже действует группа $G = (\mu_5)^5$, которая умножает на корень из 1 пятой степени. Она тоже действует на этом кольце, потому что этот идеал сохраняется. Давайте рассмотрим диагональную подгруппу: у нас есть очевидное вложение μ_5 в свою пятую степень. В случае проективной гиперповерхности действие диагональной группы будет тривиально, поскольку если в проективном пространстве мы каждый элемент умножим на одно и то же число, то получится точка с теми же самыми координатами. А на M диагональная группа действует совершенно нетривиально. Давайте рассмотрим инварианты. Если базис в кольце инвариантов был Y, то здесь будет некоторое подмножество. Я его обозначу той же буквой X, что и в предыдущем случае. Множество X состоит из мономов $\prod x_i^{n_i}$, подчиненных тому же условию, но только теперь $\sum n_i \equiv 0 \pmod 5$. Это будут инварианты. И мы можем отождествить новое множество X с предыдущим множеством X. Только там у нас были $\{a_i\}$, где $1 \leqslant a_i \leqslant 4$ и $\sum a_i \equiv 0 \pmod 5$, а тут n_i от 0 до 3. Поэтому, если мы положим $a_i = n_i + 1$, то мы совместим предыдущее X с этим. Таким образом, оказалось, что кольцо инвариантов, по крайней мере, по размеру такое же, как пространство H^3 нашей исходной гиперповерхности. Более того, давайте рассмотрим в X подмножества X_p такие, что $\sum n_i = 5p$, где p — уже некоторое конкретное число. Какие здесь есть возможности? Сумма может быть равна 0 (таких ровно 1), а может быть равна 15 (если все троечки; таких тоже 1). Это соответствует $p = 0$ и $p = 3$. А еще сумма может быть 5 и 10. Таким образом, $\sum n_i$ может быть 0, 5, 10 и 15, и больше ничего не может быть. Легко видеть, что количество наборов, для которых сумма равна 5 и 10, равно 101:

$p=0$	$p=5$	$p=10$	$p=15$
1	101	101	1

Более того, на множестве X есть инволюция, которая n_i переводит в $3 - n_i$. Она переводит X_p в X_{3-p}. И мы получили такое интересное утверждение: мы можем описать с помощью такой простой комбинаторики не только числа Бетти, но и числа Ходжа нашего многообразия. Это замечание (глубокое замечание, на самом деле) принадлежит Дворку, который очень много изучал гиперповерхности, и примерно 40 лет назад он в своем докладе сформулировал такое утверждение. Объяснение этому состоит в том, что у этих собственных подпространств правильный тип Ходжа.

Что же это за таинственные 101 элемент? Понятно, что достаточно построить что-нибудь одно, а другое будет следовать из того, что $h^{pq} = h^{qp}$. И опять-таки это делается очень просто. Квинтика Ферма —

это многообразие Калаби—Яо. И поэтому на нем канонический пучок тривиален, т. е. $\Omega^3 \cong \mathcal{O}$. Стало быть, Ω^2 можно отождествить с векторными полями $\mathcal{T}$. Выбор формы объема (выбор изоморфизма $\mathcal{O} \to \Omega^3$) индуцирует такой изоморфизм. Сколько у нас таких выборов, столько есть таких изоморфизмов. А пространство $H^1(\mathcal{H}, \mathcal{T})$ — это пространство деформаций нашего многообразия; это касательное пространство к пространству деформаций нашей гиперповерхности. И 101 элемент этого пространства легко описать: мы можем просто к нашему многочлену $f(x)$ прибавить моном $\epsilon \prod x_i^{n_i}$ — рассмотреть сдвинутую гиперповерхность. Таких мономов будет столько, сколько надо: $\sum n_i$ должна быть равна 5 (у нас однородное уравнение степени 5), и $0 \leqslant n_i \leqslant 3$. Факт состоит в том, что все деформации этим исчерпываются.

Тут у нас неожиданно появились векторные поля, и поэтому может оказаться интересным перевернуть крестик на 90 градусов, и вместо кольца $\bigoplus H^q(\mathcal{H}, \Omega^p)$ написать следующее:

$$\begin{array}{ccccc} & & h'^{3,0}=1 & & \\ & & h'^{2,1}=1 & & \\ h'^{0,0}=1 & h'^{1,1}=101 & & h'^{2,2}=101 & h'^{33}=1. \\ & & h'^{1,2}=1 & & \\ & & h'^{0,3}=1 & & \end{array}$$

Введем штрихованные $h^{p,q}$ — перевернутые числа Ходжа: $h'^{p,q} = h^{p,3-q}$; на самом деле, это — размерность пространства $H^p(\mathcal{H}, \Lambda^q(T))$. Это означает, что вместо $\bigoplus\limits_{p,q} H^p(\mathcal{H}, \Omega^q)$ здесь рассматривается зеркально симметричное кольцо $\bigoplus\limits_{p,q} H^p(\mathcal{H}, \Lambda^q\mathcal{T})$ (когомологии с коэффициентами в q-поливекторных полях нашего многообразия). Поскольку это многообразие Калаби—Яо, Ω^q (если мы фиксируем форму объема) изоморфно $\Lambda^{3-q}\mathcal{T}$. Поэтому с точки зрения размерности тут все будет одинаково. Однако структура кольца здесь будет другая. С многообразием можно связать некое другое кольцо, которое по размерам такое же, но в нем другое умножение.

Алгебры Баталина—Вилковысского

Здесь пойдет речь про скрытую симметрию, связанную со структурами Баталина—Вилковысского (BV-структуры). Вернемся к кольцу Милнора $M = A/(\partial_i f)$. Заменим его на квазиизоморфный объект — напишем резольвенту, а именно, на него очевидным образом отображается A, дальше

будет $A^5 \to A$ — умножение на частные производные, и легко эту резольвенту продолжить:

$$0 \leftarrow M = A/(\partial_i f) \leftarrow A \xleftarrow{(\partial_i t)} A^5 \leftarrow A^{10} \leftarrow A^{10} \leftarrow A^5 \leftarrow A \leftarrow 0.$$

Разумеется, тут будут внешние степени — биномиальные коэффициенты. Это то, что называется комплексом Кошуля. Обозначим его $K(A, df)$. Как векторное пространство это будет внешняя алгебра T, где T — это пространство дифференцирований A. Это будет пятимерный свободный модуль с пятью образующими. Образующие — это дифференцирования по координатам. Дифференциал — это просто умножение на df; df — это элемент двойственного пространства, это — 1-форма, и, стало быть, он индуцирует отображение $\Lambda^i(T)$ в $\Lambda^{i-1}(T)$ — свертку с 1-формой.

Рассмотрим $\Delta = d + df$. Здесь явления разного вида: d, разумеется, не A-линейно, а умножение на df — оно A-линейно. Также, разумеется, $\Delta^2 = 0$, потому что операторы на поливекторных полях (дивергенция и свертка с df) коммутируют, и квадрат каждого равен 0; следовательно, и квадрат суммы равен 0. А кроме того на этом пространстве как на внешней алгебре есть две структуры: во-первых, есть умножение — внешнее умножение векторных полей, просто как внешняя алгебра; во-вторых, на этом пространстве есть скобка Ли. Дело в том, что T — это алгебра Ли (векторных полей). Но по нечетному тождеству Пуассона эту скобку можно продолжить на внешнюю алгебру. Это будет то, что называется скобка Схоутена. Значит, у нас есть умножение, дифференциал и скобка Ли, и эти три структуры согласованы следующим образом:

$$\Delta(xy) - \Delta(x)y - (-1)^{|x|}x\Delta(y) = (-1)^{|x|}[x, y]. \tag{1}$$

Под $|x|$ я понимаю степень элемента x. Иными словами, если бы справа стоял 0, это просто означало бы, что Δ является дифференцированием этого объекта как супералгебры. Однако это не так. И отклонение от того, что это будет дифференцированием — это и есть скобка Схоутена.

Таким образом, на кольце Милнора (на когомологиях этого объекта) эта структура незаметна, но если мы заменим кольцо Милнора резольвентой, то мы увидим скрытую структуру Баталина—Вилковысского на этой резольвенте. Структура — это умножение, скобка и дифференциал, который удовлетворяет тождеству (1). Зеркально симметричная часть скрытой структуры Баталина—Вилковысского такова. Давайте рассмотрим нашу гиперповерхность Ферма (на самом деле — любое многообразие Калаби—Яо) $\mathcal{H}$, и рассмотрим пучок голоморфных или алгебраических дифференциальных форм $\Omega_{\mathcal{H}}$. На $\Omega^*_{\mathcal{H}}$, очевидно, имеется дифференциал

де Рама $d\colon \Omega^i_{\mathcal{H}} \to \Omega^{i+1}_{\mathcal{H}}$. Если мы выберем форму объема, то тогда у нас $\Omega^i_{\mathcal{H}}$ отождествится с $\Lambda^{3-i}T_{\mathcal{H}}$. Таким образом, если мы рассмотрим алгебру поливекторных полей на нашем многообразии, то на ней тоже возникнут три структуры: структура алгебры как на внешней алгебре, структура супералгебры Ли со скобкой Схоутена и дифференциал — просто дифференциал де Рама. Это все возникает после выбора формы объема. И эти три структуры тоже удовлетворяют тождеству (1) (здесь Δ нужно заменить просто на d). Таким образом, алгебра поливекторных полей — это пучок алгебр Баталина—Вилковысского на многообразии Калаби—Яо. Если мы рассматриваем когомологии этого пучка, то эта структура как бы незаметна. Но до перехода к когомологиям у нас имеется скрытая симметрия — структура алгебры Баталина—Вилковысского. Под симметрией подразумевается какая-то скрытая структура. В данном случае структура, которая видна на комплексах, но перестает быть видной на когомологиях.

Дальше я хочу немножко продвинуться к вертексным вещам. Я хочу сформулировать одно очень простое определение — переформулировать, что, собственно, означает наличие формы объема. Это определение имеет смысл для любого гладкого алгебраического многообразия. Итак, пусть X — гладкое алгебраическое многообразие. Назовем 0-вертексным алгеброидом (точнее, структурой 0-вертексного алгеброида на пучке векторных полей) отображение $c\colon \mathcal{T}_X \to \mathcal{O}_X$ из пучка векторных полей в функции, которое удовлетворяет следующим двум свойствам:

1. $c(a\tau) = ac(\tau) + \tau(a)$ (здесь τ — это векторное поле, а a — это функция); таким образом это отображение c не $\mathcal{O}_x$-линейно (иначе мы бы получили 1-формы), а вот с таким поправочным членом — действие векторного поля на функцию.

2. Вторая аксиома — условие коцикла: $c([\tau, \tau']) = \tau c(\tau') - \tau' c(\tau)$.

Вот такое очень простое определение. Зададимся вопросом: а что, собственно, такое эта структура? Утверждается, что на самом деле эквивалентное определение очень простое: 0-вертексный алгеброид — это структура правого D-модуля на функциях. Что такое структура D-модуля? Умножение на функции у нас уже есть. Мы должны ввести умножение на векторные поля, удовлетворяющее некоторым аксиомам. А именно, имея структуру правого D-модуля на функциях, мы получаем вертексный алгеброид в смысле этого определения, просто введя c по правилу $c(\tau) = -1 \cdot \tau$ (у нас есть единица, на нее нужно подействовать векторным полем). Утверждается, что это будет правый $D_{\mathcal{T}}$-модуль тогда и только тогда, когда выполняются эти две аксиомы. Заметим, что $\mathcal{O}_X$ — канонически левый D-модуль. Разумеется, это не всегда правый D-модуль; а именно, мы знаем, что понятия правого D-модуля и левого D-модуля переводятся

одно в другое умножением на формы объема. Таким образом, это — то же самое, что структура левого D-модуля на $\Lambda^n(\mathcal{T}_X)$, где n — это размерность многообразия. Это — обратимый пучок, 1-мерный. А что такое структура левого D-модуля? Это просто интегрируемая связность. Таким образом, наш 0-вертексный алгеброид — это интегрируемая связность на пучке $\Lambda^n(\mathcal{T}_X)$. Локально (в топологии Зарисского) это существует всегда — у нас есть базис векторных полей. Как описать множество всех различных интегрируемых связностей? Различные интегрируемые связности — это то, что называется торсёр, т. е. главное однородное пространство над замкнутыми 1-формами. Если у нас есть одна такая интегрируемая связность и мы к ней прибавим ω, где ω — замкнутая 1-форма, то мы получим тоже интегрируемую связность. И разность двух интегрируемых связностей — всегда замкнутая 1-форма. Таким образом, мы можем решить задачу о том, когда, собственно говоря, это существует. Это чисто когомологическая задача. Локально они существуют, на попарных пересечениях у нас возникает разность — это будет замкнутая 1-форма, и таким образом препятствие к существованию такой структуры — это элемент группы $H^1(X, \Omega^{1,\mathrm{fer}})$ с коэффициентами в пучке замкнутых 1-форм. Мы получаем чеховский коцикл, его тривиальность равносильна существованию интегрируемой связности. Что же это за элемент? Это просто первый класс Черна $c_1(\Lambda^n\mathcal{T})$. Его еще можно обозначить через $\det(\mathcal{T})$ — максимальная внешняя степень $\mathcal{T}$. Это есть класс торсёра. Класс Черна $c_1(\det(\mathcal{T}))$ — это то же самое, что $c_1(\mathcal{T})$. Легко написать очевидную явную формулу — это просто логарифмические производные функций перехода. Это первый факт — у нас есть такое когомологическое описание. С другой стороны, элементарный общий факт из теории D-модулей состоит в том, что, имея правый D-модуль, мы можем образовать то, что называется комплекс де Рама (комплекс де Рама от $\mathcal{O}_X$). Выглядит он следующим образом:

$$\Lambda^n\mathcal{T}_X \to \ldots \to \Lambda^2(\mathcal{T}_X) \to \mathcal{T}_X \to \mathcal{O}_X \to 0.$$

Это вариант обычного комплекса де Рама, только для правого D-модуля. Как алгебра это просто внешняя алгебра, как пучок это просто внешняя алгебра; а еще тут возникает дифференциал из-за структуры правого D-модуля; и на самом деле для любого X это всегда то, что называется алгебра Схоутена, т. е. у нас есть умножение и скобка Ли, удовлетворяющая нечетному тождеству Пуассона. А структура правого D-модуля дает нам дифференциал, и все это оказывается по тем же самым причинам алгеброй Баталина—Вилковысского. И эту алгебру Баталина—Вилковысского мне хочется называть обертывающей алгеброй этого 0-вертексного

алгеброида. Это 0-вертексная алгебра — обертывающая алгебра. Вот одно определение. Это некое описание ясного понятия, только немножко на другом языке.

Вертексные алгеброиды

Теперь я хочу ввести понятие 1-вертексного алгеброида, или просто вертексного алгеброида. Взаимоотношение числа 0 с числом 1 таково: число 1 — это, так сказать, история про струны — у нас есть римановы поверхности, которые живут на многообразии, которые флуктуируют на многообразии. Число 1 отражает тот факт, что речь идет о римановых поверхностях. Если бы мы рассматривали не римановы поверхности, а как бы частицы, то это соответствует числу 0; а это будет 1. Я это определение напишу параллельно.

Прежде я сделаю одно замечание: какое отношение имеют вертексные алгеброиды к вертексным алгебрам? Я не определял понятие вертексной алгебры. Но к тем вертексным алгебрам, которые тут стоят за кадром, отношение следующее: такое же, как у пучка векторных полей к пучку дифференциальных операторов. Грубо говоря, дифференциальные операторы — это обертывающая алгебра от векторных полей. Аналогично можно ввести понятие вертексной алгебры дифференциальных операторов — это некое бесконечномерное образование. Но у некоторого разумного класса таких алгебр, а именно, алгебр дифференциальных операторов, можно выделить конечномерное подпространство, которое целиком определяет эту алгебру, так что наша алгебра является ее обертывающей. Теперь я хочу определить понятие алгеброида, который является конечномерным образованием.

Вертексный алгеброид (здесь подразумевается цифра 1) — это пара (раньше был один оператор, а 1-вертексный алгеброид — это пара) $\mathcal{A} = (\langle\ ,\ \rangle, c)$, где скобка — это билинейное отображение $\langle\ ,\ \rangle\colon \mathcal{T}_X \otimes_{\mathbb{C}} \mathcal{T}_X \to \to \mathcal{O}_X$, а c — это билинейное отображение $\mathcal{T}_X \otimes_{\mathbb{C}} \mathcal{T}_X \to \Omega^1(X)$. При этом отображение скобки симметрично, а отображение c кососимметрично. Причем эти отображения не $\mathcal{O}_X$-билинейны, так же, как раньше у нас было не $\mathcal{O}_X$-линейное отображение. Эти отображения будут на самом деле дифференциальными операторами. Раньше был один локально дифференциальный оператор порядка 1, а здесь — два дифференциальных оператора порядка 2 и порядка 3, по каждому из переменных. И если раньше у нас был один оператор, удовлетворяющий двум аксиомам, то теперь будет два оператора, удовлетворяющих трем аксиомам. Аксиомы чуть более сложные, чем были раньше.

Первая аксиома:

$$\langle a\tau,\, b\tau'\rangle - a\langle \tau,\, b\tau'\rangle - b\langle a\tau,\, \tau'\rangle + ab\langle \tau,\, \tau'\rangle = -\tau'(a)\tau(b).$$

(Как скобка ведет себя при умножении аргументов на функции.) Это означает, что скобка — бидифференциальный оператор порядка 1, 1.

Вторую и третью аксиому я хочу выписать, потому что это будет, по-видимому, единственное конкретное определение, которое действительно связано с понятием вертексной алгебры. От такой структуры можно взять обертывающую алгебру и получить вертексную алгебру. Потом их расклассифицировать и т. д. Я об этом скажу два слова.

Я сначала запишу вторую аксиому, а потом объясню обозначения:

$$\mathrm{Lie}_\tau\langle\ ,\ \rangle(\tau',\, \tau'') = \langle \tau',\, c_-(\tau'',\, \tau)\rangle + \langle \tau'',\, c_-(\tau',\, \tau)\rangle. \tag{2}$$

По определению производная Ли от оператора — это просто насколько он коммутирует с действием векторных полей:

$$\mathrm{Lie}_\tau(\tau',\, \tau'') = \tau(\tau',\, \tau'') - ([\tau,\, \tau'],\, \tau'') - (\tau',\, [\tau,\, \tau'']).$$

Если производная Ли равна 0, то это значит, что оператор коммутирует с действием векторного поля. А c_- — это немножко подправленное c, а именно,

$$c_-(\tau,\, \tau') = c(\tau,\, \tau') - \tfrac{1}{2}d\langle \tau,\, \tau'\rangle,$$

где d — дифференциал де Рама. А в формуле (2) стоит просто спаривание векторного поля с 1-формой.

Наконец, третья аксиома описывает dc (здесь d — аналог дифференциала Шевалле; я объясню, что это означает). Так или иначе, dc — это функция от 3 переменных. И она должна быть равна следующему:

$$dc(\tau,\, \tau',\, \tau'') = -\tfrac{1}{6}d\big\{\langle \tau,\, [\tau',\, \tau'']\rangle + \ldots\big\}.$$

Тут еще два слагаемых — все переставляется по циклу. Здесь

$$dc(\tau,\, \tau',\, \tau'') = d_{\mathrm{Lie}}c(\tau,\, \tau',\, \tau'') + \tfrac{1}{3}d\langle \tau,\, c(\tau',\, \tau''\rangle + \ldots$$

(плюс перестановка по циклу). И наконец,

$$d_{\mathrm{Lie}}c(\tau,\, \tau',\, \tau'') = c([\tau,\, \tau'],\, \tau'') - \tau c(\tau',\, \tau'')$$

плюс цикл.

Если вы захотите немножко поиграть с понятием дифференциала де Рама (напишите через векторные поля), вы придете к такому определению — как от функции от 2 переменных получить функцию от 3 переменных. Оказывается, что это жуткое определение осмысленное.

На самом деле оно пришло из совершенно других аксиом, из аксиом вертексной алгебры, при выделении конечномерного куска.

Теперь спрашивается: каков будет аналог теоремы классификации (насчет первого класса Черна c_1)? Интегрируемые связности или 0-алгеброиды классифицируются классом c_1. Какой же аналог этой теоремы классификации здесь? Оказывается, что имеет место теорема классификации, которая аналогична тому очень простому утверждению. Эта теорема чуточку более сложная. Такие объекты тоже существуют всегда локально; но если там нам было удобно рассматривать это просто как множество — торсёр над замкнутыми 1-формами, то 1-вертексные алгеброиды образуют категорию. Я уже не буду выписывать определение: можно ввести понятие морфизма между двумя такими объектами. И эта категория будет торсёром (в смысле категорий) над следующей категорией. Обозначим $\mathrm{Gr}(\Omega^{[2,3]})$ (от слова группоид) категорию, у которой объекты — это замкнутые 3-формы на аффинном многообразии X, а морфизмы определяются следующим образом. Нужно определить морфизм из одной формы в другую. В смысле этой категории $\mathrm{Hom}(\omega, \omega')$ — это 2-формы η, для которых $d\eta = \omega' - \omega$. Итак, объекты — замкнутые 3-формы, а морфизм, связывающий два объекта, — это такая 2-форма, что ее дифференциал — это их разность. Эта категория — группоид, т. е. каждый морфизм — изоморфизм, потому что морфизм со знаком минус — это будет обратный морфизм. Кроме того, на ней есть еще структура абелевой группы в смысле категорий, потому что мы можем объекты складывать (если есть две 3-формы, мы их можем сложить). Это очень простая коммутативная моноидальная категория.

Теперь, если мы обозначим категорию вертексных алгеброидов через Alg_X, то теорема говорит, что Alg_X — это $\Omega^{[2,3]}$-Torseur (торсёр). Что это, собственно говоря, означает? Это означает, что у нас есть действие категории $\Omega^{[2,3]}$ на категории алгеброидов. Действие — это просто функтор $\Omega^{[2,3]} \times \mathrm{Alg}_X \to \mathrm{Alg}_X$. Объяснить кусочек этого действия очень просто. Дело в том, что алгеброид — это пара, состоящая из двух операторов. Пусть у нас есть замкнутая 3-форма; оставим $\langle\ ,\ \rangle$ неизменной, а к c прибавим эту замкнутую 3-форму. Небольшое чудо заключается в том, что эти дикие аксиомы сохраняются. Таким образом мы получаем действие. И более того, если мы возьмем два таких объекта с одинаковыми $\langle\cdot,\cdot\rangle$ и двумя разными c и c', то их разность будет 3-формой, да еще замкнутой (замкнутость гарантируется одной из аксиом). Значит, это — действие. А слово торсёр означает следующее: если мы фиксируем какой-нибудь алгеброид A, то получим функтор $\Omega^{[2,3]} \to \mathrm{Alg}$; он является эквивалентностью категорий. Это первая часть теоремы, которую я хочу сформулировать.

Эту структуру можно назвать gerbe — сноп. Хотя обычное определение снопа более частное; но не важно. Факт тот, что имеется очень простая абстрактная чепуха (гомологическая алгебра), которая говорит, что, если есть такая ситуация, как у нас — торсёр, то препятствие к существованию такого объекта на многообразии — это есть некоторый класс, который лежит в группе когомологий $H^2(X, \Omega_X^2 \xrightarrow{d} \Omega_X^{3,\mathrm{fer}})$. И вторая часть теоремы, которую я хочу сформулировать, состоит в вычислении этого класса: чему, собственно говоря, он равен. И вот, имея наш конкретный торсёр, у нас возникает характеристический класс, который равен 0 тогда и только тогда, когда такой объект существует; если он равен 0, то там, соответственно, эти объекты классифицируются предыдущей группой когомологий с коэффициентами в том же самом комплексе, и группа автоморфизмов — еще предыдущей. Так вот, этот класс — это в некотором правильном смысле второй характер Черна $\mathrm{ch}_2(\mathcal{T}_X)$, или, если хотите, это более или менее класс Понтрягина $p_1(X)$.

Про класс c_1 тут имеется такое забавное дополнение к этой теореме. Оказывается, 0-вертексные алгеброиды имеют отношение к этим жутким образованиям. Откуда там c_1 появляется? Давайте я сейчас просто скажу словами. Как я уже говорил, у этих вертексных алгеброидов можно образовать обертывающие алгебры и получить пучки вертексных алгебр на многообразии. Такой пучок существует тогда и только тогда, когда первый класс Понтрягина равен 0. Дальше мы можем задаться вопросом: а когда в этом пучке вертексных алгебр есть подалгебра Вирасоро? Оказывается, что локально всевозможные вложения алгебр Вирасоро — это в точности 0-вертексные алгеброиды. Иными словами, если у нас c_1 от многообразия равно 0, то внутри этой вертексной алгебры есть подалгебра Вирасоро; и более того, там эти всевозможные подалгебры Вирасоро классифицируются всевозможными интегрируемыми связностями на детерминанте касательного расслоения. Иными словами, каждому разумному вложению подалгебры Вирасоро мы можем сопоставить дифференциальный оператор с этими свойствами. Вот такое забавное как бы наличие конформной структуры на этой вертексной алгебре измеряется 0-вертексными алгеброидами.

Я хочу сформулировать один небольшой вопрос. А именно, в связи с этой технологией возникает желание понять, что такое 2-вертексная алгебра и т. д. Например, понятно, что 2-вертексный алгеброид — это 3 оператора уже от трех переменных: один — в функции, другой — в Ω^1, третий — в Ω^2. А выписать аксиомы, которым они удовлетворяют — это очень интересная задача, с моей точки зрения; задача чисто алгебраическая. Можно написать соответствующий характеристический класс — яв-

ные формулы; там получаются классы Черна—Саймонса, совершенно явно описываемые как коциклы. Но написать такой торсёр — это уже будет не gerbe, а 2-gerbe. Было бы очень интересно придумать такое определение — чисто из алгебраических соображений. Это дало бы нетривиальный пример 2-вертексных алгебр, по крайней мере, алгеброидов — это то, что описывает поверхность, флуктуирующую на многообразии. Что такое 2-вертексная алгебра, не известно. Есть некие кандидаты на определение, но, насколько я знаю, ни одного нетривиального примера не разобрано. У Саши Бейлинсона есть некое определение, которое имеет полный смысл и для любой размерности. Это один вопрос.

Второй вопрос такой. У всей этой науки есть супераналог. Все эти определения очевидно переносятся на суперслучай. Рассмотрим многообразие X, а кольцо функций будет внешняя алгебра расслоения E, с суперградуировкой, где у E нечетная градуировка. Тогда можно поставить вопрос о существовании такого сорта объектов для такого супермногообразия. Оказывается, что соответствующее препятствие по некоторым причинам будет лежать в той же группе когомологий, но будет равняться разности класса Понтрягина (второго класса Черна) касательного расслоения $\mathcal{T}_X$ и второго класса Черна расслоения E. Из этого следуют некоторые замечательные выводы. Например, если E — это касательное расслоение, то это препятствие нулевое; т. е. на таком супермногообразии такой объект всегда существует, и более того, канонически. Кроме того, этот класс четный, поэтому замена расслоения на двойственное этого класса на самом деле не меняет (просто потому, что квадрат числа — это квадрат минус этого числа). Мы можем в качестве E взять $\mathcal{T}_X$ или Ω^1_X — двойственное расслоение. В качестве соответствующей вертексной алгебры в этом случае мы получаем то, что называется керальным комплексом де Рама, который много лет назад был открыт Федей Маликовым. А в этом случае мы получаем в качестве обертывающей алгебры алгебру, которую можно было бы назвать алгеброй керальных поливекторных полей. Это пучок вертексных алгебр с некоторыми замечательными свойствами, который существует на любом многообразии X. От него можно взять когомологии. Теперь, если X — многообразие Калаби—Яо, то по той же причине, что я говорил про Вирасоро, для Калаби—Яо этот пучок вертексных алгебр (сюда еще вкладывается супервирасоро — $N=2$-симметрия) — это алгебра Ниво—Шварца, $N=2$-вирасоро вкладывается в этот пучок керальных поливекторных полей, если X — многообразие Калаби—Яо, т. е. если на каноническом пучке есть интегрируемая связность.

Я возвращаюсь к тому вопросу, который был сформулирован в начале. Есть основания полагать, что для произвольного многообразия Кала-

би—Яо X это есть керальная часть той самой σ-модели, о которой шла речь в начале. Правда, оснований немного: посчитан характер для квинтик Ферма. Федя Маликов с Колдуновым посчитали характер и доказали, что характер равен тому же, что и в той модели, которую физики сопоставили для X — квинтики Ферма. Это эллиптический род. И на этом основании высказывается смелая гипотеза, что это — керальная часть этой самой неизведанной σ-модели.

На самом деле ее пространство состояний (пространство, о котором идет речь) — это тензорное произведение, даже сумма тензорных произведений; это половинка примерно — один сомножитель. Но так или иначе, вне зависимости от того, что такое σ-модель, если это верно, то для случая каких-то конкретных интересных многообразий (типа гиперповерхности Ферма) эта гипотеза давала бы совершенно конкретное описание таких когомологий в терминах керального аналога комплекса Кошуля соответствующей особенности. Это то, что называется соответствием между Ландау—Гинзбургом и Калаби—Яо.

Модель точно решаема, потому что все раскладывается на 1-мерные мотивы: действует группа, мотив разлагается и все 1-мерно. Поэтому все считается. И это существенно, потому что для произвольной гиперповерхности, конечно, это неверно. Дворк изучал произвольные гиперповерхности, и там массу всего интересного открыл, очень близкое, на самом деле, по духу тем вычислениям, которые здесь производятся. Там это означало то, что ζ-функция явно выписывается; все корни Фробениуса выражаются через сумму Якоби. Это соответствует тому, что когомологии разлагаются на 1-мерные мотивы.

16 января 2003 г.

А. Н. Рыбко

ПУАССОНОВСКАЯ ГИПОТЕЗА ДЛЯ БОЛЬШИХ СИММЕТРИЧНЫХ КОММУНИКАЦИОННЫХ СЕТЕЙ

Описание системы

Я сначала скажу о модели, о которой идет речь. Это очень простой, очень естественный, как мне кажется, случайный процесс. Этот процесс следующий. У нас есть m ящиков и в них разложено n частиц. Эти ящики перерабатывают частицы, и после завершения обработки частицы вылетают из ящиков. Ящики перерабатывают частицы следующим образом: в любой момент времени каждый ящик работает лишь с одной частицей (если он не пуст), а остальные частицы ждут. Когда обработка данной частицы заканчивается, то эта частица вылетает из ящика и равновероятно перекладывается в любой из m ящиков. Есть случайное время, которое очередная частица в очередном ящике обрабатывается. Это время обслуживания образует последовательность независимых одинаково распределенных случайных величин. Функция распределения этих случайных величин одинакова для всех ящиков и для всех частиц. Эта произвольная функция распределения неотрицательной случайной величины является параметром нашей задачи. Вот какую ситуацию мы рассматриваем.

Каковы примеры такой ситуации? Например, какие-нибудь служебные программы вертятся в конечном множестве компьютеров, и там с одной программой в каждом компьютере что-то происходит единовременно, а когда ее обработают, то эта программа переходит в другой компьютер. Рассматриваемая сеть симметрична в том смысле, что каждая программа равновероятно переходит к любому другому компьютеру. Мы считаем, что все компьютеры между собой связаны, и более того, они связаны так, что это — полный граф. Вот другой пример: есть 25 студентов, и они должны сдать 5 экзаменов. Они приходят к преподавателю, стоят в очереди. Преподаватель каждый раз спрашивает лишь одного студента, а остальные ждут. Каждого студента он спрашивает независимое одинаково распределенное время. Студенты никогда ничего не сдают, приходят к следующему преподавателю. В таких очередях студенты крутятся бесконечно долго.

У нас есть два целых параметра: число ящиков и число частиц, и один параметр функциональный — функция распределения времени обслуживания одного студента преподавателем (функция распределения случайной величины — времени, которое тратит преподаватель, экзаменуя одного студента).

Стационарное распределение

Возникает естественный вопрос: как узнать стационарное распределение для этого случайного процесса? Первое обстоятельство банальное: наш процесс эргодичен, поэтому у него есть единственное стационарное распределение. Почему это так? Потому что периодически возникают моменты восстановления, моменты регенерации, например, когда все частицы соберутся в один ящик. Это обязательно происходит, потому что частицы равновероятно перекладываются. Точнее говоря, легко видеть, что с вероятностью 1 наступит такое событие, когда все частицы окажутся в одной очереди в одном ящике. Одна частица там обслуживается, а остальные ждут. Когда первая частица закончит обслуживание и должна будет вылететь, то в этот момент все начинается заново. И поскольку эти события регулярно происходят, то из общих теорем легко заключить, что наш процесс эргодичен, — у него есть одна-единственная стационарная мера. При этом почти не требуется каких-либо дополнительных ограничений на параметры n, m и функцию распределения $F(x)$ времени обслуживания одной частицей. Надо только, чтобы существовал конечный первый момент у $F(x)$ — чтобы было конечным среднее время обслуживания одной частицы. Тогда эти события регенерации будут происходить регулярно. Если среднее время обслуживания одной частицы бесконечно, то тогда неясно, почему наш случайный процесс эргодичен. Пусть указанное среднее времени обслуживания конечно, чему же тогда равно стационарное распределение? Иначе говоря, с какой вероятностью, глядя на эволюцию нашего случайного процесса, через миллион лет мы увидим ту или иную конфигурацию частиц, находящихся в ящиках? Первая наивная точка зрения (сразу скажу — абсолютно неверная) заключается в том, что все конфигурации равновероятны. Это не так. Вообще говоря, при некоторых распределениях времени обслуживания $F(x)$, если посмотреть на конфигурацию почти в любом ящике, то окажется, что он будет пуст, а все частицы будут сконцентрированы лишь в малой доле ящиков, как это ни странно. Все существенно зависит от распределения времени обслуживания одной частицы. И более того, в общем-то абсолютно ясно, что сколько-нибудь явной формулы для искомого стационарного распределения найти нельзя, потому что у нас слишком много параметров.

Однако мы можем исследовать асимптотическое поведение стационарного распределения. Основная гипотеза, которая имеет место не только для этого случая, но и для многих других, заключается в следующем. Нужно перейти к термодинамическому пределу. Пусть у нас имеется какая-то положительная константа, связывающая число ящиков и число частиц. И посмотрим, как будет устроен наш процесс, когда и число ящиков, и число частиц устремится к бесконечности, а их отношение равно указанной константе. Посмотрим, во что превратится тогда наша инвариантная мера. Популярная гипотеза, которая давно сформулирована Крейнруком, заключается в том, что в этом пределе возникнет следующее. Посмотрим на один ящик. Что в него будет поступать при больших значениях n и m? В него будет приходить очень сложный поток частиц из разных других ящиков. Но, поскольку общее число ящиков огромно, то этот суммарный поток из частиц, которые приходят в фиксированный один-единственный ящик, будет «почти пуассоновским» (поскольку суммарный поток поступающих частиц складывается из очень маленьких потоков различных частиц из разных ящиков). Естественно думать, что этот поток станет пуассоновским в термодинамическом пределе.

Пуассоновский поток

Пуассоновский поток означает следующее. Мы берем прямую (ось времени t) и отмечаем на ней точки, в которых приходят частицы (моменты времени прихода частиц). Пуассоновский поток прихода этих точек по определению устроен следующим образом. Если мы возьмем два непересекающихся интервала времени и захотим узнать, с какой вероятностью у нас пришла частица в один из этих интервалов и с какой вероятностью пришла частица в другой, то мы постулируем, что эти события прихода частиц независимы (если эти два интервала не пересекаются) и, следовательно, вероятность события, что оба интервала содержат точки пуассоновского потока, является произведением вероятностей того, что первый интервал содержит точки пуассоновского потока и что второй интервал содержит точки пуассоновского потока. Это первое обстоятельство. А второе обстоятельство такое. Пусть мы хотим узнать, с какой вероятностью у нас имеется частица на каком-нибудь интервале — хотя бы одна. Когда мы устремим длину этого интервала к нулю, вероятность того, что мы имеем в нем частицу, асимптотически пропорциональна длине интервала (в пределе). Эти два свойства уже однозначно задают наш случайный процесс, и можно сказать, как он устроен. Например, можно сказать, сколько у нас частиц на каком-нибудь множестве положительной

меры (по t) — как распределено это число частиц. Это число будет распределено по Пуассону.

Термодинамический предел

Я возвращаюсь к вопросу о том, что же гипотетически у нас возникнет в пределе. Гипотетически возникнет следующее: наши частицы будут приходить в любой ящик постоянным пуассоновским потоком. Эти потоки в разные ящики будут независимы между собой, поскольку они складываются из разных частиц. Потоки частиц, поступающих в разные ящики, будут одинаковой интенсивности, потому что все ящики равноправны. И тогда имеются явные формулы в теории массового обслуживания (наверно, первым их получил Хинчин) для явных вычислений вероятностных характеристик, — сколько у нас будет частиц и как они будут разложены по ящикам. Эти формулы (для длин очередей, времени ожидания и т. д.) я не буду приводить. Они длинные, и нам не нужны. Они пишутся для производящих функций, потом от них берется преобразование Лапласа и т. д. Но это некие явные формулы, имеющиеся уже в обычной теории массового обслуживания, которая занимается вероятностными процессами, описывающими поведение одного узла.

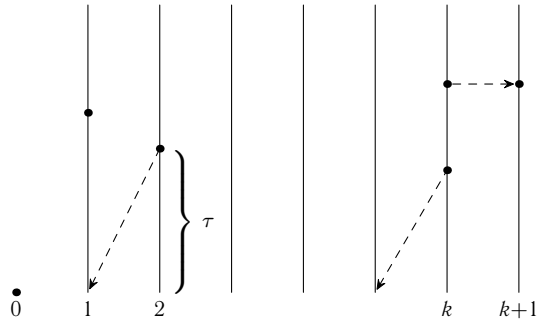
Спрашивается, как же нам доказать эту пуассоновскую гипотезу? Это давняя популярная задача. Мы занимались ею с покойным Фридрихом Израилевичем Карпелевичем много лет и доказали только слабую версию этой задачи. Я потом об этом расскажу. А теперь нам со Шлосманом удалось, наконец, ее полностью решить. Есть два побудительных мотива, почему мы так старательно ее решали. С одной стороны, была точка зрения, которую я услышал во время доклада Боровкова в Париже, лет семь назад, что это и есть центральная проблема в развивающейся теории сетей массового обслуживания.

С другой стороны, есть такая странная вещь (я о ней рассказывал на семинаре у Вершика), что это связано также и с некой интерпретацией получения специальных предельных кривых для диаграмм Юнга. Это совсем другая тематика, о которой я тоже расскажу.

