УДК 51(06)
С88

*Издание осуществлено при поддержке РФФИ, проект № 99–01–14016*

С88 **Студенческие чтения НМУ.** Вып. 1. — М: МЦНМО, 2000.— 224 с.

ISBN 5-900916-52-9

В книге представлены лекции, прочитанные в Независимом московском университете в 1997–98 г., предназначенные для широкой аудитории. Их цель — рассказать о некоторых областях математики и описать новые идеи.

Для студентов, аспирантов и преподавателей математических специальностей.

УДК 51(06)

**ISBN 5-900916-52-9**

# Предисловие

Студенческие чтения Независимого московского университета — новая традиция в математической жизни Москвы. Первая лекция серии была прочтена В. И. Арнольдом в мае 1997 года. Затем последовали лекции Ю. И. Манина (октябрь 97 г.), А. А. Кириллова (декабрь 97 г. и январь 98 г.), Д. В. Аносова (три лекции в феврале–марте 98 г.), А. А. Разборова (апрель 98 г.), С. П. Новикова (июнь 98 г.), М. Рида (Warwick Univesity, Англия; февраль 99 г.), А. Катка (Penn State University, США; март 99 г.), С. Смейла (май 99 г.), П. Картье (май 99 г.); Я. Г. Синая (июнь 99 г). Каждая лекция, за исключением лекции Манина (которую лектор, несмотря на очень широкий подтекст, называет в шутку «Об одном свойстве одного решения одного дифференциального уравнения») представляет собой обзор большого круга проблем. Лекции рассчитаны на широкую аудиторию от студентов до исследователей-профессионалов. Они не содержат никаких доказательств или технических деталей. Их цель — нарисовать панораму целой области исследований и описать новые идеи.

Предлагаемый сборник содержит лекции, прочитанные в 97–98 годах.

Как и серия «Студенческие чтения НМУ», сам Независимый университет — молодое детище московской математической школы. Он работает с 1991 года и воспитывает математиков-исследователей. Будучи совсем небольшим — 80 студентов, 13 аспирантов и около 40 преподавателей, — НМУ имеет возможность гибко менять программу обучения, сохраняя современный уровень и стиль преподавания.

Выпускники НМУ немногочисленны. Однако облик математической науки в стране определяется, как правило, творчеством очень немногих лидеров, которые создают свои собственные школы. Авторы предлагаемого читателю сборника оказали решающее влияние на облик российской математической школы в последние 30 лет. Мы надеемся, что среди их слушателей есть те, кто внесет свой скромный вклад в облик российской математики в следующее тридцатилетие.

*Ю. С. Ильяшенко*

*В. И. Арнольд*

# Таинственные математические троицы

*Лекция 21 мая 1997 года*

Я постараюсь рассказать о некоторых удивляющих меня явлениях в математике. В большинстве случаев они не формализованы. Их даже нельзя сформулировать в виде гипотез. Гипотеза отличается тем, что её можно опровергнуть: она либо верна, либо неверна.

Речь пойдёт об определённых наблюдениях, которые приводят к очень большому числу теорем и гипотез, которые уже можно проверять или опровергать. Но интерес, который они представляют, состоит в общей точке зрения.

Объясню эту общую точку зрения на простом примере — на примере линейной алгебры.

Теория линейных операторов описывается в современной математике как теория алгебр Ли серии $A_n$, т. е. $\mathfrak{sl}(n+1)$, и формулируется в терминах систем корней. Система корней сопоставляется любой группе Коксетера, т. е. конечной группе, порождённой отражениями (во всяком случае, кристаллографической). Если взять какое-нибудь утверждение линейной алгебры, относящееся к этому частному случаю этой группы $A_n$, и изгнать всё содержание из формулировки, так чтобы там больше не шло речи ни о собственных числах, ни о собственных векторах, а оставались бы только одни корни, то результат можно будет сформулировать и для других серий: $B_n$, $C_n$, $D_n$ и даже исключительных $E_6$, $E_7$, $E_8$, $F_4$, $G_2$ (а иногда даже и для всех систем Коксетера, включая и некристаллографические группы симметрий многоугольников, икосаэдра и гиперикосаэдра, живущего в четырёхмерном пространстве).

С этой точки зрения геометрии других серий ($B$, $C$, ...) — не геометрии векторных пространств с дополнительными структурами (евклидовой, сипмлектической и т. д.), каковыми они, конечно, формально являются, не дочери $A$-геометрии, а равноправные её сёстры.

Приведённая классификация простых алгебр Ли, принадлежащая Киллингу (и приписываемая поэтому Картану), имеет аналог в бесконечномерном случае, в анализе. Алгебраическая задача, решённая Киллингом, Картаном, Коксетером, имеет бесконечномерный аналог в теории алгебр Ли групп диффеоморфизмов. Если имеется многообразие $M$, то естественно возникает группа $Diff(M)$ всех диффеоморфизмов многообразия $M$. Эта группа (точнее, компонента связности единицы в этой группе) в алгебраическом смысле простая, т. е. у неё нет нормальных делителей. Имеются также другие аналогичные «простые» теории, похожие на геометрию многообразий, но от неё отличающиеся. Они тоже были расклассифицированы в своё время Картаном.[1] Наложив небольшое число достаточно естественных ограничений, он обнаружил, что имеются 6 серий таких групп:

$Diff(M)$,

$SDiff(M)$ — группа диффеоморфизмов, сохраняющих заданную форму объёма,

$SpDiff(M, \omega^2)$ — группа симплектоморфизмов.

Затем имеются комплексные многообразия и группы голоморфных диффеоморфизмов.

Есть также очень важная контактная группа, группа контактоморфизмов.

Потом имеются конформные версии некоторых из них. Я не буду описывать их подробно.

Идея, о которой я говорил, заключается в том, что здесь также имеется нечто, аналогичное переходу от теорем линейной алгебры, т. е. системы корней $A_n$, к остальным системам корней. Иными словами, для всей математики, для всей геометрии многообразий, имеются такие операции высшего уровня (например, симплектизация), которые каждому определению и каждой теореме в теории многообразий сопоставляют их аналоги в теории многообразий с элементом объёма или симплектических многообразий. Но это лишь грубо говоря, эта операция — не настоящий функтор.

Например, элемент алгебры Ли группы диффеоморфизмов есть векторное поле. Симплектизация векторного поля — это гамильтоново поле, задающее уравнение Гамильтона

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q}.$$

В других ситуациях всё сложнее. Трудно понять, во что превращается то или иное понятие линейной алгебры при переходе к другим геометриям. Ведь даже в случае группы симплектоморфизмов берётся не вся алгебра Ли этой группы, а только те векторные поля, которые задаются однозначной функцией Гамильтона. Это — коммутатор алгебры Ли, а не сама алгебра Ли группы симплектоморфизмов. И всё же, когда удаётся найти правильные аналоги каких-либо понятий одной геометрии в другой, награда бывает весьма значительной.

Разберём два примера.

---

[1] См., в частности, *Картан Э.* Избранные труды. — М.: МЦМНО, 1998. — *Прим. ред.*

1. Симплектизация. Так называемые гипотезы Арнольда (1965) о неподвижных точках симплектоморфизмов были сформулированы при попытке симплектизировать теорему Пуанкаре—Эйлера о том, что сумма индексов особых точек векторного поля на многообразии равна эйлеровой характеристике. Они оценивают через неравенства Морса (т. е. через число критических точек функции на многообразии) число замкнутых траекторий для гамильтоновых векторных полей.[2)]

Начнём с того, что сформулируем следующее более простое утверждение. Оно было высказано Пуанкаре в виде гипотезы и доказано Биркгофом.

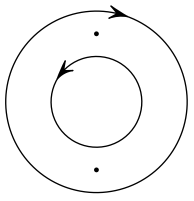


*Рис. 1*

**Теорема 1.** *Пусть диффеоморфизм кругового кольца на себя сохраняет площади, причём при его действии точки каждой из ограничивающих его окружностей перемещаются в одном и том же направлении, а точки различных окружностей движутся в противоположных направлениях* (рис. 1). *Тогда диффеоморфизм имеет хотя бы две неподвижные точки.*

Это утверждение вытекает из чуть более общей теоремы о неподвижных точках диффеоморфизмов тора.

**Теорема 2.** *Пусть диффеоморфизм $F$ тора $T^2 = \mathbb{R}^2/\mathbb{Z}^2$ задаётся формулой $x \mapsto x + f(x)$ в стандартной системе координат. Пусть $F$ сохраняет площади и «сохраняет центр тяжести», т. е. среднее значение функции $f$ (рассматриваемой как функция на торе со стандартной метрикой) равно нулю. Тогда $F$ имеет не менее четырёх неподвижных точек.*

Это связано с тем, что сумма чисел Бетти тора равна 4.

Первое доказательство этой теоремы получено Я. Элиашбергом. Но это доказательство никто не проверил. Заведомо правильное доказательство было опубликовано в 1983 г. Конли и Цендером, и это доказательство положило начало развитию целой большой теории — симплектической топологии (в работах Шаперона, Лауденбаха, Сикорава, Чеканова, Громова, Флоера, Хофера, Гивенталя и многих других).[3)] В последние месяцы появились сообщения, что и первоначальные гипотезы (о том, что число неподвижных

[2)] *Арнольд В. И.* Об одном топологическом свойстве глобально канонических отображений классической механики. // *Арнольд В. И.* Избранное — 60. — М: Фазис, 1997. С. 81—86. — *Прим. ред.*

[3)] О симплектической топологии см., например, *Арнольд В. И.* Первые шаги симплектической топологии // Там же. С. 365—389. См. также литературу, указаанную на с. XL той же книги. — *Прим. ред.*

точек точного симплектоморфизма не меньше минимального числа критических точек функции на многообразии, хотя бы для симплектоморфизмов и функций общего положения) наконец доказаны (несколькими независимыми группами в разных странах).

2. Другой пример: переход от $\mathbb{R}$ к $\mathbb{C}$. На этом примере, возможно, проще объяснить, в чём состоит дело. Речь пойдёт о переходе от вещественного случая к комплексному. Имеется вещественная геометрия и имеется комплексная геометрия. Каким образом от вещественной геометрии перейти к комплексной геометрии? Например, в вещественной геометрии есть понятие многообразия с краем, на котором основаны такие понятия, как гомологии и гомотопии. И вообще вся топология существенным образом использует понятие края.

Можно поставить вопрос: во что превращается понятие края при комплексификации?

Если мы предполагаем, что всю математику можно комплексифицировать, то, в частности, можно комплексифицировать и различные понятия математики. Давайте составим таблицу, во что превращаются разные понятия математики при комплексификации.

Комплексификацией вещественных чисел являются, очевидно, комплексные числа. Здесь дело обстоит очень просто.

В вещественном случае имеется теория Морса. У функции есть критические точки и критические значения. Теория Морса описывает, как меняется множество уровня при переходе через критическое значение. Что же получится при попытке комплексифицировать теорию Морса?

Комплексификацией вещественных функций являются голоморфные (комплексно-аналитические функции). Их множества уровня имеют комплексную коразмерность 1, т. е. вещественную коразмерность 2. В частности, они не разделяют объемлющее многообразие; дополнение к множеству уровня вовсе не будет несвязным.

В вещественном случае множество критических значений функции разделяет вещественную прямую. Поэтому при прохождении через критическое значение топология множества уровня, вообще говоря, меняется. Для комплексно-аналитических функций это не так. Их множества критических значений не разделяют плоскости комплексной переменной. Поэтому в комплексном случае множества уровня функции, отвечающие разным некритическим значениям, топологически устроены одинаково. Но при обходе вокруг критического значения возникает монодромия. Она является отображением множества уровня в себя (определённым с точностью до изотопии).

В вещественном случае дополнение к критическому значению состоит из двух компонент, т. е. его гомотопическая группа $\pi_0$ есть $\mathbb{Z}_2$. В комплексном случае дополнение к критическому значению связно и имеет фундаментальную группу $\mathbb{Z}$. Поэтому комплексификацией $\pi_0$ естественно считать $\pi_1$ и комплексификацией группы $\mathbb{Z}_2$ будем считать группу $\mathbb{Z}$.

Оказывается, эта точка зрения вполне последовательно проходит и дальше. Комплексификацией перестроек Морса (связанных с элементами группы $\pi_0$ множества некритических значений вещественной функции) является монодромия (представление группы $\pi_1$ множества некритических значений комплексной функции).

Описание монодромии даётся теорией Пикара—Лефшеца, которая является теорией ветвления интегралов.[4)] В этом смысле можно считать, что комплексификацией теории Морса является теория Пикара—Лефшеца. Комплексификацией перестройки Морса (приклеивания ручки к множеству уровня) будем считать монодромию в окрестности невырожденной критической точки голоморфной функции. Эта операция — так называемое преобразование Зейферта, т. е. перекручивание цикла на множестве уровня. Она заключается в скручивании цилиндра, при котором одно основание цилиндра остаётся неподвижным, а другое основание совершает полный оборот. В обоих случаях, вещественном и комплексном, простейшая операция соответствует особой точке, задаваемой суммой квадратов.

Можно идти и дальше. В вещественной теории есть классы Штифеля—Уитни со значениями в $\mathbb{Z}_2$. Им при комплексификации отвечают классы Черна со значениями в $\mathbb{Z}$. Всё согласовано: комплексификацией $\mathbb{Z}_2$ действительно является $\mathbb{Z}$.

Комплексификацией проективной прямой $\mathbb{R}P^1 = S^1$ является комплексная проективная прямая $\mathbb{C}P^1 = S^2$ (сфера Римана). Таким образом, сфера Римана является комплексификацией окружности. На этой сфере есть окружность — экватор. Имеется теория рядов Фурье, которые определены на этой окружности, и теория рядов Лорана, у которых имеются два полюса в полюсах сферы.

Выясним, что является комплексификацией края вещественного многообразия. Сначала следует алгебраизировать понятие края. Многообразие с краем задаётся неравенством вида $f(x) \geqslant 0$. Правильной комплексификацией этого неравенства является уравнение $f(x) = y^2$. Это уравнение определяет гиперповерхность в $(x, y)$-пространстве, стандартная проекция которой на $x$-пространство задаёт двулистное разветвлённое накрытие с ветвлением вдоль края. Таким образом, комплексификацией многообразия с краем является двулистное накрытие с ветвлением над комплексным краем.

Эта точка зрения оказалась весьма плодотворной. Трюк с накрытием я придумал в 1970 г., когда занимался 16-й проблемой Гильберта о распо-

[4)] См. *Васильев В. А.* Ветвящиеся интегралы. — М.: МЦНМО, 2000. — *Прим. ред.*

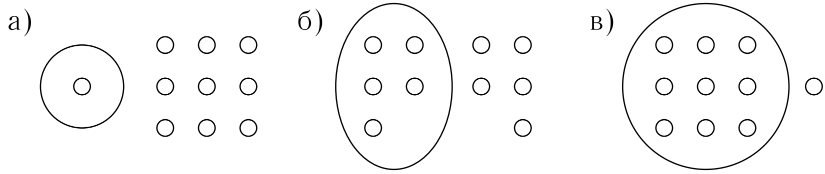



*Рис. 2*

ложении овалов алгебраической кривой заданной степени $n$.[5] Многочлен степени $n$ от двух переменных задаёт на (вещественной проективной) плоскости некоторый набор кривых. Проблема Гильберта состоит в том, чтобы выяснить, какие расположения возможны при данном $n$. В случае, когда $n = 1$ или $n = 2$, эту задачу решили древние греки. Случаи $n = 3$, 4 были исследованы Декартом, Ньютоном. Случай $n = 5$ тоже был исследован, а случай $n = 6$ не поддавался до второй половины этого века, и он уже прямо входил в проблему Гильберта. Эта часть проблемы была решена Д. А. Гудковым.

Для степени 8 вопрос о том, как могут располагаться эти овалы, открыт, даже для кривых общего положения.

Для $n = 6$ ответ на этот вопрос был получен Гудковым. А именно, максимально возможное число овалов равно 11. При этом они могут располагаться на проективной плоскости следующими тремя способами: 10 овалов имеют попарно непересекающиеся внутренности, а одиннадцатый содержит либо один, либо пять, либо девять из них внутри себя (рис. 2, *а*–*в*). (На проективной плоскости дополнение к кругу есть лист Мёбиуса. Поэтому понятие «внутри» корректно определено).

Гудков заметил, что для всех расположений, которые удалось построить, количества овалов удовлетворяют некоторому сравнению по модулю 8. Например, числа 1, 5, 9 идут через 4, а эйлеровы характеристики множеств положительности многочленов, задающих кривые, — через 8. Во всех других примерах тоже выполняются аналогичные сравнения. Это наводит на мысль, что здесь где-то рядом имеются четырёхмерные многообразия, потому что, как хорошо известно, в топологии четырёхмерных многообразий сравнения по модулю 8 и по модулю 16 играют решающую роль.

Надо искать, какое с вещественной алгебраической кривой связано четырёхмерное многообразие. После нескольких месяцев попыток построить подходящее четырёхмерное многообразие по алгебраической кривой я в конце концов сообразил, что нужно взять именно двулистное накрытие до-

[5] Литературу о 16-й проблеме Гильберта см. в кн.: *Гильберт Д.* Избранные труды. Т. II. — М.: Факториал, 1998. С. 584. — *Прим. ред.*

полнения алгебраической кривой. Применяя к этому четырёхмерному многообразию мощные методы четырёхмерной топологии удалось доказать, что вещественные алгебраические кривые, которые не удовлетворяют сравнениям Гудкова, не существуют (Арнольд, 1971[6]; Рохлин, 1972[7]). Возникшая таким образом новая область математики — вещественная алгебраическая геометрия — далеко развита позже Виро, Харламовым, Никулиным, Шустиным, Хованским и многими другими.

Для $n = 8$ доказанные ограничения оставляют примерно 90 вариантов возможного расположения двадцати двух овалов. А построено примерно 80 расположений.

Эта же самая конструкция срабатывает и в теории особенностей. Сначала были построены особенности для схем Коксетера с простыми связями, т. е. для серий $A$, $D$, $E$. Остаются ещё $B$, $C$, $F$, $G$, у которых встречаются углы между векторами, отличные от прямых и от 120°. Эти случаи при помощи особенностей не описывались, пока я не сообразил, что нужно использовать ту же самую конструкцию двулистного разветвлённого накрытия. Она позволяет построить особенности каустик и волновых фронтов, которые соответствуют остальным группам Вейля (исключая лишь $G_2$).

Правильность комплексификации подтверждают только полученные результаты. Сама по себе она не имеет никакого априорного определения. Положение здесь похоже на то, что было в прошлом веке, когда обнаружили, что теория интегральных уравнений параллельна теории линейных операторов. А ещё раньше такая же ситуация была с проективной двойственностью. До тех пор, пока не был прояснён смысл проективной двойственности, она оставалась загадочной. Точно так же, до тех пор пока не был построен функциональный анализ, оставалось загадочным сходство теорем об интегральных уравнениях и аналогичных теорем линейной алгебры — вплоть до теорем Фредгольма.

Мы и сейчас имеем похожую ситуацию, когда физики используют операцию суммирования по непрерывному индексу, а математики не желают понимать, что это значит.

Да и раньше математики тысячелетиями использовали вещественные числа, про которые ничего строго не было известно, кроме древнегреческого открытия, что не все они существуют: никакое рациональное число, возведённое в квадрат, не может равняться двойке. А так как в греческой

[6] *Арнольд В. И.* О расположении овалов вещественных плоских алгебраических кривых, инволюциях четырёхмерных гладких многообразий и арифметике целочисленных квадратичных форм // *Арнольд В. И.* Избранное — 60. — М: Фазис, 1997. С. 175–187. — *Прим. ред.*

[7] Работы В. А. Рохлина по вещественной алгебраической геометрии см. в книге: *Рохлин В. А.* Избранные работы. — М.: МЦНМО, 1999. — *Прим. ред.*

математике числа были только рациональные, то, с одной стороны, вещественными числами пользовались, а с другой стороны, не знали, что это такое, потому что не было строгого определения.

Вот примерно такое же положение в том, о чём я собираюсь рассказывать, — в явлении троичности.

## Явление троичности

Явление троичности заключается в том, что ко всем рассмотренным выше парам добавляется ещё третий член. В математике объекты очень часто встречаются тройками. И во многих случаях эти тройки образуют коммутативные диаграммы.

Простейшая тройка получается при добавлении к вещественным и комплексным числам кватернионов $\mathbb{H}$. Но есть и другие тройки.

**Пример 1.** Простые алгебры Ли $E_6$, $E_7$, $E_8$. Им отвечают диаграммы Дынкина, изображённые на рис. 3, *а—в*, соответственно. Каждая вершина

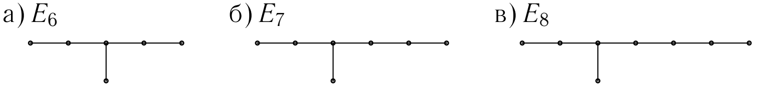



*Рис. 3*

графа соответствует вектору. Если вершины соединены ребром, то угол между соответствующими векторами равен 120°, а если не соединены, то угол равен 90°. Группа, порождённая отражениями относительно ортогональных дополнений к этим базисным векторам в евклидовом пространстве и есть группа Вейля.

**Пример 2.** В канцелярских магазинах продают угольники трёх типов: равносторонний треугольник, прямоугольный треугольник с углом 45 градусов и прямоугольный треугольник с углом 30 градусов. Оказывается, эти треугольники соответствуют диаграммам $E_6$, $E_7$, $E_8$, и образуют вместе с ними коммутативную диаграмму $3 \times 2$.

В этой мистической теории троичностей уже имеется много нетривиальных теорем и подобных коммутативных диаграмм. О том, чтобы на этой лекции всё доказывать, не может быть и речи. Я сначала выпишу примеры, а затем объясню связи между ними.

**Пример 3.** Тетраэдр, октаэдр и икосаэдр. Группы их симметрий суть, соответственно, группы Вейля $A_3$, $B_3$ и $H_3$. Их диаграммы Дынкина (ранее введённые Виттом и Коксетером) изображены на рис. 4.

Числа рёбер соответствующих многогранников равны, соответственно, 6, 12 и 30. Эти числа имеют вид $2 \cdot 3$, $3 \cdot 4$, $5 \cdot 6$. Если из первых чисел в этих

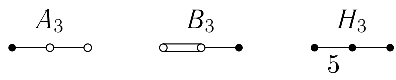



*Рис. 4*

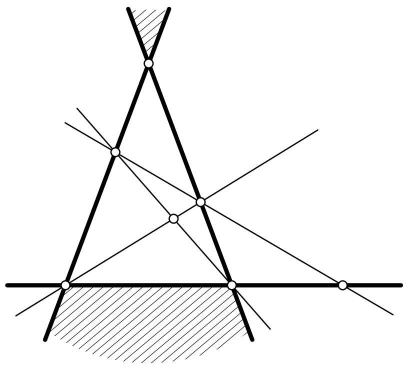


*Рис. 5*

произведениях вычесть по единице, то, как и следовало ожидать, получим 1, 2 и 4 — вещественные размерности $\mathbb{R}$, $\mathbb{C}$ и $\mathbb{H}$.

Опишем группу симметрий тетраэдра. Она порождается отражениями относительно его плоскостей симметрии, число которых равно числу рёбер, т. е. 6. Эти плоскости делят пространство на 24 части, называемые камерами Вейля. Опишем эти камеры. Все плоскости проходят через центр. Поэтому их можно изображать проективными прямыми в $\mathbb{R}P^2$. Эти прямые делят проективную плоскость на 12 частей (см. рис. 5). Возьмём одну из этих частей. Она соответствует трёхгранному углу, который на проективной плоскости изображается в виде треугольника (на рис. 5 треугольник заштрихован). Продолжим стороны этого треугольника (на рис. 5 — жирные линии). В модели в трёхмерном пространстве этим трём прямым соответствуют 3 плоскости, проходящие через начало координат. Они делят пространство на 8 частей. На проективной плоскости им соответствуют 4 части, т. е. 4 больших треугольника, составленных из 12 маленьких треугольников. В больших треугольниках содержится, соответственно, 1, 3, 3 и 5 маленьких треугольников (на рис. 5 как настоящий выглядит только один из больших треугольников, который состоит из пяти маленьких). Поэтому получаем такую формулу:

$$24 = 2(1 + 3 + 3 + 5).$$

Число камер — 24 — есть, конечно, порядок группы симметрий тетраэдра.

Аналогично порядок группы симметрии октаэдра равен

$$48 = 2(1 + 5 + 7 + 11),$$

а для икосаэдра находим

$$120 = 2(1 + 11 + 19 + 29).$$

Заметим, что если слагаемые в этих трёх формулах увеличить на 1, то мы получим системы весов (степеней базисных инвариантов) для $A_4$, $B_4$ и $H_4$, соответственно. Кстати, эти веса — в точности числа вершин, граней и рёбер тетраэдра, октаэдра и икосаэдра (двойка — это, возможно, число трёхмерных граней?)

По многогранникам, отвечающим перечисленным группам Вейля, выписываются многочлены:

| | | |
|---|---|---|
| $E_6 = x^3 + y^4 + z^2$, | $E_7 = x^3 + xy^3 + z^2$, | $E_8 = x^3 + y^5 + z^2$ |
| $\widetilde{E}_6 = x^3 + y^3 + z^3$, | $\widetilde{E}_7 = x^4 + y^4 + z^2$, | $\widetilde{E}_8 = x^3 + y^6 + z^2$ |

Соответствующие особенности называются эллиптическими и параболическими (последние связаны с так называемыми аффинными системами корней). Переход от многогранника к многочлену будет вкратце описан ниже.

Продолжим список троиц.

Характеристические классы:

| | | |
|---|---|---|
| классы Штифеля—Уитни, | классы Черна, | классы Понтрягина. |

Дальше идёт такая строчка-троица:

| | | |
|---|---|---|
| $S^1 \to S^1 = \mathbb{R}P^1$, | $S^3 \to S^2 = \mathbb{C}P^1$, | $S^7 \to S^4 = \mathbb{H}P^1$. |

Первое отображение есть двулистное накрытие края листа Мёбиуса над его базой $S^1$. Второе — расслоение Хопфа со слоем $S^1$. Из этого, кстати сказать, видно, что комплексификация — деликатная операция. Дело в том, что комплексификация $S^1$ в одном случае — $S^3$, а в другом — $S^2$. Это объясняется тем, что в накрытии листа Мёбиуса две окружности совершенно различны. Одна окружность — $\mathbb{R}P^1$, а другая — $SO(2)$. При комплексификации $\mathbb{R}P^1$ превращается в $\mathbb{C}P^1$, а $SO(2)$ превращается в $SU(2) \cong S^3$.

Третье отображение — расслоение Хопфа со слоем $S^3$.

Я выпишу ещё несколько троиц. За следующие две строчки я ещё более или менее отвечаю, а смысл идущих за ними строчек пока ещё не вполне ясен.

| | | |
|---|---|---|
| $S^0$-расслоение, | $S^1$-расслоение, | $SU(2)$-расслоение? |

В первом случае речь идёт просто о двулистном накрытии, во втором — о расслоении на комплексные прямые, в третьем — видимо, о расслоении на кватернионные прямые (снабжённые, быть может, некоторой «гиперсвязностью», комплексифицирующей комплексную связность второго случая).

Им соответствуют:

| | | |
|---|---|---|
| монодромия плоской связности, | кривизна связности (2-форма), | 4-форма (гиперкривизна?) |

Скрученность объектов, обозначенных знаком вопроса, должны измерять какие-то 4-формы, приводящие к характеристическим классам Понтрягина. Они должны измерять какую-то гиперкэлерову несогласованность комплексных структур.

Ещё одна троица:

| | | |
|---|---|---|
| многочлены, | тригонометрические многочлены, | модулярные многочлены. |

Многочлены — это такие голоморфные отображения сферы в сферу, у которых число прообразов бесконечности равно 1. Тригонометрические многочлены — это лорановские многочлены из $\mathbb{C}[t, t^{-1}]$, у которых число прообразов одной из точек равно 2. Модулярные многочлены — это рациональные функции, у которых прообраз одной из точек состоит из трёх точек: $0, 1, \infty$.

Последнюю троицу предложил Гивенталь:

| | | |
|---|---|---|
| гомологии, | $K$-теория, | эллиптические гомологии. |

Как было сказано ранее, многограннику, отвечающему группе Вейля, соответствует многочлен от трёх переменных. Опишем это соответствие на примере икосаэдра, группа симметрий которого состоит из 120 элементов. Симметрии, сохраняющие ориентацию, образуют в $SO(3) \cong \mathbb{R}P^3$ подгруппу порядка 60. Рассмотрим двулистное накрытие группы $SO(3)$ группой $Spin(3) = SU(2) \cong S^3$. Каждому элементу группы $SO(3)$ отвечают две точки из $S^3$. Рассматриваемой подгруппе из 60 точек в $SO(3)$ отвечают 120 точек на $S^3$. Эти точки и образуют бинарную группу икосаэдра.

Бинарная группа икосаэдра содержится в $SU(2)$, поэтому она действует на $\mathbb{C}^2$. Обозначим через $X$ пространство орбит этого действия. Пространство орбит $X$ — двумерная комплексная поверхность с особенностями, вложенная в $\mathbb{C}^3$. Она описывается уравнением $x^3 + y^5 + z^2 = 0$. Это доказывается с помощью теории инвариантов. А именно, мы ищем базисные инварианты. Ими будут бинарные формы с нулями в вершинах, серединах рёбер и центрах граней икосаэдра соответственно. Степени этих бинарных форм равны 12, 20 и 30. Затем ищем соотношение между ними — так называемую сизигию. Эту сизигию нашёл в прошлом веке Шварц. При надлежащей нормировке базисных инвариантов $x$, $y$ и $z$ она принимает вид $x^3 + y^5 + z^2 = 0$.

Поясним теперь, как по формулам (многочленам) можно построить $E_6$, $E_7$ и $E_8$. Вместо того чтобы приравнивать многочлены нулю, приравняем их некоторому $\varepsilon \neq 0$, т. е. рассмотрим поверхность $x^3 + y^5 + z^2 = \varepsilon \neq 0$ — так называемый слой Милнора. Для слоя Милнора можно рассмотреть гомологии средней размерности. Слой Милнора — двумерное комплексное многообразие; его вещественная размерность равна 4, поэтому средняя размерность равна 2. В двумерных гомологиях слоя Милнора действует группа монодромии, которая описывается путями в базе версальной деформации. Когда мы обходим вокруг дискриминанта соответствующего семейства многочленов, у которого такая особая точка при самом плохом значении параметра, то возникает теория Пикара—Лефшеца, которая описывает отображения слоя Милнора на себя. Оказывается, форма пересечения слоя Милнора и есть та самая евклидова структура, в которой базисные циклы задаются диаграммой Дынкина, и соответствующая простейшему обходу вокруг дискриминанта монодромия (преобразование Зейферта) как раз является отражением в соответствующем зеркале. Поэтому группы $E_6$, $E_7$ и $E_8$ — это группы монодромии, которые соответствуют особенностям, задаваемым указанными многочленами.

Наконец, объясним, как из многогранников получаются $\widetilde{E}_6$, $\widetilde{E}_7$ и $\widetilde{E}_8$. Для этого нужно взять простейшее стандартное представление соответствующей группы симметрий многогранника в двумерном пространстве и рассмотреть тензорное произведение какого-нибудь представления на стандартное. Это тензорное произведение разлагается на неприводимые слагаемые, которые входят с какими-то коэффициентами. Матрица этих коэффициентов — это «матрица Картана», описывающая соответствующую аффинную систему корней.

Вот итоговая таблица троиц:

| | | |
|---|---|---|
| $\mathbb{R}$ | $\mathbb{C}$ | $\mathbb{H}$ |
| $E_6$ | $E_7$ | $E_8$ |
| $A_3$ | $B_3$ | $H_3$ |
| $D_4$ | $F_4$ | $H_4$ |
| $x^3 + y^4 + z^2$ | $x^3 + xy^3 + z^2$ | $x^3 + y^5 + z^2$ |
| $x^3 + y^3 + z^3$ | $x^4 + y^4 + z^2$ | $x^3 + y^6 + z^2$ |
| $60°, 60°, 60°$ | $45°, 45°, 90°$ | $30°, 60°, 90°$ |
| тетраэдр | октаэдр | икосаэдр |
| $6 = 2 \cdot 3$ | $12 = 3 \cdot 4$ | $30 = 5 \cdot 6$ |

| | | |
|---|---|---|
| квадратичные формы | эрмитовы формы | гиперэрмитовы формы |
| классы Штифеля—Уитни | классы Черна | классы Понтрягина |
| $S^1 \to S^1 = \mathbb{R}P^1$ | $S^3 \to S^2 = \mathbb{C}P^1$ | $S^7 \to S^4 = \mathbb{H}P^1$ |
| двулистные накрытия | $S^1$-расслоения | $SU(2)$-расслоения |
| монодромия плоской связности | кривизна связности (2-форма) | 4-формы (гиперкривизна?) |
| многочлены | тригонометрические многочлены | модулярные многочлены |
| гомологии | $K$-теория | эллиптические гомологии |

*В. И. Арнольд*

# Принцип топологической экономии в алгебраической геометрии

*Лекция 21 мая 1997 года*

Эта часть лекции не связана с первой частью, так что её можно понимать независимо от предыдущего. Начнём с примера.

**Пример 1.** Рассмотрим в $\mathbb{C}P^n$ два алгебраических многообразия $X$ и $Y$ дополнительных размерностей. В общем положении они пересекаются по конечному множеству точек. Пусть $[X]$ и $[Y]$ — классы гомологий, реализуемые многообразиями $X$ и $Y$, $[X]\circ[Y]$ — индекс пересечения этих классов (являющийся целым числом). Он равен числу «положительных» точек пересечения $X$ с $Y$ минус число «отрицательных» пересечений. Таким образом, общее число точек пересечения $\#(X\cap Y)$ не меньше индекса пересечения $[X]\circ[Y]$ (и имеет ту же чётность). *Теорема Безу* утверждает, что $\#(X\cap Y)$ равно числу $[X]\circ[Y]$, никакого неравенства нет! Здесь дело в том, что у комплексных многообразий ориентация такова, что каждое пересечение даёт вклад $+1$, а не $-1$, в суммарный индекс пересечения. Отрицательные пересечения «неэкономны», они увеличивают число точек пересечения $X$ с $Y$ по сравнению с «топологически необходимым» их количеством. Кстати, отсюда же следует, что многочлен степени $n$ имеет ровно $n$ корней, а не больше.

Этот известный и следующие (более новые) примеры приводят к «принципу экономии», с помощью которого, в свою очередь, можно получать дальнейшие гипотезы. Эти гипотезы можно проверять в частных случаях и иногда можно доказать — тогда они становятся теоремами. Но в большинстве случаев они надолго остаются гипотезами, т. е. утверждениями, которые можно опровергать.

**Пример 2.** Следующее утверждение долго являлось гипотезой. Рассмотрим в $\mathbb{C}P^2$ риманову поверхность $X$ рода $g$, заданную неприводимым многочленом степени $n$. Её класс гомологий $[X]\subset H_2(\mathbb{C}P^2,\mathbb{Z})$ есть $n$ раз взятая образующая $[\mathbb{C}P^1]$ группы $H_2(\mathbb{C}P^2,\mathbb{Z})=\mathbb{Z}$; это число $n$ равно степени многочлена. Известно, что род $g$ поверхности $X$ можно найти по формуле Римана (некоторые, впрочем, говорят, что Риман не знал, что такое род). А именно,

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

Возникает вопрос, нельзя ли класс гомологий $[X]$ реализовать другой гладкой поверхностью, пусть не комплексной, а вещественной, но меньшего рода? Другими словами, реализует ли комплексное алгебраическое многообразие минимальный род для данного класса гомологий $[X]$? *Гипотеза Тома* утверждает, что комплексная поверхность с $g=(n-1)(n-2)/2$ ручками — это действительно самая экономная реализация класса $n\cdot[\mathbb{C}P^1]\subset H_2(\mathbb{C}P^2,\mathbb{Z})$. Любая гладкая вещественная поверхность, реализующая этот класс, имеет столько же ручек или больше. Доказательство было получено лишь недавно Кронхаймером (Kronheimer) и Мровкой (Mrowka) при помощи «тяжёлой артиллерии», а именно, теории Дональдсона, Громова, Виттена и т. д., восходящей к идеям квантовой теории поля.

**Пример 3** (гипотеза Милнора). Эта гипотеза также доказана недавно теми же авторами. Если задан произвольный узел, то всегда можно за несколько развязываний превратить его в незаузленную окружность. Развязывание — это замена прохода на переход (прохождения сверху ни прохождение снизу) на подходящей диаграмме (т. е. проекции на плоскость) данного узла. Выполняется развязывание при помощи меча; тем же методом Александр Македонский развязывал гордиев узел. Гордиево число узла — это минимальное число развязываний, необходимое для превращения заданного узла в тривиальный.

Узлы связаны с особенностями следующим образом. Рассмотрим алгебраическую кривую $K^2$ с особенностью в начале координат и маленькую сферу $S^3\subset\mathbb{C}^2$ с центром в нуле. Пересечение $N^1=K^2\cap S^3$ является узлом (или зацеплением) в $S^3$. Например, полукубическая особенность $x^2=y^3$ кривой в $\mathbb{C}^2$ соответствует трилистниковому узлу.[1] Милнор занимался вопросом, как получить гордиево число узла $N$ из алгебраических свойств многочлена, задающего кривую $K$. Он предложил следующий способ развязывания: задаём кривую $K$ параметрически, в данном случае

$$x=t^3,\quad y=t^2.$$

В общем случае подобное задание $x=t^n$, $y=f(t)$, где $f(t)$ — голоморфная функция, даётся теорией рядов Пюизо (которая была открыта Ньютоном). Приведём функции $x(t)$ и $y(t)$ в общее положение, добавив и слагаемые более низких степеней, например

$$x=t^3-\varepsilon t,\quad y=t^2.$$

Получится кривая $K'$ без особых точек (во всяком случае, вблизи начала координат), исключая некоторое число $\delta$ точек самопересечения. Это число $\delta$ (число Милнора) можно выразить через некоторые алгебраические

[1] Доказательство этого факта имеется, например, в § 1 книги: *Милнор Дж.* Особые точки комплексных гиперповерхностей. — М.: Мир, 1971.

инварианты кривой $K$. Оно является «кандидатом» на гордиево число. Дело в том, что существует способ выполнять развязывания, «соответствующие» двойным точкам: сколько двойных точек, столько и развязываний. Гипотеза Милнора состоит в том, что этот «алгебраический» способ развязывания узла особенности — наиболее экономный; за меньшее, чем $\delta$, число развязываний узел нельзя превратить в тривиальный. Это ещё одно проявление принципа экономии.

В полученном недавно доказательстве этой гипотезы применена та же «тяжёлая артиллерия» (происходящая, по существу, из квантовой физики), что и при доказательстве гипотезы Тома.

**Пример 4** (теорема Мёбиуса). Возможно, к рассмотрению листа Мёбиуса Мёбиус пришёл, занимаясь следующей задачей. Рассмотрим проективную прямую $\mathbb{R}P^1$ на проективной плоскости $\mathbb{R}P^2$. Это бесконечно вырожденная кривая: кривизна в каждой точке равна нулю. При малом возмущении возникнет кривая общего положения с конечным числом точек перегиба. Каково их наименьшее число?

Попробуем не угадывать ответ путём экспериментов, а воспользуемся принципом экономии. В соответствии с этим принципом, следует рассмотреть наиболее простую алгебраическую модель интересующего нас явления, в данном случае — возникновения точек перегиба при возмущении прямой. Вероятно, эта модель будет наиболее экономной, т. е.будет, в данном случае, содержать наименьшее возможное число точек перегиба.

Итак, надо продеформировать прямую в классе алгебраических кривых как можно более низкой степени и подсчитать количество возникших точек перегиба. Тогда по принципу топологической экономии меньшее число точек перегиба получить не удастся. Обойтись кривыми 1-й степени нельзя, так как это прямые. Кривые второй степени нам также не подходят, так как имеют совсем другую топологию: прямая $\mathbb{R}P^1$ не стягиваема в $\mathbb{R}P^2$, а окружность стягиваема (как и другие квадрики).

Кривых третьей степени уже достаточно. Например, можно прямую в аффинных координатах $x, y$ задать уравнением $y = 0$, а её возмущение — уравнением $y \cdot f(x, y) = \varepsilon$, где $\varepsilon$ — малое число, а $f(x, y)$ — многочлен второй степени, не имеющий вещественных корней. Кривые третьей степени можно исследовать, и получается, что наиболее экономной является кривая $y(1+x^2) = 1$, или $y = 1/(1+x^2)$. У неё три точки перегиба: две с абсциссами $\pm 1/\sqrt{3}$, а третья — на бесконечности. Действительно, кривая, не имеющая перегиба на бесконечности, расположена в окрестности бесконечно удалённой точки по одну сторону от касательной в этой точке, а в аффинной карте — с двух сторон от асимптоты, как гипербола. Наша же кривая при уходе на бесконечность подходит к асимптоте $y = 0$ оба раза сверху, т. е. в бесконечно удалённой точке происходит пересечение кривой и её касательной; это возможно только в точке перегиба.

Стоит заметить, что понятие точки перегиба чисто проективное. Оно не требует никакой метрики, а основано только на кратности пересечения кривой с её касательной в данной точке. Действительно, кратность пересечения двух кривых сохраняется при произвольном диффеоморфизме, а при проективных преобразованиях прямые переходят снова в прямые.

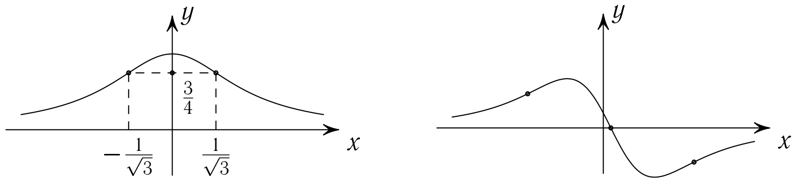



*Рис. 1*

На рис. 1 слева две конечных точки перегиба и одна — на бесконечности; справа — все три точки перегиба конечные. Во всяком случае, их общее число нечётно, что следует из неориентируемости листа Мёбиуса (окрестность $\mathbb{R}P^1$ в $\mathbb{R}P^2$ — это как раз лист Мёбиуса).

Теорема Мёбиуса утверждает, что *число точек перегиба нечётно и не меньше трёх*, *т. е. не может быть равно* 1.

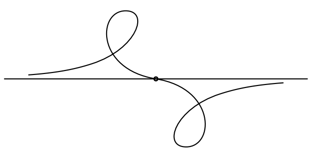


*Рис. 2*

В связи с этим сформулирую ещё одну теорему и одну гипотезу. *Если прямую деформировать сильно*, *то число точек перегиба всё равно не меньше трёх*, *пока она остается вложенной*. Продолжая деформацию далее, можно получить иммерсированную (уже не вложенную) кривую (рис. 2), у которой только одна точка перегиба. Гипотеза утверждает, что *для получения* (*регулярной гомотопией*) *кривой с одной точкой перегиба из вложенной кривой с тремя точками перегиба необходимо пройти через изменение лежандрова типа узла*, *т. е. через момент*, *когда кривая касается сама себя и ориентации двух ветвей в точке касания совпадают*.

В этот момент произойдёт самопересечение соответствующего нашей кривой лежандрова узла в пространстве контактных элементов плоскости. Итак, эта гипотеза (аналогичная гипотезе Чеканова о квазифункциях, обобщающей гипотезы о неподвижных точках и о лагранжевых пересечениях, с которых я начинал предыдущую лекцию; см. с. 6) утверждает, что число точек перегиба не может стать равным 1 до изменения лежандрова типа соответствующего узла.

На сегодняшний день доказано (Д. Панов), что нельзя уничтожить пару точек перегиба из трёх (т. е. оставить только одну) не меняя лежандров узел, если не создать как минимум девять точек перегиба на некоторой кривой в процессе гомотопии.

**Пример 5.** Для малых деформаций прямой теорема Мёбиуса является частным случаем теоремы Штурма, дающей оценку числа нулей тригонометрического ряда Фурье. Пусть $f$ — $2\pi$-периодическая функция, разложение которой в ряд Фурье начинается с гармоник $n$-го порядка:

$$f(x) = \sum_{k \geqslant n} (a_k \cos kx + b_k \sin kx).$$

Каково может быть наименьшее число нулей $f$ на одном периоде? В соответствии с общим принципом, следует взять самый простой случай, т. е.ряд, состоящий только из одной гармоники: $f(x) = \cos kx$ (или $f(x) = \sin kx$, что ничего не изменит). У этой функции имеется $2k$ нулей. Теорема Штурма (доказанная Гурвицем) утверждает, что *число нулей ряда Фурье не меньше, чем у его первой ненулевой гармоники*.

Не доказано аналогичное утверждение для интеграла Фурье: пусть

$$F(x) = \int f(k) e^{-ikx}\, dk,$$

где $f(-k) = \bar{f}(k)$, так что $F$ — вещественная функция. Предположим, что $f(k) = 0$ при $|k| \leqslant \omega$. Гипотеза Гриневича состоит в том, что *среднее число нулей $F(x)$ на длинном отрезке не меньше, чем у «первой гармоники» $\sin \omega x$, т. е. не меньше $\pi/\omega$*. Это весьма правдоподобное утверждение, иллюстрирующее общий принцип экономии, пока не доказано.

Объясню, почему теорема Мёбиуса является (при малых деформациях) следствием теоремы Штурма. Рассмотрим проективную плоскость с вложенной в неё продеформированной прямой. На сфере (двулистно накрывающей $\mathbb{R}P^2$) получим экватор, деформация которого описывается нечётной функцией $f(\varphi)$: $f(\varphi + \pi) = -f(\varphi)$, где $\varphi$ — долгота, а $f(\varphi)$ — возмущение широты в точке $\varphi$. Вычисления показывают, что точки перегиба возмущённого экватора соответствуют значениям $\varphi$, при которых

$$f''(\varphi) + f(\varphi) = 0. \qquad (*)$$

На самом деле это верно только в первом приближении; но вопрос о числе точек перегиба всё равно сводится к вопросу о числе корней уравнения вида $(*)$, только с другой функцией вместо $f$: $\tilde{f}''(\varphi) + \tilde{f}(\varphi) = 0$. Из нечётности $f$ выводится, что функция $\tilde{f}$ также нечётна, т. е. её разложение в ряд Фурье содержит только гармоники с нечётными номерами: первую, третью, пятую и т. д.Тогда ряд Фурье функции $L\tilde{f} = \tilde{f}'' + \tilde{f}$ начинается не раньше, чем с третьей гармоники, поскольку $(\sin \varphi)'' + \sin \varphi = 0$ и $(\cos \varphi)'' + \cos \varphi = 0$:

дифференциальный оператор

$$L = \frac{d^2}{d\varphi^2} + 1$$

уничтожает $\sin\varphi$ и $\cos\varphi$. Следовательно, число корней уравнения $L\tilde{f} = 0$ на отрезке $[0, 2\pi]$ не меньше шести, а число точек перегиба возмущённой прямой на $\mathbb{R}P^2$ вдвое меньше и, таким образом, не меньше трёх.

Как и теорема Мёбиуса, теорема Штурма (обобщающая её) должна иметь (и действительно имеет) обобщения на случай больших деформаций, когда и функций-то никаких нет, а рассматриваются вариации некоторых кривых в многомерных пространствах. Возникает теория выпуклых кривых в проективном пространстве, содержащая много интересных задач, (см., например, *Арнольд В. И.* Топологические проблемы теории распространения волн // Успехи мат. наук **51** (1996), №1. С. 3—50 и цитированную там литературу).

**Пример 6.** Попытаемся найти обобщение теоремы Мёбиуса на случай поверхностей и, соответственно, теоремы Штурма на функции двух переменных. Рассмотрим проективную плоскость $\mathbb{R}P^2$, вложенную в проективное пространство $\mathbb{R}P^3$. Рассмотрим малую деформацию этого вложения. Сколько «перегибов» будет у получившейся поверхности?

Сначала надо объяснить, что же такое «перегиб». Точки поверхности делятся на эллиптические, параболические и гиперболические, в зависимости от второй квадратичной формы. В эллиптической точке поверхность выпукла, в гиперболической две главные кривизны имеют разные знаки, а в параболической точке одна из двух главных кривизн равна нулю. Здесь можно обойтись без второй квадратичной формы, метрики и т. д.: эллиптические, гиперболические и параболические точки определяются проективно инвариантно. Достаточно рассмотреть пару из поверхности и касательной плоскости в данной точке. Расстояние от близкой точки поверхности до этой касательной плоскости даёт квадратичную форму, вырожденность (невырожденность) и сигнатура которой не меняются при проективных преобразованиях.

Итак, параболические точки (играющие в данном случае роль точек перегиба) образуют вполне определённые кривые на поверхностях проективного пространства. Сколько же областей эллиптичности (ограниченных параболическими кривыми) возникает при бесконечно малой деформации плоскости? Этот вопрос не решён. Однако наш метод немедленно даёт гипотезу о том, что это число всегда не меньше, чем в случае самой простой алгебраической деформации. Соответствующая алгебраическая деформация была исследована Б. Сегре в 1942 году. Оказалось, что на кубической поверхности в $\mathbb{R}P^3$, являющейся деформацией $\mathbb{R}P^2$ (т. е. не имеющей ручек; поверхности с ручками Сегре тоже исследовал) имеется 4 кривые параболических точек. Ф. Аикарди (F. Aicardi) из Триеста предположила, что параболических кривых всегда не меньше 4, и выполнила много компьютерных экспериментов, подтвердивших эту гипотезу. Гипотеза Аикарди состоит в том, что ситуация аналогична теореме Штурма: имея младшую гармонику с $2k$ нулями (кубическую деформацию с 4 линиями перегиба) и добавляя гармоники (возмущения) более высокой степени, пусть даже и с большими коэффициентами, число нулей (кривых перегиба) уменьшить нельзя.

Эта гипотеза имеет множество частных случаев. В некоторых из них её удалось доказать, тогда как к доказательству самой гипотезы не видно никаких подходов. Существует довольно много разных доказательств теорем Мёбиуса и Штурма, но все они имеют непонятное происхождение. Непонятно также, как обобщить эти доказательства на многомерный случай. Приведённая гипотеза — вероятно, простейшее многомерное обобщение утверждений этих теорем.

**Замечание 1.** На поверхности общего положения есть ещё более специальные, «перегибные», точки, чем параболические. Дело в том, что в каждой гиперболической точке имеются два асимптотических направления; прямые этих направлений имеют необычно высокий ($\geqslant 2$) порядок касания с данной поверхностью. Они являются также касательными к кривым, по которым поверхность пересекается с касательной плоскостью в данной гиперболической точке. Итак, в гиперболической области имеется поле крестиков. Попутно возникает вопрос о том, какие существуют глобальные топологические ограничения на эти поля крестиков: какие из них можно реализовать поверхностями, какие — нет. Этот же вопрос можно задать и об аффинных поверхностях, и даже о поверхностях — графиках функции $z = f(x, y)$. Эта проблема весьма мало исследована. Расскажу об одном имеющемся здесь инварианте.

При подходе к параболической кривой асимптотические направления сближаются, и поле крестиков в гиперболической области переходит в поле (асимптотических) направлений на параболической кривой. В некоторых точках оно касается самой кривой. Эти точки ещё более специальные: они доставляют вершины ласточкиных хвостов двойственной поверхности. Сегре установил, что на кубической поверхности имеется ровно 6 таких точек. Отсюда вторая гипотеза Аикарди: при любых (не обязательно кубических) деформациях проективной плоскости возникает не менее 6 точек касания асимптотического направления с параболической кривой.[2)]

---

[2)] В начале сентября 1997 г. Д. Панов сообщил, что он опроверг обе гипотезы Аикарди, построив поверхность с одной единственной параболической кривой без специальных точек. По-видимому, двойственная поверхность (фронт) довольно сложна, так что остаётся возможность обобщить теорему Мёбиуса, доказав, что его поверхность отделена от кубических каким-либо подходящим дискриминантом в дополнение к очевидно необходимому дискриминанту рождения ласточкиных хвостов на двойственной поверхности.

**Замечание 2.** Рассмотрим поверхность $M$, образованную асимптотическими направлениями исходной поверхности проективного пространства в пространстве всех прямых, касающихся исходной поверхности. Топологически она является дублем гиперболической области $N$. На ней есть поле направлений, гладкое и однозначное (с особенностями лишь в специальных точках, где асимптотическое направление касается параболической кривой), которое проектируется в поле крестиков на гиперболической области. У каждой точки внутри $N$ в $M$ имеется два прообраза, в которых проектирование $M \to N$ регулярно, а у точки на краю $N$ есть один прообраз в $M$, являющийся точкой складки. Поверхность $M$ имеет столько ручек, сколько областей эллиптичности содержится внутри $N$. На $M$ имеется гладкая динамическая система, интегральные кривые которой (они называются асимптотическими линиями) касаются данного поля направлений; при проектировании $M \to N$ (гладкие) асимптотические линии переходят в кривые с полукубическими особыми точками на параболической кривой $\partial N$. О свойствах этой динамической системы известно очень мало. Известно, что она имеет (в случае кубической поверхности) 27 периодических решений (в комплексной области) — прямых на кубической поверхности, являющихся, очевидно, её асимптотическими линиями. С этой динамической системой связано много интересных вопросов. Какими свойствами она обладает? Является ли она, например, интегрируемой? Что получится при её исследовании обычными средствами теории возмущений вблизи периодических решений? Каковы свойства отображения Пуанкаре, или отображения последования, которое переводит кривую параболических точек в себя вдоль интегральных кривых? Будет ли эта система системой общего положения, будет ли в ней наблюдаться хаос?

В заключение я покажу, каким образом из нашей задачи можно получать более простые. Надо придумать способ деформации плоскости в пространстве, при котором мы можем контролировать кривизну возникающей поверхности и следить за параболическими точками. Таких способов есть несколько; каждый из них приводит к алгоритму, дающему набор параболических кривых, а иногда и точек касания параболической кривой с асимптотическим направлением, по каким-то исходным комбинаторным данным.

Один из способов такой: возьмём триангуляцию исходной плоскости и чуть-чуть сдвинем каждую вершину с плоскости вверх или вниз. Получим многогранную поверхность, являющуюся деформацией плоскости. Её, конечно, требуется сгладить каким-либо способом, позволяющим проследить, где будут появляться параболические точки. К теореме Мёбиуса этот способ применить нетрудно, поскольку понятно, где будут рёбра перегиба у аппроксимирующей нашу кривую ломаной. Для многогранников тоже можно определить параболические кривые; можно определить их на самом многограннике, а можно на (некотором «стандартном») его сглаживании. Затем надо построить (комбинаторно!) по данному возмущению триангуляции некоторое число, которое должно оказаться не меньше, чем 4, если гипотеза Аикарди верна.

Другой способ основан на рассмотрении уравнения с частными производными и приводит (подробности я опускаю) к следующей гипотезе. Рассмотрим функции на сфере $S^2$ в $\mathbb{R}^3$, но не все, а только нечётные, т. е. обладающие свойством $f(-x) = -f(x)$, где $-x$ — точка, диаметрально противоположная $x$. Именно такие функции описывают деформации проективной плоскости. Теперь рассмотрим сферические функции, т. е. функции на сфере, являющиеся собственными для сферического оператора Лапласа. Так, $x, y, z$ — собственные функции с собственным числом $-2$, и других решений уравнение $\Delta f + 2f = 0$ не имеет, его пространство решений трехмерно. Итак, сферические функции степени 1 — это в точности все линейные однородные функции от $x$, $y$, $z$; они все нечётны. Однако оказывается, что существует другая теория сферических функций, почему-то не рассмотренная классиками. Можно рассмотреть сферические функции с особенностями, например, с полюсами кратности не выше некоторого числа. Тогда уравнение $\Delta f + 2f = 0$ имеет, помимо обычных сферических функций, ещё и обобщённые решения. Точнее, в правой части уравнения при этом оказывается не нуль, а какая-то линейная комбинация $\delta$-функций (и их производных) с носителями в тех точках, где у функции $f$ полюса. Оказывается, почти все точки поверхности, для которой деформация определена такой функцией, будут гиперболическими, и лишь при сглаживании вокруг полюсов возникнут кривые параболических точек, ограничивающие области эллиптичности. Кроме того, в некоторых отдельных точках вторая квадратичная форма такой поверхности обращается в нуль полностью. Поверхность вблизи такой точки имеет, как правило, вид обезьяньего седла. Такой подход приводит к следующей гипотезе: *число особых точек + число обезьяньих сёдел* $\geqslant 8$. Здесь $8 = 2\cdot 4$, поскольку вместо $\mathbb{R}P^2$ рассматриваем сферу $S^2$. Мы учитываем обезьяньи сёдла потому, что могут быть точки, в которых функция $f$ слишком хорошо приближается линейной функцией, т. е. точки, где второй дифференциал $d^2 f$ не просто вырожден, а обращается в нуль.

Обращение $d^2 f$ в нуль встречается обычно в изолированных точках. Простейший пример — $f(x, y) = x^3 - 3xy^2$; нетрудно подсчитать, что все точки этой поверхности (кроме нуля) гиперболические; эллиптических точек нет, и лишь в точке $(0, 0)$ форма $d^2 f$ становится нулевой.

Вообще, можно рассмотреть (в фиксированной локальной карте) отображение

$$(x, y) \longmapsto \begin{pmatrix} A & B \\ B & C \end{pmatrix} \qquad (**)$$

где

$$A = \frac{\partial^2 f}{\partial x^2}, \quad B = \frac{\partial^2 f}{\partial x\, \partial y} = \frac{\partial^2 f}{\partial y\, \partial x}, \quad C = \frac{\partial^2 f}{\partial y^2}.$$

Итак, с функцией $f(x, y)$ можно связать поверхность в трёхмерном пространстве с координатами $A, B, C$, являющемся симметрическим квадратом кокасательного пространства. Мы получили, так сказать, второе гауссово отображение.

В этом пространстве имеется конус вырожденных форм, он задаётся уравнением $AC = B^2$. Внутри него расположены положительно и отрицательно определённые формы, снаружи — гиперболические формы. Уравнение Лапласа $A + C = 0$ (и некоторые другие уравнения тоже) означает, что вся поверхность-образ при рассматриваемом отображении лежит в гиперболической области, пополненной точкой 0 (для уравнения Лапласа — в плоскости $A + C = 0$, пересекающей наш конус только в начале координат). Поэтому на поверхности-прообразе эллиптических точек нет, возможно только обнуление $d^2 f$ в некоторых точках. Если же пошевелить функцию $f$, то образ соответствующей поверхности тоже пошевелится и появятся параболические кривые (соответствующие пересечениям поверхности-образа с конусом). Поэтому обезьяньи сёдла надо учитывать наряду с полюсами: из них при малой деформации тоже возникают параболические кривые. Отсюда и получается последняя гипотеза.

Наконец, я приведу ещё одну гипотезу (М. Эрма́н), которая укладывается в рамки той же теории об экстремальности алгебраических объектов. Правда, в данном случае место алгебраического объекта занимает версальная деформация. Но известно, что в теории особенностей версальные деформации «ведут себя так же, как алгебраические объекты». Почему-то Эрман сформулировал эту гипотезу о каустиках лагранжева, или градиентного отображения. Но тот же вопрос остаётся, вероятно, открытым, если вместо градиентного отображения взять произвольное отображение из $\mathbb{R}^n$ в $\mathbb{R}^n$.

Итак, пусть $f\colon M^n \to M^n$ — отображение общего положения, или же $f\colon L \to \mathbb{R}^n$ — проекция лагранжева подмногообразия симплектического пространства $L \subset \mathbb{R}^{2n}$. Типичный пример: $y = \operatorname{grad} f(x)$, где $x \in \mathbb{R}^n$, $f\colon \mathbb{R}^n \to \mathbb{R}$. (К такому виду подходящим выбором координат приводится любое лагранжево отображение.)

Рассмотрим критические точки этого отображения, т. е. точки, где $\det(\partial y/\partial x) = 0$, и соответствующие критические значения $y$; в симплектическом случае множество критических значений образует каустику.

*Степень отображения в точке* $y$ определим следующим образом. Рассмотрим близкое к $y$ регулярное значение $y'$ и посчитаем (как обычно, с учётом ориентации) число прообразов $y'$, но не всех, а только близких к $x$.

Определённая таким образом степень не зависит от выбора точки $y'$. Предположим, что она равна нулю. Гипотеза утверждает, что *в этом случае точку* $y'$ *можно выбрать так, что у неё не будет* (*в окрестности* $x$) *вообще ни одного прообраза*. Если же степень равна $k$, то гипотеза утверждает, что *найдётся точка, имеющая ровно* $k$ *прообразов, а не больше*.

*Ю. И. Манин*

# Рациональные кривые, эллиптические кривые и уравнение Пенлеве

*Лекция 1 октября 1997 года*

Я хочу рассказать одну историю, которая, как мне кажется, может быть поучительна в разных отношениях. Это, с одной стороны, попытка решить совсем свежую задачу. С другой стороны, при этом выясняется, что пользоваться нужно очень классическими вещами, которые были сделаны в начале века, в первой трети, потом интерес к ним как-то пропал. Они были забыты, потом снова возобновились по совершенно другим причинам, и так далее.