Факторизация по действию группы $S(n)$ и $S(m)$

Итак, вот случайный процесс, описывающий эволюцию замкнутой сети. Что нам достаточно знать, чтобы задать марковский процесс, марковскую эволюцию нашего случайного процесса при фиксированных значениях n и m? Нужно знать, сколько частиц стоит в очереди в каждом из узлов. Это я обозначу как числа $n(i)$. Еще нам надо знать, сколько

времени уже обслуживается первая частица. Если мы этого не знаем, то мы не можем сказать, как распределено время до ее вылета; оно будет распределено по-разному. А если мы знаем время, в течение которого она уже обслуживалась, то мы знаем инфинитезимальную вероятность вылета частицы, и поэтому мы знаем эволюцию нашего марковского процесса. Но это не очень хорошее описание нашего процесса, потому что мы перенумеровали очереди. А ясно, что, поскольку у нас все симметрично, у нас на этих номерах действует инвариантным образом группа перестановок, и нам удобнее подействовать этой группой перестановок и рассматривать факторизованный процесс. Мы хотим забыть о номерах. Например, ясно, что для случая $m=2$, $n=8$ естественно рассматривать как одно и то же состояние две конфигурации, состояние, когда в первом ящике 5 частиц, а во втором 3 частицы, и они еще не начали обслуживаться, так и состояние, получившееся заменой нумерации ящиков, когда в первом ящике 3 частицы, а во втором 5 и т. д. Мы просто ввели напрасные обозначения, связанные с нумерацией ящиков, а нам нужно все задать инвариантным образом. Для этого мы подействуем группой перестановок на наше состояние и будем рассматривать вместо наборов очередей и первых времен обслуживания, орбиты, которые у нас получатся после действия группы перестановок. И тогда после этого у нас получится такая картина.



Р и с. 1. Гребенка

Описание факторсистемы

У нас есть гребенка (рис. 1), и на ней находятся красные точки. Всего красных точек столько, сколько у нас ящиков, т. е. их m штук. Каждая из этих точек имеет массу $1/m$. Эту картинку можно понимать как вероятностную меру на гребенке, потому что суммарная масса красных точек равна единице. Что значит одна точка? Например, точка на прямой

с номером 2 означает, что у нас имеется какой-то ящик, в котором 2 частицы (точка на прямой с номером k, соответственно, означает, что в каком-то одном ящике находится ровно k частиц, неважно в каком). Кроме того, расстояние τ обозначает, что первая частица из этих двух уже обслуживалась время τ. Теперь мы забыли про нумерацию ящиков, а только лишь указали, что в каком-то ящике уже находится, например, одна частица, в каком-то другом тоже одна, но в первом случае эта частица уже обслуживалась одно время, а в другом случае — другое. И еще есть пустые ящики; в пустых ящиках частицы не обслуживаются, но мы должны указать суммарную массу, — суммарное число пустых ящиков. У пустых ящиков вертикального параметра нет, но у них есть суммарная масса — это доля пустых ящиков.

Эволюция устроена так, что, во-первых, все точки на картинке, олицетворяющие ящики (кроме, конечно, тех, которые относятся к пустому ящику), едут вверх со скоростью 1. Это значит, просто, что время обслуживания первой частицы продолжается. Кроме того, они с некоторыми интенсивностями, которые легко указать, прыгают. Как они прыгают? Когда какая-то частица закончила обслуживание, соответствующая красная частица прыгает в начало предыдущей линии. Наглядно это значит, что в этом ящике стало на одну частицу меньше, следующая частица начинает обслуживаться, и соответствующая красная точка поедет от нуля вверх. А кроме того, в этот же момент, когда одна частица прыгнула вниз, мы равновероятно выбираем любую из наших красных точек (равновероятно, с вероятностью $1/m$) и перекладываем ее горизонтально, — передвигаем ее на следующую прямую. Это означает, что в соответствующем ящике стало на одну частицу больше. Вот такова наша динамика. Наконец, я должен определить, с какой интенсивностью в зависимости от τ происходят прыжки у одной частицы, связанные с окончанием обслуживания. Вот с такой:

$$\beta(\tau) = \Pr(\tau)/(1 - F(\tau)).$$

Здесь $\Pr(\tau)$ — плотность функции распределения $F(\tau)$, а $F(\tau)$ — сама функция распределения времени обслуживания одной частицы (я предполагаю, что плотность существует). Теперь я уже определил эволюцию марковского процесса для конечных значений m и n.

Описание факторсистемы (продолжение)

Теперь я еще раз повторю соображение, почему наш процесс эргодический. Я уже об этом говорил: потому что у нас через конечное время с вероятностью единица окажется такая картина, что все ящики пусты,

кроме одного, и в тот момент, когда частица в этом единственном ящике закончит обслуживание, все начнется заново. Теперь, когда мы задали эволюцию как случайную эволюцию атомарных вероятностных мер на гребенке, у нас уже от числа частиц и числа ящиков мало что зависит, поскольку эта гребенка при всех n и m останется такой же, и на ней красные точки, задающие случайную вероятностную меру, будут прыгать единообразно (лишь масса точек меняется и равна $1/m$). Нам, фактически, теперь безразлично, сколько частиц в нашей системе и сколько в ней ящиков. У этой динамики есть инвариант — среднее удаление красной частицы от начала координат (мы полагаем, что красная частица, расположенная на луче с номером k, удалена от начала координат на расстояние k независимо от значения ее координаты τ на k-й прямой). Легко видеть, что это среднее значение не меняется во времени, потому что каждый раз, когда происходит прыжок налево (на единицу), одновременно происходит прыжок направо на единицу, и среднее удаление не меняется во времени. Отсюда легко подсчитать, сколько у нас частиц:

$$n = \sum_k x(k)k,$$

где $x(k)$ — число красных точек на прямой k.

Так или иначе, наша эволюция при конечных m и n — это случайная эволюция вероятностных мер, которую я описал. Разумеется, эти вероятностные меры специальные (они состоят из атомов, причем каждый атом имеет массу $1/m$).

Пуассоновская гипотеза

Я повторю, что такое пуассоновская гипотеза, несколько точнее. Пусть m и n стремятся к бесконечности, а время t не будет пока стремиться к бесконечности. Просто частиц будет все больше (а каждая отдельная частица будет все меньшей массы), ящиков тоже все больше, и между ними имеется некая фиксированная пропорция, т. е. в пределе отношение числа ящиков к числу частиц стремится к некой положительной константе. Тогда довольно понятно, что вместо случайной эволюции вероятностных атомарных мер у нас в пределе возникнет детерминированная эволюция вероятностных мер. Потому что довольно понятно, что когда красные частицы будут все меньше и меньше (а их число будет все больше), то в пределе получится нечто вроде закона больших чисел, и вся картинка (описывающая расположение красных точек на гребенке) будет в пределе эволюционировать детерминированным образом. Это мы

доказали еще с Карпелевичем, я потом приведу список важных работ в этой области. И дальше можно изучать, уже позабыв про всякую случайность, как будет себя вести эта детерминированная предельная эволюция вероятностной меры. Пуассоновская гипотеза, о которой я говорил, эквивалентна следующему: при t, стремящемся к бесконечности, наша детерминированная динамическая система в пространстве вероятностных мер должна сойтись к некой вполне фиксированной известной мере на гребенке. Этот глобальный аттрактор нашей динамической системы является специальной вероятностной мерой, которая и отвечает тому обстоятельству, что в пределе у нас возникнет пуассоновский процесс с постоянной интенсивностью, описывающий поступление частиц в любой фиксированный ящик.

Мы должны доказать, что диаграмма

$$\begin{array}{ccc} x_{m,n} & \xrightarrow{t\to\infty} & Q_{m,n}(\cdot) \\ \Big\downarrow{\scriptstyle n,m\to\infty \atop n/m\to\rho} & & \Big\downarrow{\scriptstyle n,m\to\infty \atop n/m\to\rho} \\ x(t) & \xrightarrow{t\to\infty} & \delta_{\alpha(\rho)} \end{array}$$

коммутативна. У нас есть случайный процесс — $X_{m,n}(t)$ при фиксированном числе ящиков и фиксированном числе шариков. При t, стремящемся к бесконечности, вероятностная мера, имеющаяся в момент времени t на его конфигурациях (конфигурациях из m красных точек на гребенке), сходится к единственной вероятностной мере, поскольку он эргодичен. Эта стационарная мера обозначается через $Q_{m,n}$. Кроме того, мы с Карпелевичем доказали, что когда n и m стремятся к бесконечности, наш случайный процесс $X_{m,n}(t)$ сходится к детерминированной динамической системе $X(t)$ на пространстве вероятностных мер на гребенке. Нам нужно теперь доказать, что при t, стремящемся к бесконечности, какое бы ни было начальное состояние динамической системы $X(t)$ (вероятностная мера) $X(0)$, мы сойдемся к некоему единственному глобальному аттрактору $\delta_{\alpha(\rho)}$ — неподвижной точке для этой предельной динамической системы. Фазовое пространство динамической системы $X(t)$ — это все меры с фиксированным математическим ожиданием (средним расстоянием до начала координат) на гребенке.

Легко проверяется, что у нас действительно уже есть для нее неподвижная точка $\delta_{\alpha(\rho)}$, соответствующая пуассоновской гипотезе, но нужно доказать, что к ней-то мы обязательно и сойдемся, стартуя из любого начального состояния. И тогда из общих теорем получится, что инвариантная мера $Q_{m,n}$ обязана сходиться к этой неподвижной точке $\delta_{\alpha(\rho)}$, поскольку

"all4" — 2009/5/27 — 11:23 — page 113 — #113

верна следующая общая теорема: если у нас есть последовательность марковских полугрупп, которая слабо сходится к предельной полугруппе, тогда любая предельная точка из множества инвариантных мер для допредельных полугрупп является инвариантной мерой для предельной полугруппы. Но если инвариантная мера у предельной полугруппы одна-единственная, (если неподвижная точка $\delta_{\alpha(\rho)}$ — глобальный аттрактор), то тогда из предыдущей общей теоремы автоматически следует, что последовательность инвариантных мер $Q_{m,n}$ для допредельных полугрупп сходится к δ-мере с носителем в точке — глобальном аттракторе $\delta_{\alpha(\rho)}$, соответствующему пуассоновской гипотезе. После того как мы с Ф. И. Карпелевичем доказали слабую сходимость $X_{m,n}(t)$ к $X(t)$ на любых конечных интервалах времени, осталась самая трудная часть, которая, собственно, сейчас и сделана со Шлосманом. А именно, доказано, что у этой ужасной динамической системы есть только лишь один глобальный аттрактор — неподвижная точка $\delta_{\alpha(\rho)}$. Вот это и являлось основной трудностью.

Обзор литературы

Я расскажу о работах, которые предшествовали нашей работе.

1) Во-первых, есть замечательная работа Столяра, который доказал пуассоновскую гипотезу для одного простейшего случая, но тем не менее очень неожиданного. А именно, это случай, когда все частицы обрабатываются не случайное время, а фиксированное, одно и то же для всех частиц во всех ящиках, например, равное единице. Это случай, когда преподаватель тратит на экзамен для одного студента время, всегда равное единице. Тем не менее работа Столяра очень трудна, поэтому она, быть может, не так популярна, как того заслуживает. Там аккуратно доказано, что предельная динамическая система имеет ровно одну неподвижную точку, которая есть глобальный аттрактор $\delta_{\alpha(\rho)}$, и вся наша программа там выполнена. Причем ответ неожиданный из-за следующего обстоятельства: если у нас все частицы имеют время обслуживания единица, то ясно, что никакой поток в один ящик не может быть пуассоновским по следующей причине: как только все частицы соберутся в одном ящике, после этого уже все шаги нашего случайного процесса будут происходить через детерминированное время, равное единице, — времени обслуживания одной частицы; процесс синхронизуется. Поэтому поток частиц, который поступает в один ящик, не может быть пуассоновским. Частицы будут поступать только в целые моменты времени. А тем не менее, ответ — предельная инвариантная мера ровно такая, как будто бы этот поток был пуассоновским. Это странное обстоятельство, но, тем не менее, это так.

"all4" — 2009/5/27 — 11:23 — page 114 — #114

2) Затем была моя работа с Карпелевичем, где мы доказывали сходимость в термодинамическом пределе на конечных интервалах времени процессов $X_{m,n}(t)$ к специальной детерминированной динамической системе $X(t)$ в общей ситуации при произвольных распределениях $F(x)$ времен обслуживания.

3) Затем мы с Карпелевичем повторили аргументы Столяра и доказали пуассоновскую гипотезу в другом простом частном случае, — когда у нас время на обслуживание одной частицы не постоянно, а экспоненциально распределено — это самое простое распределение, для которого оказалось возможно доказать пуассоновскую гипотезу, используя те же соображения, которые применял Столяр. Эта работа опубликована в журнале «Markov Processes and Related Fields».

4) Оселедец и Хмелев также получили доказательство пуассоновской гипотезы при экспоненциальном распределении времен обслуживания, анализируя соответствующую бесконечномерную систему нелинейных дифференциальных уравнений, фактически не используя вероятностный анализ случайных процессов $X_{m,n}(t)$. Они тоже доказали, что все траектории динамической системы $X(t)$ сходятся к глобальному аттрактору $\delta_{\alpha(\rho)}$ в этом частном случае. Эта работа, как и две первые, опубликована в журнале «Проблемы передачи информации».

5) Наконец, есть еще одна из последних работ Добрушина в соавторстве с Карпелевичем и Введенской («Проблемы передачи информации», 1996), где рассматриваются похожие проблемы для совершенно других случайных процессов, — специальных сетей с экспоненциальным временем обслуживания.

Заметим, что доказательство того факта, что на конечном интервале времени в термодинамическом пределе мы сходимся к детерминированной динамической системе, легко обобщается на широкие классы случайных процессов с большой группой симметрии. Метод доказательства развит в нашей работе с Карпелевичем. Это верно для многих моделей среднего поля. А вот исследовать предельную динамическую систему оказалось совершенно невозможно прежними методами. В сколько-нибудь общей ситуации, когда время обслуживания не экспоненциально и не постоянно, многие годы мы не знали, как доказать глобальную сходимость $X(t)$ к неподвижной точке $\delta_{\alpha(\rho)}$. Далее я постараюсь объяснить идею доказательства этого факта.

Нелинейные марковские процессы

Теперь я скажу, что такое нелинейный марковский процесс. Это мало популярный объект по той причине, что это понятие кажется бесполезным.

Но сначала я скажу, что такое обычный марковский процесс. Это, наоборот, очень популярный объект, и, наверное, все его знают, но все-таки я о нем скажу, чтобы была видна разница. Итак, что такое марковский процесс, например, с конечным числом состояний и дискретным временем? Пусть мы имеем одну-единственную стохастическую матрицу — линейный оператор A в $\mathbb{R}^n$. Мы задаем вероятностные меры как векторы в $\mathbb{R}^n$ — как точки грани с суммой координат, равной, единице соответствующего n-мерного симплекса. Затем мы действуем оператором A^t на вероятностные меры и получаем по определению цепь Маркова с дискретным временем t. Самая известная теорема про конечные цепи Маркова такова: в общей ситуации, когда все коэффициенты матрицы A положительны, то при t, стремящемся к бесконечности, вся грань нашего симплекса, содержащая все вероятностные меры, будет сжиматься в одну точку — собственный вектор оператора A, соответствующий собственному значению 1. В случае непрерывного времени и однородного по времени марковского процесса с конечным числом состояний (также задающегося одной единственной матрицей — генератором соответствующей марковской полугруппы) теорема остается такой же.

Что же такое нелинейный марковский процесс? Это тоже известная вещь, хотя и менее популярная. Теперь у нас есть не одна стохастическая матрица, а бесконечно много, и каждая матрица $A(\mu)$ индексируется вероятностной мерой μ. Теперь нелинейная марковская цепь $X(t)$ с дискретным временем t задается индукцией по времени следующим образом: пусть к моменту времени t нам уже известна вероятностная мера $\mu(t)$, тогда, чтобы задать распределение $\mu(t+1)$ процесса $X(t+1)$, мы (по определению) полагаем $\mu(t+1) = A(\mu(t))[\mu(t)]$. Мы смотрим, какая мера у нас уже возникла в момент t, и в зависимости от этой меры берем свою матрицу $A(\mu(t))$ и ею действуем на меру $\mu(t)$.

Конструкция для нелинейного марковского процесса с непрерывным временем усложняется, поскольку уже невозможно строить процесс индукцией по t. Однако, тот факт, что в каждый момент времени мы применяем к вероятностной мере $\mu(t)$ инфинитезимальный оператор $A(\mu(t))$, зависящий от самой меры $\mu(t)$, сводится к исследованию соответствующей нелинейной системы дифференциальных уравнений.

Обычные системы уравнений Колмогорова для марковских процессов с непрерывным временем выглядят так. В левой части каждого уравнения стоит производная по времени вероятности находиться в момент t в какой-то фиксированной точке фазового пространства. Эта производная равна некой фиксированной линейной форме (зависящей от точки фазового про-

странства) от вероятностей в момент t. Таким образом, для марковских процессов с непрерывным временем и конечным числом состояний число линейных дифференциальных уравнений в этой системе равно числу точек в фазовом пространстве. Знаменитая теорема Маркова, о которой я упоминал, гласит, что в общем положении при t, стремящемся к бесконечности, в этой ситуации имеет место глобальная сходимость к фиксированной инвариантной мере.

Для нелинейного марковского процесса, из-за того что наша матрица зависит от самой меры в момент времени t, у нас в правой части соответствующей системы дифференциальных уравнений будут стоять нелинейные формы. Например, во многих случаях будут стоять полиномы 2-й степени от наших вероятностей. И тогда никакой содержательной теории, касающейся поведения меры $\mu(t)$ при t, стремящемся к бесконечности, в общей ситуации не может быть, просто потому что, например, асимптотическое поведение системы дифференциальных уравнений, у которых справа стоят полиномы 2-го порядка даже, например, от 5 переменных весьма нетривиально. У нас уже неверно, что при t, стремящемся к бесконечности, даже если у нас только 5 состояний, то типично мы сойдемся к неподвижной точке. Могут возникать самые разные аттракторы. Специалистам это обстоятельство хорошо известно.

Описание термодинамического предела

Когда задан случайный процесс, ему соответствует эволюция мер. Как правило, эта эволюция мер детерминированная. Сам процесс случайный, но его вероятности во времени эволюционируют детерминированным образом. И в нашем случае происходит то же самое. Мы могли бы на этом месте забыть про случайный процесс, и говорить просто о нелинейной динамической системе. Но это будет очень плохо, потому что тогда надо изучать эти ужасные уравнения — нелинейные уравнения динамической системы, описывающие детерминированную эволюцию на бесконечном пространстве вероятностных мер, с которыми аналитически разобраться нельзя. А если рассматривать случайный процесс, мера для которого описывается этими нелинейными уравнениями, то он оказывается достаточно наглядным и с ним удобнее работать. В пределе при n, m, стремящихся к бесконечности, наша динамическая система $X(t)$, описывающая детерминированную эволюцию мер, соответствует следующему нелинейному марковскому процессу.

У нас в пределе при n и m, стремящихся к бесконечности, было бесконечно много ящиков и бесконечно много частиц; а вместо этого мы

рассматриваем один-единственный ящик, в который поступает пуассоновский поток частиц интенсивности $\lambda(t)$. Возможность заменить бесконечно много ящиков одним-единственным ящиком связана со специальной двойственностью: дело в том, что поведение во времени доли ящиков (из бесконечного множества всех ящиков) с фиксированной структурой очереди частиц совпадает с вероятностью в тот же момент времени увидеть ту же самую очередь в одном-единственном ящике. Эта двойственность связана с инвариантным действием нашей группы перестановок на состояниях процесса. Интенсивность $\lambda(t)$ этого пуассоновского потока переменная, она будет зависеть от t.

Теперь давайте определим, что такое интенсивность выхода частиц из нашего ящика. Частицы в ящике стоят в очереди и обслуживаются независимо, с одинаково распределенным временем (с распределением $F(x)$), которое у нас уже встречалось. Определим формально интенсивность $b(t)$ выхода частиц следующим образом. Возьмем маленький интервал времени $\Delta(t)$ после момента t. Будем считать, что у нас уже задана вероятностная мера $\mu(t)$ на очередях в нашем единственном ящике в момент t для процесса $X(t)$. Зная $\mu(t)$, мы можем найти среднее по мере $\mu(t)$ число частиц, вылетающих из нашего ящика на интервале $[t, t + \Delta(t)]$. Поделив это среднее на $\Delta(t)$ и устремив $\Delta(t)$ к нулю, мы по определению и получим $b(t)$. Теперь закончим конструкцию — определение нелинейного марковского процесса $X(t)$, потребовав дополнительно, чтобы

$$\lambda(t) = b(t).$$

Заметим, что это определение дает нам именно нелинейный марковский процесс: действительно, инфинитезимальные вероятности прихода новых частиц $\lambda(t)$ в момент времени t зависят от всей меры $\mu(t)$ на очередях для нашего процесса $X(t)$.

Теорема о единственном аттракторе

Ранее в нашей работе с Карпелевичем доказано, что случайная эволюция вероятностных мер (при конечных n и m) сходится к именно такой детерминированной эволюции мер $\mu(t)$, которую задает нелинейный марковский процесс $X(t)$, который мы только что определили. Сходимость подразумевается в следующем смысле. Любой точке гребенки (l, τ) естественно соответствует точка фазового пространства нелинейного марковского процесса $X(t)$. Это просто означает, что точке (l, τ) соответствует очередь длины l и время τ, в течение которого первая частица в этой

очереди уже обслуживалась (для нашего единственного ящика). Меру процесса $X(t)$ опять естественно рассматривать как меру на этой же гребенке. По построению, специальные атомарные меры $\mu_{m,n}(t)$, соответствующие случайному процессу $X_{m,n}(t)$ и эволюционирующие случайным образом, тоже живут на этой же гребенке. Наша с Карпелевичем теорема заключается в следующем: если при n, m, стремящихся к бесконечности, последовательность начальных состояний $\mu_{m,n}(0)$ слабо сходится к $\mu(0)$ — начальному состоянию процесса $X(t)$, то на любом конечном интервале времени $[0, T]$ марковские процессы $X_{m,n}(t)$ сходятся к динамической системе на пространстве мер, соответствующей нелинейному марковскому процессу $X(t)$.

Теперь нам нужно доказать для этой нелинейной эволюции, что при t, стремящемся к бесконечности, $\lambda(t)$ стремится к константе. Основная наша задача теперь свелась к следующему. У нас есть уже нелинейный марковский процесс $X(t)$, описывающий эволюцию очередей в одном ящике, у него есть детерминированная эволюция мер, и нужно доказать, что $\lambda(t)$ (интенсивность внешнего потока) стремится к константе при t, стремящемся к бесконечности. С какой стати это происходит? Основная мысль следующая. Пусть у нас есть неоднородный по времени марковский процесс с произвольной функцией, задающей интенсивность поступления пуассоновского потока новых частиц $\lambda(t)$. Тогда частицы поступают в наш ящик с интенсивностью $\lambda(t)$ и вылетают с какой-то интенсивностью $\beta(t)$ (конечно, она не обязана быть равной $\lambda(t)$). Естественно предположить, что наш узел обладает следующим свойством: функция $\beta(t)$ (то, что выходит из узла) будет «регулярней», чем функция $\lambda(t)$ (то, что в него поступает). Формально нужное нам неравенство между входом и выходом означает, что осцилляция функции $\lambda(t)$ должна быть не меньше (а лучше больше), чем осцилляция функции $\beta(t)$.

Представляется, что все «нормальные» устройства, которые использует человек, в каком-то смысле должны улучшать выход по сравнению с входом, чтобы устройство устойчиво работало. И если действительно это так, то тогда уже несравненно легче увидеть, что при t, стремящемся к бесконечности, если мы имеем дополнительно равенство, что $\lambda(t) = \beta(t)$, как в нашем уравнении для процесса $X(t)$, то тогда можно надеяться, что при t, стремящемся к бесконечности, у нас $\lambda(t)$ сойдется к константе.

У нас действительно имеется свойство сглаживания для $\beta(t)$, верное даже для произвольного неоднородного по времени марковского процесса с входом $\lambda(t)$, задающего поведение очередей в одном ящике. Оказывается, что если вы задали какое-то произвольное $\lambda(t)$, то можно найти такое весьма замысловатое и сложное вероятностное распределение, которое

зависит от момента времени t и от самой функции $\lambda(t)$, что наш выход $\beta(t)$ будет сверткой $\lambda(t)$ с этим вероятностным распределением.

Это удивительное обстоятельство и дает нужное нам свойство сглаживания. В нашей работе с С. Шлосманом нужная нам формула свертки доказывается следующим образом. Сначала выписывается чрезвычайно сложная и громоздкая формула, дающая явное аналитическое выражение выхода $\beta(t)$ через вход $\lambda(t)$. В этом выражении имеются комбинаторные коэффициенты и свертки с различными ядрами. Удивительным образом оказывается, что эти комбинаторные коэффициенты таковы, что конечный результат становится именно вероятностным ядром. После того как находится такое вероятностное ядро, доказывается громоздко, но уже не так неожиданно, что при t, стремящемся к бесконечности, $\lambda(t)$ сходится к константе. Тем самым пуассоновская гипотеза доказана. Я не буду приводить здесь длинную и сложную формулу для вероятностного ядра $q_{(\cdot)}(t)$, но я объясню комбинаторную лемму, благодаря которой она находится, и приведу доказательство этой леммы.

Лемма о палочках

Я сейчас расскажу про удивительную комбинаторную формулу (по-видимому, новую — мы спрашивали многих специалистов), благодаря которой корректно определено вероятностное ядро $q_{(\cdot)}(t)$.

Прежде чем сформулировать комбинаторную лемму, приведем следующую геометрическую конструкцию: пусть у нас есть произвольные $n-1$ точек на прямой R: $X_{n-1} = x_1, \ldots, x_k, \ldots, x_{n-1}$. На прямой имеется выделенная точка, например, начало координат — точка 0. В нашем распоряжении также имеется n фиксированных ориентированных палочек с произвольными длинами $L_n = l_1, \ldots, l_i, \ldots, l_n$. Каждая палочка имеет левый конец — основание и правый — стрелку. Мы делаем следующее: сначала мы берем из множества $l_1, \ldots, l_n$ произвольное подмножество A_{n-1}, содержащее $n-1$ палочку. Затем мы выбираем произвольное взаимно однозначное отображение $G(l_i)$, $l_i \in (A_{n-1})$, $G(l_i) \in X_{n-1}$, между точками $x_1, \ldots, x_{n-1}$ и выбранными палочками из множества A_{n-1}, и основания палочек из множества A_{n-1} помещаем в точках $x_1, \ldots, x_{n-1}$, соответствующих образу этих палочек $G(A_{n-1})$. Вспомним, что поскольку множество $L_n = l_1, \ldots, l_i, \ldots, l_n$ содержало n палочек, а отображение $G(A_{n-1})$ использовало лишь $n-1$ палочку, у нас еще осталась одна свободная палочка l, $l \in L_n$, основание которой мы помещаем в произвольную точку x_0 на прямой R. Теперь положим все палочки с закрепленными основаниями из множества $G(A_{n-1}) \cup (x_0)$ горизонтально, таким образом, чтобы

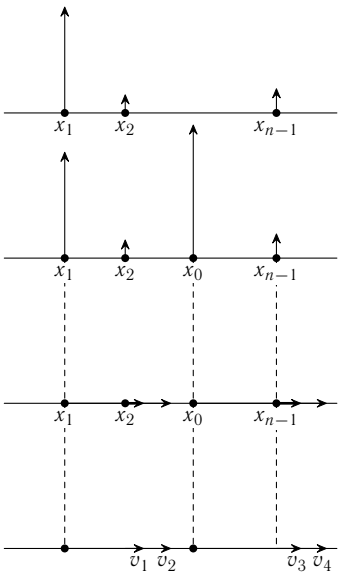


Рис. 2.

$y(\cdot)$ — стрелки (правые концы палочек) — находились в соответствующих точках прямой R: $y_0 = x_0 + l$, $y(G(l_i)) = x(G(l_i)) + l_i$, $l_i \in A_{n-1}$. После этой процедуры могут возникнуть конфликты: на прямой R часть палочек может налагаться друг на друга. Разрешим эти конфликты следующим образом: подвинем палочки, участвующие в конфликтах, направо на минимальные расстояния так, чтобы порядок на прямой R их оснований совпадал с порядком образов X_{n-1} этих оснований при отображении $G(A_{n-1})$ (см. рис. 2). Обозначим положение на прямой R оснований палочек после разрешения конфликтов через $Z_n = z_1, \ldots, z_i, \ldots, z_n$, а, соответственно, положение стрелок после разрешения конфликтов, через $V_n = v_1, \ldots, v_i, \ldots, v_n$, где $v_i = z_i + l_i$, $i = 1, \ldots, n$.

Наглядно все эти объекты означают следующее: у нас имеется n частиц, причем каждой частице i требуется время обслуживания, равное длине соответствующей палочки l_i. Множество X_{n-1} — это множество моментов времени, когда $n-1$ из этих n частиц поступают в узел. Отображение $G(A_{n-1})$ сопоставляет длины палочек — времена обслуживания частицам, поступившим в узел в моменты времени из множества X_{n-1}, а оставшаяся частица длины l поступает в узел в произвольный момент времени x_0. Множество Z_n — это множество моментов времени, когда частицы начинают обслуживаться, а множество V_n — это множество моментов времени, когда частицы вылетают из узла. Нас интересуют события, когда в выбранный фиксированный момент времени 0 какая-либо частица вылетает из узла (осуществление этого события зависит от выбора отображения $G(A_{n-1})$ и от выбора момента поступления свободной палочки — точки x_0). Иначе говоря, нас интересует, принадлежит ли точка 0 множеству V_n при заданных множестве X_{n-1} и множестве L_n.

Теперь пора сформулировать основное комбинаторное утверждение.

Т е о р е м а. *При произвольном выборе множества X_{n-1} и множества L_n возьмем число K разных способов выбора взаимно однозначного отображения $G(A_{n-1})$ и положения основания оставшейся свободной палочки l, при которых $0 \in V_n$ (какая-либо стрелка совпала с 0). Это число K не зависит ни от X_{n-1}, ни от L_n и всегда равно $n!$.*

"all4" — 2009/5/27 — 11:23 — page 121 — #121

З а м е ч а н и е. Строго говоря, мы исключаем из рассмотрения конфигурации меры 0, когда после отображения $G(A_{n-1})$ и разрешения конфликтов лишь для палочек из множества A_{n-1} сложилась такая редчайшая ситуация, что точка 0 совпала с моментом окончания обслуживания какой-либо частицы из самого множества A_{n-1}. Поскольку свободная палочка l в этом событии никак не участвует, то мы можем ее расположить на континууме свободных мест на R так, чтобы она не сдвигала конфигурацию и никак не влияла на расположение остальных палочек из A_{n-1} в этом случае.

В качестве иллюстрации утверждения теоремы рассмотрим случай $n=2$, и пусть $x_1=-1$, причем одна из палочек l_1 очень короткая, например, $l_1=1/10$, а другая l_2 — длинная, например, $l_2=5$. Если, выбирая множество A_1, мы выберем длинную палочку l_2 и положим ее основание в (единственную) точку $x_1=-1$, то ее стрелка окажется в точке 4. Ясно, что каким бы образом мы затем ни выбирали на прямой R точку x_0 — положение для основания оставшейся короткой палочки l_1, мы никак не добьемся того, чтобы какая-либо из двух стрелок из множества V_2 совпала с 0, поскольку длинную палочку, перекрывающую 0, мы можем при помощи короткой палочки подвинуть направо не более чем на ее длину $l_1=1/10$.

Если же в качестве множества A_1 мы выберем короткую палочку l_1, то найдется ровно два положения $x_0=-5{,}1$, $x_0=-5$ для основания свободной длинной палочки, при которых точка 0 совпадет с какой-либо стрелкой. В первом случае точка 0 совпадет со стрелкой сдвинутой короткой палочки, а во втором — со стрелкой длинной палочки. Таким образом, число $K=2=2!$, и в этом частном случае теорема верна.

Доказательство комбинаторного утверждения

Сначала заметим, что если длины всех палочек L_n малы, то наше комбинаторное утверждение очевидно верно: действительно, если, например, суммарная длина всех палочек из множества L_n меньше, чем минимальное расстояние между точками из множества $X_{n-1}\cup 0$, то мы можем разложить произвольным образом $n-1$ палочку в точках X_{n-1}, выбрав произвольное множество A_{n-1} и произвольное отображение $G(A_{n-1})$ и положив оставшуюся свободную палочку длины l основанием в точку $x_0=-l$. Тогда для любой такой конфигурации отсутствуют конфликты, поскольку длины всех палочек очень малы. Ясно, что стрелка свободной палочки l совпадает с точкой 0 для этой конфигурации. Также ясно, что число возможностей выбора множеств A_{n-1} и отображений $G(A_{n-1})$ в точности равно $n!$. Таким

образом, наше комбинаторное утверждение очевидно верно в ситуации, когда все палочки достаточно малы.

Теперь доказательство будет устроено следующим образом: мы убедимся, что, зафиксировав множество X_{n-1}, взяв вначале произвольный набор (длинных) палочек L_n и постепенно уменьшая длины палочек из этого набора, мы, тем не менее, не меняем при этом число K. Ясно, что этого факта достаточно, поскольку если он верен, то, уменьшая длины всех палочек, мы в конце концов придем к ситуации, когда все палочки очень малы, число K не изменилось, и, следовательно, всегда равно $n!$.

Вначале сформулируем следующую простую и полезную лемму:

Л е м м а. *Пусть для фиксированного множества X_{n-1} мы уже выбрали и зафиксировали множество палочек A_{n-1} и отображение $G(A_{n-1})$. Спрашивается, сколькими способами $K(l, G(A_{n-1}))$, мы можем расположить оставшуюся свободную палочку l на прямой R так, чтобы $0 \in V_n$? Ответ следующий*:

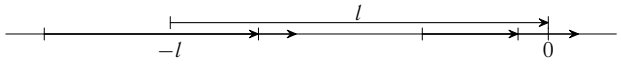
1) *Расположим палочки из множества A_{n-1} на прямой в соответствии с отображением $G(A_{n-1})$.*

2) *Разрешим все конфликты для этого семейства палочек $G(A_{n-1})$, раздвинув конфигурацию, содержащую лишь $n-1$ палочку на минимальные расстояния так, чтобы никакие палочки из семейства $G(A_{n-1})$ не пересекались на R. Обозначим получившуюся конфигурацию через $B(G)$.*

3) *Положим основание оставшейся свободной палочки длины l в точку $-l$.*

4) *Обозначим через U число стрелок в конфигурации $B(G)$ на интервале $(-l, 0)$.*

Тогда искомое $K(l, G(A_{n-1})) = U$, если точка $-l$ не лежит внутри какой-либо палочки конфигурации $B(G)$, и $K(l, G(A_{n-1})) = U + 1$, если точка $l-1$ свободна (не накрывается палочками из конфигурации $B(G)$, см. рис. 3).



Р и с. 3.

Д о к а з а т е л ь с т в о. Посмотрим, как мы можем расположить свободную палочку l, чтобы в точке 0 оказалась одна из стрелок конфигурации $B(G)$. Ясно, что, расположив основание свободной палочки l в подходящем месте x левее точки $-l$, мы можем сдвинуть направо конфигурацию

$B(G)$ на произвольное число y между l и 0. Таким образом, мы можем при помощи свободной палочки l подвинуть конфигурацию $B(G)$ направо так, чтобы произвольная стрелка в конфигурации $B(G)$, находящаяся на интервале $(-l, 0)$, попала в 0. Кроме того, если точка $-l$ свободна (сама не содержится внутри какой-либо палочки из конфигурации $B(G)$), то, расположив в самой точке $-l$ основание свободной палочки, мы ее стрелкой попадем в точности в 0. Лемма доказана.

Закончим теперь доказательство комбинаторного утверждения. Для этого достаточно доказать, что при уменьшении какой-либо палочки число K не меняется. Пусть, например, мы зафиксировали A_{n-1} и подстановку — отображение $G(A_{n-1})$, а затем стали «медленно» уменьшать длину оставшейся свободной палочки l. Тогда число $K(l, G(A_{n-1}))$, входящее слагаемым в K, не меняется, пока соответствующее количество стрелок из $B(G)$ на интервале $(-l, 0)$, принадлежащих кластеру из палочек $B(G)$, и «занятость» точки $-l$ остаются неизменным.

Согласно предыдущей лемме, в тот момент, когда точка $-l$ пересекает точку v, совпадающую с какой-либо стрелкой конфигурации $B(G)$, число $K(l, G(A_{n-1}))$ уменьшается на 1, если интервал $(v, v+\varepsilon)$ не пуст (покрывается палочками из конфигурации $B(G)$). Согласно той же лемме число $K(l, G(A_{n-1}))$ также уменьшается на 1, когда в процессе уменьшения свободной палочки точка $-l$ съезжает со свободного места и впервые попадает внутрь кластера, состоящего из палочек конфигурации $B(G)$. Рассмотрим более подробно, например, именно этот последний случай.

Укажем другое множество палочек $\widetilde{A}_{n-1}$ и отображение $G(\widetilde{A}_{n-1})$, для которого в тот самый момент, когда $K(l, G(A_{n-1}))$ уменьшилось на 1 (при уменьшении свободной палочки), а $K(\widetilde{l}, G(\widetilde{A}_{n-1}))$ увеличилось ровно на 1. Эти $\widetilde{A}_{n-1}$ и $G(\widetilde{A}_{n-1})$ строятся по A_{n-1} и $G(A_{n-1})$ следующим образом. Возьмем ту палочку $\widetilde{l}$ из множества A_{n-1}, на которую «наехала» точка $-l$ при уменьшении длины свободной палочки l. Заменим в множестве A_{n-1} и отображении $G(A_{n-1})$ палочку $\widetilde{l}$ на палочку l. Теперь для нового множества $\widetilde{A}_{n-1}$ и отображения $G(\widetilde{A}_{n-1})$ палочка $\widetilde{l}$ стала свободной палочкой, а «уменьшившаяся» палочка l оказалась на месте палочки $\widetilde{l}$. Тогда, используя предыдущую лемму, нетрудно заметить, что число $K(\widetilde{l}, G(\widetilde{A}_{n-1}))$ увеличилось на 1.

Чтобы завершить доказательство, нужно перебрать конечное множество других случаев: когда, например, уменьшая свободную палочку, мы уменьшаем число стрелок в конфигурации $B(G)$ на интервале $(-l, 0)$ и когда мы уменьшаем длину какой-либо «занятой» палочки из самого множества A_{n-1}. Все эти случаи разрешаются аналогично: строится

конфигурация $G(\widetilde{A}_{n-1})$, для которой число $K(\widetilde{l}, G(\widetilde{A}_{n-1}))$ возрастает на 1 тогда, когда $K(l, G(A_{n-1}))$ уменьшилось на 1, и наоборот. Правило выбора нового $\widetilde{A}_{n-1}$ и отображения $G(\widetilde{A}_{n-1})$ остается неизменным: нужно заменить прежнюю свободную палочку l на соответствующую «занятую» палочку из кластера конфигурации $B(G)$. Подробности этой громоздкой процедуры я опускаю.

К сожалению, невозможно объяснить за оставшееся время, каким образом, используя предыдущее комбинаторное утверждение про $n!$, строится вероятностное ядро $q_{(\cdot)}(x)$ для нужной нам формулы, связывающей интенсивность входа и интенсивность выхода частиц одного узла:

$$\beta(t) = \int_{\propto}^{t} \lambda(x) q_{(\cdot)}(x) dx.$$

Также невозможно сколь-нибудь подробно объяснить доказательство нужного нам факта, что для нелинейного марковского процесса $X(t)$, $\lambda(t)$ сходится к пределу при t, стремящемся к бесконечности. Все же укажем некоторые основные шаги доказательства этого основного утверждения.

Сначала доказывается, что семейство вероятностных ядер $q_{(\cdot)}(x)$ (зависящих от момента времени t) компактно по t: для любого ε найдутся такие T и x, что для любого $y > x$

$$\int_{y-T}^{y} q_{(\cdot)}(v) dv > 1 - \varepsilon.$$

Затем доказывается, что функция $\lambda(t)$ медленно меняется по t: какие бы ни были ε и T, существует такое t, что для любой точки $\widehat{t} > t$ функция $\lambda(u)$ на интервале $(\widehat{t}, \widehat{t} + T)$ меняется не больше, чем на ε.

И, наконец, тот факт, что $\lambda(t)$ сходится к пределу при t, стремящемся к бесконечности, доказывается, используя предыдущее утверждение. Действительно, если бы на сколь угодно длинных интервалах времени T функции $\lambda(t)$ были бы равны с точностью до малых ε разным константам, то к моментам окончаний этих больших интервалов времени среднее число частиц в нашем узле было бы различным. А построенная динамика нелинейного марковского процесса $X(t)$ такова, что из-за равенства

$$\lambda(t) = \beta(t)$$

среднее число частиц в нашем узле остается неизменным по времени. Таким образом, предположив, что $\lambda(t)$ не имеет предела, мы приходим к противоречию.

"all4" — 2009/5/27 — 11:23 — page 125 — #125

Связь с диаграммами Юнга

Теперь я в заключение скажу про обстоятельство, которое связывает эту картину с диаграммами Юнга. Это, собственно, тривиально по следующим соображениям. Когда мы раскладываем наши частицы по ящикам, это не что иное, как то, что мы раскладываем в сумму (из m слагаемых) наше общее число частиц n. Если мы упорядочим эту сумму по убыванию слагаемых, то получим диаграмму Юнга. Доказанная пуассоновская гипотеза — теорема о том, что в пределе при n, m, стремящихся к бесконечности, случайная мера на диаграммах Юнга сходится к специальной «детерминированной» мере, связанной с пуассоновской гипотезой, — может трактоваться в духе результатов, связанных со специальным законом больших чисел для диаграмм Юнга.

В замечательных работах Вершика, Керова, Ольшанского и др. рассматриваются различные естественные меры на конечных диаграммах Юнга (равномерная мера, мера Планшереля и т. д.) и доказывается, что после естественного скейлинга при n, стремящемся к бесконечности, случайные диаграммы сходятся к нетривиальной предельной кривой.

Здесь у нас похожая картина: в пределе при n, m, стремящихся к бесконечности, n/m, стремящемся к константе, соответствующие случайные меры сходятся к единственной предельной мере.

В отличие от обычной ситуации, когда изучается «статическая» картина — мера на диаграммах Юнга — и затем рассматривается соответствующий предельный переход, в нашем случае оказалось полезно добавить еще одно измерение — время, — и рассматривать соответствующую динамику: перекладывать частицы, находящиеся в ящиках, и тем самым случайно менять слагаемые в диаграммах Юнга. Изучение этой динамики помогло доказательству нужного нам закона больших чисел.

Литература

[1] *Введенская Н. Д.*, *Добрушин Р. Л.*, *Карпелевич Ф. И.* Система обслуживания с выбором наименьшей из двух очередей — асимптотический подход // Пробл. передачи информ. 1996. Т. 32, № 1. С.20—34.

[2] *Карпелевич Ф. И.*, *Рыбко А. Н.* Асимптотическое поведение симметричной замкнутой сети массового обслуживания в термодинамическом пределе // Пробл. передачи информ. 2000. Т. 36, № 2. С. 69—96.

[3] *Столяр А. Л.* Асимптотика стационарного распределения для одной замкнутой системы обслуживания // Пробл. передачи информ. 1989. Т. 25, № 4. С. 80—92.

[4] *Karpelevich F. I.*, *Rybko A. N.* Thermodynamical limit for the mean field model of simple symmetrical closed queueing network // Markov Processes Related Fields. 2000. V. 6, № 1. P. 89—105.

[5] *Rybko A. N., Shlosman S. B.* Poisson hypothesis for information networks (a study in non-linear Markov processes). Part I // Moscow Mathematical Journal. 2005. V. 5, № 3. P. 679—704.

[6] *Rybko A., ShlosmanS. B.* Poisson hypothesis for queueing Networks — Combinatorial Aspects // Problemy Peredachi Informatsii. 2005. V. 41, № 3. P. 51—57.

[7] *Rybko A., Shlosman S.* Poisson hypothesis for information networks (a study in non-linear Markov Processes). Part II // Moscow Mathematical Journal. 2005. V. 5, № 4.

13 февраля 2003 г.

С. Н. Артемов

ИНТУЦИОНИСТСКАЯ ЛОГИКА С ТОЧКИ ЗРЕНИЯ КЛАССИЧЕСКОЙ

Введение

Сначала в докладе будет неформальное обсуждение, а потом мы перейдем к математике. Но я сразу хочу вас предупредить, чтобы у вас не сложилось ложного впечатления, что всё это философия, не очень оформленная и легкая. Во-первых, речь будет идти о проблемах, которые волновали математиков в течение десятилетий. Во-вторых, сейчас это серьезная наука с приложениями.

В связи со сменой тысячелетия разные организации проводили опросы, публиковали разные списки. В частности, «Time» (по-видимому, один из самых респектабельных в мире журналов общего характера) опубликовал свой список. «100» означает 100 самых великих людей по всем наукам, по всем областям деятельности вообще, включая финансистов, актеров и — ученых. Всего 20 позиций было отдано на науку и технологию вместе взятые. По технике это изобретение аэроплана, ракеты, телевизора, транзистора, интернета; по биологии и медицине 4 позиции: психоанализ, изобретение пенициллина, ДНК и вакцина против полиомиелита; физика фактически имела 2 позиции, которые выиграли Эйнштейн и Ферми, и Хаббл — по астрономии. Математика, компьютерная наука и философия — каждая получила по 1 позиции; все три позиции выиграли логики.

Вообще, всякие ранги — кто в математике лучше, кто более великий — это несерьезно, как и в любой другой науке. Если мы спросим у алгебраического геометра или специалиста по динамическим системам — у них будет своя точка зрения, кто был самым великим в этом столетии. Но тем не менее, нам, математикам, неплохо быть в курсе того, как вся эта картина, весь этот универсум математики, взаимосвязанный и богатый, на самом деле выглядит снаружи, так сказать, с точки зрения остальной цивилизации.

Обложка этого журнала выглядит так. Фрейд дает сеанс психоанализа Эйнштейну. Рядом Карсон, которая поняла, что ДДТ — это нечто ужасное; она считается основателем науки об окружающей среде. Вот Курт Гёдель — его знаменитый снимок: он на фоне пустой доски в своем кабинете

в Принстоне, в Институте высших исследований. Это — Аллан Тьюринг — отец компьютерной науки. В его представлении сказано про связь с логикой и математикой. И Витгенштейн, которого люди в философии знают очень хорошо; в описании его вклада тоже математика и тоже логика. Я хочу показать, что на самом деле у логиков есть, что еще предложить миру.

В логике можно условно выделить три традиции. Я назову несколько имен, хотя в каждой из них десятки великих имен. Классическая традиция связана с именами Фреге, Гильберта, Гёделя, Тарского. Конструктивная — Брауэр, Гейтинг, Колмогоров, Гёдель. И есть еще «традиция явного выражения» — Сколем, Гёдель, Карри, Черч — которая по самой своей сути связывает логику, а через нее значительную часть математики, с современной компьютерной наукой. Я скажу несколько слов, характеризующих эти направления. Общее впечатление от занятия логикой состоит в том, что оно является довольно благодарным делом, если вы готовы интересоваться логикой широко, если вы ее понимаете в естественно-научном смысле, я бы сказал: в смысле колмогоровской традиции. В логике у вас есть хороший шанс увидеть результаты своего труда не только опубликованными, но и использованными какими-то другими серьезными группами людей, в частности, в компьютерной науке самым фундаментальным образом, и я дам вам хорошие примеры этого.

Классическая логика

Пропозициональная логика — это обычная булева логика (и, или, не), в которой, тем не менее, кое-что можно выразить, например, контактные схемы. С точки зрения логической это — небольшая теория, которая, конечно, непротиворечива и, конечно же, разрешима — всегда можно понять, формула выполнима или нет. Однако даже на этом уровне вычисление выполнимости уже не является легким занятием. Именно в пропозициональной логике формулируется одна из известнейших в математике проблема $\mathrm{P} = \mathrm{NP}$? — одна из проблем стоимостью в миллион долларов; в данном случае речь идет не о деньгах, конечно, а о значимости этой проблемы.

Обратной стороной или, если угодно, слабостью пропозициональной логики являются ее чрезвычайно ограниченные выразительные возможности. Обычных булевых переменных и этих связок (конъюнкция, дизъюнкция, отрицание) безнадежно мало для того, чтобы выразить что-нибудь существенное. А если удается выразить, то обычно выражение, адекватное представление задачи, бывает чрезвычайно длинным.

Логику первого порядка обычно отождествляют с логикой вообще,

потому что исследования в XX веке, в основном, концентрировалось вокруг логики первого порядка. В частности, теоремы Гёделя о полноте, о неполноте, теоремы о компактности, теория моделей — в значительной степени всё это сделано в рамках логики первого порядка. Это связано в первую очередь с математической традицией. Где-то сто лет назад люди в математике были озабочены тем, чтобы выразить как можно больше через как можно меньшее количество понятий. Пропозициональной логики не хватало, а вот логики первого порядка с одним сортом объектов (например, переменные, которые бегают по всем натуральным числам, или по множествам, или по элементам свободной группы), и где есть кванторы $\forall$ и $\exists$ в дополнение к обычным булевым связкам, уже хватало. Этого языка действительно хватает для того, чтобы выразить всю математику. Например, можно, взять теорию множеств, где единственный 2-местный предикат принадлежности $X \in Y$ — означает, что X является элементом Y, и при наличии кванторов и обычного логического материала можно, в принципе, выразить всю математику; по крайней мере, формальную ее часть.

Обратной стороной этого подхода является то, что для выражения рабочих математических конструкций требуется колоссальное количество кодирования. Это кодирование с точки зрения математики выглядит невинным. Если угодно, это было предметом гордости, что мы можем закодировать чрезвычайно сложные конструкции через очень небольшое количество понятий в относительно бедном языке. Однако кодирование разрушает попытки что-то вычислить или хотя бы формализовать. Например, известно, что кодирование по Гёделю выводов через натуральные числа имеет гиперэкспоненциальную сложность, и поэтому какое-нибудь невинное выражение длиной в строчку, если его записать как натуральное число, примет совершенно астрономический характер.

Еще один пример: разумеется, всю арифметику можно выразить через теорию множеств. Но представьте себе, что мы хотим записать условие типа $2+2=4$ на языке теории множеств. Что такое равенство? В теории множеств два множества равны, когда они состоят из одних и тех же элементов: мы определяем равенство $X = Y$ как $\forall Z\,(Z \in X \leftrightarrow Z \in Y)$. У нас появляется квантор всеобщности по всем множествам! И как мы его понимаем вычислительно, если у нас модель, в принципе, бесконечная? Это равенство — равенство просто множеств, а надо еще выделить равенство тех множеств, которые кодируют натуральные числа. В общем, естественная попытка записать, скажем, арифметику через теорию множеств приводит к совершенно разрушительным последствиям с точки зрения вычислимости, потому что даже равенство становится неразрешимым понятием. Таким образом, предмет гордости логики начала века, а именно, экономность

средств, из которых можно вывести всю математику, на самом деле сейчас является не достоинством, а, скорее, недостатком этой системы.

Я много раз делал доклады по логике в математической аудитории, и часто задавался вопрос: «А почему логика первого порядка? Почему мы должны ограничиваться только натуральными числами? Что, если я говорю: „натуральные числа", — я уже не могу говорить о функциях? Я не могу говорить о натуральных числах и о действительных числах? Или я не могу говорить о множествах натуральных чисел — внутри одной и той же теории?» В реальной математике, конечно, используют много сортов объектов, их столько, сколько надо. Формальным эквивалентом этого «столько, сколько надо» является понятие логического языка высшего порядка, например, 2-го. Этот формализм гораздо более близок к математической практике, и, мало того, практически все современные системы формализации и проверки доказательств достаточно общего характера используют языки высшего порядка.

Я помню, Алёша Сосинский когда-то читал в интернате спецкурс «Математические уроды и парадоксы». Среди уродов и парадоксов были кривая Пеано, ковер Серпинского и — теорема Гёделя о неполноте, которая тоже получила статус урода или парадокса.

Итак, когда мы естественно переходим к логике второго порядка, к языку второго порядка, всё это прекрасное здание логики, практически все эти замечательные достижения разваливаются. Все перечисленные здесь классические теоремы логики перестают быть верными при этом переходе.

Модальная логика

Я помню, когда в 1982 году А. Н. Колмогоров стал заведующим кафедрой математической логики, он меня из Стекловки позвал в университет на полставки. Я спросил у Андрея Николаевича совета, что делать и как делать. Я сначала отвечу на «как делать?». Вопрос был такой: подводить ли людей в течение спецкурса к какому-то кругу задач, чтоб они могли самостоятельно работать, или — брать широко и образовывать людей? Ответ был, разумеется, королевский: «Надо делать и то и другое» — сказал Андрей Николаевич. На второй вопрос — «Что читать?» — он сказал: «Вы знаете, Сережа, я хочу, чтобы вы учили наших студентов модальной логике — это та область, которую я не знаю», что тоже было по-своему королевским ответом.

Итак, модальная логика занимает промежуточное положение между логикой пропозициональной и логикой первого порядка. И это положение

связано с тем, что мы добавляем некоторое количество кванторов к пропозициональной логике, но немного — так сказать, ровно столько, сколько нам нужно, мы не пытаемся делать язык абсолютно универсальным. То есть кванторы появляются, но они спрятаны внутри некоторых конструкций. И в частности, в самом базисном случае мы добавляем всего одну новую связку к языку, которая традиционно обозначается $\Box$ и читается как «необходимо» ($\Box F$ — условие F необходимо). Мы тем самым различаем условие F, которое говорит, что F истинно, и $\Box F$ — это более сильное условие.

Есть два взаимосвязанных, но, в принципе, разных прочтения этого языка. Во-первых, временно́е, когда $\Box F$ означает, что F не только верно сейчас, но условие F будет верно при всех возможных развитиях ситуации. В этом плане модальная логика ухватывает динамику. Оказалось, что эта идея, когда можно выразить истинность не только сейчас, но и, так сказать, в будущих мирах, играет важную роль в попытках придумать язык для, скажем, верификации программ. И вот это темпоральное или динамическое прочтение модальности оказывается существенным для моделирования процесса вычисления. Тем более, когда у нас процесс вычисления — ветвящийся процесс, и хотелось бы выразить условие, например, что какое-то другое условие появляется бесконечно часто, или что-нибудь в этом роде, и оказывается, что простыми комбинациями этих апелляций к «всегда в будущем» это можно выразить. Положительной стороной является то, что модальная логика остается, как правило, разрешимой и обладает рядом других приятных с точки зрения пользователя свойств.

Второе прочтение модальной логики состоит в том, что $\Box F$ (связка «необходимо F») понимается как состояние знания. А именно, $\Box F$ означает, что F известно конкретному агенту, конкретной персоне (в отличие от «F истинно», где никаких предположений, кому это известно, нет).

Интуиционизм

Конструктивный подход к математике родился в одна тысяча девятисотых годах. И в начале века, в основном усилиями Брауэра, интуиционизму удалось привлечь к себе существенное внимание. Стоял вопрос, не является ли интуиционистская математика другим, более правильным и, разумеется, новым подходом к математике вообще. Наиболее активные сторонники интуиционизма предлагали вообще заменить классическую математику интуиционистской, потому что классическая математика изначально была построена на неправильных основаниях. Однако посмотрим, в чем состояла положительная программа интуиционизма. Она состояла, в частности, в следующем.

Интуиционизм утверждает, что говорить об истинности бессмысленно, если мы не указываем, кому и каким образом эта истина известна. И поэтому, как говорят интуиционисты, всюду в определении истинности мы должны в индуктивном определении валидности формулы вместо «истинно» говорить «доказуемо».

Конъюнкция доказуема (мы сейчас к этому перейдем), когда оба члена конъюнкции доказуемы; дизъюнкция доказуема тогда и только тогда, когда хотя бы один из дизъюнктивных членов доказуем.

И наконец, как понимать импликацию? Здесь проявилась элегантная гениальность А. Н. Колмогорова, который в 1932 г. сказал в явном виде, как надо понимать импликацию, если мы хотим провести в жизнь ту точку зрения, что для интуиционизма истинность — это доказуемость. Итак, читаем импликацию по Колмогорову. Доказательством импликации $A \to B$ является конструкция, которая по данному доказательству A вычисляет доказательство B. Тем самым мы не просто утверждаем, что между A и B есть связь: мы считаем, что эта связь должна носить вычислительный характер. А именно, доказательством импликации, свидетельством того, что импликация верна, является конструкция, которая вычислимым образом перерабатывает доказательство A в доказательство B. Прочтение импликации оказалось функциональным, и это наблюдение оказывает фундаментальный эффект на области, связанные с компьютерной наукой.

Таким образом, мы не можем говорить о валидности формулы, если у нас нет доказательства этой формулы. Поэтому истинностные таблицы для интуиционистской логики скорее являются таблицами вычисления доказательства. И поэтому аналогом обычных булевых таблиц истинности для классической логики в интуиционистской логике являются так называемые условия ВНК (Брауэра, Гейтинга и Колмогорова).

Брауэр дал некоторое общефилософское описание, Гейтинг начал выписывать конъюнкцию и дизъюнкцию, и А. Н. Колмогоров в 1932 г. сделал ключевой шаг, поняв, что импликацию надо тоже понимать функционально. Если вы думаете, что этот набор элементарных условий удалось тут же перевести на формальные математические рельсы, как все ожидали, как все хотели, то это не так. История формализации интуиционистской семантики заняла много времени.

Интерпретации интуиционизма

Семантика — вещь формальная. Можно, глядя на аксиомы геометрии, думать о стульях и столах, как мы хорошо знаем. Глядя на аксиомы Лобачевского, можно думать, конечно, о дугах внутри большего круга, а можно понимать их в более общем смысле.

"all4" — 2009/5/27 — 11:23 — page 133 — #133

Вот далеко не полный список того, какие интерпретации были найдены для интуиционистской логики, ни одна из которых не является формализацией изначальной семантики BHK.

1. Алгебраическая семантика (Биркгоф, 1935).
2. Топологическая семантика (Стоун, 1937; Тарский 1938).
3. Семантика реализуемости (Клини, 1945).
4. Модели Бета (Бет, 1956).
5. Диалектическая интерпретация (Гёдель, 1958).
6. Изоморфизм Карри—Ховарда (1958).
7. Логика задач (Медведев, 1962).
8. Модели Крипке (1965).
9. Интерпретация доказуемости (Кузнецов—Муравицкий—Гольдблатт, 1976).
10. Категорная семантика (1979).

Есть хорошо известная теоретико-множественная семантика обычной классической логики, состоящая в том, что каждое высказывание интерпретируется как множество, конъюнкция интерпретируется как пересечение множеств, дизъюнкция — как их объединение, отрицание — как переход к дополнению. Разумеется, есть некоторый универсум U, из которого черпаются данные множества, и критерием валидности формулы F является то, что при любой теоретико-множественной интерпретации входящих в него переменных соответствующее множество равно всему универсуму. Если угодно, каждой формуле мы приписываем множество истинности.

Аналогом такой теоретико-множественной интерпретации в интуиционистском случае является так называемая топологическая интерпретация. У нас уже не хватает просто множества с обычными теоретико-множественными операциями, а мы хотим, чтобы это множество имело структуру топологического пространства. Если есть топологическое пространство, то мы можем интерпретировать интуиционистские связки.

Множеством истинности для логической константы «ложь» является пустое множество, конъюнкция и дизъюнкция понимаются обычным образом. Оценкой импликации является следующее: мы должны взять обычную теоретико-множественную трактовку импликации, (отрицание посылки объединенное с заключением) и потом взять внутренность (в смысле топологического пространства) получившегося множества. Критерием валидности, или общезначимости, является то же самое условие, что оценкой формулы всегда является всё топологическое пространство.

Например, давайте проверим закон исключенного третьего $A \vee \neg A$. Пусть A будет интервалом (0, 1), а $\mathcal{T}$ — это вся действительная прямая.

Поскольку $\neg A$ есть $A \to \bot$, к этой формуле применимо правило интерпретации импликации. Нетрудно видеть, что интерпретацией $\neg A$ будет объединение $(-\infty, 0)$ и $(1, \infty)$. Топологическим значением формулы $A \vee \neg A$ будет являться прямая без концевых точек интервала. Тем самым закон исключенного третьего не является интуиционистской тавтологией.

Как вы понимаете, топологическая семантика оказывается весьма далека от изначального прочтения интуиционистской логики как специального исчисления, которое занимается доказательствами.

А. Н. Колмогоров в 1985 г., комментируя свою старую работу 1932 г., отметил, что когда он ее писал, он надеялся, что логика доказательств, которая фактически закодирована в условиях BHK, станет составной частью обычного курса логики в университете. Шанс на это остается, но ждать пришлось довольно долго.

Интуиционистская логика дает более богатую формальную систему, чем логика классическая. А именно, интуиционистская логика тривиальным образом содержит классическую логику, которую легко эмулировать внутри интуиционистской. Обратное, вообще говоря, не верно: при любой попытке интерпретировать интуиционистскую систему внутри классической (разумеется, при каких-то элементарных условиях сохранения смысла), если формализовать задачу, мы тут же получаем доказательство невозможности это сделать.

Некоторое время назад я позвонил знаменитому математику, с которым мы давно знакомы, и, условно говоря, сказал ему: «Иван Иванович, Вы знаете, там в старой задаче по семантике интуиционистской логики получилось продвижение, теперь мы знаем, как придать смысл тому, что...» На это последовал ответ, что в 40-х годах эту проблему решили: модель интуиционистской логики — это топологическое пространство. Для нормального математика вопрос был закрыт, хотя я надеюсь, что я дал представление о том, что на самом деле топологическая интерпретация — это корректная, но абсолютно искусственная семантика, которая к изначальным задачам Гёделя, Колмогорова и Брауэра, не имеет никакого отношения.

Система *S*4

Гёдель был первым, кто предпринял попытку дать точную математическую формулировку интуиционистской истинности как доказуемости. Он заметил, что поскольку у нас в самом определении интуиционистской истинности встроено понятие доказуемости, то, по-видимому, классическая математика, классическая логика в состоянии будет это точно

формализовать, если мы будем иметь механизм, который всюду вместо истинности будет говорить «доказуемо». Например, если мы в язык логики введем понятие доказуемости как дополнительную связку.

Гёдель использовал механизм модальной логики, который позволяет добавить новую логическую связку доказуемости к обычной классической логике. Гёдель сформулировал простую с виду систему, которая называется $S4$ — не спрашивайте, почему она называется $S4$, это отдельная история, не имеющая большого смысла.

1. *Классические аксиомы и правила*
2. $\Box(F \to G) \to (\Box F \to \Box G)$
3. $\Box F \to F$
4. $\Box F \to \Box\Box F$
5. *Правило интернализации*: $\dfrac{\vdash F}{\vdash \Box F}$

Давайте попытаемся прочесть эти условия с точки зрения доказуемости и понять, что они означают содержательно. Ну, например, второе условие означает следующее: если доказуемо, что F влечет G, и доказуемо F, то, конечно, доказуемо G. Третье условие — это фактически условие корректности: если доказуемо F, то F истинно (иначе зачем мы еще доказываем). Далее, последнее — это довольно хитрое условие: если доказуемо F, то доказуемо, что доказуемо F. Давайте попробуем понять это. $\Box F$ означает, что у F есть доказательство; а что такое доказательство — это некий конечный объект (конечная последовательность формул, дерево, или 400 страниц в «Известиях Академии Наук» — нечто конкретное), и более того, мы предполагаем, что доказательства допускают проверку, т. е. глядя на доказательство, мы в состоянии понять, доказательство это или нет. Это, если угодно, одно из требований к системе. Искать доказательство — это отдельная история. Итак, представьте себе, что у нас есть доказательство, тогда у нас есть возможность проверить это доказательство и получить утверждение, что да, действительно, вот результат проверки, доказательство верное. Следовательно, у нас есть доказательство того, что F доказуемо.

Наконец, есть правило усиления, которое говорит, что если F в рамках этой формальной системы выводимо, то, конечно, этот вывод сам по себе является достаточным основанием, чтобы считать, что F доказуемо. Это, так сказать, правило вывода.

Даже в этой очень скромной системе уже кое-что можно сказать, она кое-что замечает. Например, вот элементарное наблюдение, что конъюнкция и доказуемость коммутируют. $(\Box A) \wedge (\Box B)$ — это то же самое,

что $\Box(A \wedge B)$.

$A \to (B \to A \wedge B)$	$A \wedge B \to A$
$\Box(A \to (B \to A \wedge B))$	$\Box(A \wedge B \to A)$
$\Box A \to \Box(B \to A \wedge B)$	$\Box(A \wedge B) \to \Box A$
$\Box A \to (\Box B \to \Box(A \wedge B))$	$\Box(A \wedge B) \to \Box B$
$(\Box A \wedge \Box B) \to \Box(A \wedge B)$	$\Box(A \wedge B) \to (\Box A \wedge \Box B)$

Давайте теперь попробуем поиграть с дизъюнкцией. Таким же образом мы можем показать, что если $\Box A \vee \Box B$, то $\Box(A \vee B)$. Действительно, как мы это обосновали: если $\Box A$, то $\Box(A \vee B)$, если $\Box B$, то тоже $\Box(A \vee B)$. Таким образом из дизъюнкции доказуемости следует доказуемость дизъюнкции. Вот небольшой формальный вывод, который эту интуицию поддерживает.

$$
\begin{array}{l}
A \to A \vee B \\
\Box(A \to A \vee B) \\
\Box A \to \Box(A \vee B) \\
\Box B \to \Box(A \vee B) \\
(\Box A \vee \Box B) \to \Box(A \vee B)
\end{array}
$$

Давайте посмотрим, можно ли вынести модальность доказуемости за скобки в дизъюнкции? Если $\Box(A \vee B)$, следует ли отсюда, что $\Box A \vee \Box B$? В этом есть большие сомнения. Дизъюнкция бывает доказуема по тривиальным соображениям. Например, в классической логике $A \vee \neg A$ всегда верно, но из этого никак не следует, что у нас есть доказательство A или у нас есть доказательство $\neg A$.

Гёдель еще тогда, в 1933 г., заметил, что если мы верим в $S4$, т. е. набор элементарных принципов доказуемости, то мы уже в состоянии понять интуиционистскую логику, на том самом уровне, на котором она должна быть: на пропозициональном или на первопорядковом. И никакого дополнительного кодирования при этом не нужно. Мы берем формулу F, которую мы хотим проверить, является ли она интуиционистской тавтологией. Во-первых на каждую подформулу F навешиваем модальность; тем самым мы заставляем то устройство, которое разбирает эту формулу, каждый раз вместо истинности на конкретной подформуле говорить: «доказуемо».

Например, рассмотрим формулу $A \to A$. Формальной трансляцией этого является следующее: $\Box(\Box A \to \Box A)$. Одна формула, $A \to A$, интуиционистская, которую мы, так сказать, не понимаем, а другая формула классическая, она уже в $S4$, которую мы понимаем.

Теперь мы в состоянии, взяв интуиционистскую формулу перевести ее на классический язык, который $S4$ в состоянии понять, и просто проверить

ее в $S4$; $S4$, кстати, — разрешимая логика, там есть алгоритмы проверки истинности и выводимости в $S4$. Если $S4$ ответит: «да», тогда принимаем формулу, если он ответит «нет», мы ее отвергаем.

Гёдель показал, что этот тест является правильным для общепринятых аксиом интуиционистской логики. Точнее, он доказал это в одну сторону и сделал предположение, что в обратную сторону это тоже верно. Примерно 15 лет ушло на то, чтобы проверить, что действительно это так и есть. Между прочим, проверка вложимости интуиционистской логики в $S4$ была завершена в 1948 г. Тарским именно с помощью топологической интерпретации. Эта интерпретация носила, таким образом, искусственный, но чрезвычайно полезный характер в этой области.

Итак, что мы имеем? Если у нас есть $S4$, мы в состоянии понять интуиционистскую логику; задача свелась к тому, чтобы понять $S4$. Казалось бы, дело за малым: надо взять какую-нибудь модель доказательств, например, гёделевскую формулу доказуемости, которая фигурирует в его много раз упоминавшейся теореме о неполноте; ну и, так сказать, связать их — и всё получится. Не тут-то было.

Естественная попытка интерпретировать модальность в $S4$ как обычную формальную доказуемость, которая у Гёделя фигурирует в его теореме о неполноте, проваливается. И провал ее состоит в следующем. Формула $\Box F \to F$ — это условие рефлексивности, она, конечно, выводима в $S4$. По правилу усиления в $S4$ оказывается выводим принцип: $\Box(\Box F \to F)$. Теперь давайте вместо F вставим что-нибудь простое, например, логическую константу «ложь», $\bot$: $\Box(\Box\bot \to \bot)$. Давайте поймем эту формулу. Импликация к лжи — это отрицание. Значит, здесь стоит: «доказуемо, что не доказуема ложь». По Гёделю доказуемость понимается в какой-то фиксированной, конкретной системе, например, в арифметике, или в какой-то более обширной теории, содержащей арифметику, достаточно богатой, чтобы все кодирования сделать и чтобы записать формулу доказуемости. Итак, давайте прочтем: «не доказуема ложь». Это является формальным аналогом знаменитой формулы непротиворечивости. Таким образом, естественное прочтение этой формулы таково: внутри теории T доказуема ее собственная непротиворечивость! Это противоречит второй теореме Гёделя о неполноте. Гёдель был, разумеется, первым, кто это заметил, и сказал: проблема в том, что мы не знаем точной интерпретации $S4$.

Логика Доказательств

Пятью годами позже, в 1938 г. в Вене состоялась публичная лекции Гёделя, которая, к сожалению, была опубликована только в 1995 г., когда

уже вся история была практически закончена. Гёдель сказал, что надо перейти к логике доказательств. Надо пытаться построить модель на основе $S4$, но, тем не менее, используя формат, где объектов уже два: t есть доказательство F. То есть вместо логики доказуемости рассматривать логику доказательств.

Колмогоров и ученик Брауэра Гейтинг на самом деле с самого начала предлагали искать именно систему операций над доказательствами, а не над доказуемостью вообще. Оказывается, на этом пути всё удается сделать. На удивление простой системы вычислимых операций над доказательствами хватает для того, чтобы выстроить когерентную математическую теорию всей этой области, и тем самым ответить на вопрос Гёделя о доказуемостной интерпретации $S4$. Говоря условно, Логика Доказательств **LP** есть классическая логика + дополнительные атомы $p : F$, (p — полином доказательств, а F — формула).

$A0$. *классические аксиомы и правила*

$A1$. $t : (F \to G) \to (s : F \to (t \cdot s) : G)$

$A2$. $t : F \to F$

$A3$. $t : F \to !t : (t : F)$

$A4$. $s : F \to (s + t) : F, \quad t : F \to (s + t) : F$

$R1$. $\vdash c : A$, *где* $A \in A0$—$A4$, c — *константа доказательства*.

Мы рассматриваем термы, которые условно называются полиномами доказательств, хотя здесь кроме умножения и сложения есть еще одноместная операция проверки доказательств «!», которую условно можно называть экспонентой. Итак, у нас есть переменные, есть константы, есть две двуместные операции — умножение и сложение — и одноместная операция «!» проверки доказательств. Из них строятся полиномы доказательств.

Кроме обычных аксиом классической логики и правил классической логики у нас еще есть специальные принципы. Например, операция умножения: если t есть доказательство $F \to G$, и s — доказательство F, то t примененное к s есть доказательство G. Это как раз то, что завещал Колмогоров. Если t доказывает F, то F — истина. Заметим, что у нас есть приятное свойство перемешивания: в одном и том же языке мы можем говорить о доказательствах и об условиях. Вновь вспоминается Колмогоров, который, собственно, этого и хотел.

Теперь операция проверки доказательств уже становится явной. Помните, мы говорили, что если $\Box F$, то есть доказательство, и т. д. Теперь мы это делаем в явном виде. Мы говорим, что если t есть доказательство F, то результат применения операции проверки доказательств к t является доказательством того, что t является доказательством F.

Оказывается, что этого мало. Для того чтобы смоделировать всю $S4$, а через нее и интуиционистскую логику по Гёделю— Тарскому, нам нужна еще операция «плюс», которая в обычном модальном языке не видна.

А именно, этот плюс показывает, что, говоря о моделировании доказуемости в модальной логике, мы должны принять недерминистическую точку зрения на доказательства. Если книга в 400 страниц доказывает одну теорему, то она доказывает не только её: все леммы, которые в промежутке появились, тоже считаются в этой книге доказанными. Таким образом, мы считаем, что доказательство доказывает не только последнюю формулу, но и все формулы, входящие в доказательство. И вот эта операция «плюс» — она как раз и говорит, что если у нас s является доказательством F, то если мы что-то туда добавим, результат останется доказательством того же самого F.

Важным моментом является дисциплина применения констант. Если A — это аксиома, то мы берем любую константу и назначаем ее доказательством A, т. е. берем обязательство эту константу интерпретировать в дальнейшем как доказательство данной аксиомы A.

Мы можем применять для каждого A свою константу, но, поскольку мы договорились, что доказательство у нас объемлющее, мы можем использовать одну и ту же константу несколько раз; т. е. у нас одна и та же константа может быть доказательством многих аксиом.

И в частности выяснилось, что в одной из известных систем формализации и проверки доказательств, которая в течение 20 лет разрабатывалась в Корнельском университете, на уровне 2 константа c использовалась всегда одна и та же. То есть у них был маркер того, что есть доказательство, но разработчики не брали на себя обязательство это доказательство строить в явном виде. На уровне вложения 1 это сделать относительно легко, а на уровне 2 это довольно серьезное вычислительное обязательство, которое они не хотели на себя брать.

Примеры вывода в Логике Доказательств

Теперь я дам несколько примеров, как работает Логика Доказательств. Возьмем вывод в $S4$, который доказывает, что конъюнкция доказуемости влечет доказуемость конъюнкции. Теперь мы с неявного языка переходим на явный. Чего мы ожидаем от этого? Если доказуемо A и доказуемо B, то доказуема конъюнкция. Теперь, с точки зрения обычной математической интуиции, я бы ожидал следующего: если x является доказательством A, и y является доказательством B, то некоторый доказуемостный полином, зависящий от x и y, является доказательством $A \wedge B$. Если система на

что-то претендует, она должна уметь это делать.

1. $A \to (B \to (A \wedge B))$, *по* $A0$;
2. $c : [A \to (B \to (A \wedge B))]$, *из* 1, *по* $R1$;
3. $x : A \to (c \cdot x) : (B \to (A \wedge B))$, *из* 2, *по* $A1$ *и Modus Ponens*;
4. $x : A \to (y : B \to (c \cdot x \cdot y) : (A \wedge B))$, *из* 3, *по* $A1$ *и булевой логике*;
5. $x : A \wedge y : B \to (c \cdot x \cdot y) : (A \wedge B)$, *из* 4, *по булевой логике*.

Выведенная формула 5 содержит константу c, которая была введена в строке 2 по правилу R1. Полное прочтение результата таково:

$$x : A \wedge y : B \to (c \cdot x \cdot y) : (A \wedge B),$$

где $c : [A \to (B \to (A \wedge B))]$.

Докажем еще, что дизъюнкция утверждений о доказуемости A и B влечет доказуемость дизъюнкции $A \vee B$.

1. $A \to (A \vee B)$, *по* $A0$;
2. $a : [A \to (A \vee B)]$, *из* 1, *по* $R1$;
3. $x : A \to (a \cdot x) : (A \vee B)$, *из* 2, *по* $A1$ *и Modus Ponens*;
4. $B \to (A \vee B)$, *по* $A0$;
5. $b : [B \to (A \vee B)]$, *из* 4, *по* $R1$;
6. $y : B \to (b \cdot y) : (A \vee B)$ *из* 5, *по* $A1$ *и Modus Ponens*;
7. $(a \cdot x) : (A \vee B) \to (a \cdot x + b \cdot y) : (A \vee B)$, *по* $A4$;
8. $(b \cdot y) : (A \vee B) \to (a \cdot x + b \cdot y) : (A \vee B)$, *по* $A4$;
9. $(x : A \vee y : B) \to (a \cdot x + b \cdot y) : (A \vee B)$ *из* 3, 6, 7, 8, *по булевой логике*.

Таким образом,

$$(x : A \vee y : B) \to (a \cdot x + b \cdot y) : (A \vee B),$$

где $a : [A \to (A \vee B)]$ и $b : B \to (A \vee B)]$.

Заключение

Операция аппликации, операция сложения доказательств и операция проверки доказательств — это вычислимые операции над доказательствами. И оказывается, что больше нам ничего не надо. С точки зрения реализуемости **LP** как раз соответствует $S4$, т. е. формула выводима в $S4$ тогда и только тогда, когда она реализуема полиномами доказательств в логике доказательств. Так что **LP** замечательно легла как раз в оставшийся открытым промежуток между $S4$ и реальными доказательствами, и тем самым эту старую историю удалось завершить удовлетворительным образом.

Работа Гёделя 1933 г. оставила открытыми две проблемы. Помните, он сформулировал $S4$ с одной стороны, а с другой стороны показал, что

“all4” — 2009/5/27 — 11:23 — page 141 — #141

если мы модальность интерпретируем как доказуемость, то интерпретации не получается. Таким образом, $S4$ была система без семантики, а формальная доказуемость была семантикой без системы аксиом. Оказалось, речь идет о разных моделях доказуемости, каждая имеющая свою область применения.

Складывается впечатление, что некоторые области в компьютерной науке ждали этого математического аппарата, чтобы двинуться дальше, и сейчас мы находимся в такой ситуации, когда продвижение в области, которая начиналась неформальными рассуждениями об основаниях, привело к созданию конкретного механизма, который оказался полезным для приложений.

6 марта 2003 г.

В. И. Данилов

ЗАДАЧА ХОРНА И ДИСКРЕТНАЯ ВЫПУКЛОСТЬ

Я хочу рассказать про следующую задачу. Что можно сказать про собственные значения суммы двух вещественно-симметрических или эрмитовых матриц, если известен спектр слагаемых? Пусть A и B — симметрические матрицы, спектры которых известны; что можно сказать про спектр матрицы $A + B = C$?

Эта задача имеет довольно длинную историю (немножко я про это расскажу), но принципиальное решение этой задачи получил А. Клячко *). Конечно, с тех пор народ на нее накинулся, стал смотреть в разных других направлениях и получил много обобщений и вариаций, но я этого касаться уже не буду.

Надо сказать, что задача об описании спектра суммы двух матриц связана странным образом еще с некоторыми задачами, которые, казалось бы, довольно далеки от нее. Во-первых, это представления группы GL или симметрической группы Sym. Вторая задача, которая к этому имеет отношение — исчисление Шуберта, т. е. вычисление пересечений в грассмановом многообразии. Надо сказать, что буквально недавно в Архиве появились работы Вакила, в которых доказывается какое-то существенное продвижение в этом тоже несомненно классическом и насчитывающем более ста лет направлении: известно, что Гильберт ставил это в качестве одной из проблем. И третье: инвариантные факторы для модулей над кольцами главных идеалов. Но я тоже этой всей тематики касаться не буду. Хочу только сказать, что про это все более или менее подробно написал Фултон **). В его обзоре рассказывается и про задачу Хорна (я чуть позже подробнее скажу, что это за задача).