Это история, в которой попытка решения задачи является полуудачей. Начав с каким-то удовольствием работать — бежать, бежать по горячему следу, как гончая, ты до чего-то добегаешь, но тем не менее, задача, махнув хвостиком, скрывается где-то за поворотом. Так что есть, что делать дальше. Я очень надеюсь, что тем, кто не сможет или не захочет следить за всей математикой, будет интересно проследить за побочными вещами. А из тех, кто захочет следить за математикой, будет кое-кому интересно подумать и над самой задачей.

Задача, о которой я говорю, на техническом языке относится к попытке понять, как устроен потенциал квантовых когомологий проективной плоскости. Что это такое, я сейчас просто напишу в явном виде. Проективная плоскость, скажем, над комплексным полем, производит трёхмерное пространство своих когомологий, которое натянуто на классы когомологий всей большой клетки, гиперплоского сечения и точки. В этом трёхмерном пространстве есть координаты, которые я обозначу $x, y, z$:

$$H^*(\mathbb{P}^2, \mathbb{C}) = \{x\Delta_0 + y\Delta_1 + z\Delta_2\}.$$

Теория квантовых когомологий производит один формальный ряд (квантовый потенциал этих когомологий), который устроен следующим образом:

$$\Phi(x, y, z) = \frac{1}{2}(xy^2 + x^2z) + \sum_{d=1}^{\infty} N(d)\frac{z^{3d-1}}{(3d-1)!}e^{dy},$$

где $N(d)$ — количество рациональных кривых степени $d$ на плоскости, которые проходят через $3d-1$ точку в общем положении. $N(1)$ — количество прямых, проходящих через две точки, т. е. 1. $N(2)$ — количество коник, проходящих через 5 точек; это тоже 1. Попробуйте проверить, что $N(3)=7$; это уже немножко сложнее. То, что у этой задачи число решений конечно, легко проверяется счётом констант: кривая задаётся параметрами, потом на эти параметры накладываются линейные условия (прохождение через данные точки). Не так уж это просто сделать, тем не менее, чуть-чуть алгебраической геометрии позволяет провести это рассуждение правильно.

Поразительно, что эти числа $N(d)$ удовлетворяют совсем нетривиальным билинейным соотношениям, которые позволяют последовательно выразить их одно через другое.

Если из $\Phi(x,y,z)$ убрать классический остаточек $\frac{1}{2}(xy^2+x^2z)$, смысл которого я объяснял вчера на лекции в Московском Математическом обществе, получим функцию

$$\varphi(y,z)=\sum_{d=1}^{\infty} N(d)\frac{z^{3d-1}}{(3d-1)!}e^{dy},$$

которая от $x$ уже не зависит. Теорема, доказанная Максимом Концевичем несколько лет назад, и из обдумывания которой вышла наша с ним работа по квантовым когомологиям, выглядит следующим образом.

**Теорема 1** (Концевич). *Функция $\varphi$ удовлетворяет дифференциальному уравнению*

$$\varphi_{zzz}=\varphi_{yyz}^2-\varphi_{yyy}\varphi_{yzz}.$$

Это одно уравнение эквивалентно всем уравнениям ассоциативности для функции $\Phi(x,y,z)$. (Уравнения ассоциативности состоят в том, что третьи частные производные функции $\Phi(x,y,z)$ являются структурными константами некоторой ассоциативной алгебры.)

Это же уравнение эквивалентно следующей рекуррентной формуле:

$$N(d)=\sum_{k+l=d} N(k)N(l)k^2l\left[l\binom{3d-4}{3k-2}-k\binom{3d-4}{3k-1}\right]$$

при $d\geqslant 2$ и $N(1)=1$.

Если вы забудете про приведённую выше интерпретацию чисел $N(d)$, то эквивалентность трёх формулировок легко проверяется прямым вычислением. Но доказательство того, что так определённые числа $N(d)$ удовлетворяют рекуррентному соотношению, — это факт из алгебраической геометрии, который требует довольно крупной возни. Хотя, если разрешить себе махать руками, его можно доказать довольно правдоподобным образом.

Задача, которую я имею в виду и которая дразнит меня до сих пор, заключается в том, чтобы понять аналитическое поведение функции $\varphi$. Формальным рядом она задаётся в окрестности нуля. Несложно показать, что у неё есть ненулевая область сходимости. А дальше непонятно: у неё появляются особенности, но неясно, какого они типа. И вообще, непонятно, что это за функция. Может быть, это что-то знакомое: решение гипергеометрического уравнения, или дзета-функция, или модулярная функция. Скорее всего, ни то, ни другое, ни третье. Из попыток понять, что это за функция, как раз и возникла та работа, о которой я хочу рассказать.

Дифференциальное уравнение для функции $\varphi$, хотя и простое, но оно какого-то неклассического вида. Было, однако, известно, что это дифференциальное уравнение можно довольно заковыристыми нелинейными заменами переменных свести к одному из классических уравнений, которые называются «Пенлеве-шесть». Эти уравнения образуют четырехпараметрическое семейство, зависящее от параметров $\alpha, \beta, \gamma, \delta$. Для них имеется стандартное обозначение $\mathrm{PVI}_{\alpha,\beta,\gamma,\delta}$, которое держится уже почти 100 лет:

$$\frac{d^2X}{dt^2} = \frac{1}{2}\left(\frac{1}{X} + \frac{1}{X-1} + \frac{1}{X-t}\right)\left(\frac{dX}{dt}\right)^2 - \left(\frac{1}{t} + \frac{1}{t-1} + \frac{1}{X-t}\right)\frac{dX}{dt} + \\ + \frac{X(X-1)(X-t)}{t^2(t-1)^2}\left(\alpha + \beta\frac{t}{X^2} + \gamma\frac{t-1}{(X-1)^2} + \delta\frac{t(t-1)}{(X-t)^2}\right).$$

Я пропущу здесь объяснение того, каким образом и почему исходное уравнение сводится к PVI, а посвящу оставшееся время попытке понять, что же из этого следует для исходной загадочной функции.

В начале века Пенлеве занялся такой задачей классификации. Рассмотрим дифференциальное уравнение, для начала, скажем, такого вида:

$$\frac{dy}{dt} = F(y, t)$$

с начальным условием $y(t_0) = c$. При этом функция $F$ комплексно аналитическая. Вы начинаете аналитически продолжать решение и смотрите, где вы наткнётесь на особенность. Особенность может быть разных типов: может быть обыкновенный полюс, может быть точка ветвления, а может быть существенно особая точка. Положение этой особенности, вообще говоря, зависит от того, какую начальную константу $c$ вы подставите. Здесь бывают разные ситуации: некоторые особенности могут двигаться, а некоторые могут быть неподвижными. Вопрос был такой. (Смысл его мне до сих пор не очень понятен, но, видимо, была какая-то внутренняя логика развития анализа к тому времени, когда этот вопрос был поставлен.) Хочется понять, какие бывают дифференциальные уравнения, у которых двигаться могут только полюсы, а всё существенное (существенно особые точки и точки ветвления) от $c$ не зависит. Оказалось, что при малых порядках уравнения (при первом и даже при втором) ответ на этот вопрос приводит к классификации: можно написать не слишком большое семейство уравнений, к которым всё сводится заменами. С первым порядком всё просто, ответ был давно известен. А со вторым порядком долго мучился Пенлеве и его ученики. И когда Пенлеве написал свою окончательную работу по этому вопросу, он PVI пропустил. Вычисления были жуткие, он сделал ошибку в вычислениях. Не позже чем через год ошибку обнаружили. Первым правильно написал PVI Гамбье (B. Gambier, 1906). Вторым был Фукс, который сделал это независимо от Гамбье и совсем из других соображений.

Пропустить PVI чрезвычайно обидно, потому что в некотором смысле слова начальная часть списка состоит из классических уравнений. И вот впервые появляется неклассическое уравнение. В этом списке неклассических уравнений PVI самое общее, остальные получаются в некотором роде специализацией и предельными переходами. Пенлеве из-за своей ошибки пропустил не специальное уравнение, а общее, самое общее уравнение. Но это может случиться с каждым, и я вообще призываю вас не относиться сурово к ошибкам в хороших работах. Сам факт, что работа поставила правильный вопрос, дала полуправильный ответ и стимулировала дальнейшие исследования, много важнее, чем то обстоятельство, что автор допустил ту или иную ошибку. Все мы люди, и находящий какую-то поддержку современный американский обычай, что человеком, доказавшим теорему, считается тот, кто исправил последнюю ошибку в чужом доказательстве, я чрезвычайно не одобряю. Это неправильно и ведёт к неправильному отношению к математике, как мне кажется.

Итак, вот у нас есть уравнение Пенлеве-шесть, написанное на самом деле Гамбье и Фуксом (сыном знаменитого Фукса; его статья, в каком-то смысле, посвящена работам отца, и он о них отзывается очень уважительно). Что более интересно в статье Фукса (которую я читал с огромным удовольствием и которую я должен был прочесть 20 лет назад, но просто не знал о её существовании), так это то, что он получил это семейство уравнений совсем не таким способом, как Пенлеве. Способ Фукса мне намного ближе. Точнее говоря, Фукс получил эти уравнения двумя способами. Первый способ — это так называемые изомонодромные деформации линейных дифференциальных уравнений, но не будем на этом останавливаться. Другой способ связан с тем, что он обнаружил, что это семейство уравнений можно записать в более приятном виде и очень геометрически.[1)]

Глядя на уравнение PVI, если вы когда-нибудь работали с эллиптическими кривыми, вы сразу поймёте, что здесь должна играть какую-то роль кривая

$$Y^2 = X(X-1)(X-t).$$

Но не вполне понятно, какую именно роль: уравнение выглядит довольно

[1)] Об уравнениях Пенлеве и изомонодромных деформациях см. книгу: *Iwasaki K., Kimusa H., Shimomusa Sh., Yoshida M.* From Gauss to Penlevé. A Modern Theory of Spectral Functions. — Vieweg Verlag, 1991. — *Прим. ред.*

противно. И вот что сделал Фукс, так это то, что он записал это уравнение в совершенно замечательном виде, который я как только увидел, так сразу почти что подпрыгнул.

Прежде чем продолжать историю PVI, я хочу напомнить, с чего мы начали. У нас имеется четырёхпараметрическое семейство уравнений плюс ещё два параметра, проистекающие из начальных условий. Итого у нас имеется шестипараметрическое семейство функций, среди которых содержится функция, описывающая квантовый потенциал $\mathbb{P}^2$, с которой я начал. Что бы мы хотели сделать, чтобы описать эту функцию? Мы хотели бы вычислить, какие константы относятся к квантовому потенциалу и какие начальные условия. А уж потом, когда мы всё это вычислим, тогда можно посмотреть, что говорит уравнение Пенлеве про функцию с такими константами и такими начальными условиями. К сожалению, априори должна быть надежда только на какую-то удачу. И это ещё один интересный сюжет, связанный с историей PVI. Много сил было потрачено на точную формулировку и доказательство утверждения, которое Пенлеве формулировал так: «Почти все решения этой новой системы уравнений являются неклассическими функциями.» Это трудно сформулировать, трудно доказать и трудно поверить в то, что это утверждение полезно. Действительно, нужно дать точное определение того, что такое классическая функция. Вы определяете итеративный процесс: те функции, которые у вас уже были, разрешается брать в качестве коэффициентов нового линейного дифференциального уравнения; решения этого уравнения можно добавить. Потом их ещё разрешается брать в качестве коэффициентов алгебраического уравнения, и решения этого уравнения тоже разрешается добавить. Но вы никогда не знаете, не пропустили ли вы какую-нибудь важную операцию. Например, мне кажется, что при формулировке и доказательстве этой теоремы пропустили важнейшую операцию — взятие обратной функции. Но так или иначе, общая вера состоит в том, что большинство трансцендентностей Пенлеве являются какими-то новыми функциями. Тем не менее, не исключено, что та самая одна-единственная функция, которая нас интересует, классической всё-таки является. Потому что среди решений PVI известно огромное число классических функций. Это создаёт некоторый азарт. Нас интересует одна точка в шестимерном пространстве. Мы хотим узнать её аналитический смысл и не знаем окончательного ответа. Это именно то, что я имел в виду, когда говорил от том, что бежишь, вроде бы, по правильному следу, но задачка махнула хвостиком и скрылась за поворотом.

Теперь я возвращаюсь к теореме Фукса (1907). Рассмотрим интеграл

$$\int_{\infty}^{(X,Y)} \frac{dx}{\sqrt{x(x-1)(x-t)}}.$$

Он определён лишь с точностью до интеграла по замкнутому контуру на эллиптической кривой, отвечающей заданному значению $t$, т. е.определён с точностью до периода. Периоды эллиптической кривой удовлетворяют линейному дифференциальному уравнению, поэтому неопределённость интеграла можно снять, применив дифференциальный оператор. Давайте так и поступим и напишем уравнение

$$t(t-1)\left[t(t-1)\frac{d^2}{dt^2}+(1-2t)\frac{d}{dt}-\frac{1}{4}\right]\int_{\infty}^{(X,Y)}\frac{dx}{\sqrt{x(x-1)(x-t)}}= \\ =\alpha Y+\beta\frac{tY}{x^2}+\gamma\frac{(t-1)Y}{(x-1)^2}+\left(\delta-\frac{1}{2}\right)\frac{t(t-1)Y}{(x-t)^2},$$

где $Y^2=X(X-1)(X-t)$. Это уравнение эквивалентно предыдущей записи PVI.

Эта запись лучше потому, что левая часть приобретает теперь совершенно прозрачный алгебро-геометрический смысл, и известен очень широкий контекст, в котором такие уравнения можно обобщать — в отличие от уравнений Пенлеве, которые вкладываются естественным образом в совершенно другой контекст.

Вот эта штука уводит задачу совершенно в другую область — в область алгебро-геометрическую, что приятно: исходная задача была алгебро-геометрической, мы вложили её в какие-то нелинейные диффуры, а потом снова вернули в алгебраическую геометрию.

Я люблю смотреть на это уравнение как на неоднородное уравнение Пикара—Фукса. Уравнением Пикара—Фукса называется уравнение следующего типа. У вас есть семейство торов, или кривых, или алгебраических многообразий, и вы берёте их периоды по замкнутым циклам, которые зависят от параметра. Эти периоды как функции от параметра удовлетворяют однородным линейным дифференциальным уравнениям, в правой части которых стоит нуль. Оказывается, что если в правой части поставить не нуль, а то, что написано выше, то получится уравнение Пенлеве.

То, что я вслед за этим сделал, нужно было сделать много-много лет тому назад. Я не понимаю, почему этот сюжет не развивался дальше. Если вы получили такую эллиптическую кривую, то совершенно очевидно, что можно пересчитать это уравнение на другие геометрические представления того же самого объекта. И самое естественное геометрическое представление того же самого объекта, конечно же, такое. Вы заменяете плоскость $t$ на верхнюю полуплоскость $\tau$, интерпретируя точку верхней полуплоскости как образующую решётки периодов тора. Над точкой $\tau$ тогда растёт тор $\mathbb{C}/\langle 1,\tau\rangle$. Дифференциал превращается просто в дифференциал $dz$. Интеграл становится просто равным $z$. Поэтому, не производя никакого пересчёта, мы знаем, что написанный выше дифференциальный оператор превращается в

$d^2/d\tau^2$. Дело в том, что тот оператор был второго порядка и аннулировал два периода. У нас периоды 1 и $\tau$, поэтому аннулирующий их оператор и есть $d^2/d\tau^2$. Поэтому вся левая часть превращается в $d^2/d\tau^2$ — видите, как резко она сокращается. Правая часть не вполне тривиальна. Для того чтобы написать ответ, нужно воспользоваться серией классических формул из теории эллиптических функций. Но давайте я сразу напишу ответ, он совсем не сложен:

$$\frac{d^2z}{d\tau^2} = \frac{1}{(2\pi i)^2} \sum_{j=0}^{3} \alpha_j \wp_z \left( z + \frac{T_j}{2},\ \tau \right).$$

Здесь $T_j$ — периоды $(0, 1, \tau, 1+\tau)$; константы $\alpha_j$ — те же самые, но только правильно нормированные: $(\alpha_j) = (\alpha, -\beta, \gamma, 1/2 - \delta)$; $\wp_z$ — производная по $z$ функции Вейерштрасса

$$\wp(z, \tau) = \frac{1}{z^2} + \sum_{m,n}{}' \left( \frac{1}{(z + m\tau + n)^2} - \frac{1}{(m\tau + n)^2} \right).$$

Функция Вейерштрасса — это простейший ряд, который можно соорудить из $z$ и $\tau$, чтобы он допускал 1 и $\tau$ в качестве периодов.

Такой вид уравнения Пенлеве я лично уже в состоянии запомнить; он достаточно прост. Почему этот вид уравнения Пенлеве не был известен до моей работы — для меня абсолютная и полная загадка. Или, скорее, доказательство того, что интерес к уравнению Пенлеве по каким-то причинам надолго увял, а когда он возродился, люди не удосужились посмотреть на классические работы по PVI.

Чем приятна эта переформулировка, так это тем, что она позволяет сразу кое-что увидеть. Не будем забывать, что нас интересуют решения PVI, и даже если мы поверим в то, что большинство из них неклассические, то вдруг наше окажется классическим. И вообще, давайте посмотрим, есть ли классические решения. Одно из них сразу видно: если все $\alpha_j$ равны нулю, то мы получим уравнение $d^2z/d\tau^2 = 0$; его решения — линейные функции. Между прочим, при первоначальной записи уравнения этого решения не видно. На самом деле, этим простым замечанием можно воспользоваться абсолютно нетривиальным способом, по причинам, которые скрыты очень глубоко. Об этом я скажу позже.

**Следствие 1.** *Для* $(\alpha, \beta, \gamma, \delta) = (0, 0, 0, 1/2)$ *все решения пишутся в классических функциях.*

Те из вас, кто не хотят думать геометрически, должны просто представлять себе, что переход от $X$, $Y$, $t$ к $z$, $\tau$ — это просто подстановка, нелинейная замена переменных. Но, конечно, гораздо продуктивнее смотреть на это геометрически.

Прежде чем двигаться дальше, я сформулирую ещё одно следствие. Классические функции Вейерштрасса удовлетворяют многим тождествам, среди которых есть так называемые тождества Ландена. Эти тождества связывают функции Вейерштрасса от аргументов $z$ и $\tau$ и от аргументов $z$ (с каким-то сдвигом на полупериод) и $2\tau$. Довольно понятно, что если заменить $\tau$ на $2\tau$, то мы перейдём к функциям, которые периодичны относительно несколько меньшей решётки (подрешётки индекса 2). Усреднив такие функции по полупериодам, можно снова получить функцию, которая периодична относительно первой решётки. Это и приводит к тождествам Ландена, которые я выписывать не буду. Но если посмотреть, выдерживает ли уравнение Пенлеве переход к подрешётке индекса 2, мы придём к тому, что есть неожиданные симметрии.

**Следствие 2.** *Если у вас есть уравнение Пенлеве с константами* $(\alpha_0, \alpha_1, \alpha_0, \alpha_1)$, *то оно выдерживает преобразование Ландена, и у вас получаются формулы пересчёта от решений уравнения с этими константами к решениям уравнения с константами* $(4\alpha_0, 4\alpha_1, 0, 0)$.

Посчитав константы, соответствующие уравнению, отвечающему за квантовые когомологии, получим, что оно имеет совершенно очаровательный вид

$$\frac{d^2 z}{d\tau^2} = -\frac{1}{2\pi^2}\wp_z(z, \tau).$$

Те из вас, кто занимался симплектической геометрией и гамильтоновой механикой, глядя на это уравнение, без труда убедятся, что оно гамильтоново.

**Следствие 3.** *Уравнение* PVI *гамильтоново:*

$$\frac{dz}{d\tau} = \frac{\partial H}{\partial y}, \qquad \frac{dy}{d\tau} = -\frac{\partial H}{\partial z},$$

*где*

$$H = \frac{y^2}{2} - \frac{1}{(2\pi i)^2}\sum \alpha_j \wp\left(z + \frac{T_j}{2}, \tau\right).$$

Гамильтониан существенно зависит от $\tau$ (параметр $\tau$ играет роль комплексного времени).

То, что уравнение PVI гамильтоново, было в классике известно. Однако формулы были чудовищные, и совершенно непонятен их геометрический смысл. Это представление делает очевидным, что уравнение гамильтоново, и оно же позволяет разобрать его геометрический смысл. Точные формулировки я здесь опущу.

Основной общий результат о шестипараметрическом семействе решений уравнения Пенлеве — это в высшей степени загадочная симметрия, к описанию которой я сейчас перейду.

Вся надежда на симметрии состоит в том, что сколько-то классических решений мы знаем, и группа симметрий позволит их размножать. Удивительно, что группа симметрий здесь очень большая, и она абсолютно не очевидная. Эти симметрии многократно открывались и переоткрывались. Но до самого последнего времени их невозможно было понять и написать в каком-то приличном виде. К именам Шлезингера и Окамото я хочу добавить Аринкина и Лысенко. Это два ученика Дринфельда, которые написали очень приятную работу, отчасти проясняющую, что здесь происходит. Давайте я сформулирую ответ.

Я должен ввести новые параметры $a_i^2 = 2\alpha_i$. На пространстве с координатами $(a_i)$ действует группа симметрий $W$, порождённая следующими преобразованиями:

а) $(a_i) \mapsto (\pm a_i)$;

б) перестановки $a_i$;

в) $(a_i) \mapsto (a_i + n_i)$, где $n_i \in \mathbb{Z}$ и $\sum_{i=0}^{3n_i}$ чётна.

Главная серия симметрий — в). Эта группа симметрий поднимается до действия во всей структуре и, в частности, переводит решения уравнения Пенлеве с одними параметрами в решения с другими параметрами.

Самое главное открытие здесь принадлежит Шлезингеру (1924). Шлезингер вложил эту задачу в серию других задач, которые теперь называются уравнениями Шлезингера, или теорией изомонодромных деформаций линейных дифференциальных уравнений. Он заложил основы этой теории, и он же обнаружил огромное число дискретных преобразований от одних уравнений к другим. Частным случаем этих преобразований является и описанная выше группа симметрий. Были потом разные физические работы, которые описали следствия из всех этих симметрий. Этим много занималась японская школа. Окамото обнаружил некоторое подмножество, на котором эти преобразования образуют группу (преобразования Шлезингера образуют лишь полугруппу и они перемешивают одни уравнения с другими — это довольно непроглядная вещь). А Окамото обнаружил, что на PVI преобразования образуют группу. Аринкин и Лысенко, о которых я только что упоминал, в каком-то смысле вернулись к идеологии Шлезингера для того частного случая уравнения на $\mathbb{P}^1$ с четырьмя особыми точками, к которому сводится уравнение Пенлеве. Они показали, что в некотором разумном смысле слова эта конструкция даёт полную группу бирациональных автоморфизмов правильно определённого геометрического объекта.

**Следствие 4.** *Все* PVI *с* $(a_i) \in \mathbb{Z}^4$ *с чётной суммой* $a_i$ *имеют полностью классические решения.*

Действительно, это нам известно в случае $a_i = 0$.

Уравнением Пенлеве в последнее время много занимался Хитчин. В одной из его работ доказывается, что для констант $(0, 0, 0, 2)$ решения полностью классические, и приводятся явные формулы. Он не заметил, что эта точка всего лишь одна из точек бесконечной орбиты.

Добравшись до этого места своих занятий, я воспарил духом и решил, что конечно же не может быть, чтобы уравнение для квантового потенциала $\mathbb{P}^2$ не попало в эту сеть. Оно действительно почти попадает — оно лежит на середине дороги от одного классического решения к другому: в нашем случае получаются параметры $(a_i) = (0, 0, 0, 1)$, что лежит посредине между точками $(0, 0, 0, 0)$ и $(0, 0, 0, 2)$, про которые известно, что для них решения классические. А про середину дороги мы не знаем совершенно ничего. Может случиться, например, что у таких точек есть не двухпараметрическое семейство классических решений, а однопараметрическое семейство.

Вероятно, в этом месте пора сказать про недостающие два параметра. Четыре параметра — это место уравнения в иерархии, а ещё два параметра — это начальные условия для той одной конкретной функции, которая меня интересует.

Мне вычислять начальные условия оказалось удобнее для самой вырожденной из всех эллиптических кривых с комплексным умножением: $\tau_0 = e^{2\pi i/3}$. Интересно, что начальное условие получилось такое: $z(\tau_0)$ — точка третьего порядка на эллиптической кривой.

На этом приходится остановиться.[2)]

Кстати, эту работу про уравнение Пенлеве-шесть я долго хотел назвать, цитируя Владимира Игоревича Арнольда, «Об одном свойстве одного решения одного дифференциального уравнения» (он применяет этот термин, как известно, к работам, которые ничего не стоят и которые нужно выбросить в мусорный ящик). Я не рискнул назвать так, сам не знаю почему.

**Примечание (14 января 1998).** Уже после этой лекции я получил по электронной почте ссылку на заметку Пенлеве 1906 года, где он выводит форму своего уравнения с $\wp$-функцией Вейерштрасса. Я вздохнул с облегчением. Значит, она была открыта вовремя и просто забыта, а сто́ящие мысли, даже если их забывают, обязательно всплывают вновь.

---

[2)] Заинтересованный читатель может обратиться к книге: *Manin Yu. I.* Frobenius Manifolds, Quantum Cohomology and Moduli Spaces. — AMS, 1999. — *Прим. ред.*

*А. А. Кириллов*

# Метод орбит и конечные группы

*Лекции 28 декабря 1997 года и 3 января 1998 года*

## Часть 1

Честно говоря, я несколько смущен тем, что назвал бы overqualified присутствующей аудитории. Потому что мне были обещаны студенты второго и отчасти первого курса. Но будем надеяться, что хотя бы записи этой лекции попадут по адресу.

Я начну с не относящейся к лекции темы, а именно, что происходит нового в математике. Это отдельная тема, которой можно было бы занять все 2 часа. Но две вещи мне все же хотелось бы упомянуть. Об одной вещи я скажу совсем коротко: это последняя статья Максима Концевича, которая закрыла тему деформационного квантования и которая сейчас бурно обсуждается во всех математических центрах.[1] А вторая — это тривиальное доказательство так называемой космологической теоремы Конвея. Эта тема специально предназначена для студентов первого курса. Конвей — это совершенно нетривиальный математик, который занимается совершенно неожиданными вещами. Например, ему пришло в голову исследовать так называемый audioactive operator — оператор произнесения. А именно, представьте себе, что вы, как чукча, который едет по тундре, и что видит, о том и поет песню. И вот такой человек видит число, для простоты единицу. Он видит — одна единица, и пишет: «Одна единица (11)». Тем самым получается второй член нашей последовательности. Он видит второй член последовательности — две единицы, и пишет: «Две единицы (21)». Это третий член нашей последовательности. Читая его вслух, получаем: «Одна двойка, одна единица (1211)». Следующий член: «Одна единица, одна двойка, две единицы (111221)». Так возникает последовательность

$$1,\ 11,\ 21,\ 1211,\ 111221,\ 312211,\ \ldots$$

[1] *Kontsevich M.* Formality conjecture. Deformation theory and symplectic geometry. // Math. Phys. Stud. **20**. — Dordrecht: Kluwer Acad. Publ., 1996. P. 139–156. — *Прим. ред.*

Конвей заинтересовался, какие у этой последовательности асимптотические свойства. Довольно простое упражнение — доказать, что в этой последовательности не встретится никаких цифр, кроме 1, 2 и 3. Следующее нетрудное упражнение — если в первом и втором члене нет никаких цифр, кроме 1, 2 и 3, то и дальше никаких других цифр не встретится.

Оказывается, что у оператора произнесения есть ровно одна неподвижная точка: 22. Все остальные последовательности начинают расти, но весьма нерегулярным образом: иногда длина сильно увеличивается, а иногда не очень. Конвей доказал, что для всех последовательностей, состоящих из цифр 1, 2 и 3 (за исключением неподвижной точки 22), отношение длины слова $a_n$ к длине слова $a_{n-1}$ стремится к определенному пределу, который равен 1,301577269 ... Этот предел — алгебраическое число, и специалисты по динамическим системам, конечно, догадаются, что это — максимальное собственное значение некоторого оператора, но подробности я опускаю.

Это не относится к теме лекции, но я специально потратил несколько минут, чтобы все, кто хотел прийти, пришли бы и сели.

Теперь о теме сегодняшней лекции. Я собирался коротко рассказать о некоторых задачах, возникающих в связи с применением метода орбит к конечным группам. Поскольку лекция была рассчитана на способных первокурсников, то у меня принцип такой: я иногда произношу слова, которые слушатель не обязан понимать в данный момент, но может спросить у старшего товарища, или сообразить сам, или прочитать в книжке, и тем самым ликвидировать локальное непонимание. Идейных непониманий не должно быть, потому что каждые несколько минут будет сменяться, как говорит Юрий Иванович Манин, парадигма, и поэтому можно забыть всё, что было раньше, и начинать внимать сначала.

В целом речь идет о теории представлений. Я полагаю, что первокурсник, обучающийся в Независимом университете, представляет себе, что такое группа, что такое линейное пространство, что такое линейный оператор, что такое линейное представление этой группы (т. е., выражаясь более ученым языком, гомоморфизм группы в группу обратимых линейных операторов), и понимает, что полезно заниматься этой наукой — теория представлений имеет много разных приложений, о которых я сейчас говорить не буду.

Мой вклад в теорию представлений состоит в том, что я предложил так называемый метод орбит. Этот метод является в какой-то степени соединением двух вещей: симплектической геометрии и теории представлений. Или, если подниматься на более высокую степень абстракции, это — соединение современной матфизики с математикой. Дело в том, что матфизика уже постепенно вытесняет математику из ее канонически устоявшихся областей и придумывает новые подходы, новые задачи, попутно решая старые. И уже примерно половина математических журналов постепенно наполняются статьями, в которых матфизика играет ту или иную роль, иногда решающую.

Я потрачу несколько минут на то, чтобы объяснить, что такое симплектическая геометрия, хотя мне и говорили, что эта тема здесь в какой-то степени уже известна. Начну с того, что рассмотрю гладкое многообразие $M$. Гладкое многообразие — это то, на чем можно развивать математический анализ. А именно, это множество $M$, которое локально устроено как евклидово пространство. Это значит, что множество $M$ можно покрыть открытыми областями[2)] $U_i$ так, чтобы для каждой области существовало взаимно однозначное отображение на некоторую область в евклидовом пространстве: $U_i \xrightarrow{\varphi} V_i \subset \mathbb{R}^n$. Тогда все, что мы знаем из математического анализа, а именно, правила обращения с функциями многих переменных, дифференцирование, интегрирование, подстановки, дифференциальные уравнения и прочее, можно перенести на многообразие $M$.

Простейший пример многообразия — это окружность $S^1$. На ней нельзя ввести одной координаты. Это опять же задача для первокурсников: доказать, что любая непрерывная функция на окружности принимает по крайней мере одно значение в двух разных точках, и тем самым не может служить координатой. С другой стороны, нетрудно сообразить, что на окружности можно выделить две области, на каждой из которых такая координата существует. На всей окружности, кроме верхней точки, можно ввести координату, проектируя окружность из верхней точки на горизонтальную прямую. При этом у самой верхней точки образа нет, но можно воспользоваться другой проекцией, из нижней точки. Тогда все, кроме нижней точки, можно отобразить на ту же горизонтальную прямую. Получаем две карты, которые в совокупности покрывают наше многообразие, нашу окружность. Назовем эти локальные координаты $x_+$ и $x_-$. Для окружности радиуса 1 они связаны соотношением $x_+x_- = 1$ ($x_+x_- = r^2$ для окружности радиуса $r$).

Пока наш основной объект — многообразие, т. е. множество, на котором можно заниматься математическим анализом. Кроме того я сказал слово «группа». У группы можно перемножать ее элементы. Обычно группы возникают как группы преобразований чего-либо. Если мы объединим две структуры — группы и многообразия, то получится так называемая группа Ли. Это будет наш основной объект.

Группа Ли — это такой объект, который одновременно является многообразием и группой. Это значит, что на нем одновременно существуют локальные координаты и закон умножения, обладающий всеми положенными свойствами: ассоциативность, существование обратного элемента, существование единицы.

---

[2)] Когда я говорю об открытых областях, я подразумеваю, что $M$ — не просто множество, а топологическое пространство, для которого определены открытые и замкнутые множества.

Та же окружность является прекрасным примером группы. Закон умножения проще всего ввести так. Представим себе, что окружность имеет радиус 1 и лежит не просто на вещественной плоскости, а на комплексной плоскости. Тогда ее можно задать уравнением $|z| = 1$ и ввести групповой закон в виде обычного умножения комплексных чисел: $z_1 z_2$. Если два числа имели модуль 1, то их произведение тоже имеет модуль 1.

Я забыл сказать, что две структуры — группы и многообразия — должны быть естественным образом связаны между собой. А именно, на многообразии есть понятие гладкой функции. При определении многообразия я намеренно забыл указать, что соответствия $U_i \xrightarrow{\varphi} V_i$ произвольные, но там, где возникают две системы координат, мы должны потребовать, чтобы переход от одной системы координат к другой совершался с помощью гладких функций. Гладкость означает существование нескольких (одной, двух, трех и так далее) непрерывных производных. Как правило, удобно сразу об этом не заботиться, а считать, что все функции бесконечно дифференцируемы. Так вот, согласование аксиом группы с аксиомами многообразия очень просто: групповой закон и отображение, которое каждому элементу сопоставляет обратный элемент, должны быть гладкими функциями.

**Упражнение.** Записать групповой закон для окружности в координатах $x_+$ и $x_-$ и убедиться, что он является непрерывной функцией.

Дальше мы будем заниматься теорией представлений. Я буду рассматривать только унитарные представления. Это означает, что группа $G$ отображается в группу $\mathrm{aut}(H)$, где $H$ — гильбертово пространство. Гильбертово пространство — это обобщение линейного пространства в двух направлениях. А именно, размерность пространства не ограничивается и может быть бесконечной. И во-вторых, как правило, вместо вещественного поля рассматривается комплексное поле. Можно рассматривать и вещественные представления, но они сводятся к рассмотрению комплексных, а многие факты и теоремы над комплексным полем выглядят намного проще. Еще в гильбертовом пространстве есть наследуемое из конечномерного случая понятие скалярного произведения двух векторов. Оно позволяет ввести норму вектора и понятие ортогональности векторов. Под автоморфизмами $H$ я понимаю совокупность операторов, которые сохраняют все структуры в $H$: структуру линейного пространства и скалярное произведение. Такие операторы называют унитарными.

Итак, каждому элементу $g$ группы $G$ сопоставляется некоторый унитарный оператор $U(g)$ так, чтобы выполнялось функциональное уравнение $U(g_1 g_2) = U(g_1)U(g_2)$: произведение элементов группы переходит в произведение операторов. Есть довольно развитая наука о том, как обращаться с унитарными представлениями. В частности, есть естественные понятия эквивалентности представлений, разложения их в прямые суммы. Те представления, которые не допускают разложения в виде прямой суммы, называют неразложимыми. Они же оказываются в этом случае и неприводимыми, т. е. не имеют нетривиальных подпредставлений. И первая задача, возникающая при рассмотрении каждой группы, — это описание совокупности всех ее неприводимых унитарных представлений с точностью до эквивалентности. Множество классов эквивалентности неприводимых унитарных представлений группы $G$ обозначают $\widehat{G}$. В английской мат-физике есть очень удобный термин, сокращающий это длинное название, а именно: unirrep (UNItary IRREducible REPresentation). Эти unirrep'ы и есть наш основной объект исследования, и для каждой группы хорошо бы иметь список таких unirrep'ов с указанием их свойств: как они ведут себя при ограничении на подгруппу, или при индуцировании (что это такое, я пока не буду говорить) на бóльшую группу, каковы их матричные элементы, каковы их характеры, инфинитезимальные характеры, и прочее.

Возникает столько же задач, сколько имеется групп. А поскольку группы встречаются в любой области математики и ее приложений, то и эта задача возникает везде. Но методы решения этой задачи совершенно разные для разных областей, и иногда одно решение не похоже на другое. Скажем, для компактных групп есть один набор теорем, для полупростых групп другой набор теорем, для нильпотентных третий набор теорем, и т. д.

Метод орбит, о котором я буду говорить, имеет то преимущество, что он рассматривает все группы сразу и дает универсальный ответ, как описать множество неприводимых представлений и как ответить на большинство вопросов, с ним связанных. Подробно описывать этот метод я не буду.[3] Скажу лишь, какие понятия с ним связаны и что нужно делать. Наибольшую популярность из всех моих увлекательных дел получил так называемый User's Guide — руководство к использованию. Когда вы покупаете электрический утюг или что-то еще, к нему прилагается инструкция, как им пользоваться. Вот существует и такая инструкция, как пользоваться методом орбит. Что вам нужно — и как получить ответ: какие вещи нужно определить, какие вещи перемножить, какие сложить, чтобы получился ответ.

Я перечислю основные ингредиенты, а потом перейду к основной теме моей лекции — как применить метод орбит к конечным группам. Для начинающих математиков всегда приятнее иметь дело с чем-нибудь легко обозримым и конечным. Я, когда был первокурсником, считал, что могу решить любую задачу, в которой идет речь о конечном наборе объектов. Оказалось, что это не совсем так — есть вполне конечные задачи, которые люди до сих пор решить не могут. Некоторые такие задачи я позже еще попробую сформулировать. А сейчас я закончу описание метода орбит и скажу, каким образом его можно приложить к конечным группам.

[3] См. *Кириллов А. А.* Элементы теории представлений. — М.: Наука, 1972. — *Прим. ред.*

Следующим нашим объектом будет алгебра Ли $\mathfrak{g} = Lie(G)$. Буква $\mathfrak{g}$, маленькая готическая буква $g$, — теперь уже общепринятое обозначение для алгебры Ли, связанной с группой Ли $G$. Теория групп Ли и теория представлений претерпели долгие изменения, и у каждой эпохи были свои обозначения. Сейчас уже более или менее устоялся обычай обозначать группу Ли большой латинской буквой, а алгебру Ли — маленькой готической. Хотя есть отсталые ученые, которые используют устаревшие обозначения, и наоборот — передовые ученые, которые выдумывают свои новые обозначения. Но я буду придерживаться наиболее общепринятых обозначений.

Алгебра Ли — это то, что остается от группы, если мы рассматриваем инфинитезимальную окрестность единицы. А именно, возьмем касательное пространство $T_eG$ к группе Ли $G$ в единице $e$, и будем интерпретировать векторы этого касательного пространства как точки, бесконечно близкие к единице. Что остается от группового закона, если мы рассматриваем только точки, очень близкие к единице? Если мы введем локальные координаты с началом координат в единице, то каждый элемент группы изобразится вектором. Возьмем два таких вектора $\vec{x} = (x_1, \ldots, x_n)$ и $\vec{y} = (y_1, \ldots, y_n)$ и рассмотрим их произведение как элементов группы Ли $\vec{x} \cdot \vec{y} = \vec{z}$. Координаты полученного вектора $\vec{z}$ будут функциями от $\vec{x}$ и $\vec{y}$:

$$z_k = \varphi_k(\vec{x}, \vec{y}).$$

Оказывается, что если учесть ассоциативность умножения и то, что умножение любого элемента на единицу дает тот же самый элемент, то функция $\varphi_k$ становится совсем не произвольной, а при подходящем выборе локальных координат имеет следующий замечательный вид:

$$\varphi_k(\vec{x}, \vec{y}) = \vec{x} + \vec{y} + [\vec{x}, \vec{y}] + \ldots$$

(Многоточием обозначены члены третьего порядка и более высоких порядков.) Эта формула означает, что на всякой группе Ли групповой закон в первом приближении коммутативен, а во втором приближении задается билинейным антисимметричным выражением $[\cdot, \cdot]$, которое удовлетворяет тождеству Якоби. Эта билинейная операция превращает касательное пространство к группе в единице в так называемую алгебру Ли. С каждой группой Ли связана алгебра Ли.

Если в алгебре Ли выбрать базис $X_1, \ldots, X_n$, то для того чтобы задать структуру алгебры Ли, достаточно задать произведение базисных векторов: $[X_i, X_j] = c_{ij}^k X_k$. Здесь я использую стандартное эйнштейновское правило, что если в некотором выражении один индекс встречается сверху и снизу, то по этому индексу подразумевается суммирование. Таким образом, алгебра Ли — это просто набор структурных констант $c_{ij}^k$. И великое открытие Софуса Ли состояло в том, что в этом наборе структурных констант заключена вся информация о группе Ли. Всё, что вы захотите о ней узнать, можно, в принципе, извлечь из этого набора структурных констант. Чуть позже я скажу, как это сказывается на тех задачах, которыми мы будем заниматься.

Геометрически нам понадобится не само пространство $\mathfrak{g}$, а двойственное к нему пространство, которое обозначается $\mathfrak{g}^*$. Это — пространство линейных функционалов на пространстве $\mathfrak{g}$. Если в исходном пространстве был базис $X_1, \ldots, X_n$, то можно ввести двойственные функционалы $F^1, \ldots, F^n$ так, что каждый функционал на своем векторе равен 1, на прочих базисных векторах равен 0, а дальше распространяется по линейности.

Пространство $\mathfrak{g}^*$, в отличие алгебры Ли $\mathfrak{g}$, никакого умножения в себе не содержит — функционалы перемножать нельзя. Правда, можно их насильно заставить умножаться, и тогда мы получим очень интересную конструкцию — так называемую биалгебру. Она связана с квантовыми группами, алгебрами Дринфельда и прочим. Это отдельная наука, которой я здесь касаться не буду. Нас интересует это пространство как линейное пространство.

И последний ингредиент из общей теории групп и алгебр Ли, который мне здесь понадобится, — это действие группы Ли $G$ на алгебре Ли $\mathfrak{g}$ и двойственное действие на пространстве $\mathfrak{g}^*$. Это действие можно определять многими эквивалентными способами. Проще всего убедиться, что такое действие существует, можно так. Для любого $x \in G$ определим преобразование $A(x)\colon g \mapsto xgx^{-1}$. Это преобразование является автоморфизмом группы. Кроме того, это — гладкое преобразование многообразия $G$, оставляющее неподвижной точку $e$. Поэтому есть производное отображение, которое действует в касательном пространстве. Оно обозначается $\mathrm{Ad}\, x$ и переводит элемент $X$ касательного пространства в другой элемент касательного пространства, который я тоже условно обозначу $xXx^{-1}$. Такое обозначение оправданно, потому что подавляющее большинство групп Ли можно реализовать как подгруппы в группе матриц (если рассматривать и бесконечные матрицы, то так реализуется любая группа Ли; если же рассматривать только матрицы конечного порядка, то есть исключения). А для группы матриц эту формулу можно понимать буквально. Большинство физиков так и делает, потому что для них любая группа состоит из матриц.

Мы описали действие в самой алгебре Ли, а нам нужно действие в сопряженном пространстве. Обозначу $K(x)$ действие элемента $x$ в сопряженном пространстве $\mathfrak{g}^*$ ($K$ — от слова «коприсоединенное»: представление $\mathrm{Ad}\, x$ называют присоединенным, а представление $K(x)$ — коприсоединенным). $K(x)$ действует на функционал $F$ по следующему правилу:

$$\langle K(x)F, X\rangle = \langle F, \mathrm{Ad}\, x^{-1}X\rangle.$$

Матрица коприсоединенного представления отличается от матрицы присоединенного представления тем, что она заменяется на обратную матрицу и транспонируется. Казалось бы, разница небольшая. Но оказывается, что присоединенное представление и коприсоединенное представление выглядят совершенно по-разному. Грубо говоря, коприсоединенное представление имеет гораздо больше свойств, чем присоединенное.

Одно из геометрических свойств, которое сразу бросается в глаза, заключается в том, что все коприсоединенные орбиты — симплектические многообразия, причем действие группы сохраняет симплектическую структуру. Выражение «коприсоединенные орбиты» — это жаргон; полностью надо говорить «орбиты коприсоединенного представления». Орбита получается следующим образом: берем функционал $F$ и применяем к нему $K(x)$ для всех $x \in G$.

Теперь еще одно краткое отступление о симплектических многообразиях. Это понятие пришло в математику из механики. Оно является, в каком-то смысле, нечетным аналогом риманова многообразия. Риманово многообразие, может быть, даже ближе начинающим математикам. Это — многообразие, на котором можно измерять длины, т. е. задана положительно определенная квадратичная форма в касательном пространстве. В симплектическом многообразии задается не квадратичная форма в касательном пространстве, а кососимметрическая билинейная форма, т. е. антисимметричное скалярное произведение. Это, грубо говоря, — нечетный аналог понятия метрики. Вообще, сейчас люди склоняются к тому, что у каждого понятия должен быть свой нечетный аналог, и только рассматривая одновременно сам объект вместе с его нечетным аналогом, мы получаем правильное понимание того или иного явления.

Более точная аналогия между римановыми многообразиями и симплектическими многообразиями получается, если учесть, что у римановых многообразий самое главное качество — это кривизна. Бывают плоские пространства, а бывают неплоские, и эта неплоскость выражается в том, что метрика имеет кривизну. Если метрика заменяется на кососимметрическую квадратичную форму, то для нее тоже есть понятие кривизны. Так вот, симплектические многообразия — это нечетные аналоги плоских римановых многообразий, т. е. кривизна у них равна нулю. Координатное выражение этого таково: если мы локально запишем кососимметрическую форму в виде $\omega = \omega_{ij} dx_i \wedge dx_j$, то условие плоскости запишется в виде

$$d\omega = \frac{\partial \omega_{ij}}{\partial x_k}\, dx_i \wedge dx_j \wedge dx_k = 0.$$

Известно, что всякую плоскую метрику можно в подходящей системе координат заменить на постоянную метрику. Точно так же симплектическую структуру можно в подходящей системе координат заменить на постоянную симплектическую структуру, т. е. выбрать локальную систему координат так, что коэффициенты $\omega_{ij}$ — постоянные числа. Локально симплектическая геометрия никаких инвариантов не имеет. А глобальные инварианты она имеет, и это очень интересная наука — симплектическая топология. С ней каждый может познакомиться по недавно вышедшему собранию сочинений Арнольда.[4] Кстати сказать, оно интересно и во многих других отношениях.

Каждая орбита коприсоединенного представления каноническим образом является симплектическим многообразием, и симплектическая форма $\omega$, которая на ней существует, инвариантна относительно действия нашей группы. Это, казалось бы, очень сложно формулируемое свойство имеет одно яркое и наглядное следствие: все коприсоединенные орбиты обязательно имеют четную размерность.

Самый простой пример такой. Если в качестве исходной группы взять группу вращений трехмерного пространства, то соответствующая алгебра Ли — это обычное трехмерное пространство, присоединенное представление — это так называемое тавтологическое представление (каждое вращение трехмерного пространства и действует именно как вращение трехмерного пространства), а орбиты — это двумерные сферы или само начало координат. Как видите, встречаются только размерности 0 и 2.

Основной принцип метода орбит состоит в том, что следующие два множества не то, чтобы совпадают, но в каком-то смысле близки друг к другу и отвечают друг за друга. А именно, множество $\widehat{G}$ (классы эквивалентности унитарных неприводимых представлений, сокращенно unirreps) и множество коприсоединенных орбит $\mathfrak{g}/G$:

$$\widehat{G} \cong \mathfrak{g}/G.$$

Написанное соответствие выглядит как нечто очень универсальное: оно должно быть справедливо для любой группы Ли. Поэтому естественно воспринимать его как некий общий принцип, который справедлив всегда, когда имеют смысл обе части равенства. Например, группу Ли можно понимать в несколько обобщенном смысле. Можно считать, что это не классическое гладкое многообразие с вещественными координатами, а, скажем, комплексное многообразие, или многообразие над другим полем (алгебраическое многообразие). Можно также считать, что это бесконечномерное многообразие. Можно, наконец, считать, что это квантовая группа, которая группой вообще не является, но понятие алгебры Ли для нее есть и понятие коприсоединенного представления тоже есть. И самый интересный для меня сегодня случай, это случай, когда наша группа является алгебраическим многообразием над конечным полем. Тогда она сама как группа является конечной группой. И все теоремы и гипотезы, о которых я буду говорить, относятся к очень простому и обозримому объекту — конечным группам. Замечательно, что некоторые утверждения, которые получаются применением метода орбит, оказываются известными верными теоремами, другие оказываются хотя и верными, но новыми и неизвестными теоремами,

[4] См. литературу, указанную на с. 6. — *Прим. ред.*

третьи оказываются неверными, и, наконец, четвертые остаются проблемами. В основном о проблемах я и хотел поговорить.

Общий принцип метода орбит состоит в том, что он — линейное, или, как еще говорят, квазиклассическое, приближение к настоящей теории представлений. А именно, вместо нелинейного многообразия (группы Ли) рассматривается линейное многообразие (ее алгебра Ли) и закон умножения рассматривается только с точностью до квадратичных членов, что является первым шагом приближения. Поэтому чем ближе группа Ли к алгебре Ли, тем лучше работает этот метод.

Между группой Ли и алгеброй Ли существует некоторая связь, а именно, имеется естественное отображение из алгебры Ли в группу Ли. Если у вас есть касательный вектор, т. е. направление движения из единицы, то можно двигаться в этом направлении наиболее естественным образом — так, чтобы траектория движения была однопараметрической подгруппой. Параметр на этой кривой можно выбрать так, чтобы выполнялись следующие два свойства:

1) $g_t g_s = g_{t+s}$ (это можно сделать в силу общей теоремы: любая одномерная группа локально есть аддитивная группа прямой);

2) $\dot{g}_0 = X$, где $X$ — заданный вектор.

Система обыкновенных дифференциальных уравнений имеет единственное решение с таким начальным условием, и $g_1$ по определению обозначается $\exp X$. Это обозначение не случайно: для матричной группы Ли $\exp X = \sum_{k\geqslant 0} X^k/k!$, т. е. мы получаем настоящую экспоненту.

Отображение exp определено на всей алгебре Ли. Но образ этого отображения покрывает не всю группу Ли. Поэтому обратное отображение, которое естественно было бы назвать логарифмом, не везде определено, не везде однозначно, и вообще обладает гораздо худшими свойствами, чем прямое отображение exp.

Метод орбит проще всего работает в том случае, когда отображение exp взаимно однозначно. Самый замечательный пример, когда это отображение взаимно однозначно, я и хочу сегодня рассмотреть в качестве основного объекта.

Обозначим через $G_n$ алгебраическую группу верхнетреугольных матриц. Я сделаю краткое отступление об алгебраических многообразиях. Это очень важное понятие, и даже не столь важно само понятие, как идеология, с ним связанная. На заре этой науки алгебраическим многообразием называлось множество совместных решений алгебраической системы уравнений. Со временем было понято, что это неправильная точка зрения. Например, если вы рассматриваете вещественные числа, то уравнения $x^2 + y^2 = -1$ и $x^2+y^2 = -3$ эквивалентны, потому что ни то ни другое решений не имеет. В то же время ясно, что это немножко разные уравнения, они должны описывать разные множества. И действительно, если мы рассмотрим комплексные решения, то решения этих двух уравнений окажутся разными. Или рассмотрим еще другое уравнение $x+y=x+y+2$. Оно тоже не имеет решений, поэтому должно рассматриваться как эквивалентное двум первым уравнениям. Но ясно, что никакого преобразования, переводящего их друг в друга, придумать нельзя. Это разные объекты. Если первые уравнения можно спасти, рассматривая комплексные решения, то это уравнение не имеет и комплексных решений. Но можно, например, рассматривать решения в поле вычетов по модулю 2. Тогда это уравнение задает всю плоскость — любая пара $(x, y)$ является решением.

Системы уравнений имеют разные множества решений, если их рассматривать над разными полями, или, более общо, над разными алгебрами. Постепенно люди пришли к такому пониманию: система уравнений — это функтор из категории алгебр в категорию множеств. А именно, если у вас есть какая-то система уравнений с коэффициентами из какого-то поля $k$ и если вы зададите какую-то алгебру $A$ над этим полем, то множество решений над этой алгеброй будет каким-то множеством $X_A$. Поэтому мы получаем соответствие $A \mapsto X_A$, и это соответствие есть функтор. Что такое функтор, я объяснять не буду. Каждый должен спросить у умного соседа.

Это уже правильное понимание того, что такое алгебраическое многообразие. Какие нужно рассматривать алгебры — это отдельный вопрос. Иногда достаточно рассматривать комплексные числа, иногда достаточно рассматривать более узкие образования. А иногда полезно рассматривать также некоммутативные алгебры $A$. Это уже начала так называемой некоммутативной алгебраической геометрии.

Алгебраическая группа — это одновременно группа и алгебраическое многообразие. Только надо помнить, что алгебраическое многообразие не является множеством. Оно становится множеством лишь после того, как вы выберете некоторую алгебру $A$. Так и алгебраическая группа не является группой в обычном смысле этого слова по той простой причине, что она не является множеством. А группа, как учат в теории групп, является множеством.

Алгебраическая группа — это нечто, что становится группой после того, как вы выберете алгебру $A$ и рассмотрите множество решений соответствующей системы уравнений над алгеброй $A$. Это, как вы понимаете, не одна группа, а целое семейство групп, зависящих от выбора алгебры.

После этой подготовки я напишу пример алгебраической группы. Группа $G_n$ будет алгебраическим подмногообразием в множестве всех матриц: $G_n \subset \subset \mathrm{Mat}_n$. И будет она задаваться такой системой уравнений: $x_{ij}=0$ при $i>j$ и $x_{ij}=1$ при $i=j$. Можно рассматривать решения этой системы над любым полем и над любой алгеброй. Я утверждаю, что $G_n$ — алгебраическая группа.

Что это такое, я не буду аккуратно определять. Проверю лишь, что как только мы рассматриваем решения над какой-то алгеброй, то в результате получается группа. Пусть для простоты $n = 3$. Тогда множество решений этой системы над заданной алгеброй $A$ выглядит так:

$$\begin{pmatrix} 1 & a_{12} & a_{13} \\ 0 & 1 & a_{23} \\ 0 & 0 & 1 \end{pmatrix},$$

где $a_{ij} \in A$. Такие матрицы называют строго верхнетреугольными (строго — потому что на диагонали стоят единицы). Для любой алгебры $A$ множество строго верхнетреугольных матриц является группой относительно обычного матричного умножения.

Теперь уже я могу сформулировать основной предмет изучения сегодняшней лекции. А именно, что может сказать метод орбит про теорию представлений этой группы в случае, когда $A = \mathbb{F}_q$ — конечное поле из $q$ элементов. Здесь $q = 2, 3, 4, 5, 7, 8, 9, 11, 13, \dots$ Не существует полей, состоящих из 6, 10, 12 элементов. Самый интересный контрольный пример — поле из 4 элементов: существует поле из 4 элементов, но оно не является полем вычетов по модулю 4.

Интересные задачи здесь возникают, когда начинаешь заниматься так называемой асимптотической теорией представлений. Это значит, что у нас есть не одна конечная группа, а целая бесконечная серия конечных групп, и мы интересуемся асимптотическими свойствами. Например: нельзя ли эти асимптотические свойства проинтерпретировать как какие-то результаты по теории представлений какого-то идеального объекта, находящегося в бесконечности.

Для группы $G_n(\mathbb{F}_q) \subset \mathrm{Mat}_n(\mathbb{F}_q)$ есть два параметра: $n$ и $q$. Они играют совершенно разную роль, хотя и есть любопытные соображения, по которым есть связь между этими двумя параметрами. Можно действовать двумя способами. Во-первых, можно фиксировать $n$ и менять $q$. Это означает, что мы рассматриваем одну и ту же алгебраическую группу над разными полями. Оказывается, что метод орбит к этому прекрасно приспособлен и дает универсальное описание представлений всех этих групп единообразным образом с некоторыми дополнительными поправками, зависящими от $n$. А вот если мы фиксируем $q$ и $n$ устремляем к бесконечности, то эта задача намного интереснее, и даже самые простые вопросы пока еще не имеют ответов. Вот именно эту задачу я и хочу обсудить: как устроена теория представлений группы верхнетреугольных матриц очень большого порядка с элементами из фиксированного конечного поля. Что касается поля, то здесь во многих случаях от поля мало что зависит, и результат примерно одинаков для всех полей. Поэтому проще всего ограничиться простейшим полем, а именно полем $\mathbb{F}_2 = \{0, 1\}$, состоящим из двух элементов.

Поле $\mathbb{F}_2 = \{0, 1\}$ является простейшим полем в силу аксиомы, утверждающей, что $0 \neq 1$. Эта и только эта аксиома запрещает полю состоять из одного элемента. И в некоторых случаях имеет смысл говорить о поле, состоящем из одного элемента. Тогда эта аксиома нарушается, но других препятствий нет. И любопытно, что некоторые формулы имеют смысл и в том случае, когда $q$ обращается в 1. Есть даже такая метаматематическая философия, которая говорит, что случай $q = 1$ в какой-то степени соответствует бесконечному полю, а именно, комплексному полю $\mathbb{C}$. Более того, $q = -1$ соответствует полю $\mathbb{R}$. Но это отдельная тема, о которой я не хочу сейчас говорить из-за недостатка времени. Желающие могут посмотреть рассуждения на эту тему в «Сочинениях» Арнольда.[5)]

В дальнейшем я часто буду ограничиваться простейшим полем из двух элементов, хотя интересные вопросы будут и для общих полей.

Мы начнем с вопроса о том, как описать коприсоединенную орбиту. Уж если мы верим, что метод орбит работает, то тогда неприводимые представления соответствуют коприсоединенным орбитам, поэтому хорошо бы научиться как-то описывать коприсоединенные орбиты. И тут возникает следующий вопрос. В алгебраической геометрии есть понятие элемента общего положения. И орбиты общего положения описываются обычно легко, во всяком случае, для треугольной группы они описываются легко. А дальше идут вырожденные орбиты, и чем вырожденнее класс орбит, тем труднее их классифицировать. Этот вопрос пока еще не решен до конца. Более того, он не решен ни для какого поля. Для вещественного поля этот вопрос возник очень давно, лет 30 тому назад. Уже тогда такая задача была известна. Уже тогда было непонятно, как классифицировать представления треугольной группы с вещественными элементами. После появления метода орбит стало ясно, что это то же самое, что классифицировать коприсоединенные орбиты. Задача, казалось бы, существенно упростилась: вместо бесконечномерных представлений в гильбертовом пространстве возник вполне конечный объект — конечномерные треугольные матрицы и конечномерная группа, действующая на них указанным образом. Тем не менее, классификация орбит не поддается легкому решению. И даже есть соображения, почему тривиального решения здесь и не должно быть. С другой стороны, те, кто занимался этой классификацией, рано или поздно приходили к убеждению, что если такая классификация и получится, то от поля она не будет сильно зависеть. Для комплексного поля, для вещественного, для конечного — классификация примерно одна и та же. Неизвестно только, как она устроена. Поэтому случай простейшего поля вызывает наибольший интерес.

Теперь я напишу несколько простых формул, показывающих, как устроено коприсоединенное представление для треугольной группы. Сама

[5)] *Арнольд В. И.* Избранное — 60. — М.: Фазис, 1997. — *Прим. ред.*

группа $G_n$ состоит из элементов

$$g = \begin{pmatrix} 1 & * & * & * & * \\ 0 & 1 & * & * & * \\ 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

(Для примера здесь $n = 5$.) На диагонали стоят единицы, под диагональю — нули, а над диагональю — любые элементы.

Алгебра Ли — это касательное пространство. Хотя наше множество $G_n$ теперь не многообразие, и формального касательного пространства у него нет, но для любого алгебраического многообразия можно определить касательное пространство. В нашем случае касательное пространство состоит из таких же элементов, только с нулями на главной диагонали:

$$\mathfrak{g}_n \ni X = \begin{pmatrix} 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Элементы двойственного пространства удобно записывать в виде нижнетреугольных матриц:

$$\mathfrak{g}_n^* \ni F = \begin{pmatrix} 0 & . & . & . & . \\ * & 0 & . & . & . \\ * & * & 0 & . & . \\ * & * & * & 0 & . \\ * & * & * & * & 0 \end{pmatrix}.$$

На диагонали стоят нули, под диагональю стоят какие-то элементы, а над диагональю мне не хочется ставить нули — я поставил точки. Смысл в этом вот какой. Если некоторое пространство является подпространством в пространстве матриц, то двойственное пространство будет факторпространством в двойственном пространстве матриц. Переход к двойственному пространству — это контравариантный функтор, который меняет подпространства и факторпространства. Пространство матриц само себе двойственно. След $\operatorname{tr}(F \cdot X)$ — билинейная функция, зависящая от матриц $F$ и $X$. При фиксированном $F$ она будет линейным функционалом от $X$, а при фиксированном $X$ она будет линейным функционалом от $F$. При этом каждая линейная функция от матрицы $X$ записывается в таком виде для подходящей матрицы $F$. Так что пространство $\mathrm{Mat}_n(\mathbb{F}_q)$ само себе двойственно. Более того, эта двойственность, что особенно важно, инвариантна относительно сопряжения: если заменить $X$ и $F$ на $xXx^{-1}$ и $xFx^{-1}$, то результат не изменится. Поэтому двойственность выдерживает присоединенное действие на $X$

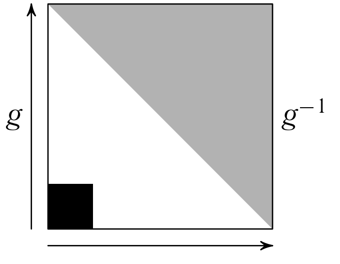
и одновременно коприсоединенное действие на $F$. Но если $X$ пробегает не все матрицы, а только часть — верхнетреугольные матрицы, тогда в роли $F$ нужно рассматривать факторпространство всех матриц по тем матрицам, которые аннулируют верхнетреугольные матрицы. Точки над главной диагональю означают, что там стоят любые числа, на которые мы не обращаем внимания.