Здесь я буду делать упор на дискретную выпуклость, что, может быть, еще менее известно, потому что в основном это мы с Глебом Кошевым придумали. Глеб Кошевой является соавтором большей части того, что

*) *Klyachko A. A.* Stable bundles, representation theory and Hermitian operators // Selecta Math. 1998. V. 4. P. 419—445.

**) *Fulton W.* Eigenvalues, invariant factors, heighest weights, and Schubert calculus // Bull. Amer. Math. Soc. 2000. V. 37. P. 209—249.

я буду рассказывать здесь. И я постараюсь как можно более элементарными средствами рассказать именно про задачу Хорна.

Давайте сначала рассмотрим одну симметрическую матрицу. Самое главное, что про нее нужно знать, — это то, что она диагонализируется, т. е. у нее существует ортогональный базис из собственных векторов, и собственные значения вещественны. И по традиции эти собственные значения упорядочивают в порядке убывания: $\alpha(1) \geqslant \alpha(2) \geqslant \ldots \geqslant \alpha(n)$. Я всегда буду предполагать, что n — это размер матрицы. Такой набор α естественно назвать спектром этой матрицы.

Спектром обычно особенно интересуются в приложениях. Например, в квантовой механике собственные значения — это значения наблюдаемых величин (правда, там обычно занимаются эрмитовыми матрицами, а я из соображений простоты не хочу рассматривать эрмитовы матрицы; но пока то, что я буду говорить, верно и для эрмитовых матриц). А в механических приложениях это связано с частотой колебаний систем типа маятника или чего-то подобного. Поэтому задачи о том, что произойдет, когда мы две системы как бы накладываем одну на другую, и как при этом изменятся собственные частоты, давно интересовали людей.

Теперь я хочу от одной матрицы перейти к двум. Допустим, что у нас есть две матрицы A и B, и мы знаем спектр одной (α) и спектр другой — (β). Что можно сказать про спектр их суммы, который я буду обозначать через γ? Есть две ситуации, когда довольно легко ответить на этот вопрос.

Первая ситуация — это когда A и B коммутируют. В этом случае у них есть общий собственный базис. Поэтому собственные значения для одинаковых собственных векторов складываются. Ниоткуда не следует, что $\alpha(1)$ и $\beta(1)$ относятся к одному и тому же собственному вектору. Мы должны взять точку α, а что касается β, то мы должны как-то переставить эти значения:

$$\gamma(i) = \alpha(\sigma(i)) + \beta(\tau(i)), \qquad \sigma, \tau \in \mathrm{Sym}(n).$$

Второй случай, который входит почти во все учебники — это то, что называется теория возмущений. Это тот случай, когда предполагается, что, скажем, матрица B мала по сравнению с A. Надо более точно говорить, что значит мала, потому что мала она не по сравнению с A, конечно, а по сравнению с различием между спектром: можно сказать, что мала, если $\beta(1) - \beta(n) \ll \alpha(i) - \alpha(i+1)$ для любого $i = 1, \ldots, n-1$; предполагается, что спектр матрицы A простой. В этом случае есть формула, как будет выглядеть спектр суммы. Собственные значение в этом случае не складываются. Там возникают довольно сложные формулы. И обычно считается, что есть малый параметр ε. Пусть B пока не очень малая, но когда мы

умножаем на ε, делаем матрицу малой. И тогда раскладывают в ряд по ε, как будут меняться собственные векторы. Потому что, когда мы прибавим к матрице A какое-то небольшое возмущение, собственные векторы тоже немножко изменятся, чуть-чуть повернутся, и теория возмущения говорит, на сколько или в каком направлении повернутся эти собственные векторы и на сколько изменятся собственные значения.

Это я просто упомянул. В дальнейшем я буду рассматривать общий случай.

Вся трудность в этой задаче состоит в том, что собственный базис для матрицы A и для матрицы B торчат в разных направлениях, и поэтому довольно сложно представить, что произойдет с собственными значениями для суммы. Можно поставить и более общий вопрос: как изменятся собственные векторы. Но это уже очень сложный вопрос: непонятно даже, как формулировать ответ.

Несмотря на то, что ответ о том, как будет выглядеть спектр суммы, довольно сложный, одно соотношение почти очевидно. А именно, если мы возьмем и сложим все числа $\gamma(i)$ — это будет сумма всех чисел $\alpha(i)$ и всех чисел $\beta(i)$: $\gamma(1)+\ldots+\gamma(n)=\alpha(1)+\ldots+\alpha(n)+\beta(1)+\ldots+\beta(n)$. Объяснение этому, конечно, очень простое: сумма всех собственных значений — это то же самое, что след матрицы. А след суммы равняется сумме следов, потому что обычно след определяют как сумму диагональных элементов, и конечно, диагонали складываются. Поэтому одно такое соотношение есть, и про него нужно помнить.

Тут сразу же стоит сказать, что если мы будем прибавлять к матрице A или B единичную матрицу (или даже скалярную матрицу), и соответственно будет меняться и C, то задача меняется довольно просто. Если мы прибавляем единичную матрицу, все $\alpha(i)$ увеличиваются на единицу. Это частный случай того, что я говорил про коммутирующие матрицы — единичная матрица (или кратная единичной) коммутирует с любой матрицей. Поэтому ситуация с прибавлением единичной матрицы легко контролируется. Я не буду особенно на это напирать, но можно, пользуясь этим, считать, что следы всех матриц равны 0. Это ничего в задаче не меняет. Если мы так сделаем, мы можем ограничиться матрицами с нулевым следом.

Помимо такого линейного соотношения есть еще одно простое соотношение типа неравенства. А именно: $\gamma(1)\leqslant\alpha(1)+\beta(1)$. Давайте будем считать, что A, B и C, скажем, положительно определенные (временно прибавив матрицу, кратную единичной матрице). Тогда $\gamma(1)$ (максимальный коэффициент) показывает, насколько у нас вектор увеличивает длину. Мы берем какой-нибудь вектор, действуем на него матрицей A — он как-то изменяется. Мы даже примерно знаем, что происходит, потому что у нас

есть собственный ортонормированный базис. По определению $\alpha(1)$ — это во сколько раз этот вектор увеличивается в этом направлении. Он увеличивается еще во втором направлении, но уже меньше. Таким образом первое, старшее, собственное значение говорит, насколько увеличивается вектор. Пока что в одном направлении, но довольно ясно, что, так как в остальных направлениях он увеличивается меньше, то и любой вектор будет увеличиваться не больше, чем в $\alpha(1)$ раз. Его длина после применения оператора увеличивается не более, чем в $\alpha(1)$ раз. Если мы теперь хотим посмотреть, что произойдет, когда мы подействуем матрицей C, то у нас этот вектор (его длина) под действием оператора A увеличится самое большее в $\alpha(1)$ раз, под действием оператора B она увеличится в $\beta(1)$ раз. Значит, суммарно длина по неравенству треугольника не может увеличиться больше, чем в $\alpha(1)+\beta(1)$ раз. Так получаем это тривиальное соотношение.

Видно, что соображение с увеличением длин хорошее. Развивая это соображение, Г. Вейль в 1912 г. доказал обобщение этого неравенства, которое выглядит таким образом: $\gamma(k)\leqslant\alpha(i)+\beta(j)$, если $k+1=i+j$. И не только такое соотношение, но еще целый ряд соотношений на собственные значения. Такие неравенства я буду называть неравенствами Вейля.

Надо сказать, что если $n=2$, то соотношения Вейля фактически дают ответ на эту задачу. То есть γ удовлетворяет таким соотношениям и, наоборот, если у нас задана какая-то тройка чисел, которые связаны такими соотношениями, то существуют и матрицы, которые имеют такой спектр. Может быть, стоит на случае $n=2$ остановиться чуть подробнее, потому что он в миниатюре показывает, как устроено дело и в общем случае.

Пусть A — матрица 2×2. Как я уже говорил, можно считать без особого ущерба, что след этой матрицы равен 0. Тогда эта матрица будет иметь такой простой вид: $A=\begin{pmatrix} a & a' \\ a' & -a\end{pmatrix}$. Она, как видно, полностью определяется своей первой строчкой, поэтому в каком-то смысле можно говорить не про матрицы, а просто про векторы. Пусть $(a,\,a')\in\mathbb{R}^2$ — какой-то вектор на плоскости, и мы уже можем видеть наглядно, что происходит. Матрица B — тоже какой-то вектор; а C, соответственно, — сумма этих векторов, потому что там все складывается.

Теперь давайте посмотрим, как устроен спектр такой матрицы. Для этого надо написать характеристическое уравнение. Так как след равен 0, то члена при λ не будет, а будет стоять только определитель матрицы A. И характеристическое уравнение имеет такой вид: $\lambda^2+\det A=0$. А определитель матрицы 2 на 2 посчитать несложно. Он равен $\det A=-a^2-a'^2$. Таким образом $\lambda_{1,2}=\pm\sqrt{a^2+a'^2}$. А что такое корень квадратный из

суммы квадратов? Матрицу A я представляю в виде вектора (a, a'); собственные значения этой матрицы — это длина этого вектора и минус длина. Минусом, по понятным соображениям, мы не особенно интересуемся.

Теперь мы берем аналогичный вектор для матрицы B; его длина — это $\beta(1)$. И спрашивается: какова будет длина суммы векторов.

Чему равняется, сказать нельзя, но мы можем легко понять, какие значения может принимать длина этого вектора. Будем считать, что A длиннее, чем B. Проведем окружность радиуса $\beta(1)$ с центром на расстоянии $\alpha(1)$ от начала координат. Всевозможные расстояния от начала координат до точек этой окружности годятся в качестве ответа. Ясно, что длина C, если у нас известны длины A и B, задается неравенством треугольника: она не может быть больше, чем сумма длин A и B, не может быть меньше, чем их разность. Фактически это в точности тот ответ, который дается неравенством Вейля.

Я хотел только обратить внимание на то, что всевозможные значения для $\gamma(1)$ заполняют отрезок от $\alpha(1)+\beta(1)$ до $\alpha(1)-\beta(1)$. И второе обстоятельство, которое тоже видно на этом чрезвычайно простом примере, состоит в том, что крайние значения (максимум $\gamma(1)$ и минимум $\gamma(1)$) достигаются в тех случаях, когда A и B пропорциональны. Это обстоятельство имеет место не только в этом частном случае; при правильной формулировке оно верно в общем случае. Я про это еще чуть подробнее скажу.

Этот пример наводит на мысль, что множество возможных значений γ, во-первых, является выпуклым многогранником (в данном случае — отрезком); и во-вторых, граничные значения этого многогранника соответствуют какому-то специальному соотношению между матрицами A и B. Потом я про это скажу. Можно и сейчас сказать, что A и B находятся в выделенном положении, если у них есть нетривиальное общее собственное подпространство. В частном случае, где все делается вручную, я обрисовал ситуацию и думаю, что в случае $n=2$ все предельно ясно.

После того как Вейль установил свое неравенство, было некоторое затишье, а следующая вспышка интереса была где-то между 1949 и 1962 г., и она затронула Москву. А именно, И. М. Гельфанд заинтересовался этой задачей, как можно прочитать у Зелевинского. Зелевинский говорит, что Гельфанд первым поставил задачу об описании множества всех значений γ, которые возможны в этой задаче. Раньше народ больше интересовался задачами такого типа: каким может быть, скажем, второе собственное значение $\gamma(2)$, или что можно сказать про третье — какие есть ограничения на конкретные собственные числа. Гельфанд первым поставил задачу более правильно: как, вообще, множество всех этих γ устроено?

Может быть, здесь сразу стоит ввести обозначения. Множество таких γ, которые могут встретиться в качестве ответа в этой задаче, я буду обозначать символом $H(\alpha, \beta)$, где α и β — спектры матриц, которые мы рассматриваем. В зависимости от того, как их собственные реперы расположены, может получаться тот или иной ответ для γ. Это множество всевозможных ответов я буду обозначать H.

Я уже упоминал Гельфанда. Тут надо упомянуть и Лидского, который, видимо, по поручению Гельфанда, пытался более элементарно взглянуть на эту задачу, и у него была сравнительно короткая заметка в «Докладах», где почти ничего не доказывалось, а формулировались эти утверждения. Еще нужно упомянуть Хорна. Я отметил 1962 г. потому, что в этом году вышла статья Хорна, в которой он восстановил эти не приведенные доказательства Лидского. А главное, что он в этой же работе выписал гипотетически все условия.

В нашем простом примере мы видели, что H — это многогранник. Хорн, видимо, чувствовал, что и в общем случае это будет многогранник. А раз это многогранник, он задается какими-то линейными неравенствами. И задача состояла в том, чтобы эти неравенства явно выписать.

Если мы рассматриваем двумерный случай, то это полуплоскость; в общем случае у нас есть несколько неравенств. Хотелось бы сказать, что это конус, камера Вейля. Все это действительно становится камерой Вейля, если мы наложим еще условие нулевого следа; тогда мы получим действительно настоящую камеру. Тут есть лишняя размерность, которая не нужна, так что правильно рассматривать матрицы с нулевым следом. Пока для меня это не очень важно.

Если фиксированы α и β, то у нас появляется множество $H(\alpha, \beta)$ возможных значений для γ. Я рассматриваю задачу о том, как устроено это множество, что про него можно сказать. Хорн догадывался, что это будет многогранник, и пытался написать, или угадать те линейные неравенства, которые его задают. Он дал список неравенств, которые должны давать этот многогранник. Я их выписывать не буду, а только хочу подчеркнуть, что они имеют вид $\gamma(K) \leqslant \alpha(I) + \beta(J)$, где I, J и K — это какие-то подмножества $\{1, \ldots, n\}$, причем $|I| = |J| = |K|$; а $\alpha(I)$ — это просто сумма $\sum\limits_{i \in I} \alpha(i)$. Случай Вейля — это тот случай, где эти множества I, J, K состоят из одного элемента; но, вообще говоря, они состоят из нескольких элементов. В гипотезе Хорна содержится, что мощности этих трех множеств одни и те же. Точно так же, как и раньше, где мы брали по единичке. Конечно, I, J и K не произвольны. И Хорн предложил некоторую рекурсивную процедуру, как эти равномощные тройки строить.

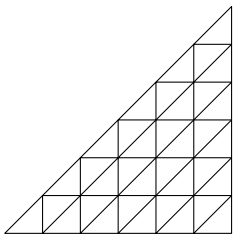
Я не хочу объяснять, как устроена эта рекурсивная процедура, потому что у меня такое впечатление, что именно она была каким-то препятствием для доказательства — она немножко запутывала дело. И может быть, именно поэтому ни Хорну, ни Лидскому не удалось доказать эту гипотезу. Я в своем докладе хочу сказать, что, в принципе, тех средств, той техники, которой обладали эти люди (надо сказать, что, конечно, они значительно превосходили те возможности, которыми я, скажем, обладаю; т. е. они довольно виртуозно владели всей этой техникой), было достаточно для доказательства. Мне кажется, что они смогли бы доказать эту гипотезу, если бы у них была более правильная формулировка. Очень сложно работать с такими рекурсивно определяемыми множествами. К более правильной формулировке я постепенно и перейду.

Нужно теперь сказать по поводу этой более правильной формулировки. Народ, который занимался этой задачей, постепенно понял, что условия Хорна интерпретируются в терминах пересечения циклов Шуберта, и смог быстро доказать, что, действительно, все такие неравенства выполняются, т. е. множество H лежит в многограннике, который задается этими неравенствами. Более трудная задача состояла в том, чтобы построить эту матрицу C. Точнее, не явно ее построить, но доказать, что если мы возьмем любую точку γ, которая удовлетворяет этим неравенствам (т. е. лежит в многограннике, который задан этими неравенствами), то можно так подобрать A и B, что их сумма будет иметь спектр γ. Как я уже говорил, задачу в обратную, более трудную на мой взгляд сторону решил Клячко. Потом народ немножко занимался осмыслением его результата. И мне кажется, важную роль тут сыграл Зелевинский, который на какой-то конференции рассказывал про эту задачу и, в частности, привлек внимание к так называемым паттернам Беренштейна—Зелевинского, про которые я тоже говорить не буду. И уже эти паттерны как-то подвигли Кнутсона и Тао — главных упрощателей этой ситуации к хоникомбной (honeycombe) модели. Что такое хоникомб, я тоже рассказывать не хочу, но это что-то вроде того, что идут 3 пучка прямых по трем направлениям, и потом эти прямые сталкиваются. Я про это рассказывать не буду, потому что буду говорить как бы на двойственном языке.

Когда мы с Глебом увидели такую картинку, мы с ним занимались дискретной выпуклостью. Это задача, относящаяся к дискретной математике, к целочисленной математике. В простейшем случае, а именно, в двумерном, те картинки или те образы, которые у нас возникали, имели как раз примерно такой вид. Когда мы увидели эту картинку, мы сразу поняли, что здесь пахнет как-то дискретной выпуклостью, и предприняли усилия, чтобы понять, как одно с другим связано.

Теперь я скажу немножко про дискретную выпуклость и формулировку в этих терминах ответа на задачу Хорна. Про хоникомбы не хотелось бы говорить, потому что хоникомбы не мы придумали, а к дискретной выпуклости все-таки мы руки приложили.

Некоторое время я буду вводить понятия, связанные с дискретной выпуклостью, но в минимуме, который нужен именно для формулировки нового ответа на эту задачу. Для этого мы снова зададимся некоторым числом n (мы знаем, что это — ранг матрицы).



Р и с. 1. Грид $\Delta(6)$

Рассмотрим то, что я называю «треугольный грид» (grid или, просто, решетка; но так как слово «решетка» довольно сильно избито, то лучше называть гридом). Я рассмотрю множество $\Delta(n) = \{(i,\ j) \in \mathbb{Z}^2 \colon 0 \leqslant j \leqslant i \leqslant n\}$ целых точек, т. е. пар целых чисел $(i,\ j)$. Итак, рассмотрим целые точки, которые удовлетворяют таким ограничениям. Это будут целые точки в треугольнике. На рис. 1 я изобразил грид $\Delta(6)$. Рисунок нужно еще дополнить прямыми $\{i - j = 1,\ i - j = 2,\ \ldots\}$. Таким образом, мы имеем треугольник, который разбит на малые треугольники. (Правильнее, конечно, этот грид рисовать в гексагональном виде, но давайте нарисуем так; это совершенно не важно, как он там расположен, можно его аффинно преобразовывать, от этого ничего не меняется.) Через $\Delta(n)$, я уже сказал, я обозначаю множество целых точек треугольника, но иногда я буду говорить про это как про треугольник, тогда будет иметься в виду, что это — настоящий треугольник, а не только конечное множество.

Меня будут интересовать функции, которые заданы на этом гриде, на этом конечном наборе точек: $f \colon \Delta(n) \to \mathbb{R}$. Но не все функции, а функции, которые я буду называть дискретно вогнутыми. Что это означает? Имеется в виду вот что. Назовем ромбом в этом гриде два малых треугольника $\{a,\ b,\ c\}$ и $\{b,\ c,\ d\}$, которые соприкасаются друг с другом по ребру $\{b,\ c\}$. Условие дискретной вогнутости записывается так: $f(b) + f(c) \geqslant f(a) + f(d)$. Оно означает, что если мы возьмем сумму чисел, стоящих на одной диагонали, то она должна быть больше или равна, чем сумма чисел, разделенных другой диагональю. И вот если это неравенство выполнено для всех ромбиков, которые здесь можно себе представить, то такую функцию я называю дискретно вогнутой. Дискретно — понятно, почему: потому что она задана на дискретном множестве, в целых точках. А почему вогнутая? Вот по какой причине. Возьмем маленький треугольник, в трех его вершинах заданы значения этой функции; естественно ее на этот треугольник проинтерполировать линейным образом (каким еще другим

способом можно функцию на треугольнике нарисовать, если известны ее значения в 3 вершинах?); и тогда это условие означает, что у этой функции (уже проинтерполированной, т. е. мы вместо f рассматриваем $\widetilde{f}$) над этим ромбиком есть излом вниз. Так как она вдоль каждого ребра (внутреннего, конечно, для граничных у нее никаких условий нет) ломается вниз, то она и выглядит таким вот вогнутым образом; т. е. она напоминает функцию $y = -x^2$ (я, правда, нарисовал 1-мерную). Если мы нашу функцию f как бы естественно продолжим по линейности, то это получится вогнутая функция. Этим объясняется термин «вогнутая». Что можно сказать про такие функции?

Перво-наперво можно посмотреть, как устроена эта функция на нижнем основании треугольника. Рассмотрим приращение, скажем, на отрезке $((i, 0), (i+1, 0))$, и сравним его с приращением на отрезке $((i+1, 1), (i+2, 1))$. В силу неравенства, которое я написал, приращение на $((i+1, 1), (i+2, 1))$ будет меньше, чем приращение на $((i, 0), (i+1, 0))$; а если мы теперь рассмотрим приращение здесь на $((i+1, 0), (i+2, 0))$, то оно будет меньше, чем на $((i+1, 1), (i+2, 1))$; таким образом приращение на $((i+1, 0), (i+2, 0))$ будет меньше, чем приращение на $((i, 0), (i+1, 0))$. Чтобы это все культурно оформить, давайте для такой функции f через $\alpha(f)$ будем обозначать приращения на нижнем основании:

$$\alpha(f) = (f(1, 0) - f(0, 0),\ f(2, 0) - f(1, 0),\ \ldots,\ f(n, 0) - f(n-1, 0)) = \\ = (\alpha(1),\ \alpha(2),\ \ldots,\ \alpha(n)).$$

Из того, что я сейчас объяснил с помощью нехитрого сравнения приращений на этих трех отрезках, видно, что $\alpha(1) \geqslant \alpha(2) \geqslant \ldots \geqslant \alpha(n)$, т. е. это будет убывающая последовательность чисел. Приращение на нижнем основании я буду обозначать через α. Аналогично можно посмотреть приращения, когда мы будем двигаться такими единичными отрезочками или стрелочками по правой стороне. Это я назову β. Наконец, я могу рассмотреть приращения на гипотенузе, и то, что будет происходить на гипотенузе, обозначу через γ. Таким образом, когда имеется дискретно вогнутая функция, у нас возникают три набора чисел α, β и γ, которые упорядочены по убыванию. А это уже наводит на мысль про спектры.

Далее, давайте посмотрим, чему будет равна сумма всех α — сложим все α, потом сложим все β. Это будет приращение нашей функции вот на этом пути; ну и, как для любой функции, это будет равно приращению на любом другом пути, в том числе на гипотенузе, т. е. это будет сумма всех γ. И это соотношение — это в точности то соотношение, которое я выписывал для матриц. Если для матриц оно было, может быть, чуточку

нетривиально, то для функций это абсолютно банальное утверждение. Тем не менее, такое совпадение или близость свойств, которыми обладают эти α, β и γ, наводит на мысль, что все это какое-то имеет отношение к той задаче, с которой я начинал.

Чтобы усилить это впечатление, рассмотрим простейший случай, когда $n=2$. Можно считать, что в вершинах функция равна 0, это соответствует тому, что мы считаем след равным нулю; тогда $f(1,\,0)=\alpha(1)$, $f(2,\,1)=\beta(1)$, $f(1,\,1)=\gamma(1)$; и условие дискретной вогнутости, если его применить, к каждому из трех имеющихся при $n=2$ ромбиков, даст, что $\alpha(1)+\beta(1)\geqslant\gamma(1)$, $\alpha(1)+\gamma(1)\geqslant\beta(1)$, $\beta(1)+\gamma(1)\geqslant\alpha(1)$. Это — в точности условия Хорна, даже условия Вейля. Таким образом, случай $n=2$ показывает, что дискретно вогнутая функция дает в точности тот ответ, который мы знаем и для матриц.

Это все подводит к тому, что эти две задачи должны быть тесно связанными, и сейчас я сформулирую ответ. Ответ выглядит так: в задаче Хорна могут встретиться те и только те γ (спектры сумм при заданных α, β), которые бывают при ограничении на гипотенузу некоторой дискретно вогнутой функции, у которой значения приращений на катетах равны α и β.

Более аккуратно: введем множество $\Gamma(\alpha,\,\beta)$ — множество всех $\gamma(f)$, где f дискретно вогнутые и $\alpha(f)=\alpha$, $\beta(f)=\beta$:

$$\Gamma(\alpha,\,\beta)=\{\gamma(f)\colon\ f\text{ — дискретно вогнутая, }\alpha(f)=\alpha,\ \beta(f)=\beta\}.$$

То, что я ранее сказал словами, я теперь написал в виде значка Γ, и основной результат состоит в том, что $H(\alpha,\,\beta)=\Gamma(\alpha,\,\beta)$.

Т е о р е м а. $H(\alpha,\,\beta)=\Gamma(\alpha,\,\beta)$ — *многогранник*.

Это наша формулировка решения задачи Хорна. Но я бы не сказал, что это слишком уж большое изобретение, потому что это — немножко в переписанном виде ответ, который у Кнутсона и Тао дается в двойственных терминах через хоникомбы. А может быть, в еще более завуалированном виде — уже у Клячко и других людей.

Выписать Γ — тоже, конечно, некоторая задача. Я сейчас скажу про это. Но принципиальная вещь состоит в том, что у нас здесь имеется равенство. Утверждается, что множество H, в формулировке которого участвуют какие-то симметрические матрицы, какие-то нелинейные объекты типа спектров, выражается в терминах каких-то детсадовских вещей: нарисовали какую-то решетку, расставили какие-то числа, проверили какие-то неравенства, посмотрели, что может получиться на гипотенузе.

Я сформулировал теорему. Теперь я хочу перейти к доказательной части. Надо сказать, что имеется два подхода к доказательству гипотезы Хорна; и оба из них в высшей степени нетривиальны. Один подход,

который нашел Клячко, связан с тем, что он работал с флагами, действовал на них группой GL. Я уже говорил, что у каждой симметрической матрицы главная ее часть, в которой сидит спектр — это ее собственный базис. А собственный базис (особенно, если α упорядочено) — это, фактически, флаг. Клячко рассматривал действие на них полной линейной группы и неравенства Хорна он интерпретировал как условия стабильности — стабильности в смысле теории инвариантов действия этой группы на этом многообразии. И есть другой подход, который опирается на теорию отображения моментов и симплектическую редукцию. И тот, и другой подход, конечно, как я уже сказал, в высшей степени нетривиальны, и, чтобы это рассказывать, надо или знать эту теорию, или довольно долго излагать основы этой теории. Мы предлагаем некий более элементарный подход.

Если посмотреть на то, что нужно доказывать, то ясно, что множество Γ гораздо более понятно и просто устроено. Прежде всего ясно, что это — многогранник, потому что дискретно вогнутые функции можно складывать: неравенства вогнутости линейные, и если мы возьмем две функции, удовлетворяющие этим неравенствам, мы их можем сложить. Поэтому ясно, что множество Γ выпуклое. Так как здесь участвует конечное число неравенств, ясно, что Γ — многогранник. Если мы докажем равенство $H = \Gamma$, мы получим, что множество Хорна H — тоже многогранник. Я на это уже намекал, но в теореме это явно формулируется. И утверждение о том, что H — выпуклый многогранник, конечно, факт довольно нетривиальный. В случае $n = 2$ я показывал, почему это так.

К сожалению, здесь приходится апеллировать к теории отображения моментов. Ссылаясь на высокую науку, я сформулирую важную лемму.

Лемма. 1) $H(\alpha, \beta)$ — *выпуклый многогранник.*

2) *Если* $A + B$ *имеет простой спектр* γ *и* $\gamma \in \partial H(\alpha, \beta)$, *т. е. границе* $H(\alpha, \beta)$, *то* A *и* B *имеют общее собственное подпространство.*

Второе утверждение леммы говорит: пусть у нас γ — спектр суммы двух матриц A и B, все $\gamma(i)$ различны, и эта точка γ лежит на границе нашего многогранника (первая часть условия — это более-менее условие общего положения, а второе условие, что γ лежит на границе нашего многогранника — важно). Тогда (я уже про это говорил) A и B имеют общее собственное подпространство, т. е. расположены довольно специальным образом относительно друг друга.

То, что H — многогранник, не выводится из того, что Γ — многогранник; это нужно независимо. Это главное неэлементарное утверждение. Что H — многогранник, надо знать заранее. Это доказывается с помощью теории отображения моментов в работе Кнутсона. А вообще-то, это какие-то общие факты про отображение моментов; если орбита относительно

действия некоторой группы Ли отображается с помощью специального отображения моментов, то образ является выпуклым многогранником. Вы видели справедливость этого утверждения в случае $n=2$. Это утверждение доказывается в том предположении, что γ не лежит на границе камеры Вейля. Это важная оговорка, потому что в противном случае утверждение неверно, но это оговорка, на которую я обращать внимания особенно и не буду.

Я этим воспользуюсь для того, чтобы показать включение $H(\alpha, \beta) \subset \subset \Gamma(\alpha, \beta)$. Я сейчас хочу доказать сравнительно легкую часть гипотезы Хорна: что любая точка γ, которая появляется как спектр суммы, лежит в этом многограннике, т. е. удовлетворяет линейным неравенствам, которые, как мы потом увидим, имеют прямое отношение к неравенствам Хорна.

У нас есть выпуклое множество Γ (выпуклость Γ уже объяснялась). Теперь возьмем какую-нибудь точку на границе H; я покажу, что любая точка на границе H будет принадлежать многограннику Γ. Мне нужно доказать, что множество H лежит внутри Γ; а я докажу более слабую вещь, а именно, что любая граничная точка H лежит внутри; тогда отсюда будет следовать, что и весь он лежит внутри. Потому что H — это компактное множество, спектр не может расти до бесконечности, уже $\gamma(1)$ зажато. Поэтому, если все граничные точки лежат внутри Γ, то и все H будет лежать внутри Γ, и я получу это включение. Здесь я пользуюсь тем, что Γ — выпуклый многогранник.

Берем точку γ на границе $H(\alpha, \beta)$. Конечно, здесь надо как-то пошевелить задачу, чтобы у нас все было не на границе камеры Вейля, но это только отвлекает нас, поэтому я на это не буду обращать внимания. Если я беру такую точку γ, которая является спектром $A+B$ и лежит на границе, то тогда A и B имеют собственное подпространство. Это означает, что они приводятся к блочному виду, т. е. A имеет вид $A_1 \oplus A_2$, B имеет вид $B_1 \oplus B_2$, где, конечно, A_1 и B_1 действуют в одном и том же подпространстве. То есть можно выбрать такой базис, что они имеют блочный вид с блоками одинакового размера.

Тогда что я делаю? Я беру матрицу $A_1 + B_1$; у нее есть спектр γ_1 — если мы рассмотрим только часть задачи, которая относится к этому подпространству. Все уже имеет меньшую размерность, поэтому по индукции γ_1 лежит в своем многограннике. А более правильно сказать, что γ_1 происходит из какой-то функции f_1 — я должен взять какой-то не такой большой грид, а поменьше; если у меня это подпространство имеет размер k, то я должен взять какую-то функцию f_1, дискретно вогнутую на гриде $\Delta(k)$. Аналогично для второй пары: γ_2 будет происходить из какой-то

функции f_2 на гриде $\Delta(n-k)$. И теперь все завершается тем, что мне нужно построить какую-то дискретно вогнутую функцию f на большом гриде, и делается это с помощью еще одной операции, которую я не упомянул, но которая играет важную роль — это операция свертки (конволюции). Я сейчас скажу, что это такое. Это операция, которая из двух функций на маленьких треугольниках строит некоторую новую функцию на сумме этих двух треугольников — сумме по Минковскому, конечно. Когда мы возьмем f_1 на $\Delta(k)$ и f_2 на $\Delta(n-k)$, то получится некоторая функция $f = f_1 * f_2$ на сумме. Как она устроена? $f(x) = \max\limits_{x_1+x_2=x} (f_1(x_1) + f_2(x_2))$ — т. е. ее значение в точке x — это максимум $f_1(x_1) + f_2(x_2)$, где $x_1 + x_2 = x$. Ведь эта точка по определению есть сумма какой-то точки x_1 и какой-то точки x_2; мы их сложим, и получится точка x. В качестве значения функции в этой точке нужно взять, грубо говоря, сумму значений в этих точках, а более аккуратно надо взять всевозможные разложения x в такие суммы, и взять соответствующий максимум.

Легко проверяется, что свертка двух дискретно вогнутых функций тоже будет дискретно вогнутой. Собственно, этим свойством и объясняется интерес того, что мы рассматриваем именно гриды такого специального вида, функции именно такого специального вида и т. д. Можно сказать так: дискретно вогнутые функции чем-то похожи на настоящие вогнутые функции, которые встречаются в выпуклом анализе. С вогнутыми функциями можно делать довольно много разных операций, но главные из них две: вогнутые функции можно складывать, и сумма двух вогнутых снова вогнутая, и второе — с функциями можно делать свертку, и она тоже будет вогнутой функцией.

Если у нас есть две вогнутые функции, то что значит их свертка? Нужно взять их подграфики и сложить, получится снова некоторое вогнутое множество, надо взять его огибающую, это и будет свертка. То есть фактически свертка функций — это сумма по Минковскому некоторых множеств. Вообще, в том, что называется теория вогнутых или выпуклых функций, сами функции не очень существенны, потому что фактически работа идет просто с выпуклыми множествами специального вида, которые с каждой точкой содержат весь луч, который идет вниз.

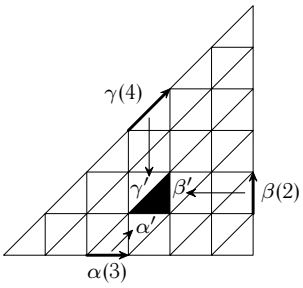
Нетрудно убедиться, что, когда мы делаем свертку, мы получим в точности нужный набор чисел γ. У каждой функции на гипотенузе было то γ, которое нужно было; а теперь, когда мы произведем свертку, эти наборы чисел γ так друг в друга вставятся, что получится как раз то, что нужно.

Итак, я доказал (или, по крайней мере, объяснил, как доказывается), что граница H лежит в Г.

"all4" — 2009/5/27 — 11:23 — page 155 — #155

Теперь надо провести доказательство в обратную сторону, и это более сложная задача. Рассуждение в каком-то смысле очень похожее. Я буду доказывать, что, наоборот, $\partial\Gamma \subset H$. Из этого следует, что $\Gamma \subset H$, поскольку H выпуклое по лемме. А вместе с предыдущим утверждением это будет означать равенство.

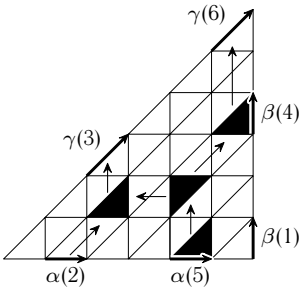
Здесь уже мы должны взять какую-то точку γ на границе Г. Тут мы приходим к необходимости разбираться в том, какими уравнениями задается наш многогранник Г. То, что это многогранник, мы и так понимали, а теперь приходится обращаться к описанию границы, или к описанию тех неравенств, которыми задается Г. Какие могут получиться γ, если известно, что это ограничение на гипотенузу дискретно вогнутой функции, что это граничные значения дискретно вогнутой функции. Тут тоже появляются некоторые интересные вещи.



Р и с. 2.

Пусть $n = 6$. Нарисуем грид и возьмем следующий малый треугольник в нем.

Пусть α', β' и γ' — приращения дискретно вогнутой функции на его сторонах. Теперь возьмем третий отрезок на основании, в данном случае $\alpha(3)$. Аналогично возьмем $\beta(2)$ и $\gamma(4)$, как на рис. 2. После этого я поставлю здесь стрелки. Пока не очень понятно, что они означают, но вот идут такие стрелки. Рассматриваются проекции черного треугольничка на стороны большого треугольника: на горизонтальный катет — вдоль гипотенузы, на вертикальный катет — горизонтальная, на гипотенузу — вертикальная. Теперь посмотрим приращение функции на нижней грани этого треугольника. Из соображений дискретной вогнутости, я уже говорил: $\alpha' \leqslant \alpha(3)$. Ну и, наоборот: $\gamma(4) \leqslant \gamma'$. Так как



Р и с. 3.

$$\gamma' = \alpha' + \beta',$$

мы получаем, что

$$\gamma(4) \leqslant \alpha(3) + \beta(2).$$

Аналогично, можно понять, что $\beta' \leqslant \beta(2)$. Если посмотреть на числа, которые здесь стоят — 4, 3 и 2 — то это как раз те числа, которые должны фигурировать в неравенстве Вейля.

Легко понять, что все неравенства Вейля получаются именно таким образом: мы берем какой-то один треугольничек, и они тут возникают.

Более общие неравенства получаются, если мы будем рассматривать более хитрые конфигурации. Возьмем, например, конфигурацию из рис. 3.

Тут имеется 4 черных треугольника, соединенных коридорами. На коридорах нарисованы стрелки; двигаясь по этим стрелкам, мы переходим к соответствующим параллельным сторонам этих зачерненных треугольников. Каждый раз у нас при движении по этой стрелке, так сказать, поток или сумма может только убывать, а внутри черных треугольников никакой потери не происходит, потому что сумма приращений на двух катетах равна приращению на гипотенузе. Из этих соображений легко понять, что получается неравенство $\gamma(3)+\gamma(6)\leqslant\alpha(2)+\alpha(5)+\beta(1)+\beta(4)$. По форме оно похоже на неравенство Хорна (и фактически является им), но здесь уже не одно слагаемое фигурирует, а два.

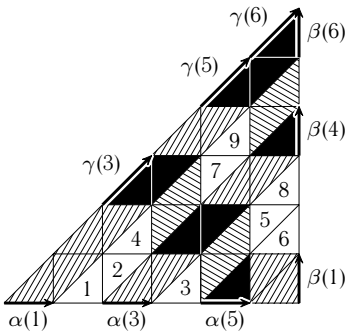
Естественно, что надо обобщить все это и рассматривать более сложные конфигурации. Такие конфигурации Кнутсона и Тао называли пазлами, мы предпочитаем называть их галереями.

Галерея состоит из черных треугольников, белых треугольников и коридоров между ними. Эти коридоры устроены так, что у них эта сторона и эта сторона черные, эта сторона и эта сторона белые; и правило такое, что черные стороны должны с черными соседствовать, а белые — с белыми. Так что здесь белый треугольник примыкает сюда. В общем, весь этот наш грид получается замощением; и это действительно похоже на пазл, т. е. когда какую-то картину нужно из каких-то маленьких кусочков замостить. Она у нас получается замощением из черных треугольников (таких и перевернутых), белых треугольников и, скажем, таких ромбических фишек. Главное, что у этих ромбических фишек некоторые стороны черные, которыми они должны примыкать к черным треугольникам или к черным же фишкам (потому что эти фишки могут одна за одной идти, как здесь было, одна к одной примыкают). Так что у нас вся эта поверхность делится на 3 группы: черные залы, белые залы и промежуточные коридоры между ними, которые нельзя отнести и не следует относить ни к черным, ни к белым, а это как бы переходы между этими залами. Можно представлять, что черные находятся на нижнем уровне, белые находятся выше, а вдоль коридоров мы можем ходить из черных в черные, а в этом же коридоре можно сверху ходить из белых в белые. Они раздваиваются на двух уровнях: на одном уровне можно ходить белым людям, а на другом — черным, они друг с другом не пересекаются.

Вот еще пример галереи, чуть более сложной.

Любая галерея дает некоторое неравенство. Например, галерея из рис. 4 дает неравенство $\alpha(1)+\alpha(3)+\alpha(5)+\beta(1)+\beta(4)+\beta(6)\geqslant\gamma(3)+{}$ ${}+\gamma(5)+\gamma(6)$. Причем я хочу заметить, что черных выходов на α-сторону

будет столько же, сколько выходов на β-сторону, и столько же, сколько выходов на γ-сторону; т. е. как у Хорна.



Р и с. 4.

Простая часть состоит в том, что любая такая галерея дает некоторое линейное неравенство на α, β и γ, причем того вида, как у Хорна: $\gamma(K) \leqslant \alpha(I) + \beta(J)$; I, J и K — это выходы черных коридоров или черных залов на границу, на эти 3 стороны. Это совсем элементарная вещь. А менее очевидная вещь состоит в том, что галерейные неравенства дают необходимые и достаточные условия того, чтобы α, β и γ продолжались до дискретно вогнутой функции. То есть что галерейные неравенства задают многогранник Г. Это утверждение, требующее довольно длительной возни и некоторой изобретательности. Но тут уже нет никаких матриц и спектров; это чисто задача полиэдральной комбинаторики, чисто задача линейного программирования. Тут никакой высокой науки использовать не нужно.

Как же этим свойством можно воспользоваться для доказательства обратного включения? Предположим, что наша точка γ лежит на границе Г. Это означает, что она удовлетворяет одному из галерейных равенств. Галерейные неравенства задают все неравенства для Г. И если у нас точка лежит на границе, значит, какое-то галерейное неравенство обращается в равенство. А теперь подумаем: а что означает обращение в равенство какого-то галерейного неравенства? Когда мы рассматриваем какой-то коридор, мы сравниваем приращение на входе этого коридора и на выходе этого коридора. И у нас были из-за дискретной вогнутости неравенства, что здесь приращение меньше, чем здесь. И эти приращения накапливаются. Если у нас теперь галерейное неравенство превращается в равенство, то это значит, что потерь никаких нет; что при движении по коридору у нас приращения равны. В частности, на концах коридора приращения одинаковые.

И это наводит на мысль, что мы теперь можем сделать некоторую процедуру, обратную к свертке. А именно, всю эту картинку, которая здесь нарисована, можно дезагригировать. Мы должны сделать вот что. Я говорил, что задать галерею — это значит разбить все на черные треугольники, белые треугольники и коридоры. Коридоры мы убираем, а все черные треугольники сдвигаем друг к другу; у нас здесь образуется такой чисто черный треугольник. Все белые треугольники тоже сдвигаем. Коридоры нам как бы не нужны, мы их никуда не сдвигаем, мы их просто

выбрасываем. А они нам нужны только для того, чтобы сказать, что здесь наша гипотетическая функция должна испытывать какое-то приращение, и здесь она испытывает приращение; мы эти два треугольника когда склеили, то у нас приращения на склеенных сторонах одно и то же. Поэтому все склеится в некоторую «черную» функцию f_1. Приращения, которые были на белых сторонах, склеятся в «белую» функцию f_2. И нетривиальное, но достаточно простое, утверждение состоит в том, что наша функция f будет сверткой черной функции f_1 и белой функции f_2, и каждая из этих функций будет дискретно вогнутая.

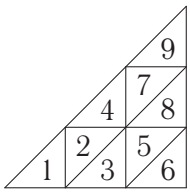


Рис. 5.

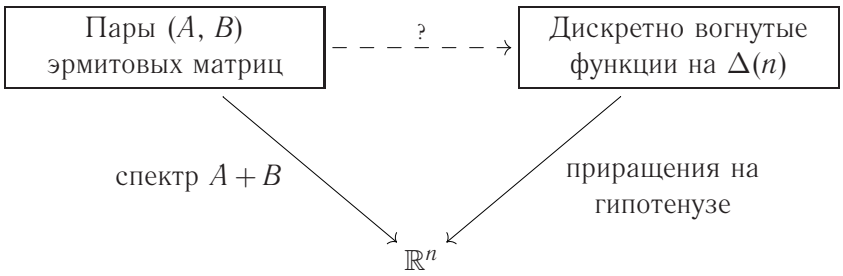
На рис. 5 я изобразил, как соберутся вместе белые треугольнички из рис. 4 и образуют треугольник размера 3.

Это означает, что я просто то, что я раньше говорил про разложение, проделываю в обратную сторону. У меня была γ, это было граничное значение функции f, когда я так разложил эту функцию, у меня получилось γ_1 и γ_2; это граничные значения некоторой дискретно вогнутой функции. По индукции мы знаем, что они происходят из некоторых эрмитовых матриц; нам же нужно от функций перейти к эрмитовым матрицам. Как это происходит? По индукции мы построим матрицы A_1 и B_1, которые соответствуют черной ситуации. Аналогично берем матрицы A_2 и B_2, которые соответствуют белой ситуации. Если взять прямую сумму, $A = A_1 \oplus A_2$ и $B = B_1 \oplus B_2$, то уже довольно легко убедиться, что мы здесь построили пару матриц A и B такую, что спектр их суммы равен γ.

Вот такая схема доказательства. Обсудим теперь кратко полученный ответ.

Мне этот ответ кажется каким-то недоговоренным. Мы получили совпадение двух множеств; мы получили, что пара эрмитовых матриц имеет какое-то смутное, не очень понятное отношение к дискретно вогнутым функциям. Когда мы с Глебом читали Кнутсона и Тао, мы никак не могли понять: ну а где же они по матрицам строят дискретно вогнутую функцию или хоникомбы? И где они по хоникомбу строят матрицы? Этого ничего не было. Мы просто видим, что у нас есть два множества, у которых образ один и тот же. Я сейчас даже не очень хочу объяснять, что это за множества. Вот одно множество, связанное с матрицами; другое множество — это какие-то дискретно вогнутые функции. Мы и то, и другое отображаем в камеру Вейля — пространство наборов чисел; и получаем от одной H, от

другой мы получаем Г.



С помощью каких-то хитрых (или грязных, как сказал бы Клячко) уловок, связанных с рассмотрением граничных точек — из-за того, что граница одного совпадает с границей другого, и пользуясь еще индукцией, я показал, что эти два множества H и Г совпадают.

Настоящее объяснение должно заключаться в том, что существует какое-то, по-видимому, более или менее явно выписываемое отображение, которое по паре матриц строит дискретно вогнутую функцию. И чтобы образ был один и тот же, конечно, это отображение должно быть сюръективным. Но, к сожалению, нам не удалось это доказать. А я только хотел бы предложить явную конструкцию такого отображения. Придумать его мы придумали, но не умеем для него доказывать то, что нужно. Это отображение описывается следующим образом.

Пусть у нас даны две матрицы A и B — симметричные или эрмитовы. По этим двум матрицам я на этом гриде определю некоторую функцию. Она определяется таким явным образом. Пусть дана точка с координатами (i, j). Рассмотрим максимум из следов (ясно, что следы какую-то должны играть роль) матриц $AP+BQ$, где P и Q — ортогональные проекторы и выполняется такое условие: что след (или ранг) P равняется i, $\operatorname{tr} Q = j$, и $P \geqslant Q$. Неравенство $P \geqslant Q$ означает, что разность между ними — это положительно определенная матрица.

Если на эту формулу посмотреть, то станет ясно, что приращение на одном основании — это в точности спектр A, на втором основании — спектр B, а на гипотенузе — спектр $C = A + B$.

Рассмотрим, к примеру, приращения на гипотенузе. В этом случае $i = j$. Значит, $P = Q$, потому что это два проектора с одинаковыми размерностями, и один содержит другой. Поэтому Q можно заменить на P и написать просто $AP+BP$; P можно вынести за скобку, и получится $(A+B)P$. А то, что такая вещь и дает в точности спектр или то, что нужно для этого — это известное свойство собственных значений.

Итак, мы привели довольно естественную формулу, определяющую

некоторую функцию на гриде. И наша гипотеза состоит из двух частей: первая — что это будет дискретно вогнутая функция, как мы уже ее определили; вторая — что любая дискретно вогнутая функция получается таким образом.

По виду тут все очень гладко. Мы с Глебом долго возились с доказательством. В нескольких частных случаях мы доказали, что это дискретно вогнутая функция *), но в общем случае это, так сказать, остается висеть в воздухе.

13 марта 2003 г.

*) Подробности можно найти в нашей статье «Дискретная выпуклость и эрмитовы матрицы» (Труды МИАН. Т. 241. 2003. С. 68—89).

А. М. Бородин

СЛУЧАЙНЫЕ ПЕРЕСТАНОВКИ, СЛУЧАЙНЫЕ СЛОВА И РАЗНОСТНЫЕ УРАВНЕНИЯ ПЕНЛЕВЕ

Случайные перестановки

Я начну свой рассказ с результата, который был доказан 5 лет назад, в 1998 г. Он обычно называется теоремой Байка—Дейфта—Йоханссона (Baik—Deift—Johansson 1998). Этот результат породил целый ряд исследований в соответствующем направлении и, в общем, сам по себе является замечательным. Чтобы его сформулировать я начну с симметрической группы — группы перестановок n символов.

Каждый элемент симметрической группы S_n удобно записывать в виде перестановки чисел от 1 до n. Например, 32145 можно рассматривать как элемент симметрической группы, переставляющий 5 символов. Когда мы записываем перестановку в таком виде, мы можем рассматривать возрастающие подпоследовательности такой последовательности. Например, 15 — это возрастающая подпоследовательность. Нас будут интересовать самые длинные возрастающие подпоследовательности. В данном случае, например, 145 — одна из самых длинных возрастающих подпоследовательностей; 245 имеет такую же длину. Таким образом для каждой перестановки $\sigma \in S_n$ я буду обозначать через $l_n(\sigma)$ длину максимальной возрастающей подпоследовательности.

Давайте теперь предположим, что все перестановки являются равновероятными. И тем самым сделаем l_n случайной величиной. Вопрос о распределении l_n достаточно заслуженный. В 1961 г. Станислав Улам высказал гипотезу по поводу поведения l_n. А именно, проведя какие-то эксперименты на компьютерах, что тогда еще было экзотикой, он предположил, что ожидание l_n ведет себя как константа, умноженная на корень из n: $\mathbb{E}l_n \sim c\sqrt{n}$. Вопрос оказался непростым. Сначала были частичные результаты. Наконец, в 1977 г. независимо Вершик и Керов в Ленинграде, а Логан (Logan) и Шепп (Shepp) в Америке доказали более сильный результат, а именно, не только, что математическое ожидание ведет себя как корень из n, вообще, случайная величина $l_n/\sqrt{n}$ сходится по вероятности

к константе, которая к тому же равна 2: $\frac{l_n}{\sqrt{n}} \to 2$. Это достаточно тонкий результат, на что указывает и промежуток времени, потребовавшийся для его получения.

Следующий вопрос, который хочется задать, — как ведет себя отклонение этой случайной величины от ее среднего; как асимптотически ведет себя дисперсия. Этот вопрос оказался еще более тяжелым. Ответ на него был найден как раз в работе Байка—Дейфта—Йоханнсона в 1998 г.; в ней было доказано, что если поделить случайную величину на $n^{1/6}$, то будет существовать предел распределения, который к тому же был посчитан. Я это напишу в таком виде:

$$\lim_{n\to\infty} P\left\{\frac{l_n - 2\sqrt{n}}{n^{1/6}} \leqslant x\right\} = F_2(x).$$

Это вероятность того, что центрированная нормированная случайная величина не превосходит x. Это некоторая аналитическая функция от x, которая, как это ни удивительно, возникала в науке раньше.

Такое распределение обычно называется распределением Трейси—Видома. Впервые оно было получено в 1993 г. как распределение максимального собственного числа случайной эрмитовой матрицы, когда размер матрицы стремится к бесконечности. Об этом я в данный момент рассказывать не буду — замечательное совпадение, которое имеет глубокие причины. Вообще, связь между перечислительными задачами комбинаторики и теорией случайных матриц оказалась богатой. На эту тему было написано несколько десятков работ.

Я же сконцентрируюсь в настоящий момент на этой функции F_2. Что же это за функция? Общая вера состоит в том, что нельзя написать формулу для этой функции, эта функция трансцендентна. Лучшее, что можно сделать — это написать некоторое дифференциальное уравнение, из которого эту функцию можно получить.

Дифференциальное уравнение Пенлеве II

Дифференциальное уравнение удобно писать не на саму функцию F_2, а на корень из ее второй производной. Возьмем вторую логарифмическую производную от функции F_2. Оказывается, такая функция отрицательна. Если подставить знак минус, то можно извлечь корень: $-(\ln F_2)'' = u^2$. Тогда новая функция u, определенная таким образом, удовлетворяет простому (на первый взгляд) дифференциальному уравнению $u''(x) = 2xu + u^3$. Это нелинейное дифференциальное уравнение второго порядка, которое обычно называется вторым уравнением Пенлеве.

Недостаточно знать уравнение, надо знать еще какие-то граничные условия. Это можно делать разными способами. Обычно их задают таким образом, что асимптотически $u(x) \sim -\mathrm{Ai}(x)$ при $x \to +\infty$. Можно показать, что тогда решение существует, единственно и соотносится с предельным распределением F_2 таким вот образом.

Надо сказать пару слов о том, откуда вообще берутся уравнения Пенлеве и почему они достойны собственного имени. Задача возникла в начале XX века. Пенлеве, французский математик, родился в 1863 г. Его работа относится примерно к первому десятилетию XX века. Удивительно, что потом он стал премьер-министром Франции. Даже, говорят, в Сорбонне, в Париже, есть небольшая, но очень симпатичная площадь Пенлеве.

Пенлеве ставил следующую задачу: давайте рассмотрим дифференциальное уравнение второго порядка, разрешенное относительно старшей производной $u'' = R(u', u, x)$, где правая часть является рациональной функцией по первой производной, алгебраической функцией от неизвестной функции и аналитической функцией от x. И мы хотим следующего. Вообще говоря, решения такого уравнения могут быть сколь угодно плохие. Однако мы требуем, чтобы для произвольных начальных условий решения нашего уравнения существовали во всей комплексной плоскости (мы рассматриваем x в комплексной плоскости), и при этом единственными возможными особенностями таких решений могут быть только полюсы, т. е. у решения не может появляться существенно особой точки, никакой точки ветвления, кроме каких-то наперед заданных точек. Например, легко понять, что для этого уравнения на бесконечности у функции u обязана быть существенная особенность. Неочевидное утверждение про это уравнение состоит в том, что для любых начальных условий, скажем $u(0)$ и $u'(0)$, существует единственное мероморфное решение этого уравнения во всей комплексной плоскости. Это очень общее условие. На другом языке оно часто выражается фразой, что у уравнения нет движущихся существенных особенностей — движущихся по отношению к начальным условиям. Если мы меняем начальные условия, то, в принципе, у нас могло бы меняться положение существенной особенности; этого происходить не должно.

Удивительно, что, после того как сформулированы граничные условия, все уравнения второго порядка, обладающие таким свойством (оно называется свойством Пенлеве), можно классифицировать. Точное утверждение состоит в том, что всякое такое уравнение, не сводящееся к уравнению первого порядка, приводится к одному из 6 уравнений, которые называются уравнениями Пенлеве (с соответствующими номерами — I, II, III, IV,

V, VI). Уравнение VI — это самое общее уравнение, при некотором предельном переходе оно спускается в уравнение V, и дальше есть некоторая простая схема вырождения.

Случайные перестановки и случайные матрицы

Для асимптотики нашей последовательности случайных величин l_n получился такой замечательный ответ. Вопрос, который будет меня сегодня занимать, это можно ли придумать что-то в таком же духе до того, как мы сделаем предельный переход. Мы можем написать функцию распределения для l_n; например, обозначим через $P_k^{(n)} = P\{l_n \leqslant k\}$ вероятность того, что $l_n \leqslant k$. Можно ли написать что-то похожее, какое-то разностное уравнение или что-нибудь вообще, для того чтобы посчитать эту функцию? Оказывается, что удобно работать не с самой последовательностью $P_k^{(n)}$, а с ее производящей функцией. Введем такое обозначение:

$$P_k^{(\theta)} = e^{-\theta^2} \sum_{n \geqslant 0} \frac{P_k^{(n)}}{n!} \theta^{2n}$$

где θ — это произвольное положительное число. Мы берем экспоненциальную производящую функцию; удобнее взять θ^2. Такая производящая функция имеет свою личную вероятностную интерпретацию и, вообще, оказывается значительно более удобной для рассмотрения.

Вероятностная интерпретация следующая. Возьмем первый квадрант и рассмотрим в нем процесс Пуассона с плотностью 1. Это означает, что у нас есть какая-то случайная точечная конфигурация в квадранте и при этом на единицу площади приходится в среднем по одной точке. Теперь возьмем в этом квадранте точку с координатами (θ, θ) и рассмотрим все ломаные, идущие из точки (0, 0) в точку (θ, θ) так, что они всегда идут на север и на восток и имеют местами излома как раз точки нашей случайной точечной конфигурации. Из всех таких ломаных нас интересует ломаная, которая по пути соберет наибольшее количество точек. Тогда утверждение состоит в том, что вероятность того, что количество точек на максимальной ломаной не превосходит k, будет в точности $P_k^{(\theta)}$.

Таинственного тут ничего нет. Легко увидеть связь между этой картинкой и перестановками. Если мы занумеруем точки в нашем квадрате, скажем, по координате x по порядку и посмотрим на координату y, то они каким-то образом переставятся. (В данном случае x — 5, 2, 1, 6, 4 и 3.) У нас, таким образом, возникнет некоторая перестановка, в данном случае на 6 символах. Количество точек является случайным. На самом деле, распределено экспоненциально. Как раз здесь производящая функция —

соответствующая функция распределения. И длина такой наибольшей ломаной — это в точности длина наибольшей возрастающей последовательности для перестановки.

Перед тем как формулировать соответствующее утверждение, я сформулирую еще одну интерпретацию чисел $P_k^{(\theta)}$. А именно, $P_k^{(\theta)}$ можно описать как некоторое среднее по унитарной группе размера k по мере Хаара. Среднее надо брать от такой простой величины: $e^{\theta^2} P_k^{(\theta)} = \mathbb{E}_{U(k)} e^{\operatorname{tr} \theta(U+U^{-1})}$. И это тут же дает, если воспользоваться формулой для радиальной части меры Хаара, представление этой величины в виде тёплицева определителя: $e^{\theta^2} P_k^{(\theta)} = \det[\varphi_{i-j}]_{i,j=1,\dots,k}$. Это — определитель матрицы $k \times k$, матричные элементы зависят только от диагонали, и при этом производящую функцию матричных элементов легко написать: $\sum \varphi_n z^n = e^{\theta(z+z^{-1})}$ — ровно то же самое, что стоит в этом среднем.

Можно написать еще два или три определения для $P_k^{(\theta)}$. Я напишу еще одно. $P_k^{(\theta)}$ имеет некоторый явный смысл с точки зрения теории представлений симметрической группы. А именно, это число можно представить как сумму по всем диаграммам Юнга произвольного размера с единственным ограничением, что первая строка по длине не превосходит k. Вес соответствующей диаграммы Юнга — это квадрат размерности неприводимого представления симметрической группы, отвечающего диаграмме λ. Еще нужно поделить на количество клеток диаграммы факториал в квадрате. И производящая функция та же самая, что и раньше. Опять мне надо поставить нормализационный множитель:

$$e^{\theta^2} P_k^{(\theta)} = \sum_{\lambda \leqslant k,\, \lambda \in \mathbb{Y}} \frac{\dim^2 \lambda}{|\lambda|!^2} \theta^{2|\lambda|}.$$

Такого рода веса называются мерой Планшереля для симметрической группы.

Разностное уравнение Пенлеве

Утверждение, которое я хочу сформулировать, состоит вот в чем. Давайте определим последовательность $v_0, v_1, v_2, \dots$ с помощью начальных условий $v_0 = 1$, $v_1 = -\frac{\varphi_1}{\varphi_0} = -\frac{I_1(2\theta)}{I_0(2\theta)}$. φ — это те φ, которые появлялись выше. Явно они пишутся через бесселевы функции. Я напишу в традиционных обозначениях, рассматривая их как коэффициенты соответствующего ряда. И есть рекуррентное соотношение, которое имеет следующий вид: $v_{n+1} + v_{n-1} = \frac{n v_n}{\theta(v_n^2 - 1)}$, $n \geqslant 1$. Тогда вторая логарифмическая производная чисел $P_k^{(\theta)}$, которая сейчас будет заменена на разностную

логарифмическую производную, есть $1-v_k^2$: $\frac{P_{k+1}P_{k-1}}{P_k^2}=1-v_k^2$, $k\geqslant 1$. (Тут везде стоит символ θ, который я опускаю, чтобы не загромождать обозначений.)

Разностное уравнение, которое здесь написано (нелинейное разностное уравнение второго порядка), тоже, как выяснилось, появлялось раньше. И не удивительно, что называется оно на самом деле разностное уравнение Пенлеве II. Я расскажу чуть больше об этом уравнении через некоторое время.

Пока я хочу сказать, что если взять просто это утверждение и произвести формальный предельный переход, то нетрудно убедиться, что в пределе возникнет как раз непрерывное уравнение Пенлеве II. И на самом деле такой способ может быть основой для эмпирического угадывания, какого рода формулировку здесь надо поставить и что получится в ответе, если мы хотим считать асимптотику P_k. Надо сказать, что задача вычисления асимптотики при $n\to\infty$ для $P_k^{(n)}$ эквивалентна (это не совсем тривиально, но не очень сложно) асимптотике $\theta\to\infty$ для $P_k^{(\theta)}$. Я раньше написал, что $P_k^{(n)}$, где k у меня будет $2\sqrt{n}+xn^{1/6}$, стремится к $F_2(x)$. Я утверждаю, что если заменить здесь n на θ, то получится эквивалентное утверждение. И зная о разностном уравнении Пенлеве, можно угадать, какого рода предельный переход надо брать, чтобы получился осмысленный предел у уравнения.

Что же такое разностное уравнение Пенлеве? Каким образом можно пытаться обобщить классическое свойство Пенлеве для непрерывных уравнений? Вопрос неочевидный, и, может быть, поэтому разностные уравнения Пенлеве не возникали или не были идентифицированы вплоть до начала 90-х годов.

Вот одно свойство этого рекуррентного соотношения, которое легко проверить, но которое, с другой стороны, совершенно не является очевидным, если его не знать. Очевидно, что если у меня $v_n=\pm 1$, тогда возникают некоторые трудности с определением v_{n+1}, поскольку в том выражении я вынужден делить на 0. Давайте смошенничаем. Слегка возмутим начальное условие таким образом, что v_n у меня будет ± 1 плюс малая величина ε. Тогда если мы разложим v_{n+1} в ряд по этому ε, то сначала у меня будет член с $1/\varepsilon$, а дальше будет ряд, ограниченный при ε, стремящемся к 0: $v_{n+1}=\varepsilon^{-1}+O(1)$.

Нетривиальное утверждение состоит в том, что v_{n+2}, v_{n+3} и все остальные члены имеют пределы при $\varepsilon\to 0$. То есть решение нашего уравнения продолжается через особую точку.

Я вас уверяю, что если вы попробуете написать произвольное нелинейное рекуррентное соотношение второго порядка, то такого свойства

вы не обнаружите, ваша особенность будет себя повторять. Это забавное свойство. По-английски оно называется confinement of singularities. Какое-то время предполагалось, что это есть правильный аналог условия интегрируемости в дискретной ситуации. Есть разные варианты того, что надо считать интегрируемостью для дискретных уравнений. Какое-то время был популярен этот ответ, но сейчас, видимо, считается, что это слишком слабое условие. Однако оно не такое уж слабое.

Это утверждение в случае конкретного уравнения, которое написано на доске, имеет очень изящную геометрическую интерпретацию. Интерпретация следующая. Давайте рассмотрим это уравнение как отображение из $\mathbb{C}^2$ в $\mathbb{C}^2$, которое берет пару v_{n-1}, v_n и отображает ее в пару v_n, v_{n+1}. У нас есть $\mathbb{C}^2$, которое бирациональным образом отображается опять-таки в $\mathbb{C}^2$. Оказывается, что можно придумать некоторую надстройку над этим $\mathbb{C}^2$; по-научному это будет называться $\mathbb{C}P^2$, раздутое в 9 точках. Я должен поставить здесь номер, поскольку раздутие будет зависеть от шага. Если это отображение будет f_n, то это f_n поднимется на моем раздутии до некоторого изоморфизма алгебраических многообразий. Если это отображение будет бирациональное, то можно явным образом разрешить особенности; то многообразие, на котором особенности разрешаются, это $\mathbb{C}P^2$, раздутое в 9 точках:

$$\begin{array}{ccc} \Sigma_n & \xrightarrow[\widetilde{f}]{} & \Sigma_{n+1} \\ \downarrow & & \downarrow \\ \mathbb{C}^2 & \xrightarrow[\widetilde{f}]{} & \mathbb{C}^2 \end{array}$$

Давайте я скажу, как это можно представлять себе геометрически. У нас есть $\mathbb{C}^2$, у нас есть какая-то такая динамическая система здесь, в какой-то момент мы не знаем, куда нам прыгнуть. Оказывается, что мы не знаем, потому что у нас неполное пространство: сюда нужно вклеить сферу, и тогда на следующем шаге мы прыгнем на эту самую сферу, а потом на следующем шаге ровно так же из нее и выпрыгнем. Это говорит о том, что v_{n+2}, v_{n+3}, … прекрасно определены. И тогда, если вы вклеите 9 сфер в $\mathbb{C}P^2$, то почему-то окажется, что все хорошо, дальше наша динамическая система всюду определена.

Здесь можно сказать больше: если взять $\mathbb{C}P^2$, раздуть его в 4, 5, 6, 7, 8 точках, то группа автоморфизмов соответствующего многообразия будет конечна; 9 точек — это первое число, где группа становится бесконечной. А если эти 9 точек находятся в общем положении, то тогда группа автоморфизмов — это аффинная группа $E_8^{(1)}$. В данном случае эта ситуация сильно вырожденная: одна точка кратности 5, одна двойная и две простые.

Случайные слова

Я немножко обобщу сейчас рассматриваемую ситуацию, с тем чтобы ответить на вопрос, насколько часто мы можем ожидать того, что эта комбинаторная задача типа перестановок ведет к алгебро-геометрической картинке. Я теперь хочу рассматривать не случайные перестановки, а случайные слова — всевозможные слова длины n, составленные из алфавита с l буквами. После этого я ровно так же определю $q_k^{(n)}$ — это будет распределение самой длинной неубывающей подпоследовательности моего слова (я предположу, что мой алфавит упорядочен — мои буквы это $a_1, \ldots, a_l$). Слова у меня будут опять равновероятны. Величина $q_k^{(n)}$ будет вероятность того, что длина не превосходит k. И это аналогично $P_k^{(n)}$. Ровно так же я хочу определить $q_k^{(\theta)}$ — ровно такую же производящую функцию.

На самом деле несложно доказать, что когда $l \to \infty$, q_k будут сходиться к P_k. Есть простое объяснение для этого: если у нас длинное слово и очень много букв, то вероятность того, что буква будет повторяться, очень мала, поэтому наше слово — это, практически, перестановка.

Теперь я хочу сказать, что есть аналог такого утверждения. Разница состоит только в том, что здесь в правой части у меня опять будет тёплицев определитель, в правой части я должен поставить функцию $e^{\theta z}(1+z^{-1})^l$. Есть аналог определения из теории представлений. Это определение выглядит так. Это будет сумма по диаграммам Юнга с первой строчкой, не превосходящей k. Разница состоит в том, что у меня будет не квадрат размерности, соответствующей симметрической группе, а будет произведение размерностей, одна из которых будет соответствовать симметрической группе, а вторая будет соответствовать $\mathrm{GL}(l)$, где l — это количество букв. И все, в общем-то, похоже, но только тут чуть более общее:

$$e^{\theta^2} q_k^{(\theta)} = \sum \frac{\dim_{S(\lambda)} \lambda \, \dim_{\mathrm{GL}(l)} \lambda}{l^{|\lambda|}|\lambda|!} \theta^{2|\lambda|}.$$

У нас другая нормировочная константа. Изменение одного факториала на $l^{|\lambda|}$ объясняется посчитанным количеством слов фиксированной длины: здесь l^n, а количество перестановок — это просто $n!$.

Вопрос в том, есть ли какое-то разностное уравнение, которое эту модель обслуживает. Утверждение состоит в том, что опять мы можем вычислить эту последовательность с помощью некоторого рекуррентного соотношения, которое, тем не менее, нельзя так уж легко записать. По крайней мере, не так изящно. Я напишу его на доске, тем не менее, чтобы просто было видно, насколько сложными получаются формулы.

Опять рекурренция будет 2-мерная, в том смысле, что ее можно описать как бирациональное отображение $\mathbb{C}^2 \to \mathbb{C}^2$. Но я уже не смогу написать его с помощью одной последовательности, для которой я просто беру два последовательных члена. У меня будет две последовательности, я их назову, скажем, f_s и g_s; и у меня будет отображение, которое из них делает новую пару $(f_{s+1},\ g_{s+1})$. Теперь давайте я напишу формулу:

$$f_s f_{s+1} = \frac{\theta^2 g_s}{(g_s - s - 1)(g_s + l - s - 1)},$$
$$g_s + g_{s+1} = \frac{\theta^2}{f_{s+1}} - \frac{s+1}{1 - f_{s+1}} - l + 2s + 3.$$

Удивительным образом после того, как эти уравнения написаны, оказывается, что такие соотношения опять-таки появляются не впервые. И более того, даже имеют специальное название. Нетрудно угадать, что это какое-то дискретное уравнение Пенлеве; в данном случае оно называется дискретное уравнение Пенлеве IV.

Каким же образом они появились раньше? Впервые такого рода соотношения были выписаны буквально полтора или два года назад японским математиком по фамилии Сакаи, который исследовал следующий вопрос. Он брал $\mathbb{C}P^2$, раздутое в 9 точках (некоторые из них могут совпадать) и рассматривал автоморфизмы этой поверхности. Оказывается, при некоторой конкретной конфигурации точек на $\mathbb{C}P^2$ это и есть формула, которая дает при соответствующем поднятии изоморфизм этого $\mathbb{C}P^2$, раздутого в 9 точках. У Сакаи на самом деле были более общие уравнения: дискретные уравнения Пенлеве V и VI. Если взять точки в общем положении, то возникает то, что Сакаи назвал эллиптическим уравнением Пенлеве VI. Рекурренция пишется через эллиптические функции, и применения этой рекурренции в другой науке я ни разу не видел. Нумерация, надо сказать, несколько произвольная. Сакаи ее получает просто из группы симметрий соответствующей поверхности, но, например, одно и то же дискретное уравнение Пенлеве может вырождаться в разных режимах в разные непрерывные уравнения Пенлеве. В общем, эта наука более сложная, чем непрерывная.

Я должен закончить утверждение, которое состоит в том, что если я напишу опять вторую логарифмическую производную, т. е. $\frac{q_{k+1}q_{k-1}}{q_k^2}$, то это пишется как некоторая явная рациональная функция от последовательности $f_k,\ g_k$.

Матричные полиномы

Последующий шаг будет состоять в том, что я обобщу ситуацию еще дальше. Чтобы объяснить, как я это сделаю, я дам еще одну интерпретацию для последовательности $q_k^{(\theta)}$. Интерпретация очень близка той формуле, которая тут написана, единственно, теперь я хочу переписать эту формулу в терминах строчек моих диаграмм Юнга. Я введу новые координаты: $\lambda_i + l - i$ я обозначу через x_i, и тогда окажется, что то, что мне нужно суммировать, это такое выражение:

$$q_k = \sum_{x_1 > x_2 > \ldots > x_l,\quad x_1 \leqslant k+l-1} \prod (x_i - x_j)^2 \prod w(x_i).$$

Это определитель Вандермонда (взятый от этих x_i) в квадрате, умноженный на некоторую мультипликативную функцию от координат x_i. Суммирование берется по всем монотонным целым последовательностям из l иксов, где первое максимальное не превосходит $k+l-1$. Я не сказал, что такое w. $w(x)$ в данном случае просто $\frac{\theta^{2x}}{x!}$, $x_i \in \mathbb{Z}_+$. Выражения такого рода на самом деле часто появляются в теории случайных матриц. Если взять гауссовскую меру на эрмитовых матрицах и спроектировать ее на собственные значения, то якобиан как раз будет определитель Вандермонда в квадрате.

И общее утверждение, которое я хочу сделать, состоит в том, что для некоторого общего класса весовых функций w, входящих в эту формулу, существует рекуррентное соотношение, позволяющее получить соответствующую последовательность. Если наш вес имеет логарифмическую производную, которая есть рациональная функция $\frac{w(x-1)}{w(x)} = \frac{P(x)}{Q(x)}$ ($P(x)$ и $Q(x)$ — многочлены), то существует некоторая рекурренция, которая в частных случаях позволяет вычислять такое распределение первой частицы по вероятностной мере такого вида. Как же эту рекурренцию строить? Уже для такого простого случая я вынужден писать на доске довольно нелегкие формулы. Оказывается, что правильный язык, на котором нужно разговаривать, это язык матриц, более того, матричных многочленов.

Рекурренция устроена так:

$$(A_s,\, M_s(z)) \to (A_{s+1},\, M_{s+1}(z)).$$

Шаг рекурренции состоит в том, что у нас есть пара: нильпотентная матрица A_s размера 2×2 и матричный многочлен $M_s(z) = M_s^{(0)} z^m + \ldots + M_s^{(m)}$ (это просто есть формальная линейная комбинация коэффициентов, каждый из которых есть некоторая матрица); некоторое отображение, которое

по такому объекту строит следующий объект, и рецепт, как же это происходит. Рецепт состоит в одной строчке. Сейчас я его выпишу:

$$M_s(z)(z + A_s - s) = (z + A_{s+1} - s)M_{s+1}(z).$$

То, что я сейчас пишу, это некоторое тождество матричных многочленов; z — это формальная переменная, которая со всем коммутирует. Идея состоит в том, что если я знаю левую часть (нильпотентную матрицу 2×2 и матричный многочлен), то тогда это тождество единственным образом определяет нильпотентную матрицу и матричный многочлен в правой части. Чтобы это было более конкретно, я скажу, что в этом случае, с которого я начал, матрица M_s линейна по z. В общем случае на самом деле степень M_s есть максимум степени P и степени Q; в данном случае эта степень равна 1. Старший член — линейный член, его легко написать. Дальше идет некоторая неизвестная матрица.

В случае, когда $w(x) = \dfrac{\theta^{2x}}{x!}$, M_s линейно по z.

Вообще говоря, A_s — это какая-то нильпотентная матрица, можно написать ее матричные элементы. После этого у меня возникнет 6 переменных — 4 для M_s и 2 для A_s:

$$M_s(z) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} z + \begin{pmatrix} \alpha_s & \beta_s \\ \gamma_s & \delta_s \end{pmatrix}, \quad A_s = \begin{pmatrix} p_s & q_s \\ r_s & -p_s \end{pmatrix}, \quad p_s^2 = -r_s q_s.$$

Если я разрешу матричное уравнение, записанное в таком компактном виде, то оказывается, что если я далее правильно параметризую мои матрицы, мне удастся исключить 4 из 6 переменных, и оставшиеся 2 переменные будут эволюционировать согласно этой динамике, которая затем ведет к некоторой алгебраической геометрии.

Тем не менее в общем случае, конечно, явно посчитать ничего не удается. Исключать переменные можно в каждом конкретном случае. Но общий случай ведет к такой матричной динамике. И чтобы как-то прояснить, почему такого рода матричное соотношение позволяет определить динамику, я хочу сформулировать некоторую лемму, которая для меня была неожиданной.

Это чисто линейно-алгебраическое утверждение. Для почти всех матриц произвольного порядка $S, T \in \mathrm{Mat}(m, \mathbb{C})$ (я еще специально потребую, чтобы $\mathrm{Sp}\ S \cap \mathrm{Sp}\ T = \varnothing$) существуют единственные матрицы $\widehat{S}$ и $\widehat{T}$ такие, что $\mathrm{Sp}\ S = \mathrm{Sp}\ \widehat{S}$, $\mathrm{Sp}\ T = \mathrm{Sp}\ \widehat{T}$ и имеет место равенство следующих матричных многочленов:

$$(z - S)(z - T) = (z - \widehat{T})(z - \widehat{S}).$$

Если приравняем коэффициенты, то мы увидим, что это равносильно тому, что $S + T = \widehat{S} + \widehat{T}$ и $ST = \widehat{T}\widehat{S}$.

Если мы вернемся к тому, что написано здесь, и предположим, например, что M_s линейны по z, то тогда как раз это есть перестановка двух линейных множителей.

Изомонодромные преобразования

Во второй части доклада я хочу рассказать, каким образом матричные соотношения возникают совсем в другом сюжете, связанном с тем, как на дискретную ситуацию обобщить изомонодромные преобразования.

Два слова о том, что такое изомонодромное преобразование в обычном, классическом случае. Задача восходит, по-видимому, к Риману. Мы рассмотрим матричное обыкновенное дифференциальное уравнение

$$\frac{dy(\zeta)}{d\zeta} = \sum \frac{B_i(x)}{\zeta - x_i} y(\zeta).$$

Оно будет иметь, конечно, особенности в точках (x_i). Решение будет ветвиться: если мы обойдем вокруг точки x_i, то, если мы возьмем фундаментальное решение, оно домножится справа на некоторую матрицу; она зависит только от гомотопического класса соответствующей петли: $y \to y \cdot M_\gamma$.

Не вполне очевидно, как такого рода картинка должна обобщаться на дискретную ситуацию, поскольку уравнение, которое я хочу теперь рассматривать — что-нибудь в таком роде: $Y(z+1) = A(z)Y(z)$. После этого у меня не будет никакого ветвления. Тем не менее, понятие изомонодромной деформации остается.

Я не сказал, что такое изомонодромная деформация. Изомонодромная деформация — это когда мы хотим менять положение точек (x_i) и хотим при этом, чтобы матрицы монодромии M_γ оставались постоянными. Для этого нам надо потребовать, чтобы коэффициенты зависели от x_i. И тогда задача состоит в том, можно ли вообще это сделать. Задача была решена в 1912 г. Шлезингером. Он написал некоторую систему уравнений с частными производными на B_i, которая позволяет это сделать.

Во второй части доклада я попытаюсь рассказать, что надо делать в дискретной ситуации, как надо понимать монодромию и изомонодромную деформацию как матричные соотношения.

Линейные разностные уравнения

Теперь я абстрагируюсь от первой части своего доклада. Я перейду просто к обсуждению разностных матричных линейных уравнений вида $Y(z+1) = A(z)Y(z)$, где Y и A — это некоторые квадратные матрицы произвольного порядка, причем A рациональная. Легко сообразить, что на самом деле мы не обязаны рассматривать рациональную A, доста-

точно рассматривать полиномиальную, потому что когда мы умножим Y на подходящую Г-функцию, тем самым A умножится или разделится на любой линейный множитель от z. Поэтому все знаменатели, которые есть в A, можно сократить. То есть с этого момента я буду предполагать, что A — многочлен. Более того, если я все сопрягу и предположу, что оно в общем положении, то можно считать, что у меня старший коэффициент диагонален:

$$\begin{pmatrix} \rho_1 & & \\ & \ddots & \\ & & \rho_n \end{pmatrix} z^m + A_1 z^{m-1} + \ldots + A_m.$$

Хороший пример такого уравнения, который уже несет с собой много интересных вещей, это просто Г-функция, которая удовлетворяет известному уравнению: $\Gamma(z+1) = z \cdot \Gamma(z)$. Давайте я уберу отсюда Γ и напишу Y, чтобы не было соблазна заменять общее решение частным: $y(z+1) = zy(z)$.