Теперь коприсоединенное действие можно записать следующим образом:

$$K(x)F = [xFx^{-1}]_{\text{нижняя часть}}.$$

Матрицу $F$ при этом тоже можно себе представлять как нижнюю часть какой-то матрицы: нижняя часть преобразованной матрицы зависит только от нижней части исходной матрицы.

После этой предварительной подготовки я могу выписать явные формулы, как выглядят орбиты действия коприсоединенного представления. Матрицу $F$ мы умножаем на верхнетреугольную матрицу $g$ слева и на верхнетреугольную матрицу $g^{-1}$ справа:




Как известно, умножение слева на треугольную матрицу $g$ указанного вида сводится к тому, что последняя строка остается неизменной, к предпоследней строке добавляется последняя с каким-то коэффициентом, к предпредпоследней — линейная комбинация двух предыдущих строк, и т. д. Левая стрелочка показывает, что действие происходит снизу вверх. Точно так же, когда умножаем на матрицу $g^{-1}$ справа, неизменным остается первый столбец. Ко второму столбцу добавляется первый, умноженный на какой-то коэффициент, и т. д. Нижняя стрелочка показывает, в какую сторону происходит действие. Видно, что влияние распространяется из левого нижнего угла вправо и вверх. В частности, сам левый нижний элемент остается неизменным. Это — пример так называемого инварианта. Точно так же минор, составленный из двух последних строк и двух первых столбцов, тоже остается неизменным. К его второму столбцу добавляется первый с каким-то коэффициентом, но определитель при этом не меняется. То же самое и для строк. Левые нижние миноры, которые я обозначу $\Delta_1, \Delta_2, \ldots, \Delta_{[n/2]}$, являются инвариантами: при коприсоединенном действии эти величины не меняются. Оказывается, что если мы приравняем эти инварианты константам, то система уравнений

$$\Delta_1 = c_1, \quad \Delta_2 = c_2, \quad \ldots, \quad \Delta_{[n/2]} = c_{[n/2]}$$

задает инвариантное относительно коприсоединенного действия множество, которое при общих (типичных) значениях констант состоит ровно из одной орбиты. Что здесь означает «типичные»? Например, достаточно, чтобы все эти константы, кроме последней для нечетного $n$, были отличны от нуля. Если же какие-то константы обращаются в нуль, то эта система уравнений по-прежнему задает инвариантное множество, но оно может распадаться на более мелкие орбиты. Картина распадения большого инвариантного множества на мелкие орбиты усложняется по мере увеличения количества инвариантов, обращающихся в нуль, т. е. по мере того как падает размерность соответствующей коприсоединенной орбиты. Структура напоминает ветвящееся дерево. Имеется инвариантное множество общего положения, как правило, это просто одна орбита. Некоторые исключительные множества оказываются не орбитами, и там возникают дополнительные инварианты, которые являются инвариантами только на этом множестве: вообще говоря, они не инварианты, но на данном множестве инварианты. Происходит дальнейшее разбиение, где опять же типичные слои являются орбитами, а исключительные слои допускают дополнительные инварианты третьего уровня, и т. д. Этот ветвящийся процесс распространяется все дальше и дальше, пока он не доходит до орбит нулевой размерности, т. е. до точек. И пока еще не придуман адекватный аппарат для описания всего этого процесса.

Теперь я вступаю в новую область, которая связана с двумя вещами. Первая — это так называемая экспериментальная математика. Человек сидит и что-то считает и получает какой-то результат. Потом он еще что-то считает и получает другой результат. А когда получено несколько результатов, он думает, нельзя ли их объединить какой-нибудь теорией. Обычно успех зависит от того, сколько предварительных вычислений было сделано. В прошлом веке люди сидели годами, исписывали килограммы бумаги. Сейчас люди работают на компьютере, и не годами, а месяцами, потому что надо писать работы, надо участвовать в конкурсах на замещение должности, и т. д. Жизнь ускоряется, времени на обдумывание остается все меньше и меньше. Но тем не менее, экспериментальная математика процветает. Может быть, отчасти, потому, что есть люди, которые уже не озабочены своим положением и могут думать неограниченное количество времени. По крайней мере, Конвей и Коксгер — два замечательных математика, которые настолько прочно сидят на своих местах, что им не нужно думать о вещах насущных, и они могут думать о смысле жизни и о смысле математики. И они оба придумали в последнее время несколько совершенно неожиданных вещей, которые можно придумать, только если у тебя есть свободное время и ты можешь его тратить в неограниченном количестве. Последнее изобретение Кокстера очень пригодно для математической олимпиады.

Экспериментальная математика — это одна часть. Вторая — это наука, которая еще, может быть, даже и не существует как наука, но направление такое уже есть. Оно называется частично полностью интегрируемые системы. Интегрируемые системы — это такие механические системы, где решение, грубо говоря, можно выписать в явном виде. Если речь идет не о классических, а квантовых системах, то имеются в виду матрицы, собственные значения которых и собственные векторы можно выписать явно. Таких матриц немного, и обычно это связано с какой-то симметрией. Если есть симметрия, то это облегчает нахождение собственных значений и собственных векторов. Но лет 10 тому назад было обнаружено, что есть такие матрицы, для которых удается написать несколько первых собственных значений и несколько первых собственных векторов, а дальше не получается. И не потому не получается, что человек туп, а потому, что природа так устроена: первые значения хороши, а дальше наступает какой-то хаос, где уже сказать ничего нельзя.

Что-то похожее наблюдается и в той науке, о которой я рассказываю. Внешне это проявляется следующим образом. Применим экспериментальную математику, взяв самое простое поле из двух элементов и проделав тупые вычисления. Возьмем матрицы порядка 1, 2, 3, и т. д., и вычислим там количество присоединенных орбит. Получается последовательность, начало которой я сейчас выпишу. Чтобы она была убедительна, проделаем здесь вычисление до тех пределов, пока это легко сделать.

Начнем с матриц порядка 1. Группа Ли состоит и одной матрицы с элементом 1, алгебра Ли состоит из одной матрицы с элементом 0. Эта матрица и является единственной орбитой.

Для матриц порядка 2 алгебра Ли состоит из матриц

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}.$$

Двойственное пространство состоит из матриц

$$\begin{pmatrix} \cdot & \cdot \\ x & \cdot \end{pmatrix}.$$

Над полем из двух элементов получаем 2 орбиты, потому что группа в данном случае коммутативна и ее действие тривиально, а само $x$ принимает 2 значения.

Для третьего порядка нужно уже немножко считать. Алгебра Ли состоит из матриц

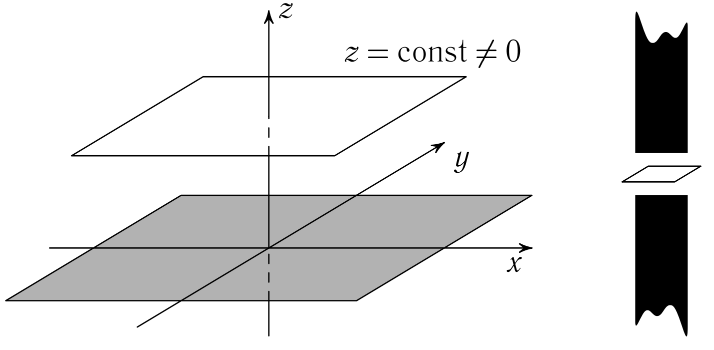
$$\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}.$$

Двойственное пространство состоит из матриц

$$\begin{pmatrix} \cdot & \cdot & \cdot \\ x & \cdot & \cdot \\ z & y & \cdot \end{pmatrix}.$$

Преобразование выглядит так: $(x, y, z) \mapsto (x + az, y - bz, z)$.

Я нарисую, как выглядит множество орбит. Нас интересует поле из двух элементов, но картинка годится для любого поля, поэтому я нарисую картинку над полем вещественных чисел. Орбиты выглядят вот как: $z$ является инвариантом (это тот самый инвариант $\Delta_1$, который обсуждался в общем случае). Поэтому каждая плоскость $z = \mathrm{const}$ есть инвариантное множество. Если $z \neq 0$, то за счет подбора $a$ и $b$ мы можем сделать $x$ и $y$ произвольными. Это означает, что вся плоскость состоит из одной орбиты. Это соответствует общему геометрическому замечанию о том, что все орбиты должны быть четномерными. Если же $z = 0$, то все добавки тоже нулевые и все точки неподвижные. Это означает, что координатная плоскость $z = 0$ разбивается на орбиты, состоящие из отдельных точек. Размерность точки тоже четна.




Наглядно множество орбит можно изобразить так. Нарисуем толстую ось $z$, состоящую из толстых точек — они соответствуют орбитам. Одну точку $z = 0$ нужно вырезать и вместо нее вклеить целую плоскость, состоящую из тощих точек. То, что получится, и есть пространство орбит вместе с его естественной топологией: если мы стремимся к нулю по толстой оси $z$, то в качестве предела получаются все тощие точки. Такая топология типична для пространства орбит. Она же, кстати сказать, присутствует и в теории представлений. У множества неприводимых представлений тоже есть своя естественная топология, и оказывается, что она совпадает с топологией пространства орбит.

Теперь вернемся к конечному полю. Количество толстых орбит равно $q - 1$, а количество тощих орбит равно $q^2$. Всего будет $q - 1 + q^2$ орбит.

Замечу, что мы получили многочлен от $q$. Это факт общий, для любого $n$ количество орбит будет многочленом от $q$.

Продолжая такие вычисления, получаем табличку

| $n$ | $-1$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|
| $O(n)$ | 1 | 1 | 1 | 2 | 5 | 16 | 61 |

Я дополнил табличку двумя начальными членами: числом орбит для $n = 0$ и для $n = -1$. Дело в том, что такая последовательность появляется и из совсем других соображений, и тогда она как раз начинается с трех единиц.

Теперь я сделаю отступление о частично точно решаемых задачах. Оказывается, что написанная выше последовательность ровно до того места, до которого она выписана, очень хорошо описывается другой последовательностью, весьма замечательной и встречающейся во многих других ситуациях. Неприятность заключается в том, что у хорошо известной последовательности на следующем месте должно стоять 272, а количество орбит при $n = 6$ равно 275. Формулу для последовательности $K_n$, о которой идет речь, несложно написать:

$$\sum_{n\geqslant 0} K_n \frac{x^n}{n!} = \operatorname{tg} x + \sec x.$$

Количество орбит начинает хорошо, но потом неожиданно отклоняется от этой хорошей последовательности.

В таком состояние дело находилось уже довольно давно, первый мой доклад на эту тему был уже года два тому назад. Тогда были известны эти две последовательности и было известно их расхождение в шестом члене. Оказывается, что если произвести более детальные исследования, а именно, рассмотреть не только поле из двух элементов, но и поле из $q$ элементов, то можно придумать $q$-аналог последовательности $K_n$. Членами этой последовательности будут многочлены от $q$. Они тоже, как и числа $K_n$, задаются так называемым треугольником Эйлера—Бернулли. Он опять же имеется в «Сочинениях» Арнольда.[6] Треугольник Эйлера—Бернулли имеет замечательный $q$-аналог, элементы которого являются многочленами от $q$. На самом деле там есть еще вспомогательный параметр $t$, поэтому получаются многочлены даже от двух переменных: они превращаются в многочлены от $q$ при $t = 1$. Так вот, оказалось, что несовпадение, если рассматривать его над полем с произвольным $q$, делится на $(t-q)(t-q^2)$. В частности, если параметр $t$ принимает исключительные значения $q$ и $q^2$, то невязка пропадает.

Объясню теперь это всё подробно. Мы занимаемся разбиением множества $\mathfrak{g}_n^*(\mathbb{F}_q)$ на орбиты (здесь $n$ — порядок треугольных матриц). Это множество является линейным пространством над полем $\mathbb{F}_q$, размерность

---

[6] *Арнольд В. И.* Избранное — 60. — М.: Фазис, 1997. С. 688. — *Прим. ред.*

которого равна числу ненулевых элементов в треугольных матрицах. Поэтому $|\mathfrak{g}_n^*(\mathbb{F}_q)| = q^{n(n-1)/2}$. Это конечное множество распадается на орбиты. Каждая орбита $\Omega$ четномерна (над конечным полем это имеет вполне определенный смысл: орбита $\Omega$ состоит из $q^{2r}$ точек). Мы можем классифицировать орбиты по их размерности. Это довольно насильственная классификация, но какой-то смысл она имеет. Пусть $a_n^r(q)$ — число $2r$-мерных орбит в $\mathfrak{g}_n^*(\mathbb{F}_q)$. Для последовательности $a_n^r(q)$, зависящей от параметров $n$ и $r$, естественно ввести производящую функцию

$$P_n(t,q) = \sum a_n^r(q)t^r.$$

Теперь у нас есть последовательность многочленов, которые зависят от двух переменных: $t$ и $q$. Эта последовательность многочленов может быть включена в обобщенный треугольник Эйлера—Бернулли вплоть до $n = 5$. При $n = 6$ между ними получается разница, но эта разница делится на $(t-q)(t-q^2)$. При $t = q$ или $q^2$ эта разница пропадет. Для $t = q^2$ это не удивительно. Грубо говоря, $t$ классифицирует орбиты по размерности, и когда мы полагаем $t = q^2$, это означает, что мы заменяет каждую орбиту размерности $2r$ группой из $q^{2r}$ точек. Это — просто число точек в орбите, а общее число точек легче считать, чем число орбит: оно известно. Поэтому делимость на $t - q^2$ доказывается совсем просто. А вот когда $t = q$, мы получаем квадратный корень из числа точек на орбите. Согласно методу орбит это число имеет истолкование: оно равно размерности неприводимого представления. Значит, мы получаем не число орбит, а сумму размерностей неприводимых представлений. Эта сумма размерностей ведет себя уже лучше, чем число орбит. Например, она дальше совпадает с предсказаниями теории. Оказывается, что если положить $t = q$, то совпадение с теорией продолжается дальше, и проверено оно вплоть до $n = 11$. Это потребовало большой вычислительной работы, на которую я лично не способен, но я нашел себе соавтора, Аню Мельникову из института Вейцмана, которая смогла это сделать. Она проделала все эти вычисления и убедилась, что до $n = 11$ обе формулы совпадают. Мы уже было обрадовались, но пришло известие, что для $n = 13$ формула не может быть верна. Дело в том, что из нее вытекает некое следствие, которое я сейчас приведу, а это следствие опровергается неким примером.

Это не совсем противоречащий пример — это противоречащий пример к гораздо более общему утверждению, я бы даже сказал более нахальному утверждению, которое состоит в том, что метод орбит работает *буквально* для рассматриваемой последовательности групп. Тогда, в частности, из метода орбит вытекает явная формула для характеров неприводимых представлений. А если эта формула верна, то все характеры неприводимых представлений вещественны (для поля из двух элементов). А уже из этого, как знает каждый специалист по теории групп, вытекает, что любой элемент группы сопряжен своему обратному: $g \sim g^{-1}$. Классы сопряженности различаются характерами, а характер на элементе $g$ — это сопряженное значение характера на элементе $g^{-1}$. Для вещественных чисел они совпадают.

Возникает задача, которую можно сформулировать для первокурсника. Имеется треугольная матрица с элементами из $\mathbb{F}_2$. Спрашивается: сопряжена ли она своей обратной матрице внутри треугольной группы или не сопряжена? Если бы речь шла обо всей группе матриц, то ответ положителен. Дело в том, что треугольная матрица приводится к набору жордановых блоков, а жордановы блоки у самой матрицы и у ее обратной одни и те же. Но внутри треугольной группы ответ не такой очевидный. Для матриц порядка 13 с помощью компьютера показано, что имеется ровно один противоречащий пример. Поэтому неверна та формула для характеров, которая получается совсем тупым применением метода орбит. Это жаль, потому что если бы такое применение было верно, то можно было бы доказать совпадение всех многочленов.

На самом деле, я немножко поспешил. Из того, что имеется противоречащий пример, еще не вытекает, что при $t = q$ многочлены различны. Верно лишь то, что доказательство, которое я имел в виду, не годится. Сами же многочлены настолько красивы, что, может быть, они и продолжают годиться.

## Часть 2

Эта лекция хотя и будет продолжением первой, но ее можно слушать независимо. Я надеюсь рассказать более конкретные вещи. Я хочу рассказать о задачах, связанных с представлениями конечных групп, которые возникают из метода орбит. Хотя сам метод орбит относится к бесконечным (и даже бесконечномерным) группам, оказывается, что правильное его истолкование применимо и к конечным группам тоже.

Начну я с формулировки некоторой гипотезы, для понимания которой не нужно вообще ничего. Это будет в стиле тех занятий с молодыми математиками, которые ведутся с молодыми математиками здесь в Независимом университете, и будет также продолжением традиции математических кружков. Я буду рассказывать про некоторую замечательную последовательность многочленов. Я уверен, что она замечательная, и одним из оправданий этого является то, что эта последовательность имеет по крайней мере 5 разных определений, про которые не доказано, что эти определения совпадают.

Мы будем рассматривать последовательности многочленов, которые обозначаются $A_n(q)$, $B_n(q)$, $C_n(q)$, $D_n(q)$ и $Y_n(q)$. Эти многочлены не имеют отношения к простым алгебрам Ли серий $A$, $B$, $C$, $D$. Тем, кто был в прошлый раз, я напомню, что мы рассматривали теорию представлений. Для конечной группы $G$ одной из теоретико-представленческих характеристик является набор размерностей неприводимых комплексных представлений этой группы. Сам набор неприводимых комплексных представлений группы $G$, рассматриваемых с точностью до эквивалентности, я буду обозначать $\widehat{G}$. Элемент множества $\widehat{G}$ я буду обозначать $\lambda$, а то представление, класс эквивалентности которого обозначается $\lambda$, будет обозначаться $\pi_\lambda$. Таким образом, $\pi_\lambda$ — это гомоморфизм $G \to \mathrm{GL}(d(\lambda), \mathbb{C})$. Группа конечна, поэтому ее любое конечномерное представление эквивалентно унитарному представлению, и можно считать, что $\pi_\lambda : G \to U(d(\lambda)) \subset \mathrm{GL}(d(\lambda), \mathbb{C})$.

Набор целых положительных чисел $d(\lambda)$ обладает несколькими замечательными свойствами. Например, все они являются делителями порядка группы. Кроме того, имеется знаменитое тождество Бернсайда

$$\sum d^2(\lambda) = \#G.$$

(Здесь $\#G$ обозначает число элементов группы $G$). А если взять сумму нулевых степеней, то получим

$$\sum d^0(\lambda) = \#\widehat{G}.$$

Хотелось бы эти два тождества проинтерполировать и вставить между ними сумму самих $d(\lambda)$: $\sum d(\lambda) = {?}$ В действительности не известно, обладает ли эта сумма какими-либо хорошими свойствами. Но есть частный случай, когда эта сумма все-таки обладает хорошими свойствами. Это тот случай, когда все представления (которые априори комплексные) являются вещественными.

Я сделаю отступление и напомню тем, кто знает, и сообщу тем, кто не знает, что комплексные неприводимые представления группы разбиваются на три типа:

*вещественный тип*: представление $\pi_\lambda$ эквивалентно вещественному, т. е. при подходящем выборе базиса все операторы представления записываются матрицами с вещественными коэффициентами;

*комплексный тип*: комплексно сопряженное представление $\overline{\pi}_\lambda$ не эквивалентно представлению $\pi_\lambda$ (для каждого представления можно построить комплексно сопряженное представление, выбрав базис и заменив в каждой матрице представления каждый элемент на комплексно сопряженный: комплексное сопряжение является автоморфизмом комплексного поля, поэтому все тождества сохраняются и представление остается представлением);

*кватернионный тип*: представление $\pi_\lambda$ не эквивалентно вещественному представлению, но $\overline{\pi}_\lambda \sim \pi_\lambda$.

Кватернионный тип, можно сказать, самый интересный тип представлений. Он возникает тогда, когда комплексная размерность представления четна и это четномерное комплексное пространство получается сужением поля скаляров из кватернионного пространства вдвое меньшей размерности. При этом сами операторы представления можно записывать кватернионными матрицами.

Самый известный пример — группа $SU(2)$, которая сама по себе является группой кватернионов, по модулю равных 1. Эта группа бесконечная, но у нее имеется много конечных подгрупп. Тавтологическое представление группы $SU(2)$ как одномерных кватернионных матриц дает пример кватернионного представления.

Как известно, каждый кватернион можно изобразить матрицей порядка 2 с комплексными элементами, поэтому каждое кватернионное представление размерности $n$ можно при желании рассматривать как комплексное представление размерности $2n$. Это будет представление кватернионного типа.

Назовем индексом представления $\lambda$ число

$$\operatorname{ind}\lambda = \begin{cases} 1, & \text{если представление } \lambda \text{ вещественного типа;} \\ 0, & \text{если представление } \lambda \text{ комплексного типа;} \\ -1, & \text{если представление } \lambda \text{ кватернионного типа.} \end{cases}$$

Существует замечательная формула Германа Вейля, позволяющая этот индекс сосчитать:

$$\operatorname{ind}\lambda = \frac{1}{\#G} \sum_{g\in G} \operatorname{tr} \pi_\lambda(g^2).$$

В качестве задачи я предлагаю доказать это равенство самостоятельно. Сделать это непросто. Даже не все специалисты по теории представлений, которых я спрашивал, могли это равенство сразу доказать, если они его раньше не знали. Поэтому здесь необходимо указание. Я не знаю, существует ли много разных доказательств. То доказательство, которое я сам придумал, мне очень нравится, поэтому я хочу его пропагандировать.

Рассмотрим пространство $\mathbb{C}[G]$ всех комплекснозначных функций на нашей группе. Для любой конечной группы существует преобразование Фурье, которое отображает это пространство в пространство матричных функций на двойственном объекте, на $\widehat{G}$. Обозначим это пространство $\mathbb{C}\mathrm{Mat}[\widehat{G}]$. Это соответствие взаимно однозначное, как и полагается преобразованию Фурье. А точная его запись такова: функции $f \in \mathbb{C}[G]$ соответствует функция

$$\tilde{f}(\lambda) = \sum_{g\in G} f(g)\pi_\lambda(g).$$

Для каждого класса $\lambda$ представление $\pi_\lambda$ сопоставляет элементу $g$ матрицу $\pi_\lambda(g)$ порядка $d(\lambda)$. Можно проверить, что если взять не конечную группу, а окружность или прямую, то эта процедура дает, соответственно, ряд Фурье или интеграл Фурье.

Из взаимной однозначности преобразования Фурье следует, в частности, совпадение размерностей пространств $\mathbb{C}[G]$ и $\mathbb{C}\mathrm{Mat}[\widehat{G}]$. Это, кстати сказать, и есть тождество Бернсайда. Преобразование Фурье является не только изоморфизмом пространств, но и изоморфизмом алгебр: свертка функций переходит в произведение матриц.

Рассмотрим в пространстве $\mathbb{C}[G]$ линейный оператор $V$, который переводит функцию $f$ в функцию $\check{f}$, где по определению $\check{f}(g) = f(g^{-1})$. Вычислим след этого оператора двумя разными способами. А именно, можно вычислить след оператора в исходном пространстве $\mathbb{C}[G]$, выбрав там естественный базис $\{\chi_g\}$: функция $\chi_g$ равна 1 на элементе $g$ и равна 0 на всех остальных элементах. След — это сумма диагональных элементов, поэтому он равен количеству тех элементов группы, которые равны своему обратному элементу: только эти базисные элементы переходят сами в себя, а все прочие базисные элементы переходят в другие базисные элементы и дают только недиагональные элементы матрицы. Равенство $g = g^{-1}$ можно записать в виде $g^2 = 1$. Преобразования, квадрат которых тождествен, называют инволюциями. Итак, если $\mathrm{Inv}(G)$ — множество всех инволюций в группе $G$, то $\mathrm{tr}\, V = \#\,\mathrm{Inv}(G)$.

По-другому вычислить след оператора $V$ можно, сделав преобразование Фурье. Тогда элементами пространства будут уже не функции на группе, а наборы матриц. Количество этих матриц равно количеству неприводимых представлений, а их порядки равны размерностям неприводимых представлений. Функции $f$ соответствует набор матриц $\tilde{f}(\lambda_1), \ldots, \tilde{f}(\lambda_n)$ порядков $d_1 = d(\lambda_1), \ldots, d_k$. Выясним, как оператор $V$ действует на эти матрицы.

Если представление вещественного типа, то оператор $V$ переводит матрицу в транспонированную матрицу: $A \mapsto A^t = {}^T\!A$. Легко указать, какой вклад в след этого оператора дает каждое вещественное представление. Недиагональные элементы переходят в другие недиагональные элементы и не дают вклада, а диагональные элементы дают вклад, равный 1. Поэтому каждое вещественное представление дает вклад, равный своей размерности.

С представлениями комплексного типа разобраться еще проще: они не дают вообще никакого вклада. Представления $\overline{\pi}_\lambda$ и $\pi_{\overline{\lambda}}$ не эквивалентны. Преобразование $V$ просто переставляет значения в разных точках: представления $\lambda$ и $\overline{\lambda}$ не эквивалентны, это две разные точки пространства $\widehat{G}$. Преобразование $V$ переставляет совершенно разные элементы матрицы, и это не дает никакого вклада в след.

С представлениями кватернионного типа разобраться сложнее. Это нужно считать, нужно знать, как кватернионы после комплексификации дают матрицы второго порядка, и что с ними происходит. Я скажу без доказательства результат. Каждое комплексное представление кватернионного типа дает вклад, равный своей размерности со знаком минус.

Сравнивая всё это с определением индекса, мы видим, что в целом получается следующий результат:

$$\# \operatorname{Inv}(G) = \operatorname{tr}(V) = \sum_{\lambda\in\widehat{G}} \operatorname{ind}(\lambda) d(\lambda).$$

Число инволюций в группе выражается в виде знакопеременной суммы размерностей неприводимых представлений. Особенно просто эта формула выглядит в случае, когда все представления группы вещественны. Тогда индекс тождественно равен 1, и получается просто сумма размерностей неприводимых представлений. Таким образом, если все неприводимые представления вещественны, то

$$\# \operatorname{Inv}(G) = \sum_{\lambda\in\widehat{G}} d(\lambda).$$

Всё это было присказкой к определению первой последовательности многочленов. Сегодняшняя моя лекция будет состоять из определений пяти последовательностей многочленов и формулировки гипотезы о том, что эти пять последовательностей совпадают. Я уже почти готов к тому, чтобы определить первую последовательность многочленов, но все-таки почти. Нужны еще некоторые сведения о конечных полях. На прошлой лекции я ввел обозначение $\mathbb{F}_q$ для конечного поля из $q$ элементов, где $q = p^k$ и $p$ — простое число. В дальнейшем мы будем рассматривать многочлены от $q$. И довольно странно было бы, что буква, изображающая независимую переменную, могла бы принимать только отдельные выбранные значения. Есть разные попытки интерпретировать значения этих многочленов для $q \neq p^k$. Но это задача не только не решенная, но даже и не сформулированная. Поэтому я пока про нее говорить не буду. Хотя, как отмечал Арнольд в своей книжке, самые интересные задачи — это те задачи, которые еще даже и не сформулированы. Вот это — пример не сформулированной задачи.

Рассмотрим группу $G_n(\mathbb{F}_q)$, состоящую из верхнетреугольных матриц $g$ порядка $n$, у которых на диагонали стоят единицы, а над диагональю стоят элементы поля $\mathbb{F}_q$. Порядок этой группы равен $q^{n(n-1)/2}$, потому что каждый из $n(n-1)/2$ наддиагональных элементов может принимать ровно $q$ различных значений.

Вообще говоря, для этой группы возможны представления всех трех типов — вещественные, комплексные и кватернионные. Но если $q = 2^l$, то при малых $n$ все представления этой группы вещественны, а при $n = 13$ есть пример не вещественного представления. В первом приближении мы закроем глаза на этот факт и будем считать, что все представления вещественны (в случае, когда $q = 2^l$). Тогда равенство

$$\# \operatorname{Inv}(G) = \sum_{\lambda\in\widehat{G}} d(\lambda)$$

дает выражение для числа инволюций в группе через размерности неприводимых представлений. А для поля из четного числа элементов инволюции очень легко вычислить. А именно, запишем матрицу $g$ в виде $g = 1_n + X$, где $1_n$ — единичная матрица порядка $n$. Тогда $g^2 = 1_n + 2X + X^2 = 1_n + X^2$, так как в поле из четного числа элементов $2X = 0$. Поэтому уравнение $g^2 = 1_n$ эквивалентно уравнению $X^2 = 0$.

Поставим такую задачу: сколько существует верхнетреугольных матриц $X$ с элементами из поля $\mathbb{F}_q$, для которых $X^2 = 0$.

Теперь можно дать определение первой последовательности многочленов. Определим $A_n(q)$ как число решений уравнения $X^2 = 0$ в верхнетреугольных матрицах порядка $n$ с элементами из поля $\mathbb{F}_q$. Здесь имеются в виду не только четные $q$, а и все другие $q$, для которых существует поле из $q$ элементов.

**Задача.** Доказать, что $A_n(q)$ — многочлен от $q$.

Это — хорошая комбинаторная задача. А именно, у вас есть некоторый конечный объект, и нужно вычислить количество точек в этом объекте. Это и есть основная задача комбинаторики. Комбинаторика в этом смысле не является вполне наукой, потому что для каждой задачи изобретается своя собственная теория. Но в этом смысле многие другие науки тоже науками не являются. Когда-то Юрий Иванович Манин объяснял мне, что алгебраическая геометрия не является наукой, потому что она состоит из набора решенных задач, а каждая задача требует своего собственного метода. Есть, конечно, какие-то общие понятия и общие приемы, но, как правило, они действуют в двух или самое лучшее в трех случаях, а так, чтобы больше, то такое редко бывает. Наиболее интересные результаты в алгебраической геометрии относятся к индивидуальным приемам. Вот так и комбинаторика. Она тоже умеет решать несколько конкретных задач. Слово «несколько» может здесь означать несколько тысяч, но тем самым имеется несколько тысяч конкретных задач, которые умеет решать комбинаторика.

Как же решить нашу конкретную комбинаторную задачу? Естественно попытаться получить рекуррентное выражение $A_{n+1}(q)$ через $A_n(q)$. Но я этого делать не умею, и, насколько мне известно, этого никто пока делать не умеет. Тогда остается один из немногих стандартных приемов, а именно, давайте усложним задачу, тогда решить ее будет легче. Разобьем множество всех решений на типы по рангу матрицы:

$$A_n^r(q) = \{X \in A_n(q) \mid \operatorname{rk} X = r\}.$$

Казалось бы, задача усложнилась, потому что вместо одного числа мы должны вычислить сразу много чисел. С другой стороны, как и во многих других случаях, такое усложнение приводит к упрощению, потому что для чисел $A_n^r(q)$ уже существует рекуррентное соотношение.

Это рекуррентное соотношение получается следующим образом. Рассмотрим матрицу $\mathscr{X}=\begin{pmatrix}X & x\\ 0 & 0\end{pmatrix}$. Ясно, что $\mathscr{X}^2=\begin{pmatrix}X^2 & Xx\\ 0 & 0\end{pmatrix}$. Поэтому $\mathscr{X}^2=0$ тогда и только тогда, когда $X^2=0$ и $Xx=0$. Число решений уравнения $X^2=0$ мы, по предположению, уже знаем: оно было получено на предыдущем шаге. А число решений уравнения $Xx=0$ зависит от ранга матрицы $X$. Поэтому в общем случае число решений этого уравнения не известно. Но если известно, что $\operatorname{rk} X=r$, то число решений тоже известно. Более того, зная кое-что из линейной алгебры, нетрудно догадаться, каким будет ранг матрицы $\mathscr{X}$. Он может либо не измениться, либо увеличиться на 1. Что именно происходит, зависит от соотношения между $X$ и $x$, которое легко контролируется. На этом я заканчиваю свои объяснения и выписываю рекуррентное соотношение:

$$A_{n+1}^{r+1}(q)=q^{r+1}A_n^{r+1}(q)+(q^{n-r}-q^r)A_n^r(q).$$

Теперь можно забыть о том, что $q$ принимает лишь весьма специальные значения, и пользоваться рекуррентным соотношением. Тогда оно определяет многочлены $A_n^r$, которые можно по очереди вычислять, если знать самый первый многочлен $A_0^0$. А он, естественно, равен 1. Получается таблица многочленов совершенно общего вида, у которых все коэффициенты отличны от нуля и про которые ничего хорошего сказать нельзя. Но если рассмотреть $\sum\limits_r A_n^r(q)=A_n(q)$, то происходит магическое сокращение, почти все члены взаимно уничтожаются. В качестве доказательства я нарисую таблицу многочленов $A_n$:

$$\begin{aligned}
&A_0=1, && A_6=5q^9-5q^7+q^5,\\
&A_1=1, && A_7=14q^{12}-14q^{11}+q^7,\\
&A_2=q, && A_8=14q^{16}-20q^{14}+7q^{12},\\
&A_3=2q^2-q, && A_9=42q^{20}-48q^{19}+8q^{15}-q^{12},\\
&A_4=2q^4-q^2, && A_{10}=42q^{25}-75q^{23}+35q^{21}-q^{15},\\
&A_5=5q^6-4q^5, && A_{11}=132q^{30}-165q^{29}+44q^{25}-10q^{22}.
\end{aligned}$$

Я напомню смысл $A_2(q)$ — это число решений уравнения $X^2=0$ в верхнетреугольных матрицах второго порядка. Равенство $X^2=0$ выполняется для всех верхнетреугольных матриц второго порядка (с нулями на диагонали). Это и соответствует тому, что $A_2(q)=q$.

В табличке сначала идут 3 многочлена с одним мономом, затем идут 3 многочлена с двумя мономами, затем 3 многочлена с тремя мономами, затем 3 многочлена с четырьмя мономами. Примерно два дня, или даже три, я смотрел на эту таблицу многочленов (она была выписана до размерности 20) и нашел в ней много свойств, которые позволяли продолжать эту таблицу неограниченно. И только потом догадался, какой правильный ответ.

Старшие коэффициенты многочленов — числа Каталана. Все знают, как записывается треугольник Паскаля. Давайте сделаем такой же треугольник, но только поставим зеркало и запретим за это зеркало заходить:

$$\begin{array}{ccccccc}
1 & & & & & & \\
& 1 & & & & & \\
1 & & 1 & & & & \\
& 2 & & 1 & & & \\
2 & & 3 & & 1 & & \\
& 5 & & 4 & & 1 & \\
5 & & 9 & & 5 & & 1
\end{array}$$

В первом столбце стоят как раз числа Каталана. Но зеркало мешает. Хорошо бы его убрать, а закон сохранить. Как хорошо известно из курса физики, зеркало заменяется отражением. В зеркальном мире еще нужно поменять знаки. Тогда получается треугольник, который образуется в точности по тому же правилу, что и треугольник Паскаля. На самом зеркале всюду стоят нули:

$$\begin{array}{ccccccccccccccc}
& & & & & & -1 & & 1 & & & & & & \\
& & & & & -1 & & 0 & & 1 & & & & & \\
& & & & -1 & & -1 & & 1 & & 1 & & & & \\
& & & -1 & & -2 & & 0 & & 2 & & 1 & & & \\
& & -1 & & -3 & & -2 & & 2 & & 3 & & 1 & & \\
& -1 & & -4 & & -5 & & 0 & & 5 & & 4 & & 1 & \\
-1 & & -5 & & -9 & & -5 & & 5 & & 9 & & 5 & & 1
\end{array}$$

Легко сообразить, что это просто разность двух треугольников Паскаля: один треугольник Паскаля растет из 1, а другой — из $-1$.

Теперь я предлагаю вам продолжить треугольник Каталана и сравнить его с табличкой многочленов. Это прекрасный пример экспериментальной математики, когда можно обнаружить формулу. Но это еще не значит, что ее можно доказать. Доказана она была сравнительно недавно, и это потребовало усилий двух крупных комбинаториков. Они использовали для этого разработанную ими ранее теорию. Сейчас имеется книжка под названием «$A = B$». Содержание этой книжки состоит в построении некоего алгоритма: на входе пишется формула $A = B$, а на выходе пишется доказательство этой формулы при условии, что формула верная. Один из авторов этой книги является моим партнером по университету в Филадельфии. Когда я ему сообщил, что у меня есть формула, которую я не могу доказать, он ответил, что у нас есть книжка, в которой доказана эта формула. Я предложил им засунуть мою формулу в их книгу. Они засунули. Алгоритм сразу не сработал, но они поднатужились и через некоторое время довели доказательство до конца.

Если считать, что в первой строке треугольника Каталана стоят числа $c_{1,-1} = -1$ и $c_{-1,1} = 1$, то для чисел $c_{k,l}$ можно написать рекуррентную формулу

$$c_{k,l} = c_{k-1,l-1} + c_{k-1,l+1}.$$

Чтобы рисовать треугольную таблицу, нужно иметь два индекса, причем хорошо бы, чтобы эти индексы имели одинаковую четность.

Введем теперь новую последовательность многочленов, которую я в честь Каталана обозначаю $C_n$:

$$C_n(q) = \sum_{s \equiv n+1(2) \equiv (-1)^n(3)} c_{n+1,s}\, q^{\frac{n^2}{4} + \frac{1-s^2}{12}}.$$

Это — экспериментальное правило, какие числа из треугольника Каталана нужно взять, чтобы получить коэффициенты полиномов $A_n$.

**Теорема 1.** $A_n(q) = C_n(q)$.

Эта теорема доказана с помощью книжки «$A = B$».

Мы уже выполнили значительную часть программы — определили две последовательности многочленов. Про многочлены $C_n$ мы знаем явную формулу, но не знаем никакой другой интерпретации. Про многочлены $A_n$ мы знаем интерпретацию: если $q$ — степень двойки, то $A_n(q)$ — это число решений уравнения $X^2 = 0$ в верхнетреугольных матрицах над полем $\mathbb{F}_q$. Но мы не знаем для этих многочленов явной формулы. И вот теперь есть теорема, которая дает и явную формулу и интерпретацию. Я, кстати сказать, не очень удовлетворен доказательством этой теоремы, хотя бы потому, что я не могу ни слова понять в этом доказательстве. Поэтому если кто-нибудь придумает более прозрачное доказательство, то это будет вполне содержательный шаг в науке, и мы с удовольствием это где-нибудь опубликуем.

Наряду с треугольником Паскаля есть гораздо более хитрый треугольник, о котором я лично узнал из лекции Арнольда, прочитанной 4 года назад. А из новой книжки Арнольда я узнал, что этот треугольник известен уже более ста лет. Но кому известно и где известно, не сказано. Косвенно можно сделать вывод, что не Арнольд это придумал, потому что ему ста лет еще нет. Это — так называемый треугольник Эйлера—Бернулли. Он строится способом, похожим на построение треугольника Паскаля, но с помощью челночного движения. Правило такое: нужно двигаться по нечетным строкам слева направо, по четным строкам справа налево, и каждый раз нужно начинать с нуля. А во всем остальном нужно следовать правилам треугольника Паскаля.

Нулевая строка состоит из одного элемента 1. Первая строка состоит из двух элементов 0 и 1. Вторую строку начинаем уже не слева, а справа. Ставим 0 и движемся влево: берем сумму предыдущего числа той же строки и предыдущего числа строки сверху, т. е. числа строго справа и числа справа вверху. Затем начинаем строку слева. Сначала всегда ставим 0, а затем берем суммы пар полученных ранее чисел:

$$
\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 0 & & 1 & & & & & \\
 & & & & 1 & & 1 & & 0 & & & & \\
 & & & 0 & & 1 & & 2 & & 2 & & & \\
 & & 5 & & 5 & & 4 & & 2 & & 0 & & \\
 & 0 & & 5 & & 10 & & 14 & & 16 & & 16 & \\
61 & & 61 & & 56 & & 46 & & 32 & & 16 & & 0
\end{array}
$$

В левой части треугольника через раз по определению идут нули. А оставшиеся числа $(1, 1, 5, 61, \dots)$ известны в науке. Они называются числами Эйлера. В любом справочнике по специальным функциям есть таблица чисел Эйлера. Они обозначаются $E_n$ и определяются соотношением

$$\sum_{n\geqslant 0} E_n \frac{x^n}{n!} = \frac{1}{\cos x}.$$

Кажется, первоначальное определение Эйлера было как раз такое.

Если взять правую диагональ, то мы получим другие числа, а именно, нормированные числа Бернулли. Сами числа Бернулли, как известно, рациональные. А у нас получаются так называемые нормированные числа Бернулли, которые являются целыми. Нормированное число Бернулли $\widetilde{B}_n$ получается из обычного числа Бернулли $B_n$ по формуле

$$\widetilde{B}_n = B_n \frac{2^{2n}(2^{2n}-1)}{2n}.$$

Нормированные числа Бернулли определяются соотношением

$$\sum_{n>0} \widetilde{B}_n \frac{x^n}{n!} = \operatorname{tg} x.$$

Это факт, который обычно скрывают от студентов мехмата, что не только синус и косинус имеют простые разложения в ряд Тэйлора, но и $1/\cos x$ и тангенс тоже имеют разложения в ряд Тэйлора, которые можно описать в явном виде (через числа Эйлера и через числа Бернулли).

Отдельно числа Эйлера и числа Бернулли считать трудно, а весь треугольник вместе считать легко.

Можно написать производящую функцию для всего треугольника. Это тоже хорошее упражнение.

Тот факт, который я в эту науку внес, состоит в том, что правильное понимание треугольника Эйлера—Бернулли состоит в том, что нужно заменить числа, туда входящие, многочленами. Более того, многочленами от двух переменных $q$ и $t$. Заменять числа многочленами — хорошо известный принцип математики. Его иногда называют принципом $q$-аналога. Многие математические величины, которые первоначально кажутся целыми числами, на самом деле нужно интерпретировать как многочлены. Один из способов это объяснить такой. Целое число в математике встречается чаще всего как размерность некоторого пространства. С другой стороны, пространства часто бывают градуированными, т. е. представляются в виде суммы подпространств, занумерованных целыми числами. Тогда, чтобы учесть градуировку, нужно писать не глобальную размерность всего пространства, а писать отдельно размерность каждой однородной компоненты. А это уже можно кодировать многочленами: свободный член означает размерность пространства с градуировкой 0, коэффициент при $q$ — размерность пространства с градуировкой 1, и т. д. Поэтому каждый раз, когда у вас есть градуированное пространство, вместо его размерности можно рассматривать градуированную размерность, т. е. многочлен.

Хорошо известный пример — это $n!$. Правильной интерпретацией числа $n!$ является многочлен $\prod\limits_{k=1}^{n} \dfrac{1-q^k}{1-q}$. Степень этого многочлена равна $n(n-1)/2$. При $q=1$ этот многочлен равен $n!$. (Если рассматривать дроби $\dfrac{1-q^k}{1-q}$, то нужно вычислять предел при $q$ стремящемся к 1.) В отличие от самого $n!$ выражение $\prod\limits_{k=1}^{n} \dfrac{1-q^k}{1-q}$ имеет глубокий хорошо известный геометрический смысл — это размерность группы гомологий пространства флагов. Группы гомологий градуированы по размерности, поэтому вместо числа возникает многочлен, который задается этой формулой. Пространство флагов — это одно из самых замечательных многообразий, которые сейчас имеются. Про него очень много всего известно. Это в какой-то степени объясняет, почему $n!$ встречается в разных формулах. Почему, например, $x^n$ любит, чтобы его делили на $(n!)$? Потому, что пространство флагов существует. Я не буду сейчас это объяснять. Есть масса книг, в которых это объясняется.

Я объясню еще один пример замены чисел на многочлены. Давайте возьмем другой пример — число сочетаний

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Это выражение имеет $q$-аналог. Например, каждый факториал можно заменить приведенным выше многочленом. На самом деле $q$-аналог биномиального коэффициента был придуман раньше, чем $q$-аналог числа $n!$. Это было придумано Гауссом следующим образом. Что такое число сочетаний из $n$ по $k$? Это число $k$-точечных подмножеств в $n$-точечном множестве. Но это объект плохой — конечные множества никакой структуры не имеют. Это слишком слабый объект, чтобы им занимались математики. Давайте возьмем объект поинтереснее, например, линейные пространства. Но если взять вещественные или комплексные пространства, то объект будет бесконечным, сосчитать число точек будет невозможно. Выход такой — возьмем пространство над конечным полем. Тогда в нем будет уже конечное число точек. Что является правильным аналогом $q$-точечного подмножества в $n$-точечном множестве? Этим аналогом будет $q$-мерное подпространство в $n$-мерном пространстве над полем $\mathbb{F}_q$. Число $q$-мерных подпространств в $n$-мерном пространстве над полем $\mathbb{F}_q$ является многочленом от $q$. При $q = 1$ этот многочлен равен биномиальному коэффициенту. В пределе можно понимать так: $n$-мерное линейное пространство над полем из одного элемента — это просто $n$ точек. Это одна из возможных интерпретаций предела $q \to 1$, но далеко не единственная.

Существует еще по крайней мере два или три разных способа заменять число на многочлен. Этот многочлен по традиции считается многочленом от $q$. Но в разных подходах буква $q$ играет совершенно разную роль. Только что $q$ было числом элементов конечного поля. В примере с пространством флагов число $q$ было размерностью группы гомологий. Бывают и другие ситуации. Иногда $q$ — это корень из единицы определенной степени. Иногда — малый параметр. Пожалуй, самый замечательный факт состоит в том, что в разных подходах возникают одни и те же выражения. Вот это уже область экспериментальной математики. Я объясняю это тем, что число разумных выражений невелико, а природа необъятна. Поэтому она обязана пользоваться одним и тем же выражением много раз, просто потому, что не хватает разумных выражений для всех природных явлений.

После всего этого отступления я хочу сказать, каким образом нужно превратить треугольник Эйлера—Бернулли в треугольник, элементами которого являются многочлены от $q$. Аккуратное определение требует многочлена от двух переменных $q$ и $t$. Но из тех соображений, о которых я кратко сказал на предыдущей лекции, нужно положить $t = q$ и тогда получается многочлен от одной переменной. Сейчас я напишу правило, по которому строится $q$-аналог треугольника Эйлера—Бернулли. Небольшая модификация правила образования треугольника Эйлера—Бернулли приводит к правилу

$$b_{k,l} = q^{-1} b_{k-1,l+1} + (q^{l+1} - q^l) b_{l,k-1}$$

для $k > 0$. Кроме того, $b_{0,l} = q^l b_{l-1,0}$ и $b_{0,0} = 1$.

Набор этих правил позволяет восстановить весь треугольник. Обратите внимание, что в рекуррентной формуле у последнего члена индексы $k$ и $l$ поменялись местами. Это — следствие челночного прохода: каждое число зависит от предыдущих чисел либо слева, либо справа. От этого $k$ и $l$ меняются местами.

Здесь я хочу сформулировать две нерешенные задачи.

1. Рекуррентное соотношение есть, треугольник можно написать, но формула для него не известна.

2. Не известна даже асимптотика роста членов треугольника.

Вторая задача, пожалуй, наиболее интересна: насколько быстро растут члены треугольника при фиксированном $q$.

Третья последовательность многочленов определяется так: $B_n(q) = b_{n-1,0}(q)$.

Теперь из пяти последовательностей многочленов мы определили три. Остались самый интересные многочлены $D_n$ и $Y_n$. Чтобы заинтриговать слушателей, я сразу напишу для $D_n$ формулу:

$$D_n(q) = \zeta_{G_n(\mathbb{F}_q)}(-1), \quad \text{где } q = 2^l.$$

Здесь $G_n(\mathbb{F}_q)$ — группа строго верхнетреугольных матриц порядка $n$ с элементами из $\mathbb{F}_q$.

Я напомню определение классической дзета-функции Римана: $\zeta(s) = \sum_{n\geqslant 1} n^{-s}$. Эта функция очень знаменита. Последняя нерешенная великая задача в математике — нахождение нулей дзета-функции. По известной гипотезе Римана они лежат на прямой $\operatorname{Re} s = 1/2$. Поэтому многие люди пытались обобщить дзета-функцию, чтобы примазаться к славе Римана. Специалисту по теории представлений обобщить дзета-функцию ничего не стоит, потому что натуральные числа — это размерности неприводимых представлений простейшей некоммутативной компактной группы, а именно $SU(2)$. Эта группа имеет ровно одно неприводимое представление в каждой размерности (с точностью до эквивалентности). Поэтому если у вас имеется любая группа $G$, у которой неприводимые представления конечномерные (например, конечная группа или компактная группа), то вы можете определить дзета-функцию, связанную с этой группой:

$$\zeta_G(s) = \sum_{\lambda\in\widehat{G}} d(\lambda)^{-s}.$$

Как ни странно, эта функция до сих пор не привлекала к себе большого внимания. Может быть, из-за того, что здесь содержательные примеры либо сводятся к обычной дзета-функции, либо про них нельзя сказать ничего содержательного. Хотя, конечно, некоторые известные тождества можно переписать как тождества для значений дзета-функции. Например, тождество Бернсайда, которое я выписывал, означает, что для конечной группы $G$ выполняется равенство

$$\zeta_G(-2) = \#G.$$

В начале лекции я обсуждал, что для конечных групп еще имеет смысл

$$\zeta_G(0) = \#\widehat{G}.$$

Кроме того, иногда имеет смысл

$$\zeta_G(-1) = \sum d(\lambda).$$

А именно, в некоторых случаях выполняется равенство $\zeta_G(-1) = \# \operatorname{Inv} G$. В определении многочлена $D_n$ участвует как раз $\zeta_G(-1)$.

При нашем определении совершенно не ясно, почему $D_n$ — многочлен.

Последнее, что я хочу сказать, это определение пятой последовательности многочленов $Y_n$. Эти многочлены связаны с описанием коприсоединенных орбит для группы верхнетреугольных матриц над конечным полем. Что такое коприсоединенная орбита я объяснял на прошлой лекции. Это некоторая траектория, описываемая точкой под действием группы. Если бы поле было вещественным, то эти траектории были бы довольно простыми алгебраическими многообразиями. А есть замечательная теорема, которая не доказана в полной общности, что если у вас есть хорошее алгебраическое многообразие, т. е. множество, задаваемое алгебраическими уравнениями, и если эти уравнения имеют смысл над конечным полем (например, если коэффициенты уравнений — целые числа), то тогда число решений этих уравнений над полем $\mathbb{F}_q$ является многочленом от $q$.

Например, уравнение $x^2 + y^2 = 1$ над полем $\mathbb{F}_q$ имеет конечное число решений, и это число решений является многочленом от $q$. Это — очень красивая задача. Я очень рекомендую тем, кто хочет поближе познакомиться с конечными полями, попробовать эту задачу решить.

Есть еще гипотеза, которая для многообразий размерности 1 была высказана Андре Вейлем, одним из самых замечательных математиков нашего века, и доказана сравнительно недавно другим замечательным математиком, Пьером Делинем. Гипотеза состоит в том, что число решений является многочленом, коэффициенты которого выражаются через структуру множества решений над комплексным полем.

То же самое уравнение $x^2 + y^2 = 1$ над комплексным полем задает множество, гомеоморфное $\mathbb{C} \setminus \{0\}$. Естественная компактификация этого множества — двумерная сфера $S^2 \simeq \mathbb{C}P^1$.

К сожалению, в полном объеме эта теорема не доказана. Не известно, какие алгебраические многообразия обладают таким замечательным свойством, что число точек этого алгебраического многообразия над полем $\mathbb{F}_q$ является многочленом от $q$. Не известно также, какой смысл имеют коэффициенты этого многочлена. Но мнемоническое правило известно. Эти коэффициенты должны соответствовать размерностям групп гомологий этого многообразия над вещественным или комплексным полем. Точнее говоря, $q = 1$ отвечает комплексному полю, а $q = -1$ отвечает вещественному полю. Неформальное объяснение этому примерно такое. Многообразие в каком-то смысле склеивается из клеток, которые устроены как обычные аффинные пространства. Над вещественным полем это дает вклад в гомологии размерности $n$, равный $(-1)^n$ (здесь $n$ — размерность клетки). Над комплексным полем клетка дает вклад в гомологии размерности $2n$, всегда равный 1. А над конечным полем клетка размерности $n$ дает вклад, равный $q^n$.

Поэтому если многообразие допускает клеточное разбиение, согласованное с его структурой алгебраического многообразия (например, окружность допускает разбиение на точку и прямую), то мы получаем доказательство. Это «доказательство» очень простое, убедительное, но неправильное, потому что слишком мало многообразий допускает такое клеточное разбиение. А сам факт верен и в гораздо более общей ситуации, когда разбиения на клетки не существует.

В интересующем меня случае я не знаю, есть ли теорема, обеспечивающая это утверждение, или нет такой теоремы. И те специалисты, которых я об этом спрашивал, тоже не могут сказать сразу. Так вот, последний многочлен $Y_n(q)$ описывает число коприсоединенных орбит для нашей группы. Это число можно интерпретировать как число точек на некотором алгебраическом многообразии над конечным полем. И если к этому многообразию применимы те или иные теоремы, которые в этой области известны, то тогда $Y_n$ будет многочленом от $q$.

Многочлен $Y_n$ имеет интерпретацию в духе современной квантовой статистической физики. В этой теории запас слов довольно бедный. Нужно знать такие слова: состояние, гибсовское распределение, энергия и статсумма. Больше ничего знать не нужно. Система состояний — это нечто, когда рисуется решетка, система узлов, или что-то, где имеются клеточки, ребра, вершины, какое-либо двумерное образование или, в крайнем случае, трехмерное. Состояния обычно получаются, если каждому элементу системы приписать какое-либо конкретное значение, чаще всего 0 или 1.

В моем случае системой можно назвать строго верхнетреугольную матрицу, а состояние системы заключается в том, что в каждую клеточку над диагональю можно поставить элемент конечного поля. Еще интересно рассмотреть составную систему, в которую входят верхнетреугольные матрицы $X$ и $Y$ и нижнетреугольная матрица $F$. Количество состояний этой системы равно $q^{3n(n-1)/2}$. Это число довольно большое. Для первого интересного примера $q = 2$ и $n = 13$. Для него имеется $2^{228}$ состояний.

Когда у вас есть система и заданы состояния, нужно придумать некоторую функцию, которая была бы явно лучше всех остальных функций. Обычно, если такую функцию вы придумали, то вы называете ее энергией и обозначаете $H$. Потом берете сумму по всем состояниям $s$:

$$\sum_s e^{-\beta H(s)}.$$

Здесь $\beta$ — параметр, который обычно интерпретируют как величину, обратно пропорциональную температуре: $\beta = 1/kT$, где $T$ — абсолютная температура в градусах Кельвина, а $k$ — постоянная Больцмана. Сосчитав такую сумму, вы получаете функцию от $\beta$. При всей неуправляемости этого процесса, про такую функцию от $\beta$ иногда можно что-нибудь сказать. Например, иногда можно доказать, что эта функция гладкая. Тогда с точки зрения физики система неинтересная — у нее нет фазовых переходов. А иногда эта функция не является гладкой. Тогда у системы есть фазовые переходы. В очень редких случаях эту функцию можно сосчитать явно и увидеть фазовые переходы невооруженным глазом: ответ задается одной формулой на одном отрезке и другой формулой на другом отрезке.

Вернемся к нашему примеру. Какая наиболее интересная функция может быть построена из трех матриц, две из которых верхнетреугольные, а одна нижнетреугольная? Я утверждаю, что наиболее интересная функция — это $\mathrm{tr}(F\cdot[X, Y])$. Значения этой функции лежат в конечном поле. Но надо придать смысл не самой функции, а ее экспоненте. Что такое экспонента, здесь хорошо известно. Экспонента — это функция, удовлетворяющая функциональному уравнению: сложение переходит в умножение. Для конечного поля — это так называемый аддитивный характер $\chi$. Получаем сумму

$$\sum_{X,Y,F} \chi(\mathrm{tr}(F\cdot[X, Y])).$$

Есть предположение, что эта сумма равна $q^{n(n-1)/2}Y_n(q)$. Это означает, что функция очень сильно осциллирует и почти все члены взаимоуничтожаются. Поэтому считать сумму необычайно трудно. Обычно для вычисления таких сумм используют метод Монте-Карло, а для сильно осциллирующих функций он непригоден.

Теперь я могу сформулировать общую гипотезу: все 5 последовательностей многочленов на самом деле совпадают.

Что по этому поводу известно? Результаты таковы:

1. $A_n = C_n$ для всех $n$. Я повторяю, что доказательство этого утверждения имеется и опубликовано.[7)] Модный современный журнал, в котором оно опубликовано, физического тела не имеет, он не на бумаге печатается, а существует только в виде файлов. Но любой подписчик может иметь его на своем компьютере. При желании можно распечатать доказательство, что я и проделал, но, надо сказать, ничего из него не понял. Так что если кто-нибудь мне объяснит, как это доказать, не обращаясь к той статье, то будет очень приятно.

---

[7)] *Ekhad S. B., Zeilberger D.* The number of solutions of $X^2 = 0$ in triangular matrices over GF($q$) // Electronic J. Combin. **3** (1996), #R2.

Заинтересованный читатель может найти подробности в работах:

*Кириллов А. А.* О числе решений уравнения $X^2 = 0$ в треугольных матрицах над конечным полем // Функц. анализ и его прилож. **29** (1995), №1. С. 82–87.

*Kirillov A. A.* Merits and demerits of the orbit method // Bull. of the AMS **36** (1999), №4. P. 433–488. — *Прим. ред.*

2. Утверждение о том, что $A_n = B_n$, проверено экспериментально для $n \leqslant 26$. На мой взгляд, это не оставляет сомнений в том, что это верно всегда. Но науке известны разные примеры подобных совпадений. Например, многочлен $n^2 - n + 41$ для первых сорока значений $n$ дает простое число.

3. Равенство $A_n = B_n = C_n = D_n = Y_n$ экспериментально проверено для $n \leqslant 11$.

4. По-видимому, $A_n \neq D_n$ для $n = 13$. Это объяснялось на предыдущей лекции. Недавно выяснилось, что существует треугольная матрица порядка 13 над полем из двух элементов, которая не сопряжена своей обратной. Структура ее жордановых блоков такова: 4, 5, 4.

На этом я хочу закончить лекцию. Но если кто-нибудь заинтересуется этим и готов это обсуждать, то я всегда доступен по электронной почте:

`kirillov@math.upenn.edu`

*Д. В. Аносов*

# О развитии теории динамических систем за последнюю четверть века

*Расширенная запись трёх лекций, прочитанных весной 1998 года*

**Содержание**



## Предисловие

Настоящий очерк достижений теории динамических систем (ДС), смежных разделов теории обыкновенных дифференциальных уравнений и других математических дисциплин за последние примерно 25 лет[1)] является очень кратким и неполным, особенно по сравнению с изложением аналогичного материала за более ранние периоды в известных изданиях ВИНИТИ «Итоги науки» и «Современные проблемы математики. Фундаментальные направления» (более поздние из которых частично захватывают и наш период). В связи с этим сошлюсь на недавний обзорный доклад Йоккоза [1], написанный под другим углом зрения и в значительной степени дополнительный к

[1)] По ходу дела затрагиваются и работы более раннего времени, а для некоторых тем вообще более естественным является несколько бóльший промежуток времени, и было бы нецелесообразно искусственно «отсекать» его начало.

имеющемуся в этой статье перечню.[2] Обширный материал, в том числе и нового происхождения, имеется в объёмистой книге [2], недавно переведённой на русский язык.

Не только Йоккоз, но и ряд других докладчиков, делавших пленарные и большие секционные доклады на Международных математических конгрессах, говорили о динамических системах. Все эти доклады могут быть рекомендованы как авторитетные освещения различных сторон предмета, дающие и перспективу, и новейшие для своего времени сведения. Я выделяю особо доклад Йоккоза потому, что он отличается широтой охвата предмета. Позднее другой тематически довольно широкий доклад сделал Ю. Мозер [3].

Признаки, по которым отобран материал первых двух параграфов, явствуют из их названий. Этот отбор был осуществлён на основании чётких формальных критериев; в этом отношении, мне кажется, он свободен от субъективизма. Эти критерии надёжны в том смысле, что соответствующие результаты в любом случае заслуживали бы некоторого обсуждения: безусловно надо было сказать о возникновении или возрождении целых больших частей теории ДС; а если некая проблема упоминается вместе с именем известного учёного, за этим, скорее всего, стоит не просто тот факт, что он её поставил, но и признание её важности[3]. (Ведь редко знаменитое имя упоминается по поводу не столь уж значительных задач, которые носитель этого имени наверняка формулировал для своих студентов и аспирантов. Конечно, любая задача имеет своего автора, — в крайнем случае того же, кто её и решил, — но авторами проблем из §2 были весьма известные математики, в связи с чем и сами проблемы обычно упоминались и упоминаются вместе с их именами.)