Хочется задать вопрос, что такое монодромия для уравнения, соответствующего Г-функции? Как вообще ввести понятие чего-то типа изомонодромного преобразования, сохраняющего ветвление?

Теперь я потрачу несколько минут на изложение некоторой теории, которая принадлежит Джорджу Биркгофу. Она была развита им в двух работах, опубликованных в 1911 и 1913 годах. Это замечательные работы, которые остались не очень известными; их читали, но мало.

Утверждение первое. Существует единственное формальное решение уравнения, имеющее вид

$$\widehat{Y}(z) = \left(I + \frac{\widehat{Y}_1}{z} + \frac{\widehat{Y}_2}{z^2} + \ldots \right) \begin{pmatrix} \rho_1^z z^{d_1} & & \\ & \ddots & \\ & & \rho_n^z z^{d_n} \end{pmatrix} z^{nz} e^{-nz}.$$

У решения $\widehat{Y}(z)$ есть сначала часть, которая выглядит как голоморфная функция в окрестности ∞, затем некоторая диагональная часть, которую можно угадать, и на диагонали еще появляются степенные множители, плюс некоторые скалярные множители z^{nz} и e^{-nz}. Что я подразумеваю под формальным решением уравнения? Если взять выражение такого вида, подставить его в уравнение и разложить обе части в формальный степенной ряд по $1/z$, то тогда можно подобрать (единственным образом) такие константы — скалярные матрицы $\widehat{Y}_1$, $\widehat{Y}_2$, …, а также константы d_1, …, d_n, что это уравнение превратится в тождество.

Для уравнения на Г-функцию такой степенной ряд выглядит следующим образом:

$$\widehat{y}(z) = \left(1 + \frac{1}{12z} + \ldots \right) e^{-z} z^{z-1/2}.$$

Это на самом деле есть не что иное, как асимптотическое разложение Г-функции, с точностью до константы.

Неприятность состоит в том, что степенной ряд, который здесь появляется, всюду расходится. Конечно, Г-функция не голоморфна в окрестности бесконечности. Тем не менее, она имеет такое асимптотическое разложение. Тем самым, это действительно решение формальное.

Второе утверждение Биркгофа состоит в том, что в левой или правой полуплоскости существует и единственно голоморфное обратимое решение (я это буду обозначать Y^l или Y^r) нашего уравнения, которое асимптотически равно $\widehat{Y}$, т. е. моему формальному ряду.

Конечно, если решение существует в полуплоскости, то оно существует всюду: мы можем использовать уравнение, чтобы его шаг за шагом продлить налево или направо. Однако важно, что мы можем вычислить асимптотику нашего решения только лишь в полуплоскости. Если мы потребуем, чтобы наше решение было асимптотически равно такому выражению справа или слева, ответ получится разным, и решения получатся разные.

Для уравнения на Г-функцию

$$\frac{\Gamma(z)}{\sqrt{2\pi}} = y^r(z), \qquad \frac{\sqrt{2\pi}e^{i\pi z - i\pi/2}}{\Gamma(1-z)} = y^l(z).$$

Решение этого скалярного уравнения, которое в правой полуплоскости имеет такую асимптотику, это просто есть Г, отнормированная на $\sqrt{2\pi}$ — это Y^r. $Y^l(z)$ похоже, но чуть-чуть другое: на самом деле это $\sqrt{2\pi}/\Gamma(1-z)$, но только еще подправленное на экспоненту. Все решения определены с точностью до периодических функций: всегда можно умножить решение Y справа на периодическую матрицу, от этого решение останется решением. Поэтому домножение на периодические функции не меняет решения. Зачем я сейчас умножаю на периодическую функцию, я объясню позже.

Единственность появляется после того, как я зафиксирую ветви логарифма, которые мне нужны, чтобы определить асимптотическое поведение степеней z, которые здесь участвуют. Как только я зафиксирую ветви логарифма, решение становится единственным. Если я меняю ветви логарифма, то решения домножаются справа на некоторую фиксированную периодическую функцию.

Утверждение третье. Давайте при определении левого и правого решения зафиксируем ветви логарифма таким образом, что они совпадают в верхней полуплоскости. У меня есть левая и правая полуплоскости, мне надо зафиксировать две ветви. Я зафиксирую их таким образом, что они совпадают при уходе на бесконечность. После этого давайте рассмотрим

отношение левого и правого решения. Надо взять правое решение в минус первой степени и умножить на левое: $(Y^r)^{-1}Y^l = P$. Тогда такое отношение есть периодическая матрица, потому что, если мы сдвинем это на единицу, то здесь появится матрица коэффициентов $A(z)$, а здесь появится такая же матрица, и они сократятся. Это утверждение аналогично утверждению об обыкновенном дифференциальном уравнении, а именно, о том, что две фундаментальные системы решений всегда отличаются на постоянную матрицу. Здесь постоянная матрица заменяется периодической.

Опять-таки в случае Г-функции несложно посчитать, что если я возьму отношение этих двух решений, то у меня получится $P = 1 - e^{-2\pi iz}$.

На самом деле можно показать, что матричные элементы этой матрицы всегда многочлены от $e^{\pm 2\pi iz}$. И можно точно указать степень этого многочлена: будет видно, какое количество как бы свободных констант есть в этом многочлене.

Я описал некое соответствие, которое по матрице коэффициентов $A(z)$ производит на свет периодическую матрицу $P(z)$. Совершенно неочевидное утверждение состоит в том, что такое преобразование является аналогом преобразования или отображения, которое в непрерывном случае ставит матрице коэффициентов уравнения в соответствие матрицу монодромии, матрицу ветвления. Я не смогу объяснить, почему это действительно аналог. Одним из оправданий служит то, что в предельном переходе, который переводит разностное уравнение в непрерывное, по предельному поведению периодической матрицы P можно получить матрицу монодромии соответствующего предельного непрерывного уравнения. Второе объяснение состоит в том, что преобразования разностного уравнения, сохраняющие эту периодическую матрицу P, в предельном переходе опять-таки переходят в потоки, сохраняющие монодромию.

Теперь я буду интересоваться тем, могу ли я каким-то образом поменять мою матрицу $A(z)$ так, что $P(z)$ останется ровно той же. Еще одно утверждение Биркгофа состоит в следующем. Если у нас есть $A'(z)$, $A''(z)$, которые соответствуют одной и той же P, то существует такая рациональная матрица $R(z)$, что левое и правое решение одной задачи связано с левым и правым решением другой задачи просто домножением на эту рациональную матрицу. И соответственно, матрица коэффициентов простым образом связана с той — связана почти сопрягающим преобразованием:

$$(Y')^l = R(Y'')^l, \quad (Y')^r = R(Y'')^r,$$
$$A''(z) = R(z+1)A'(z)R^{-1}(z).$$

То есть если мы верим, что $P(z)$ — это аналог монодромии, то всякое изомонодромное преобразование получается с помощью некоторой раци-

ональной матрицы. Это объясняет, как должно быть устроено изомонодромное преобразование, но не говорит о том, есть ли они вообще. Значит, если оно существует, оно устроено именно так.

Я повторю утверждение. Если у меня есть две матрицы коэффициентов, которые дают одинаковую матрицу монодромии, то тогда существует некоторая рациональная матрица R, которая связывает одну и другую.

Я пытаюсь привести пример R для популярного уравнения, отвечающего Γ-функции. Я возьму $r(z)=z$. Тогда я должен получить что-то в таком духе: новое правое решение есть z, умноженное на старое правое решение, а новое левое решение есть z, умноженное на старое левое решение. Это решение было практически Γ-функция, только нормированная: $z\Gamma(z)/\sqrt{2\pi}$, т. е. $\Gamma(z+1)/\sqrt{2\pi}$; $\Gamma(z+1)$ удовлетворяет новому уравнению:

$$\Gamma(z+1+1)=(z+1)\Gamma(z+1).$$

Поэтому моя новая матрица коэффициентов есть $z+1$ — я начал со старой z, получил новую $z+1$. Сопрягающая матрица r — это просто функция z в данном случае.

Следующее утверждение, которое я сейчас сформулирую, это классификация всех таких матриц R, которые дают изомонодромное преобразование.

Мы фиксируем (в наших обозначениях) $A'(z)$ и хотим найти все рациональные матрицы R так, что это выражение дает $A''(z)$, которое будет многочленом того же порядка и будет иметь такой же вид. Конечно, эта матрица всегда будет рациональная, однако мы хотим, чтобы новая матрица находилась ровно в том же пространстве, что и старая. Старший коэффициент диагональный, и главное — многочлен, он не имеет полюсов в конечной плоскости.

Сейчас я объясню, как такие матрицы, вообще, получаются. Для этого надо посмотреть на нули определителя $\det A(z)=c(z-a_1)\ldots(z-a_{mn})$. Тогда утверждение состоит вот в чем. Для любых целых сдвигов $a_i\mapsto a_i+\varkappa_i$, $\varkappa_i\in\mathbb{Z}$ (теперь я вспомню, что у меня еще были некоторые показатели d, про которые я не очень много говорил; давайте рассмотрим $d_i\mapsto d_i+\delta_i$, δ_j тоже произвольные целые числа; таким образом, что $\sum\varkappa_i+\sum\delta_i=0$) существует и единственна матрица $R(z)$ такая, что выражение

$$\widetilde{A}(z)=R(z+1)A(z)R^{-1}(z)$$

отвечает $\widetilde{a}_i=a_i+\varkappa_i$; $\widetilde{d}_j=d_j+\delta_j$.

Это утверждение не Биркгофа, а мое. Впрочем, я несколько рискую, когда это говорю, потому что в 1923 г. кто-то запросто мог его доказать.

Просто на самом деле возможности поиска в литературе 20-х годов довольно ограниченные. Все, что я видел, не указывает ни на одну работу, ни на одного человека, который вообще задавался вопросом рассмотрения всевозможных преобразований, сохраняющих P.

Обращаясь опять к уравнению для Г-функции, $y(z+1)=zy(z)$, мы видим, что у матрицы коэффициентов один единственный нуль, равный нулю. И мы можем делать произвольные сдвиги этого нуля на целые числа. На самом деле все уравнения, которые мы можем получить из этого уравнения, есть просто сдвиг матрицы коэффициентов на целое число: $y(z+1)=(z+k)y(z)$. Здесь мы сделали сдвиг на единичку. Если в данном случае мы посмотрим на асимптотический ряд, то увидим, что $d=-1/2$. У нас одно d, одно a, мы сдвигаем a вправо, автоматически возникает сдвиг d влево. Это все преобразования для уравнения с Г-функцией.

Теперь я хочу выделить некоторую подгруппу: $\mathbb{Z}^{mn+n-1}\supset\mathbb{Z}^m$. У меня есть mn нулей, у меня еще есть n штук d; я должен вычесть единицу, поскольку сумма равна 0. Я хочу выделить некоторую замечательную подгруппу $\mathbb{Z}^m$. Здесь n — это размер квадратной матрицы, а m — это степень многочлена $A(z)$. Я хочу выделить m-мерную подгруппу в этой группе, которая проще, чем общие преобразования. В частности, я буду в состоянии указать более явно, каким образом строится этот загадочный $R(z)$, про который я ничего пока не рассказал.

Что же это за подгруппа? Мне надо выделить подгруппу в этих целых сдвигах. Давайте я начну с того, что множество нулей я разобью на группы, по n собственных чисел в каждой группе. В первой группе у меня будет от 1 до n, и так до m-й группы:

$$\{a_i\}=\{a_1^{(1)},\ \ldots,\ a_n^{(1)}\}\sqcup\ldots\sqcup\{a_1^{(m)},\ \ldots,\ a_n^{(m)}\}.$$

Я это делал произвольным образом; я предполагаю, что все собственные значения различны. Теперь я захочу сдвигать собственные числа в каждой группе на одно и то же число. Я буду рассматривать целые сдвиги, теперь я прибавляю ко всем числам в группе одно-единственное число. Значит, $a_i^{(k)}$ сдвигается на какое-то μ_k для всех $i=1,\ \ldots,\ n$. Таким образом, возникает m разных сдвигов. Я еще хочу, чтобы все d тоже сдвигались на одно и то же число, которое обязано быть минус суммой μ_k, потому что общая сумма сдвигов равна 0.

Чем же приятна такая подгруппа? Приятна она, например, вот чем. Давайте возьмем $\mu_1=1$, $\mu_2=\ldots=\mu_m=0$ (все μ кроме первого равны 0). Как же тогда описать $R(z)$? Тогда $R(z)$, оказывается, есть просто линейная функция, матричный линейный многочлен, я напишу его в виде $z-B$, где

это есть единственный правый делитель матрицы $A(z)$ с заданным спектром — со спектром $\{a_1^{(1)}, \ldots, a_n^{(1)}\}$ (первая группа собственных значений). Я беру матрицу $A(z)$ и хочу отфакторизовать линейный множитель справа $A(z) = \widehat{A}(z)(z - B)$ таким образом, чтобы спектр матрицы B в правой части был предписанным с самого начала. Он, конечно, должен находиться среди нулей определителя моей исходной матрицы, но как только я фиксирую это подмножество нулей, я могу это сделать единственным образом. И оказывается, что это $z - B$ и есть соответствующий правый делитель.

Давайте это проверим. Я посчитаю $\widetilde{A}(z)$. Тогда я должен написать $R(z + 1)$, т. е. я должен написать $z + 1 - B$ умножить на $A(z)$, которое я напишу в виде $A(z) = \widehat{A}(z)(z - B)$, и умножить на $R^{-1}(z)$, т. е. на $(z - B)^{-1}$, который благополучно сокращается. Таким образом, моя матрица $\widetilde{A}(z)$ имеет такой вид:

$$\widetilde{A}(z) = (z + 1 - B)\widehat{A}(z)(z - B)(z - B)^{-1} = (z + 1 - B)\widehat{A}(z).$$

Если я хочу продолжать, например, если я опять хочу взять $\mu_1 = 1$, то тогда мне опять в этой матрице $\widetilde{A}$ нужно выделить правый делитель. Я это должен написать в виде $\widehat{\widetilde{A}}(z)$ умножить на какой-то $z + 1 - \widetilde{B}$. Посмотрим на это соотношение, забудем про все остальное, просто сравним два выражения. То, что здесь написано, есть матричный многочлен; линейный матричный множитель. И мы этот линейный матричный множитель проносим через наш матричный многочлен и получаем какой-то другой результат:

$$\widetilde{A}(z) = (z + 1 - B)\widehat{A}(z) = \widetilde{\widehat{A}}(z)(z + 1 - \widetilde{B}).$$

Именно такого рода соотношение появлялось в динамике, нужной мне для вычисления вероятностной величины, о которой я говорил. Это не случайно. На самом деле всякая вероятностная модель из описанных ранее естественным образом приводит сначала к некоторой линейной задаче, к разностному линейному уравнению, а потом к соответствующей изомонодромной деформации этого уравнения. И как раз уравнения на изомонодромные деформации приводят к рекурренциям, которые в частных случаях дают разностное уравнение Пенлеве, т. е. то, с чего я начал.

Ответы на вопросы и обсуждение

Разбиение нулей определителя не нужно фиксировать каждый раз заново. Когда я выделяю эту подгруппу, я уже фиксирую разбиение на некоторые группы; теперь я каждую из этих групп сдвигаю на свое целое число. Я взял первую группу, сдвинул ее на 1 один раз, потом я хочу

сдвинуть ее на 1 еще раз, и для этого мне нужно выделить в той же матрице еще один правый делитель.

Уравнение Пенлеве, вообще говоря, не возникает так, оно возникает как самоподобная редукция для КдФ, это не лично одно из уравнений КдФ. Есть некая деятельность здесь, связанная с тем, чтобы получать, так сказать, разностные уравнения Пенлеве из каких-то дискретных аналогов интегрируемых моделей. Эта деятельность пока на мой взгляд не привела к большому успеху, но это отчасти потому, что не очень понятно, что такое разностный аналог Пенлеве; имеется разногласие, по-моему так же, как и имеется разногласие, каким методом следует строить разностные уравнения Пенлеве. Разные люди подразумевают под этим словом разные вещи.

Случайные последовательности, задача о вычислении распределения наибольшей длины возрастающей подпоследовательности и т. д. могут быть сформулированы в контексте общей задачи. Суть состоит вот в чем. Мы рассматриваем вероятностную меру на наборах целых чисел, которая выглядит следующим образом. У нас есть определитель Вандермонда от позиции этих точек в квадрате, умноженный на произведение весовой функции от этих точек:

$$\prod (x_i - x_j)^2 \prod w(x_i).$$

Это мера на конечных последовательностях иксов. И мы хотим узнать о том, как по мере такого вида распределен максимум этих иксов. Чтобы это вычислить, оказывается, нужно сделать следующее: нужно просто вычислить ортогональный многочлен. Я предполагаю, что мои иксы — это $0, 1, \ldots$ Я хочу вычислить вероятность того, что максимум не превосходит какого-то k. Тогда я это могу сделать, например, таким образом. Если я умею вычислять ортогональные многочлены на интервале от 0 по k с весовой функцией w, то это дало бы мне ответ на такую задачу. К сожалению, мы, как правило, такие многочлены вычислить не можем. Даже при классическом весе это не вычисляется. Но такие ортогональные многочлены заведомо удовлетворяют некоторому разностному линейному уравнению, если наш вес достаточно хороший. Теперь же если мы возьмем разные значения k, например k и $k+1$, то у нас получатся две разные системы ортогональных многочленов и две разные системы разностных линейных уравнений, связанных с этими ортогональными многочленами. Эти две системы оказываются в одном классе эквивалентности в этом смысле, связаны некоторым преобразованием такого вида. Таким образом, чтобы предъявить какую-то рекурренцию по k для распределения этого максимума, $k, k+1, \ldots$, мы можем вместо этого рассматривать изо-

монодромные преобразования такого рода системы. И это оказывается правильным способом описания.

Есть следующая гипотеза. Если я рассмотрю меру такого вида, забуду про то, что иксы дискретны и посажу их на вещественную прямую, возьму здесь гауссовский вес, рассмотрю распределение точек по этому весу и возьму правильным образом предел при количестве точек, стремящемся к бесконечности, то чудесным образом оказывается, что ответ совпадает с предельным распределением нулей ζ-функции Римана, если мы по критической прямой уходим в бесконечность. Это связано ровно таким образом. Гипотеза частично доказана, но полного доказательства нет. И есть общая вера в то, что нули ζ-функции Римана ведут себя как собственные числа случайных матриц. Объяснения же этому, в настоящий момент, по крайней мере, нет. Есть какое-то объяснение для ζ-функции над функциональными полями. Но над комплексным полем мне неизвестно никаких концептуальных работ, которые бы это объяснили.

27 марта 2003 г.

О. В. Шварцман

50 ЛЕТ ТЕОРЕМЕ ШЕВАЛЛЕ

Этот доклад посвящен теореме Шевалле, доказанной где-то между 53-м и 54-м годом прошлого века, и некоторым ее более поздним обобщениям.

Начну издалека: пусть на комплексном многообразии X действует дискретная группа его автоморфизмов $\Gamma \subset \operatorname{Aut} X$. Под дискретностью действия понимается следующее: для любого компакта $K \subset X$ множество $\{\gamma \in \Gamma\colon \gamma K \cap K \neq \varnothing\}$ конечно, т. е. почти все элементы γ группы Γ «уводят» компакт K так далеко, что он не пересекается со своим образом γK. Это относится, в частности, и к любой точке $x \in X$. Таким образом, стабилизатор $\Gamma_x = \{\gamma \in \Gamma\colon \gamma x = x\}$ конечен для любой точки $x \in X$.

Группа Γ называется кокомпактной (иногда кристаллографической), если факторпространство $Y = X/\Gamma$ или $X \pmod \Gamma$ компактно. Я не предполагаю, что группа Γ действует без неподвижных точек (скоро мы поймем, почему это важно). Поэтому Y является комплексным орбифолдом.

Через $\operatorname{Fix} \gamma$ обозначим множество неподвижных точек элемента γ в X. Главная роль в моем рассказе досталась отражениям. Вот их определение: элемент $\gamma \in \operatorname{Aut} X$ назовем *комплексным отражением* (в дальнейшем просто *отражением*), если его порядок конечен и $\operatorname{codim}_{\mathbb{C}} \operatorname{Fix} \gamma = 1$.

Примеры. а) симметрия в любой точке комплексной прямой $\mathbb{C}$;

б) линейное преобразование $\mathbb{C}^{n+1}$ с собственными значениями

$$\Big(\underbrace{1, \ldots, 1}_{n}, \exp\big(2\pi\sqrt{-1}/N\big)\Big).$$

Вернемся к компактному комплексному орбифолду $Y = X/\Gamma$. Общая теория предсказывает, что, как правило, Y оказывается проективным алгебраическим многообразием. Хотелось бы, чтобы это алгебраическое многообразие было как можно «проще»: например, комплексным проективным пространством $\mathbb{C}P^n$ или его «скрученным» обобщением — взвешенным проективным пространством WP^n. Возникает законный вопрос: для каких пространств X и групп Γ такое бывает?

Скажу сразу — вопрос не из легких. Классическая теорема Шевалле полностью на него отвечает в том случае, когда X — это проективное пространство $\mathbb{C}P^n$, а Γ — конечная группа его автоморфизмов.

Теорема Шевалле

В удобной для наших целей формулировке она звучит так:

Теорема 1. *Факторпространство $\mathbb{C}P^n/\Gamma$ тогда и только тогда изоморфно взвешенному проективному пространству WP^n, когда группа Γ является группой, порожденной отражениями* (*или просто группой отражений*)

Чтобы не пришлось оправдываться в дальнейшем, договоримся, что тривиальная группа — это группа, порожденная пустым множеством отражений.

Таким образом, в случае $X = \mathbb{C}P^n$ простота (в нашем понимании) факторпространства обеспечивается конечными проективными группами комплексных отражений. Все такие группы в $\mathbb{C}P^n$ при $n > 1$ были найдены еще в начале XX века. Их классификации мы в первую очередь обязаны Митчеллу (Mitchell), Блихфельду (Blichfeldt) и Бернсайду (Burnside). Классификация конечных подгрупп в $\mathrm{PSL}(2, \mathbb{C})$ — результат классический.

Приведенная формулировка несколько отличается от авторской, в которой утверждается следующее: пусть в $\mathbb{C}^{n+1}$ действует конечная линейная группа Γ. Тогда факторпространство $\mathbb{C}^{n+1}/\Gamma$ тогда и только тогда изоморфно комплексному аффинному пространству, когда группа Γ есть группа отражений.

Объясним связь двух предложенных версий. Прежде всего заметим, что любое проективное отражение поднимается в линейную группу $\mathrm{GL}_{n+1}(\mathbb{C})$ отражением того же порядка (этот подъем однозначен, если $n > 1$).

Через $\widehat{\Gamma}$ обозначим подгруппу группы $\mathrm{GL}_{n+1}(\mathbb{C})$, порожденную «подъемами» всех отражений из Γ. Несложно проверить, что $\widehat{\Gamma}$ — это конечное (центральное) расширение группы Γ и, разумеется, линейная группа отражений. Итак, мы построили конечную линейную группу отражений $\widehat{\Gamma}$, рождающую проективную группу Γ.

Теперь стоит напомнить, что само проективное пространство $\mathbb{C}P^n$ есть факторпространство $\mathbb{C}^{n+1} - \{0\}$ по действию группы $\mathbb{C}^*$: $z \to tz$, $t \in \mathbb{C}^*$. Такое действие перестановочно с действием линейной группы, и, как следствие, мы получаем коммутативную диаграмму действий

$$\begin{array}{ccc} \mathbb{C}^{n+1} - \{0\} & \xrightarrow{\widehat{\Gamma}} & \mathbb{C}^{n+1}/\widehat{\Gamma} - \{0\} \\ \Big\downarrow{\scriptstyle \mathbb{C}^*} & & \Big\downarrow{\scriptstyle \mathbb{C}^*} \\ \mathbb{C}P^n & \xrightarrow{\Gamma} & Y = ? \end{array}$$

Согласно классической версии теоремы Шевалле, правая вертикальная стрелка связана с действием группы $\mathbb{C}^*$ на комплексном аффинном пространстве без точки. Давайте разберемся с этим действием.

Для этого было бы неплохо обзавестись удобными (для его изучения) аффинными координатами на факторпространстве $\mathbb{C}^{n+1}/\widehat{\Gamma}$. Это можно сделать, если перевести классическую теорему Шевалле на язык теории инвариантов. С точки зрения этой теории то обстоятельство, что факторпространство $\mathbb{C}^{n+1}/\widehat{\Gamma}$ есть аффинное комплексное пространство, означает ровно следующее: подалгебра $\widehat{\Gamma}$-инвариантных многочленов $\mathbb{C}[z_0, \ldots, z_n]^{\widehat{\Gamma}}$ является свободной полиномиальной алгеброй $\mathbb{C}[f_0, \ldots, f_n]$ от такого же числа переменных f_i, которые можно выбрать однородными со степенями $\deg f_i = k_i$. Тогда отображение

$$\mathbb{C}^{n+1} - \{0\} \to \mathbb{C}^{n+1}/\widehat{\Gamma} - \{0\}$$

задается формулой

$$(z_0, \ldots, z_n) \mapsto (f_0(z), \ldots, f_n(z)),$$

т. е. однородные многочлены f_i служат аффинными координатами на факторпространстве $\mathbb{C}^{n+1}/\widehat{\Gamma}$. Теперь уже можно явно указать искомое действие $\mathbb{C}^*$ на этом факторпространстве: $t(f_0(z), \ldots, f_n(z)) = (f_0(tz), \ldots$ $\ldots, f_n(tz)) = \big(t^{k_0} f_0(z), \ldots, t^{k_i} f_i(z), \ldots\big)$.

Окончательный вывод таков: на $\mathbb{C}^{n+1}$ без нуля группа $\mathbb{C}^*$ действует взвешенным образом, умножая i-ю координату на t^{k_i}, $t \in \mathbb{C}^*$. Такое взвешенное действие группы $\mathbb{C}^*$ на $\mathbb{C}^{n+1} - \{0\}$ как раз и дает в качестве факторпространства взвешенное проективное пространство PW с весами $k_0, \ldots, k_n$.

Итак: теорема 1 фактически есть проективизация классической теоремы Шевалле. В любой своей формулировке теорема указывает на решающую роль, которую играют группы отражений в поисках «простых» орбифолдов. Из чисто топологических соображений доказывается обратное утверждение: если комплексный орбифолд «прост» (например, является взвешенным проективным пространством), то группа орбифолда Γ есть группа комплексных отражений.

Наконец, можно уточнить и туманный вопрос о «простых» факторпространствах. Вот одна из его возможных точных постановок: назовем пару (X, Γ), состоящую из эрмитова симметрического пространства X и кокомпактной дискретной группы отражений $\Gamma \subset \operatorname{Aut} X$, парой Шевалле, если орбифолд X/Γ изоморфен взвешенному проективному пространству.

З а д а ч а. Описать все пары Шевалле с точностью до изоморфизма.

Если $X = P^n(\mathbb{C})$, то эта задача полностью решена.

Аффинные пары Шевалле: гипотеза и примеры

Комплексное аффинное эрмитово пространство я буду обозначать через $A_{\mathbb{C}}$. Речь пойдет о кристаллографических группах отражений Γ, т. е. о дискретных группах движений $A_{\mathbb{C}}$, порожденных отражениями, с компактным факторпространством $X = A_{\mathbb{C}}/\Gamma$, изоморфным взвешенному проективному пространству.

Но прежде давайте немного поговорим о кристаллографических группах Γ в $A_{\mathbb{C}}$. Пусть L — ассоциированное с $A_{\mathbb{C}}$ комплексное векторное пространство сдвигов. У каждого аффинного преобразования имеется дифференциал $d\gamma$ — это линейное (унитарное в нашем случае) преобразование пространства L. Рассмотрим гомоморфизм взятия дифференциала $d\colon \Gamma \to d\Gamma \subset \mathrm{Aut}\, L$. Если группа Γ является кристаллографической, то по теореме Бибербаха

а) группа линейных частей $d\Gamma$ конечна;

б) ядро d является решеткой сдвигов T полного ранга.

Таким образом, кристаллографическая группа Γ в $A_{\mathbb{C}}$ является почти абелевой: ее костяк составляет решетка параллельных переносов полного ранга. Кристаллографическая группа Γ называется неприводимой, если группа ее линейных частей $d\Gamma$ действует в L неприводимо.

Далее, мы хотим, чтобы группа Γ была группой отражений. Но тогда, как легко проверить, группа линейных частей $d\Gamma$ необходимо будет конечной линейной группой отражений.

Имеется гипотеза Бернштейна—Шварцмана (около 1976 г.), что пара ($A_{\mathbb{C}}$, неприводимая кристаллографическая группа отражений Γ) является парой Шевалле. Гипотеза эта по сей день не доказана, хотя для ее подтверждения осталось исследовать конечное число случаев.

Я расскажу об одном примере, где гипотеза подтверждается, и это можно объяснить, не используя сложной техники.

Начнем с одномерного комплексного аффинного пространства $A_{\mathbb{C}}$, в котором действует кристаллографическая группа $\Gamma = T_\tau \rtimes d\Gamma$ — полупрямое произведение решетки сдвигов T с базисом $\langle 1, \tau\rangle$ и группы линейных частей $d\Gamma = \langle -1\rangle$ (умножение координаты на -1). Группа Γ порождается отражениями: полуцелые точки решетки T_τ — суть зеркала ее комплексных отражений. Конструкцию факторпространства удобно разбить на два этапа: сначала одномерное комплексное аффинное пространство $\mathbb{C}$ факторизуется по действию решетки сдвигов T_τ, что приводит к одномерному комплексному тору $E_\tau = \mathbb{C}/T_\tau$ или эллиптической кривой. Затем остается профакторизовать эллиптическую кривую E_τ по действию инволюции $z \to -z$. В результате получается кривая рода 0 — од-

номерное проективное пространство $\mathbb{CP}^1$. В этом простейшем случае все хорошо.

Рассмотренный пример легко обобщается, превращаясь из одномерного в n-мерный. Выбрав начало координат, отождествим комплексное аффинное пространство $A^n_{\mathbb{C}}$ с $\mathbb{C}^n$. Тогда, вместо одного, у нас будет n комплексных координатных направлений: $z_1, \ldots, z_n$. Пусть в каждом комплексном направлении действует своя решетка сдвигов T_τ. В результате в $\mathbb{C}^n$ будет действовать решетка $T = \underbrace{T_\tau \oplus \ldots \oplus T_\tau}_{n}$ полного ранга.

Теперь нужно «скрестить» решетку T с подходящей группой линейных частей, причем последняя должна быть группой отражений. В качестве такой группы выберем группу Вейля $W(B_n)$. Ничего страшного я не сказал: речь идет о группе, порожденной всеми перестановками координат $z_1, \ldots, z_n$, а также изменениями знака у любой из них. Легко заметить, что группа $W(B_n)$ есть полупрямое произведение $\mathbb{Z}_2{}^n \rtimes S_n$.

Нужная нам кристаллографическая группа Γ есть полупрямое произведение решетки T и группы $W(B_n)$: $\Gamma = T \rtimes W(B_n)$. Как и в одномерном случае, переход к факторпространству $\mathbb{C}^n/\Gamma$ можно осуществить поэтапно, пользуясь «расщепимостью» группы Γ: сначала посмотрим на $\mathbb{C}^n \pmod T$. Затем, то, что получится, профакторизуем по действию группы перемен знака $\mathbb{Z}_2{}^n$ и, наконец, — по действию симметрической группы S_n.

На первом этапе мы получим произведение эллиптических кривых:

$$\mathbb{C}^n/T \simeq E_\tau \oplus \ldots \oplus E_\tau$$

(в самом деле, в каждом одномерном комплексном направлении $\mathbb{C}$ независимо действует своя решетка T_τ). Группа перемен знака независимо действует умножением на -1 на каждой из кривых E_τ. Поэтому на этом этапе в качестве факторпространства мы получим произведение n экземпляров комплексной проективной прямой $\mathbb{CP}^1$.

Наконец, как хорошо известно (кажется, это называется теоремой Виета),

$$\underbrace{\mathbb{CP}^1 \times \ldots \times \mathbb{CP}^1}_{n}/S_n = \mathbb{C}P^n$$

(группа S_n естественно действует перестановками множителей).

Следовательно, $A^n_{\mathbb{C}}/\Gamma = \mathbb{C}P^n$. Таким образом, мы смастерили кристаллографическую группу отражений в n-мерном аффинном комплексном пространстве и убедились, что факторпространство $A^n_{\mathbb{C}} \pmod \Gamma$ есть комплексное n-мерное проективное пространство.

Аффинные кокстеровские пары Шевалле: классификация

Гипотеза доказана для одного обширного класса кристаллографических групп, и вот его описание.

Пусть Γ — неприводимая кристаллографическая группа отражений в $A_{\mathbb{C}}$. Она называется кокстеровской, если группа ее линейных частей $d\Gamma$ является группой Вейля. Это некоторое ограничение на линейную группу отражений $d\Gamma$ в комплексном линейном пространстве L. Состоит оно в том, что в некотором базисе L группа $d\Gamma$ должна записываться вещественными матрицами. Вообще говоря, комплексных групп линейных отражений намного больше, чем групп Вейля.

Так вот, справедлива следующая

Теорема 2. *Пусть* Γ *— неприводимая кристаллографическая кокстеровская группа отражений в* $A_{\mathbb{C}}$. *Тогда* $(A_{\mathbb{C}}, \Gamma)$ *— пара Шевалле*.

И. Бернштейн и докладчик сначала классифицировали все аффинные кокстеровские группы отражений (с точностью до аффинной эквивалентности). Затем теорема была доказана для каждой серии таких групп с точным указанием весов факторпространства $A_{\mathbb{C}}/\Gamma$.

Скажу коротко, как выглядят кокстеровские группы отражений и какие, к примеру, взвешенные проективные факторпространства здесь встречаются.

Оказалось, что конструкция кристаллографической кокстеровской группы Γ связана с понятием аффинной системы корней. Напомню, что это такое. На этот раз дело происходит в вещественном аффинном евклидовом пространстве $A = A_{\mathbb{R}}$ с его вещественным пространством сдвигов $L = L_{\mathbb{R}}$. Через A^* обозначим линейное пространство линейных функций на A. У каждой линейной функции f на A есть дифференциал $\tilde{f} = df$, лежащий в пространстве L^*. Так вот, аффинная система корней S — это, прежде всего, подмножество в $A^* - \{\text{const}\}$. Любая линейная функция $\alpha(x)$ в A, отличная от константы, определяет аффинную гиперплоскость $\pi_\alpha = \{x \in A\colon \alpha(x) = 0\}$. Выберем вектор $h_\alpha \in L$ так, чтобы

а) вектор h_α был ортогонален зеркалу π_α;

б) $\tilde{\alpha}(h_\alpha) = 2$.

Условиями а) и б) вектор h_α определяется однозначно. И, конечно, с каждой непостоянной линейной функцией α связано отражение r_α, действующее в A^* по формуле

$$r_\alpha(\beta) = \beta - \tilde{\beta}(h_\alpha)\alpha$$

Подмножество $S \subset A^* - \{\text{const}\}$ называется аффинной системой корней, если для любых α и β из S выполнены условия: $r_\alpha(\beta) \in S$ и $\tilde{\beta}(h_\alpha) \in \mathbb{Z}$.

"all4" — 2009/5/27 — 11:23 — page 187 — #187

Как и в случае конечных систем корней, известна A-B-C-D-E-F-G-классификация неприводимых аффинных систем корней. Пусть S — аффинная система корней из этого списка. Выберем комплексное число τ, лежащее в верхней полуплоскости H_+. В комплексном аффинном пространстве $A_{\mathbb{C}} = A_{\mathbb{R}} + L_{\mathbb{C}}$ рассмотрим набор зеркал $\pi(\alpha, k) = \{\tau\alpha = k\}$, $\alpha \in S$, $k \in \mathbb{Z}$. С каждым таким зеркалом связано отражение $r(\alpha, k)$ порядка 2. Все такие отражения порождают кокстеровскую кристаллографическую группу $W(S, \tau)$ и, обратно, любая неприводимая кристаллографическая кокстеровская группа отражений изоморфна группе $W(S, \tau)$ для некоторой неприводимой аффинной системы корней S и некоторого $\tau \in H_+$.

Если, например, вы выберете систему корней A_n (или C_n, $n \geqslant 2$) и профакторизуете комплексное аффинное пространство по группе $W(A_n, \tau)$ (или $W(C_n, \tau)$), то получите взвешенное проективное пространство с весами $(1, \ldots, 1)$, т. е. обычное проективное пространство. Выбор системы корней B_n ($n \geqslant 3$) приведет в итоге к взвешенному проективному пространству с весами $(1, 1, 1, 2, \ldots, 2)$.

Что дальше

Таким образом, про пары Шевалле, связанные с аффинным эрмитовым комплексным пространством, известно немало и есть надежда полностью с ними разобраться. Но, как хорошо известно, комплексные отражения присутствуют и в группе движений комплексного гиперболического пространства $\mathbb{H}^n_{\mathbb{C}}$ или комплексного шара.

Здесь имеется лишь конечное число примеров групп Шевалле, которые были в разное время и из разных соображений построены Пикаром, Хирцебрухом, Делинем и Мостовым, ... Но нет никакой развитой теории или даже намека на классификацию.

15 мая 2003 г.

Д. Б. Фукс

УЗЛЫ В КОНТАКТНОЙ ГЕОМЕТРИИ

В Америке я должен был бы начать с каких-то бессмысленных фраз того типа, что я очень благодарен организаторам семинара и прочее, и прочее. Теперь я всё это пропущу. Но всё же я не могу не сказать, как я счастлив быть опять среди друзей.

Лежандровы кривые

Трудно сказать, как это получилось, что я оказался в ряду специалистов по контактной геометрии. Сказалось влияние окружения. И в результате я раз за разом оказываюсь автором нескольких работ и всякое прочее.

Мой сюжет очень прост: я буду рассказывать о лежандровых узлах в стандартном контактном пространстве. Я одним словом объясню, что это такое. Мои объяснения, конечно, скорее приспособлены для американской аудитории, потому что для них главная проблема — проблема вождения машины; здесь, может, это — проблема не первая, а, скажем, вторая. Но всё же давайте, представьте себе, что вы ведете машину по улице с односторонним движением и вам необходимо припарковаться. И естественно, весь ряд занят уже припаркованными машинами, но я вдруг замечаю пустое место, и хочу поставить сюда машину. Как это сделать? Ясно, что это можно сделать разными способами: можно подать немножко вперед и поставить машину; другая возможность — можно проехать немножко вперед, потом назад, потом поставить ее (рис. 1).

Р и с. 1. Парковка машины

Кривые, которые я рисую на доске, на самом деле уместно рассматривать не на плоскости, потому что совершенно ясно, что значение имеют не только координаты машины (центра или там чего), но и направление. Таким образом, фактически мое движение происходит в 3-мерном про-

странстве с тремя координатами; координаты — это координаты точки на плоскости, которые, кстати, я обозначаю нетрадиционным образом через x и z, и еще одна координата — y недостающая, которая есть наклон, угловой коэффициент касательной, т. е. тангенс угла. Разумеется, когда мы движемся, как я нарисовал, по этой кривой, то в пространстве x, y, z мы проделываем путь, между прочим, по кривой без всяких особенностей, без всяких каспидальных точек, по гладкой кривой, но подчиненной серьезному условию. Дело в том, что y — это угловой коэффициент, и поэтому, если эту кривую рассматривать как кривую в пространстве, то она подчинена уравнению $y\dot{x} - \dot{z} = 0$ — это как раз выражает, что y есть угловой коэффициент касательной. Еще раз хочу сказать: хотя кривая на плоскости имеет видимые особенности, соответствующая кривая в пространстве, если это каспы общего положения, особенностей не имеет — гладкая кривая. А то, что мы видим здесь — это ее проекция на плоскость x, z. Кстати, наличие у этой кривой особенностей видно уже из этого уравнения, потому что ясно, что если $\dot{x} = 0$, то и $\dot{z} = 0$. Здесь $x(t)$ — это функция, и естественно, когда ее производная обращается в нуль, то там и $\dot{z} = 0$; каждый раз, когда производная x обращается в нуль, мы получаем в проекции особенность, хотя ее и нет в пространстве. Кривые в пространстве, которые удовлетворяют такому уравнению, это лежандровы кривые.

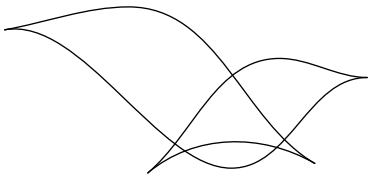
Можно этот простой пример замутить: сказать, что в действительности имеется дифференциальная форма α в пространстве x, y, z. Форма эта будет $y\,dx - dz$, и наша лежандрова кривая — это просто кривая, на которую ограничение этой формы равняется нулю, вот и вся премудрость. В действительности вся эта теория может быть заключена в более ученые слова: можно сказать, что α — это контактная форма. Она, в принципе, могла бы быть и другой. Единственное существенное условие — что эта форма нигде не интегрируема. Если бы форма была интегрируема, то это означало бы что $\alpha \wedge d\alpha = 0$ — это одна из форм записи условия Фробениуса. И если бы было так, то α равнялась бы некоторой функции умноженной на дифференциал другой функции: $\alpha = f\,dg$; функция g вдоль кривой была бы постоянна: вдоль кривой сужение α равняется 0, и таким образом эта кривая идет по поверхности уровня функции g. Наша же форма удовлетворяет противоположному условию, которое и есть условие, делающее ее контактной формой. Здесь α — контактная форма, т. е. $\alpha \wedge d\alpha$ нигде не обращается в нуль — форма объема. Только благодаря этому свойству мы, собственно, можем припарковать машину в любом месте, и вообще, можно машину, которая как-нибудь расположена, перевести в любое другое положение, избегая при этом вертикальных касательных. Потому что неинтегрируемость формы, собственно, и позволяет находить

интегральные кривые этого уравнения, которые соединяют две любые точки. Более того, можно любую гладкую кривую C^0-аппроксимировать кривой, удовлетворяющей этому уравнению, так что можно перевести одну точку в другую более или менее близко от любого пути.

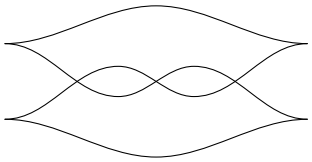
Итак, я сказал, что такое лежандрова кривая, и одновременно мы получили главный инструмент обращения с лежандровыми кривыми — проекцию на плоскость x, z, которая еще называется *фронтальная проекция*. Давайте переделаем определение, которое дано, в определение, использующее фронтальные проекции. Лежандрова кривая полностью описывается кривой с каспами на плоскости x, z, которая не имеет вертикальных касательных (вертикаль на плоскости — это ось z). Этот запрет я могу рассматривать просто как правила уличного движения, которые не подвергаются сомнению. Но если я хочу, чтобы y — тангенс угла — был числом (чтобы всё происходило в 3-мерном пространстве), то нужно, чтобы угол не равнялся 90 градусам. Мои кривые в 3-мерном пространстве; если я разрешу вертикальные касательные, будет не 3-мерное пространство. И проекция кривой, удовлетворяющей требуемому условию, никогда не имеет вертикальных касательных. Потому что угловой коэффициент касательной — это y, а y — это, естественно, число не равное бесконечности.

Лежандровы узлы

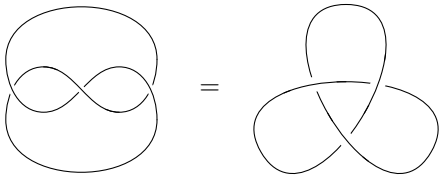
Итак, лежандровы кривые можно просто отождествить вот с чем: это кривая на плоскости, которой разрешается иметь каспы общего положения и не разрешается иметь вертикальных касательных; во-вторых, она гладкая вне каспов, в каспах тоже есть касательная. Вот что такое — лежандрова кривая. Определение лежандрова узла естественное: узел — это замкнутая кривая без самопересечений. Как же описать лежандров узел с помощью проекции на плоскость x, z? Очень просто: это — замкнутая кривая на плоскости x, z, и эта кривая может, конечно, иметь самопересечения (это самопересечения в плоскости x, z, игреки там будут разные), но не может иметь самокасания — вот это запрещено, не разрешается. Если кривая себя коснулась, то это уже будет являться точкой самопересечения в пространстве. Итак, проекция на плоскость x, z полностью определяет лежандрову кривую, определяет узел, если это — замкнутая кривая, и мы можем рисовать узлы, пользуясь этим кодом (рис. 2).



Р и с. 2. Лежандров узел



Р и с. 3. Трилистник



Р и с. 4. Трилистник в пространстве

Вот, скажем, такая картинка (рис. 3). Рекомендую: это — узел, который в топологии называется трилистник. Вот вспомните, пожалуйста, что обычно на диаграммах узлов показывают, какое перекрестье. А здесь не нужно этого делать, потому что координата y (отсутствующая здесь координата) — это угловой коэффициент наклона, поэтому в точке самопересечения я знаю, как расположены две эти ветви.

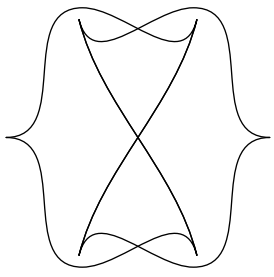
Здесь, правда, вступает в силу обстоятельство, которое является источником для бесконечного количества недоразумений в этой науке. Это обстоятельство заключается в том, что если это правда ось x, это ось z, и вы желаете ориентировать пространство при помощи правила правой руки, то ось y будет направлена не на вас, а от вас. Вы поверьте мне, это так. Туда направлена. Поэтому то, что ближе к вам, имеет меньший y, а не больший y. По этой причине вот такое перекрестье, как здесь нарисовано, здесь угловой коэффициент больше, чем здесь, поэтому y больше, поэтому эта ветвь дальше от вас, чем эта, и рисовать надо так, как на рис. 4. Если бы вы пожелали перерисовать узел на рис. 3, как обычно рисуется диаграмма узла, то это делается вот так; ну и дальше пожалуйста.

Вы можете распознать на моей картинке трилистник — обычный трилистник с тремя перекрестьями.

Наличие каспов мы считаем необходимым обстоятельством. Но когда мы смотрим на диаграмму, когда мы сличаем эту диаграмму с диаграммой узлов, то мы как бы игнорируем каспы — в конце концов каспы ничего не значат.

Ясно, что замкнутая кривая не может не иметь вертикальной касательной, если она гладкая. Но мы компенсируем отсутствие вертикальных касательных наличием каспов.

Вот еще одна картинка (рис. 5). Это тоже трилистник, но не совсем. Потому что трилистник — это узел, обладающий таким странным свойством, что он не изотопен своему зеркальному образу.

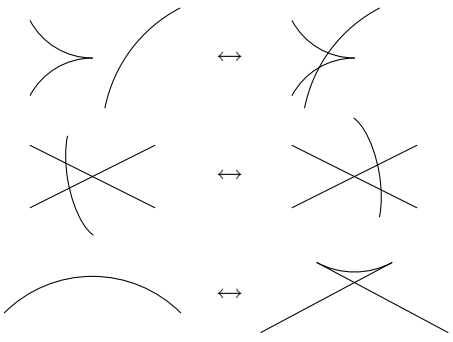


Р и с. 5. Зеркальный образ трилистника

Два эти узла — есть трилистник (настоящий трилистник, который есть в таблицах), а есть зеркальный трилистник. И вообще, большинство узлов, которые есть в таблицах — они не являются зеркальными: поэтому в таблицах обычно представлены одной диаграммой два фактически разных узла. Некоторые узлы зеркальны, т. е. изотопны своим зеркальным образам (например. «восьмерка»), но их относительно немного.

Теперь следующее обстоятельство, которое очень важно в теории узлов. Для узлов имеется понятие изотопии: один из узлов можно как-то продеформировать в другой, чтобы он не утратил своих свойств. Есть точное определение, которое я не буду давать. Лежандровы узлы тоже могут быть изотопны. Но у них есть две возможности: они могут быть изотопны как топологические узлы, а могут быть изотопны как лежандровы узлы, т. е. в процессе изотопии узел (меняющийся узел) всё время остается лежандровым. Опять-таки это не очень точное определение, но оно достаточно точно для моего рассказа. Например, узлы на рис. 4 и 5 не изотопны даже топологически, и уж никак не изотопны лежандрово.

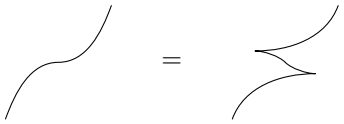
Сейчас я хочу кое-что сказать про лежандрову изотопию. Допустим, что мы вообще ничего не желаем знать, только рисовать такие картинки. Пусть есть две картинки; что означает, что они лежандрово изотопны? В топологии имеются способы перестройки диаграммы, которые сохраняют топологический тип узла (они называются движения Райдемайстера). Точно так же и в этой теории имеются свои движения, которые называются движениями Святковского. Святковский — это польский математик, который опубликовал то, что я сейчас собираюсь сказать, около 90-го года. Но надо сказать, что на семинаре Арнольда никого не впускали в аудиторию за десять лет до этой публикации, кто этого не знал. По-видимому, впервые я услышал об этом от Арнольда в одном из его установочных докладов лет за десять до появления работы Святковского. Но, в конце концов, Святковский в этом не виноват — это никогда не публиковалось.



Р и с. 6. Движения Святковского

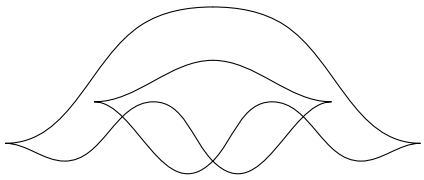
На рис. 6 изображены преобразования диаграммы, которые оставляют узел лежандрово изотопным тому, что было; и любые два узла лежандрово изотопны тогда и только тогда, когда их можно преобразовать друг в друга последовательностью таких преобразований. Первое

преобразование такое: если имеется касп и еще ветвь, то можно этот касп продеть сквозь эту ветвь (и обратно, конечно, тоже). Второе преобразование, которое имеется и среди движений Райдемайстера, таково: если имеется почти что тройное пересечение, то можно одну ветвь пропустить через перекрестье. Третье преобразование менее очевидное, но оно тоже существует. Это преобразование таково: если есть ветвь, в которой просто ничего нет, то к этой ветви можно приделать два каспа, но именно таким образом, как на рис. 6. Преобразование, изображённое на рис. 7, является категорически запрещенной операцией.



Р и с. 7. Запрещённая операция

Вообще-то бывают неприятности, когда есть два узла и мы пытаемся один в другой преобразовать такими шагами. Первое упражнение, которое я хочу предложить, вот какое. Возьмём узел, который изображён на рис. 3, и перерисуем его так, как на рис. 8. Вот упражнение: доказать, что два эти узла лежандрово изотопны — один может быть получен из другого при помощи преобразований Святковского, причём нужны все три преобразования.



Р и с. 8. Другая диаграмма трилистника

Числа Беннекина и Маслова

Как и в стандартной теории узлов, имеются инварианты, которые позволяют различать теперь уже лежандровы узлы. Есть огромный спектр топологических инвариантов, и если мы желаем знать, будут ли два узла лежандрово изотопны, мы, конечно, прежде должны посмотреть, будут ли они топологически изотопны, и для этого имеется обширная теория, которой я касаться почти не буду.

Но еще есть два классических инварианта — два числа, два целых числа, которые ассоциируются с лежандровым узлом и которые тоже являются лежандровыми инвариантами. Позвольте мне привести их определение.

Первое из этих чисел называется число Беннекина или более подробно — число Тёрстона—Беннекина. Беннекин — это математик, который, видимо, не придумал этот инвариант (этот инвариант уже был известен); его выдающаяся работа знаменита неравенством, которое приводится ниже. Проще всего описать, что такое число Беннекина, пользуясь пространственной картинкой. Пусть у нас есть лежандров узел (замкнутый). Нам задана контактная структура — форма α, значит, в каждой точке имеется плоскость, имеется 2-мерное направление — нули этой формы; и наша кривая касается всюду этих плоскостей. А на самом деле все эти плоскости можно некоторым совместным образом ориентировать. Узел тоже можно ориентировать (хотя от ориентации узла это число не будет зависеть). Затем в каждой плоскости берем положительную нормаль к нашей кривой и сдвигаем наш узел Γ в направлении этой нормали. Он, кстати, в результате обязательно перестает быть лежандровым. Но получатся две замкнутые ориентированные кривые Γ и Γ^+, и мы имеем число, которое называется индекс зацепления: сколько раз одна кривая пересечет другую, если мы будем их растаскивать по разным углам комнаты; при этом мы считаем их соударения со знаком. Такое число зацепления есть у любых двух ориентированных замкнутых кривых в 3-мерном пространстве. Число зацепления $\mathrm{lk}(\Gamma, \Gamma^+)$ называется числом Тёрстона—Беннекина узла Γ. Оно много от чего не зависит: не зависит от ориентации, не зависит от того, в какую сторону сдвигать (если сдвинем в другую сторону, то будет то же самое).

Прежде чем сказать, как это число соотносится с диаграммой, я хочу определить другое число, которое называется число вращения или число Маслова. Правильное число Маслова, собственно говоря, будет вдвое больше (тут терминология не вполне правильная). Это число определяется вот как. Можно во всех плоскостях контактной структуры одновременно выбрать вектор. Например, можно сказать так: все плоскости контактной структуры (т. е. плоскости, которые касаются лежандрова узла) параллельны оси y — так получается; и поэтому можно взять направление положительной оси y в каждой плоскости. Есть лежандров узел, плоскость V — это плоскость контактной структуры, которая определяется уравнением $\alpha = 0$. И во всех плоскостях одновременно можно выбрать вектор, во всем пространстве. После этого вы берете какую-нибудь точку, измеряете угол поворота по часовой стрелке от выбранного направления до направления узла (здесь мы узел ориентируем — это существенно в этом определении). Получится число, которое определено с точностью до 2π умножить на целое число. Если вы обходите весь узел и возвращаетесь обратно, и выбираете непрерывную ветвь этого угла, то, конечно, результат

после возвращения будет отличаться от того, с чего вы начинали, на кратное 2π. И это число, деленное на 2π, называется числом вращения или числом Маслова $\mu(\gamma)$. Таким образом, есть угол $\varphi(t)$; если мы обойдем вокруг, то $\varphi(1)$ будет равняться $\varphi(0)$ плюс $2\pi\mu(\gamma)$.

Эти два целых числа между собой более или менее независимы. Единственное, что можно сказать, это то, что, если это действительно узел (связный узел), то их сумма всегда нечетна — так получается, и это легко доказать. В остальном это два независимых числа. И они оба являются лежандровыми инвариантами: если два узла лежандрово изотопны, то эти числа одинаковые. Такое описание этих чисел, конечно, хорошо дать, но на самом деле оно не нужно, потому что по диаграмме ничего не стоит оба эти числа найти. Сейчас я скажу, как это делается.

Давайте возьмем какую-нибудь диаграмму. У нее есть каспы и перекрестья. Давайте выберем ориентацию узла (результат не будет зависеть от нее) и сделаем следующее. Каждому левому каспу мы припишем число -1. Правые каспы мы игнорируем; их количество равно числу левых каспов, как легко усмотреть. Что касается перекрестий, то мы различаем два их типа. Дело в том, что касательная не вертикальна, поэтому каждая ветвь направлена направо или налево. Если они направлены в одну сторону, я пишу $+1$; а если они направлены в разные стороны, я пишу -1. После чего суммирую всё по всей диаграмме.

Давайте проделаем вычисления для двух диаграмм на рис. 3 и 5. Начнем с рис. 3. Выберем какую-нибудь ориентацию. Как видите, здесь есть 2 левых каспа $(-1, -1)$ и 3 перекрестья, которым всем соответствует значение $+1$ (направленных в одну сторону). Сумма этих чисел равна $+1$ — это число Беннекина этого узла, которое обозначается $\tau\beta$.

Кстати сказать, ясно, что если ориентацию изменить на противоположную, то это число не изменится, потому что в каждом месте обе стрелки будут направлены в другую сторону.

Возьмем теперь вторую диаграмму (рис. 5). Для нее число Беннекина равно -6.

Число μ тоже можно найти по диаграмме, причем нужно совершенно не смотреть на перекрестья, а смотреть только на каспы. И делать следующее: выбрать ориентацию — от нее будет зависеть число (если вы поменяете ориентацию, число изменит знак). Каждый касп будет находиться либо в направлении вверх, либо в направлении вниз; мне всё равно — левый он или правый, я смотрю только на то, направлен он вверх или вниз. И я беру d — число каспов, направленных вниз, и вычитаю из него u — число каспов, направленных вверх. Полученное число будет четным, потому что сумма их четна, поскольку это число

всех каспов — и очень естественно это четное число поделить на 2: $\mu = \frac{1}{2}(d - u)$.

Посмотрим опять на эти картинки. Для верхней диаграммы $\mu = 0$, а для нижней $\mu = \pm 1$, в зависимости от ориентации.

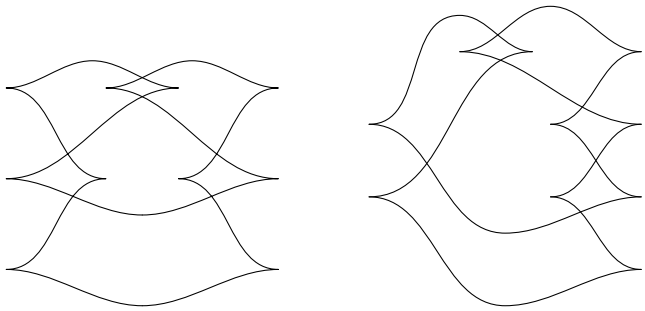
Я хочу сказать, что такой способ вычисления по диаграмме чисел дает то, что нужно. Это очень легко доказать. Тоже в какой-то момент кто-то это придумал. Первая публикация известная мне, в которой содержатся оба правила, которые я сформулировал, — это работа примерно 86 года Сережи Табачникова, которая на самом деле посвящена другому предмету, но там это написано; может быть, это было известно раньше, я этого просто не помню.

Итак, имеются два инварианта. Я еще хочу сказать вот что. На рис. 7 был нарисован запрещенный шаг. Теперь ясно, почему он запрещен. Если мы действительно сделаем это, то у нас появится лишний левый касп без какого бы то ни было другого изменения (правый касп, конечно, тоже появится, но правые каспы я не считаю). Число Беннекина упало на единицу, стало другим (кстати, μ тоже). Таким образом, этот шаг меняет число Беннекина и уж никак не может сохранять лежандров тип узла.

О работе Чеканова и Элиашберга

Естественно возникает вопрос: если есть два лежандровых узла, которые топологически изотопны и имеют эти числа одинаковые, то можно ли в этом случае ожидать, что они будут лежандрово изотопны? Эта проблема довольно долго была открытой, и была решена в отрицательную сторону, скажем, в 97 г., когда это было сделано достоянием широкой публики, в двух работах двух разных авторов — Юры Чеканова и Яши Элиашберга, работавших независимо. К сожалению, по сей день ни одна из этих работ по-настоящему не опубликована, хотя есть препринт Юры Чеканова, который, как теперь говорят, имеется на сети. Это великолепно написанная работа, опубликованная в 99 году. Краткое изложение работы Элиашберга имеется в трудах Берлинского конгресса. Между прочим, там приводятся некоторые интересные обобщения основной конструкции. В общем, его работа осталась как бы устной. Я сам слышал ее изложение, когда я еще не знал, что Чеканов этим занимается.

Эти два автора придумали новый инвариант — мы его будем обсуждать. И каждый из них привел пример двух узлов, которые не различаются привычными инвариантами, а различаются их инвариантом. Далеко не сразу стало ясно, что в этих двух работах пример тот же самый, потому что он нарисован был совсем по-разному. И я нарисую его здесь еще третьим



Р и с. 9. Два узла

способом (рис. 9), который по-дружески сообщил мне Яша Элиашберг. Топологически они довольно простые. Как вы знаете, есть таблица узлов, все узлы нумерованы, и это узел, который в таблице называется 5_2 — четвертый по сложности узел.

Конечно, эти узлы топологически одинаковы. Мгновенный подсчет показывает, что число Беннекина для обоих узлов равно $+1$, а число μ равно 0. Вы можете попытаться преобразовать один узел в другой, и если вам это удастся, то всё, о чем я собираюсь говорить, утратит всякий смысл. Но согласно этой теории эти два узла не могут быть лежандрово изотопны. Я постараюсь здесь объяснить, как вычислить эти инварианты по диаграмме, но это чуть позже. Прежде чем обратиться к этим инвариантам, я хочу сказать о другой проблеме, которая тоже будет здесь существенна.

О множестве значений чисел Беннекина на классе изотопных узлов

Есть другая проблема: пусть имеется топологический тип узлов (например, трилистник — что-нибудь такое). Спрашивается, какие значения могут принимать в пределах этого класса числа Тёрстона—Беннекина и Маслова? Оказывается, что число Беннекина (именно число Беннекина) всегда ограничено сверху — в пределах топологического типа узлов, в стандартном контактном пространстве. И, собственно, Беннекин это и доказал. Его работа (насколько я понимаю, вообще, более-менее единственная его работа; после этого математику он как-то оставил) содержит следующий результат. Как вы знаете, для каждого узла есть поверхность Зайферта — это вложенная в пространство ориентируемая поверхность, границей которой является этот узел (обычно рассматриваются связные поверхности; во всяком случае, необходимо предположить, что

поверхность не имеет сферических компонент). Допустим, что имеется такая поверхность Зайферта S; она имеет эйлерову характеристику χ (целое число). Можно взять ее со знаком минус, потому что она отрицательна. И число Тёрстона—Беннекина будет меньше или равно $-\chi$. Например, для трилистника эта поверхность будет тор с дыркой, ее легко нарисовать, и эйлерова характеристика будет равна -1. Тогда это число равно $+1$, поэтому отсюда вытекает утверждение, что для трилистника (правда, для зеркального трилистника тоже — при таком подходе это совершенно не различается) число Беннекина не может превосходить $+1$. Для Беннекина существенно было то, что если лежандров узел топологически тривиален (вообще узлом не является) то тогда, естественно, он ограничивает диск, и поверхность Зайферта имеет эйлерову характеристику $+1$. Из этого вытекает, что лежандров узел, топологически тривиальный, всегда имеет отрицательное число Беннекина. Пользуясь этим результатом, Беннекин смог доказать (собственно, пример был известен), что в 3-мерном пространстве имеются экзотические контактные структуры. Имеется контактная структура, которую легко описать, которая не может быть диффеоморфна (контактноморфна) стандартной контактной структуре, потому что там это свойство не выполняется. Там можно построить лежандров узел (этот узел на самом деле — плоская окружность), который имеет число Беннекина 0. Так получается, что хотя контактные структуры локально все одинаковы (есть такая теорема), но глобально они могут различаться даже в 3-мерном пространстве. Существенно, однако, что хотя эта оценка очень хорошо работает для трилистника, для зеркального трилистника она катастрофически не работает. И, таким образом, она довольно грубая. Есть другие оценки. Мне будет существенна в моем рассказе одна из этих оценок.

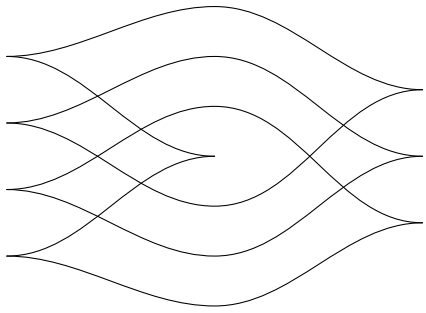
Вы, наверное, наслышаны о том, что есть такое явление, которое называется многочлены узлов. Их довольно много. Я совершенно не хочу углубляться в определения, но поверьте, что с каждым узлом связано много многочленов, которые являются их топологическими инвариантами. И один из них, который мне будет важен, это так называемый многочлен Кауфмана — многочлен от 2 переменных. Я говорю «многочлен», хотя никакой он не многочлен, потому что там обе переменных могут быть в отрицательных степенях. Эти переменные традиционно обозначаются буквами z и α. И этот многочлен имеет максимальную и минимальную степень по α — два числа. Я еще раз повторю, что степени могут быть отрицательные. И минимальная степень по α полинома Кауфмана узла обязательно будет строго больше, чем число Тёрстона—Беннекина узла. Странно, что эта теорема не была известна давно из-за незнакомства

контактных геометров с теорией узлов. Эта теорема была, видимо, впервые опубликована (в несколько более слабой форме) в моей работе с Сережей Табачниковым в 96 г.; и в такой форме это потом было доказано Чмутовым и Горюновым. И еще есть очень короткое изящное доказательство Табачникова. Вот эти три работы более или менее содержат доказательство этого факта. Факт замечательный. Для обычного трилистника мы опять получаем, что число Беннекина не превосходит $+1$, а для зеркального трилистника получаем оценку -6, именно -6, т. е. эта оценка Кауфмана оказывается точной.

Я рассмотрел узел, который имеет топологический тип 5_2. Есть, конечно, другие узлы. Про некоторые узлы установлено, что их лежандров тип полностью определяется их топологией и этими двумя числами. Насколько мне известно, это доказано для всех торических узлов, как «алгебраических», так и «зеркальных», включая и трилистник. Еще раньше это было доказано для топологически не заузленных лежандровых узлов. И еще это отдельно доказано для узла, который называется восьмерка. Этот перечень делает узел 5_2 первым узлом, для которого это не доказано, и для него это оказывается неверным. Мне еще будут нужны торические узлы в моем рассказе. Для них известно всё, известна полная лежандрова классификация. А именно, известно, какие значения принимают числа, и известно, что они полностью определяют лежандров тип узла.

Насколько точна вышеприведенная оценка? Оказывается, она не точна. И имеются узлы, для которых доказано, что она не точна. Первый пример такой. Торический узел описывается двумя взаимно простыми числами p и q (пусть для определенности $p > q$). Если q четно, то приведенная оценка для зеркального торического узла p, q точна; если q нечетно, то она не точна.

Таким образом, первый узел, для которого оценка Кауфмана не дает правильный результат — это торический узел 4, 3 (зеркальный). Его диаграмма устроена так, как показано на рис. 10. Максимальное число Беннекина этого узла равняется -12 (это, собственно, 4 умножить на 3). А Кауфман дает оценку -11 — это расхождение на единицу, первое расхождение оценки Кауфмана и фактического числа Беннекина.



Р и с. 10. Диаграмма зеркального торического узла

Алгебра Чеканова—Элиашберга

Теперь я хочу сказать несколько слов о том, в чем же заключается инвариант Чеканова—Элиашберга. Это довольно сложная алгебраическая конструкция. То, что я сейчас скажу, не содержится ни в работе Чеканова, ни в работе Элиашберга. Это модификация их конструкции, принадлежащая молодому человеку Lenny Ng с легко произносимым именем Ленни и трудно произносимой фамилией Нг, которая в Америке обычно читается как Инь.

Эта конструкция, в том виде, в каком она изложена в работе Ленни, такова. Берем диаграмму узла, и на ней рассматриваем следующие точки: перекрестья и еще правые каспы (их взять удобнее, чем левые). Их можно обозначить буквами: a, b, c. Рассматривается свободная ассоциативная алгебра с единицей, порожденная этими образующими. По-другому я могу сказать так: некоммутативные полиномы от этих букв с коэффициентами в поле из 2 элементов (поле, которое топологи обозначают $\mathbb{Z}_2$). Это настоящие полиномы, с положительными степенями. Получается алгебра, и в этой алгебре определяется дифференциал. Это самая сложная вещь: с помощью некоей конструкции (я совершенно не буду ее касаться) определяется дифференциал, который удовлетворяет правилу Лейбница. Фактически достаточно определить его на образующих (перекрестьях и каспах). Эта конструкция сложная. Она настолько сложная, что когда вы, пользуясь ей, вычисляете, то вы обязательно получите с первого раза неправильный ответ, а со второго — уже неизвестно. И Чеканов, и Элиашберг для своего примера нечто вычисляли; оба вычисления были неверны. Потом оба они исправили свои вычисления.

Существенно, что есть способ приписать всем этим вот объектам степень — некое целое число, которое называется степенью. Построить его проще всего для каспов. Степень каспа (правого) по определению равна $+1$. Степень же перекрестья определяется вот как. Как вы помните, у нас есть способ вычисления числа Маслова через эти самые верхние и нижние каспы. Можно вычислить эту степень как бы не полностью: т. е. мы начинаем с перекрестья, идем в произвольном направлении (что влево, что вправо — это всё равно), и идем до тех пор, пока не вернемся снова в исходную точку. И по дороге вычисляем те же самые проходы вверх и вниз. То есть берем ту же самую формулу $d - u$. Но это число уже не обязательно четно, и мы теперь не делим его на 2. Это число и называется степенью перекрестья. Существенно вот что: можно было бы начать обход в другую сторону. При этом может получиться другое число. Но разность этих чисел равняется удвоенному числу Маслова. Поэтому, если число

Маслова не равняется 0, то эта степень определена как вычет по модулю удвоенного числа Маслова. Если же оно равняется 0, как в случае этих узлов, то это будет целое число.

Я описал степень образующей. Когда перемножаются образующие, то степени складываются. И вот что существенно: дифференциал, который я не описал, имеет степень -1. Так что получается, как говорят, дифференциальная градуированная алгебра с дифференциалом степени -1. И отсюда возникает, собственно, инвариант. Например, можно вычислить гомологии; гомологии тоже будут кольцо, и оно является инвариантом узла. Имеется более точная теорема Чеканова: что же, собственно, является инвариантом. Я сейчас тоже не буду про это подробно говорить. Сама по себе эта алгебра инвариантом, конечно, не является: просто даже число образующих не является инвариантом, потому что можно добавить образующие. Как я сказал, при пользовании этими инвариантами главная проблема заключается в том, что довольно трудно вычислить этот дифференциал, потому что вам нужно на диаграмме кое-что находить, что прямо, в общем-то, может быть и не видно.

Аугментация алгебры Чеканова—Элиашберга

В работе Юры Чеканова, однако, имеется некоторое усовершенствование этой конструкции. Оно заключается в следующем. Это алгебра с единицей, и алгебраисты знают, что можно говорить о вещи, которая называется аугментация. Давайте через A обозначим алгебру Чеканова—Элиашберга; это алгебра с единицей. Аугментация — это отображение ε этой алгебры в поле (пусть будет $\mathbb{Z}_2$), которое мультипликативно и, конечно, переводит единицу в единицу. Но это дифференциальная алгебра, и я налагаю дополнительное условие, что композиция ε и d равняется 0. Совершенно ясно, что, поскольку ε мультипликативно, это отображение ε полностью определяется своими значениями на образующих. Так что фактически просто мы каждому перекрестью и каждому каспу приписываем значение 0 или 1 — вот, собственно, всё, что происходит. Но это условие довольно серьезное.

Если такая аугментация существует, то мы производим то, что называется в алгебре линеаризацией, более или менее стандартным образом, т. е. мы берем ядро аугментации (это идеал в A), и затем факторизуем этот идеал по его квадрату. Так сказать, если считать аугментацию точкой, то я беру касательное пространство в этой точке. И дифференциал, который имеется в A, переносится на линеаризацию (естественно, если он правильно определен). Более того, у A все эти a, b и c, и прочее —

мультипликативные образующие, свободные; у этого фактора они же являются базисом над полем $\mathbb{Z}_2$. То есть эта линеаризация довольно-таки маленькая. В дополнение ко всему этому потребуем следующее. Если a — это образующая, и $\varepsilon(a) \neq 0$, т. е. равняется 1, то давайте потребуем, чтобы это означало, что степень a равняется 0. Пусть аугментация в этом отношении будет чистая, т. е. она может принимать ненулевые значения только на перекрестьях степени 0. Тогда градуировка, которую мы ввели в A, (степень) сохраняется здесь, т. е. будет также и дифференциал, имеющий степень -1. Единственная разница — что комплекс этот довольно маленький, т. е. его аддитивными образующими являются перекрестья и каспы. И опять-таки гомологии этого комплекса являются инвариантом.

Здесь можно сказать более точно. Дело в том, что разных аугментаций может быть много. И эти гомологии могут быть разные. Полностью нужно сказать так: если два узла лежандрово изотопны, и у одного из них есть аугментация, то у другого тоже есть аугментация, так что соответствующие линеаризованные комплексы имеют одинаковые гомологии. Мне не очень важна такая точность, и я не знаю примеров, когда бы действительно для разных аугментаций гомологии были бы разные.

Необходимое и достаточное условие существования аугментации

Ясно, что имея такое упрощение, нам, вероятно, можно будет обращаться с диаграммами гораздо проще. И вот, собственно, первая загадка. Имеется легко проверяемое по диаграмме необходимое и достаточное условие существования аугментации. Его история такова. Достаточность условия, которое я сейчас опишу, доказана в моей работе, которая до сих пор пребывает в печати в журнале «Геометрия и физика». А необходимость установлена мной совместно с моим учеником Тиграном Ишхановым, и наша работа готовится к печати.

Это необходимое и достаточное условие для существования аугментации геометрическое, оно формулируется для такого рода диаграмм. Причем его можно сформулировать как с этим условием, связанным с градуировкой, так и без него. Условие заключается в следующем. Представьте себе, что имеется некоторая фронтальная диаграмма узла. У диаграммы имеется некоторое количество каспов, левых и правых, одинаковое количество, к слову сказать (мы ведь идем поочередно: налево, направо, налево, направо, налево, направо — получается одинаковое число левых и правых каспов). Первое, что нужно сделать — это поставить каспы друг другу в соответствие, чтобы каждый левый касп знал свой правый. Скоро

будут другие условия, но пока ставим в соответствие как угодно. После этого нужно внутри диаграммы выбрать два непересекающихся (не считая концов) пути, соединяющих два эти каспа. Каждый из этих путей каспов не содержит, кроме концов. То есть путь как вышел, так и идет слева направо, нигде не поворачивая назад. Отсюда, кстати сказать, вытекает, что эти пути все вместе однократно покрывают всю диаграмму. Вся диаграмма оказывается разложенной таким образом.

К этому есть дополнительные условия, которые оказываются очень важными. Я не выдумывал этих условий. У меня была некая конструкция аугментации, я просто смотрел, при каких условиях она проходит, и всё это так и получилось. Мне стыдно сказать, но теорема о необходимых и достаточных условиях, которую я сейчас формулирую, имеет довольно замысловатое доказательство. Я, как и многие из вас, не люблю красивые доказательства, потому что красота сама по себе означает, что вы доказательство не понимаете. Я не понимаю это доказательство. Мое доказательство довольно замысловатое, я не умею доказывать просто.

А сейчас я сформулирую теорему. Имеется еще некое условие вот на что. Я сказал, что можно перейти в точке перекрестья с одной линии на другую. И теперь я хочу сказать, что на самом деле это можно сделать не всегда. А именно, есть следующее условие. Пусть один из путей, соединяющих два каспа, действительно совершает переход с одной ветви на другую. Это означает, что оставшаяся половина тоже составляет путь, соединяющий какие-то два каспа; причем не те же самые, потому что пути, соединяющие те же самые каспы, не могут пересекаться. Так что один путь соединяет два каспа, и другой путь соединяет два каспа. И для каждой пары есть другой путь, который их соединяет. Я не смотрю на эти пути полностью, но я хочу посмотреть на них в узкой вертикальной щели (рис. 11). Требование заключается в том, что заштрихованные фигуры либо не пересекаются (я не считаю угол пересечением), либо одна содержит другую. На самом деле, если в это вникнуть более подробно, есть 6 возможных расположений двух этих путей. Из этих 6 возможностей 3 разрешены, 3 запрещены.

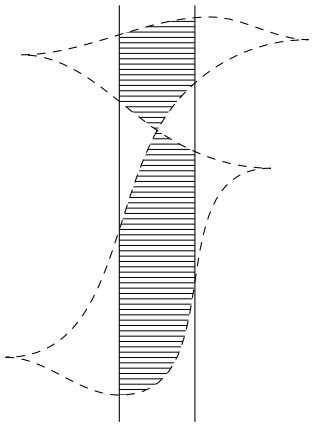


Рис. 11. Щель

Теорема 1. а) *Аугментация* (*я пока игнорирую условие на степень*) *существует в том и только в том случае, если всю диаграмму можно разбить с выполнением этого условия.*

б) *Если я непременно хочу, чтобы моя аугментация была полностью согласована с градуировкой, со степенями таким вот образом, то это означает, что я должен иметь всё то же самое, но делать такие перескоки только лишь на перекрестьях степени* 0. *Если дополнительно запретить такие перескоки на перекрестьях какой-либо степени, кроме* 0, *то эта аугментация будет согласована с степенями.*

В диаграммах на рис. 9 все перекрестья, кроме двух верхних, имеют степень 0; а эти перекрестья — здесь имеют степень 0, а здесь — 2 и -2. Вообще, это вот как: тут у меня 3 перекрестья и одно; если вы это обозначите буквами p и q, то здесь будет стоять $p-q$, $q-p$. Ну и еще вы помните, что степень 1 имеют правые каспы, это довольно много. Абсолютно очевидно, что для этих диаграмм структура, описанная здесь, существует. Имеется 2 узла, имеется 2 комплекса. Какие степени этих образующих? Здесь степень -2, потом масса нулей, потом масса единиц и одна двойка. А в другом комплексе только нули и единицы. Каков бы ни был дифференциал, у этих комплексов не могут быть одинаковые гомологии. Из-за этой минус двойки: она никуда не денется. Эта минус двойка приведет к нетривиальным гомологиям. А в другом комплексе гомологий нет. Поэтому сразу автоматически получаем: гомологии разные и узлы не совпадают.

Если известны размерности пространств цепей, но не известны дифференциалы, и мы хотим, чтобы гомологии были одинаковы, то на этот счет имеются неравенства Морса, которые всегда могут быть применены. Для двух комплексов, имеющих одинаковые гомологии, выполняются некие неравенства между размерностями. А если эти неравенства нарушаются, то гомологии не могут быть одинаковые. Здесь можно обойтись без неравенств Морса, всё обеспечивает изолированная минус двойка.