Если говорить о более или менее широких новых (или обновлённых) направлениях, то мне кажется, что они все названы в §1, но что касается различных довольно значительных достижений, то они не всегда состоят в решении какой-нибудь «именной» проблемы и потому отнюдь не исчерпываются тем, о чём сказано в §2[4]. В §3 я вкратце упоминаю некоторые

---

[2] Прежде всего, я не писал или писал короче о том, о чём сказано у Йоккоза. Кроме того, понятно, что я лучше знаком с работами русских, а он — западных математиков.

[3] Начав с «именной» проблемы, я заодно говорю и о некоторых смежных вопросах. Иногда им отводится даже больше места, так что «именная» проблема используется как повод для освещения целой области.

[4] И не обязаны уступать по важности результатам §1, потому что события более мелкого «таксономического» уровня могут быть по своему внутреннему содержанию столь же значительными, как и более высокого уровня (достаточно напомнить о создании теории КАМ (Колмогорова—Арнольда—Мозера) и «гиперболической революции» в 60-е гг., которые не выходили за рамки гладких ДС).

из таких достижений. Такого чёткого критерия для отбора материала, как раньше, здесь не было, поэтому выбор этого материала неизбежно был неполным (я не стал писать даже о том, чем сам много занимался в последние годы — о потоках на поверхностях и смежных геометрических вопросах) и, вероятно, в какой-то степени субъективным. Кроме того, при всём желании я не мог бы написать в §3 обо всём столь же подробно, как в первых параграфах (если изложение в §§1, 2 можно назвать подробным), уложившись в разумные объём и срок. Но я надеюсь, что даже там, где я, по существу, ограничиваюсь простым упоминанием некоторых результатов и имён и даю как бы слегка аннотированный (и очень не полный) список литературы, даже и это может быть полезным как некоторая исходная информация о том, что делается на свете. Если читатель чем-то заинтересуется, даже те неполные литературные ссылки, которые я мог дать, помогут ему, по крайней мере, начать более серьёзно знакомиться с предметом.

Я старался учесть известные мне работы не только о гладких, но и о топологических или чисто-метрических (в смысле меры) системах. Но темы, освещаемые подробнее, оказались посвящёнными преимущественно гладким системам. Это не связано с тем, что именно они находятся в центре моих интересов, — надеюсь, что по крайней мере с наиболее яркими результатами о других системах я знаком (ведь яркие работы часто упоминаются в беседах с коллегами). Если бы я писал о более ранних работах, баланс был бы другим, но вот в эти годы с топологическими или чисто-метрическими системами не оказались связаны ни новые направления (за одним исключением), ни «именные» проблемы. Кое-что о них сказано только в конце §3, если не считать раздела о групповых действиях в §1.

Во время работы над этой статьёй я неоднократно наводил справки по различным вопросам у знакомых (а иногда и лично не знакомых) математиков. Прежде всего, конечно, это были участники моих семинаров в МИАН и МГУ, а также нижегородские математики[5)], но кое-какие сведения я получил и от других коллег, в том числе зарубежных. Благодарю их всех и надеюсь, что, распорядившись полученной информацией по своему разумению и в соответствии со своими планами, я её не переврал. (Если же переврал, то это моя вина.)

О литературных ссылках. Я старался ссылаться на последние публикации или на публикации, которые имеют обзорный характер или содержат более или менее подробное (не обязательно оригинальное) изложение некоего круга вопросов (то, что по-английски называется expository papers). Это позволило сэкономить, но прошу отнестись с пониманием к тому, что я порой экономил на ссылках на первопроходцев.

[5)] С. Смейл не случайно «записал» меня в нижегородскую школу, хотя знал, что я москвич.

Для понимания этой статьи нужна, во-первых, общематематическая подготовка в объёме примерно того, что сообщается студентам-математикам на первых трёх курсах университета (на старших курсах подготовка обычно становится специализированной) и, во-вторых, некоторое знакомство с ДС, преимущественно гладкими[6]. Мне пришлось излагать содержание этой статьи в лекциях для студентов, специальная часть подготовки которых включала именно гладкие ДС. Поэтому, говоря об эргодической теории, я делал больше пояснений, чем в других случаях. Эта особенность лекций сохранилась и в статье.

По эргодической теории довольно разнообразным по содержанию, хотя и не новым, является учебник [4].

[6] Помимо учебников, рекомендую вниманию читателя обзорные статьи в уже упоминавшихся изданиях ВИНИТИ (некоторые из них здесь цитированы, но не все), а также статьи в «Математической энциклопедии», где помимо определений и краткой сводки сведений приводятся довольно многочисленные литературные ссылки.

# §1. Новые или «обновлённые» направления

**1.1. Симплектическая геометрия.** К началу данного периода прочно сложилось подразделение теории ДС (включая сюда и смежные вопросы) на четыре основные части, характеризуемые наличием (и использованием) соответствующей структуры в фазовом пространстве, с которой в определённом смысле согласована рассматриваемая ДС, — дифференциальная динамика (теория гладких ДС)[7], топологическая динамика (теория топологических ДС), эргодическая теория (в которой фазовое пространство предполагается измеримым, чаще даже пространством с мерой) и аналитическая

---

[7] По существу, та же самая ветвь теории ДС известна под названием «качественной теории обыкновенных дифференциальных уравнений»; употребление того или иного названия — это скорее вопрос традиции той или иной научной школы. Ниже много раз встречается название «локальная теория». Так обычно называют часть теории гладких ДС, посвящённую поведению траекторий потока вблизи положения равновесия или периодического решения или траекторий итерируемого отображения вблизи неподвижной или периодической точки. В «глобальной теории» речь идёт о поведении траекторий во всём фазовом пространстве или по крайней мере в какой-то «более или менее обширной» его области.

Поскольку зашла речь о различии между локальной и глобальной теориями, то надо сказать, что оно отчасти условно: если координаты вектора фазовой скорости $\mathbf{v}(x)$ (локального) потока в $\mathbb{R}^n$ являются однородными многочленами степени $k$ от координат $x$, то исследование поведения траекторий возле начала координат (являющегося положением равновесия) в значительной степени эквивалентно глобальному исследованию некоторого потока в $(n-1)$-мерном проективном пространстве $\mathbb{RP}^{n-1}$, куда траектории «проектируются» по радиусам. (Собственно говоря, при проектировании траекторий в $\mathbb{RP}^{n-1}$ там получается не векторное поле, а поле направлений (касательных прямых). Но это хотя и требует известного внимания, в конечном счёте не нарушает сказанного ниже.) Таким путём можно получить (при подходящем $k$) потоки в $\mathbb{RP}^{n-1}$ достаточно общего вида, чтобы (при $n \geqslant 4$) для них реализовывались различные известные варианты сложного поведения траекторий, известные в глобальной теории. Можно также обеспечить, чтобы соответствующие траектории в $\mathbb{R}^n$ целиком лежали возле 0 и вели себя столь же сложно. Таким образом, $n$-мерная локальная теория как бы содержит внутри себя $(n-1)$-мерную глобальную или по крайней мере весьма значительную её часть и не может быть проще неё.

Но практически это принципиальное замечание не имеет большого значения. Дело в том, что при $k > 1$ положение равновесия 0 является вырожденным, причём коразмерность вырождения (характеризующая «его степень») быстро растёт с ростом $k$. Со столь вырожденными случаями локальная теория практически не имеет дела. Сказанное означает только то, что (по крайней мере в многомерном случае) локальная теория не может претендовать на полное исследование всех возможностей, ибо в глобальной ситуации они необозримы. Но таких претензий никогда и не высказывалось.

теория (в которой фазовое пространство и «время» — независимая переменная — предполагаются комплексными). Разумеется, встречались и другие структуры, но их учитывали, так сказать, не выходя за пределы той или иной из этих четырёх частей. Так, гамильтоновы системы имеют свою специфику, связанную с симплектической структурой, но их изучение всегда рассматривалось как часть теории гладких ДС.

За последние 20 лет возникла новая дисциплина — симплектическая геометрия (или симплектическая топология), которая в системе математических дисциплин имеет по меньшей мере тот же уровень, что и названные четыре традиционных подразделения теории ДС, а отчасти даже выходит за пределы последней. Новые дисциплины такого уровня формируются не часто, поэтому такое событие привлекает особое внимание.

Симплектическое многообразие — это гладкое многообразие $M$ вместе с заданной на нём внешней дифференциальной замкнутой 2-формой $\omega$, которая (во всех точках $M$) является невырожденной. Замкнутость $\omega$, как обычно для внешних дифференциальных 2-форм, означает, что $d\omega = 0$, где $d$ — внешний дифференциал. Понятие невырожденности вводится для любых (не только кососимметричных) билинейных форм на векторном пространстве (в данном случае — на касательном пространстве $T_xM$). Такая форма называется невырожденной, если невырождена её матрица коэффициентов. Здесь подразумевается использование координат; в линейной алгебре приводятся эквивалентные бескоординатные формулировки. Замечу сразу же, что кососимметрическая форма может быть невырожденной только в чётномерном случае ($\dim T_xM = 2n$). Наконец, подразумевается, что в терминах локальных координат коэффициенты $\omega$ являются гладкими функциями своих аргументов (насколько гладкими — это может зависеть от рассматриваемой задачи, но часто и даже, вероятно, в большинстве случаев можно считать их гладкими класса $C^\infty$).

Таким образом, симплектическое многообразие — это не просто многообразие $M$, а пара $(M, \omega)$. Часто всё-таки в обозначении опускают явное упоминание об $\omega$ и говорят о симплектическом многообразии $M$.

Приведённое определение аналогично определению риманова многообразия. В последнем вместо кососимметрической формы $\omega$ имеется симметричная билинейная дифференциальная форма $g$, обычно не только невырожденная, но и положительно определённая[8)] (и то, и другое возможно и при нечётной размерности), на которую не накладывается никаких дифференциальных условий (тогда как от $\omega$ требуется замкнутость).

---

[8)] Это значит, что положительно определена соответствующая квадратичная форма. Если же положительной определённости нет, то часто говорят о псевдоримановом многообразии и о псевдоевклидовом скалярном произведении в касательных пространствах.

Однако аналогия между римановой и симплектической геометрией кончается на уровне исходных определений. В римановом случае различные многообразия могут быть локально устроены по-разному: имеется обширная система локальных инвариантов — тензор кривизны и его ковариантные производные. В симплектическом же случае, согласно теореме Г. Дарбу[9], возле любой точки $x \in M$ имеются так называемые симплектические или канонические координаты (координаты Дарбу) $p_1, \ldots, p_n, q_1, \ldots, q_n$, в терминах которых $\omega$ локально выражается в виде

$$\omega = \sum_{i=1}^{n} dp_i \wedge dq_i. \tag{1}$$

Поэтому любые два симплектических многообразия $(M_1, \omega_1)$ и $(M_2, \omega_2)$ одинаковой размерности локально устроены одинаково: для любых двух точек $x_1 \in M_1$, $x_2 \in M_2$ имеются такие их окрестности $U_1$, $U_2$ и такой диффеоморфизм $f\colon U_1 \to U_2$, который переводит эти формы друг в друга (именно, $f^*\omega_2 = \omega_1$)[10]. Кстати, такой диффеоморфизм называют симплектоморфизмом. В данном случае он, в понятном смысле, локальный, но бывают и глобальные симплектоморфизмы — диффеоморфизмы $f\colon M_1 \to M_2$, для которых $f^*\omega_2 = \omega_1$.

Хотя никаких локальных инвариантов у самих симплектических многообразий нет (кроме размерности), подмногообразия симплектического многообразия могут быть различными и при совпадении размерностей. Так, большую роль играют лагранжевы подмногообразия — подмногообразия $N$ половинной размерности, ограничение на которые $\omega|_N$ формы $\omega$ тождественно равны нулю[11]; в то же время имеются подмногообразия той же размерности с ненулевым ограничением $\omega|_N$. Всё же и здесь инвариантов намного меньше, чем в римановой или евклидовой геометрии. Например, если $N_1$ и $N_2$ — подмногообразия симплектического многообразия $(M, \omega)$, $U_i \subset N_i$ — окрестности точек $a_i \in N_i$ и если существует диффеоморфизм $f\colon U_1 \to U_2$,

[9] Она была доказана также и Г. Фробениусом, но в общепринятом названии теоремы о нём почему-то не упоминают.

[10] В этом отношении ситуация аналогична с ситуацией для плоских римановых многообразий. Таким образом, имеется некоторая аналогия между условием замкнутости формы $\omega$ и условием обращения в нуль тензора кривизны. Однако она не кажется глубокой. Начать хотя бы с того, что замкнутость — это дифференциальное условие первого порядка, а нулевая кривизна — второго. Если это различие может показаться формальным, то вот и вполне содержательное различие: группа изометрий связного риманова многообразия конечномерна, а группа симплектоморфизмов симплектического многообразия бесконечномерна.

[11] Лагранжевы подмногообразия, как и ряд родственных объектов, неявно присутствовали в математическом аппарате классической аналитической механики, но в явном виде они были выделены В. П. Масловым и В. И. Арнольдом.

для которого $f(a_1) = a_2$ и $f^*(\omega|_{U_2}) = \omega|_{U_1}$, то существует симплектоморфизм $g\colon V_1 \to V_2$ некоторых окрестностей $V_i$ точек $a_i$ во всём $M$, локально продолжающий $f$ в том смысле, что он совпадает с $f$ на $V_1 \cap U_1$. В данном случае внутренняя геометрия подмногообразия локально полностью определяет его внешнюю геометрию (сравните это с кривыми в $\mathbb{R}^n$!)

Думаю, что эта скудность локальных инвариантов задержала возникновение симплектической геометрии, — заранее непонятно, чего, собственно, здесь изучать?

Симплектические многообразия встречаются в математике главным образом в связи с гамильтоновыми системами классической механики и в связи с кэлеровыми многообразиями. Обычная гамильтонова система — это система дифференциальных уравнений для $2n$ неизвестных $p_1, \ldots, p_n$ (импульсы) и $q_1, \ldots, q_n$ (координаты) вида

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \qquad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \tag{2}$$

где $H$ — некоторая функция от неизвестных и, возможно, времени (гамильтониан). Векторное поле фазовой скорости $V$ этой системы следующим образом связано с дифференциалом $dH$ и симплектической формой (1): для любого вектора $U$

$$\omega(V, U) = -dH(U) = -U \cdot H,$$

где в $U \cdot H$ вектор $U$ известным образом действует на $H$ как линейный дифференциальный оператор первого порядка (если этот оператор хотят отличать от вектора, то оператор обозначают через $D_U$ или $\mathscr{L}_U$). Иными словами, с помощью $\omega$ мы переходим от вектора $V$ к ковектору (линейному функционалу, линейной форме на векторах) $U \mapsto \omega(V, U)$; этот ковектор, с точностью до знака, совпадает с $dH$. В общем случае с функцией $H$ на симплектическом многообразии $(M, \omega)$ можно по точно такому же правилу связать векторное поле $V$ на $M$; оно называется глобально гамильтоновым векторным полем (с гамильтонианом $H$). В координатах Дарбу координаты (компоненты) этого поля суть в точности правые части системы (2). В классической механике исходная форма $\omega$ часто имеет вид (1), но даже и тогда при последующих преобразованиях (особенно когда с помощью симметрий понижают размерность) может произойти переход к другим $\omega$. (Локальный) поток с гамильтоновым векторным полем фазовой скорости тоже называют гамильтоновым. Такой поток $\{\varphi_t\}$ сохраняет форму $\omega$, т. е. отображения $\varphi_t$ являются симплектоморфизмами ($\varphi_t^*\omega = \omega$). Обратно, если поток $\{\varphi_t\}$ с полем фазовой скорости сохраняет $\omega$, то локально поток и поле являются гамильтоновыми — любая точка многообразия $M$ лежит в некоторой области $W$, в которой поле $V$ получается описанным выше способом из некоторой функции (локального гамильтониана) $H_W$, определённой в $W$. Но, вообще говоря, от системы локальных гамильтонианов $\{H_W\}$, которые определены в областях $W$, в совокупности покрывающих $M$, нельзя перейти к единому «глобальному» гамильтониану (который был бы определён на всём $M$ и всюду определял бы по описанному способу поле $V$).

Теперь о другом источнике симплектических многообразий. Здесь предполагается известным понятие комплексно-аналитического многообразия. Для такого многообразия $M$ естественно вместо обычной римановой метрики рассматривать так называемую эрмитову метрику. Именно, поскольку касательное пространство к комплексному многообразию $M$ само естественным образом является комплексным векторным пространством, то вместо евклидова скалярного произведения в нём естественно рассматривать эрмитово скалярное произведение (полуторалинейную форму по терминологии Бурбаки). Если такая форма задана во всех касательных пространствах и, будучи выражена в локальных координатах, имеет достаточно гладкие коэффициенты, то будем говорить об эрмитовой структуре на $M$ или об эрмитовом многообразии. Вещественная часть $g$ эрмитовой формы — это евклидово скалярное произведение; таким образом, эрмитово многообразие автоматически является также и римановым. (Здесь может всюду стоять приставка «псевдо».) А мнимая часть эрмитовой формы — это невырожденная кососимметрическая 2-форма $\omega$. Если она замкнута, то эрмитово многообразие называется кэлеровым. Такое определение, с одной стороны, является коротким; с другой стороны, именно оно главным образом и используется при работе с кэлеровыми многообразиями. Но поначалу оно кажется каким-то формальным, непонятно чем мотивированным. Поэтому я приведу и другое определение, геометрически достаточно естественное. Риманова метрика $g$ известным образом определяет связность Леви—Чивита на $M$. Последняя позволяет осуществлять параллельное перенесение векторов касательного пространства вдоль любой гладкой кривой $\gamma(t)$ на $M$. Получаются линейные отображения $T_{\gamma(t_1)}M \to T_{\gamma(t_2)}M$. Они сохраняют евклидово скалярное произведение, но поскольку мы имеем дело с комплексной ситуацией, то хотелось бы, чтобы они сохраняли и всё эрмитово скалярное произведение, а также были линейными не только над $\mathbb{R}$, но и над $\mathbb{C}$ (т. е., в понятном смысле, сохраняли бы умножение векторов на комплексные числа). Оказывается, это пожелание в точности эквивалентно кэлеровости.

Кэлеровых многообразий довольно много. На комплексном проективном пространстве имеется естественная кэлерова метрика (метрика Фубини—Штуди); она индуцирует кэлерову метрику и на алгебраических подмногообразиях этого пространства. (С другой стороны, существуют и такие кэлеровы многообразия, которые не являются алгебраическими многообразиями.)

Тот факт, что «симплектические» соображения и результаты составляют некий относительно самостоятельный и единый комплекс понятий и методов, неоднократно подчёркивал, начиная с середины 60-х гг., В. И. Арнольд, которому много пришлось иметь дело с этим комплексом, причём (в отличие от других исследователей) по различным поводам (гамильтоновы ДС, лагранжевы перестройки, асимптотические методы в теории уравнений с частными производными)[12]. Своего рода «предвестником» новой дисциплины, впоследствии органически вошедшим в её состав, была возникшая в предыдущем 20-летии теория лагранжевых и лежандровых перестроек и особенностей, к которой примыкает также ряд других вопросов. В этой области особенно много сделали В. И. Арнольд и его школа, а также А. Вайнстейн. Её изложение имеется в [5]. В принципиальном отношении на этом этапе, по сравнению с современным, недоставало открытия глобальных инвариантов самих симплектических многообразий (а не инвариантов каких-то связанных с ними объектов).

Стимулирующую роль в возникновении симплектической геометрии сыграла работа Ч. Конли и Э. Цендера, посвящённая доказательству для $n$-мерного тора $\mathbb{T}^n$ следующей гипотезы В. И. Арнольда: у симплектического диффеоморфизма $g_1$ компактного симплектического многообразия $M$, гомотопного тождественному диффеоморфизму $g_0$ в классе симплектических диффеоморфизмов посредством деформации $\{g_t\}$, скорость которой $\dfrac{dg_t}{dt}$ при всех $t$ имеет однозначный гамильтониан[13], существует по меньшей мере столько неподвижных точек, сколько существует критических точек у гладкой функции на $M$ (при $M = \mathbb{T}^n$ их $2n+1$, а если они невырожденные — $2^{2n}$). В доказательстве сочетались две ранее возникшие идеи (уже использовавшиеся к тому времени в других задачах): идея П. Рабиновица о новом вариационном подходе к периодическим решениям гамильтоновых систем, и идея Ч. Конли о топологической характеризации поведения потока возле некоторых (так называемых изолированных или локально-максимальных) инвариантных множеств посредством некоторого обобщения классического индекса Морса положений равновесия градиентного потока (сам Морс говорил о критических точках функции, что эквивалентно).

Теперь гипотеза Арнольда доказана и в ряде других случаев, а несколькими авторами анонсировано её доказательство в общем виде.

Последний шаг, после которого симплектическая топология уже несомненно приобрела характер автономной дисциплины, сделал М. Громов [6]. Прежде чем говорить о его подходе, укажу некоторые из его результатов.

---

[12] В конечном счёте, естественно, и у него появление симплектических многообразий происходило в духе указанных выше двух источников, но непосредственные причины обращения к ним были разнообразными.

[13] Это одна из формулировок условия на $g_1$, накладываемого в этой гипотезе. Чаще приводят другую формулировку, связанную с так называемыми асимптотическими циклами — гомологическим аналогом числа вращения Пуанкаре, относящегося к гомеоморфизмам окружности.

Громов ввёл несколько симплектических инвариантов (т. е. инвариантов относительно симплектоморфизмов). Единственным инвариантом, известным ранее, был объём. Полезность новых инвариантов видна из следующего утверждения Громова[14)], которое кажется удивительным и имеет наглядную формулировку, а также выразительное название «теорема о непрохождении симплектического верблюда через игольное ушко»: при $n > 1$ в $\mathbb{R}^{2n}$ с координатами $p_i, q_i$ $(i = 1, \ldots, n)$ не существует непрерывного семейства симплектоморфизмов $\{\psi_t;\ 0 \leqslant t \leqslant 1\}$, сохраняющих (1) на с. 80, которое переводило бы шар $B$ радиуса $R$ из полупространства $p_1 < 0$ в полупространство $p_1 > 0$ таким образом, чтобы в процессе деформации пересечение

$$\psi_t B \cap \{(p, q);\ p_1 = 0\}$$

всё время находилось бы строго внутри некоторого $(2n - 1)$-мерного шара радиуса $r \leqslant R$. (Осьминог, подчинённый только условию сохранения объёма, конечно, может проползти через маленькое отверстие, что подтверждается опытом содержания этих животных в неволе.) Отмечу также, не вдаваясь в детали, что с помощью своего подхода Громов получил неожиданную информацию о глобальных свойствах лагранжевых подмногообразий, даже подмногообразий обычного евклидова пространства (с координатами $p_i, q_i$ и формой (1)), когда a priori не видно, с какой стати они должны иметь какие-то особенные свойства.

Как уже говорилось, на кэлеровом многообразии $M$ имеется симплектическая структура, «хорошо» связанная с соответствующими комплексной и римановой структурами. По терминологии Громова, соответствующая псевдокомплексная структура[15)] на $M$ «доминируется» («is tamed by») симплектической (Громов точно описывает, какие именно соотношения он подразумевает под доминированием). Таких псевдокомплексных структур много, и всем известно, насколько настоящая комплексная кэлерова структура лучше всех прочих. Оказывается, что на симплектическом многообразии существует много псевдокомплексных структур, доминируемых симплектической; не пытаясь выделить из них какую-то одну «хорошую», Громов предложил рассматривать все их сразу (при этом естественно возникают и римановы метрики). Совокупность «плохих» структур оказалась в какой-то степени подходящим заменителем для одной «хорошей»! Для каждой из них можно говорить о псевдоголоморфных отображениях круга в наше многообразие

14) Первый набросок доказательства был дан Громовым совместно с Я. М. Элиашбергом.

15) То есть структура комплексного пространства во всех касательных пространствах $T_xM$, в понятном смысле гладко зависящая от $x$; в данном случае она получается из исходной комплексной структуры на $M$, но в общем случае может получаться и как-нибудь иначе.

(ради краткости говорят о псевдоголоморфных кругах) — они определяются путём непосредственного обобщения обычной голоморфности. Условие псевдоголоморфности записывается в виде некоторой квазилинейной системы уравнений в частных производных, обобщающей классическую систему Коши—Римана. Громов исследовал эту систему и доказал, грубо говоря, что, как и в классическом случае, она имеет много решений. Его симплектические инварианты определяются с помощью псевдоголоморфных кругов, отвечающих всевозможным псевдокомплексным структурам, доминируемым исходной симплектической.

Как видно, подход Громова связан не с ДС, а с уравнениями с частными производными. Но часть его результатов (а также идей и методов) оказывается полезной для теории ДС. Другими авторами дана более близкая к теории ДС трактовка этой части или, вернее, чего-то аналогичного. На этом пути получены лучшие результаты о периодических решениях гамильтоновых систем (К. Витербо, Х. Хофер и Э. Цендер; родственным примером может служить теорема Хофера, упомянутая в конце п. 2.5). Эта сторона дела хорошо отражена в [7]. Сборник [8] посвящён более широкому кругу вопросов симплектической топологии, в том числе там освещены и некоторые связи с ДС, не затронутые в [7]. (Это особенно касается инвариантных лагранжевых многообразий, в основном торов, для которых до проникновения в теорию ДС новых симплектических идей ряд вопросов качественного характера даже не ставился (не считая, конечно, частного случая инвариантных кривых двумерных диффеоморфизмов, о которых говорил ещё Дж. Биркгоф). Насколько известно, обзоров на последнюю тему нет. В дополнение к сказанному можно сослаться на [9], [10].) Учебником по собственно симплектической топологии является [11]. В последнее время доклады по симплектической геометрии (как в связи с ДС, так и независимо от них) делались на международных и европейских математических конгрессах, так что в соответствующих трудах можно найти много информации о состоянии дел в этой дисциплине.

Отметим ещё два примыкающих цикла работ (хотя они — особенно второй — именно только в той или иной степени примыкают к нашей теме).

Продолжая и развивая начатые ранее исследования, В. И. Арнольд и его сотрудники занялись группой вопросов, получивших название «нелинейной теории Морса» [12]. В ней речь идёт о свойствах колеблемости и пересечений некоторых кривых или более общих многообразий, которые можно рассматривать как обобщения известных теорем Штурма и Морса; последние при этом выступают как относящиеся к тем частным случаям, когда эти кривые являются графиками некоторых функций или даже решений линейных дифференциальных уравнений.

Второй цикл — это два доказательства теоремы, что на двумерной сфере с любой римановой метрикой существует бесконечное число замкнутых геодезических (прежний результат, в основном восходящий к работам Л. А. Люстерника и Л. Г. Шнирельмана конца 20-х гг., но «дотянутый до конца» позднее [13], [14] — три замкнутых геодезических; правда, эти три геодезических отличаются отсутствием самопересечений). Одно доказательство начинается с принадлежащей В. Бангерту редукции задачи к двум случаям, для которых доказательства были даны самим Бангертом и Дж. Френксом [15], [16]. Здесь можно отметить замечательное сочетание новизны и преемственности (а также своего рода «элементаризации»). На случай, который рассмотрел Френкс, обратил внимание ещё Дж. Биркгоф. Это тот случай, когда имеется такая «простая» (не имеющая самопересечений) замкнутая геодезическая $L$, что любая пересекающая $L$ геодезическая спустя некоторое время снова пересекает $L$. (Он охватывает, в частности, вполне классический случай метрик положительной кривизны.) Биркгоф отметил, что в этом случае вопрос сводится к исследованию некоторого отображения кольцеобразной области в себя. Френкс доказал, что данное отображение имеет бесконечное число неподвижных точек, а это и означает существование бесконечного числа замкнутых геодезических. (Как известно, самому Биркгофу принадлежит некоторый результат об отображениях кольца — доказательство гипотезы, высказанной А. Пуанкаре и известной под названием «последней теоремы Пуанкаре», хотя Пуанкаре опубликовал её именно как гипотезу. Кстати, эта теорема стимулировала формулировку упомянутой выше гипотезы Арнольда, а Френкс начал свою работу в данной области с другого доказательства теоремы Биркгофа и родственных результатов.) Френкс использовал некоторые результаты М. Хендела (которые связаны с упоминавшейся выше гипотезой Арнольда; поскольку Хендел не опубликовал доказательств, Френкс привёл таковые при достаточных для его целей дополнительных предположениях. Затем Френкс ещё раз вернулся к этому кругу вопросов уже для того, чтобы наряду с теоремой Хендела дать независимое от общей симплектической теории доказательство гипотезы Арнольда для диффеоморфизмов поверхностей[16]) [17]. (Позднее Ш. Мацумото, продолжая эту линию исследования, предложил полное доказательство в общем случае гомеоморфизмов как гипотезы Арнольда для поверхностей[17], так и анонсированной Хенделом теоремы [18].) Вскоре после появления статей Френкса и Бангерта Н. Хингстон предложила второе доказательство теоремы о бесконечном числе замкнутых геодезических,

[16]) Здесь идёт речь об этой гипотезе для гомеоморфизмов ориентируемых поверхностей рода $> 1$, сохраняющих площадь. «Общими симплектическими» методами она к тому времени уже была доказана А. Флёром.

[17]) В двумерном случае симплектичность отображения сводится к сохранению площади, а о нём можно говорить применительно не только к диффеоморфизмам, но и к гомеоморфизмам.

основанное на вариационных соображениях [19]. Оно ближе по духу к традициям дифференциальной геометрии (и тем самым ещё дальше от темы настоящего раздела).

Надо сказать, что в России к симплектической геометрии относят также цикл исследований А. Т. Фоменко и его учеников о топологии интегрируемых гамильтоновых систем (см. их последнюю книгу [20]). Конечно, это геометрия и притом связанная с соответствующей симплектической структурой (без которой нельзя было бы говорить о гамильтоновых системах), но по своему содержанию этот цикл исследований довольно далёк от темы данного раздела; я бы отнёс его скорее к теории интегрируемых систем (о других аспектах этой теории см. п. 3.5).

В книге Ш. Кобаяши «Группы преобразований в дифференциальной геометрии» (имеется русский перевод) почти в самом начале говорится: «Не все геометрические структуры сотворены равными: некоторые являются творениями природы, в то время как другие — продукты человеческого разума. Среди первых риманова и комплексная структуры выделяются своей красотой и богатством.» Эта книга вышла в 1972 г. Похоже, что тогда симплектические многообразия были скорее продуктами человеческого разума, но теперь усилиями последнего они постепенно становятся творениями природы, «освежающе непохожими»[18) ] на другие её творения.

**1.2. Конформная динамика.** Другое направление, интенсивно развивавшееся в эти годы, но не выходящее за пределы теории динамических систем, а составляющее часть дифференциальной динамики — это конформная динамика, т. е. исследование итераций аналитических функций в области комплексного переменного. Фазовым пространством здесь служит некоторая область $D$ на плоскости комплексного переменного (причём очень интересными оказываются уже те случаи, когда $D$ есть вся плоскость $\mathbb{C}$ или плоскость, расширенная до сферы Римана) и речь идёт об определённой в $D$ динамической системе с дискретным временем $\{f^n\}$, где $f$ — аналитическая функция (хотя бы многочлен, даже многочлен второй степени!)

Это направление не является новым — оно восходит к классическим работам П. Фату и Г. Жюлиа начала века (если не говорить о ещё более ранних работах по локальным вопросам[19)]). Уже тогда было обнаружено «сложное»

---

[18)] Выражение В. И. Арнольда по более скромному поводу — в связи со сравнением линейной алгебры в симплектическом пространстве $(\mathbb{R}^n, \omega)$ с постоянной (имеющей постоянные коэффициенты) $\omega$ и обычной евклидовой геометрии.

[19)] Основной локальный вопрос: пусть $f(a) = a$; при каких условиях возле $a$ конформное отображение $f$ сопряжено со своим линейным приближением, т. е. с линейным отображением $x \mapsto a + f'(a)(x - a)$, и если сопряжения нет, то что этому мешает? Если $|f'(a)| \neq 1$, сравнительно просто доказывается существование сопряжения, в противном случае ситуация гораздо сложнее, см. [1].

поведение траекторий, напоминающее таковое в гиперболической теории. Последней тогда ещё не существовало, но уже было обнаружено (впервые — Ж. Адамаром) «сложное» поведение геодезических на поверхностях отрицательной кривизны; однако в то время эти два случая «сложного» поведения не сопоставлялись. С течением времени систематическая работа в конформной динамике прекратилась и это направление впало в длительную «спячку» (хотя отдельные работы время от времени появлялись; особенно надо отметить блестящий локальный результат К. Зигеля 1942 г.[20)]), от которой около 1970 г. его пробудило, по-видимому, осознание связей или аналогий с уже возникшей к тому времени гиперболической теорией. Большое значение имели также численные эксперименты, обнаружившие ряд новых явлений и позволившие сформулировать ряд гипотез, работа над доказательствами которых значила очень много для всей этой области. Инициатива в таких экспериментах принадлежит Б. Мандельброту, позднее же, с появлением достаточно мощных персональных компьютеров, они стали широко доступными (хотя порой это требует осторожности).

В конформной динамике (как, впрочем, и при исследовании вещественных одномерных отображений, которыми тоже много занимались в рассматриваемое время) мы встречаемся с почти уникальной возможностью достаточно полного исследования сложного поведения ДС — поведения как в смысле качественной картины в фазовом пространстве, так и в смысле зависимости от параметров. Удаётся исследовать ДС, у которых в разных частях фазового пространства реализуются противоположные типы поведения траекторий — в одной части оно может быть гиперболическим (причём в конформной динамике нередко имеют дело с ослабленными вариантами гиперболичности, которые в данном случае успешно исследуются), в другой — квазипериодическим или напоминающим таковое.

Некоторое понятие о данном направлении может дать книга [21]. Если в обычных математических книгах рисунки иллюстрируют текст, то в [21] дело обстоит, пожалуй, наоборот — книга в значительной степени представляет собой альбом с полученными на компьютере красивыми картинками, которые поясняются в тексте. При этом обычно даются точные формулировки теорем, но не всегда приводятся доказательства. Даются также многочисленные литературные ссылки, приводится подробная информация об истории, — в этом отношении [21] вполне удовлетворяет традиционным высоким требованиям к изложению учебно-обзорного характера. (В ней также имеются замечания о возникающих в некоторых случаях вычислительных трудностях.) Здесь можно указать также имеющиеся на русском языке обзоры [22], [23] и готовящуюся к переводу книгу [24]; наконец, новейшая информация имеется в докладах на последних Международных математи-

[20)] Первый положительный результат для $|f'(a)| = 1$.

ческих конгрессах, на которых конформной динамике неизменно уделялось заметное внимание.

Несколько рисунков из [21] могут дать куда лучшее представление о множествах, возникающих в конформной динамике, чем несколько лекций без рисунков. Я же в порядке примера отмечу только следующее. В некотором смысле, вся «сложность» динамики $\{f^n\}$ «сосредоточена» на так называемом множестве Жюлиа $J = J_f$, которое может выглядеть довольно сложно и удивительно красиво. Но если не говорить о красоте, а только о сложности, то сложно устроенные объекты, играющие сходную роль, известны и в других разделах теории ДС. Пусть теперь $f = f_c$ зависит от параметра $c$ (вполне содержательный и по сей день исследованный не вполне до конца пример: $f_c(z) = z^2 + c$). Тогда возникает вопрос о бифуркациях, т. е. об изменениях качественной картины при изменении $c$. В основном соответствующие значения $c$ — это точки так называемого множества Мандельброта $M$. Вот здесь другие разделы теории ДС далеко уступают конформной динамике — ничего подобного подробным рисункам $M$ из [21] в них нет.

Чем-то $M$ удивительно похоже на множества Жюлиа, хотя это сходство — какое-то неуловимое, да оно и не может иметь простого описания — ведь для указанного выше семейства $\{f_c\}$ множеств Жюлиа бесконечно много, а $M$ одно. Глобально $M$ и $J$ совсем различны, а сходство проявляется в их локальном строении. В помещённой в [21] статье Лей Тан (имя, фамилия) — ученицы одного из видных деятелей в данной области А. Дуади — объясняется, что возле некоторых точек $c \in M$ локальное строение $M$ и локальное строение $J_{f_c}$ в окрестности некоторых точек $z \in J$ асимптотически одинаковы: если рассматривать эти множества в микроскоп, то чем больше увеличение, тем больше их сходство в поле зрения микроскопа. Это пример теоремы, на формулировку которой навели компьютерные эксперименты. Работа Лей Тан была выполнена почти 15 лет назад; в данной области это большой срок, и по нынешним меркам её теорема может считаться простой. А вот пример весьма непростой «компьютерно-стимулированной» гипотезы: для указанного (внешне крайне простого!) семейства $\{f_c\}$ множество Мандельброта локально связно («гипотеза *MLC*»). Она привлекла самое серьёзное внимание специалистов (её формулировка сама по себе интересна, но кроме того оказывается, что будь эта гипотеза доказана, то получались бы интересные следствия, на чём я уже не буду останавливаться), и всё же существование сколь угодно малых связных окрестностей доказано не для всех точек $M$ [21)]. В имеющемся же доказательстве для части точек $M$ снова обыгрывается некое локальное сходство между $M$ и $J$ (первым значительного продвижения здесь достиг

---

[21)] На конференции по теории ДС в Триесте осенью 1998 г. Р. Перец-Марко анонсировал доказательство гипотезы *MLC* в общем случае.

Ж.-К. Йоккоз, а последние известные мне результаты принадлежат М. Любичу [25]).

Сам Б. Мандельброт был весьма обрадован, получив в своё время первые (по нынешним меркам, плохие) изображения множеств Жюлиа и введённого им и носящего ныне его имя множества $M$, так как они доставили новые и важные примеры того, что он называет фрактальными множествами или, короче, фракталами, укрепив его уверенность в их важности и доставив ценный пропагандистский материал. Слово «фрактал» происходит от «fraction» и связано с тем, что соответствующие множества обычно имеют дробную хаусдорфову размерность. Но если хаусдорфова размерность — точное математическое понятие, то фрактал таковым не является. По словам Мандельброта, „все фигуры, которые я исследовал и назвал фракталами, в моём представлении обладали свойством быть «нерегулярными, но самоподобными». Слово «подобный» не всегда имеет классический смысл «линейно увеличенный или уменьшенный», но всегда находится в согласии с удобным и широким толкованием слова «похожий».“ («Самоподобие» понимается примерно в том же смысле, как выше понималось «одинаковое локальное строение».) Это, конечно, не определение, а неформальное описание. Мандельброт поясняет, что множество дробной хаусдорфовой размерности, конечно, «нерегулярно», но оно не обязано быть самоподобным, а множество целой хаусдорфовой размерности вполне может быть «нерегулярным» (скажем, «изрезанным») и при этом обладать или не обладать свойством локального «самоподобия», поэтому «фрактальность» не надо понимать буквально. Так что последняя является каким-то не вполне точно определённым свойством, более или менее подробно описанным с привлечением многочисленных примеров как из математики, так и из естествознания. Фрактальны скалистое побережье, горная цепь, колеблющееся пламя, облако, колония плесени... Мандельброт даже указывает количественные характеристики самоподобия в различных естественнонаучных примерах. Конечно, в этих примерах, в отличие от математических, самоподобие имеет место только в некотором ограниченном диапазоне масштабов. По словам Мандельброта, в неодушевлённой природе концы этого диапазона могут различаться примерно в $10^4$ раза, а в биологических примерах — в $10^2$ раз.

Эти идеи и наблюдения Мандельброта приобрели известную популярность, и с его лёгкой руки различные другие авторы тоже стали обнаруживать фракталы в природе. Видимо, наряду с объектами, форма которых с высокой степенью точности изображается привычными «евклидовыми» фигурами, в природе действительно встречаются объекты «фрактальной» формы. Конечно, это только феноменология; встаёт вопрос, почему тот или иной объект имеет фрактальную форму и притом с такими-то количественными характеристиками. Насколько мне известно, в «математическом естествознании» (понимаемом в самом широком смысле) с ответом на этот вопрос дела пока обстоят неважно. Но можно напомнить, что и с пониманием природы привычных «евклидовых» форм прогресс был довольно медленным и ещё не завершился[22)], так что не удивительно, что применительно к новым формам соответствующие вопросы остаются пока открытыми. В то же время фракталы стали «модными», со всеми вытекающими отсюда последствиями[23)] (ср. с «теорией катастроф» п. 1.4).

Вернёмся к математике. Подчёркивая «самоподобие» фракталов, Мандельброт обратил внимание математиков на существенное свойство ряда «сложно устроенных, нерегулярных» множеств, которые встречаются в математике. Конечно, «самоподобие» может быть «артефактом», т. е. быть созданным руками человека. Такова ситуация в классических примерах вроде нигде не дифференцируемых функций К. Вейерштрасса и Б. Ван дер Вардена, кривых Дж. Пеано, Х. Коха и Д. Гильберта, ковра В. Серпинского. Во всех этих примерах некий объект строился с помощью бесконечного числа шагов и, конечно, авторы стремились сделать эти шаги похожими друг на друга, чтобы построение было легче описать. Но в других случаях вроде $J$ и $M$ «самоподобие» не содержится непосредственно в исходном определении или построении и является немаловажным.

В начале века при возникновении конформной динамики использовались методы «чистой» теории функций комплексного переменного. При

[22)] Так, правильная форма кристаллов известна с древности; уже в прошлом веке были известны их количественные характеристики и тот факт, что данное вещество кристаллизуется всегда в одну и ту же кристаллическую форму или в одну из нескольких кристаллических форм со свойственными данному веществу характеристиками; давно возникло предположение, что дело здесь в расположении атомов или молекул в узлах некоторой кристаллической решётки; количественная реализация последнего предположения была дана Е. С. Фёдоровым и Шёнфлисом немногим более 100 лет назад. Однако только около 90 лет назад это предположение было доказано с помощью дифракции рентгеновских лучей на кристаллах, а что касается самого факта, что большинство веществ при низких температурах переходят в кристаллическую фазу, то лишь сравнительно недавно в статистической термодинамике стали получаться результаты, идущие в этом направлении (между прочим, они отчасти имеют отношение к некоторой части теории ДС; см. упоминание о фазовых переходах в п. 1.3). Наконец, экспериментальные исследования механизма роста кристаллов, в частности открытие дислокационного механизма роста — это почти исключительно послевоенный период (важная идея дислокации появилась раньше, но по другому поводу). О том же, чтобы получить механизм роста «из первых принципов», пока, кажется, вообще не было речи.

[23)] Пожалуй, на этом поприще вершины своей карьеры фракталы достигли в следующей фразе: «Можно сказать, что Иисус Христос находится в центре многомерного фрактала, распространяющегося по некоторым правилам порождения, которые могут быть описаны в бинарных терминах» (М. Элиадис, И. Кулиано, «Словарь религий, образов и верований», М.: Рудомино и СПб: Университетская книга, 1997).

исследовании локальных вопросов использовались разложения в ряды и мажоранты, в глобальных вопросах большую роль играли соображения компактности различных множеств аналитических функций (то, что примерно тогда же было оформлено П. Монтелем в теорию «нормальных семейств аналитических функций», которая имеет и другие применения). Начиная с К. Зигеля, применяются различные приёмы борьбы с малыми знаменателями — более искусно построенные мажоранты у самого Зигеля, позднее метод КАМ (А. Н. Колмогорова—В. И. Арнольда—Ю. Мозера). В нём вместо разложений в ряды используется сходящийся быстрее более сложный процесс с бесконечным числом замен переменных; другой аспект этого метода состоит в некоторой аналогии с методом Ньютона решения нелинейного уравнения (по словам самого Колмогорова, он отчасти руководствовался этой аналогией). Эти два приёма имеют «общематематическое» значение; позднее для той же цели были разработаны более специальные приёмы (связанные с так называемой «перенормировкой» и с «геометрией» соответствующей теоретико-числовой задачи о «малости» знаменателя). Будучи специализированными, они позволили получить лучшие результаты, но (пока что?) только для одномерных задач (вещественных или комплексных). Расширение круга методов в 70-х гг. отчасти связано с развитием теории функций комплексного переменного (например, пригодились пространства Тайхмюллера, о которых уже много было известно). Очень удачным оказалось привлечение квазиконформных отображений. Соответствующий приём вначале смахивает на глупую авантюру. Желая построить конформное отображение с некоторыми свойствами, начинают с некоторого отображения $f$, которое всем было бы хорошо, если бы оно было конформным, но которое, увы, не является таковым; впрочем, оно конформно по отношению к некоторой нестандартной комплексной структуре, вполне заслуживающей ругательные названия, из которых самыми мягкими являются «нелепая» и «никому не нужная». Вот если бы удалось подправить $f$, сделав его конформным в обычном смысле... И тут на сцене появляется deus ex machina — «измеримая теорема Римана»[24]), утверждающая, что $f$ квазиконформно сопряжено с некоторым настоящим конформным отображением. Отмечу, наконец, уже упоминавшуюся перенормировку — приём, не связанный специфически с

[24]) Так её называют в конформной динамике, вообще же она называется «теоремой о существовании квазиконформного гомеоморфизма» (какого именно, в названии не уточняется, и я тоже не буду этого делать). Теорема постепенно доказывалась во всё бóльшей общности; для конформной динамики нужен именно последний вариант, принадлежащий Л. Альфорсу, Л. Берсу, И. Н. Векуа и Ч. Морри. А идея использовать её в конформной динамике принадлежит Д. Сулливану. По-видимому, он же указал и на возможность её использования в теории клейновых групп, что попало в печать несколько быстрее. Тем временем, начиная с 1981 г., измеримую теорему Римана независимо начали использовать в локальных вопросах (С. М. Воронин,

одномерной комплексной ситуацией[25]. В какой-то степени перенормировка связана с общей идеей локального самоподобия, но применяемой не столько к множествам, сколько к отображениям. Впрочем, я не уверен, что на тех, кто разрабатывал этот приём, повлияла общая идея Мандельброта. Вроде бы должна была повлиять, ибо этот приём использовался как раз для исследования упоминавшихся выше вопросов о $J$ и $M$, но дело в том, что ещё раньше этот приём появился сперва в некоторых математических вопросах статфизики, а затем в работах по одномерной динамике в вещественной области.

Кстати, я ничего не говорил о последней. Она тоже получила значительное развитие, начиная с 70-х гг. Но, мне кажется, конформная динамика сейчас является во всех отношениях более крупным направлением и производит бо́льшее впечатление, поэтому, отнюдь не желая умалить достоинства одномерной вещественной динамики, я всё же ограничусь простым упоминанием о ней и ссылкой на книгу [26], которая переводится на русский язык (см., впрочем, конец п. 2.1). Сравнивая обе теории, нельзя не сказать, что обе они существенно используют специфические свойства изучаемых объектов, но в конформной динамике эта специфика составляет предмет богатой и глубокой теории функций комплексного переменного, а в вещественной специфика — это всего лишь свойства прямой. В последнем случае приходится удивляться, что на такой скромной базе всё-таки возникла и продолжает успешно развиваться достаточно содержательная теория. Стоит добавить, что некоторые из недавних её успехов связаны с выходом на время исследования в комплексную область[26] (см., например, конец п. 2.1). Ещё раньше выход в комплексную область использовался в связи с «универсальностью Фейгенбаума»[27] [26]. В самом широком плане такой приём, мягко говоря,

---

Б. Мальгранж и Ж. Мартине—Ж. П. Рамис). Вслед за Сулливаном его идею в других задачах конформной динамики использовали А. Дуади и Дж. Хаббард, после чего она, так сказать, «внедрилась» в сознание исследователей в этой области.

[25] Выше, казалось бы, утверждалось обратное, но там речь шла не об общей идее перенормировки, а о некоем специфическом приёме, использующем перенормировку вместе с «геометрией» некоей теоретико-числовой задачи; в многомерной ситуации в первую очередь недостаёт понимания этой «геометрии».

[26] Иногда этот приём удаётся использовать даже для неаналитических отображений, которые продолжаются в комплексную область неаналитически, но, грубо говоря, так, что они мало отличаются от аналитических.

[27] Речь идёт о некоторой закономерности (тоже своего рода самоподобии) в бесконечной последовательности бифуркаций удвоения периода одномерного отображения. (При такой бифуркации периодическая точка теряет устойчивость и из неё «рождается» устойчивая периодическая точка удвоенного периода.) Доказано, что эта закономерность является «свойством общего положения», сам же М. Фейгенбаум обнаружил её в ходе численных экспериментов. (По его несколько ирони-

не нов (в алгебре ему почти 450 лет), но вот в теории ДС до сих пор выход в комплексную область удачно срабатывал только в локальных и близких к ним вопросах[28]. Видимо, здесь на ограниченном «пятачке» одномерной динамики мы являемся свидетелями начала проникновения этой старой, но вечно юной идеи в куда более молодую теорию динамических систем.

**1.3. Неклассические группы преобразований.** В бóльшей части настоящей статьи под ДС понимается действие группы или полугруппы $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{Z}_+$, $\mathbb{R}_+$ [29] на фазовом пространстве $X$ (причём $X$ имеет определённую структуру и действие определённым образом согласовано с ней). Но можно рассматривать действие других групп. Тогда говорят о «группах преобразований», но если желать подчеркнуть, что нас интересуют вопросы, во многом сходные с рассматриваемыми для обычных ДС, то можно говорить о «ДС с неклассическим временем» (пробегающим группу или полугруппу $G$), тогда как ДС в прежнем смысле — это «ДС с классическим временем». Такая терминология тем более оправдана, что по историческим причинам название «группы преобразований» чаще всего используют, говоря либо о непрерывном действии компактной группы, либо об алгебраическом действии

---

ческому высказыванию, ему помогло то, что он вначале работал всего-навсего с программируемым микрокалькулятором, вследствие чего приходилось более тщательно продумывать ход вычислений, нежели бы это было нужно при использовании более мощного компьютера). Хотя непосредственно здесь говорится об одномерных отображениях, соответствующие последовательности бифуркаций встречаются и для многомерных ДС (включая потоки), потому что для них вполне может случиться, что «основные события» будут всё-таки «одномерными». См. рассказ самого Фейгенбаума о его открытии и (частично гипотетической) роли последнего [27].

[28] Под вопросами, в каком-то отношении близкими к локальным, здесь понимаются некоторые вопросы о бифуркациях, более или менее глобальные по фазовому пространству, но локальные по параметру. См. конец п. 2.3.

[29] Левое действие (полу)группы $G$ на $X$ — это такое отображение

$$G \times X \to X \qquad (g, x) \mapsto \varphi_g(x),$$

что для единицы $e$

$$\varphi_e(x) = x \quad \text{при всех } x$$

и

$$\varphi_g \circ \varphi_h = \varphi_{gh} \quad \text{при всех } g, h.$$

Аналогично определяется правое действие. Для коммутативных $G$ это одно и то же, но в общем случае — нет. По поводу обозначений $\mathbb{Z}_+$, $\mathbb{R}_+$ стоит уточнить, что согласно стандарту Бурбаки, эти множества состоят из неотрицательных элементов $\mathbb{Z}$, $\mathbb{R}$ (так что значок $+$ не надо понимать слишком буквально). Ниже встречается также $\mathbb{N} = \mathbb{Z}_+ \setminus \{0\}$.

Замечу для дальнейшего, что по аналогии со случаем классического времени множество $\{\varphi_g(x);\ g \in G\}$ называют траекторией (или орбитой) точки $x$ (под действием $G$).

алгебраической группы, удовлетворяющем сильному условию регулярности; изучаемые вопросы в обоих случаях имеют другой характер.

Направление, о котором будет идти речь, — это эргодическая теория ДС с неклассическим временем. Но сперва надо сделать одно отступление.

Ниже придётся упоминать аменабельные группы. Хотя их определение дал Дж. Нейман ещё в 1929 г., оно пока не стало общеизвестным. Поэтому я скажу о двух эквивалентных более поздних определениях[30], ограничиваясь только случаем дискретной группы $G$, во втором определении даже конечно порождённой. (В общем случае речь шла бы о локально компактной топологической группе или полугруппе.)

В определении Е. Фёлнера постулируется существование некоторой системы $\{F_n\}$ подмножеств $G$, которую теперь называют фёлнеровской системой. Она играет примерно такую же роль, какую играют отрезки $[-n, n]$ для $\mathbb{Z}$ — при возрастании $n$ они покрывают всю $G$; по ним можно осреднять; при сдвиге на фиксированный элемент группы среднее по достаточно большим множествам фёлнеровской системы почти не меняется.

Другое определение (восходящее к Х. Кестену и усовершенствованное Р. И. Григорчуком) годится для конечно порождённой группы $G$. Если она имеет $n$ образующих, то её можно представить как факторгруппу $G = F_n/H$ свободной группы с $n$ образующими по некоторой нормальной подгруппе $H$. Обозначим через $f(k)$ число несократимых слов из $F_n$ длины $\leqslant k$, а через $h(k)$ — число тех из них, которые лежат в $H$. Почти очевидно, что $f(k) = 2n(2n-1)^{k-1}$. Группа $G$ аменабельна, если $\lim\limits_{k\to\infty} \sqrt[k]{h(k)} = 2n - 1$, т. е. если в $H$ число слов длины $\leqslant k$ растёт (с ростом $k$) практически столь же быстро, как и в $F_n$. Факторизация $F_n$ по $H$ является, стало быть, весьма основательной, — в $G$ очень много слов от образующих равно 1! Тем не менее кое-что остаётся, — как-никак, коммутативные, нильпотентные и даже разрешимые группы аменабельны.

Об аменабельных группах см. в [28] (впрочем, второе из приведённых определений является более новым).

Теперь мы можем перейти к нашей теме. Вскоре после появления первых эргодических теорем («статистической» Дж. Неймана и «индивидуальной» Дж. Биркгофа) их попробовали перенести на другие группы и полугруппы преобразований с инвариантной мерой. Довольно быстро были доказаны аналоги эргодических теорем для $\mathbb{Z}^n$, $\mathbb{R}^n$, $\mathbb{Z}^n_+$, $\mathbb{R}^n_+$, затем перешли к другим или к более общим группам. По-видимому, естественным классом групп (полугрупп), для которых можно рассчитывать на эргодические теоремы,

---

[30] Они выдержаны в более классическом духе, чем первоначальное определение Неймана, в котором речь идёт о возможности осуществления в пространстве ограниченных функций на $G$ некоторого трансфинитного построения (дающего так называемое левоинвариантное банахово среднее).

более или менее непосредственно напоминающие теоремы для классического времени, является класс локально компактных аменабельных групп (полугрупп). Статистическая эргодическая теорема доказана именно в такой общности. С индивидуальной эргодической теоремой положение сложнее. Ограничимся дискретными группами[31]. Пусть $(X, \mu)$ — пространство с конечной[32] мерой, в котором действует посредством сохраняющих меру преобразований $\{\varphi_g\}$ дискретная аменабельная группа $G$. По аналогии с классическим случаем, в индивидуальной эргодической теореме должны фигурировать средние $f_n(x) := \frac{1}{\#F_n} \sum_{g \in F_n} f(\varphi_g(x))$, где $F_n$ — множество из фёлнеровской системы $\{F_n\}$, а $\#F_n$ — число его элементов. Оказывается, не всегда можно утверждать, что последовательность $f_n(x)$ сходится почти всюду. Достаточным для этого является следующее дополнительное условие об $\{F_n\}$: система $\{F_n\}$ — возрастающая (т. е. $F_n \subset F_{n+1}$) и, главное,

$$\sup_{n \in \mathbb{N}} \frac{\#(F_n^{-1} F_n)}{\#F_n} < \infty$$

(последнее условие указано А. Кальдероном). Фёлнеровские системы, удовлетворяющие условию Кальдерона, существуют для почти нильпотентных групп (групп, имеющих нильпотентную подгруппу конечного индекса)[33]. Они существуют также для счётных локально конечных групп. Неизвестно, нет ли ещё какого-нибудь значительного класса групп, для которых суще-

---

[31] В литературе рассматривались также действия локально компактных групп и полугрупп. Предположение о дискретности делается для простоты изложения (здесь заметных различий между дискретным и локально компактным случаями не видно, хотя в дальнейшем таковое кое-где возникает).

[32] Для классического времени хорошо известно, что эргодические теоремы имеют место и в пространствах с бесконечной мерой, причём при переходе к ним и формулировки, и доказательства почти не меняются, хотя при дальнейшем развитии теории случаи конечной и бесконечной меры оказываются существенно различными и сейчас это развитие в основном относится к случаю конечной меры. Но поскольку в случае неклассического времени ситуация с индивидуальной эргодической теоремой неясна, естественно пока что ограничиться случаем конечной меры, что обычно и делают.

[33] В [29] сказано, что счётные группы, для которых существует фёлнеровская система, удовлетворяющая рассматриваемым там условиям, основным из которых является условие Кальдерона, «в существенном (essentially)» суть группы с полиномиальным ростом, т. е. конечно порождённые группы с полиномиальным ростом введённой выше функции $f(k)$, а такие группы, как доказал М. Громов, почти нильпотентны. В [29] не сказано, что конкретно означают слова «в существенном»; поскольку говорить о полиномиальном росте можно только для конечно порождённых групп, то возможно, что именно они и имеются в виду. Но, насколько известно, даже и для них не доказано, что из условия Кальдерона следует полиномиальность роста $f(k)$.

ствуют такие фёлнеровские системы, а также нельзя ли заменить условие Кальдерона каким-нибудь более общим. Таким образом, индивидуальная эргодическая теорема доказана в формулировке, достаточно близкой к классической, для почти нильпотентных и для локально конечных групп $G$. Для других групп в литературе имеются некоторые анонсы, подозрительно долго остающиеся не поддержанными подробными публикациями.

Об эргодических теоремах для аменабельных групп см. [30], [29].

Недавно И. Линденштраусс анонсировал следующий общий результат. В любой счётной аменабельной группе имеется такая фёлнеровская система $\{F_n\}$, что для любого действия с конечной инвариантной мерой для осреднений по $\{F_n\}$ имеет место индивидуальная эргодическая теорема. (Таким образом, «некоторые фёлнеровские системы являются более фёлнеровскими, чем другие».) Подробной публикации пока нет, но так как анонс был недавним, слова «подозрительно долго» к нему не могут относиться.

Но, оказывается, и за пределами класса аменабельных групп могут иметь место эргодические теоремы. Во всяком случае, Р. И. Григорчук получил индивидуальную эргодическую теорему для конечно порождённых свободных групп (полугрупп) и некоторых близких к ним, — а ведь считается (да отчасти и видно из второго определения), что свободные группы максимально далеки от аменабельных[34)]. Всё дело в том, что Григорчук изменил способ осреднения по «времени», т. е. по группе. Я бы сравнил это с переходом в теории рядов от обычной сходимости к чезаровским средним. Как известно, там имеется много различных способов суммирования. Быть может, нечто аналогичное ожидает нас и в эргодической теории? А может быть, хотя бы для конечно порождённых групп какой-то способ окажется универсальным?

Но я заговорил о неклассическом времени не ради одних только эргодических теорем (если бы речь шла только о них, это едва ли подошло бы под название данного параграфа). В эргодической теории систем с классическим временем эргодические теоремы составляют только часть. Исторически эта часть возникла первой и дала название всей теории, но уже полстолетия это именно часть и даже не половина. В «абстрактных», «чисто метрических»

---

[34)] Первоначально Григорчук опубликовал сообщение о своих результатах только в трудах провинциальной конференции [31] (ограничиваясь свободными группами, но не требуя конечности инвариантной меры). Оно не привлекло внимания, и позднее появились аналогичные работы других авторов [32]. Начали также рассматриваться аналогичные теоремы для действий некоторых полупростых групп Ли [33], [34].

Я не буду обсуждать здесь более ранних работ [35], [36]. Их пионерское значение бесспорно, но в них речь идёт о специальных ситуациях и у них имеются специфические особенности, которые хотелось бы понять с общих позиций, к чему я, — думаю, не я один, — не готов. (Например, в [36] осреднение по разрешимой группе производится таким образом, как если бы она была свободной.) Отмечу только статистическую эргодическую теорему для свободной группы, доказанную в [37].

разделах эргодической теории обсуждаются различные свойства эргодических систем, вплоть до частичной классификации последних. При этом обычно приходится вместо произвольных пространств с мерой (хотя бы и нормированной) рассматривать пространства Лебега[35], что, впрочем, практически не сужает общности результатов с точки зрения возможности их приложения к конкретным примерам. В других разделах, которые условно можно назвать «прикладными», исследуются свойства конкретных систем или некоторых их классов. (Эти системы часто являются гладкими или топологическими, так что, в принципе, их изучение можно было бы отнести к топологической или гладкой динамике, но когда речь идёт о свойствах, известных из «абстрактной» эргодической теории, соответствующие исследования (классов) конкретных систем обычно тоже относят к эргодической теории.)

Для систем с неклассическим временем отдельные исследования такого рода появлялись и раньше, но широкое развитие они получили примерно после 1970 г. При этом в одних случаях выявлено сходство с классическим случаем, а в других — отличие от него, причём граница, отделяющая привычную ситуацию от необычной, проходит в различных вопросах по-разному. Часто, как и выше, ситуация для аменабельных групп аналогична классической; иногда уже $\mathbb{Z}^2$ разительно отличается от $\mathbb{Z}$.