Обсуждение

Дальше возникают еще некоторые обстоятельства, которые делают весь этот предмет весьма непонятным. Когда я занимался этой работой, я считал, что я нахожусь как бы в вакууме. Но потом оказалось, что вакуума никакого не существует, и что это условие, которое я формулировал, на самом деле не новое. Удивительным образом, новой является моя теорема, а само условие уже появлялось в других работах. Раньше всего (правда, без условия, сформулированного перед теоремой 1) оно появилось в последней работе Яши Элиашберга, опубликованной по-русски в журнале «Функциональный анализ». Там было сформулировано,

что вот такая штука, правда, без этого условия, которое ему не нужно, существует для любого лежандрова узла, лежандрово изотопного простейшему лежандровому узлу (с двумя каспами и без перекрестий).

Более или менее одновременно со мной не кто иной, как Юра Чеканов, работы которого, собственно, я пытался изучать, вывел то же самое условие, занимаясь абсолютно другой задачей. Точнее говоря, это была совместная работа Юры Чеканова и Пети Пушкаря. Я не считаю, что здесь нужно рассказывать работы Чеканова и Пушкаря, вам они доступны больше, чем мне — и тот, и другой. Работа Чеканова и Пушкаря — это совершенно замечательная работа, в которой доказана так называемая гипотеза Арнольда. Нужно сказать, что у Арнольда есть много гипотез; и когда говорят «гипотеза Арнольда», обычно имеется в виду некое утверждение о неподвижных точках симплектоморфизмов. Это другая гипотеза; и её формулировка очень простая. Она заключается в следующем. Рассмотрим окружность, скажем, с семейством нормалей, направленных внутрь, и начнем ее деформировать вместе с этими нормалями, причем разрешаются более или менее те же перестройки, что при перестройке лежандровых узлов, с маленькой вариацией. В результате этих перестроек нужно вывернуть окружность наизнанку (чтобы нормали были направлены наружу). Гипотеза Арнольда утверждает, что при этом обязательно по дороге будет кривая, имеющая по крайней мере 4 каспа. Теорема, похожая по формулировке на теорему о 4 вершинах, но это другая теорема. Гипотеза была очень известна, она стояла несколько лет (года четыре по крайней мере). И Чеканов и Пушкарь доказали это, пользуясь буквально той же самой конструкцией.

Эксперимент показывает, что если диаграмма лежандрова узла действительно обладает разложением с тем условием нормальности, которое я сформулировал, то в пределах топологического класса этот лежандров узел имеет максимальное возможное число Беннекина. Во всех случаях это так. Этот факт имеет такое полуподтверждение. Как может быть, что число Беннекина не является максимальным? Возьмите какую-нибудь диаграмму и в каком-нибудь месте сделайте такой зигзаг, как на рис. 7. Число Беннекина сразу упадет на единицу. Это, конечно, никак не повлияет на топологию узла. В пределах топологического класса число Беннекина будет не максимально. Но уж такая диаграмма никак не может иметь моей структуры. В самом деле, где же найти два пути, исходящие из этого каспа, идущие в другой касп? Из этого каспа мы можем перейти только в этот касп; но нет другого пути.

Надо сказать, что многие вещи в теории лежандровых узлов не известны. В частности, не известно вот что: правда ли, что всякий узел не

с максимальным числом Беннекина получается из узла с максимальным числом Беннекина прибавлением некоторых таких зигзагов.

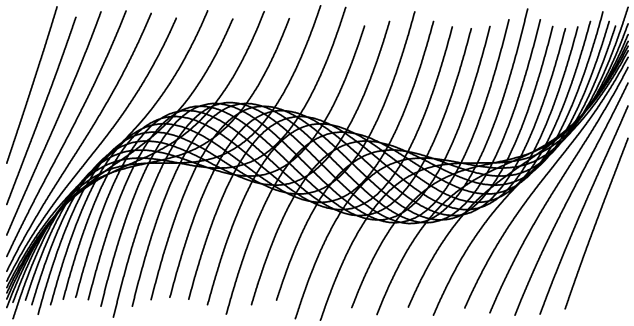
Возникает гипотеза, которая подтверждается другими зеркальными торическими узлами, что не просто число Беннекина должно быть максимально, а оно должно равняться степени полинома Кауфмана минус единица, т. е. оно не просто максимально, а оценка должна показывать его максимальность. Но я должен сказать, что материал, подтверждающий эту гипотезу, довольно скромный.

Дело в том, что имеется весьма обширная таблица узлов. Есть таблицы в книгах, но это далеко не всё. Опять же на так называемой сети имеется некая программа, которая умеет рисовать диаграммы всех узлов, по крайней мере, с 16 перекрестьями. Ровно на этом же уровне находится вычисление полиномов. Я слышал, что сейчас какие-то люди работают над диаграммами с 17 перекрестьями; все вместе пожелаем им успеха.

Но наличие или отсутствие такой структуры, конечно, ни в каких таблицах найти нельзя. Моя дочь написала на эту тему курсовую работу в Беркли. Из ее работы вытекает, что для диаграмм с 9 перекрестьями такую структуру не удается построить для 2 диаграмм. Одна из них — та, которую я рисовал; это 8 перекрестий. Другая диаграмма относится к узлу 9_{49}. Узел 9_{49} — это романтическая вещь. Его диаграмма не симметрична. Тот факт, что узел не является зеркальным, как правило, демонстрируют его полиномы, потому что при зеркальном отражении многочлен узла преобразуется очень простым образом. И поэтому можно понять по диаграмме, будет ли узел зеркальным. Вот для этого узла классические полиномы, т. е. Кауфмана, HOMFLY, как раз не показывают, что он не симметричный. А в действительности он не симметричный, и это вытекает из других полиномов. И вот из этих двух диаграмм для одной не удается построить ни того, ни другого: ни такой структуры, ни диаграммы, у которой бы число Тёрстона—Беннекина равнялось, чему положено равняться по оценке с полиномом Кауфмана. То есть это тоже как-то является косвенным подтверждением, что одна и та же трудность мешает тому и другому.

Так что есть некая связь всего этого с проблемой значений числа Беннекина. Есть однако еще абсолютно другая вещь. Мне не хочется говорить о ней подробно, потому что это работа Пети Пушкаря. Она, можно сказать, не написана, а то, что написано — это далеко не полная работа. Там совершенно замечательные результаты, но в конце концов, я не могу за них ручаться.

Есть еще одно важное свойство лежандровых узлов, которое известно под названием производящее семейство функций. Это вот что. Давайте вернемся к моим x, z-диаграммам, и скажем, что самая простая



Р и с. 12. Семейство функций

лежандрова кривая, которую можно себе вообразить — это просто график функции. Пусть есть график функции $z = f(x)$. Там нет никаких каспов, просто кривая — график функции. И это тоже есть лежандрова кривая в пространстве x, y, z. Но это не лежандров узел. А теперь представим себе, что у нас есть не одна функция, а однопараметрическое семейство функций (рис. 12). Вы легко можете написать формулу (кубический полином). Это семейство кривых имеет огибающую. Если у вас есть функция f от x и еще параметра от t, то можно взять уравнения $f(t, x) = 0$ и $\frac{\partial f}{\partial t}(t, x) = 0$, исключить t из двух этих уравнений, и получится (как знаете из анализа) уравнение огибающей. И как многие тоже знают, это будет кривая с каспом. Это выглядит, как диаграмма лежандрова узла.

Можно эту конструкцию усложнить. А именно, взять не один параметр, например, в качестве множества параметров взять какое-нибудь компактное многообразие. Нарисовать это уже трудно. Теперь будет $f(x, t_1, t_2, \ldots, t_k)$, и все частные производные по t нужно приравнять нулю. Получится некоторая кривая на плоскости. Эта кривая, если повезет, будет лежандровым узлом. А если параметры составляют компактное многообразие, то она точно будет лежандровым узлом. Вопрос такой: если есть лежандров узел, то можно ли таким образом его представить? Ответ фантастический: можно в том и только в том случае, когда существует то же самое, что нужно для существования аугментации. Эти два факта оказываются равносильными. Но первое впечатление — что между ними нет никакой связи. Это можно переформулировать так: лежандров узел тогда и только тогда обладает этим свойством, когда алгебра Элиашберга—Чеканова имеет аугментацию. Более того, как я сказал, в этой конструкции замешано некое компактное многообразие — параметры принимают значения в некотором компактном многообразии. С многообразиями связано такое явление, как комплекс Морса. В действительности конструкция

более сложная (нужно взять не это многообразие, а его декартов квадрат). В общем, есть некий комплекс Морса, который возникает при наличии управляющего семейства, и этот комплекс тоже, как и его гомологии, является лежандровым инвариантом узла. Более того, этот комплекс градуирован, и его образующие тоже соответствуют правым каспам и перекрестьям, и имеют те же самые степени. Конечно, поверить в то, что это разные комплексы, невозможно. По всей видимости, дело обстоит так, что вся эта линеаризованная конструкция Чеканова—Элиашберга как бы излишняя. Существуют узлы, которые задаются производящим семейством; тогда это всё распространяется на высшие размерности (как, впрочем, и в конструкции Элиашберга—Чеканова). Существуют наряду с лежандровыми узлами многомерные лежандровы узлы. Контактное пространство не обязательно имеет размерность 3, оно может иметь любую нечетную размерность. И если оно имеет размерность $2n+1$, в нем имеется много n-мерных лежандровых многообразий — они так же определяются. И для них есть вся та же самая наука. И в частности, производящие семейства тоже. И эта конструкция Пети Пушкаря тоже работает. Как, впрочем, и конструкция Чеканова—Элиашберга. Между ними есть какая-то немыслимая связь, о которой я не имею никакого понятия.

В конструкции Чеканова—Элиашберга есть такое обстоятельство, что она перестает работать, как только вы к узлу приделываете зигзаг (рис. 7). Просто всё делается нулем. Если вы возьмете лежандров узел и приделаете один крошечный зигзаг, то конструкция Элиашберга—Чеканова перестает работать. И поэтому могла бы возникнуть гипотеза, что нет инвариантов, различающих узлы с такими зигзагами. Эти узлы оказываются важными в некотором контексте, о котором я сейчас не буду говорить. Я расскажу про одну действительно занятную вещь. Рассмотрим эти инварианты. Они совершенно не работают при наличии такого зигзага. Однако, спрашивается, можно ли вообразить, что действительно, если взять два узла с одинаковыми числами Беннекина и Маслова, и к каждому из них приделать по такому зигзагу, то они уже будут лежандрово изотопны? В частности, если взять два узла, которые различаются в работе Чеканова и Элиашберга. И мы можем двумя способами преобразовать с помощью движений Святковского. Оказывается, что после такого присоединения не работает инвариант Чеканова—Элиашберга, и узлы оказываются изотопны. Но если взять эти два зигзага и их повернуть в другую сторону, то не получится построить изотопию.

13 июня 2003

Е. Б. Дынкин

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И АНАЛИЗ

Вступительное слово Ю. С. Ильяшенко

Мне хотелось сказать, что, по моему ощущению, Евгений Борисович, которого московская математическая школа помнит всегда, у которого учатся московские и российские, и, в общем-то, математики мирового математического сообщества, у которого мы учились всегда, — Евгений Борисович никогда не терял связи с нами. Вот здесь, в этой аудитории сидят его школьники, которые несколько, может быть, выросли с тех пор, как у него учились, но сохранили, я думаю, прежнее отношение к своему учителю; сидят ученики Евгения Борисовича. И мы очень счастливы тем, что мы можем поставить маленькую вешку в этой истории связи Евгения Борисовича с нашей математической школой, с нашим математическим сообществом. Спасибо.

Уравнение Лапласа и броуновское движение

Начиная с 1976 года, я почти не употреблял русского языка в моих лекциях и докладах. Сегодня я буду говорить по-русски, пользуясь тем, что все мои слушатели владеют этим языком. Однако я буду использовать английские слайды, приготовленные с учетом того, что через неделю я должен выступать на близкую тему в Англии. Так как всегда лучше недооценить, чем переоценить подготовку аудитории, я не буду предполагать ничего за пределами университетского курса анализа (включая интеграл Лебега) и некоторого представления о том, что такое вероятность.

Я начну с классического предмета — связи между проблемой Дирихле и случайным блужданием. Речь идет о гармонических функциях, то есть решениях уравнения Лапласа $\sum\limits_{i=1}^{d} \frac{\partial^2 u}{\partial x_i^2} = 0$ в области D d-мерного

Статья представляет собой расширенное и дополненное изложение доклада, прочитанного на заседании семинара «Глобус» 18 июня 2003 года. В статье отражено развитие предмета за годы, прошедшие после доклада.

пространства. Чтобы можно было рисовать картинки, мы будем предполагать, что $d=2$. Мы ищем гармоническую функцию u в области D, которая равняется заданной функции f на границе ∂D.

Вероятностный подход к этой задаче содержится уже в известной работе Куранта, Фридрихса и Леви, опубликованной в 1928 г. Авторы начинают с дискретного аналога проблемы. Функция u задается на решетке $\mathbb{Z}$, а условие $\Delta u=0$ заменяется на условие, что функция u в точке x равна среднему значению во всех соседних точках:

$$u(x)=\frac{1}{4}\sum u(x'), \tag{1}$$

где x' — четыре соседа x. Для решения проблемы Дирихле авторы используют случайное блуждание.

Отправляясь от какой-то точки $x\in D$, мы движемся с равными вероятностями в четырех возможных направлениях и попадаем в одну из соседних четырех точек, которую мы обозначаем x_1. Повторяя эту процедуру, мы попадаем последовательно в точки x_2, x_3, … В популярных книжках это интерпретируется как прогулка по городу с шахматной планировкой. Прогулка продолжается до тех пор пока мы впервые не выйдем из D. При этом мы оказываемся в точке $\xi_\tau\in\partial D$, где τ — это время первого выхода. Решение дискретной проблемы Дирихле, отвечающее функции f, можно получить, умножая значения функции f в каждой точке границы на вероятность нашего пути ξ к этой точке и беря затем сумму таких произведений по всем точкам границы. Вероятность каждого пути длительности τ равна $\left(\frac{1}{4}\right)^\tau$, так что решение нашей задачи дается формулой

$$u(x)=\sum f(\xi_\tau)\left(\frac{1}{4}\right)^\tau.$$

Распределение вероятностей в пространстве путей зависит от исходной точки x, и мы обозначаем его через Π_x. Значение $f(\xi_\tau)$ является функцией пути ξ. Обозначая через $\Pi_x f(\xi_\tau)$ интеграл этой функции по мере Π_x, мы приходим к выражению

$$u(x)=\Pi_x f(\xi_\tau). \tag{2}$$

Следующий шаг состоит в переходе к пределу, когда решетка становится все более густой. Точнее, мы заменяем решетку $\mathbb{Z}$ на $\varepsilon\mathbb{Z}$ и устремляем ε к нулю. В пределе дискретное уравнение (1) становится уравнением Лапласа, а функция u, заданная формулой (2), стремится к решению задачи Дирихле для уравнения Лапласа. Эта процедура (известная как метод сеток) применима к широкому классу дифференциальных уравнений с частными производными.

Однако желательно иметь для решения задачи Дирихле формулу, не содержащую предельного перехода. Этого можно достичь, используя вместо случайного блуждания броуновское движение, которое возникает как предел случайных блужданий по решетке $\varepsilon\mathbb{Z}$ при стремящемся к нулю ε. При этом надо считать, что каждый переход занимает время δ и что отношение $\delta^2/d\varepsilon$ стремится к единице (напомним, что d — это размерность пространства, где происходит блуждание).

Мера Π_x в пространстве дискретных путей превращается в пределе в вероятностную меру в пространстве непрерывных путей, начинающихся в точке x. Для этой меры (впервые введенной Минером) мы сохраняем обозначение Π_x. Для каждого фиксированного момента t

$$\Pi_x f(\xi_t) = \int\limits_{\mathcal{R}^d} \frac{1}{(2\pi t)^{d/2}} e^{-|y-x|^2/2t} f(y)\, dy.$$

При такой интерпретации Π_x формула (2) применима и к уравнению Лапласа. И с этой формулой можно работать весьма эффективно, так как интеграл Лебега совершенно нечувствителен к тому, в каком пространстве задана мера.

Для того, чтобы интеграл был определен, достаточно предположить, что функция f измерима и ограничена. Мы будем считать, что она непрерывна. Нужны также некоторые условия относительно границы области. Их необходимость видна из следующего примера. Пусть D — это круг с выколотым центром. Граница состоит из внешней окружности и центра. Так как броуновская траектория, начинающаяся внутри D, с вероятностью 1 не проходит через центр, то решение u, соответствующее f по формуле (2), равно единице всюду, если $f = 1$ на внешней окружности. Следовательно, мы не можем задать произвольно значение f в центре. Чтобы исключить подобную ситуацию, вводится понятие регулярной точки. Существуют вероятностное и аналитическое определения этого понятия. Вероятностное определение (которое представляется более наглядным) таково: точка границы регулярна, если траектория, начинающаяся из этой точки, с вероятностью 1 выходит из области за сколь угодно малое время (начальное положение при этом не принимается во внимание). Отсюда немедленно следует простое и удобное достаточное условие: точка границы регулярна, если ее можно коснуться снаружи маленьким прямолинейным отрезком. *) (Легко доказывается, что броуновская траектория, выходящая из конца отрезка, пересекает этот отрезок с вероятностью 1 за сколь угодно малое время). Ясно, что в примере с кругом центр не является

*) Критерий с отрезком применим к размерности $d = 2$. Если $d > 2$, то вместо отрезка надо рассмотреть $(d-1)$-мерный симплекс.

регулярной точкой. Если граница гладкая, то все ее точки регулярны. Это первый пример взаимодействия между теорией вероятностей и теорией дифференциальных уравнений.

Связь между дифференциальными уравнениями и броуновским движением (и более общими так называемыми диффузионными процессами) известна уже давно. Значительная роль в установлении такой связи принадлежит работе Колмогорова «Об аналитических методах в теории вероятностей», напечатанной в 1931 г. В те времена речь шла о применении анализа к теории вероятностей. С дальнейшим развитием теории случайных процессов (Ито, Маллиавен, ...) и теории меры в функциональных пространствах (Колмогоров, Дуб, ...) стало возможным говорить и о вероятностных методах в анализе. В каком-то смысле теория случайных процессов может рассматриваться как часть анализа. Вероятностные идеи играют важную роль и в современной физике. Об отношении физиков к этим идеям, говорит, например, следующее посвящение в монографии Барри Саймона «The $P(\phi_2)$ Euclidean (Quantum) Field Theory»: «Посвящается Эду Нелсону, которому я обязан пониманием того, сколь неестественно воспринимать теорию вероятностей как что-то неестественное».

Нелинейное уравнение $\Delta u = \psi(u)$ и суперброуновское движение

Сегодня я буду говорить об одной задаче, которой я занимаюсь последние годы. В настоящее время нелинейные дифференциальные уравнения находятся в центре внимания аналитиков. Наш предмет — простейшее нелинейное уравнение

$$\Delta u = \psi(u), \tag{3}$$

где ψ — положительная нелинейная функция, например, u^2. Уравнения, в которых нелинейность не затрагивает производных высшего порядка, называют полулинейными. Наше уравнение принадлежит этому классу.

Оказывается, что это уравнение (и схожие с ним другие уравнения) можно изучать с помощью случайного процесса, который я предложил называть суперброуновским движением. Конечно, приставкой «супер» очень злоупотребляют: суперструны, супермартингалы, супермаркеты, ... Но тем не менее это название привилось. Дело в том, что альтернативное название очень длинно, а так каждому марковскому процессу можно сопоставить суперпроцесс: броуновскому движению — суперброуновское движение, диффузии — супердиффузию и т. п.

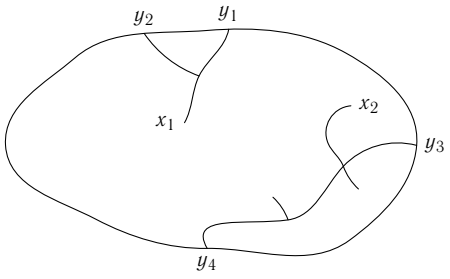
Что же такое суперброуновское движение? В отличие от броуновского движения, которое описывает случайное движение частицы, суперброуновское движение — это модель, которая описывает случайную эволюцию облака, состоящего из многих мелких частиц. Предельный переход (подобный переходу от случайного блуждания к броуновскому движению) приводит к модели, где состояния облака описываются недискретными мерами, характеризующими распределение масс. Специальную роль играют выходные меры X_D. На эвристическом уровне X_D — это мера на границе области D, которая возникает, если все частицы замораживаются, когда они достигают границы ∂D.

Формула для решения задачи Дирихле похожа на формулу (2). Значение $f(\xi_\tau)$ в точке выхода надо заменить на интеграл $\langle f, X_D\rangle$ функции f по выходной мере X_D. Нелинейная зависимость решения u от граничной функции f отражается в появлении показательной и логарифмической функций. Мера Π_x в пространстве траекторий броуновского движения, исходящих из точки x, заменяется на меру P_x в пространстве траекторий суперброуновского движения, начинающихся с единичной меры δ_x, сосредоточенной в точке x. Формула, задающая решение, выглядит следующим образом:

$$u(x) = -\log P_x e^{-\langle f, X_D\rangle}. \tag{4}$$

Как я уже говорил, суперброуновское движение получается предельным переходом из модели, описывающей эволюцию системы конечного числа частиц. Каждая частица совершает броуновское движение, пока она не погибнет или не достигнет границы области. В первом случае она производит в момент смерти m потомков с вероятностью p_m. Во втором случае она замораживается.

Рисунок 1 объясняет на примере, как возникает выходная мера *).



Р и с. 1. Ветвящееся броуновское движение

В начальный момент имеется две частицы, расположенных в точках x_1, x_2. В момент смерти первая частица оставляет двух потомков, которые совершают броуновское движение независимо друг от друга пока не достигают границы ∂D в точках y_1, y_2. Из трех непосредственных потомков второй частицы один попадает на границу в точке y_3, другой

*) Конечно, это только схема. Траектории броуновского движения крайне нерегулярны, что не показано на нашем рисунке.

умирает бездетным, а третий производит две частицы, из которых только одна достигает границы (в точке y_4).

В нашем примере выходная мера

$$X_D = \delta_{y_1} + \delta_{y_2} + \delta_{y_3} + \delta_{y_4}.$$

Траектория каждой частицы зависит от места ее рождения, но в остальном она независима от фамильной истории.

В общем случае вначале имеется m частиц, находящихся в точках $x_1, \ldots, x_m$. Предполагая, что масса каждой частицы равна β, мы имеем начальную меру

$$\mu = \sum \beta\delta_{x_i}$$

и выходную меру

$$X_D = \sum \beta\delta_{y_i}.$$

Таким образом мы определяем семейство случайных мер (X_D, P_μ), где D — произвольная область, а μ — произвольная мера со значениями $0, \beta, 2\beta, \ldots$ Далее мы переходим к пределу, когда начальная мера стремится к μ, а время жизни каждой частицы и β стремятся к нулю в определенной пропорции. Мы снова получаем семейство случайных мер, однако μ и X_D уже не дискретны. Это семейство и задает суперброуновское движение, которое наследует от ветвящегося броуновского движения два важных свойства. Первое из них — ветвящееся свойство — выражается формулой

$$P_\mu e^{-\langle f, X_D\rangle} = e^{-\langle u, \mu\rangle}, \tag{5}$$

где u определено формулой (4).

Второе — марковское свойство гласит, что «прошлое» и «будущее» независимы при условии, что известно «настоящее». При этом роль «прошлого» и «будущего» играют события, определенные, соответственно, поведением системы внутри и вне области D, а «настоящее» задается значением X_D *).

Функция (4) удовлетворяет интегральному уравнению

$$u(x) + \Pi_x \int_0^\tau \psi[u(\xi_s)]ds = \Pi_x f(\xi_\tau), \tag{6}$$

*) Технически это выражается формулой

$$P_\mu\{C\colon \mathcal{F}_{\subset D}\} = P_{X_D}(C) \quad P_\mu\text{-a.s.}$$

для всех $C \in \mathcal{F}_{\supset D}$. Через $\mathcal{F}_{\subset D}$ и $\mathcal{F}_{\supset D}$ обозначены σ-алгебры, порожденные, соответственно, $X_{D'}$, $D' \subset D$ и $X_{D''}$, $D'' \supset D$.

где τ — это момент первого выхода из области D, а $\psi(u)$ получается определенным предельным переходом из производящей функции $\varphi(u)=\sum_0^\infty p_n u^n$ распределения вероятностей для числа потомков каждой частицы. Если D — ограниченная область с гладкой границей, а функция f непрерывна, то уравнение (6) влечет за собой, что u является решением задачи Дирихле для уравнения (3), отвечающим граничной функции f.

Полезность формулы (3) можно продемонстрировать на следующем простом примере. Будем искать решения уравнения (2), стремящиеся к бесконечности при приближении к границе (аналитики называют их «большими решениями»). Заметим, что предел e^{-na} при $n\to\infty$ равен нулю, если $a>0$, и равен 1, если $a=0$. Функция (3) при $f=n$ задает решение u_n, равное n на границе. По теореме о монотонном предельном переходе под знаком интеграла Лебега

$$u=\lim u_n=-\log P_x \lim e^{-n\langle 1,X_D\rangle}=-\log P_x\{X_D=0\}.$$

Легко видеть, что u является «большим решением», по крайней мере, если область D ограничена. (Из вероятностных соображений следует, что для таких областей $P_x\{X_D=0\}>0$.)

Положительные решения

Я остановлюсь подробнее на одном направлении в теории полулинейных дифференциальных уравнений — описании всех положительных решений таких уравнений. В результате усилий ряда вероятностников и аналитиков за последние годы в этом направлении были достигнуты весьма значительные результаты.

Для уравнения Лапласа решение давно известно. Существует взаимно однозначное соответствие между положительными гармоническими функциями h в ограниченной гладкой области D и конечными мерами на границе ∂D. Оно задается интегралом Пуассона

$$h(x)=\int\limits_{\partial D} k_D(x,\,y)\,\nu(dy). \tag{7}$$

Мы называем меру ν граничным следом гармонической функции h. Мы пишем $\mathrm{tr}(u)=\nu$ и мы обозначаем гармоническую функцию со следом ν через h_ν.

Функция $k_D(x,\,y)$ называется ядром Пуассона. Для шара радиуса 1 с центром в начале координат

$$k_D(x,\,y)=c_d\frac{(1-|x|^2)}{|x-y|^d},$$

где c_d — константа, зависящая только от d. Вероятностный смысл ядра Пуассона виден из формулы

$$\Pi_x\{\xi_\tau \in B\} = \int\limits_B k_D(x, y),$$

где B — произвольное борелевское подмножество границы и интеграл берется относительно площади поверхности $\sigma(dy)$.

Обратимся теперь к изучению множества $\mathcal{U}$ всех положительных решений уравнения (3). Специальная роль принадлежит решениям u, ограниченным сверху гармоническими функциями. Мы называем их умеренными и мы обозначаем множество всех умеренных решений через $\mathcal{U}_1$. Для каждого $u \in \mathcal{U}_1$ существует минимальная гармоническая функция $h = h_\nu$, мажорирующая u. По определению $\mathrm{tr}(u) = \mathrm{tr}(h_\nu) = \nu$. Умеренное решение со следом ν обозначается u_ν. Меры ν, соответствующие умеренным решениям, образуют подкласс $\mathcal{N}_1$ класса конечных мер на границе, а $\nu \to u_\nu$ — это монотонное взаимно однозначное отображение $\mathcal{N}_1$ на $\mathcal{U}_1$.

Другое важное подмножество класса $\mathcal{U}$ — это множество $\mathcal{U}_0$ σ-умеренных решений. Мы говорим, что u σ-умеренно, если существует неубывающая последовательность умеренных решений u_{ν_n}, сходящаяся поточечно к u. Для всякого B $\nu_1(B) \leqslant \nu_2(B) \leqslant \ldots$ и поэтому ν_n сходятся к σ-конечной мере ν. Мы пишем $u = u_\nu$ и мы обозначаем через $\mathcal{N}_0$ класс мер, которым соответствуют по этому рецепту σ-умеренные решения u_ν. Отображение $u \to u_\nu$ продолжается однозначно в монотонное отображение некоторого класса $\mathcal{N}_0$ σ-конечных мер на $\mathcal{U}_0$. (Мера ν принадлежит $\mathcal{N}_0$, если она является пределом неубывающей последовательности ν_n, и u_ν определяется как предел ν_n.) Следует отметить, что из равенства $u_\nu = u_\mu$ не следует, что $\nu = \mu$.

След произвольного решения u — это пара (Γ, ν), где борелевское множество $\Gamma \subset \partial D$ — это множество $\mathrm{SG}(u)$ сингулярных точек u, а ν — σ-конечная мера на ∂D, не заряжающая Γ. Мы начнем с определения сингулярных точек.

Мы говорим, что точка $y \in \partial D$ сингулярна для решения u, если $\psi'(u)(x) \to \infty$ достаточно быстро, когда $x \to y$. Почему возникает производная ψ'? Дело в том, что $\mathcal{U}$ — это бесконечномерное многообразие в функциональном пространстве, и его касательная гиперплоскость в точке u задается линейным уравнением $\Delta v = \psi'(u)v$. *) Для произвольной непрерывной положительной функции $a(x)$ мы определяем точки быстрого

*) Другими словами, уравнение $\Delta v = \psi'(u)v$ является линеаризацией уравнения (3) в точке u.

роста как точки $y \in \partial D$, такие что

$$\int_0^\tau a(\xi_t)dt = \infty \quad \Pi_x^y\text{-почти наверное для всех } x \in D. \tag{8}$$

Здесь Π_x^y означает условное распределение вероятностей для траектории броуновского движения, начинающегося в точке $x \in D$, при условии, что она выходит на границу в точке y. (Это важное понятие было введено Дубом.) Таким образом, сингулярное множество $\mathrm{SG}(u)$ определяется как множество точек y, подчиняющихся (при любом x) условию

$$\int_0^\tau \psi'[u(\xi_t)]dt = \infty \tag{9}$$

для всех траекторий условного броуновского движения, за исключением множества Π_x^y меры 0.

Остается определить вторую часть следа — меру ν. Обозначим через $\mathcal{N}_1(\Gamma)$ множество всех мер класса $\mathcal{N}_1$, сосредоточенных на Γ. Мера ν задается формулой

$$\nu(B) = \sup\{\mu \in \mathcal{N}_1(\Gamma) \colon u_\mu \leqslant u\}. \tag{10}$$

Для того чтобы классифицировать положительные решения уравнения (3), необходимо выяснить, какие пары являются следами, и описать все решения с данным следом. Для этого вводятся следующие понятия.

(1) Частичный порядок $u \leqslant v$ в $\mathcal{U}$. Доказывается существование для любого множества $\widetilde{\mathcal{U}} \subset \mathcal{U}$ супремума $\operatorname{Sup} \widetilde{\mathcal{U}} \in \mathcal{U}$ относительно этого порядка.

(2) Операции

$$u \oplus v = \operatorname{Sup}\{w \in \mathcal{U} \colon w \leqslant u + v\},$$
$$u \vee v = \operatorname{Sup}\{u, v\}.$$

Доказывается, что если $\widetilde{\mathcal{U}}$ замкнуто относительно операции $\vee$, то существует неубывающая последовательность $u_n \in \widetilde{\mathcal{U}}$, сходящаяся поточечно к u.

(3) Класс полярных множеств: множество $B \subset \partial D$ называется полярным, если $\nu(B) = 0$ для всех $\nu \in \mathcal{N}_1$. Мы говорим, что борелевские множества B_1, B_2 эквивалентны, и мы пишем $B_1 \sim B_2$, если симметрическая разность B_1 и B_2 полярна.

(4) Семейство решений

$$u_B = \operatorname{Sup}\{u_\nu \colon \nu \in \mathcal{N}_1(\mathcal{D})\},$$

отвечающее борелевским множествам B. Очевидно, $u_{B_1} = u_{B_2}$, если $B_1 \sim B_2$.

(5) Класс борелевских множеств B, таких что все сингулярные точки u_B принадлежат B. Мы называем их f-замкнутыми множествами. (Этот класс определяет f-топологию в $\mathcal{U}$.)

Если B_1 и B_2 f-замкнуты, то равенство $u_{B_1} = u_{B_2}$ влечет за собой, что $B_1 = B_2$.

Доказывается, что:

А. След (Γ, ν) любого решения u удовлетворяет условиям:

А.1. Γ — f-замкнутое борелевское множество.

А.2. ν — мера класса $\mathcal{N}_0$ такая, что $\nu(\partial D \setminus \Gamma) = 0$ и $\mathrm{SG}(u_\nu) \subset \Gamma$.

Б. Если пара (Γ, ν) удовлетворяет условиям А.1—А.2, то существует решение u со следом (Γ', ν), где $\Gamma' \sim \Gamma$.

Специальную роль играют решения $u_{\Gamma,\nu} = u_\Gamma \oplus u_\nu$.

В. Если $\mathrm{tr}(u) = (\Gamma, \nu)$, то $u_{\Gamma,\nu}$ — это максимальное σ-умеренное решение, не превосходящее u.

Г. Если (Γ, ν) удовлетворяет условиям А.1—А.2, то $\mathrm{tr}(u_{\Gamma,\nu}) = (\Gamma', \nu)$, где $\Gamma' \sim \Gamma$, и $u_{\Gamma,\nu}$ является минимальным решением с этим свойством. Более того, $u_{\Gamma,\nu}$ единственное σ-умеренное решение в этом классе.

Введем класс $\mathcal{U}^{(s)}$ сингулярных решений, определяемый условием

$$u \oplus u = u.$$

Доказывается, что этот класс совпадает с совокупностью решений u_Γ, соответствующих f-замкнутым множествам Γ. Определим сингулярную часть $u^{(s)}$ решения u как супремум всех сингулярных решений, не превосходящих u. Устанавливается, что $u^{(s)} = u_\Gamma$, где $\Gamma = \mathrm{SG}(u)$ и, следовательно, $u^{(s)} = u_{\Gamma'}$ для всех множеств Γ', эквивалентных Γ. Таким образом, каждому решению u соответствует класс эквивалентных σ-замкнутых множеств и мера ν, не заряжающая ни одного из этих множеств. Каждая из этих пар удовлетворяет условиям А.1—А.2. Тем самым устанавливается взаимно однозначное соответствие между σ-умеренными решениями и классами эквивалентных пар, удовлетворяющих условиям А.1—А.2. (Две пары (Γ, ν) и (Γ', ν) считаются эквивалентными, если $\Gamma \sim \Gamma'$.)

Этим решается задача классификации σ-умеренных решений. Доказано, что все решения уравнения

$$\Delta u = u^\alpha, \quad \text{где } 1 < \alpha \leqslant 2, \tag{11}$$

σ-умеренны. Следовательно, для этого уравнения мы умеем описывать все положительные решения уравнения (3). В общем случае остается

открытым кардинальный вопрос: существуют ли не σ-умеренные положительные решения?

Исторический очерк

Систематическая теория суперпроцессов была заложена Синдзю Ватанабе [10], который построил такие процессы посредством предельного перехода от систем частиц, движущихся по траекториям марковского процесса и размножающихся. Он же установил связь между этими процессами и нелинейными дифференциальными уравнениями.

Глубокие свойства траекторий суперброуновского движения, отвечающего уравнению $\Delta u = u^2$, были открыты Даусоном и Перкинсом. Обзор этих результатов содержится в [1] и [9].

В [2] Дынкин ввел выходные меры, что позволило применять суперпроцессы к исследованию граничных задач для нелинейных дифференциальных уравнений. Описанный в предыдущей секции след — это тонкий след, определенный и изученный в работе Дынкина и Кузнецова [5]. В случае уравнения (11) с $\alpha < (d+1)/(d-1)$ он совпадает с грубым следом, рассматривавшимся до этого в литературе. Систематическое изложение теории тонкого следа можно найти в главе 11 монографии [3]. В эпилоге к этой монографии формулируется критическая проблема: доказать или опровергнуть, что все решения уравнения (3) σ-умеренны. Для уравнения $\Delta u = u^2$ положительный ответ на этот вопрос был дан в диссертации Б. Мселати [7], написанной под руководством Ж.-Ф. Ле Галля. Этот результат был распространен на уравнение (11) в [4]. В общем случае проблема остается открытой.

В [8] Маркус и Верон ввели понятие точного следа. Это одна из выше описанных эквивалентных пар (Γ, ν), и она характеризуется дополнительным свойством: Γ равно объединению $\mathrm{SG}(u_\Gamma)$ и множества точек взрыва меры ν. *) Эта связь между тонким и точным следами установлена в [6].

Литература

[1] *Dawson D. A.* Measure-valued Markov processes // École d'Été de Probabilités de Saint Flour, 1991. Springer, 1993. (Lecture Notes in Math.; V. 1541). P. 1—260.

[2] *Dynkin E. B.* A probabilistic approach to one class of nonlinear differential equations // Probab. Th. Rel. Fields. 1991. V. 89. P. 89—115.

[3] *Dynkin E. B.* Diffusions, superdiffusions and partial differential equations. Providence, RI: AMS, 2002.

[4] *Dynkin E. B.* Superdiffusions and positive solutions of nonlinear partial differential equations. Providence, RI: AMS, 2004.

*) y является точкой взрыва ν, если $\nu(O) = \infty$ для всякой f-окрестности O точки y.

[5] *Dynkin E. B., Kuznetsov S. E.* Fine topology and fine trace on the boundary associated with a class of semilinear differential equations // Comm. Pure Appl. Math. 1998. V. 51. P. 897—936.

[6] *Dynkin E. B., Kuznetsov S. E.* A class of boundary traces for solutions of the equation $Lu = \psi(u)$ // Journal of Functional Analysis. 2007. V. 252. P. 696—709.

[7] *Mselati B.* Classification and probabilistic representation of the positive solutions of a semilinear elliptic equation // Memoirs of the American Mathematical Society. Providence, RI: AMS, 2004. V. 168, № 798.

[8] *Marcus M., Véron L.* The precise boundary trace of positive solutions of the equation $\Delta u = u^q$ in the supercritical case // Perspectives in Nonlinear Partial Differential Equations // In honor of Haim Brezis. Providence, RI: AMS, 2007. (Contemp. Math.; V. 446.); arxiv.org/math/0610102.

[9] *Perkins E.* Dawson-Watanabe Superprocesses and Measure-valued Diffusions // École d'Été de Probabilités de Saint Flour, 1999. Springer, 2002. (Lecture Notes Math.; V. 1781).

[10] *Watanabe S.* A limit theorem of branching processes and continuous state branching processes // J. Math. Kyoto Univ. 1968. V. 8. P. 141—167.

18 июня 2003 г.

Оглавление

ГЛОБУС
Общематематический семинар. Выпуск 4

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

Подписано в печать 24.04.2009 г. Формат $70 \times 100\ {}^{1}/_{16}$. Бумага офсетная.
Печать офсетная. Печ. л. 14. Тираж 800 экз. Заказ №

Издательство Московского центра непрерывного математического образования.
119002, Москва, Большой Власьевский пер., 11. Тел. (499) 241–74–83.

Отпечатано по CtP-технологии в ОАО «Печатный двор» им. А. М. Горького.
197110, Санкт-Петербург, Чкаловский проспект, 15.

Книги издательства МЦНМО можно приобрести в магазине «Математическая книга»,
Большой Власьевский пер., д. 11. Тел. (499) 241–72–85. E-mail: biblio@mccme.ru