Как и в предыдущей части §1, невозможно в двух словах рассказать о целой науке. Приведу только несколько примеров (поскольку я пытаюсь охарактеризовать данную область в целом, некоторые из этих примеров относятся к более раннему периоду, нежели последняя четверть века). Ограничусь только действиями дискретных групп (хотя многое — но не всё — проходит для сепарабельных локально компактных полугрупп[36]) и вначале

---

[35] Пространство Лебега — это пространство с мерой, изоморфное (в том смысле, как это естественно понимать для пространств с мерой) стандартному объекту — отрезку $[0, 1]$ с нормированной мерой Лебега—Стильтьеса. Эквивалентное определение: пространство Лебега изоморфно отрезку $[0, a]$ с мерой Лебега, к которому добавлено не более чем счётное множество «атомов» с мерами $p_n$, причём $a+\sum p_n = 1$ (нормированность). Это определение учитывает некоторые полезные особенности большинства встречающихся в эргодической теории конкретных примеров, позволяющие при анализе чисто метрических вопросов продвинуться значительно дальше, нежели это было бы возможно для общих пространств с мерой. О том, что класс пространств Лебега достаточно велик, свидетельствует следующий факт: метризуемый компакт с нормированной мерой (определённой на его борелевских подмножествах) является пространством Лебега.

Теорию пространств Лебега разработали Дж. Нейман и П. Халмош (употреблявшие другое название) и особенно В. А. Рохлин.

[36] Уже для некоммутативных топологических групп могут возникать затруднения из-за различия между левоинвариантной и правоинвариантной мерами Хаара. Это есть нечто новое по сравнению с дискретными группами.

(в пп. *а*—*в*) буду считать, что группа $G$ действует на пространстве Лебега $(X, \mu)$ и что все преобразования сохраняют меру $\mu$.

*а*) Один из разделов эргодической теории — спектральная теория — начинается с того, что элементам $g \in G$ сопоставляются операторы $U_g$ в $L^2(X, \mu)$:

$$(U_g f)(x) = f(\varphi_g^{-1}(x)). \tag{3}$$

При этом операторы $U_g$ получаются унитарными и $U_{gh} = U_g U_h$, так что мы имеем дело с унитарным представлением[37)] группы $G$. Возникает задача — исследовать свойства этого представления и выяснить, насколько полно они отражают свойства исходной ДС. При переходе от одной ДС к другой ДС, метрически изоморфной исходной, соответствующее представление заменяется на унитарно эквивалентное. В функциональном анализе разработана система понятий, относящихся к свойствам унитарных представлений, инвариантных относительно унитарной эквивалентности. В простейшем случае, когда речь идёт о степенях $U^n$ унитарного оператора $U$ (так что мы имеем дело с представлением $\mathbb{Z}$), всё сводится к спектру этого оператора; поэтому и в более общем случае соответствующие понятия, инварианты, свойства называют спектральными. Константы всегда инвариантны относительно (3) и они образуют одномерное подпространство; поэтому молчаливо ограничиваются рассмотрением этого представления на ортогональном к константам подпространстве $H \subset L^2(X, \mu)$ и соответствующие спектральные свойства называют также свойствами ДС $\{\varphi_g\}$. Надо пояснить, что в данном случае даже для одного оператора $U$ спектр понимается в более тонком смысле, нежели в элементарных курсах функционального анализа, где под спектром линейного оператора $A$ понимают совокупность тех $\lambda \in \mathbb{C}$, для которых оператор $A - \lambda I$ не имеет всюду определённого ограниченного обратного. В смысле этого определения спектр ДС $\{\varphi^n\}$ во всех практически интересных случаях совпадает с единичной окружностью; таким образом, он доставляет нам инвариант, который практически всегда одинаков и потому бесполезен. Но для различных специальных (и в то же время важных) классов операторов в функциональном анализе имеются более тонкие варианты понятия спектра; для нас наиболее существенно такое уточнённое понятие спектра для унитарного оператора $U$ и самосопряжённого оператора $A$ (последний выступает на сцену в случае потока как производящий оператор соответствующей однопараметрической группы унитарных операторов). Недостаток места заставляет отослать читателя по поводу этого понятия к менее элементарным учебникам по функциональному анализу (оно

---

[37)] Если бы мы в правой части (3) взяли $\varphi_g$ вместо $\varphi_g^{-1}$, то получилось бы антипредставление: $U_{gh} = U_h U_g$, что в общем ничуть не хуже, но менее привычно. Для полугруппы пришлось бы иметь дело именно с антипредставлением, но кроме того (что существеннее) операторы $U_g$ были бы не унитарными, а изометрическими.

приводится также, хотя бы отчасти, в некоторых книгах по эргодической теории)[38]. Ограничусь указанием, что говорят о непрерывном спектре, когда у $U$ или $A$ нет собственных функций (на $H$); о дискретном спектре — когда собственные функции образуют полную систему (в $H$ и тогда во всём $L^2(X,\mu)$); о смешанном спектре — в остальных случаях.

Следует предупредить, что если для двух ДС соответствующие представления унитарно эквивалентны, то отсюда, вообще говоря, не следует, что ДС метрически изоморфны. Поэтому свойства ДС, вообще говоря, не исчерпываются её спектральными свойствами. Но иногда это так. В классической ситуации метрического автоморфизма $\varphi$ или потока с инвариантной мерой $\varphi_t$ таков тот случай, когда ДС эргодична[39] и спектр соответствующего $U$ или $A$ дискретный. В этом случае спектр полностью определяет ДС с точностью до метрического изоморфизма. В неклассической ситуации Дж. Макки[40] выделил случай, который во многом аналогичен и потому получил то же название: это тот случай, когда рассматриваемое представление разлагается в дискретную прямую сумму неприводимых конечномерных представлений. За описанием этого разложения условно сохраняется название «спектр». Если $G$ коммутативна, то ситуация вполне аналогична классической, но для общих групп это не совсем так; в частности, дискретный спектр может не определять ДС с точностью до метрического изоморфизма.

Если спектр не дискретный, то даже для классического времени едва ли можно говорить о какой-то общей теории, а скорее о примерах и классах примеров. Для неклассического времени известно ещё меньше. Отмечу один только факт, контрастирующий с ситуацией для классического случая. В своё время в классической ситуации С. Банах обратил внимание на то,

---

[38] См., впрочем, п. 3.9.

[39] В «абстрактной» эргодической теории ДС с классическим временем доказывается, что каждая система в некотором естественном смысле разлагается и притом единственным образом на эргодические компоненты, поэтому принимается точка зрения, что свойства «общих» ДС как бы сводятся к свойствам их эргодических компонент и оттого в основном надо заниматься именно эргодическими ДС (тогда как в «прикладной» теории приходится, конечно, для каждой конкретной ДС стараться выяснить, эргодична она или нет). Для ДС с неклассическим временем сказанное остаётся в силе, если $G$ — локально компактная группа со счётной базой (причём мера в фазовом пространстве может быть не инвариантной, а квазиинвариантной). При более общих $G$ различные определения эргодичности и разложения на эргодические компоненты становятся неэквивалентными (см. статью «Метрическая транзитивность» в «Математической энциклопедии»); всё же в какой-то степени указанная точка зрения остаётся применимой.

[40] Надо сказать, что в первые две трети 60-х гг. Макки был едва ли не единственным, кто пропагандировал изучение ДС с неклассическим временем, выходящее за пределы эргодических теорем.

что во всех известных ему примерах с лебеговским спектром этот спектр счётнократный, и поставил вопрос, обязательно ли это должно быть так. Теперь разобрано больше примеров и во всех них подтверждается наблюдение Банаха, но ответ на его вопрос по-прежнему неизвестен[41]. А вот для мультипликативной группы ненулевых рациональных чисел известен пример действия с однократным лебеговским спектром (М. Новодворский).

Отмечу заодно ещё один пример контраста между классическим и неклассическим аменабельным временем, хотя этот пример и не «спектральный», по крайней мере, непосредственно не является таковым[42]. В. А. Рохлин высказал предположение, что если ДС с классическим временем обладает свойством перемешивания, то она обладает и свойством перемешивания любой кратности. Этот вопрос тоже остаётся открытым, тогда как для $G=\mathbb{Z}^2$ Ф. Ледраппье указал пример, где перемешивание есть, а кратного перемешивания нет.

*б*) Другим большим разделом эргодической теории является энтропийная теория, содержащая не только определение метрического инварианта «энтропия» и исследование его свойств, но и ряд родственных вопросов. Она в значительной степени переносится на аменабельные группы, начало чему положил А. М. Стёпин, определивший в конце 60-х гг. энтропию $h_\mu$ для действия аменабельной группы $G$ с инвариантной нормированной мерой $\mu$. Позднее совместно с А. Т. Таги-заде для непрерывного действия $G$ на метрическом компакте он определил топологическую энтропию и доказал, что она совпадает с $\sup_\mu h_\mu$, где верхняя грань берётся по всем инвариантным нормированным мерам этого действия. (Тем самым подразумевается, что последние имеются; см. п. *г*.) Это является обобщением относящейся к классическому времени теоремы Е. И. Динабурга—Т. Гудмана—Л. Гудвина. См. также п. *е*.

*в*) Вот ещё один пример, где класс аменабельных групп тоже оказывается тем естественным классом, на который переносится известный результат, первоначально полученный для классического времени: любые два эргодические действия двух счётных дискретных аменабельных групп на пространстве Лебега $X$, сохраняющие меру, траекторно эквивалентны, т. е. существует такой метрический изоморфизм $X \to X$, который переводит траектории одной

---

[41] Если спектр имеет не только лебеговскую, но и сингулярную компоненту, то кратность первой может быть конечной, см. [38], где имеются также ссылки на более ранние примеры такого рода.

[42] Перемешивание, о котором здесь идёт речь, является «спектральным» свойством, т. е. определяется некоторым свойством спектра соответствующих $\{U\}$ или $\{U_t\}$, а для кратного перемешивания характеристики в терминах спектра не имеется (хотя из формулируемой ниже гипотезы Рохлина в конечном счёте следовало бы, что такая характеристика есть).

системы в траектории другой (А. Конн, Дж. Фельдман и Б. Вейсс, 1981 г.). С чисто метрической точки зрения имеется, стало быть, только одно разбиение на траектории эргодических действий — в качестве «стандартного образца» можно взять, скажем, разбиение окружности на траектории поворота на «иррациональный» угол[43].

У неаменабельной группы имеются траекторно неэквивалентные эргодические действия. Особенно резко контрастирует с теоремой Конна—Фельдмана—Вейсса следующий результат Р. Циммера (1980). Пусть $G, H$ — две связные полупростые группы Ли ранга больше 1, без центра и конечных факторгрупп. Пусть $\{\varphi_g\}$ и $\{\psi_h\}$ — их действия на пространстве Лебега $(X, \mu)$, которые сохраняют меру $\mu$, эргодичны и остаются таковыми при ограничении на любую неединичную нормальную подгруппу, причём любой элемент группы, отличный от единицы, сдвигает все или почти все[44] точки $X$. Если эти действия траекторно эквивалентны, то $G$ и $H$ изоморфны и после соответствующего отождествления их элементов оба действия становятся метрически изоморфными. Иными словами, существуют такие изоморфизм групп Ли $f\colon G \to H$ и метрический автоморфизм $h\colon X \to X$, что $h\varphi_g = \psi_{f(g)}$ для всех $g \in G$.

Когда нечто может реализовываться, по существу, единственным образом (т. е. единственным образом с точностью до очевидных модификаций вроде «подкручивания» на $f$ и $h$ выше), то часто говорят о «жёсткости»[45]. К моменту появления работы Циммера уже был получен ряд результатов (прежде всего, Г. Мостова и Г. А. Маргулиса) о жёсткости дискретных подгрупп групп Ли; они, как отмечает сам Циммер, оказали на него влияние (в данном случае, возможно, не столько в смысле непосредственной логической зависимости, сколько в отношении идейных ассоциаций[46]). Позднее появились работы (прежде всего, А. Катка и его сотрудников) о некоторых

---

[43] То есть на угол, несоизмеримый с «полным» (360°). При реализации окружности как $\mathbb{R}/\mathbb{Z}$ это есть сдвиг на иррациональное (в обычном смысле слова) число. Замечу, что именно разбиение окружности на траектории этого сдвига встречается при построении наиболее известного примера неизмеримого множества.

[44] В первом случае говорят о свободном действии. Во втором случае Макки и Циммер называют действие «существенно свободным» («essentially free»; более точным переводом было бы «в сущности свободным»).

[45] Этот термин пришёл, по-видимому, из геометрии поверхностей, где, впрочем, некоторые из наиболее авторитетных специалистов предпочитали говорить об «однозначной определённости» и только для некоего инфинитезимального варианта последней употребляли название «жёсткость».

[46] Однако у Циммера есть теорема (к сожалению, формулируемая слишком громоздко), следствиями которой являются как его результат о жёсткости, так и часть результатов Мостова и Маргулиса, так что формальные связи здесь тоже имеются.

явлениях жёсткости в более гладкой ситуации, смыкающейся с гиперболической теорией.

*г*) Широко известна теорема Н. М. Крылова—Н. Н. Боголюбова, которая находится на грани между топологической динамикой и эргодической теорией. Она утверждает, что топологическая динамическая система с классическим временем, фазовое пространство которой компактно, имеет хотя бы одну нормированную инвариантную меру. Вскоре после её опубликования Н. Н. Боголюбов указал, что точно такая же теорема справедлива для непрерывных действий аменабельных групп[47)] на компактах (об этой его работе 1939 г. см. [39]). Можно доказать, что и обратно, если для любого непрерывного действия локально компактной группы на компактах имеется конечная инвариантная мера, то группа аменабельна.

*д*) Эта теорема Боголюбова, как и относящаяся к «классическому» случаю теорема Крылова—Боголюбова, ничего не говорит о свойствах инвариантной меры — в различных примерах они могут быть различными; может также случиться, что одна система имеет много нормированных инвариантных мер, в том числе и эргодических, свойства которых существенно различаются. Одна мера может быть сосредоточена в одной точке (неподвижной при всех $\varphi_g$), а другая — быть положительной для всех открытых множеств. Первая мера эргодична[48)], но никаких содержательных утверждений о системе с такой мерой (кроме того, что у системы есть неподвижная точка), разумеется, сделать нельзя. Во втором случае мера может быть, а может и не быть эргодической; если она эргодическая, то она может обладать или не обладать более сильными свойствами «квазислучайного» характера (в «классической» ситуации — перемешивание, положительная энтропия и т. д.). Иногда существование инвариантной меры, заведомо заслуживающей особого внимания, известно заранее (с этим мы сталкиваемся в «классической» ситуации, рассматривая гамильтоновы системы), иногда же заранее никакой «привилегированной» инвариантной меры не имеется и специально выясняется вопрос о существовании не просто инвариантных мер, а мер с теми или иными интересными свойствами. Естественно, это делается в гораздо более конкретной ситуации, нежели в общих теоремах Крылова—Боголюбова или Боголюбова.

Интересный и важный класс примеров, для которых исследуется данный вопрос, возникает из статистической физики или по крайней мере подсказывается ею. Представим себе, что в каждой точке решётки $\mathbb{Z}^m$ находится

---

47) В то время Н. Н. Боголюбов называл такие группы «банаховыми», ибо они характеризуются существованием банахова среднего.

48) Когда одна и та же топологическая ДС рассматривается одновременно с различными инвариантными мерами, то вместо «эргодичности ДС по отношению к мере $\mu$» часто говорят об «эргодичности меры $\mu$».

частица, которая может находиться в одном из $k$ возможных состояний («решётчатая система»). Обозначим фазовое пространство состояний одной частицы через $A$ (например, можно взять $A = \{1, \ldots, k\}$, если нет особых причин обозначать состояния частицы как-то иначе). Тогда состояние всей нашей бесконечной системы частиц описывается функцией $\xi\colon \mathbb{Z}^m \to A$, где $\xi(g)$ для точки $g \in \mathbb{Z}^m$ есть состояние находящейся в этой точке частицы. Я нарочно обозначаю точку $\mathbb{Z}^m$ через $g$, чтобы сразу перейти от $\mathbb{Z}^m$ к произвольной группе $G$ (собственно, в бóльшей части этого пункта можно было бы взять и полугруппу, но ограничимся группой, чтобы потом не делать оговорок). Конечно, в случае $G \neq \mathbb{Z}^m$ мы уже не можем представлять себе некий кристалл, расположенный в $\mathbb{R}^m$, но по-прежнему можем рассматривать функции $\xi\colon G \to A$. Совокупность $\Omega := A^G$ всех таких функций есть фазовое пространство нашей бесконечной системы. Оно естественно снабжается топологией (тихоновская топология прямого произведения бесконечного числа экземпляров пространства $A$)[49] и несколько менее естественным образом — метрикой. Несколько меньшая естественность метрики состоит в том, что имеется много метрик, индуцирующих эту топологию, и *a priori* нет оснований предпочесть одну из них другим. Опыт показывает, что в случае $G = \mathbb{Z}^m$ для наших целей хороша следующая метрика. Взяв какую-нибудь (безразлично, какую) метрику $d$ в $A$ (например, можно принять, что расстояние между различными точками $A$ равно 1, или положить $d(i, j) = |i - j|$), взяв какое-нибудь (безразлично, какое) число $a \in (0, 1)$, и обозначив

$$|g| = \sum_r |g_r| \quad \text{для } g = (g_1, \ldots, g_k) \in \mathbb{Z}^m,$$

положим

$$\rho(\xi, \eta) = \sum_{g \in \mathbb{Z}^m} a^{|g|} d(\xi(g), \eta(g)).$$

Легко проверяется, что топология в $\Omega$, индуцируемая этой метрикой, есть как раз тихоновская топология. Для других групп тоже можно взять нечто аналогичное, заменив $a^{|g|}$ на подходящую функцию от $g$, которая, так сказать, достаточно быстро убывает при удалении $g$ от единицы группы; я не буду на этом останавливаться. Таким образом, $\Omega$ — метрический компакт.

Наконец, имеется естественное действие $G$ на $\Omega$:

$$(g, \xi) \mapsto \varphi_g(\xi), \quad \text{где} \quad (\varphi_g(\xi))(h) = \xi(gh) \quad \text{при всех } h \in G. \tag{4}$$

Здесь мы обозначили групповую операцию как умножение $(g, h) \mapsto gh$, а не как сложение $(g, h) \mapsto g + h$, потому что $G$ может быть некоммутативной.

[49] Читатель, не знакомый с тихоновской топологией, может принять по определению, что в данном случае это есть топология, индуцируемая вводимой ниже метрикой $\rho$.

(Когда имеют дело с $\mathbb{Z}^m$, то, конечно, имеют дело с обычным сложением и записывают его с помощью знака $+$.) Отображения $\varphi_g$ — гомеоморфизмы $\Omega$.

Полученный объект — метризуемый компакт $\Omega$ с действующей на нём согласно сказанному группой $G$, — называют топологической ДС Бернулли или топологическим бернуллиевским действием группы $G$. Основанием для такого названия служит то, что при $G = \mathbb{Z}$ элемент $\xi \in \Omega$ можно интерпретировать как запись результатов бесконечной последовательности испытаний, имеющих одни и те же возможные исходы, которые образуют множество $A$: $\xi(n)$ — это исход испытания в момент времени $n$ (время дискретно и пробегает $\mathbb{Z}$). При этом $\xi(n)$ с $n \leqslant 0$ — это результаты уже проведённых испытаний (в частности, $\xi(0)$ — это результат испытания, проведённого «сейчас»), а $\xi(n)$ с $n > 0$ — это результаты будущих испытаний; первые нам уже известны, вторые нет. Чаще говорят о двусторонне бесконечной последовательности

$$\ldots, \xi(-n), \xi(-n+1), \ldots, \xi(-1), \xi(0); \xi(1), \xi(2), \ldots, \xi(n), \ldots, \tag{5}$$

представляя себе её элементы записанными именно так, как выше, — в порядке возрастания номера слева направо. Чтобы выделить положение элемента с номером ноль, использована точка с запятой. Действие $\mathbb{Z}$ на $\Omega$ сводится к итерированию «топологического сдвига Бернулли»

$$\sigma\colon \Omega \to \Omega, \quad \xi \mapsto \sigma\xi, \quad (\sigma\xi)(n) := \xi(n+1)$$

и обратного к нему отображения $\sigma^{-1}$. При этом сдвиге последовательность (5) сдвигается налево относительно точки с запятой; прежнее $\xi(1)$ становится элементом с нулевым номером, так что можно представить себе, что мы ещё раз повторили испытание и теперь уже знаем этот элемент. Обозначим проекцию бесконечного произведения

$$\Omega = A^{\mathbb{Z}} = \ldots A \times A \times \ldots \times A \times \ldots$$

на его «нулевой» сомножитель $A$ через $\pi_0$, так что $\pi_0\xi = \xi(0)$. Тогда можно сказать, что при последовательных испытаниях в момент времени $n$ наблюдается исход $\pi_0(\sigma^n\xi)$.

Из сказанного видно, что при $G = \mathbb{Z}$ один и тот же объект $(\Omega, \sigma)$ может интерпретироваться двумя способами. С точки зрения статфизики, с которой мы начали, речь идёт о состояниях бесконечной цепочки. Ни о какой динамике в обычном смысле — в смысле эволюции системы с изменением времени — при этом нет речи, ибо о настоящем времени вообще не говорится. Число $n$ в (5) — это номер элемента цепочки, а действие $\sigma$ соответствует пространственному сдвигу цепочки на одно «звено». С вероятностной точки зрения, к которой мы перешли, $n$ может считаться «настоящим, физическим» временем: за время $n$ точка $\xi$ фазового пространства переходит в $\sigma^n\xi$. Это соответствует тому, что мы представляем себе испытания, входящие в данную последовательность испытаний, осуществляющимися один за другим; сюда добавляется ещё указание, что в опыте наблюдается не $\xi$, а только $\pi_0\xi$, так что с изменением времени наблюдаются величины $\pi_0\sigma^n\xi$, которые в данном случае совпадают с $\xi(n)$. (В более общем случае фазовое пространство $\Omega$ и отображение $\sigma\colon \Omega \to \Omega$ могут быть другими, а наблюдаемая величина может быть некоторой функцией $f\colon \Omega \to \mathbb{R}$, так что с течением времени из наблюдений получается последовательность $f(\sigma^n\xi)$.) Статфизическая точка зрения годится и при $G = \mathbb{Z}^m$ с $m > 1$; «динамика» опять-таки означает не изменение состояния со временем (которого нет), а действие пространственных сдвигов (трансляций) на рассматриваемую систему. Говорить же о последовательности испытаний больше не приходится, хотя можно вообразить испытания, почему-то пронумерованные элементами $\mathbb{Z}^m$ (чему, пожалуй, ещё можно дать (квази)реалистическую интерпретацию) или даже элементами общей группы $G$. Для $G \neq \mathbb{Z}^m$ статфизическая интерпретация тоже становится условностью речи, навеянной аналогией с физически реальным случаем $G = \mathbb{Z}, \mathbb{Z}^2, \mathbb{Z}^3$.

Хотя мы отчасти использовали вероятностный язык и упомянули о Бернулли, пока никаких вероятностей у нас не было. С именем Якова Бернулли, как известно, связано исследование последовательности одинаковых независимых испытаний. В этом случае каждый возможный исход испытания $a \in A$ имеет некоторую вероятность $p(a)$, а независимость испытаний проявляется в том, что результат испытания в момент времени $n$ не зависит от результатов испытания в другие моменты времени. Стало быть, вероятность того, что в моменты $n_1, \ldots, n_l$ получатся исходы $a_1, \ldots, a_l$, есть $p(a_1) \cdot \ldots \cdot p(a_l)$. Вот мы фактически и описали меру $\mu$ в $\Omega$, по отношению к которой последовательность случайных величин $\pi_0(\sigma^n\xi)$ описывает последовательность независимых одинаковых испытаний. Повторим описание ещё раз, перенося его на общий случай $\Omega = A^G$, когда вероятностная интерпретация (в терминах последовательности испытаний) теряет непосредственный смысл, хотя условно можно продолжать использовать вероятностный язык. (В теории вероятностей в этих случаях говорят о случайных полях, что соответствует «интуитивному смыслу» этих слов, когда $G = \mathbb{Z}^m$.)

Элементам $a \in A$ должны быть заранее приписаны вероятности или меры $p(a) \geqslant 0$; при этом сумма последних должна равняться 1. Тем самым подмножеству $B \subset A$ приписывается мера $p(B) = \sum\limits_{a \in B} p(a)$ и мы получаем меру на конечном множестве $A$, что, конечно, тривиально. Мера на $\Omega$, которая будет построена, зависит от задания этих $p(a)$, т. е. этой меры $p$ на $A$.

Цилиндрическим подмножеством пространства $\Omega$ называется множество вида

$$C = \{\xi;\ \xi(g_1) \in B_1,\ \ldots,\ \xi(g_l) \in B_l\}, \tag{6}$$

где $g_1, \ldots, g_l \in G$, $B_1, \ldots, B_l \subset A$. Ему приписывается мера

$$\mu(C) = p(B_1) \cdot \ldots \cdot p(B_l). \tag{7}$$

В теории меры доказывается, что на $\sigma$-алгебре борелевских подмножеств компакта $\Omega$ существует и притом единственная мера $\mu$, которая для цилиндрических множеств (6) принимает значения (7). С точки зрения теории меры, она является прямым произведением

$$p^G = \ldots \times p \times p \times \ldots \times p \times \ldots$$

мер $p$ на сомножителях $A$ прямого произведения $A^G$. Почти очевидно, что эта мера инвариантна относительно действия (4) группы $G$ и что действие эргодично по отношению к этой мере.

В классическом случае сдвиг $\sigma$, рассматриваемый по отношению к построенной мере, называется метрическим автоморфизмом Бернулли. Саму меру естественно тоже называть бернуллиевской. Свойства ДС $\{\sigma^n\}$ с этой мерой резко отличаются от свойств эргодической ДС с дискретным спектром. Система $(\{\sigma^n\}, \mu)$ имеет перемешивание всех степеней, счётнократный лебеговский спектр и положительную, но конечную (метрическую) энтропию. Вообще, автоморфизм Бернулли — это как бы образец ДС с «квазислучайными» свойствами (что и понятно ввиду его происхождения) и даже, так сказать, крайний такой образец, если не считать ДС с бесконечной энтропией. В случае неклассического времени бернуллиевские ДС играют аналогичную роль.

Само по себе определение бернуллиевской ДС и бернуллиевской меры не зависит от того, аменабельна ли $G$. Для дальнейшего же в этом и следующем пп. требуется аменабельность.

Значительным достижением теории ДС является теорема Д. Орнстейна, согласно которой автоморфизмы Бернулли с одинаковой энтропией метрически изоморфны. Д. Орнстейн и Б. Вейсс показали, что эта теорема переносится на бернуллиевские действия аменабельных групп [40], [41].

*е*) Мера на $A$ (т. е. система чисел $p(a) \geqslant 0$, удовлетворяющая условию $\sum p(a) = 1$) может выбираться произвольно; таким образом, у топологической ДС Бернулли имеется континуум инвариантных эргодических мер. Но оказывается, что помимо построенных выше, эта ДС имеет очень много других инвариантных нормированных мер, в том числе и эргодических. Некоторые из них безусловно заслуживают особого внимания.

При $G = \mathbb{Z}$ некоторые меры, как и бернуллиевские, выделяются (из общего огромного множества инвариантных мер) или вводятся (независимо от того, что мы знаем о прочих мерах) на основании вероятностных соображений. В теории вероятностей рассматриваются некоторые последовательности одинаковых испытаний, которые не являются независимыми. Особенно важны те случаи, когда $\{\pi_0(\sigma^n\xi);\ n \in \mathbb{Z}\}$ оказывается марковским процессом. Им соответствуют новые инвариантные нормированные меры $\mu$ в $A$, которые иногда эргодичны, иногда нет. Но при переходе от $\mathbb{Z}$ к $G$ определение марковского свойства теряет смысл. Некое разумное видоизменение этого определения выработано для $G = \mathbb{Z}^m$, но пока что соответствующие объекты, насколько известно, играют заметную роль в теории случайных полей, а не в эргодической теории. Мы их оставим в стороне и обратимся от теории вероятностей к статистической физике, с которой мы начали, но от которой затем ушли.

С физической точки зрения бернуллиевские меры отвечают такой ситуации, когда состояния частиц в вершинах решётки $\mathbb{Z}^m$ не зависят друг от друга. Так, конечно, может быть при отсутствии взаимодействия между ними, но интересен как раз тот случай, когда взаимодействие имеется. В этом случае соображения, заимствованные из статфизики, приводят к некоторым новым нормированным инвариантным мерам. Их определению, как мы увидим, можно придать такую форму, что оно уже не будет относиться специально к топологической бернуллиевской ДС, а будет иметь смысл для действия группы $G$ (по-прежнему аменабельной) на метрическом компакте. Правда, при этом может случиться, что не существует мер, удовлетворяющих этому определению. Но если они существуют, то заслуживают внимания. Следует предупредить, что наша цель здесь всё-таки не статфизика, а теория ДС, поэтому в излагаемых далее построениях даже применительно к решётчатым системам встречается некоторое несоответствие со статфизикой. Можно доказать, что на окончательные результаты для таких систем оно не влияет. Но для нас указания, заимствованные из статфизики, имеют скорее эвристический характер, так что на это несоответствие можно вообще не обращать внимания.

В статистической физике для системы, имеющей конечное число состояний $\alpha_1, \ldots, \alpha_N$ с энергиями $E(\alpha_j)$, вводят так называемую статистическую сумму

$$Z = \sum_{j=1}^{N} e^{-\beta E(\alpha_j)}, \tag{8}$$

где $\beta$ — величина, обратная температуре, выраженной в подходящих единицах; поскольку же температуру издавна измеряют в градусах, то для перехода к «подходящим» единицам её надо умножить на константу Больцмана $k$; итак, $\beta = \frac{1}{kT}$. Для системы макроскопических размеров рассматриваемая в феноменологической термодинамике свободная энергия $F$ выражается через $Z$:

$$F = -\frac{\ln Z}{\beta}. \tag{9}$$

При термодинамическом равновесии (отвечающем данной температуре $T$) состояние $\alpha_j$ имеет вероятность

$$p(\alpha_j) = e^{-\beta E(\alpha_j)}/Z. \tag{10}$$

Одна из возможных характеризаций этого распределения вероятностей такова: при данных $\beta$ и $E(\alpha_j)$ величина

$$\sum p_i E(\alpha_i) + \frac{1}{\beta}\sum p_i \ln p_i \tag{11}$$

достигает минимума по всем распределениям вероятностей $(p_1, \ldots, p_N)$ в точности когда $p_i = p(\alpha_i)$. А если подставить эти $p_i$ в (11), получится как раз $F$.

При $p_i = 0$ в (11), естественно, подразумевается, что $p_i \ln p_i = 0$, поскольку $\lim\limits_{p\to 0} p \ln p = 0$. Значит, можно считать, что в (11) фигурируют только положительные $p_i$. Опустив же в (8) слагаемые с теми $j$, для которых $p_j = 0$, мы можем только уменьшить статсумму $Z$, и если для этой уменьшенной $Z$ будет доказано, что

$$\sum p_i E(\alpha_i) + \frac{1}{\beta}\sum p_i \ln p_i \geqslant -\frac{1}{\beta}\ln Z \quad \text{при } p_i > 0,\ \sum p_i = 1, \tag{12}$$

то тем более это будет справедливо для первоначальной $Z$. Поскольку $\ln x$ — вогнутая функция, т. е. её график обращён выпуклостью вверх, то при любых $x_i > 0$

$$\sum p_i \ln x_i \leqslant \ln \sum p_i x_i.$$

В частности,

$$\sum p_i \ln(e^{-\beta E(\alpha_i)}/p_i) \leqslant \ln \sum p_i e^{-\beta E(\alpha_i)}/p_i = \ln \sum e^{-\beta E(\alpha_i)},$$

а это равносильно (12).

Заметим кстати, что при подстановке в (11) $p_i = p(\alpha_i)$ первое слагаемое получается равным энергии нашей системы в данном равновесном состоянии, а второе — равным $-TS$, где $S := -k\sum p_i \ln p_i$ — её энтропия (измеренная в макроскопических единицах, используемых в феноменологической термодинамике). Таким образом, $F = E - TS$. В феноменологической термодинамике это соотношение является определением $F$.

После этого краткого экскурса в статфизику вернёмся к нашей решётчатой системе или, более общо, к ДС Бернулли (с аменабельной группой $G$). Сперва сделаем одно терминологическое замечание. Бывает, что в разных науках одно и то же слово имеет совсем разный смысл. Если науки далеки друг от друга, это не приводит к недоразумениям, — трудно представить себе, как можно было бы спутать «клетки» в топологии с «клетками» в биологии. Но бывает и так, что слово имеет разный смысл в довольно близких разделах науки, которые со временем соприкасаются и даже частично перекрываются. Тогда терминологию стоит уточнять. Такое случилось с термином «состояние». В теории ДС состояние — это точка фазового пространства. В статфизике состояние — это распределение вероятностей, т. е. мера, в фазовом пространстве. (В конце предыдущего абзаца слово «состояние» один раз употреблялось именно в таком смысле.) Если принять последнюю терминологию, то не надо говорить о точках фазового пространства как о состояниях. В теории решётчатых систем их часто называют «конфигурациями»[50]. Но эта статья написана с иных позиций, и «не назвать ли нам меру мерой?»

Суммарная энергия всей бесконечной решётчатой системы, вообще говоря, бесконечна и работать с ней нельзя. Но разумно считать, что в ней можно как бы выделить часть, приходящуюся на одну частицу, и что для частицы, находящейся в точке $g$, этот вклад является некоторой функцией $E(g,\xi)$ от $g$ и состояния всей решётчатой системы $\xi$. Мы считаем, что взаимодействие инвариантно относительно левых групповых сдвигов. Значит, если в точке $g$ частица находится в состоянии $\xi(g)$, а в точках $gh$ (со всевозможными $h$) — в состояниях $\xi(gh)$ (как это и есть для состояния всей системы $\xi$), то вклад в энергию, приходящуюся на «$g$-ю» частицу, таков же, как вклад в энергию, приходящийся на «$e$-ю» частицу ($e$ — единица группы), если она находится в состоянии $\xi(g)$, а «$h$-е» частицы — в состоянии $\xi(gh)$. Последнее имеет место при состоянии всей системы $\varphi_g(\xi)$. Итак, $E(g,\xi)=E(e,\varphi_g(\xi))$, и нам достаточно рассматривать функцию $f(\xi):=-E(e,\xi)$, которую мы будем считать непрерывной. (Знак «минус» не имеет физической мотивировки, но принят в соответствующих математических работах, слегка упрощая часть формул. Излагаемая теория хотя и навеяна статфизикой, но в конечном счёте является математической, и фигурирующие в ней величины в общем случае никак не могут иметь такого физического смысла, как в обсуждаемом сейчас примере.) Тогда

$$E(g,\xi)=-f(\varphi_g(\xi)). \tag{13}$$

(При более реалистическом подходе для случая $G=\mathbb{Z}^m$ начинают с того, что взаимодействие сводится к парным, тройным, ..., $r$-арным, ... взаимодей-

[50] На что имеются известные основания. В классической механике конфигурация системы — это расположение её частей, без учёта скоростей их движения. А сейчас мы имеем дело с равновесной статфизикой, где о движениях не говорится. (При желании можно вообразить, что если какие-то движения есть, то они являются «скрытыми», «спрятанными внутри состояний» и только эффективно учитываются в фигурирующей ниже энергетической характеристике $f$. Впрочем, физические задачи с конечным числом состояний частиц решётки бывают связаны со спином, а спин вообще не описывается в классических терминах с различением координат и импульсов и с разделением энергии на потенциальную и кинетическую. Раз уж зашла об этом речь, надо сказать, что на самом деле квантовомеханическое описание спина сложнее обсуждаемой здесь примитивной ситуации, где говорится только, что у частицы имеется конечное число состояний (с соответствующей энергетической характеристикой). Тем не менее иногда — и, по-видимому, чаще, чем можно было бы ожидать, — описываемая примитивная модель оказывается дающей достаточно хорошее приближение.)

ствиям; однако в предположении, что все взаимодействия быстро убывают с увеличением числа частиц и расстоянием, результат в конечном счёте оказывается таким же.)

План наших действий будет состоять в том, чтобы сперва рассмотреть конечный «кусок» $S_n$ решётчатой системы, образованный частицами, расположенными в тех точках решётки $\mathbb{Z}^m$, которые лежат в кубе или, более общо, воображаемыми частицами, занумерованными элементами фёлнеровского множества $F_n$. Затем мы хотим сделать некоторый предельный переход при $n \to \infty$. Формула (9) подсказывает, что разумно взять

$$-\frac{\ln Z_{F_n}}{\beta\,\#F_n}, \tag{14}$$

где $\#F_n$, как и ранее, — число элементов $F_n$. В пределе при $n \to \infty$ (существование которого, конечно, надо доказать) получается свободная энергия, приходящаяся на одну частицу. («Суммарная» свободная энергия всей бесконечной решётчатой системы, под которой следовало бы понимать предел $-\ln Z_{F_n}/\beta$, скорее всего, будет бесконечной).

При осуществлении этого плана некоторые детали конкретизируются не совсем так, как это могло бы показаться более естественным.

Во-первых, в (14) опускают знак минус и соответствующий предел называют не «свободной энергией» (приходящейся на одну частицу), а «давлением» и соответственно обозначают через $P$. Если не говорить о «житейском» смысле слова «давление», то в термодинамике для макроскопической непрерывной физической системы под давлением понимают следующее: надо выразить $F$ через объём системы $V$ и температуру $T$ (возможно, и через какие-то ещё параметры); тогда $P = -\partial F(V,T)/\partial V$. Называя взятый с обратным знаком предел (14) давлением, по-видимому, исходят из того, что для решётчатой системы вместо производной $F$ по объёму естественно взять приращение $F$, получающееся при добавлении одной частицы, а это и приводит к (14) (причём как раз без минуса, если приращение, как и производную в случае непрерывной системы, брать с обратным знаком). Однако в самой статфизике о давлении для решётчатых систем, кажется, не говорят.

Во-вторых, обычно полагают $\beta = 1$. Это соответствует изменению функции $f(\xi)$ на постоянный множитель, что не принципиально. (Существенным это становится только тогда, когда по каким-то причинам надо рассматривать не одну $f$, а семейство $\beta f$ с параметром $\beta$. В физике для этого есть серьёзное основание — надо с самого начала иметь в виду, что одна и та же система будет рассматриваться при разных температурах. В теории ДС нет причин с самого начала уделять особое внимание семейству $\beta f$; если же когда-нибудь в этом возникнет нужда (что действительно бывает), ничто не мешает тогда и заменить $f$ на $\beta f$.)

В-третьих (это более серьёзно), $Z_{F_n}$ определяют так. Состояния $S_n$ суть элементы $A^{F_n}$ (т. е. функции $F_n \to A$). Каждую такую функцию $\alpha$ мы каким-нибудь образом продолжаем до отображения $\xi_\alpha : F \to A$ и полагаем (ср. (13))

$$Z_{F_n} := \sum_{\alpha \in A^{F_n}} e^{-\sum\limits_{g \in F_n} E(g,\xi_\alpha)} = \sum_{\alpha \in A^{F_n}} e^{\sum\limits_{g \in F_n} f(\varphi_g(\xi_\alpha))}. \tag{15}$$

В показателе здесь учитывается взаимодействие «$g$-й» частицы (с $g \in F_n$) со всеми остальными частицами «решётки», не только отвечающими элементам $F_n$. С точки зрения статфизики можно было бы, рассматривая выделенный «кусок» $S_n$ как изолированную систему, учесть только взаимодействия частиц из $S_n$ друг с другом. (Тогда и продолжать $\alpha\colon F_n \to A$ до $\xi_\alpha : F \to A$ не понадобилось бы.) Повторяю, что на самом деле (при естественных предположениях о взаимодействиях) это не меняет окончательного результата и что для нас это вообще не так уж важно. $Z_{F_n}$ зависит от конкретного выбора продолжений $\alpha$ до $\xi_\alpha$. Но оказывается, что при любом таком выборе существует предел

$$P(f) = \lim_{n\to\infty} \frac{1}{\# F_n} \ln Z_{F_n}, \tag{16}$$

который не зависит ни от конкретного выбора $\xi_\alpha$, ни от использованной фёлнеровской системы. Число $P(f)$ называют топологическим давлением, отвечающим непрерывной функции $f\colon \Omega \to \mathbb{R}$. (Оно, конечно, зависит от $G$ и $A$, а в более общей ситуации, к которой мы вскоре перейдём, — от рассматриваемой ДС, но обычно они считаются фиксированными и потому не указываются явно в обозначении для $P$.) Выражение (11), как будет пояснено ниже, тоже имеет естественный аналог для решётчатой системы («приходящийся на одну частицу»), в котором вместо конечного распределения вероятностей $(p_1, \ldots, p_N)$ фигурирует мера $\mu$ в фазовом пространстве, инвариантная относительно $\{\varphi_g\}$. Естественно сравнить этот аналог с $P(f)$ (т. е. с точностью до знака, со свободной энергией, приходящейся на одну частицу), и обратить внимание на те меры, при которых эти две величины совпадают (если такие меры существуют).

Пока что речь шла о топологической ДС Бернулли (это использовалось, когда говорилось о её «куске» $S_n$ и о продолжении состояния $\alpha$ последнего до $\xi_\alpha$). Теперь мы перейдём к общей топологической ДС с аменабельной группой $G$ и компактным фазовым пространством $X$, считая также заданной непрерывную $f\colon X \to \mathbb{R}$.

Пусть $\mathscr{U} = \{U_1, \ldots, U_l\}$ — конечное открытое покрытие $X$. Положим $\varphi_g^{-1}\mathscr{U} := \{\varphi_g^{-1}U_1, \ldots, \varphi_g^{-1}U_l\}$,

$$\mathscr{U}_{F_n} := \bigvee_{g \in F_n} \varphi_g^{-1}\mathscr{U},$$

где использовано следующее обозначение: для нескольких конечных покрытий $\mathscr{U}_t$, $t \in T$, через $\bigvee_t \mathscr{U}_t$ обозначается покрытие, элементы которого суть всевозможные непустые пересечения $\bigcap_{t\in T} U_{k_t}$ с $U_{k_t} \in \mathscr{U}_t$. Положим

$$Z_{F_n,\mathscr{U}} := \sum_{U\in\mathscr{U}_{F_n}} e^{\sup\limits_{x\in U}\sum\limits_{g\in F_n} f(\varphi_g x)}. \tag{17}$$

$$P(f) := \sup_{\mathscr{U}} \lim_{n\to\infty} \frac{Z_{F_n,\mathscr{U}}}{\#F_n}, \tag{18}$$

где $\sup_{\mathscr{U}}$ берётся по всевозможным конечным покрытиям пространства $X$. Такое определение, в отличие от (15), (16), уже не связано с какими-то специфическими фазовым пространством и групповым действием. В то же время для бернуллиевской ДС оно приводит к тому же числу $P(f)$. (Различия между (17), (18) и (15), (16) в основном сводятся к тому, что для бернуллиевской ДС используется только одно покрытие

$$\mathscr{U} = \{U_a;\ a \in A\}, \quad \text{где} \quad U_a := \{\xi;\ \pi_0(\xi) = a\}; \tag{19}$$

заметим, что при этом

$$\mathscr{U}_{F_n} = \{U_\alpha;\ \alpha \in A^{F_n}\}, \quad \text{где} \quad U_\alpha := \{\xi;\ \xi|_{F_n} = \alpha\},$$

так что прежние $\xi_\alpha \in U_\alpha$. Оказывается, что в данном случае на $\mathscr{U}$ из (19) как раз и достигается фигурирующая в (18) верхняя грань $\sup_{\mathscr{U}}$. Кроме того, в правой части (15) фигурируют произвольно выбранные $\xi_\alpha \in U_\alpha$, а в (17) в показателе берётся верхняя грань по всем таким $\xi_\alpha$; оказывается, что для данных ДС и $\mathscr{U}$ это не меняет предела $\dfrac{1}{\#F_n}\ln Z_{F_n,\mathscr{U}}$.)

Мы получили некий аналог «свободной энергии на одну частицу». Теперь займёмся аналогом (11). Сумма $\sum p_i \ln p_i$ — это, конечно, взятая с обратным знаком энтропия[51)] распределения вероятностей $(p_1, \ldots, p_N)$. Аналогом последнего для общей ДС является инвариантная нормированная мера $\mu$, а аналогом энтропии является упомянутая в п. *б* энтропия $h_\mu$. Аналогом же (средней) энергии, приходящейся на одну частицу, является $-\int_X f\,d\mu$ (знак «минус» вызван «минусом» в определении $f$). Итак, аналогом (11), взятым с обратным знаком, является величина

$$h_\mu + \int_X f\,d\mu. \tag{20}$$

(об энтропии $h_\mu$ мы уже говорили в пункте *б*). Для частного случая ДС Бернулли можно было бы более детально проследить переход от (11) к (20), аналогично тому как это было сделано для $P(f)$. Но для $P(f)$ это давало

---

[51)] В теории информации естественно использовать не натуральный, а двоичный логарифм. Но в статфизике, как и в более аналитических математических вопросах, используют натуральный логарифм, от чего соответствующие величины только умножаются на некоторый постоянный множитель.

эвристическую мотивировку определения (которое иначе казалось бы «взятым с потолка»), в чём нет нужды для обоих слагаемых в (20). (На самом деле нечто в этом роде можно было бы привести в связи с опущенным в *б* определением $h_\mu$. Я не стал этого делать, поскольку для ДС с классическим временем определение $h_\mu$ достаточно широко известно, а переход в определении к общему случаю является непосредственным.)

После всего сказанного читатель, вероятно, найдёт естественной следующую теорему (которую, однако, всё же надо доказывать, и это отнюдь не просто):

$$\sup_\mu \Big( h_\mu + \int_X f\, d\mu \Big) = P(f), \tag{21}$$

где верхняя грань берётся по всем инвариантным нормированным мерам рассматриваемой топологической ДС. В отличие от простого случая системы с конечным числом состояний, верхняя грань не всегда достигается, а если достигается, то не всегда на единственной мере. Если верхняя грань достигается, то соответствующие меры называют равновесными мерами (а часто также и состояниями равновесия, опять-таки используя слово «состояние» в смысле статфизики, а не теории ДС). Соотношение (21) называют «вариационным принципом для топологического давления». Изучение соответствующих вопросов составляет предмет так называемого «термодинамического формализма» для ДС.

Я отмечу здесь только одно. В случае ДС Бернулли с $G = \mathbb{Z}^m$ существование равновесной меры доказано для любой непрерывной функции $f$, но при этом имеется существенное различие между случаями $m = 1$ и $m > 1$. В первом случае равновесная мера для гёльдеровской $f$ единственна, а во втором — не обязательно. Для решётчатой системы неединственность равновесной меры отвечает известному (не только людям, но и животным[52]) физическому явлению — фазовому превращению.

Для ДС с классическим временем основы термодинамического формализма в духе только что намеченного подхода были заложены П. Уолтерсом и Д. Рюэллем в начале 70-х гг.; в частности, они доказали вариационный принцип. Для Рюэлля это было тем более естественно, что он ещё раньше много занимался математическими вопросами статистической физики. Поэтому он был также первым или одним из первых, кто стал распространять эту теорию на действия $\mathbb{Z}^m$. Переход к общим аменабельным группам $G$ был осуществлён А. М. Стёпиным и А. Т. Таги-заде в [42].

Имеется другой подход к выделению и исследованию «интересных» мер, предложенный для ДС с классическим временем Я. Г. Синаем тоже около

[52] Правда, животным оно известно для непрерывной системы (вода), а с фазовым превращением в решётчатых системах познакомились только люди, когда стали изучать магнитные и электрические явления в кристаллах.

1970 г. [43] (причём он сразу отметил, что этот подход годится и для неклассического времени, и в виде примера указал, что для ДС Бернулли так можно получить меры, возникающие в статфизике). Соответствующие меры называются «гиббсовскими».[53] Гиббсовская мера для ДС $\{\varphi_g;\ g \in G\}$ в компакте $X$ строится по заданным нормированной инвариантной мере $\mu_0$ и функции $f \in L^\infty(X, \mu_0)$. По этим данным строится последовательность мер $\mu_n$, абсолютно непрерывных относительно $\mu_0$ с «плотностью» (т. е. производной Радона—Никодима)

$$\frac{e^{\sum\limits_{g\in F_n} f(\varphi_g x)}}{\int_X e^{\sum\limits_{g\in F_n} f(\varphi_g y)}\, d\mu_0(y)}.$$

Гиббсовские меры — это предельные точки последовательности мер $\mu_n$ (в смысле слабой сходимости мер). Они не обязательно инвариантны[54]. Но если гиббсовская мера является не предельной точкой, а пределом, то она инвариантна. Инвариантные гиббсовские меры являются равновесными, обратное же не обязательно. Для ДС с классическим временем и гиперболическим поведением траекторий обратное имеет место для «хороших» $f$.

При исследовании равновесных и гиббсовских мер ДС с гиперболическим поведением траекторий используется приём «кодирования», позволяющий в некотором смысле представить рассматриваемую ДС как подсистему бернуллиевской ДС. Буквально подсистема последней была бы инвариантным подмножеством $A$ в соответствующем $\Omega$; в топологическом контексте естественно рассматривать замкнутые $A$. Оговорка «в некотором смысле» связана с тем, что такое $A$ нульмерно, так что система с фазовым пространством бóльшей размерности никак не может быть топологически эквивалентной ограничению $\{\varphi_g|_A\}$ ДС Бернулли на $A$. Но она вполне может быть факторсистемой последней, т. е. получаться из неё при факторизации по некоторому отношению эквивалентности, инвариантному относительно ДС. В удачных случаях $A$ и факторизация допускают достаточно подробное описание, причём при факторизации точки $A$ хотя и «склеиваются» друг с

[53] Они действительно похожи на гиббсовские распределения в классической статфизике и ещё более — на «гиббсовские ДЛР-меры» для бесконечных систем типа $A^{\mathbb{Z}_n}$, введённые Добрушиным, Ланфордом и Рюэллем. Для ДС с классическим временем приводимое ниже определение уже «прошло проверку практикой», но для неклассического времени такая проверка в бóльшей степени ещё предстоит (кроме случая, известного из статфизики; но и в этом случае, может быть, предстоит выяснить различные аспекты связи приводимого определения с конструкцией ДЛР).

[54] В статфизике меры, не являющиеся трансляционно инвариантными, могут быть вполне естественными: достаточно представить себе, что в одном полупространстве имеется одна фаза, а в другом — другая. Это столь же реально, как и тот случай, когда всё пространство заполнено одной фазой и соответствующая мера трансляционно инвариантна.

другом, но это происходит не очень часто — «большинство» точек ни с чем не склеивается. Этот приём (когда он срабатывает) позволяет сводить ряд вопросов об исследуемой ДС к вопросам, относящимся к $\{\varphi_g|_A\}$. Успех, конечно, зависит от того, удастся ли найти «удачное» кодирование. Для систем с наиболее чётко выраженным гиперболическим поведением траекторий (системы Аносова и базисные гиперболические множества) удачное кодирование связано с так называемыми марковскими разбиениями, которые для гиперболических автоморфизмов двумерного тора были введены Р. Адлером и Б. Вейссом, для систем Аносова — почти одновременно Я. Г. Синаем и затем в усовершенствованном виде, пригодном также и для базисных гиперболических множеств — Д. Рюэллем и Р. Боуэном [44].

Синай, Рюэлль и Боуэн не только построили новые инвариантные меры, но и специально выделили те случаи, когда эти меры представляют особый интерес. Впоследствии их подход[55)] переносился на другие типы систем с несколько (ненамного) «ухудшенной» гиперболичностью — биллиарды, псевдоаносовские гомеоморфизмы поверхностей, аттракторы типа аттрактора Лоренца. Что же касается действий аменабельных групп, то сколько-либо продвинутых применений этого подхода за пределами решётчатых систем мне не известно.

Хотя выше почти всё время отмечалось особое положение аменабельных групп, на самом деле работы по эргодической теории (или близким вопросам статфизики), где группы не аменабельны, не исчерпываются теми немногими работами такого характера, которые упоминались (или подразумевались) выше. Имеется ряд работ, в которых, в отличие от того, что делается для аменабельных групп, используется какая-то дополнительная структура. Например, могут фиксироваться образующие элементы группы[56)], какое-нибудь разбиение или покрытие фазового пространства. (Ведь и для $\mathbb{Z}^m$, рассматривая бернуллиевскую ДС, мы начали с использования фиксированного покрытия (оно же в данном случае и разбиение) (19) — вернее, там это было настолько естественно, что мы вначале даже не отмечали этого специально. О других разбиениях мы заговорили только при переходе к более общим групповым действиям.) Однако пока что было бы затруднительно охарактеризовать исследования такого рода какими-то общими чертами.

---

[55)] Собственно, формально их подходы не совсем совпадают — у Боуэна вообще нет общего понятия гиббсовской меры (которое само по себе, как и понятие равновесной меры, не связано с кодированием), а он просто называет так те меры, которые строит для рассматриваемых им систем. Но с более широкой точки зрения это всё-таки один подход.

[56)] Уже упоминавшийся выше изменённый способ осреднения по группе, предложенный Р. И. Григорчуком, формально зависит от выбора её образующих. Зависит ли от этого результат, в общем случае неизвестно (но в эргодическом случае не зависит).

Новых обзоров по теме п. 1.3 я не знаю. О ситуации в начале рассматриваемого периода см. [45].

**1.4. Бифуркации.** За последнюю четверть века преобразился облик и изменилась роль одного из разделов теории гладких ДС — теории бифуркаций. Буквально это слово означает «раздвоение»; в этом смысле оно употребляется, например, в анатомии («бифуркация бронха»). В математике этот термин употребляется в более широком смысле — для обозначения качественных изменений рассматриваемых объектов при изменении параметров, от которых эти объекты зависят. Более точных общих формулировок здесь дать нельзя, потому что рассматриваемые объекты и интересующие исследователей свойства этих объектов могут быть самыми различными. Точные формулировки относятся к тем или иным конкретным задачам.

Первоначально в математике о бифуркациях говорили применительно к фигурам равновесия вращающейся жидкости. Речь идёт о задаче: при каких условиях тело, состоящее из однородной жидкости, на частицы которой действуют только силы их взаимного притяжения по закону тяготения Ньютона, может вращаться как твёрдое тело? Соответствующую фигуру и называют равновесной. Как точные решения этой задачи известны только некоторые эллипсоидальное фигуры (эллипсоиды Маклорена и Якоби) и кольца, но кроме того известно, что имеются другие фигуры, близкие к названным. Эти фигуры обнаружены с помощью бифуркационных соображений — начинают с эллипсоидальной фигуры равновесия $E_\lambda$, непрерывно зависящей от некоторого параметра $\lambda$, от которого зависит исследуемая задача; оказывается, что когда $\lambda$ проходит некоторое (как говорят, «бифуркационное») значение $\lambda_0$, то возникает новая (уже не обязательно эллипсоидальная) фигура равновесия $E'_\lambda$, которая тем ближе к $E_\lambda$, чем ближе $\lambda$ к $\lambda_0$, так что можно сказать, что семейство фигур $E'_\lambda$ «ответвляется» при $\lambda = \lambda_0$ от семейства $E_\lambda$. В этом случае слово «бифуркация» употребляется в смысле, достаточно близком к буквальному — при увеличении параметра $E_\lambda$ как бы раздваивается на $E_\lambda$ и $E'_\lambda$. Аналитически вопрос сводится к исследованию некоторого интегрального уравнения[57)], и поэтому естественно, что «бифуркационная» терминология стала применяться вообще при исследовании интегральных уравнений, зависящих от параметра. А поскольку видным деятелем в теории фигур равновесия вращающейся жидкости был А. Пуанкаре, то не удивительно, что он перенёс эту терминологию в качественную теорию обыкновенных дифференциальных уравнений, при этом он стал применять её в более широком смысле, для любых качественных изменений.

По инициативе Р. Тома вместо бифуркаций говорят о «катастрофах». Это слова тоже не надо понимать буквально. Приведу примеры, действительно

[57)] Довольно необычного: в нём подинтегральное выражение известно, а неизвестной является область интегрирования.

серьёзно рассматривавшиеся в работах по «теории катастроф»: если нарушается устойчивость упругой конструкции, то это, скорее всего, катастрофа, но если солнечные лучи, преломляясь в воде, образуют на дне ручья яркие линии — это едва ли кого-нибудь волнует, кроме разве детей, видящих это впервые. Как возможный (но не доведённый до математической модели[58)]) пример «катастрофы» упоминают о резком изменение в течении болезни, после которого больной почти на глазах начинает поправляться; это если и катастрофа, то только для бактерий.

Если катастрофа — синоним бифуркации, то можно спросить, какой термин удачнее. Как ясно из сказанного, ни тот, ни другой не приходится понимать буквально. Но «катастрофа» — слово обычного (литературного и разговорного) языка, имеющее определённый и притом весьма эмоционально окрашенный смысл, а о первоначальном значении слова «бифуркация» знает намного меньше людей, и даже у них с ним едва ли связаны какие-то эмоции. Поэтому для науки более подходит нейтральное слово «бифуркация», а для массовых изданий — «катастрофа».

Весьма содержательное математические идеи Р. Тома об особенностях гладких отображений и бифуркациях критических точек функций (он продолжал пионерские работы Х. Уитни, но пошёл значительно дальше) и по содержанию, и по времени заметно выходят за пределы данной статьи; последующие работы в этой области тоже выходят за эти пределы (если не по времени, то по содержанию). Их непосредственные применения в теории динамических систем таковы. Рассматриваются системы, у которых одни переменные (скажем, $y$; это, вообще говоря, не число, а вектор) изменяются быстро, а другие (скажем, $x$) — медленно (такие системы действительно встречаются и являются достаточно важными); предполагается, что при фиксированным $x$ для «быстрых» переменных $y$ получается градиентная система

$$\frac{dy(t)}{dt} = -\nabla f(x, y)$$

(так тоже бывает, но уже реже), вследствие чего $y(t)$ быстро приближается к устойчивому положению равновесия последней, т. е. к критической точке $f$ как функции от $x$, являющейся локальным минимумом. Далее движение происходит таким образом, что $x(t)$ постепенно изменяется, при этом изменяется и соответствующая критическая точка функции $y \mapsto f(x, y)$, с каковой почти точно совпадает $y(t)$. В один прекрасный момент с критической точкой происходит бифуркация, — скажем, она сливается с другой критической точкой и исчезает, после чего $y(t)$ должно устремиться к другой критической точке. Ясно, что в такой ситуации результаты о бифуркациях

[58)] Точнее, математические модели подобного явления предлагались, но, кажется, они не были связаны с «теорией катастроф» и исследовались численно.

критических точек функций весьма существенны для понимания качественной картины, но на данном уровне приближения речь идёт только о довольно тривиальной более или менее непосредственной ссылке на эти результаты. Далее утверждается, что и не зная дифференциальных уравнений, а только предполагая, что они имеют описанный характер, и наблюдая изменения в реальной «физической» системе[59], можно сделать качественные выводы об особенностях функции $f$ и тем самым понять наиболее существенные свойства данной системы. В конечном счёте речь идёт о том, что предлагается некоторая гипотетическая интерпретация экспериментальных данных. Никаких иных оснований не приводится (легко поверить, что имеются быстрые и медленные переменные, но какие экономические, психологические, социальные законы позволяют думать, что быстрые переменные должны двигаться по градиенту какой-то функции? если так, то она, видимо, сама имеет какой-то экономический, биологический, психологический, социологический смысл?) Всё это имеет подозрительно натурфилософский стиль[60], а натурфилософия фактически устарела уже при Ньютоне, хотя процветала ещё более века[61].

---

[59] Она может относиться и не к физике, а, скажем, к экономике, биологии, даже, как говорят, к психологии или социологии. Для систем, относящихся к физике, обычно более или менее известны их математические модели или по крайней мере общий характер последних. В этих случаях применения «теории катастроф» не вызывают сомнений.

[60] Конечно, психологические, экономические или социологические вопросы в старину к натурфилософии не относили, как явствует из самого её названия. Но я и говорю только о стиле.

[61] Раз уж зашла об этом речь, отмечу ещё следующее. Имеется известный приём иллюстрации взаимоотношений между логическими понятиями — круги Эйлера—Венна. Скажем, мы хотим проиллюстрировать взаимоотношение между понятиями «домашние животные», «млекопитающие», «кошки», «собаки». Рисуем два частично пересекающихся «круга» (которые могут быть и овалами, если их удобнее будет рисовать) — «млекопитающие» и «домашние животные», внутри первого рисуем два круга — «кошки» и «собаки», которые не пересекаются друг с другом, но оба частично пересекаются с «домашними животными». При этом никто не думает, будто бы множество животных в каком-то смысле двумерно и что четыре интересующих нас множества суть и вправду круги или овалы. Не могут ли иногда те рисунки, которые рисуют «катастрофисты», интерпретироваться как какие-то условные изображения каких-то взаимоотношений, не обязательно связанных с теми конкретными моделями, о которых при этом говорят, вроде того как для кругов Эйлера—Венна несущественно, что это именно круги и притом на плоскости? Это можно предполагать, если фактическая сторона дела в экономических и прочих вопросах действительно такова, как описывают «катастрофисты», а модели в этих случаях сомнительны, как говорят их критики. Том как раз и полагал, что теория катастроф даёт нам некий новый язык форм и что более сложные системы и взаимоотношения между ними

В брошюре В. И. Арнольда [46] имеется несколько страниц, где всё это описано несколько подробнее, но со столь же критических позиций, с которых написаны предыдущие строки[62]. Там же имеются литературные ссылки. С иных позиций написана популярная статья [48], где отношение к нефизическому использованию «катастрофической» идеологии положительное. Наконец, в книге [49] авторы на базе подробно и элементарно изложенных в первой её половине сведений из теории особенностей рассмотрены разнообразные приложения теории катастроф — от не вызывающих принципиальных возражений до куда более «натурфилософских» (по стилю).

Из сказанного не следует, будто теория особенностей гладких отображений и бифуркаций критических точек функций мало что дала для теории ДС, а там, где дала, это получалось путём довольно непосредственного использования достижений топологов. Значительным оказалось идейное влияние первой теории на вторую, провозвестником которого около 1970 г. выступил В. И. Арнольд. Надо сказать, что Том с самого начала утверждал, что та «теория катастроф», о которой говорилось выше, — это только первая, элементарная часть некоей более обширной теории (вот она-то уж точно будет универсальной). Но это общее указание им никак не конкретизировалось, и Арнольд отправлялся не от этого неопределённого указания (заклинания), а от конкретного фактического содержания теории особенностей гладких отображений. Он никогда не претендовал на универсальность соответствующего подхода (к моменту его выступления уже были известны факты, исключающие таковую), но резонно указывал, что он имеет довольно широкую, хотя и не безграничную, область применений, а что касается её границ, то в некоторых случаях их даже можно было уже тогда довольно чётко очертить. Своего рода первым манифестом нового веяния была его статья [50], в которой объяснялось, что в теорию локальных бифуркаций ДС можно естественным образом перенести ряд понятий, первоначально возникших в теории особенностей или вообще в гладкой топологии — коразмерность, стратификация, трансверсальность, универсальные и версальные семейства, модули и их число, бифуркационные диаграммы, конечная определённость.

---

как бы строятся из элементарных блоков, описываемых согласно сказанному выше посредством отдельных систем с отдельными особенностями из некоего определённого списка. (Кстати, сам этот список оказался неполным.) Но если нет оснований считать, что изучаемая физическая система действительно такова, как сказано (т. е. с медленными и быстрыми переменными, причём последние, часть которых нам неизвестна, изменяются по градиенту), то как ещё могли бы интерпретироваться эти блоки и рисунки?

[62] Основное содержание этой брошюры, конечно, состоит не в критическом обсуждении теории катастроф, а в изложении понятий теории особенностей, её приложений, её истории и ранней истории теории бифуркаций ДС. Ещё меньше о теории катастроф и больше об остальном сказано в его обзоре [47].

Я немного остановлюсь только на одной (и, вероятно, самой простой) из пропагандировавшихся в [50] идей: вырождения коразмерности $k$ неустранимым образом встречаются только в $k$-параметрических семействах[63], поэтому целесообразно рассматривать такие положения равновесия в соответствующем бифуркационном контексте[64]. (Впрочем, могут быть и другие причины появления вырождений довольно высокой коразмерности, — прежде всего, наличие симметрий. Множество положений равновесия, имеющее большую коразмерность в классе «общих» ДС, может быть множеством положений равновесия «с симметриями», коразмерность которого в классе ДС с соответствующими симметриями невелика или даже равна 0. Но всегда ли мы можем быть уверены в точности этих симметрий, особенно если они не являются следствиями законов природы? Не могут ли симметрии в действительности слегка нарушаться, и не стоит ли поинтересоваться, что происходит при таком нарушении?[65]) Поэтому теория локальных бифуркаций — не какое-то внешнее дополнение к локальной качественной теории, а существенная её часть. Уже одно это изменение точки зрения (пусть и не совсем новое, но впервые чётко намеченное в столь широких масштабах) существенно повлияло на облик не только теории бифуркаций, но и всей локальной качественной теории.

Разумеется, в [50] могла быть реализована только часть намеченной программы. (Хотя ко времени написания [50] группой Арнольда уже была проведена определённая работа в этом направлении. Кроме того, в [50] было указано, как в новую теорию входят ранее полученные значительные результаты[66].) После публикации [50] как группой Арнольда, так и другими математиками была проведена большая работа по конкретной реа-

---

[63] Это значит, что существуют такие $k$-параметрические семейства ДС, что и для них, и для всех достаточно близких к ним семейств при некотором значении параметра в соответствующей ДС имеется вырождение рассматриваемого типа. («Близость» понимается как близость в смысле $C^m$, где $m$ зависит от того конкретного типа вырождений, о котором идёт речь.) Если же имеется $(k-1)$-параметрическое семейство, то путём сколь угодно малого его возмущения можно получить семейство ДС, ни одна из которых не имеет вырождения данного типа.

[64] Несомненно, что авторы «классических» работ (в понятных обозначениях, появившихся в период $(-\infty, t_{[50]})$) это «чувствовали» (как «чувствовали» они и многое другое из новой парадигмы). Но они давали на сей счёт разъяснения (если вообще их давали) только применительно к интересовавшим их задачам и к рассматривавшимся при этом случаям малой коразмерности. По большей части это как-то проплывало мимо сознания более широких кругов.

[65] Некоторая работа такого рода проводилась как в специальном контексте теории катастроф (информацию см. в [49]), так и вне него.

[66] Частично этот материал более детально изложен в конце учебника Арнольда [51].

лизации нового подхода. Для потоков на плоскости «типичные» локальные бифуркации[67] в двух- и трёхпараметрических семействах изучены почти с такой же степенью подробности, с какой раньше были изучены локальные бифуркации в однопараметрических семействах. В случаях бо́льшей размерности картина до сих пор не столь исчерпывающая, но всё же довольно полная, потому что вопрос о локальных бифуркациях в таких семействах в значительной степени сводится к аналогичному вопросу для потоков на плоскости[68].

Наряду с локальными бифуркациям можно говорить о «глобальных» бифуркациях, меняющих фазовую картину в целом и не локализованных возле положений равновесия или периодических траекторий. В большинстве случаев правильнее бы было говорить о «полулокальных» бифуркациях, имея в виду, что при исследовании таковых мы обращаем внимание не только на то, что происходит при возмущении возле какого-то «локального» объекта, но на то, что происходит в некоторой области фазового пространства («вдали» от локального объекта, если таковой вообще играет какую-то роль), обычно в некоторой окрестности некоторого инвариантного множества невозмущённой системы, не сводящегося к положению равновесия и т. п. Однако при этом, вообще говоря, не идёт речи о полном контроле за всем фазовым пространством, так что название «глобальная бифуркация» является, пожалуй, несколько преувеличенным. С другой стороны, в классических работах А. А. Андронова и его сотрудников (восходящих ещё к довоенным временам, хотя частично опубликованных позднее) рассматривались изменения всего «фазового портрета» потока на всей фазовой плоскости, т. е. самые настоящие глобальные бифуркации. Впрочем, хотя это восходит к упомянутым старым вопросам, по-видимому, только в последние 25 лет или около того для потоков на плоскости окончательно выяснилось, что надо добавить к прежним сведениям о локальных или полулокальных бифуркациях, чтобы получить полное описание изменений глобальной качественной картины при бифуркациях в типичных однопараметрических семействах.

Изложение результатов, достигнутых по этим (локальным и полулокальным) вопросам к середине 80-х гг., имеется в [52]; более новых публикаций со сколько-либо обширными сводками результатов я не знаю. Для «типичных»

[67] Т. е. бифуркации, относящиеся к объектам локальной качественной теории — положениям равновесия и периодических траекторий потоков, неподвижным и периодическим точкам диффеоморфизмов.

[68] В ряде случаев благодаря использованию центрального многообразия автоматически получается полная редукция к потокам на плоскости, но иногда центральное многообразие имеет размерность 3 или 4; в этом случае ситуация сложнее. Всё же искусственным образом удаётся произвести некоторую редукцию к двумерному случаю, хотя и не столь полную.

двух- и трёхпараметрических потоков на плоскости переход от локальных и полулокальных бифуркаций к «настоящим» глобальным, кажется, не исследован. Для таких потоков после [52] были подробно исследованы бифуркации так называемых «полициклов», обобщающие бифуркации замкнутых сепаратрисс [53].

Другое значительное изменение в теории бифуркаций связано с исследованием таких полулокальных (а иногда и глобальных) бифуркаций в размерностях $> 2$ для диффеоморфизмов и $> 3$ для потоков, которые связаны со сложным поведением траекторий[69]. Это — принципиально новое направление. Ему уделено должное внимание в [52], однако здесь последующее развитие представляется более значительным. К сожалению, приходится повторить сказанное по другому поводу: более новых обзоров сравнимой полноты нет.

К началу 70-х гг. динамические системы со сложным поведением траекторий более или менее успешно исследовались в тех случаях, когда это поведение определялось гиперболичностью. Уже тогда, а тем более теперь гиперболическая теория в какой-то степени приобрела известную завершённость, хотя в ней остаются нерешённые задачи и работы на эту тему появляются и поныне. Баланс работ определённо сместился в сторону бифуркационной тематики. Это связано с открытием комплекса явлений, связанных с нетрансверсальными гомоклиническими траекториями. Он настолько обширен и разнообразен, что его изучение далеко от завершения хотя бы в первом приближении. Первым обратил внимание на принципиально новые бифуркационные явления, так или иначе связанные с нетрансверсальными гомоклиническими траекториями, Л. П. Шильников в совместной с Н. Г. Гавриловым работе 1972 г. [54] (а отчасти в 1970 г.). Конкретно там были обнаружены четыре различных типа бифуркаций, из которых один происходит на границе систем с простой динамикой — систем Морса—Смейла. В этом случае дано символическое описание возникающих гиперболических множеств и были указаны вторичные бифуркации, при которых рождаются и исчезают устойчивые периодические траектории. Чем ближе значение параметра к критическому, тем большее число таких траекторий может появиться[70].

---

[69] Исследовались, конечно, и полулокальные бифуркации, более или менее непосредственно аналогичные соответствующим бифуркациям для потоков на плоскости, которые с таким поведением траекторий не связаны. Их исследование — необходимая составная часть теории (и о них тоже говорится в [52]), но для нас они не столь интересны.

[70] В [54] рассматривался трёхмерный поток (что фактически покрывает и случай двумерного диффеоморфизма). Многомерные аналоги ситуации из [54] были рассмотрены в [55], [56], [57], причём в двух последних работах исследование идёт дальше

После этого наметилась возможность изучения сложного поведения траекторий в некоторых ситуациях, находящихся за пределами «чистой» гиперболической теории. Едва ли будет преувеличением сказать, что в настоящее время теория бифуркаций стала основным (хотя и не единственным) источником примеров ДС со сложным поведением траекторий, в каком-то смысле выдерживающим малые возмущения (если не все, то «многие»). Для контраста можно отметить, что в предшествующий период гиперболические множества были открыты вне связи с теорией бифуркаций; в основном то же относится и к упоминаемому ниже аттрактору Лоренца. Он был обнаружен в результате исследования соответствующей ДС в широком диапазоне значений параметра, от которого она зависит, и при этом были отмечены различные происходящие в ней бифуркации, в том числе и связанные с аттрактором Лоренца, но его свойства рассматриваются безотносительно к бифуркациям[71)].

Примерно тогда же Ш. Ньюхаус (вначале — только для двумерного диффеоморфизма) обратил внимание на нетрансверсальные гомоклинические траектории гиперболических множеств, не сводящихся к периодическим траекториям [59]. Позднее он назвал гиперболическое множество $A$ (обычно локально максимальное[72)] и топологически транзитивное), которое не сводится к периодическим траекториям и у которого имеется нетрансверсальная гомоклиническая траектория (так что неустойчивое и устойчивое многообразия некоторой траектории из $A$ где-то касаются друг друга), диким гиперболическим множеством (а о существовании в ДС такого множества говорят как о дикой гиперболичности). Определение «дикое» намекает на тот факт, что с таким множеством связан ряд неожиданных бифуркационных явлений. Сперва Ньюхаус обнаружил пример, в котором дикая гиперболичность сохраняется при любом малом

---

аналогов результатов [54]. В [54], как и в ряде других работ в этой области, предполагалось, что отображение последования для исходной периодической траектории гладко сопряжено с линейным; в [55], [56], [57] такого предположения не делается.

[71)] Отчасти сказанное относится и к весьма важному аттрактору Эно, введённому независимо от теории бифуркаций. Однако со временем он оказался тесно связанным с последней: во-первых, он возникает (и играет важную роль) при некоторых бифуркациях; во-вторых (исторически это обнаружилось раньше), описание его свойств получается при некоторых значениях параметров, от которых он зависит; нет сомнений, что при других значениях параметров свойства могут быть другими, хотя (пока?) это подробно не исследовано. Я не говорю здесь об аттракторе Эно единственно по той причине, что о нём говорится у Йоккоза [1]; см. также его доклад на семинаре Бурбаки [58].

[72)] Это значит, что $A$ является максимальным инвариантным множеством в некоторой своей окрестности. Локально максимальные инвариантные множества (не обязательно гиперболические) называют также изолированными.

возмущении[73)], хотя само это множество $A$ — вроде канторового, так что между устойчивыми многообразиями его траекторий имеются не содержащие таких многообразий «полосы», куда, казалось бы, можно при малом возмущении «убрать» все дуги неустойчивых многообразий траекторий из $A$, имеющие «неподходящие» направления (эти многообразия ведь тоже «не идут сплошняком»). В $A$ плотны периодические траектории, и можно показать, что у них при всюду плотном множестве значений параметров имеются нетрансверсальные гомоклинические траектории. Таким образом, Ньюхаус обнаружил, что в пространстве ДС имеются области, в которых всюду плотны ДС с нетрансверсальными гомоклиническими траекториями. Теперь их называют «областями Ньюхауса». В понятном смысле говорят также об «областях Ньюхауса» в «типичных» конечнопараметрических семействах ДС[74)].

Дальнейшие исследования были направлены как на обнаружение областей Ньюхауса в тех или иных ситуациях, так и на исследование свойств ДС из таких областей. Оба вопроса были отчётливо поставлены в [60], где утверждалось, что: а) при определённых условиях вблизи системы с гомоклиническим касанием существуют области Ньюхауса; б) в таких областях существуют ДС со счётным множеством устойчивых периодических траекторий, причём таких ДС в некотором смысле «много». (Очевидно, само по себе это сильнее и удивительнее, чем обнаруженное в [54] увеличения числа таких решений при приближении параметра к бифуркационному значению, но если сопоставить [54] и утверждение «а», то это уже не покажется таким уж удивительным).

После Ньюхауса бифуркации, связанные с нетрансверсальными гомоклиническими траекториями, стали популярны во всём мире. Однако в [60] утверждение «а» было скорее угадано (тогда как «б», видимо, можно считать более или менее доказанным, хотя, как выяснилось позднее, и при излишне ограничительных предположениях). Очень существенный прогресс был достигнут в фундаментальной работе Ньюхауса [61] (см. также дополняющую её статью К. Робинсона [62]). После [61], [62] описанные выше утверждения для двумерных диффеоморфизмов можно считать в основном доказанными и даже улучшенными благодаря снятию излишних условий (хотя, по-видимому, некоторые из более поздних публикаций можно квалифицировать

[73)] В этом пункте я скорее описываю результаты, чем даю точные формулировки. В частности, малость возмущения понимается в смысле некоторого $C^r$, но я ничего не говорю об этом $r$ (в различных случаях оно может быть различным)

[74)] Должен предупредить, что в точных формулировках, относящихся к областям в пространстве ДС и к областям в семействах, имеются некоторые различия. Приводя только приблизительное описание этих результатов, я игнорирую эти различия как «технические».

как окончательную доработку вопросов, восходящих к середине 70-х гг.). Работы [63] и [64] посвящены некоторым многомерным аналогам той же группы вопросов (ослабление условий [63] достигнуто в [65]).

Параллельно был исследован ряд других вопросов о нелокальных бифуркациях. Сложность качественной картины в подобных задачах делает сомнительным, чтобы здесь можно было бы получить полное описание этой картины при фиксированных значениях параметра и полное описание её изменения при изменении параметра. Часто описание, не вполне полное, даётся только для некоторого множества значений параметра; это, конечно, особенно интересно, если это множество оказывается в каком-то смысле «значительным». В этом отношении современные работы часто принципиально не могут претендовать на ту полноту описания качественной картины, которая во многих случаях достигнута в теории потоков на плоскости или при исследовании систем с равномерно гиперболическим поведением траекторий. Естественно, при описании изменений качественной картины теперь нередко речь может идти только об изменении каких-то существенных её особенностей. В этом отношении результаты исследования сложных бифуркаций не имеют той полноты, какая была свойственна теории во времена Андронова. Но это, по-видимому, связано с существом дела. Некоторый дефект современного состояния дел скорее состоит в отсутствии чётких общих формулировок, какие черты качественной картины привлекают наше внимание. Это указывается в конкретных задачах, но каждый раз по-своему. Похоже, что для обобщающих формулировок пока не пришло время.

Информация о состоянии теории нелокальных бифуркаций в середине 80-х гг. имеется в [52]. В докладе Йоккоза [1] приведено несколько более новых результатов и даны литературные ссылки. Особо обращаю внимание на (цитированную там) книгу Ж. Палиса и Ф. Такенса [66]. Дополнительно укажу несколько новых работ группы Л. П. Шильникова, потому что работы этой группы (в том числе и прежние) пока недостаточно известны[75].

В названии статьи [67] фигурирует бифуркационное явление, при котором некоторая периодическая траектория неограниченно удлиняется и «в пределе» исчезает, причём при этом от неё не остаётся видимого следа вроде петли сепаратриссы или чего-нибудь в этом роде (она «бесследно исчезает в голубом небе», откуда и название «катастрофа голубого неба», данное (вначале, видимо, в шутку) таким бифуркациям). Для двумерных потоков такое явление (без указанного названия) было открыто Ф. Фуллером и подробнее изучено (уже под придуманным к тому временем названием) В. С. Медведе-

[75] В частности, в [66] имеется ссылка только на одну раннюю работу Шильникова, написанную до его основных «бифуркационных» работ (но уже выходящую за рамки равномерной гиперболичности).

вым. В [67] же эта бифуркация рассмотрена в трёхмерном случае. Она имеет коразмерность 1, т. е. является «типичной», и происходит при пересечении некоторой гиперповерхности в пространстве всех систем. Одновременно описывается и другая бифуркация коразмерности 1, быть может более интересная.

Представляет очевидный интерес вопрос: каким образом в гладкой системе дифференциальных уравнений может появиться сложно устроенный аттрактор[76)]? Этот вопрос связан со следующим наблюдением: до сих пор ни в одной прикладной задаче не обнаружены аттракторы типа базисных множеств потоков с аксиомой А (не сводящиеся к периодическим траекториям). Естественно поэтому попытаться построить такие аттракторы бифуркационными методами. В данной работе в этом направлении решается следующая задача: как через простую бифуркацию гладкого векторного поля можно из простого векторного поля (Морса—Смейла) получить поле со странным аттрактором[77)]? В работе обнаружена и изучена бифуркация коразмерности 1, приводящая к появлению гиперболического странного аттрактора, носитель которого является соленоидом Смейла—Вильямса. Интересно, что при приближении к бифуркационной гиперповерхности соленоид не претерпевает бифуркаций, а длина любой замкнутой траектории в нём стремится к бесконечности.

Другой пример появления странного аттрактора при бифуркации рассмотрен в [68]. Этот аттрактор, существующий в некотором интервале значений параметра возмущения, при всех его значениях содержит некоторое дикое

---

[76)] Это слово происходит от *attract* — притягивать, привлекать, и употребляется как название множеств, которые как бы притягивают к себе близкие траектории. В литературе встречаются различные формализации этого свойства «притягивать траектории». В настоящей статье по большей части под аттрактором понимается компактное инвариантное множество $A$, которое устойчиво по Ляпунову (т. е. для любой его окрестности $V$ имеется такая окрестность $W$, что ни одна положительная полутраектория, начинающаяся в $W$, никогда не выйдет из $V$) и таково, что все положительные полутраектории, начинающиеся в некоторой окрестности $A$, с возрастанием времени неограниченно приближаются к $A$. (Это почти дословная перефразировка принадлежащего А. М. Ляпунову определения асимптотической устойчивости положения равновесия.)

[77)] Насколько я могу судить, «странные» и «хаотические» аттракторы — это не точные термины, а несколько неопределённые названия (ср. с «фракталами» Мандельброта), к тому же несущие в себе оттенок эмоционального отношения, связанного с первоначальным удивлением. Под «странными аттракторами» обычно понимаются аттракторы (т. е. множества, как бы притягивающие траектории), устроенные в каком-то смысле «сложно» и «странно»; они не являются многообразиями или чем-нибудь в этом роде (скажем, не состоят из нескольких «кусков» многообразий). «Хаотический» аттрактор характеризуется «квазислучайным» поведением его гиперболическое множество, что гарантирует изобилие соответствующих бифуркационных явлений[78].

В [56], [69] показано, что в области Ньюхауса всюду плотны ДС со сколь угодно высокой кратностью седлоузловых периодических траекторий. Опираясь на эти работы, В. Ю. Калошин [70] показал, что в области Ньюхауса имеются ДС со сколь угодно высокой скоростью роста при $T \to \infty$ числа $N_{\mathrm{per}}(T)$ периодических траекторий с периодом $\leqslant T$. Более того, для любой функции $f(T)$ такие ДС, для которых $N_{\mathrm{per}}(T) \geqslant f(T)$ при достаточно больших $T$, образуют в этой области с $C^n$-топологией (с любым натуральным $n$) множество второй категории, так что нельзя сказать, что речь идёт о каком-то исключительном явлении. Это даёт ответ на вопрос, возникший около 30 лет назад. Правда, для аналитических систем вопрос о возможности сверхэкспоненциального по $T$ роста числа периодических траекторий с периодом $\leqslant T$ остаётся открытым.

В [71], [72] показано, что в некоторых случаях при наличии негрубого гетероклинического контура возможны бифуркационные явления, похожие на связанные с нетрансверсальными гомоклиническими траекториями.

В недавнем обзорном (и отчасти полупопулярном) докладе Шильникова [73] (который предполагается опубликовать также и по-русски) можно найти дополнительную информацию (с литературными ссылками) по бифуркациям, связанным с нетрансверсальными гомоклиническими траекториями.

Ещё одним значительным достижением, связанным с полулокальными бифуркациями, является исследование бифуркаций «аттрактора Лоренца». Последний, в соответствии с его названием, был открыт Э. Лоренцом в результате численного эксперимента с некоторой конкретной системой 3-го порядка, внимание к которой мотивировалось гидродинамическими соображениями. Открытие Лоренца привлекло внимание только примерно 10 лет спустя, причём прикладников (в широком смысле слова) и математиков-те-

---

траекторий. Ещё в начале 70-х гг. В. М. Алексеев предложил формализовать «квазислучайность» как положительность топологической энтропии. (Но имеются и другие (неэквивалентные) формализации (см. [66]), позволяющие рассматривать и некоторые аттракторы с нулевой энтропией как всё-таки хаотические.) Понимаемые в таком смысле, «хаотичность» и «странность» — не синонимы: аттрактор может быть многообразием и в то же время иметь положительную энтропию; он может иметь нулевую энтропию и в то же время не быть многообразием и вообще быть устроенным более или менее сложно. Другое дело, что большинство встречающихся «странных» аттракторов являются и «хаотическими» (что связано с присущей им некоторой гиперболичностью, пусть и более слабой, чем у «настоящих» гиперболических множеств).

[78] В [66] имеются ссылки на более ранние работы западных авторов, тоже отмечавших возникновение странных аттракторов при некоторых бифуркациях. Обычно это были не равномерно гиперболические аттракторы, а аттракторы типа аттрактора Лоренца, о котором говорится ниже.

оретиков оно заинтересовало по различным причинам. Прикладников оно убедило в реальном существовании странных (или хаотических) аттракторов — открытые математиками гиперболические странные аттракторы большинству из них не были известны, поскольку примеров таковых в задачах естественнонаучного происхождения не встречалось. Для математиков же аттрактор Лоренца представляет интерес не как ещё одна демонстрация уже известных возможностей поведения траекторий, а, наоборот, как объект, хотя и близкий по ряду своих свойств (как раз тех, которые произвели особое впечатление на прикладников) к гиперболическим аттракторам, но в то же время отличный от них по другим свойствам (для математиков именно эти тонкости и интересны). Математическая интерпретация результатов Лоренца была начата Д. Рюэллем и Ф. Такенсом в статье под названием «Странный, странный аттрактор». Повторение слова «странный» (помимо того, что оно навеяно фильмом С. Крамера) связано с тем, что: а) данный аттрактор является странным в том смысле, что он имеет динамически сложную структуру траекторий (счётное множество седловых периодических траекторий, континуум устойчивых по Пуассону, и т. д.) и в то же время сохраняется при малых возмущениях (всё это похоже на уже известные в то время равномерно гиперболические аттракторы, которые, однако, многими тоже воспринимались как нечто странное); б) он не являются равномерно гиперболическим (хотя определённая гиперболичность в нём наблюдается); в) в отличие от равномерно гиперболических множеств, его внутренняя структура не остаётся неизменной при малых возмущениях; точнее, она непрерывно изменяется в любом однопараметрическом семействе систем общего положения с таким аттрактором. Значит, если любое качественное изменение считать бифуркаций, то они происходят при всех значениях параметра семейства, причём это явление является «неустранимым» в том смысле, что оно присуще и всем достаточно близким семействам ДС. Первое само по себе не ново, поскольку то же самое имеет место и в давно известном примере семейства потоков на торе, имеющего в координатах $x, y \bmod 1$ вид $\dot{x} = 1$, $\dot{y} = \lambda$. Но в этом случае сколь угодно малым изменением семейства можно обеспечить, чтобы бифуркации происходили только тогда, когда $\lambda$ принадлежит к дополнению к некоторому открытому всюду плотному множеству (стало быть, с топологической точки зрения бифуркационные значения параметра являются «исключительными»[79]). Второе к моменту

[79] Здесь возникает отмеченная по несколько другому поводу А. Н. Колмогоровым коллизия между топологической и метрической точками зрения: при достаточной гладкости рассматриваемых семейств множество бифуркационных значений параметра всегда имеет положительную меру. Подобные вопросы входят в компетенцию теории КАМ, а применительно к данному примеру — также более специальной теории М. Эрмана и Ж.-К. Йоккоза, на доклад которого я сослался в самом начале.

появления статьи Рюэлля и Такенса тоже не было совершенно новым, так как аналогичное явление в другом примере было раньше обнаружено С. Смейлом и Р. Абрагамом[80], но здесь это явление было обнаружено в ином примере[81].

В связи со сказанным, применительно к аттрактору Лоренца под «бифуркациями» понимают не просто изменения его внутренней структуры, а изменения, в каком-то смысле «существенные». Таковы заведомо те бифуркации, при которых аттрактор Лоренца возникает или исчезает, но также и некоторые другие. Конкретное уточнение, какие качественные изменения являются существенными, связано с конкретным описанием самого аттрактора Лоренца и в этом смысле специфически привязаны к определённой ситуации.

Некоторая модель аттрактора Лоренца была предложена Р. Вильямсом. Приняв её, можно изучать внутреннее строение данного аттрактора (включая поведение лежащих в нём траекторий); можно также доказать, что действительно существуют системы третьего порядка с таким аттрактором и что он выживает при малых возмущениях. Однако при этом накладываются такие условия на рассматриваемую систему, которые хотя и могут выполняться для каких-то систем, но не выполняются в исходной системе Лоренца, так что, строго говоря, остаётся неясным, действительно ли Лоренц имел дело с аттрактором, названным его именем.

Эти вопросы были сняты в работе [74]. В ней указаны более широкие условия, гарантирующие существование аттрактора Лоренца. Это условия геометрического характера на некоторое отображение последования. Проверка этих условий для самой системы Лоренца до сих пор отчасти опирается на численный эксперимент, но таковый производился неоднократно, в различных группах и с различными программами, так что результаты не вызывают сомнений[82]. Внутренняя структура рассмотренного в [74] аттрактора (если не говорить об изменениях при некоторых бифуркациях) — такая, как у Вильямса, но свойства системы возле аттрактора — другие («притяжение»

---

[80] Ср. с «принцессой на горошине», ещё ранее описанной Г.-Х. Андерсеном. Она всю ночь пыталась путём приведения себя в общее положение сделать беспокоившие её ощущения исключительными, но это ей не удалось.

[81] Из сказанного об областях Ньюхауса ясно, что там это явление тоже наблюдается. Но, как говорилось, для них в середине 70-х гг. имелись скорее догадки, чем строгие результаты.

[82] Поясняю ещё раз: принципиальный факт существования ДС с аттрактором Лоренца был ясен уже после первых публикаций Дж. Гукенхеймера и Р. Вильямса; несложные конкретные системы дифференциальных уравнений, в которых при некоторой бифуркации рождается аттрактор Лоренца, указаны в работах, цитированных в [66]; ни то, ни другое не связано с численными экспериментами. На них отчасти опирается «только» проверка наличия аттрактора Лоренца в рассматривавшейся Лоренцом системе.

траекторий к аттрактору может быть более слабым). В последующих работах других авторов (посвящённых, главным образом, эргодическим свойствам аттрактора Лоренца — другие вопросы в основном закрыты[83)]) последний понимается именно в смысле [74]. Кроме того, в данной работе более подробно изучена его внутренняя структура (в частности, дано его символическое описание, приспособленное также и к задачам бифуркации).

Для нас здесь важно, что в [74] для их модели не только подтверждены утверждения о сохранении аттрактора Лоренца при малых возмущениях и постоянно происходящем при этом изменении его структуры, но и описан сценарий появления такого аттрактора, т. е. последовательность бифуркаций, приводящая к его рождению, а также рассмотрены бифуркации, происходящие после его рождения, и в том числе указан сценарий его разрушения.

## §2. «Именные» проблемы

**2.1. Грубые системы.** Уже из простейших примеров видно, что иногда качественные свойства ДС могут изменяться при сколь угодно малых возмущениях, а иногда — нет. Тривиальный пример — если в ДС никаких движений на самом деле не происходит, т. е. если речь идёт о тождественном диффеоморфизме или о потоке с нулевым векторным полем фазовой скорости. Тогда, конечно, при сколь угодно малом возмущении могут получиться различные качественные картины. Менее вырожденный пример — когда имеется неподвижная точка диффеоморфизма или положение равновесия потока (нуль соответствующего векторного поля), которые асимптотически устойчивы, но не экспоненциально устойчивы. Тогда при сколь угодно малом (в смысле $C^1$) возмущении в одних случаях неподвижная точка (положение равновесия) может вообще исчезнуть, а в других случаях исчезновения не происходит, но устойчивость может нарушиться. Если же неподвижная точка экспоненциально устойчива, то при $C^1$-малом возмущении она не исчезает и остаётся асимптотически устойчивой; то же относится и к экспоненциально устойчивому положению равновесия. В данном случае речь идёт о сохранении локальной качественной картины (возле неподвижной точки или положения равновесия). Те случаи, когда качественная картина во всём фазовом пространстве выдерживает малые возмущения, заслуживает особого внимания. В связи с этим по идее А. А. Андронова и Л. С. Понтрягина (1937 г.) вводят следующее определение.

Диффеоморфизм $f$ гладкого замкнутого многообразия $M$ называется грубым, если для любого достаточно $C^1$-близкого к нему диффеоморфизма $g$ существует гомеоморфизм $\chi\colon M \to M$, сопрягающий $f$ и $g$ в том смысле, что

$$\chi \circ f = g \circ \chi. \tag{22}$$

---

[83)] Информацию см. в [75].

Поток $\{\varphi_t\}$ на гладком замкнутом многообразии $M$, задаваемый гладким векторным полем фазовой скорости $\mathbf{v}$ (так что $\frac{d}{dt}\varphi_t(x) = \mathbf{v}(\varphi_t(x))$), называется грубым, если любое достаточно $C^1$-близкое к $\mathbf{v}$ поле $\mathbf{w}$ определяет поток $\{\psi_t\}$, эквивалентный потоку $\{\varphi_t\}$ в том смысле, что существует гомеоморфизм $\chi\colon M \to M$, переводящий траектории первого потока в траектории второго потока с сохранением направления движения по ним. Стоит отметить некоторые особенности этого определения (более подробное обсуждение имеется в [76]). В случае непрерывного времени не требуется, чтобы гомеоморфизм $\chi$ сопрягал невозмущённый и возмущённый потоки в том смысле, что

$$\chi \circ \varphi_t = \psi_t \circ \chi \quad \text{при всех } t \tag{23}$$

(что казалось бы естественным аналогом (22)). Дело в том, что если у потока имеется замкнутая траектория, то при возмущении её период может измениться, а тогда (23) невозможно; между тем изменение периода мы не считаем изменением качественной картины. Не требуется, чтобы $\chi$ было диффеоморфизмом, потому что если у диффеоморфизма $f$ имеется неподвижная точка (которая у грубого диффеоморфизма должна сохраняться при малом возмущении), то при возмущении собственные значения соответствующей матрицы линейного приближения могут измениться, а если бы $\chi$ в (22) было диффеоморфизмом, этого не могло бы произойти. Для потоков аналогичные соображения оформляются несколько сложнее (потому что мы требуем эквивалентности, а не выполнения (23)), но вывод по-прежнему состоит в том, что, вообще говоря, $\chi$ не может быть диффеоморфизмом. В определении Андронова и Понтрягина дополнительно требовалось, чтобы при достаточной близости $g$ к $f$ или $\mathbf{w}$ к $\mathbf{v}$ гомеоморфизм $\chi$ был $C^0$-близким к тождественному. Позднее М. М. Пейксото предложил этого не делать, так что имеются два логически различных варианта грубости — по Андронову—Понтрягину и по Пейксото. Первое формально более ограничительно, чем второе, но теперь известно, что на самом деле эти варианты эквивалентны. Поэтому я позволю себе их не различать, хотя на самом деле их эквивалентность — весьма нетривиальный факт (о чём в своём месте будет сказано отдельно). Вместо грубости часто (особенно за рубежом) говорят о «структурной устойчивости».

После того как в классической работе А. А. Андронова и Л. С. Понтрягина было введено понятие грубой системы и охарактеризованы грубые потоки на плоскости (точнее, на двумерной сфере), естественно возник вопрос о качественной характеризации поведения траекторий грубых систем в других случаях. В определении говорится о том, что происходит при возмущении; в характеризации, о которой идёт речь, говорится только о поведении траекторий невозмущённой системы. М. М. Пейксото перенёс теорему Андронова—Понтрягина на потоки на замкнутых поверхностях; формулировка при этом почти не изменилась. Во всех этих случаях грубые потоки образуют открытое всюду плотное множество в пространстве всех потоков с $C^1$-топологией. Естественно, встал вопрос, какие системы являются грубыми в многомерных случаях (а для систем с дискретным временем — уже и в двумерном[84)]). Наиболее непосредственное обобщение условий грубости двумерных потоков приводит к так называемым системам Морса—Смейла; кроме того, по аналогии с этим случаем можно было бы думать, что грубые системы образуют открытое всюду плотное множество в пространстве всех ДС на рассматриваемом многообразии с $C^1$-топологией. Но всё это оказалось неверным, за исключением предположения о грубости систем Морса—Смейла. Не лишённая драматизма история исследований в этой области неоднократно описывалась, в том числе и мной (см. ниже), поэтому я сразу перейду к ответу. В ходе «гиперболической революции» С. Смейл высказал предположение, что для грубости необходимо и достаточно, чтобы множество неблуждающих точек было гиперболическим[85)] (это главное), чтобы периодические точки были в нём плотны (совокупность этих двух условий называется «аксиомой А» Смейла) и чтобы соответствующие устойчивые и неустойчивые многообразия имели только трансверсальные пересечения («сильное условие трансверсальности»). Достаточность была доказана (в полной общности — Р. К. Робинсоном) в конце предыдущего 20-летия, необходимость же доказана только теперь, хотя очень важный шаг был уже давно сделан Ч. Пью, доказавшим на первый взгляд простую, а на самом деле трудную лемму о замыкании. Она утверждает, что если данная гладкая система имеет неблуждающую точку $x$, то сколь угодно малым в $C^1$-смысле возмущением можно обеспечить, чтобы $x$ стала периодической. Несколько упрощённое и уточнённое по сравнению с первоначальным доказательство имеется в [79], где также показано, что аналогичная лемма верна и в классе гамильтоновых систем. (В последнем классе, как и в классе ДС, сохраняющих объём, $C^k$-аналог этой леммы при достаточно большом $k$ неверен! Ссылки на соответствующие работы М. Эрмана и Дж. Ксиа см. в [1].). Но и после того, как эта лемма была доказана, прошло немало времени, пока не

---

[84)] Грубые диффеоморфизмы окружности были описаны А. А. Майером вскоре после появления работы Андронова—Понтрягина. Они тоже образуют открытое всюду плотное множество в пространстве всех $C^1$-диффеоморфизмов.

[85)] Определение систем Морса—Смейла и гиперболических множеств приводится в учебниках по гладким динамическим системам, обзорах [77] и [78], а также в соответствующих статьях в «Математической энциклопедии». Стоит обратить внимание, что применительно к положениям равновесия терминология, установившаяся в теории гладких динамических систем с 60-х гг., отличается от прежней: теперь экспоненциально устойчивые (неустойчивые) фокусы и узлы тоже относят к числу гиперболических положений равновесия, а раньше гиперболическими назывались только сёдла.

удалось доказать необходимость условий грубости, гипотетически указанных Смейлом. Для динамических систем с дискретным временем необходимость была доказана Р. Мане, для потоков — Ш. Хаяши. (В [76] упоминаются некоторые «промежуточные» работы, отчасти тоже сыгравшие роль.) Стоит заметить, что Мане и Хаяши пришлось добавить к лемме о замыкании ещё некоторые утверждения сходного характера (столь же «очевидные»).

Упоминавшаяся выше эквивалентность грубости по Андронову—Понтрягину и по Пейксото следует из того, что условия, необходимые для второй, являются достаточными для первой. Более простого доказательства эквивалентности не известно, хотя, казалось бы, этот факт должен быть более элементарным.

Несколько более слабым, чем грубость, свойством является так называемая $\Omega$-грубость, состоящая, грубо говоря, в сохранении при малых возмущениях множества неблуждающих точек вместе с динамикой на нём. Точное определение см. в любом учебнике по гиперболической динамике или в обзоре о гиперболических множествах [77]. Название связано с тем, что множество неблуждающих точек часто обозначают через $\Omega$. Здесь мы тоже примем это обозначение. Необходимое и достаточное условие $\Omega$-грубости было высказано в виде гипотезы С. Смейлом и Дж. Палисом. Главным в нём является гиперболичность множества неблуждающих точек; сверх того нужны ещё плотность в нём множества периодических точек и так называемая ацикличность этого множества. Имеются эквивалентные условия, в которых требуется либо гиперболичность и ацикличность некоторого другого множества, которое *a priori* могло бы быть меньше чем $\Omega$, либо одна гиперболичность некоторого множества (множества цепно рекуррентных точек), которое *a priori* могло бы быть больше $\Omega$; на самом деле при выполнении упомянутых условий все эти множества совпадают с $\Omega$. Эквивалентность этих условий и их достаточность доказали С. Смейл и его сотрудники; по теперешним меркам это сравнительно несложно (достаточность доказывается легче, чем достаточность соответствующего условия для грубости). Доказательство необходимости оказалось столь же трудным, как и для грубости; для систем с дискретным временем его дал Дж. Палис сразу же после работы Мане, а для потоков — Ш. Хаяши.

Обзор [76] отражает ситуацию вплоть до начала 80-х гг. (в историческом отношении его дополняют воспоминания М. Пейксото [80]). Достижения следующих лет: [81], [82], [83], [84].

В определении грубой системы говорится о возмущениях, малых в смысле $C^1$. Если рассматривать диффеоморфизмы $f, g, \ldots$ (векторные поля $\mathbf{v}, \mathbf{w}, \ldots$) класса $C^k$ и если в определении грубости исходной ДС (определяемой $f$ или $\mathbf{v}$) «близость» $g$ к $f$ ($\mathbf{w}$ к $\mathbf{v}$) понимать как $C^k$-близость, то получится определение свойства, которое естественно назвать $C^k$-грубостью. (От гомеоморфизма $\chi$ по-прежнему гладкости не требуется. При этом по-прежнему можно требовать или не требовать близости $\chi$ к тождественному преобразованию, так что получаются два варианта $C^k$-грубости — $C^k$-грубость по Андронову—Понтрягину и $C^k$-грубость по Пейксото.) В новых терминах прежняя грубость — это $C^1$-грубость. Что можно сказать о $C^k$-грубости при $k > 1$?

Пока что ничто не противоречит предположению, что $C^k$-грубость эквивалентна $C^1$-грубости (не считая, конечно, того, что о $C^k$-грубости можно говорить только применительно к ДС класса $C^k$). Однако положительных результатов на сей счёт всего два: это так а) в размерности 1 (и для потоков, и для диффеоморфизмов); б) для потоков на ориентируемых двумерных замкнутых многообразиях и на трёх простейших неориентируемых двумерных замкнутых многообразиях — тех, у которых эйлерова характеристика равна 1 (проективная плоскость), 0 (тор) или $-1$. Во всех этих случаях $C^k$-грубые системы образуют открытое всюду плотное множество в пространстве всех ДС класса $C^k$ на данном многообразии. В остальных случаях попытка найти необходимое условие $C^k$-грубости на том пути, который привёл к успеху при $k = 1$, упирается, прежде всего, в вопрос о справедливости $C^k$-варианта леммы о замыкании. В свете сказанного выше, перспективы здесь вызывают сомнения. Можно ещё добавить, что если несколько усилить утверждения леммы о замыкании (казалось бы, совершенно естественным способом), то уже её $C^2$-аналог не будет справедлив. См. цитированную в [1] статью К. Гутьерреса, а также [85] (усиление леммы, которое, как показано в [85], при $k = 2$ неверно, близко к утверждениям, доказанным и использовавшимся при $k = 1$ Мане и Хаяши).

Можно определить понятие грубости для систем с некомпактным фазовым многообразием и для потоков на компактных многообразиях с краем. Ситуация в этих случаях далеко не полностью выяснена. Известно, что она отчасти отличается от описанной выше.

По темам последних трёх абзацев я не знаю более новых работ, чем указанные в [76].

Наконец, надо остановиться на грубости гладких отображений отрезка $[0, 1]$ в себя, которые не являются взаимно однозначными (для диффеоморфизмов отрезка вопрос тривиален). Определение грубости дословно переносится на этот случай (с заменой слова «диффеоморфизм» на «гладкое отображение»). Сразу же бросается в глаза новое обстоятельство: при наличии критических точек (точек $x$, где $f'(x) = 0$) отображение $f$ не может быть $C^1$-грубым. Действительно, путём возмущения, сколь угодно малого в смысле $C^1$, можно столь заметно изменить характер критической точки, что это «почувствуется» даже при нашем довольно грубом подходе, когда всё рассматривается с точностью до топологического сопряжения. Например, можно обеспечить, чтобы возмущённое отображение переводило некоторый отрезок в одну точку; или, наоборот, чтобы каждая точка имела только конечное число прообразов. Поэтому при наличии критических точек в содержательной теории речь должна идти о $C^k$-грубости с $k > 1$. В свете сказанного выше ясно, что это создаёт немалые трудности.

Используя специфику одномерного случая и пользуясь выходом в комплексную область (где, к счастью, уже имелись результаты, которые удалось использовать), О. С. Козловский смог справиться с этими трудностями в простейшем нетривиальном случае так называемых унимодальных отображений, т. е. гладких отображений отрезка в себя, имеющих ровно одну критическую точку [86]. Результат состоит в том, что при любом $k > 1$ унимодальное $C^k$-гладкое отображение отрезка в себя является $C^k$-грубым в том и только том случае, когда оно удовлетворяет аксиоме A (подходящим образом переформулированной применительно к одномерным отображениям с критическими точками) и его критическая точка является невырожденной (вторая производная в ней отлична от нуля). Козловский доказал также, что унимодальные отображения, удовлетворяющие аксиоме A (к которой уже тривиальным образом можно добавить условие невырожденности критической точки), всюду плотны в пространстве всех унимодальных отображений класса $C^k$. Собственно, в этом и состоит его основной результат, из которого уже сравнительно легко получается необходимость приведённого выше условия грубости, тогда как его достаточность была известна раньше.

**2.2. 21-я проблема Гильберта.** В этом пункте мы будем иметь дело с системой линейных обыкновенных дифференциальных уравнений в комплексной области

$$\frac{dy}{dx} = C(x)y, \qquad y = (y^1, \ldots, y^p) \in \mathbb{C}^p. \tag{24}$$

За одним исключением, которое будет оговорено, она подразумевается голоморфной во всей расширенной плоскости комплексного переменного (сфере Римана) $\overline{\mathbb{C}}$, кроме нескольких особых точек $a_1, \ldots, a_n$. Голоморфность системы в точке $x \in \mathbb{C}$ означает просто голоморфность в этой точке коэффициентов матрицы $C(x)$. Если среди $a_i$ нет $\infty$, то система (24) должна быть голоморфной в точке $\infty$; это означает следующее: перепишем (24) в терминах новой независимой переменной $\zeta := 1/z$; коэффициенты полученной системы

$$\frac{dy}{d\zeta} = -\frac{1}{\zeta^2}\, C\Big(\frac{1}{\zeta}\Big) y \tag{25}$$

должны иметь в точке $\zeta = 0$ устранимую особенность. Решения системы (24), вообще говоря, ветвятся (простейший пример:

$$p = 1, \quad n = 2, \quad S := \overline{\mathbb{C}} \setminus \{0, \infty\}, \quad \frac{dy}{dx} = \frac{\alpha}{x}\, y; \tag{26}$$

решения суть $y = Cx^\alpha$). Но их можно рассматривать как функции на универсальной накрывающей поверхности $\widetilde{S}$ области $S$, и там они являются однозначными голоморфными функциями. (У линейной системы любое решение действительно продолжается на всю поверхность $\widetilde{S}$, что доказывается примерно так же, как и вероятно известное читателю утверждение, что в вещественной области решения линейной системы продолжается на весь тот интервал, полупрямую или всю прямую $\mathbb{R}$, где определены и непрерывны коэффициенты этой системы. Для нелинейных систем ни комплексный, ни вещественный варианты утверждения о продолжении, вообще говоря, не верны.) Точки $\widetilde{S}$ я буду обозначать с добавлением тильды, при этом $\widetilde{x}$ лежит над точкой $x \in S$; поэтому для решений (24) лучше писать $y(\widetilde{x})$. Вполне может случиться, что для какого-то решения какой-то конкретной системы (24) однозначность достигается уже при его подъёме на какую-то ме́ньшую, чем $\widetilde{S}$, накрывающую поверхность области $S$ (иногда решение однозначно уже в самой $S$). Но ничто не мешает всё равно поднять его на $\widetilde{S}$ — «каши маслом не испортишь».

Проекцию $\widetilde{S} \to S$, переводящую точку $\widetilde{x} \in \widetilde{S}$ в накрываемую ею точку $x \in S$, обозначим через $\pi$. Гомеоморфизмы $\sigma\colon \widetilde{S} \to \widetilde{S}$, для которых $\pi \circ \sigma = \sigma \circ \pi$ (т. е. которые переставляют между собой точки $\widetilde{x}$, лежащие над одной и той же $x$), называют скольжениями. Группа всех скольжений обозначается через $\Delta$. Она изоморфна более конкретному объекту — фундаментальной группе $\pi_1(S, x_0)$, которая в данном случае является свободной группой с $n-1$ образующими, отвечающими обходам вокруг каких-нибудь $n-1$ из точек $a_i$. Однако группы $\pi_1(S, x_0)$, в отличие от $\Delta$, отчасти зависят от $x_0$ (при различных $x_0$ они изоморфны, но «стандартного» изоморфизма между ними нет, как нет и «стандартного» изоморфизма $\pi_1(S, x_0) \to \Delta$).

Сумма двух решений линейной системы — снова решение, произведение решения на постоянный скаляр — решение, поэтому решения (24) образуют векторное пространство $\mathfrak{Y}$, в данном случае комплексное. Из того, что решение однозначно определяется своим значением при каком-нибудь $\widetilde{x}_0$, легко следует, что $\mathfrak{Y}$ $p$-мерно. Базис в $\mathfrak{Y}$ — это то, что в теории дифференциальных уравнений называется фундаментальной системой решений. Наряду с векторно-матричной системой (24) рассматривается матричная система

$$\frac{dY}{dx} = C(x)Y, \tag{27}$$

где $Y$ — квадратная матрица $p$-го порядка. Её столбцы суть решения (24), и их линейная независимость означает невырожденность матрицы $Y(\widetilde{x})$ (при любом $\widetilde{x} = \widetilde{x}_0$, для чего достаточно доказать её невырожденность при каком-нибудь $\widetilde{x}_0$); в этом случае матрицу $Y(\widetilde{x})$ называют фундаментальной матрицей системы (24) (так что фундаментальная матрица — это матрица, столбцы которой образуют фундаментальную систему решений). Общее решение (24) имеет вид $y = Y(\widetilde{x})c$ с постоянным $c \in \mathbb{C}^p$. При использовании в $\mathfrak{Y}$ базиса, образованного столбцами $Y(\widetilde{x})$, координаты этого $y$ даются вектор-столбцом $c$.

Всё это аналогично определениям и утверждениям из теории линейных систем в вещественной области. А вот и нечто новое: если $y(\widetilde{x})$ — решение (24) или $Y(\widetilde{x})$ — решение (27) и $\sigma \in \Delta$, то $\widetilde{x} \mapsto y(\sigma\widetilde{x})$ или $\widetilde{x} \mapsto Y(\sigma\widetilde{x})$ — тоже решение (24) или (27). Действительно, у любой точки $\widetilde{x}_0 \in \widetilde{S}$ имеется в $\widetilde{S}$ окрестность $\widetilde{U}$, которую $\pi$ гомеоморфно отображает на окрестность $U := \pi(\widetilde{U})$ точки $x_0 = \pi\widetilde{x}_0$ в $S$; тогда $\sigma\widetilde{U}$ — окрестность $\sigma\widetilde{x}_0$ в $\widetilde{S}$, которую $\pi$ тоже гомеоморфно отображает на $U$; ясно, что

$$\sigma(\pi|_{\widetilde{U}})^{-1} = (\pi|_{\sigma\widetilde{U}})^{-1}. \tag{28}$$

Утверждение, что $y(\widetilde{x})$ является решением (24) в $\widetilde{U}$, означает, что $y_1(x) := y((\pi|_{\widetilde{U}})^{-1}x)$ — решение (24) в $U$. Утверждение, что $\widetilde{x} \mapsto y(\sigma\widetilde{x})$ является решением (24) в $\widetilde{U}$, означает, что $y_2(x) := y(\sigma(\pi|_{\widetilde{U}})^{-1}x)$ — решение (24). Но ввиду (28) имеем $y_2(x) = y((\pi|_{\sigma\widetilde{U}})^{-1}x)$, а это действительно решение (24), ибо $y(\widetilde{x})$ является решением в $\sigma\widetilde{U}$ (как и на всей $\widetilde{S}$). Итак, $y(\sigma\widetilde{x})$ действительно является решением (24) возле любой точки $\widetilde{x}_0 \in \widetilde{S}$. Пользуясь изоморфизмом $\Delta \approx \pi_1(S, x_0)$, переход от $y(\widetilde{x})$ к $y(\sigma\widetilde{x})$ можно описать как изменение решения (24) при его аналитическом продолжении по цепочке кругов, взятых вдоль замкнутого пути, отвечающего $\sigma$. Но это подразумевает несколько отождествлений различных объектов — $\Delta$ и $\pi_1(S, x_0)$, голоморфных функций на $\widetilde{S}$ и наборов их элементов по Вейерштрассу.

Таким путём возникает преобразование

$$\sigma_*\colon\ \mathfrak{Y} \to \mathfrak{Y}, \qquad (\sigma_* y)(\widetilde{x}) = y(\sigma^{-1}\widetilde{x}),$$

очевидно, линейное и невырожденное. При этом $(\sigma_*\tau_*) = \sigma_*\tau_*$ (если бы мы определили $(\sigma_* y)(\widetilde{x})$ как $y(\sigma\widetilde{x})$, то получилось бы $(\sigma_*\tau_*) = \tau_*\sigma_*$; то же самое отмечалось в п. 1.3, *а*). Обозначая через $\mathrm{GL}(\mathfrak{Y})$ группу невырожденных линейных преобразований пространства $\mathfrak{Y}$, получаем представление

$$\Delta \to \mathrm{GL}(\mathfrak{Y}), \qquad \sigma \mapsto \sigma_*, \tag{29}$$

которое называется представлением монодромии (для (24)). Взяв в $\mathfrak{Y}$ какой-нибудь базис, т. е. фундаментальную систему решений, объединённую в фундаментальную матрицу $Y(\widetilde{x})$, можем перейти к матричному представлению $\chi\colon \Delta \to \mathrm{GL}(p, \mathbb{C})$, для которого $\chi(\sigma)$ описывает изменение координат в этом базисе элемента $y \in \mathfrak{Y}$ при переходе к $\sigma_* y$. Как легко видеть, $Y(\widetilde{x}) = Y(\sigma\widetilde{x})\chi(\sigma)$, что является эквивалентным определением $\chi$. Представление $\chi$ тоже называют представлением монодромии. Оно определяется заданием $n-1$ невырожденной матрицы — образов образующих $\Delta$ (значит, эти матрицы описывают аналитическое продолжение решений при обходах вокруг $n-1$ особой точки), так что является более конкретным объектом, нежели (29). Но самой системой (24) $\chi$ определено не совсем единственным образом: при изменении базиса в $\mathfrak{Y}$ представление $\chi$ заменяется на сопряжённое представление $C\chi C^{-1}$, где $C$ — некоторая постоянная матрица.

В связи с этим монодромией называют также и весь класс сопряжённых представлений $\{C\chi C^{-1}; C \in \mathrm{GL}(p, \mathbb{C})\}$ (он уже определяется системой (24)).

Нас будут интересовать системы (24), имеющие особенности сравнительно «слабого» типа. Особая точка $a_i \neq \infty$ называется фуксовой, если $C(z)$ имеет в этой точке полюс первого порядка; особая точка $\infty$ называется фуксовой, если коэффициенты (25) имеют в точке $\zeta = 0$ полюс первого порядка. Система, у которой все особые точки фуксовы, называется фуксовой системой. Можно доказать, что у фуксовой системы, у которой особые точки суть $a_1, \ldots, a_n$ и, возможно, $\infty$, матрица $C(x)$ имеет вид

$$C(x) = \sum_{i=1}^{n} \frac{B_i}{x - a_i},$$

где $B_i$ — постоянные матрицы. При этом точка $\infty$ тоже будет особой в том и только том случае, когда $\sum B_i = 0$.

«Слабость» фуксовых особенностей проявляется и в поведении решений возле особых точек. Если $x$ стремится к фуксовой особой точке $a$, то $|y(x)|$ может расти или убывать не быстрее некоторой степени $|x - a|$. Сказанное нуждается в некотором уточнении, потому что даже для (26) при $\operatorname{Im}\alpha \neq 0$ можно обеспечить сколь угодно большой рост $|y|$ с уменьшением $|x|$, если приближаться к 0 по спирали, вдоль которой $|y|$ убывает медленно по сравнению с числом оборотов, сделанных вокруг 0 (тогда $|y|$ изменяется главным образом за счёт этих оборотов). Нужное уточнение очень просто: $|y|$ растёт или убывает не быстрее некоторой степени от $|x - a|$, когда $x$ стремится к $a$, оставаясь внутри фиксированного угла с вершиной в $a$.

Изолированная особая точка $a$ системы (24) называется регулярной, если для неё выполняется то же самое свойство не более чем степенно́го роста или убывания решений при приближении $x$ к $a$ (с тем же уточнением об угле). Оказывается, в регулярной особой точке коэффициенты правой части системы имеют полюс, но не обязательно первого порядка, так что существуют регулярные особые точки, не являющиеся фуксовыми. Но уже при полюсе второго порядка особая точка может не быть регулярной. Лишь немногим более 10 лет назад удалось указать алгоритм, позволяющий (если удастся реально проделать все вычисления) определять, является ли особая точка регулярной. Он оказался весьма громоздким. Было бы неосторожно утверждать, что его совсем нельзя упростить, но по-видимому среди всех систем, коэффициенты которых имеют полюс в $a$, регулярные системы расположены каким-то сложным образом, так что «распознающий» их алгоритм не может быть слишком простым.

Если все особые точки системы регулярные, то и вся система называется регулярной.

Теперь можно перейти к нашей теме — 21-й проблеме Гильберта (её называют также «проблемой Римана—Гильберта»). Она гласит: показать, что «всегда существует линейное дифференциальное уравнение (на самом деле имеется в виду не одно уравнение, а система. — *Д. А.*) фуксова типа с заданными особыми точками и заданной группой монодромии» (теперь говорят точнее о представлении монодромии). Гильберт вполне мог иметь в виду не только те системы, которые мы теперь называем фуксовыми, а (по теперешней точной терминологии) регулярные системы. Независимо от неясности с терминологией начала века, здесь реально имеются две проблемы — «фуксова» и «регулярная». Последняя была вскоре положительно решена Й. Племелем (между прочим, фактически это было первое удачное использование теории сингулярных интегральных уравнений, основы которой именно в данной связи и были заложены Племелем, хотя формально в тот момент Племель о них не говорил). Племель пытался также вывести отсюда положительную разрешимость фуксова варианта, но на самом деле его редукция проходит при некотором дополнительном условии на представление монодромии, т. е. фактически он получил некоторое достаточное условие положительной разрешимости фуксовой 21-й проблемы. Оно состоит в том, что хоть одна из матриц монодромии, отвечающих обходам вокруг особых точек $a_1, \ldots, a_n$, приводится к диагональному виду. Но это было осознано много позже. По словам Ю. С. Ильяшенко, он заметил пробел у Племеля в 1975 г., когда говорил о фуксовых системах в своих лекциях. В печати неполнота рассуждений Племеля была отмечена, по-видимому, только в 1985 г. в обзоре [87]. Но тогда ещё оставалась надежда, что для фуксовых систем ответ всё-таки всегда положителен. Неожиданностью явилось открытие А. А. Болибрухом противоречащего примера [88][86]. Продолжая работу в этом направлении, Болибрух, с одной стороны, построил серию контрпримеров различного характера, а с другой — нашёл новые достаточные условия различной степени общности, гарантирующие положительный ответ; в этой работе участвовали и другие авторы. Вот одно из новых достаточных условий, принадлежащее В. П. Костову и А. А. Болибруху: представление $\chi$ неприводимо.

Новые методы нашли применение и к некоторым другим задачам аналитической теории (задача Биркгофа о стандартной форме системы в окрестности нерегулярной особой точки). Следует отметить, что в работах Болибруха вместо интегральных уравнений обычно используются векторные расслоения; такой геометрический подход впервые появился в работе Х. Рорля, датированной в точности началом предыдущего 20-летия (и содержащей, в частности, другое доказательство теоремы Племеля); другим постоян-

[86] Это, несомненно, была одна из лучших работ, опубликованных в «Математических заметках» (если не лучшая из них). Но хотя «Заметки» вообще-то переводятся на английский язык, эта работа не была переведена — в то время «краткие научные сообщения» (где как раз и публикуются последние новинки) не переводились!

ным «ингредиентом» его работ является принадлежащее А. Левелю (1961) усовершенствование классической локальной теории, построенной ещё в прошлом веке (главным образом, Л. Фуксом и А. Пуанкаре). Изложение этого направления (хотя теперь уже не совсем полное) см. в [89].

Замечу, что можно сформулировать (и исследовать) нелинейный аналог задачи Римана—Гильберта [90].

**2.3. Гипотеза Дюлака.** Эта гипотеза состояла в том, что система двух уравнений в $\mathbb{R}^2$

$$\dot{x} = f(x, y), \quad \dot{y} = g(x, y), \tag{30}$$

в которой $f$ и $g$ суть многочлены (скажем, $n$-й степени), может иметь только конечное число предельных циклов. Ниже оно обозначается через $L(f, g)$. Сам А. Дюлак рассматривал утверждение «$L(f, g) < \infty$» не как гипотезу, а как теорему, которую ему удалось доказать (1923 г.). В действительности Дюлак правильно понял, что доказательство сводится к анализу отображения последования вдоль полицикла[87] и что сложность проблемы связана с неаналитичностью этого отображения; он установил некоторые свойства последнего, однако их недостаточно для требуемого заключения, что у него может быть только конечное число неподвижных точек. В 1977 г. Ф. Дюмортье высказал сомнения в полноте рассуждений Дюлака. Летом 1981 г. Р. Муссю сделал эти сомнения предметом широкого обсуждения, написав о них нескольким коллегам. Независимо, тем же летом, Ю. С. Ильяшенко обнаружил ошибку в мемуаре Дюлака (он занял более категорическую позицию, прямо говоря об ошибке в определённом месте), после чего «теорема Дюлака» была переименована в гипотезу. В печати это переименование совершилось, по-видимому, в препринте Ю. С. Ильяшенко, цитированном в уже упоминавшемся обзоре [87], а затем — в самом этом обзоре. Статья [91] подводит итоги этому этапу критического освоения мемуара Дюлака (к которому, конечно, добавилось много другого).

Теперь гипотеза Дюлака полностью доказана. Для $n = 2$ это сделал Р. Бамон, а в общем случае — Ю. С. Ильяшенко и Ж. Экаль [92], [93]. Методы Ильяшенко и Экаля различны; оба метода применялись и в других

[87] Полицикл — это замкнутая кривая $L$, состоящая из сепаратрис, соединяющих некоторые положения равновесия, и самих этих положений равновесия, причём направления движения (с ростом $t$) по сепаратрисам отвечают одному и тому же направлении обхода по $L$. При наличии у $L$ точек самопересечения (таковыми могут быть только положения равновесия) фактически дополнительно требуется, чтобы путём малой деформации ориентированной кривой $L$ могла получиться замкнутая кривая без самопересечений (иначе не приходится говорить о рождении предельного цикла из ориентированной кривой $L$).

Название «полицикл» получило распространение в последние годы, хотя, кажется, ещё не стало стандартным. Раньше часто говорили «сепаратрисный контур (многоугольник)», «сложный цикл» и т. п.

задачах. Теорию Ильяшенко называют «геометрической», тогда как подход Экаля связан с новым процессом суммирования расходящихся рядов, хорошо приспособленным для использования в аналитической теории обыкновенных дифференциальных уравнений. Краткое изложение подхода Экаля имеется в [94]. Предшественниками Экаля в отношении его процесса суммирования — но не в задаче Дюлака — были Ж.-П. Рамис и В. Бальзер. Книга последнего [95] является упрощённым изложением этого процесса (не совсем с позиций Экаля) и ряда его применений (к тем задачам, которыми интересовался сам Бальзер). Она заканчивается доказательством теоремы Б. Брааксмы, согласно которой формальные решения нелинейных мероморфных обыкновенных дифференциальных уравнений всегда суммируемы с помощью нового процесса, так что в этом отношении ничего лучшего желать не приходится!

В этом пункте уместно сказать о 16-й проблеме Гильберта. Её первая половина относится к вещественной алгебраической геометрии, а вторая — к теории дифференциальных уравнений. В обоих половинах указаны удачно поставленные крупные научные задачи, но, насколько можно судить, они не только формально относятся к различным разделам математики, но и по существу далеки друг от друга, о чём Гильберт, по-видимому, не вполне отдавал себе отчёт. Нас интересует вторая часть 16-й проблемы. Она состоит в том, чтобы получить оценку сверху числа предельных циклов системы (30), где $f, g$ — многочлены $n$-й степени, в виде явно указанной функции от $n$. Сама постановка этого вопроса подразумевает, что для любого $n$ число

$$H(n) := \max\bigl\{L(f,g);\ f, g \text{ — всевозможные многочлены } n\text{-й степени}\bigr\} < \infty. \qquad (31)$$

До сих пор неизвестно, так ли это.

В своё время И. Г. Петровский и Е. М. Ландис [96], [97], [98] полагали, что им удалось доказать (31) и получить некоторую оценку для $H(n)$; в частности, они утверждали, что $H(2) = 3$. Однако теперь известно, что система (30), в которой $f, g$ — многочлены второй степени, может иметь четыре предельных цикла (Ши Сонглин. Его пример изложен в [91], [99].)

Конечно, нередко бывало, что кто-нибудь намечал правильный путь исследования, но допускал ошибку при его реализации, — иногда это был просто досадный просмотр, иногда же из виду упускалось нечто существенное. Но в данном случае ситуация оказалась более сложной. В [96], [97] имеются ценные идеи, оказавшие влияние на дальнейшее развитие теории обыкновенных дифференциальных уравнений в комплексной области, но что касается собственно 16-й проблемы Гильберта, намеченный там подход встречает препятствия, не преодолённые до сих пор.

В [96], [97] использовался выход в комплексную область, где геометрическим образом, отвечающим системе (30), является не система кривых (траекторий), а двумерное слоение (слоение с двумерными слоями) с особенностями (последние суть те точки, где $f = g = 0$). Обсуждались некоторые свойства этого слоения, для чего была введена относящаяся к нему система понятий, в значительной степени перекликающаяся с общей теорией слоений, которая зародилась несколько раньше и примерно в то же время стала успешно развиваться. В других местах [96], [97] отмечались специфические особенности рассматриваемых слоений, отвечающих именно многочленам $f$, $g$ в (30), в частности, было начато обсуждение типичных свойств таких слоений. Само по себе всё это весьма содержательно. Но оказывается, что в комплексной области система (30) может иметь бесконечное число предельных циклов (Петровский и Ландис по-видимому допускали такую возможность, аккуратно же это выяснил Ю. С. Ильяшенко. Ссылку см. в [87]). Поэтому Петровский и Ландис попытались доказать, что из этих комплексных предельных циклов только сравнительно небольшое явно оцениваемое число может попасть в вещественную область. Систематическая проверка, организованная в конце 60-х гг. С. П. Новиковым на специально посвящённом этому семинаре (при моём активном участии в роли «адвоката дьявола») показала, что эта часть [96], [97] несостоятельна (независимо к тому же выводу пришёл Ю. С. Ильяшенко). А собственно для 16-й проблемы Гильберта эта часть является решающей[88].

Некоторый прогресс достигнут в исследовании двух локальных (по параметру) вариантов 16-й проблемы Гильберта. В первом, исследование которого было начато Ю. С. Ильяшенко, речь идёт об оценке (в зависимости от степени соответствующих многочленов) наибольшего числа предельных циклов, которые могут родиться при полиномиальном возмущении полиномиальной гамильтоновой системы, обычно — системы с полиномиальным гамильтонианом. Ситуация на середину 80-х гг. освещена в [87] (раздел об «ослабленной проблеме Гильберта»); после того работа продолжалась, но более новой сводки результатов, насколько известно, нет. Другой вариант был предложен В. И. Арнольдом. Его первоначальная формулировка оказалась чересчур оптимистической [101]. Сейчас предметом исследований является более слабый вариант, который и называют «проблемой Гильберта—Арнольда», хотя формально ни у Гильберта, ни у Арнольда такой задачи нет: «доказать, что в типичном $k$-параметрическом семействе векторных полей на плоскости $\mathbb{R}^2$ (или, после естественных доопределений, на проективной

[88] В той части [96], [97], где говорится об общих свойствах рассматриваемых слоений, имеются пробелы и неточности (что было сразу же отмечено несколькими людьми), но в этой части дефекты более или менее поправимы, требуя главным образом уточнений формулировок. В основном это указано в [98]. С критикой же, касающейся перехода от общих вопросов к самой 16-й проблеме Гильберта, они согласились позднее [100] (только после упомянутого семинара).

плоскости $\mathbb{RP}^2$) из полицикла рождается только конечное число предельных циклов, оцениваемое сверху константой, зависящей только от $k$». Она доказана для полициклов, содержащих только элементарные особые точки (т. е. такие, для которых у матрицы линейного приближения хоть одно из собственных значений ненулевое). В. Ю. Калошин получил для числа возможных предельных циклов, рождающихся в такой ситуации, оценку вида $e^{ck^2}$ с некоторой константой $c$. (Предположительно должна существовать полиномиальная по $k$ оценка.)

**2.4. Однородные потоки и гипотеза Рагунатана.** Значительным достижением в теории ДС алгебраического происхождения — так называемых однородных потоков — было полученное М. Ратнер доказательство гипотезы М. Рагунатана и её метрического аналога (ввиду наличия здесь аналогии говорят о метрической гипотезе Рагунатана, что, строго говоря, неправильно. Сам Рагунатан сформулировал свою гипотезу только в частном случае; общая формулировка и метрический аналог были предложены Ш. Дани). Приведу необходимые определения, сообщая заодно ещё некоторые сведения об однородных потоках.

Однородный поток — это ДС на однородном пространстве $G/D$, где $G$ — группа Ли, $D$ — её замкнутая подгруппа, определяемая левым[89] действием на $G/D$ некоторой подгруппы $H \subset G$: под действием элемента $h \in H$ смежный класс $gD$, служащий элементом $G/D$, переходит в $hgD$. Собственно, на $G/D$ действует слева вся $G$, но нас (в конечном счёте) интересует только ограничение этого действия на $H$. Ниже всегда предполагается, что $G/D$ имеет конечный объём; это значит, что на $G/D$ имеется конечная мера $\mu$, инвариантная относительно указанного действия $G$. (То, что известно о более общем случае — это скорее различные примеры.) В наиболее популярных (и старых) примерах подгруппа $H = \{h_t;\ t \in \mathbb{R}\}$ — однопараметрическая, так что получается поток в обычном смысле слова (с классическим временем); однако можно рассматривать и многомерные $H$.

Обозначим через $\mathfrak{g}$ алгебру Ли группы $G$, а через $\mathrm{GL}(\mathfrak{g})$ — группу Ли всех линейных преобразований $\mathfrak{g}$ как векторного пространства. В теории групп Ли вводят так называемое присоединённое представление Ad, являющееся некоторым гомоморфизмом групп Ли

$$\mathrm{Ad}\colon\ G \to \mathrm{GL}(\mathfrak{g}), \qquad g \mapsto \mathrm{Ad}_g$$

(и определяющее некоторое действие $G$ в $\mathfrak{g}$: $(g, X) \mapsto \mathrm{Ad}_g X$). Для его определения в общем случае потребовалось бы привести ряд начальных сведений из теории групп Ли, что потребовало бы места, а читателю, имеющему эти

[89] Подразумевается, что $G/D$ состоит из левых смежных классов $gD$. Если бы речь шла о правых смежных классах $Dg$ (их совокупность лучше обозначать через $D\backslash G$, но пишут и $G/D$), то однородный поток определялся бы правым действием.

сведения, определение Ad тоже должно быть известным. Но несложно привести определение Ad в важном частном случае, когда $G$ — матричная группа Ли, т. е. подгруппа группы невырожденных матриц $\mathrm{GL}(n,\mathbb{R})$ или $\mathrm{GL}(n,\mathbb{C})$ (с некоторым $n$), являющаяся гладким подмногообразием последней. $\mathrm{GL}(n,\mathbb{R})$ или $\mathrm{GL}(n,\mathbb{C})$ является открытым подмножеством в пространстве $\mathrm{Mat}(n,\mathbb{R})$ или $\mathrm{Mat}(n,\mathbb{C})$ всех матриц (в том числе и вырожденных) соответствующего порядка. Последнее как векторное пространство (т. е. отвлекаясь от умножения матриц) изоморфно $\mathbb{R}^{n^2}$ или $\mathbb{C}^{n^2}$ (рассматриваем коэффициенты матрицы как обычные координаты в $\mathbb{R}^{n^2}$ или $\mathbb{C}^{n^2}$, только занумерованные иначе). Итак, в данном случае $G$ является подмножеством векторного пространства, и ясно, что имеют в виду, говоря, что $G$ — гладкое подмногообразие. Оказывается, что $\mathfrak{g}$ можно считать просто касательным пространством к $G$ в точке, являющейся единичной матрицей $I$, только это пространство лучше перенести параллельно самому себе из $I$ в 0 (нулевая матрица). (Таким образом, обычными касательными прямыми к $G$ в $I$ служат прямые $I+tA$, $t\in\mathbb{R}$ или $\mathbb{C}$, $A\in\mathfrak{g}$.) Тогда действие $G$ в $\mathfrak{g}$ сводится просто к сопряжению матриц:

$$\mathrm{Ad}_g X = gXg^{-1} \qquad (g\in G\subset \mathrm{GL}(n,\mathbb{R}) \text{ или } \mathrm{GL}(n,\mathbb{C}),\ X\in\mathfrak{g}).$$

Элемент $g\in G$ называется унипотентным, если все собственные значения преобразования $\mathrm{Ad}_g$ равны 1. Подгруппа Ли $U\subset G$ называется унипотентной, если все её элементы унипотентны. Частным случаем являются так называемые орисферические подгруппы. Подгруппа $H\subset G$ называется орисферической, если имеется такой элемент $g\in G$, что $g^n h g^{-n}\to e$ при $n\to\infty$ для всех $h\in H$ ($e$ — единица группы $G$). Название вызвано тем, что в одном частном случае такие подгруппы тесно связаны с так называемыми орисферами в геометрии Лобачевского. (Объяснение этой связи потребовало бы слишком длительного отступления — пришлось бы описывать теоретико-групповую трактовку геодезических потоков на многообразиях (хотя бы только поверхностях) постоянной отрицательной кривизны. Если же читатель знаком с этой трактовкой, то ему, скорее всего, известна и упомянутая связь.) Однопараметрическая орисферическая подгруппа называется орициклической (по причине той же связи).

Однородный поток (вообще говоря, с многомерным временем) на факторпространстве $G/D$ группы Ли $G$, получающийся при действии унипотентной подгруппы $U\subset G$ на смежные классы посредством левых сдвигов, называют унипотентным. Если подгруппа орисферическая (орициклическая), то и поток называют орисферическим (орициклическим).

Гипотезы Рагунатана—Дани относятся к унипотентному потоку (возможно, с многомерным «временем») на факторпространстве $G/D$ конечного объёма. Первая гипотеза гласит, что замыкание каждой траектории такого потока является однородным подпространством конечного объёма, а вторая — что любая эргодическая мера для этого потока, конечная на компактах, сосредоточена на таком же подпространстве и имеет там простое алгебраическое описание, происходя из меры Хаара на подгруппе, с которой связано однородное подпространство. Вторая гипотеза доказана Ратнер и в таком варианте, когда не предполагается, что $G/D$ имеет конечный объём, а требуется только конечность рассматриваемой эргодической меры.

В доказательстве важную роль играет выделенное Ратнер некоторое свойство — его называют «свойством Ратнер» — однопараметрических унипотентных подгрупп. Судя по высказываниям самой Ратнер, нащупать это свойство ей помогли её предыдущие работы по однородным потокам, формально посвящённые совсем другим вопросам. (В них было обнаружено явление жёсткости орициклических потоков: в определённых случаях из метрического изоморфизма двух таких потоков следует, что этот изоморфизм имеет алгебраическое происхождение, в основном получаясь из некоторого внутреннего автоморфизма $G$ в сочетании со сдвигом на постоянное время вдоль траекторий.) С другой стороны, можно отметить ряд работ, где были доказаны частные случаи обоих гипотез. Большинство из них относится к орисферическим подгруппам; первый результат такого рода был получен Г. А. Хедлундом ещё в 1936 г. Однако по современным меркам для орисферических потоков ситуация слишком проста, в связи с чем для них, собственно, не столько доказывались гипотезы Рагунатана и Дани, сколько вообще проводилось далеко идущее исследование их топологических и метрических свойств. Г. А. Маргулис установил справедливость топологической гипотезы Рагунатана в значительно более сложном случае — для некоторых однопараметрических[90] унипотентных (но не орициклических) потоков на $\mathrm{SL}(3,\mathbb{R})/\mathrm{SL}(3,\mathbb{Z})$. Это позволило ему доказать известную в теории чисел гипотезу Оппенгейма—Давенпорта о квадратичных формах (идея о привлечении унипотентных потоков для её доказательства была высказана ещё Рагунатаном). Затем Маргулис и Дани доказали топологическую гипотезу Рагунатана для всех унипотентных потоков «общего положения» в том же пространстве и получили некоторые дополнения к предыдущему теоретико-числовому результату Маргулиса. (Вообще, создаётся впечатление, что в работах Маргулиса и др. возникает новый раздел геометрической теории чисел, в котором «местом действия» является не евклидово пространство и тор, а группы Ли и их однородные пространства. Наиболее близкие к евклидову случаю нильпотентные группы начали играть соответствующую роль ещё до Маргулиса, однако тогда речь шла не столько о новых теоретико-числовых результатах, сколько о «групповой и динамической» интерпретации уже известных фактов.)

После того как обе гипотезы (топологическая и метрическая) были доказаны, удалось установить справедливость некоторых их обобщений. Правда,

[90] То есть получающихся при действиях однопараметрических подгрупп.

в одном отношении это не совсем обобщения, потому что дополнительно требуется связность $U$ (насколько это дополнительное требование необходимо — этот вопрос обсуждается), зато в других отношениях условия на $U$ ослабляются до следующих: а) $U$ порождена унипотентными элементами (Ратнер); б) $U$ порождена квазиунипотентными элементами, т. е. элементами $g$, для которых собственные значения $\mathrm{Ad}_g$ равны 1 по модулю (А. Н. Старков). В случае *а* заключения остаются теми же, что и выше, а в случае *б* несколько модифицируются (например, замыкание траектории — по-прежнему многообразие, но уже не обязательно являющееся однородным пространством, а только определённым образом связанное с таковыми).

По контрасту с теоремой Ратнер стоит заметить, что давно известны случаи, когда замыкание некоторой траектории не является многообразием. Г. А. Маргулис указал, что такое явление заведомо имеет место, если однопараметрический однородный поток обладает известным в теории гладких динамических систем свойством равномерной частичной гиперболичности (в данном случае это эквивалентно тому, что он не квазиунипотентный).

Для полноты картины упомяну несколько более старых результатов об однородных потоках. В известном смысле теория однородных потоков сводится к теории эргодических однородных потоков: в неэргодическом случае Старков описал разбиение $G/D$ на эргодические компоненты, которые хотя и могут не быть однородными пространствами, но накрываются таковыми (с конечной кратностью накрытия); при этом ограничение исходного потока на компоненту поднимается до однородного потока на накрытии. Поэтому эргодические потоки заслуживают первостепенного внимания; накоплена значительная информация об их свойствах, относящихся к эргодической теории. Имеется критерий эргодичности однородного потока, формулирующийся в терминах алгебраических условий на определяющие поток «входные данные». В основном он был получен в итоге работ Л. Ауслендера (разрешимый случай), К. Мура (полупростой случай) и Ш. Дани (общий случай); некоторые завершающие штрихи независимо внесли Старков и Д. Витте. Описаны спектры однородных потоков. Для однородных потоков доказана гипотеза В. А. Рохлина, упомянутая в п. 1.3, *а* (Старков на базе метрической теоремы Ратнер). В специальном случае, когда $G = \mathrm{SL}(2, \mathbb{R})$, о свойствах однородных потоков (т. е. геодезических и орициклических потоков на поверхностях $G/D$ постоянной отрицательной кривизны) довольно много известно и без предположения о конечности объёма (в данном случае — площади) поверхности $G/D$ (хотя вопрос далеко не исчерпан). Это тесно связано с действиями фуксовых групп на круге Пуанкаре, а последние привлекает внимание уже более 100 лет. В [102] имеется сводка результатов, позволяющих сопоставить свойства потоков и действий.

Однородным потокам посвящены доклад Ратнер [103], обзоры [104], [105], [106] и книга [107].

**2.5. Гипотеза Зейферта.** Эта гипотеза состояла в том, что гладкий поток без положений равновесия на трёхмерной сфере $\mathbb{S}^3$ обязательно имеет замкнутую траекторию. Основанием для этой гипотезы послужила теорема Г. Зейферта, согласно которой замкнутая траектория имеется у всех потоков, получающихся при малом возмущении «потока Хопфа», который сейчас будет описан.

Реализуем трёхмерную сферу $\mathbb{S}^3$ как множество тех точек $(z, w)$ двумерной комплексной плоскости $\mathbb{C}^2$ (с вещественной точки зрения она четырёхмерна), для которых $|z|^2 + |w|^2 = 1$. Фазовая скорость потока Хопфа — это векторное поле, сопоставляющее точке $(z, w)$ вектор $(iz, iw)$. Траектории потока Хопфа суть окружности $\{e^{it}z, e^{it}w\}$; разбиение $\mathbb{S}^3$ на эти окружности — это известное в топологии расслоение Хопфа[91)], откуда и название потока. Помимо доказательства, данного самим Зейфертом, имеются по крайней мере два других доказательства его теоремы, принадлежащие Ф. Б. Фуллеру и М. Боттколу (ссылки и изложение идеи Фуллера см. в [78]). Фуллер использовал «индекс Фуллера» — введённую им топологическую характеристику поведения траекторий возле замкнутой траектории[92)], а Ботткол — своеобразный вариант теории возмущений (предложенный в одной работе Ю. Мозера о периодических решениях возле положения равновесия); таким образом, методы этих работ имеют более широкое значение (чего, по-видимому, нельзя сказать о доказательстве самого Зейферта). Но они тоже относятся только к малым возмущениям потока Хопфа.

С гипотезой Зейферта связана «гипотеза о торе»: если на границе «полнотория» («сплошного тора») $\mathbb{D}^2 \times \mathbb{S}^1$ векторное поле фазовой скорости направлено всюду внутрь (или всюду вовне) полнотория и в нём нет положений равновесия, то в нём имеется замкнутая траектория. Интуитивно кажется, что последняя должна, так сказать, делать один оборот вдоль полнотория; поэтому построенный Фуллером пример, когда, правда, имеется замкнутая траектория, но она гомотопна нулю в полнотории, послужил «тревожным сигналом», предостерегающим от доверия к наивной интуиции.

В настоящее время обе гипотезы опровергнуты даже для аналитических потоков. В основном это заслуга К. Куперберг [108], [109], построившей $C^\infty$-контрпримеры, после чего У. Тёрстен и Э. Гис указали аналитическую модификацию построения. Стоит отметить и вклад предыдущих авторов. Ф. В. Вильсон построил контрпримеры к многомерным аналогам гипотезы

---

91) Точнее, расслоение Хопфа — это отображение $\mathbb{S}^3 \to \mathbb{S}^2$, получающееся при отождествлении в точку каждой из указанных окружностей. С ним связано (неожиданное в то время) открытие Х. Хопфа, что гомотопическая группа $\pi_3(\mathbb{S}^2) \neq 0$.

92) Этот индекс не имеет ничего общего (кроме самого слова «индекс») с упомянутым в п. 1.1 индексом Конли. Если говорить о соответствующих классических корнях, то индекс Фуллера связан не с индексом Морса, а с индексом Кронекера—Пуанкаре.

Зейферта. Это, в общем, не удивительно, — довольно понятно, что в многомерном случае квазипериодические траектории вполне могут «заменить» периодическую, что и имеет место у Вильсона, — однако часть его технических приёмов пригодилась последующим авторам. Неожиданностью был контрпример П. Швейцера[93] (уже «настоящий», трёхмерный) с потоком гладкости $C^1$ (1974 г.). После этого более гладкие варианты обоих гипотез не внушали большого доверия, однако даже поднять гладкость в контрпримерах до $C^2$ удалось далеко не сразу (Дж. Харрисон).

Интересно, что, как доказал Х. Хофер, гипотеза Зейферта верна для так называемых контактных потоков [111]. Контактные потоки бывают на нечётномерных многообразиях $M^{2n+1}$. Такой поток определяется с помощью так называемой контактной формы $\lambda$, т. е. пфаффовой формы, для которой $(2n+1)$-мерная форма

$$\lambda \wedge \underbrace{d\lambda \wedge \ldots \wedge d\lambda}_{n \text{ раз}} \tag{32}$$

всюду отлична от нуля. (Я опускаю уточнения о требуемой степени гладкости $\lambda$.) Это определение напоминает определение симплектической структуры на чётномерном многообразии[94], но по существу здесь имеется по крайней мере одно существенное различие. Задание симплектической формы никак не выделяет никаких направлений в точках $M$. Контактная же форма в каждой точке $x$ определяет некоторое одномерное направление, — именно, направление вырождения формы $d\lambda$: вектор $X \in T_xM$ имеет это направление, если

$$d\lambda(X, Y) = 0 \quad \text{для любого } Y \in T_xM.$$

Вдобавок фиксируется имеющий это направление вектор $X_x \in T_xM$, для которого $\lambda(X_x) = 1$. Векторное поле $X$ иногда называют полем Ж. Риба. Контактный поток — это поток с векторным полем фазовой скорости $X$.

В действительности теорема Хофера является более общей: если у трёхмерного замкнутого многообразия $M$ одномерная группа когомологий

[93] На русском языке пример Швейцера изложен в книге Тамуры [110].

[94] Сходство усиливается, если учесть следующую теорему Г. Дарбу: в окрестности каждой точки $x$ можно ввести такие локальные координаты $(x_1, y_1, \ldots, x_n, y_n, z)$, в терминах которых $\lambda = dz + \sum y_i\, dx_i$. Однако контактной структурой обычно называют не пару $(M, \lambda)$ с контактной формой $\lambda$, а определяемое последней на $M$ поле $2n$-мерных касательных подпространств $E_x \subset T_xM$, где

$$E_x = \{X \in T_xM;\ \lambda(X) = 0\}.$$

Уточню, что только такое поле $E_x$ $2n$-мерных касательных подпространств называют контактной структурой, которое получается указанным образом при помощи некоторой контактной формы. Сказанное равносильно тому, что контактная форма $f\lambda$, где $f$ — скалярная функция, определяет ту же самую контактную структуру, что и $\lambda$.

$H^1(M, \mathbb{R}) = 0$, то любой контактный поток на $M$ имеет замкнутую траекторию. Это частный случай гипотезы А. Вайнстейна, в которой многообразие не предполагается трёхмерным.

Форму (32) можно принять за форму объёма на $M$. (Для читателя, который связывает понятие «объём» только с римановой геометрией, замечу, что если на $m$-мерном гладком многообразии $M$ задана нигде не обращающаяся в нуль $m$-мерная внешняя форма $\Omega$, то на $M$ существует такая риманова метрика, для которой «элемент объёма» выражается как раз формой $\Omega$.) Легко доказать, что контактный поток сохраняет задаваемый этой формой объём. Возникает вопрос, не верна ли гипотеза Зейферта для любых потоков, сохраняющих объём? Г. Куперберг (сын К. Куперберг) построил противоречащий пример к этому предположению (ссылку см. в [109]).

## §3. Некоторые другие достижения

Повторяю, что в этом параграфе я часто ещё более краток, чем в предыдущих, и нередко вместо того, чтобы объяснять результаты, только называю их. Но я по-прежнему указываю литературу, где можно найти изложение части важнейших результатов и ссылки на другие работы.

**3.1.** Прежде всего, имеются области, которые привлекали внимание много более четверти века назад и которые по сей день остаются областями активных исследований; последние ведутся более или менее в тех же направлениях, что и раньше, хотя и обогатились значительными новыми идеями, понятиями, методами и т. д. В докладе Йоккоза [1] речь идёт как раз о двух таких направлениях, различающихся характером исследуемых движений. Одно из них посвящено квазипериодическим траекториям и траекториям, в чём-то близким к квазипериодическим (например, сюда относятся «канторторы»), а другое — гиперболическому поведению траекторий, куда опять-таки включается не только «чистая гиперболичность» в том виде, как она оформилась в 60-е гг., но и чем-то похожие на неё случаи. Как уже говорилось, я не собираюсь дублировать его доклад. Я сделаю только несколько небольших литературных добавлений.

*а*) Своеобразную часть «гиперболической» теорией составляют результаты (главным образом, французских математиков, хотя почин здесь положил В. Тёрстен) о трёхмерных потоках Аносова и связанных с ними геометрических вопросах. О ситуации на 1991 г. см. в [77]. Одна из последних работ в данной области [112].

*б*) В 60-е гг. при исследовании такого классического геометрического объекта, как геодезические потоки на многообразиях отрицательной кривизны, роль геометрии явно уступала роли теории ДС. За последние 20 лет роль геометрии заметно возросла, особенно когда кривизна предполагается не строго отрицательной, а только неположительной. См. обзор [113] и, как более новые примеры (и источники других ссылок), статьи [114][95], [115].

*в*) Йоккоз уделил сравнительно мало внимания специальным свойствам инвариантных мер при гиперболическом поведении траекторий. Об относительно более старых результатах (частично подразумеваемых известными в [1]) см. [2]. Отмечу, далее, недавнюю статью [116] (где можно найти ряд литературных ссылок на предшествующие работы), завершившую многолетние усилия по исследованию некоторых давно известных вопросов о гиперболических мерах. Инвариантная относительно диффеоморфизма $f\colon M \to M$ (который далее будет класса $C^{1+\varepsilon}$) нормированная мера $\mu$ с компактным носителем называется гиперболической, если почти всюду (в смысле этой меры) под действием итераций отображения $f$ (точнее, его «касательного расширения», «дифференциала» или «производной») $Tf$ касательные векторы изменяются «экспоненциально»: при почти всех $x \in M$ для всех $X \in T_xM$

$$\varlimsup_{n\to\infty} \frac{\ln \|T_x f^n X\|}{n} \neq 0.$$

Основной результат [116] — метрический аналог известного топологического факта о локальном строении гиперболических множеств, о котором можно прочитать в любом учебнике или обзоре по гиперболической теории. Локально-максимальное гиперболическое множество $A$ локально (в окрестности любой своей точки $x$) устроено как прямое произведение некоторого множества $B$ в локальном неустойчивом слое $W^u_{\mathrm{loc}}(x)$ на некоторое множество $C$ в локально устойчивом слое $W^s_{\mathrm{loc}}(x)$; при этом точке $(y, z) \in B \times C$ соответствует точка множества $A$, лежащая на пересечении проходящего через $y$ локального устойчивого слоя $W^s_{\mathrm{loc}}(y)$ с проходящим через $z$ локально неустойчивым слоем $W^u_{\mathrm{loc}}(z)$. (Я опускаю детальные уточнения.) Это утверждение постоянно используется, а некий ослабленный его вариант имеет

[95] В ней для одного класса замкнутых многообразий неположительной кривизны — так называемых многообразий ранга 1 — исследуется введённое ранее разбиение единичного касательного расслоения на два инвариантных множества, на первом из которых поведение траекторий в инфинетезимальных терминах (терминах уравнений в вариациях) является, так сказать, более гиперболическим, чем на втором. Если инфинитезимальные характеристики не обманывают, то можно ожидать, что поведение траекторий на первом множестве является «более стохастическим», чем на втором. В [114] показано, что так и есть в двух (тесно связанных) отношениях: в отношении топологической энтропии и в отношении асимптотики по $T$ числа замкнутых геодезических длины $\leqslant T$. Хотя внешне статья [114] выглядит столь же аналитической, как и работы 60-х гг., в ней, помимо самого понятия ранга 1 и упомянутых двух множеств, используются ещё следующие объекты геометрического происхождения: функция Буземана, абсолют и некоторые меры на нём, строящиеся с помощью приёма типа рядов Пуанкаре в теории автоморфных форм.

место и при неравномерной гиперболичности. Имея локальную структуру прямого произведения, естественно поинтересоваться, не представляется ли при этом и каждая гиперболическая инвариантная мера в виде прямого произведения меры на $W^u_{\mathrm{loc}}(x)$ и меры на $W^s_{\mathrm{loc}}(x)$. Для негиперболических мер ожидать этого было бы неестественно, но и для гиперболических мер в общем случае не приходится надеяться на утверждение такого рода. Однако в [116] установлен некоторый ослабленный аналог локального представления гиперболических мер в виде таких прямых произведений: грубо говоря, представление имеет место с ошибкой, выражаемой дополнительным множителем, который с уменьшением размера шаровой $\varepsilon$-окрестности точки $x$ меняется медленнее сколь угодно малой степени $\varepsilon$. При этом получена некоторая дополнительная информация о свойствах мер-«сомножителей».

В связи с исследованием этого вопроса в [116] доказано также, что гиперболические меры имеют некоторые «хорошие» свойства, которыми произвольные инвариантные меры динамических систем обладать не обязаны. Например, известно, что имеются различные понятия размерности множества; каждое из них разумным образом отражает какое-то из тех свойств, которые естественно считать «размерностными». В общем случае различные размерности не совпадают, но они совпадают для «хороших» множеств. Менее известно, что совершенно аналогично имеются различные понятия размерности меры. Оказывается, что для гиперболических мер ряд размерностей совпадает.

Эти результаты могут показаться по самой своей формулировке «техническими» (тогда как в других местах я приводил результаты, формулировки которых звучат более «непосредственно»), но поскольку я сыграл в своё время некоторую роль в создании гиперболической теории, то, думаю, я мог себе позволить положиться на возникшее у меня впечатление об их важности. Впрочем, я уже говорил, что отбор материала для §3 более субъективен, чем для §§1, 2,

*г*) Дополнительно к сказанному в [1] о канторторах можно отметить ещё прежние работы об аналогичных объектах для некоторых потоков [117], [118], [119], а также некоторые статьи, в которых развиваются три разных подхода, отличных от освещаемого в [1]. А. Б. Каток [120] отметил возможность исследования соответствующих инвариантных множеств для двумерных диффеоморфизмов с помощью давнишних идей Дж. Биркгофа. А. Фати [121] рассматривает канторторы как своего рода обобщённые решения (типа последнего варианта таковых — так называемых «вязких решений») для уравнения Гамильтона—Якоби. Наконец, Р. Мане обратил внимание на некоторые новые обстоятельства, связанные с суперлинейными лагранжианами; они касаются не только канторторов. См. [122], [123].

Другой пример успешного продолжения деятельности, начавшейся ранее — работы А. Д. Брюно и его сотрудников по локальной теории (см. [87],

[124]). Начиная по крайней мере с Пуанкаре, важнейший метод локальной теории состоит в построении формального (т. е. представляющегося с помощью формальных степенных рядов) «нормализующего» преобразования, приводящего (локальный) поток возле положения равновесия или диффеоморфизм возле неподвижной точки к некоторому более простому виду — так называемой нормальной форме. В работах группы Брюно широкий круг соответствующих вопросов был изучен с наибольшей полнотой и ими было рассмотрено много приложений. Вопросы, касающиеся нормальных форм, включают: построение нормализующих преобразований (одна из заслуг А. Д. Брюно, относящаяся к более раннему времени, состоит в предложенном им в многомерном случае геометрическом приёме, обобщающем многоугольник Ньютона[96)]); обсуждение структуры нормальной формы и возможностей, связанных с некоторой неоднозначностью построения (когда она имеется); анализ сходимости построенных преобразований. Приложения относятся не только к самим обыкновенным дифференциальным уравнениям (в том числе зависящим от параметров, что существенно для теории бифуркаций), но и к некоторым уравнениям в частных производных.

При хорошей разработанности данного метода можно оценивать «степень нетривиальности» той или иной задачи по тому, насколько полно её удаётся исследовать с его помощью и в какой мере нужны дополнительные соображения. Не все это понимают, с чем и связан эпиграф, которым Брюно начал свою книгу: «Мой дядя... удивлялся, куда деваются неудавшиеся механики, оружейники, сапожники, слесари, инженеры...» (Марк Твен). Если Марк Твен полагал, что они идут в часовые мастера, то Брюно намекает, что их «способности» находят и иное применение...

**3.2. Теория сингулярных возмущений.** Это ещё один пример успешного продолжения начатых ранее исследований. Речь идёт о системах обыкновенных дифференциальных уравнений вида

$$\varepsilon\dot{x} = f(x, y, \varepsilon), \quad \dot{y} = g(x, y, \varepsilon), \tag{33}$$

где $\varepsilon$ — малый параметр. «Сингулярность» состоит в том, что малый параметр входит как множитель при производной, а не просто как параметр, от которого зависит правая часть системы (в (33) она тоже может зависеть от $\varepsilon$, но это менее существенно). Конечно, можно ввести новое («медленное») время $s := t/\varepsilon$ и, обозначая дифференцирование по нему штрихом, переписать (33) как

$$\dot{x} = f(x, y, \varepsilon), \quad \dot{y} = \varepsilon g(x, y, \varepsilon). \tag{34}$$

---

[96)] Метод может применяться также для локального исследования систем алгебраических уравнений; как известно, многоугольник был предложен Ньютоном как раз для этого в частном случае одного уравнения с двумя неизвестными.

Теперь $\varepsilon$ входит только в правую часть, но зато в новых терминах отрезку времени $t$ конечной длины $T$ соответствует отрезок медленного времени $s$ длины $T/\varepsilon$, неограниченно удлиняющийся при $\varepsilon \to 0$ (а мы, естественно, хотим изучить поведение решений (33) по крайней мере на конечном отрезке времени, не уменьшающемся при $\varepsilon \to 0$), так что всё равно получается не совсем обычная задача о возмущениях.

Так или иначе, ясно, что $x$ меняется намного быстрее $y$, и поэтому свойства решений зависят, прежде всего, от свойств «системы быстрых движений»

$$\dot{x} = f(x, y, 0), \tag{35}$$

в которую $y$ входит как постоянный параметр. При тех или иных предположениях о (35) можно попытаться описать более медленное изменение $y$.

Простейший случай здесь тот, когда решения (35) стремятся (при $t \to \to \infty$) к экспоненциально устойчивому положению равновесия $x_0(y)$, вообще говоря, зависящему от «параметра» $y$. Тогда естественно предположить, что в системе (33) это $y$ изменяется со временем приблизительно как решение системы

$$\dot{y} = g(x_0(y), y, 0). \tag{36}$$

Если же решения (36) со временем стремятся к экспоненциально устойчивым положению равновесия $y_0$ или замкнутой траектории $l_0$, то и решения (33) стремятся к положению равновесия $(x_0(y_0), y_0)$ или, соответственно, к замкнутой траектории $L_\varepsilon$, близкой к кривой $L_0 := \{(x_0(y), y);\ y \in l_0\}$. Результаты такого типа (разумеется, при надлежащих уточнениях условий на систему) были получены около 50 лет назад. Помимо доказательства утверждения о пределе траекторий при $\varepsilon \to 0$, можно провести более точное исследование их зависимости от $\varepsilon$, а именно, получить для них асимптотическое разложение с точностью до любой степени $\varepsilon$. Для периодической траектории $L_\varepsilon$ получается асимптотический ряд по степеням $\varepsilon$, для траектории же с фиксированным начальным значением $(x', y')$, где $x' \neq x_0(y')$, соответствующее разложение содержит также члены с $\ln \varepsilon$. Доказывается также единственность замкнутой траектории $L_\varepsilon$, близкой к $L_0$. Это результат несколько иного характера, чем асимптотика $L_\varepsilon$. Ведь *a priori* могли бы существовать две замкнутые траектории $L'_\varepsilon$ и $L''_\varepsilon$, «расстояние» между которыми имело бы более высокий порядок малости, чем любая степень $\varepsilon$, скажем, экспоненциальный порядок $O(e^{-1/\varepsilon})$ или $O(e^{-1/\sqrt{\varepsilon}})$. Упомянутое асимптотическое разложение с точностью до любого $\varepsilon^n$ этого попросту «не могло бы почувствовать».

Более сложная ситуация возникает, когда при некоторых («бифуркационных») значениях $y$ сливаются два положения равновесия системы (35) — устойчивое $x_0(y)$ и неустойчивое, но при этом где-то в стороне от этого «места слияния» у (35) имеется устойчивое положение равновесия $x_1(y)$, к которому после прохождения бифуркационного значения $y$ может «притянуться» наша траектория. Допустим, что те $y$, при которых происходит такое слияние, образуют гладкую гиперповерхность $M$ в пространстве $y$-ов (точнее, накладываются аналитические условия на $M$, при которых это так) и что траектории системы (36) попадают на $M$, имея трансверсальное к $M$ направление (это формулируется в виде явных условий на $f$ и $g$). Если это так для траектории системы (36), к которой близка $y$-компонента интересующей нас траектории системы (33), то естественно предположить, что после этого последняя траектория быстро переходит к $x_1(y)$ и далее её $y$-компонента близка к траектории $y(t)$ системы (36), в которой $x_0(y)$ заменено на $x_1(y)$, а $x$-компонента близка к $x_1(y(t))$. Этот процесс может повторяться. Тогда можно ожидать, что траектория системы (33) будет состоять из дуг двух типов — одни из них аналогичны описанным выше в более простой ситуации и близки к дугам вида $\{(x_i(y_i(t)), y_i(t))\}$, где $y_i$ суть решения систем вида (36) (с какими-то $x_i(y)$ вместо $x_0(y)$), а другие близки к каким-то дугам, каждая из которых идёт от двух сливающихся положений равновесия системы (35) к некоторому устойчивому положению равновесия. Дуги первого типа проходятся за конечное время, а второго — очень быстро. Нетрудно сообразить, что такая система дуг может «замыкаться», образуя замкнутую кривую $L_0$. Тогда можно ожидать, что (опять-таки при надлежащих уточнениях условий) при малых $\varepsilon$ у (33) будет существовать замкнутая траектория $L_\varepsilon$, близкая к $L_0$. Стоит заметить, что подобные объекты известны в физике, доставляя математическое описание некоторых типов релаксационных колебаний[97].

Исследование этих вопросов началось в 40-х гг. с того случая, когда $x$ и $y$ «одномерны» (Ж. Хааг, А. А. Дородницын). В 50-х гг. Л. С. Понтрягин и Е. Ф. Мищенко получили существенные результаты в многомерном случае. В этих работах были получены первые члены асимптотических разложений для различных дуг траекторий, для периодической траектории $L_\varepsilon$ и её периода. Эти асимптотические разложения оказались значительно сложнее, чем в предыдущем случае (появляются дробные степени $\varepsilon$) и их едва ли можно было предвидеть заранее. Несколько позднее Н. Х. Розов выяснил

[97] Первоначальный смысл последнего термина был связан с физической природой колебательной системы. Релаксационные системы противопоставлялись системам другого (более обычного) характера (которые одно время называли системами «томсоновского типа»), примером которых является обычный радиогенератор. В нём имеется колебательный контур, «подпитываемый» энергией; с другой стороны, энергия «уходит» из контура, главным образом ввиду излучения генератора (ради чего он и создан) и отчасти ввиду наличия сопротивления. Устойчивая периодическая траектория, описывающая генерируемые колебания, близка к одной из траекторий, описывающей свободные колебания контура, а её амплитуда определяется балансом между поступающей и уходящей энергией. В релаксационной колебательной системе колебания возникают иначе: сперва в некоторой части системыкаким-то образом накапливается энергия, а потом она «разряжается» через другую часть

структуру всего асимптотического разложения в случае «одномерных» $x$, $y$. В последнем случае единственность замкнутой траектории $L_\varepsilon$ доказывается очень просто, так что в этом случае теория приобрела известную законченность, тогда как в многомерном случае оставались открытыми вопросы о единственности $L_\varepsilon$ и о полной структуре асимптотического разложения. (Подробно о состоянии дел к началу рассматриваемого периода см. [125].)

Ответы на эти вопросы были получены около 10 лет назад. Единственность $L_\varepsilon$ была независимо доказана К. Боне и группой четырёх русских математиков — Е. Ф. Мищенко, Ю. С. Колесов, А. Ю. Колесов, Н. Х. Розов. См. [126].

Напомню, что выше одно из условий состояло в том, что при попадании траектории системы (36) на гиперповерхность $M$ (где сливаются устойчивое и неустойчивое положения равновесия системы (35)) вектор $g(x_0(y), y, 0)$ (задающий направление этой траектории) трансверсален к $M$. Возникает вопрос, как ведут себя траектории системы (33) возле той точки, где $g(x_0(y), y, 0)$ касается $M$? Существование таких точек — явление, в понятном смысле достаточно «типичное», за исключением того случая, когда $y$ «одномерно», так что от этого вопроса нельзя отмахнуться как от относящегося к какой-то исключительной ситуации. Если же $y$ одномерно (так что $M$ — просто точка), то сам по себе данный вопрос относился бы к исключительной ситуации, но если при этом (33) зависит от некоторого параметра $a$, то такая ситуация может возникать при некотором $a = a_0$ и это уже является достаточно «типичным». В таком случае, естественно, ставится вопрос об исследовании поведения решений (33) не только при $a = a_0$, но и при значениях $a$, близких к $a_0$. Если в правой части (33) заменить $\varepsilon$ на 0, то у полученной «слегка упрощённой» системы точка $M$ будет положением равновесия. Когда $a$ пробегает некоторый отрезок, в системе (33) (уже не упрощённой) происходит бифуркация положения равновесия, близкого к $M$. Особенно интересен случай, когда это бифуркация Хопфа; интересно проследить, как с изменением $a$ рождающийся предельный цикл, вначале маленький и почти

---

системы (отсюда и название, происходящее от *relaxation* — отдых, разрядка). В зависимости от устройства системы, этот процесс может быть «сбалансирован» таким образом, что колебания получатся похожими на гармонические (хотя никакого колебательного контура нет); но вместо этого «разрядка» может происходить весьма быстро по сравнению с более медленным накоплением энергии, тогда колебание получается «близким к разрывному» («разрыв» соответствует разрядке). В последнем случае математическое описание системы часто даётся системой вида (33), в которой с изменением «медленной» переменной $y$ происходят вкратце описанные выше явления (слияние двух положений равновесия, и т. д.); малый параметр $\varepsilon$ «отвечает» за скорость разрядки. В математической литературе установилась традиция использовать название «релаксационные колебания» только в этом последнем случае.

эллиптической формы, вырастает до совсем не похожего на него «почти разрывного» предельного цикла, о котором шла речь выше.

Сперва был исследован именно этот последний вопрос, что осуществила в начале рассматриваемого периода группа французских математиков, связанных со Страсбургом (Э. Бенуа, Ф. и М. Дьене, Ж.-Л. Калло, А. Трёш, Е. Урлаше и др.); их работа была начата по предложению Ж. Риба (известного своим вкладом в теорию слоений). См. [127], [128], [129], [130]. Кому-то показалось, что «растущий» (при изменении $a$) предельный цикл при некоторых $a$ похож на летящую утку; различные его части получили соответствующие названия, от «клюва» до «хвоста». Вскоре все исследуемые в этих задачах (в том числе и при отсутствии параметра $a$, но при «не-одномерном» $y$) траектории стали называть «траекториями-утками» (даже если они незамкнутые, не говоря уже об отсутствии сходства с утками — летящими, плавающими, ходящими или жареными), а вся эта деятельность получила название «охота на уток». Своеобразной особенностью французских работ было систематическое использование нестандартного анализа, что отражено в [127] и [128]. Французские авторы, как и авторы [128], явно считают язык нестандартного анализа более удобным для проведения рассуждений в этой области (включая построение различных асимптотических разложений). Формулировки же окончательных утверждений если и не даются на стандартном языке, то обычно легко переводятся на таковый, к тому же авторы [127] и [128] местами специально поясняют, как это делается. Впрочем, статья [129] положила начало использованию при «охоте на уток» одной только стандартной математики. В рамках последней в [126] исследован ряд вопросов об «утках» при «не-одномерном» $y$.

Очевидно, возможен и такой случай, когда при изменении $y$ положение равновесия $x_0(y)$ системы (35) претерпевает бифуркацию иного рода — от него ответвляется устойчивый предельный цикл, а само оно остаётся, но становится неустойчивым. Казалось бы, тогда можно ожидать, что траектория системы (33) быстро переходит к предельному циклу. Но действительность оказывается более сложной. Возможно (и в некотором смысле является достаточно типичным) явление «затягивания», когда и после этой бифуркации $x$-компонента траектории долго остаётся возле $x_0(y)$ (а $y$ при этом по-прежнему изменяется приблизительно согласно (36)). Исследование этого явления было начато в одном частном (но вполне «представительном») случае Л. С. Понтрягиным и М. А. Шишковой в 1973 г. Достаточно законченные результаты в общем случае были получены спустя немногим более 10 лет А. И. Нейштадтом. Ссылки см. в [52]; дополнительно см. [131], [132], [130]. В [133] обсуждается сходное явление, связанное с потерей устойчивости циклом.

Далее, траектории (35) могут при всех рассматриваемых $y$ стремиться не к положению равновесия, а к устойчивой замкнутой траектории $C(y)$. В этом случае естественно ожидать, что $x$-компонента решения (33) с начальным значением $(x', y')$ быстро приближается к $C(y')$ и в дальнейшем всегда остаётся вблизи $C(y)$, где $y$ — это $y$-компонента рассматриваемой траектории, а эволюция $y$ приближённо описывается с помощью осреднения уравнения $\dot{y} = g(x, y, 0)$ вдоль $C(y)$. Опять-таки можно рассмотреть тот случай, когда осреднённая система имеет экспоненциально устойчивые положение равновесия или замкнутую траекторию. По существу, основная работа здесь была выполнена Н. М. Крыловым и Н. Н. Боголюбовым (а в частных случаях — также и рядом других авторов) задолго до начала рассматриваемого периода, но постановка задачи у них была несколько иной; в духе излагаемого здесь подхода задача была исследована Л. С. Понтрягиным и Л. В. Родыгиным в 1960 г. (см. [125]).

В предыдущем случае в системе (35) устанавливается одночастотный колебательный режим. Вполне возможна и ситуация, когда в ней с самого начала (без переходного процесса) имеют место многочастотные колебания. Если они не зависят от $y$, вопрос сводится к поведению решений уравнения $\dot{y} = g(t, y, \varphi)$ с «многочастотной» зависимостью $g$ от $t$. Существенные результаты о задачах такого типа были получены (тоже до начала интересующего нас периода) Н. Н. Боголюбовым и его сотрудниками (прежде всего, Ю. А. Митропольским). Иной характер имеет ситуация, когда многочастотные колебания в (35) зависят от $y$. Казалось бы, в этом случае нельзя рассчитывать на сколько-либо общие результаты, потому что в достаточно реалистических задачах при сколь угодно малом изменении начальных значений $(x', y')$ характер соответствующей траектории (35) может существенно измениться. Скажем, то она плотна на торе большой размерности, то является замкнутой траекторией. По чему же осреднять уравнения для $y$?

В 1960 г. я указал, что можно получить вполне удовлетворительный ответ в достаточно общем случае[98], если интересоваться поведением не всех траекторий, а только их «большинства». «Исключительные» траектории, для которых метод осреднения в его естественной (или «наивной»?) формулировке «не работает», отвечают множеству начальных значений, мера которого при $\varepsilon \to 0$ стремится к нулю. Точнее, при фиксированных $\delta > 0$ и $T > 0$ к нулю стремится мера $\mu_{\delta,T}(\varepsilon)$ множества тех $(x', y')$, для которых ошибка метода осреднения на отрезке $[0, T]$ превосходит $\delta$. Можно сказать, что речь идёт о сходимости решений (при $\varepsilon \to 0$) по мере начальных значений. Одновременно родственные, но (по крайней мере формально) менее общие результаты получил Т. Касуга.

[98] Собственно, речь идёт даже не о многочастотных колебаниях в системе (35), а о наличии у неё при каждом $y$ «хорошей» инвариантной меры, «хорошей» системы первых интегралов $I_1(x, y), \ldots, I_k(x, y)$ и эргодичности системы на почти каждой поверхности $I_j = \mathrm{const}$.

Я не дал оценки $\mu_{\delta,T}(\varepsilon)$, а только доказал, что эта величина стремится к 0 при $\varepsilon \to 0$. Это связано со значительной общностью теоремы, при которой рассчитывать на такую оценку (во всяком случае, на сколько-либо удовлетворительную оценку) не приходится. Спустя примерно 15 лет, т. е. уже в начале рассматриваемого периода, выяснилось, что при разумном снижении общности такую оценку можно получить. Речь идёт о тех случаях, когда в (35) происходят «настоящие» многочастотные колебания (решения являются квазипериодическими). Наиболее существенный шаг здесь сделал А. И. Нейштадт, рассмотревший важный случай, когда базисные частоты этих колебаний невырожденным образом зависят от $y$. Он показал, что $\mu_{\delta,T}(\varepsilon)$, самое большее, пропорциональна $\sqrt{\varepsilon}/\delta$, причём в классе степенных оценок этот результат является окончательным — имеются примеры, где $\mu_{\delta,T}(\varepsilon)$ отличается от $\sqrt{\varepsilon}/\delta$ только на не очень существенный логарифмический множитель. Позднее с помощью тех же методов результаты такого рода были распространены (с определённым изменением формулировок) на некоторые системы с вырожденной зависимостью частот от $y$; наибольших успехов здесь добился В. И. Бахтин [134]. Ссылки на более ранние работы см. в [135].

**3.3. Экспоненциально малые эффекты в теории возмущений.** Собственно, уже многие результаты из п. 3.2 выходят за рамки «степенно́й» теории возмущений, потому что даже для аналитических систем соответствующие ряды по степеням малого параметра обычно расходятся, и потому любое утверждение о существовании и единственности замкнутой траектории (или, тем более, квазипериодического решения) связаны с каким-то выходом за рамки этой теории. Но это — утверждения качественного характера. За последние 25 лет был получен ряд количественных результатов, относящихся к экспоненциально малым эффектам. Результаты о потере устойчивости в п. 3.2 отчасти имеют такой характер. Вот ещё несколько направлений, где были получены такие результаты:

а) задачи о разделении движений;

б) задачи о расщеплении сепаратрис;

в) задачи о сохранении адиабатических инвариантов;

г) теория Нехорошева;

д) задачи о диффузии Арнольда.

Некоторую информацию и литературные ссылки можно найти в [135], [52] и [136]. Дополнительные ссылки (упоминаются только работы последнего 25-летия, но в них можно найти упоминания о сравнительно немногочисленных более ранних работах, не указанных в цитированных выше книгах):

а) [131], [133], [137];

б) [138], [139], [140], [141];

г) [142], [143], [144];

д) [143].

**3.4. Формула для энтропии.** Ещё до начала рассматриваемого периода возникла гипотеза, что для любых компактного гладкого многообразия $M$ и гладкого отображения $f\colon M \to M$ топологическая энтропия $h_{\mathrm{top}}(f)$ не меньше спектрального радиуса индуцированного отображения $f_*$ в «полной» группе гомологий $H_*(M,\mathbb{R}) := \bigoplus H_i(M,\mathbb{R})$ с вещественными коэффициентами. О ситуации в начале этого периода см. [145]. Теперь эта гипотеза доказана Йомдиным [146], [147], [148].

**3.5. Интегрируемые и неинтегрируемые системы.** Задача об интегрировании дифференциальных уравнений (т. е. о нахождении их решений) так же стара, как стара сама теория дифференциальных уравнений. Когда-то (до Коши) теория дифференциальных уравнений (которую в то время не выделяли формально в отдельный раздел анализа) в основном состояла в нахождении приёмов интегрирования, годных для тех или иных уравнений или классов таковых. (Впрочем, почти одновременно в небесной механике стали разрабатываться методы теории возмущений. Но это в то время, по-видимому, воспринималось именно как часть небесной механики, а не теории дифференциальных уравнений.) Иногда находятся не решения, а первые интегралы. (Авторы, говорящие «решение» вместо «интеграл», часто сокращают «первый интеграл» просто до «интеграла».)

Здесь уместно пояснить происхождение терминологии. Она связана с тем, что процесс решения дифференциального уравнения можно рассматривать как некое обобщение обычного процесса интегрирования ($\int f(x)\,dx$ есть решение простейшего дифференциального уравнения $\dfrac{dy}{dx} = f(x)$). Поэтому функции, являющиеся решениями дифференциального уравнения (или наборы функций, являющихся решениями системы дифференциальных уравнений), назвали его (её) интегралами. Теперь их чаще называют просто решениями (как, в частности, и делается в настоящей статье), хотя при этом одним и тем же словом называется и функция, удовлетворяющая уравнению, и процесс отыскания такой функции. (Однако Бурбаки сохранили старинную терминологию.) В отличие от интегралов, первый интеграл — это функция, постоянная вдоль решения. Название связано с тем, что часто первый этап процесса нахождения решений («интегралов») состоит в нахождении первого интеграла или, в случае системы, «полной» системы первых интегралов — такой их системы $F_1, \ldots, F_k$, что любое решение полностью получается как решение системы уравнений

$$F_1 = c_1 = \mathrm{const}, \quad \ldots, \quad F_k = c_k = \mathrm{const} \tag{37}$$

при каких-то $c_1, \ldots, c_k$. Если такие $F_i$ найдены, то решение системы дифференциальных уравнений сводится к решению «конечной» системы (37), что считалось более простым делом и во всяком случае не относится к предмету теории дифференциальных уравнений. В действительности, если первые интегралы получились сколько-либо сложными, то этот второй этап решения отнюдь не прост. В практически решённых задачах, кроме совсем простых, при этом обычно приходилось привлекать эллиптические и родственные им более сложные функции. И можно представить себе, что при ещё более сложных первых интегралах решать (37) можно только с помощью компьютера и это, вообще говоря, ничуть не проще (а часто и сложнее), чем численное интегрирование исходной системы дифференциальных уравнений. Но даже когда это так, знание первых интегралов может позволить получить такие качественные заключения о поведении решений, которые никак не видны из системы дифференциальных уравнений самой по себе.

Естественно, что на раннем этапе развития теории было проинтегрировано большое число относительно простых задач; часто это была просто удача. В любом справочнике содержатся многочисленные примеры такого рода, являющиеся, так сказать, «разрозненными» в том смысле, что каждый из них был проинтегрирован сам по себе, без связи с остальными. Многие из проинтегрированных задач относятся к аналитической механике, которая долгое время была основным «потребителем» теории дифференциальных уравнений, да и вообще анализа.

Имеется некоторое различие между тем, как обычно понимается интегрируемость в «общей» теории обыкновенных дифференциальных уравнений и в аналитической механике (где приходится иметь дело с уравнениями Эйлера—Лагранжа или Гамильтона). В первой, говоря об интегрируемости, обычно подразумевают, что интересующая нас функция $f$ (решение или первый интеграл; в случае системы обыкновенных дифференциальных уравнений речь может идти о нескольких первых интегралах, образующих полную систему таковых) получается путём определённых операций, исходя из простейших функций — многочленов. Эти операции суть: алгебраические операции; дифференцирование; (неопределённое) интегрирование; потенцирование (переход от $g$ к $e^g$). Для такой $f$, очевидно, получается явное выражение, вообще говоря, содержащее интегралы («квадратуры»), которые не обязательно «берутся», в связи с чем говорят, что $f$ выражается в квадратурах или что дифференциальное уравнение (система) интегрируется в квадратурах. В списке операций не упоминается о ряде «элементарных функций» — логарифме, прямых и обратных тригонометрических функциях. Однако легко видеть, что их использование при построении $f$ можно заменить использованием указанных выше операций. Например, $\ln g = \int \frac{g'}{g}\,dx$. Иногда к списку добавляют ещё одну операцию — решение алгебраического уравнения, коэффициентами которого являются ранее построенные функции. Тогда явного выражения для $f$, вообще говоря, не получается. В этом случае говорят о представлении с помощью обобщённых квадратур.

В аналитической механике обычно говорят об интегрируемости при наличии полной системы первых интегралов, являющихся либо сравнительно простыми функциями координат в фазовом пространстве — алгебраическими, либо любыми аналитическими функциями этих координат. Не выделяется отдельно случай, когда первый интеграл выражается в квадратурах. В классических и новых проинтегрированных задачах первые интегралы часто являются алгебраическими (а то и рациональными) функциями удачно введённых координат. Отрицательные результаты — что такая-то система не имеет первых интегралов или не имеет полной системы первых интегралов — обычно намного легче доказываются для алгебраических интегралов, чем для аналитических. В то же время алгебраичность или неалгебраичность функции на фазовом пространстве зависит от используемых координат, поэтому отрицательные результаты об алгебраических первых интегралах связаны с координатами, и заранее нельзя сказать, не может ли случиться, что при использовании каких-то других фазовых переменных система станет интегрируемой.

Имеется ещё один вариант понятия интегрируемости, который восходит к С. В. Ковалевской. Общеизвестно, что она нашла новый случай, когда уравнения движения тяжёлого твёрдого тела допускают четвёртый интеграл[99] и в конечном счёте — полную систему первых интегралов. Менее известно, что в основном она решала задачу: в каких случаях все решения, рассматриваемые в комплексной области, являются мероморфными функциями времени[100]? (Тот факт, что в найденном ею новом случае (как и в ранее известных) имеется четвёртый интеграл, получился как побочный результат.) В связи с этим позднее стали говорить об «интегрируемости по Ковалевской».

Общих методов интегрирования, которые были бы столь же универсальными, как правила дифференцирования, не существует. Всё же ряд важных задач аналитической механики (математически формулирующихся в виде уравнений Эйлера—Лагранжа или Гамильтона) был в своё время решён на основании всего двух (частично перекрывающихся) способов. Во-первых, при наличии непрерывных групп симметрий у этих задач имеются соответствующие первые интегралы (теорема Э. Нётер). Во-вторых, можно перейти к некоторому уравнению в частных производных (уравнение Гамильтона—Якоби) и попытаться подобрать координаты, в которых переменные разделяются. Нет никаких гарантий, что в том или ином конкретном случае

[99] Если педантично соблюдать различие между «интегралами» и «первыми интегралами», то надо говорить о «четвёртом первом интеграле», как бы странно не звучало словосочетание «четвёртый первый».

[100] Ковалевская указала, что её результаты относятся и к более общему вопросу: когда все решения являются однозначными функциями комплексного времени? Справедливость этого утверждения подтвердил А. М. Ляпунов.

удастся применить какой-нибудь из этих способов, но, действуя в обратном направлении и стараясь найти такие задачи, в которых эти способы применимы, удалось найти ряд интегрируемых задач аналитической механики. Заметим, что для задач этого типа решение (когда его удаётся найти) обычно особенно чётко разбивается на два указанных выше этапа — нахождение достаточно полной системы первых интегралов и получение с их помощью явных выражений для зависимости координат и импульсов от времени. На первом этапе большую роль играет теорема Ж. Лиувилля[101)], согласно которой для «механической» ДС с $n$ степенями свободы достаточно найти такие $n$ функционально независимых первых интегралов $F_1, \ldots, F_n$, для которых все скобки Пуассона $\{F_i, F_j\} = 0$ (в таком случае говорят, что эти интегралы находятся в инволюции). Уже на этом этапе можно сделать далеко идущие качественные заключения о поведении траекторий, что давно делалось в конкретных примерах и что в общем случае отметил В. И. Арнольд (в связи с чем соответствующую теорему часто называют теоремой Лиувилля—Арнольда). При разделении переменных, в принципе, можно было бы избежать первого этапа, по крайней мере, первые интегралы могли бы и не выступать явно. Однако практически они появляются и особо отмечаются.

Незадолго до начала рассматриваемого периода П. Лакс предложил новый, третий способ интегрирования (в смысле нахождения системы первых интегралов) — «метод $(L, A)$-пары». Может быть лучше будет сказать, что это третий способ нахождения интегрируемых задач. Он не связан специфически с аналитической динамикой, но практически большинство его применений относится к гамильтоновым системам и доставляет системы, интегрируемые по Лиувиллю. Метод применим к уравнениям в частных производных; на самом деле он был предложен именно в связи с одним из таких уравнений (уравнением Кортевега—де Фриса), причём этому предшествовали исследования этого уравнения с иных позиций. (Вначале оно исследовалось численными методами и это исследование было связано с более ранними работами, где речь шла о других уравнениях и других вопросах; это поучительная и драматичная история, но она увела бы нас слишком в сторону. Начатые в связи с этим аналитические исследования со временем привели к открытию другого метода интегрирования этого уравнения,

---

[101)] Как я понимаю, в основном для данного круга задач случае автономной (т. е. не содержащей явно времени) системы уравнений механики эта теорема в случае $n$ степеней свободы была впервые опубликована малоизвестным математиком Э. Буром (Bour), а Лиувилль указал её обобщение на неавтономный случай, которое не играет заметной роли. (Правда, Лиувилль ссылался на более ранний свой устный доклад. Кроме того, раньше он указал частный случай этой теоремы для $n = 2$, но последний по существу до Лиувилля был известен К. Якоби и С. Пуассону.) Здесь подтверждаются слова, что имущему добавится, а у неимущего отнимется. При выполнении условий теоремы Лиувилля говорят, что система «интегрируема по Лиувиллю».

связанного с обратной задачей теории рассеяния. Лакс, по-видимому, хотел с другой точки зрения осмыслить полученные результаты.) Но Лакс сразу понял, что его метод имеет более широкую область применимости.

Вероятно, сейчас большинство применений метода $(L, A)$-пары и его модификаций, в том числе и большинство наиболее интересных применений, относится к уравнениям в частных производных. Однако метод дал довольно много и для обыкновенных дифференциальных уравнений.

Непосредственно метод $(L, A)$-пары отвечает указанному выше первому этапу (нахождение первых интегралов). Но очень быстро П. Лакс и С. П. Новиков добавили весьма существенные соображения, относящиеся уже ко второму этапу.

Если попытаться получить явные выражения для координат и импульсов как функций времени на основании того, что непосредственно содержится в доказательстве теоремы Лиувилля, то оказывается, что применительно к ряду классических задач механики поначалу кажется, что приходится обращаться к сравнительно сложным аналитическим функциям, однако, привлекая различные конкретные соображения, можно выразить ответ в терминах сравнительно более простых функций. Но это обычно выглядит результатом каких-то случайных совпадений. Хотелось бы иметь какие-то общие формулировки, более непосредственно указывающие на те функции, обращение к которым действительно необходимо. Едва ли это можно сделать сколько-либо удовлетворительным и «естественным» образом в тех терминах, в которых формулируется теорема Лиувилля (если бы это было возможно, такие формулировки нашли бы примерно век назад). Но это оказалось возможным с помощью некоторой модификации метода $(L, A)$-пары — речь идёт о так называемых $(L, A)$-парах со спектральным параметром. Вообще говоря, они не всегда существуют, а если и существуют, найти их труднее, чем просто какие-нибудь $(L, A)$-пары. Зато если уж они имеются, они дают намного больше.

Как и «обычный» метод $(L, A)$-пары, метод спектральной $(L, A)$-пары интенсивно применялся не только к обыкновенным дифференциальным уравнениям, но и к уравнениям в частных производных. Эта сторона дела приводит к глубокому изучению алгебро-геометрических аспектов интегрируемых систем — настолько глубокому, что с ним связано открытие новых фактов даже в самой алгебраической геометрии. Правда, последнее было связано более с уравнениями в частных производных, да и вообще в ряде публикаций по алгебро-геометрическим аспектам интегрируемых систем (из числа наиболее важных) более значительное внимание уделяется уравнениям в частных производных[102]. Но теперь появилась книга, в которой можно

[102] Если бы я писал обзор по последним, то о тематике настоящего пункта надо было бы сказать в параграфе «Новые или «обновлённые» направления». Но применительно к теории ДС я, как видно, отнёс её в §3.

найти изложение, с самого начала ориентированное на гамильтоновы системы [149]. Обращаю внимание читателя на недавно появившийся сборник переводов на русский язык более старых статей Ю. Мозера об интегрируемых системах [150].

Наряду с открытием новых интегрируемых систем и исследованием их свойств за последние 25 лет был получен ряд результатов в «противоположном» направлении — доказано, что ряд конкретных систем не интегрируем[103]. Первые результаты такого рода были получены около 1900 г., но большинство результатов того времени носило довольно ограниченный характер. Обычно доказывалось, что система механического происхождения не имеет полной системы алгебраических первых интегралов. Очевидно, алгебраичность первого интеграла зависит от используемых координат. Конечно, когда говорят, что в той или иной механической задаче нет алгебраических первых интегралов, кроме уже известных в механике, то подразумевают использование тех координат, которые естественным образом появляются при самой формулировке этой задачи. И тем не менее такой результат оставляет впечатление незавершённости, ибо не исключает такой возможности, что «дополнительные» (к уже известным) первые интегралы всё-таки существуют, но в «естественных» координатах они выражаются трансцендентными функциями.

Для того чтобы доказать «настоящую» неинтегрируемость, нужны не столько алгебраические, сколько геометрические соображения. Хотя они появились уже у Пуанкаре, систематическая трактовка этих вопросов началась с работ В. М. Алексеева, появившихся немногим менее чем за 10 лет до начала рассматриваемого периода. Затем значительный вклад внесли В. В. Козлов и С. Л. Зиглин, см. обзор Козлова [151] и [135]. В более новой книге [152] имеются и более новые результаты о неинтегрируемости (наряду с рядом «положительных» сведений об интегрировании ряда конкретных систем[104]). Специально геодезическим потокам посвящены важные работы И. А. Таймaнова и Г. Патернайна. См. [153].

В ряде случаев речь идёт о несуществовании аналитических первых интегралов. Насколько существенна аналитичность? В одних случаях существенна, в других — нет. Крайний пример последнего: у систем Аносова с «хорошей» инвариантной мерой даже и измеримых первых интегралов нет (ввиду эргодичности). Но это, как сказано, крайний пример. Для систем с двумя степенями свободы имеются довольно хорошие результаты, гарантирующие (при определённых предположениях) отсутствие «дополни-

---

[103] Здесь и далее под неинтегрируемостью имеется в виду отсутствие интегрируемости по Лиувиллю.

[104] Метод $(L, A)$-пары фигурирует в [152] под названием «представление Гейзенберга». Спектральные $(L, A)$-пары там не рассматриваются.

тельного» первого интеграла с мало-мальски приличной гладкостью. При этом основную роль, как и предвидел Пуанкаре, играют гомоклинические траектории (и связанные с ними гиперболические множества), а прочие условия, накладываемые в теоремах о неинтегрируемости, являются как бы дополнительными. Когда же число степеней свободы больше двух, то ситуация существенно усложняется: если не говорить об аналитических первых интегралах, то достаточно удобных дополнительных условий пока не известно.

Л. Батлер [154] построил ряд (родственных друг другу) примеров аналитических римановых метрик с интегрируемыми геодезическими потоками, для которых «дополнительные» (к интегралу энергии) первые интегралы — класса $C^\infty$, но не аналитические, и при этом разбиение фазового пространства на области, заполненные инвариантными торами (фигурирующими в теореме Лиувилля—Арнольда), не обладает теми геометрическими свойствами, которые следовали бы из аналитичности первых интегралов и которые используются для вывода ограничений на геометрию фазового пространства интегрируемых систем. Соответственно, и сами эти ограничения не выполняются. Используя идею Батлера, А. В. Болсинов и И. А. Тайманов [155] построили «усовершенствованные» примеры, в которых топологическая энтропия интегрируемого аналитического геодезического потока положительна. Более того, ограничение этого потока на некоторое «исключительное» инвариантное подмногообразие является потоком Аносова (даром что по своим свойствам потоки Аносова и интегрируемые потоки — это как бы две крайние противоположности).

**3.6. Теория Конли.** Недавно я написал предисловие к намечаемому русскому изданию книги по этой теории [156], которым позволю себе воспользоваться здесь. Примерно до 1970 г. сколько-либо далеко идущее использование топологических методов производилось только для систем нескольких специальных типов (например, имеющих вариационную природу). Когда же речь шла о системах, так сказать, общего характера, то применения топологии были довольно примитивными (что, конечно, не означает, будто они были не важными). В основном они так или иначе были связаны с вращением векторного поля на границе области и индексом Пуанкаре—Кронекера[105)] нулей последнего. По существу, при этом речь идёт о чисто топологических (и притом довольно простых) понятиях и утверждениях —

[105)] Как известно, в математике имеется целый ряд объектов различной природы, носящих название «индекс» (не считая тех индексов, которые являются значками при основных буквах), поэтому во избежание путаницы приходится к слову «индекс» добавлять различные уточняющие слова, нередко напоминающие об авторах соответствующего понятия. Когда же из контекста ясно, о чём идёт речь, говорят просто «индекс».

ничего специфически связанного с динамическими системами в них нет[106]. Имелся ещё принцип Важевского, (с которым тесно связаны идеи Конли). Но этот принцип долго занимал особое, изолированное положение.

Новые понятия, введённые Конли, уже существенным образом связаны с динамическими системами (что, впрочем, относится и к принципу Важевского), причём заранее не делается никаких предположений о специфике системы[107]. Конли подчёркивает другую сторону дела — своего рода «грубость» (в том же смысле, что и в выражении «грубо, но надёжно») соответствующих объектов. В его теории даётся далеко идущее обобщение классического понятия индекса Морса. Сам М. Морс отправлялся от вариационных задач и говорил об индексе невырожденной критической точки функции, но давно известно, что на его индекс можно посмотреть с иных позиций — с точки зрения теории динамических систем. Именно, критическая точка функции является положением равновесия соответствующего градиентного потока и её индекс Морса естественно интерпретируется в терминах свойств потока возле этого положения равновесия. В теории Конли речь идёт не только о положениях равновесия, но о широком классе так называемых изолированных (или локально-максимальных) компактных инвариантных (т. е. состоящих из траекторий) подмножеств фазового пространства потока, не обязательно градиентного. Для этих множеств вводятся некоторые топологические характеристики, по-прежнему называемые индексами; точнее говоря, эти индексы характеризуют не только и не столько само это множество в смысле его «внутреннего» строения, сколько некоторые особенности поведения траекторий возле этого множества. Новые индексы являются более общими и сложными объектами, нежели прежние индексы Морса, но они по существу сводятся к последним в том случае, с которым имел дело Морс, чем и объясняется использование прежнего названия.

Уже при самом возникновении новой теории начали появляться работы, посвящённые её приложениям к математическим вопросам небесной механики (задача трёх тел и смежные вопросы) и к задачам, связанным с бегущими волнами; позднее появились и другие применения в теории уравнений с частными производными. Неожиданным было использование «индексных» соображений в работах Ч. Конли и Э. Цендера, упомянутых в п. 1.1 и, как говорилось, сыгравших заметную роль в формировании симплекти-

---

[106] Хотя работа, нужная для того, чтобы «подвести» исследуемый вопрос под соответствующие топологические утверждения, может быть весьма нетривиальной и существенно связанной с теми или иными особенностями рассматриваемой динамической системы.

[107] Другое дело, что применение теории Конли к заданной системе может оказаться бессодержательным или практически невозможным. С такой возможностью приходится считаться при любой попытке использования любой общей теории.

ческой геометрии как отдельной дисциплины высокого таксономического ранга. Правда, теперь, по-видимому, роль «индексных» соображений в этой дисциплине значительно уменьшилась.

Название «теория Конли» обусловлена ролью Ч. Конли в возникновении и развитии этой теории. Сам он отмечал, что при её возникновении значительную роль сыграл также Р. Истон. Позднее над ней и её приложениями работал целый ряд авторов.

Краткое изложение исходных понятий теории Конли имеется в [78]. Этой теории посвящены лекции Конли [156] и Мишайкова [157].

**3.7. Особенности в задаче *n* тел** (движение материальных точек (частиц), притягивающихся по закону Ньютона). Математическое описание такой системы даётся некоторой гамильтоновой системой в $6n$-мерном фазовом пространстве с переменными $p_i, q_i$, $i = 1, \ldots, 3n$. Здесь $q_{3i-2}, q_{3i-1}, q_{3i}$ суть обычные координаты $i$-й частицы в обычном (физическом) пространстве $\mathbb{R}^3$, а $p_{3i-2}, p_{3i-1}, p_{3i}$ — проекции её импульса на координатные оси в $\mathbb{R}^3$. Даже не выписывая системы, легко понять, что она имеет особенности в тех точках, где при каких-нибудь $i \neq j$ координаты $i$-й частицы совпадают с координатами $j$-й. Обозначим через $\Sigma$ множество соответствующих точек в конфигурационном пространстве (пространстве переменных $q_i$); тогда особые точки системы суть точки множества $\Sigma \times \mathbb{R}^{3n}$. Если начальные значения $(p(0), q(0))$ взяты вне этого множества, то локально существует соответствующее решение $(p(t), q(t))$. Оно не обязательно продолжается на всю положительную полуось $t$. Если максимальный интервал существования решения конечен, скажем, является интервалом $[0, T)$, $T < \infty$ (мы рассматриваем только положительные $t$, хотя на самом деле это несущественно), то говорят, что при $t = T$ у решения имеется особенность. (Вектор-функция $(p(t), q(t))$ и впрямь имеет особенность при $t = T$.) Легко доказать, что $q(t)$ при $t \to T$ неограниченно приближается к $\Sigma$. Если существует предел $\lim q(t)$ при $t \to T$ (неизбежно принадлежащий $\Sigma$), то говорят, что в системе при $t = T$ происходит столкновение. (Этимология здесь очевидна: в пределе при $t \to T$ две или более частиц оказываются в одной и той же точке физического пространства $\mathbb{R}^3$). До недавнего времени оставался открытым вопрос, бывают ли особенности, отличные от столкновений? Известно, что при $n \leqslant 3$ таких особенностей нет. Для $n = 2$ это просто, а для $n = 3$ было доказано около 100 лет назад.

Если столкновение двойное, то, как обнаружил ещё Л. Эйлер, можно разумным (и вполне наглядным) образом определить дальнейшее движение системы (в течение некоторого времени)[108]. Эйлеровское доопределение

---

[108] За это Эйлера критиковал Вольтер, говоривший — и, конечно, правильно говоривший, — что эйлеровское описание движения после столкновения физически нереально. Однако реально движение частиц, которые пролетают очень близко друг

движения после двойного столкновения увеличивает максимальный интервал существования решения, но (за исключением случая $n = 2$) всё-таки не обязательно делает его бесконечным. Так, при $n = 3$ всё может кончиться тройным столкновением (до которого могут происходить двойные столкновения, но в течение конечного времени их может произойти только конечное число)[109].

Перед самым началом последнего 25-летия Дж. Мезер и Р. Мак-Гехи [158] обнаружили, что уже при $n = 4$ возможны особенности иного типа. В их примере за конечное время $T$ происходит бесконечное число двойных столкновений. Все четыре тела движутся по одной прямой, причём в итоге при $t \to T$ три тела уходят в бесконечность — одно в одну сторону, а два других, неограниченно сближаясь между собой (что и даёт энергию для всего процесса) — в другую; четвёртое же тело осциллирует между этими двумя, попеременно сталкиваясь то с одним, то с другим. Возник вопрос, существуют ли такие особенности, которые не являются столкновениями и которыми завершается некоторый интервал $[0, T)$, на котором не происходит столкновений? Здесь уже нельзя обойтись одномерным случаем, что усложняет задачу. В 1992 г. Дж. Ксиа [159] показал, что при $n \geqslant 5$ подобное явление возможно[110]. Вопрос остаётся открытым только при $n = 4$.[111]

**3.8. Устойчивая эргодичность.** $C^2$-гладкие системы Аносова с «хорошей» инвариантной мерой эргодичны и остаются таковыми при малых (уже при только $C^1$-малых) возмущениях, не нарушающих ни $C^2$-гладкости системы, ни наличия у неё «хорошей» инвариантной меры. Это свойство — система остаётся эргодической при малых (в смысле некоторого $C^r$) возмущениях, не нарушающих «хорошей» инвариантной меры, — можно назвать «устойчивой эргодичностью». Спрашивается, не существует ли других устойчиво эргодических систем?

Положительный ответ на этот вопрос является заслугой Ч. Пью и М. Шуба (к которым присоединились и другие авторы). Вначале речь шла об от-

---

к другу, но всё-таки не сталкиваются. Эйлеровское движение после столкновения описывает предел такого движения, если рассматривать частицы, пролетающие всё ближе и ближе друг к другу. Впрочем, сам Эйлер мотивировал своё доопределение движения при $t > T$ иначе — на основании его аналитических свойств.

[109] Помимо этой качественной стороны дела, имеется ещё и аналитическая — о характере функций, описывающих особенности того или иного типа столкновений. Поэтому в целом ситуацию с особенностями и при $n = 3$ нельзя резюмировать несколькими фразами. Сводку сведений, имевшихся примерно 15 лет назад, см. в [135].

[110] Ксиа указывает, что Дж. Джервер иным способом установил возможность особенностей, не сводящихся к столкновениям, в задаче $n$ тел с некоторыми большими $n$.

[111] Напрашивается мысль посмотреть движения, близкие к движениям в примере Мезера—Мак-Гехи, но уже не «умещающиеся» на прямой, а происходящие на плоскости или в $\mathbb{R}^3$. Однако до сего времени это не привело к успеху.

дельных примерах, потом — уже о некоторых (хотя и довольно специальных) классах устойчиво эргодических систем. Разумеется, представляет интерес исследование их прочих эргодических свойств (и того, насколько они изменяются при возмущениях). По крайней мере в некоторых случаях, как и для систем Аносова, имеет место не только устойчивая эргодичность, но и (аналогично определяемое) «устойчивое К-свойство» и даже «устойчивая бернуллиевость». Несомненно, в гладкой эргодической теории открылась новая глава.

Во время написания настоящей статьи многие работы в этой области имелись только в виде препринтов. Известные мне публикации: [160], [161], [162], [163]. Близкому вопросу посвящены статьи [164], [165]. В них рассматриваются ДС специального типа — некоторые косые произведения — и речь идёт о сохранении эргодичности и более сильных свойств при малых возмущениях в классах таких систем.

**3.9. «Абстрактная» (чисто-метрическая) эргодическая теория.** Мои интересы связаны преимущественно с гладкими динамическими системами, что в самом общем плане неизбежно обуславливает моё отношение к данной тематике. В предыдущий период главным достижением в чисто-метрической эргодической теории было развитие «энтропийной» теории и затем примыкающей к ней теории Орнстейна; влияние этих достижений на «гладкую» теорию трудно переоценить. Достижения последних 25 лет, насколько я могу судить, такого влияния не имели. О работах предшествующего времени см. [4], [45], [166].

**О строении систем с инвариантной мерой.** Ниже под ДС мы будем понимать динамическую систему с дискретным временем $\{\varphi^n\}$ в «хорошем» пространстве $X$ с нормированной инвариантной мерой $\mu$ (определённой на $\sigma$-алгебре $\mathfrak{B}$ подмножеств $X$). Г. Фюрстенберг и Р. Циммер [167], [168], [169], [170] разработали теорию, в которой ДС представляется в виде «обратного предела» последовательности (возможно, трансфинитной) ДС, которая начинается с тривиальной ДС в пространстве, сводящемся к одной точке, и в которой каждая следующая ДС является так называемым «примитивным расширением» предыдущей (а ДС, занумерованная трансфинитным числом $\alpha$, являющимся пределом возрастающей последовательности трансфинитов, — обратный предел последовательности ДС с номерами $\beta < \alpha$). Здесь фигурируют три новых понятия: «обратный (он же проективный) предел», «расширение» и «примитивное расширение». Ни одного из них я точно определять не буду, но всё же попытаюсь хотя бы отчасти объяснить их суть, пренебрегая некоторыми деталями и пользуясь аналогиями.

Сперва надо сказать о расширениях пространств с мерой. Это в какой-то степени аналог «расслоения» в топологии. Пространство с мерой $(X, \mathfrak{B}, \lambda)$ является расширением пространства с мерой $(Y, \mathfrak{C}, \mu)$ (а последнее в духе топологии можно было бы назвать «базой», но чаще говорят, что $(Y, \mathfrak{C}, \mu)$

является «фактором» $(X, \mathfrak{B}, \lambda)$), если нам задана фиксированная сюръекция (отображение «на») $\pi\colon X \to Y$ («проекция на базу (на фактор)»), которая измерима (если $A \in \mathfrak{C}$, то $\pi^{-1}A \in \mathfrak{B}$) и сохраняет меру (в том же предположении $\lambda(\pi^{-1}A) = \mu(A)$). По аналогии с топологией, множества $\pi^{-1}(y)$, $y \in Y$, называют слоями. Примером может служить проекция прямого произведения $Y \times Z$ пространств с мерой $(Y, \mathfrak{C}, \mu)$ и $(Z, \mathfrak{D}, \nu)$ с соответствующей $\sigma$-алгеброй измеримых множеств и мерой $\lambda := \mu \times \nu$ на один из «сомножителей» $Y$ или $Z$, скажем, на $Z$. В этом примере слой $\pi^{-1}(z) = Y \times \{z\}$ и на нём имеется естественная мера $\lambda_z$ — это просто мера $\mu$ в $Y$, «перенесённая» в слой $\big(\lambda_z(A \times \{z\}) := \mu(A)\big)$. По теореме Фубини, для измеримого $A \subset Y \times Z$

$$\lambda(A) = \int_Z \lambda_z(A \cap \pi^{-1}(z))\, d\nu(z). \tag{38}$$

Если меру интерпретировать как вероятность, то в духе теории вероятностей $\lambda_z(A \cap \pi^{-1}(z))$ можно интерпретировать как условную вероятность «события» $A$ при условии $z$. (Заметим, что, вообще говоря, множество этих «условий» (т. е. точек $z \in Z$) несчётно, поэтому в данном случае понятие условной вероятности является довольно деликатным.) Вместо этого в «чистой» теории меры можно говорить об «условной мере». Оказывается, что в «хороших» пространствах с мерой в слоях расширения $\pi\colon X \to Y$ тоже возникает система «условных мер» — в каждом слое $\pi^{-1}(y)$ своя мера $\lambda_y$, причём аналогично (38)

$$\lambda(A) = \int_Y \lambda_y(A \cap \pi^{-1}(y))\, d\mu(y).$$

Я опускаю здесь и далее необходимые уточнения о пренебрежении множествами меры нуль и об измеримости различных объектов. С подобной небрежностью для функции $f$ на $X$ можно говорить об $L_2$-нормах её ограничений на различные слои, т. е. о

$$\|f\|_y := \int_{\pi^{-1}(y)} |f(x)|^2\, d\lambda_y(x).$$

Ниже встречается ещё аналог «послойного произведения» двух расслоений (которое наиболее известно для векторных расслоений, в каковом случае оно называется «суммой Уитни»). Пусть $(X, \mathfrak{B}, \lambda)$ и $(Y, \mathfrak{C}, \nu)$ — расширения $(Z, \mathfrak{D}, \nu)$ с проекциями $\pi$ и $\rho$. Их послойное произведение

$$X \times_Z Y := \{(x, y);\ x \in X,\ y \in Y,\ \pi(x) = \rho(y)\}.$$

Подразумевается, что оно снабжается мерой $\lambda \times_Z \mu$, где

$$\lambda \times_Z \mu(A) := \int_Z \lambda_z(A \cap \pi^{-1} z) \mu_z(A \cap \rho^{-1}(z))\, d\nu(z).$$

Измеримые подмножества пространства $X \times_Z Y$ определяются таким образом, чтобы для них последнее определение имело смысл; я не буду пояснять этого подробнее. Заметим, что $X \times_Z Y$ естественным образом является некоторым расширением $Z$.

ДС $\{\psi^n\}$ в фазовом пространстве $(X, \mathfrak{B}, \lambda)$ называется расширением ДС $\{\varphi^n\}$ в $(Y, \mathfrak{C}, \mu)$, если первое пространство представлено как расширение второго с проекцией $\pi$ и последняя коммутирует с соответствующими преобразованиями: $\varphi \circ \pi = \pi \circ \psi$. Если ДС $\{\chi^n\}$ в $X$ и $\{\psi^n\}$ в $Y$ являются расширениями ДС $\{\varphi^n\}$ в $Z$, то их послойное произведение $\{(\chi \times_Z \psi)^n\}$ — это ДС в $X \times_Z Y$, для которой

$$(\chi \times_Z \psi)(x, y) := (\chi(x), \psi(y)).$$

Она сохраняет послойное произведение соответствующих мер и является некоторым расширением ДС $\{\varphi^n\}$ в $Z$.

Вводятся некоторые классы расширений ДС. Нам нужны два класса: «слабо перемешивающие» и «компактные». В том случае, когда база сводится к одной точке, соответствующие ДС суть слабо перемешивающие ДС и ДС, метрически изоморфные групповым сдвигам компактных коммутативных групп. (ДС последнего типа можно охарактеризовать ещё так: орбиты соответствующего унитарного оператора $U_\varphi$ в гильбертовом пространстве $L_2(X, \lambda)$ (здесь $(U_\varphi f)(x) := f(\varphi x)$) условно компактны. Отсюда и название. Если мы применяем общее определение компактного расширения к данному частному случаю, непосредственно получается именно это последнее свойство.) Известно, что слабое перемешивание эквивалентно непрерывности спектра ДС (т. е. оператора $U_\varphi$, рассматриваемого на ортогональном дополнении к константам); кроме того, если ДС эргодична, то компактность эквивалентна дискретности спектра. Таким образом, если ДС эргодична, то введённые два класса отвечают выделению ДС с непрерывным и, соответственно, дискретным спектром, однако сейчас это делается не в спектральных терминах.

В общем случае, когда база нетривиальна, в определении слабо перемешивающего расширения ДС $\{\varphi^n\}$ в $Y$ говорится, что это есть такое расширение $\{\psi^n\}$ в $X$, «послойный квадрат» которого $\{(\psi \times_Y \psi)^n\}$ не имеет иных инвариантных множеств, кроме полных прообразов инвариантных множеств ДС $\{\varphi^n\}$ (рассматриваемой как фактор ДС $\{(\psi \times_Y \psi)^n\}$). Когда $Y$ сводится к одной точке, получаем известное определение непрерывности спектра: декартов квадрат $\psi \times \psi$, действующий в $X \times X$, эргодичен. А вот одно из определений компактного расширения. В $L_2(X, \mathfrak{B}, \lambda)$ плотны такие функции $f$, что ограничения функций $U_{\varphi^n} f$ со всевозможными $n$ на слои $\pi^{-1}$ составляют условно компактные подмножества

$$\{U_{\varphi^n} f|_{\pi^{-1}(y)};\ n \text{ любые}\} \subset L_2(\pi^{-1}(y), \lambda_y), \tag{39}$$

и даже более того: $\varepsilon$-сети для таких множеств с различными $y$ можно выбрать, в известном смысле, согласованным образом — при любом $\varepsilon > 0$ существуют такие $g_1, \dots, g_k \in L_2(X, \mathfrak{B}, \lambda)$, что их ограничения на слои являются $\varepsilon$-сетями для множеств (39). Как видно, оба определения — слабо перемешивающего и компактного расширений — являются как бы «послойными» модификациями определений слабого перемешивания и условной компактности орбит $\{U_{\varphi^n} f\}$.

Примитивное расширение — это расширение, являющееся либо слабо перемешивающим, либо компактным. Наконец, предел обратного спектра понимается практически так же, как это принято в алгебре и топологии, с единственным различием, что теперь мы имеем дело с другими структурами (пространства с мерой, в которых действуют преобразования), и надо позаботиться о соответствующей структуре в предельном пространстве.

Сказанное (и определения, и основная структурная теорема) обобщается на ДС с неклассическим временем, пробегающим коммутативную группу $G$ конечного ранга. Наиболее существенное изменение касается определения примитивных расширений. В этом случае расширение называется примитивным, если $G$ можно представить как прямое произведение $G_1 \times G_2$ таким образом, что расширение оказывается компактным, если его рассматривать как расширение ДС со временем из $G_1$, и слабо перемешивающим, если его рассматривать как расширение со временем из $G_2$. Стало быть, мы не можем сделать так, чтобы каждый шаг нашей последовательности расширений был одного из двух простейших типов, но можем обеспечить, чтобы на каждом шаге эти два типа сочетались только некоторым простым способом.

Применяя идеи из эргодической теории или топологической динамики — иногда структурную теорему [167], [168], иногда заметно более простые соображения [171], [168], — к ДС Бернулли, удалось получить сравнительно простые и единообразные доказательства ряда теорем теории чисел, включая как известную теорему Ван дер Вардена (здесь структурной теоремы ещё не требуется), так и теорему Е. Семереди, дающую далеко идущее развитие теоремы Ван дер Вардена и высказанную в виде гипотезы П. Эрдешем и П. Тураном. (Формулировки этих двух теорем приводятся ниже.) Многие из этих теорем были известны ранее, но, например, $n$-параметрический аналог теоремы Семереди оказался новым; его доказательство в духе первоначальных рассуждений Семереди (если таковое возможно) было бы, по-видимому, весьма громоздким.

Я поясню характер таких применений теории ДС на примере теоремы Ван дер Вардена, утверждающей, что если $\mathbb{Z}_+$ (нам сейчас чуть удобнее начинать с нуля) разбито на $m$ непересекающихся подмножеств $A_i$, то для любого $l \in \mathbb{N}$ существует такое $i$, что в $A_i$ содержится «отрезок» арифметической прогрессии длины $l$. Довольно легко доказать следующее утверждение: если $\varphi\colon X \to X$ — непрерывное отображение метрического компакта в себя, то для любого $l \in \mathbb{N}$ найдутся такие последовательность $n_k \to \infty$ и точка $x \in X$, что $\varphi^{in_k} x \to x$ при всех $i = 1, \ldots, l$. Возьмём в $\{1, \ldots, r\}^{\mathbb{Z}_+}$ точку $\xi = \{\xi_n\}$, для которой $\xi_n$ равно номеру того из множеств $A_i$, которому принадлежит $n$, и применим данное утверждение к ограничению топологического сдвига Бернулли $\sigma$ на замыкание $X$ траектории $\{\sigma^n \xi\}$. Для соответствующей точки $x = \{x_n\}$ и достаточно больших $k$ нулевая координата каждой из точек $\sigma^{in_k} x$ с $i = 1, \ldots, l$ совпадает с нулевой координатой точки $x$, т. е.

$$x_0 = x_{n_k} = x_{2n_k} = x_{3n_k} = \ldots = x_{ln_k}.$$

А раз $x$ лежит в замыкании множества $\{\sigma^n \xi\}$, то при некотором $j$ первые $ln_k$ координат точки $\sigma^j \xi$ совпадают с первыми $ln_k$ координатами точки $x$. Значит,

$$\xi_j = \xi_{j+n_k} = \xi_{j+2n_k} = \xi_{j+3n_k} = \ldots = \xi_{j+ln_k},$$

а это означает, что числа $j, j+n_k, j+2n_k, j+3n_k, \ldots, j+ln_k$ (как раз и образующие отрезок арифметической прогрессии длины $l$) лежат в одном и том же $A_i$.

Теорема Семереди утверждает, что если подмножество $A \subset \mathbb{Z}_+$ имеет «положительную верхнюю плотность», т. е. если при некоторых $a_n, b_n \in \mathbb{Z}_+$ с $b_n - a_n \to \infty$

$$\varlimsup_{n \to \infty} \frac{\text{число элементов } A \cap [a_n, b_n]}{b_n - a_n} > 0,$$

то в $A$ содержатся сколь угодно длинные отрезки геометрических прогрессий. Возникает мысль, что её можно доказать с помощью некоего метрического (в смысле меры) аналога того топологического утверждения, из которого так легко получается теорема Ван дер Вардена. Эта догадка верна, но доказательство соответствующего метрического утверждения[112] намного сложнее, чем топологического. Приходится использовать структурную теорему, «продвигаясь» шаг за шагом по соответствующей «последовательности» расширений (а так как она, вообще говоря, трансфинитная, то сказанное надо понимать cum grano salis).

**Кратности спектров.** Один из вариантов спектральной теоремы для унитарного оператора $U$ в сепарабельном гильбертовом пространстве $H$ состоит в том, что для $U$ имеется некоторая модель — некоторый довольно конкретно описываемый оператор $V$ в столь же конкретном гильбертовом пространстве $K$. $V$ «служит моделью» для $U$ в том смысле, что он сопряжён с $U$ посредством некоторого унитарного изоморфизма $W\colon H \to K$. В этой «модели» $K$ строится из конечного или бесконечного числа взаимно ортогональных «блоков» — пространств $L^2(\mathbb{S}^1, \nu)$, где $\mathbb{S}^1 := \{\lambda \in \mathbb{C};\ |\lambda| = 1\}$,

[112] Оно гласит: если $\varphi$ — эндоморфизм пространства Лебега, то для любого $l > 0$ и любого измеримого $A$ с $\mu(A) > 0$ имеется такое $i > 0$, что

$$\mu(A \cap \varphi^i A \cap \ldots \cap \varphi^{li} A) > 0.$$

$\nu$ — мера на $\mathbb{S}^1$. Меры $\nu$ для различных «блоков» либо ортогональны, либо совпадают; имеется $n_1$ «блоков», отвечающих мере $\nu_1$, $n_2 > n_1$ «блоков», отвечающих мере $\nu_2$, и т. д.; возможно наличие бесконечного числа блоков, отвечающих мере $\nu_\infty$. Наконец, оператор $V$ таков, что «блоки» инвариантны относительно него, и если $g$ принадлежит одному из блоков, скажем, $g \in L^2(\mathbb{S}^1, \nu)$, то $(Vg)(\lambda) = \lambda g(\lambda)$. Набор $\{n_1, n_2, \dots\}$ (который может быть конечным или бесконечным; он может кончаться символом $\infty$ или не содержать такового) называется набором спектральных кратностей для $U$. Он однозначно определён этим оператором и одинаков для всех операторов, унитарно сопряжённых с $U$ (тогда как каждая мера $\nu_i$ определена с точностью до перехода к эквивалентной мере).

Как только автоморфизму $\varphi$ пространства Лебега был сопоставлен унитарный оператор $U_\varphi$ (что сделал Б. Купмен в конце 20-х гг.), встал вопрос, какие операторы могут при этом получиться, по крайней мере в основном для «абстрактной» эргодической теории случае эргодического $\varphi$. Частью этого вопроса является вопрос, какими могут быть соответствующие наборы спектральных кратностей. Вопрос о соответствующих мерах (точнее, классах эквивалентных мер) в этой части не ставится, но ниже я буду различать случаи дискретного, непрерывного и смешанного спектра.

В случае дискретного спектра ответ почти очевиден: только $\{1\}$. Давно известны примеры непрерывного (даже лебеговского[113)]) спектра, для которых набор кратностей есть $\{\infty\}$. После войны постепенно набралось довольно много примеров с другими наборами спектральных кратностей, прежде всего, с набором $\{1\}$ («однократный (или простой) спектр»), а примерно в последние 10 лет был достигнут весьма значительный прогресс. Легко показать, что для смешанного спектра набор кратностей должен начинаться с 1 (это связано с его дискретной компонентой); оказывается, любой такой набор реализуется для некоторого $\varphi$. Это доказано различными методами в работе Я. Квятковски и М. Леманчика, цитируемой в [172], и в работе О. Н. Агеева [173]. В случае непрерывного спектра вопрос пока не выяснен до конца, но доказано, что может реализоваться любой набор, начинающийся с 1 [174]. Для наборов, начинающихся с $n_1 > 1$, пока не всё ясно; из различных примеров особого упоминания заслуживает следующий. Когда появились примеры с $\{1\}$, В. А. Рохлин поставил вопрос, имеются ли примеры $\varphi$ с «двукратным» непрерывным спектром (с набором кратностей $\{2\}$); недавно Агеев и В. В. Рыжиков получили положительный ответ. По этой тематике имеется новый обзор [172].

**Аппроксимации периодическими преобразованиями и джойнинги.** Самые известные идеи и понятия эргодической теории — спектральные и энтропийные. С простейшими прообразами первых мы встречаемся в

113) Напомню о проблеме Банаха из п. 1.3, *а*.

квазипериодических колебаниях, вторых — в случае последовательности независимых случайных испытаний. Сами по себе эти прообразы не относятся специфически к эргодической теории и были известны задолго до её возникновения. В результате далеко идущего развития соответствующих идей последние получили более широкие применения, причём не только в «абстрактной» (чисто-метрической), но и в «прикладной» эргодической теории, что и не удивительно, если вспомнить, откуда в конечном счёте возникли эти идеи. (В какой-то степени неожиданным для постороннего наблюдателя могло бы показаться использование идей вероятностного происхождения при исследовании ДС с гиперболическим поведением траекторий, но ведь и это можно считать реализацией пророческих замечаний Дж. К. Максвелла и А. Пуанкаре о том, что неустойчивость порождает стохастичность.)

Примерно за последние тридцать лет в эргодической теории сформировалось некое новое направление, новая система понятий и идей, происхождение которых является уже внутренним. Быть может, по этой причине данное направление пока что почти не имеет применений к гладким ДС; основной массив соответствующих примеров — это примеры комбинаторного характера (скажем, в пространствах последовательностей). Кажется, единственный известный пример гладкой ДС, подпадающей под данное направление — это орициклический поток. (Имеются также искусственные гладкие примеры.) Но ведь известно, что чаще всего новая идея находит широкие применения вне области своего зарождения через полвека, а пока прошла всего половина этого срока.

Я сформулирую только два понятия из этого направления (на самом деле их намного больше).

Пусть $\varphi$ — автоморфизм пространства Лебега $(X, \mu)$, и пусть измеримое подмножество $A \subset X$ таково, что множества

$$A,\ \varphi A,\ \ldots,\ \varphi^h A \tag{40}$$

попарно не пересекаются. Тогда говорят, что они образуют башню Рохлина (для $\varphi$) высоты $h$. Если $\mu\left(X \setminus \bigcup\limits_{i=0}^{h} \varphi^i A\right) < \varepsilon$, то я позволю себе говорить об $\varepsilon$-башне Рохлина. Чтобы пояснить смысл этого понятия, на минуту искусственно замкнём «цепочку» отображений

$$A \xrightarrow{\varphi} \varphi A \xrightarrow{\varphi} \ldots \xrightarrow{\varphi} \varphi^h A,$$

отобразив точку вида $\varphi^h x,\ x \in A$ в $x$. Получится периодическое преобразование $\psi$ пространства $\bigcup\limits_{i=0}^{h} \varphi^i A$, совпадающее с $\varphi$ по крайней мере на множестве $\bigcup\limits_{i=0}^{h-1} \varphi^i A$, мера которого $> 1 - (\varepsilon + 1/n)$. Если $h$ велико, а $\varepsilon$ мало, то $\psi$ как бы аппроксимирует $\varphi$ с большой точностью.

Оказывается, практически для любого $\varphi$, за некоторыми тривиальными исключениями, существуют $\varepsilon$-башни Рохлина высоты $h$ со сколь угодно большими $h$ и малыми $\varepsilon$. В. А. Рохлин использовал этот факт, доказывая, что множество слабо перемешивающих $\varphi$ «массивно» в пространстве всех $\varphi$ с равномерной (и тем более со слабой) топологией. В конце 60-х гг. А. Б. Каток, В. И. Оселедец и А. М. Стёпин использовали аппроксимационные соображения для изучения некоторых индивидуальных $\varphi$ или построения $\varphi$ с определёнными свойствами. Их формулировки несколько отличались от непосредственного использования башен Рохлина и включали условие количественного характера о скорости аппроксимации периодическими преобразованиями. (См. статью «Аппроксимация периодическими преобразованиями» в «Математической энциклопедии» или [166]. Если бы никаких условий не накладывалось, то $\varphi$ могло бы быть практически любым, а тогда нельзя было бы сказать ничего специального о его свойствах.) Затем Д. Орнстейн предложил модификацию той же идеи, в которой последнее требование заменено одним условием качественного характера. Условие Орнстейна таково: для любого $\varepsilon > 0$ и любого измеримого $A$ существует такая $\varepsilon$-башня Рохлина (40), что множество $A$ можно с точностью до $\varepsilon$ аппроксимировать объединением $A'$ некоторых из множеств (40), а именно, мера симметрической разности $\mu(A \triangle A') < \varepsilon$. При выполнении этого условия говорят, что $\varphi$ (и ДС $\{\varphi^n\}$) имеет ранг 1.

Автоморфизмы $\varphi$ ранга 1 имеют ряд общих свойств. Так, все они являются $LB$-автоморфизмами[114] и имеют простой спектр; автоморфизм $\psi$, коммутирующий с $\varphi$, является слабым пределом некоторой последовательности $\varphi^{n_i}$, откуда легко вывести, что централизатор $\varphi$ либо сводится к $\{\varphi^n\}$ (оказывается, это заведомо так, если $\varphi$ перемешивает), либо несчётен (в последнем случае имеется классификация таких $\varphi$). В то же время во многих отношениях свойства ДС ранга 1 весьма разнообразны (что отчасти уже отразилось в предыдущей фразе). Например, эргодический автоморфизм с дискретным спектром — ранга 1, но обратное не обязательно. Автоморфизм $\varphi$ ранга 1 может иметь квадратный корень[115] (даже континуум попарно неизоморфных корней), а может и не иметь такового, причём при этом может случиться, что $\varphi^2$ имеет корни всех степеней. Два автоморфизма ранга 1 могут быть слабо изоморфны (т. е. каждый из них изоморфен фактору другого), не будучи изоморфными. Многие неожиданные примеры такого рода (некоторые из них противоречат предположениям, которые ранее могли казаться

---

[114] $LB$ — сокращение от loosely Bernoulli. Первоначально такие автоморфизмы назывались также «стандартными», но теперь это название оставлено. О них см. параграф об эквивалентности ДС в смысле Какутани в [175].

[115] Т. е. существует такой автоморфизм $\psi$, что $\psi^2 = \varphi$. Аналогично понимаются и корни других степеней.

естественными) представляют интерес просто как примеры автоморфизмов пространства Лебега, независимо от того, что при этом построение примера доставляет нам автоморфизм ранга 1. Надо сказать, что обычно при построениях непосредственно используется не приведённое выше определение ДС ранга 1, а эквивалентные ему другие, более конструктивные определения. Они являются более длинными, но в них явно фигурируют параметры конструкций, доставляющих всевозможные $\varphi$ ранга 1. Успех построения различных примеров связан с тем, что удаётся проконтролировать влияние этих параметров на свойства строящегося $\varphi$.

Раз имеются автоморфизмы ранга 1, то должны быть и автоморфизмы других рангов. Идея состоит в том, что такие $\varphi$ хорошо аппроксимируются в терминах нескольких башен Рохлина.

Подробно о ДС конечного ранга см. в [176]. Отмечу только три факта. Для $k \in \mathbb{N}$

$$\text{ранг}\,\varphi^k = k \cdot \text{ранг}\,\varphi.$$

Элементы набора спектральных кратностей не превосходят ранга, но даже простота спектра не гарантирует конечности ранга. Для автоморфизмов конечного ранга доказана упоминавшаяся в п. 1.3 гипотеза Рохлина, что перемешивание влечёт за собой перемешивание всех степеней.

Второе понятие, на котором я остановлюсь — это так называемые джойнинги, которые восходят к Г. Фюрстенбергу и Д. Орнстейну, а систематическое использование которых началось по инициативе Д. Рудольфа.

Джойнинг автоморфизма $\varphi$ пространства с мерой $(X, \mu)$ — это такая инвариантная нормированная эргодическая меры $\nu$ в декартовой степени

$$\varphi^{\times n}\colon X^n \to X^n, \quad (x_1, \ldots, x_n) \mapsto (\varphi x_1, \ldots, \varphi x_n),$$

которая проектируется в $\mu$ при естественном проектировании $X^n$ на любой из сомножителей $X$. (Следует заметить, что здесь, строго говоря, приходится отходить от ортодоксальной концепции, согласно которой в чисто метрической теории всё происходит в пространстве Лебега. Ведь в теории пространств Лебега $X^n$ рассматривается только с мерой

$$\mu^{\times n} = \underbrace{\mu \times \ldots \times \mu}_{n\ \text{раз}},$$

которая является одним из джойнингов, но прочие джойнинги слабо перемешивающего $\varphi$ сосредоточены на множествах $\mu^{\times n}$-меры нуль, которыми с ортодоксальной точки зрения надо пренебрегать. Приходится временно реализовывать пространство $(X, \mu)$ не как пространство Лебега, а как пространство иного типа, обычно как «стандартное пространство» Макки, а потом проверять, что результат (информация о джойнингах) не зависит от выбора конкретной реализации.)

«Джойнинговые» различия между ДС состоят в том, что у одних ДС джойнингов «мало», а у других «много». ДС $\{\varphi^n\}$ всегда имеет джойнинги, являющиеся прямыми произведениями мер вида

$$A_1 \times \ldots \times A_k \mapsto \mu(\psi_1 A_1 \cap \ldots \cap \psi_k A_k),$$

где $\psi_i$ — коммутирующие с $\varphi$ автоморфизмы. Если других джойнингов нет (и в этом смысле их «мало»), то автоморфизм $\varphi$ называется простым. Как доказала М. Ратнер, для орициклического потока $\{\varphi_t\}$ на замкнутой поверхности постоянной отрицательной кривизны (и в некоторых других случаях) все $\varphi_t$ с $t \neq 0$ являются простыми. «Особенно мало» джойнингов, если $\varphi$ прост и его централизатор сводится к $\{\varphi^n\}$; в этом случае говорят, что $\varphi$ имеет минимальные самоприсоединения.

Само по себе понятие джойнинга не имеет отношения к аппроксимации периодическими преобразованиями. Однако в настоящее время найти джойнинги или хотя бы получить о них сколько-либо существенную информацию удаётся обычно в тех случаях, когда $\varphi$ допускает быструю (в том или ином смысле) аппроксимацию. Работа Ратнер в этом отношении является исключением, равно как и полученное Б. Остом с помощью джойнингов доказательство гипотезы Рохлина для $\varphi$ с сингулярным спектром.

Джойнингам посвящён обзор [177], который дополняет статья [178].

## Литература

[1] Ж.-К. Йоккоз. Недавнее развитие динамики // Международный конгресс математиков в Цюрихе, 1994 г. М.: Мир, 1999. С. 349—380.

[2] А. Б. Каток, Б. Хасселблат. Введение в современную теорию динамических систем. М.: Факториал, 1999.

[3] J. Moser. Dynamical systems—past and present // Proc. Internat. Congr. Math., Berlin 1998. Vol. I: Plenary lectures and ceremonies. Bielefeld, Germany: Univ. Bielefled, 1998. P. 381—402.

[4] И. П. Корнфельд, Я. Г. Синай, С. В. Фомин. Эргодическая теория. М.: Наука, 1980.

[5] В. И. Арнольд, А. Б. Гивенталь. Симплектическая геометрия // Итоги науки и техн. Соврем. пробл. матем. Фунд. напр. **4**: Динамические системы — 4. М.: ВИНИТИ, 1985. С. 5—139.

[6] M. Gromov. Pseudo-holomorphic curves on symplectic manifolds // Invent. Math. **82** (1985), №2. P. 307—347.

[7] H. Hofer, E. Zehnder. Symplectic invariants and Hamiltonian dynamics. Basel: Birkhäuser, 1994.

[8] M. Audin, J. Lafontaine (ed.). Holomorphic curves in symplectic geometry. Basel: Birkhäuser, 1994.

[9] M. Bialy, L. Polterovich. Hamiltonian diffeomorphisms and Lagrangian distributions // Geom. Funct. Anal. **2** (1992), №2. P. 173—210.

[10] M. Bialy, L. Polterovich. Invariant tori and symplectic topology // Sinai's Moscow seminar on dynamical systems. Amer. Math. Soc. Transl., Ser. 2 **171**. Providence, RI: Amer. Math. Soc., 1996. P. 23—33.

[11] D. McDuff, D. Salamon. Introduction to symplectic topology. Oxford Mathematical Monographs. New York: Clarendon Press, 1995.

[12] В. И. Арнольд. Топологические проблемы теории распространения волн // Успехи матем. наук **51** (1996), №1. С. 3—50.

[13] W. Ballmann. Der Satz von Lyusternik und Schnirelmann // Beitrage zur Differentialgeometrie. Bonner Math. Schriften **102**. Bonn: Univ. Bonn, 1978. P. 1—25.

[14] И. А. Тайманов. Замкнутые экстремали на двумерных многообразиях // Успехи матем. наук **47** (1992), №2. С. 143—185.

[15] V. Bangert. On the existence of closed geodesics on two-spheres // Internat. J. Math. **4** (1993), №1. P. 1—10.

[16] J. Franks. Geodesics on $S^2$ and periodic points of annulus homeomorphisms // Invent. Math. **108** (1992), №2. P. 403–418.

[17] J. Franks. Rotation vectors and fixed points of area preserving surface diffeomorphisms. // Trans. Amer. Math. Soc. **348** (1996), №7. P. 2637–2662.

[18] Sh. Matsumoto. Arnold conjecture for surface homeomorphisms. Препринт.

[19] N. Hingston. On the growth of the number of closed geodesics on the two-sphere // Internat. Math. Research Notices **9** (1993). P. 253–262.

[20] А. В. Болсинов, А. Т. Фоменко. Введение в топологию интегрируемых гамильтоновых систем. М.: Наука, 1997.

[21] Х. О. Пейтген, П. Х. Рихтер. Красота фракталов. М.: Мир, 1993.

[22] М. Ю. Любич. Динамика рациональных преобразований: топологическая картина // Успехи матем. наук **41** (1986), №4. С. 35–95.

[23] А. Э. Ерёменко, М. Ю. Любич. Динамика аналитических преобразований // Алгебра и анализ **1** (1989), №3. С. 1–70.

[24] L. Carleson, T. W. Gamelin. Complex dynamics. New York: Springer, 1993.

[25] M. Lyubich. Dynamics of quadratic polynomials I–II // Acta Math. **178** (1997), №2. P. 185–297.

[26] W. de Melo, S. van Strien. One-dimensional dynamics. Berlin: Springer, 1993.

[27] М. Фейгенбаум. Универсальность в поведении нелинейных систем // Успехи физ. наук **141** (1983), №2. С. 343–374.

[28] Ф. Гринлиф. Инвариантные средние на топологических группах. М.: Мир, 1979.

[29] U. Krengel. Ergodic theorems. De Gruyter studies in Math. **6**. Berlin—New York: De Gruyter, 1985.

[30] А. А. Темпельман. Эргодические теоремы на группах. Вильнюс: Мокслас, 1986.

[31] Р. И. Григорчук. Индивидуальная эргодическая теорема для действий свободных групп // XII школа по теории операторов в функциональных пространствах. Тезисы докладов (Тамбов, 14–20 сентября 1987 г.). Часть 1. Тамбов: Тамбовский госпединститут, 1987. С. 57.

[32] A. Nevo, E. M. Stein. A generalization of Birkhoff's pointwise ergodic theorem // Acta Math. **173** (1994), №1. P. 135–154.

[33] A. Nevo, E. M. Stein. Analog of Wiener's ergodic theorems for semi-simple Lie groups I // Ann. Math. **145** (1997), №3. P. 565–595.

[34] G. M. Margulis, A. Nevo, E. M. Stein. Analog of Wiener's ergodic theorems for semi-simple Lie groups II. Препринт.

[35] В. И. Арнольд, А. Л. Крылов. Равномерное распределение точек на сфере и некоторые эргодические свойства решений линейных обыкновенных дифференциальных уравнений // Докл. АН СССР **148** (1963), №1. С. 9–12.

[36] Д. А. Каждан. Равномерное распределение на плоскости // Труды Моск. матем. о-ва **14** (1965). С. 299–305.

[37] Y. Guivarc'h. Généralisation d'un théorème de von Neumann. C. R. Acad. Sci. Paris, Sér A **268** (1969), №18. P. 1020–1023.

[38] О. Н. Агеев. Динамические системы с чётнократной лебеговской компонентой в спектре // Матем. сб. **136** (1988), №3 (7). С. 307–319.

[39] Д. В. Аносов. О вкладе Н. Н. Боголюбова в теорию динамических систем // Успехи матем. наук **49** (1994), №5. С. 5–20.

[40] D. Ornstein, B. Weiss. Ergodic theory of amenable group actions I. The Rohlin lemma // Bull. Amer. Math. Soc. **2** (1980), №1. P. 161–164.

[41] D. Ornstein, B. Weiss. Entropy and isomorphism theorems for actions of amenable groups // J. d'Analyse Math. **48** (1987). P. 1–141.

[42] А. М. Стёпин, А. Т. Таги-заде. Вариационная характеризация топологического давления аменабельных групп преобразований // Докл. АН СССР **254** (1980), №3. С. 545–549.

[43] Я. Г. Синай. Гиббсовские меры в эргодической теории // Успехи матем. наук **27** (1972), №4. С. 21–64.

[44] Р. Боуэн. Методы символической динамики. М.: Мир, 1979.

[45] А. М. Вершик, И. П. Корнфельд. Периодические аппроксимации и их приложения. Эргодические теоремы, спектральная и энтропийная теория для действий общих групп // Итоги науки и техн. Соврем. пробл. матем. Фунд. напр. **2**: Динамические системы — 2. М.: ВИНИТИ, 1985. С. 70–89.

[46] В. И. Арнольд. Теория катастроф. М.: Наука, 1990.

[47] В. И. Арнольд. Теория катастроф // Итоги науки и техн. Соврем. пробл. матем. Фунд. напр. **5**: Динамические системы — 5. М.: ВИНИТИ, 1986. С. 219–277.

[48] А. В. Чернавский. Применения теории катастроф в психологии // Сб. «Число и мысль», №2. М.: Знание, 1979.

[49] Т. Постон, И. Стюарт. Теория катастроф и её приложения. М.: Мир, 1980.

[50] В. И. Арнольд. Лекции о бифуркациях и версальных семействах // Успехи матем. наук **27** (1972), №5. С. 119–184.

[51] В. И. Арнольд. Дополнительные главы теории обыкновенных дифференциальных уравнений. М.: Наука, 1978.

[52] В. И. Арнольд, В. С. Афраймович, Ю. С. Ильяшенко, Л. П. Шильников. Теория бифуркаций // Итоги науки и техн. Соврем. пробл. матем. Фунд. напр. **5**: Динамические системы — 5. М.: ВИНИТИ, 1986. С. 5–219.

[53] С. И. Трифонов. Цикличность элементарных полициклов типичных гладких векторных полей // Дифференциальные уравнения с вещественным и комплексным временем / Под ред. Ю. С. Ильяшенко. Труды МИАН **213**. М.: Наука, 1997. С. 152–212.

[54] Н. К. Гаврилов, Л. П. Шильников. О трёхмерных динамических системах, близких к системам с негрубой гомоклинической кривой I, II // Матем. сб. **88** (1972), №4. С. 475–492; **90** (1973) №1. С. 139–156.

[55] С. В. Гонченко. Об устойчивых периодических движениях систем, близких к системе с негрубой гомоклинической кривой // Мат. заметки **33** (1983), №5. С. 745–755.

[56] С. В. Гонченко, Д. В. Тураев, Л. П. Шильников. Динамические явления в многомерных системах с негрубой гомоклинической кривой Пуанкаре // Докл. РАН **330** (1993), №2. С. 144–147.

[57] S. V. Gonchenko, L. P. Shilnikov, D. V. Turaev. Dynamical phenomena in systems with structurally unstable Poincare homoclinic orbits // Chaos **6** (1996), №1. P. 15–31.

[58] Ж.-К. Йоккоз. Квадратичные многочлены и аттрактор Эно // Труды семинара Н. Бурбаки за 1991 г. / Под ред. В. А. Васильева и М. И. Монастырского. М.: Мир, 1998. С. 121–140.

[59] Sh. Newhouse. Non-density of axiom A(a) on $S^2$ // Proc. A.M.S. symp. pure math. **14** (1970). P. 191–202.

[60] Sh. Newhouse. Diffeomorphisms with infinitely many sinks // Topology **13** (1974), №1. P. 9–18.

[61] Sh. Newhouse. The abundance of wild hyperbolic sets and nonsmooth stable sets for diffeomorphisms // Publ. Math. IHES **50** (1979). P. 101–151.

[62] C. Robinson. Bifurcations to infinitely many sinks // Comm. Math. Phys. **90** (1983), №3. P. 433–459.

[63] J. Palis, M. Viana. High dimension diffeomorphisms displaying infinitely many periodic attractors // Ann. Math. **140** (1994), №1. P. 207–250.

[64] С. В. Гонченко, Д. В. Тураев, Л. П. Шильников. О существовании областей Ньюхауса вблизи систем с негрубой гомоклинической кривой Пуанкаре (многомерный случай) // Докл. РАН **329** (1993), №4. С. 404–407.

[65] N. Romero. Persistence of homoclinic tangencies in higher dimensions // Ergod. Theory and Dyn. Dystems **15** (1995), №4. P. 735–757.

[66] J. Palis, F. Takens. Hyperbolicity and sensitive chaotic dynamics at homoclinic bifurcations. Cambridge: Cambridge Univ. Press, 1993.

[67] Л. П. Шильников, Д. В. Тураев. О катастрофах голубого неба // Докл. РАН **342** (1995), №5. С. 596–599.

[68] Д. В. Тураев, Л. П. Шильников. Пример дикого странного аттрактора // Матем. сб. **189** (1998), №2. С. 137–160.

[69] S. V. Gonchenko, L. P. Shilnikov, D. V. Turaev. On models with non-rough Poincare homoclinic curves // Physica D **62** (1993), №1-4. P. 1–14.

[70] V. Kaloshin. Generic diffeomorphisms with superexponential growth of number of periodic orbits. SUNY Stony Brook Inst. for Math. Sc. Препринт №1998-8.

[71] С. В. Гонченко, Д. В. Тураев, Л. П. Шильников. Об областях Ньюхауса двумерных диффеоморфизмов, близких к диффеоморфизму с негрубым гомоклиническим контуром // Динамические системы и смежные вопросы. Труды МИАН **216**. М.: Наука, 1997. С. 76–125.

[72] С. В. Гонченко, Л. П. Шильников. О двумерных аналитических сохраняющих площадь диффеоморфизмах со счётным множеством эллиптических устойчивых периодических точек // Регулярная и хаотическая динамика **2** (1997), №3/4. С. 106–123.

[73] Л. П. Шильников. Homoclinic orbits: Since Poincare till today. Berlin: Weierstrass-Inst. für angew. Analysis und Stochastik, 2000. Препринт №571.

[74] Л. П. Шильников, В. С. Афраймович, В. В. Быков. О притягивающих негрубых предельных множествах типа аттрактора Лоренца // Труды Моск. матем. о-ва **44** (1982). С. 150–212.

[75] Р. В. Плыкин, Е. А. Сатаев, С. В. Шлячков. Странные аттракторы // Итоги науки и техн. Соврем. пробл. матем. Фунд. напр. **66**: Динамические системы — 9. М.: ВИНИТИ, 1991. С. 100–148.

[76] Д. В. Аносов. Грубые системы // Топология, обыкновенные дифференциальные уравнения, динамические системы / Под ред. Е. Ф. Мищенко. Тр. МИАН **169**. М.: Наука, 1985. С. 59–93.

[77] Д. В. Аносов, В. В. Солодов. Гиперболические множества // Итоги науки и техн. Соврем. пробл. матем. Фунд. напр. **66**: Динамические системы — 9. М.: ВИНИТИ, 1991. С. 12–9.

[78] Д. В. Аносов, И. У. Бронштейн, С. Х. Арансон, В. З. Гринес. Гладкие динамические системы // Итоги науки и техн. Соврем. пробл. матем. Фунд. напр. **1**: Динамические системы — 1. М.: ВИНИТИ, 1985. С. 151–242.

[79] Ch. C. Pugh, R. C. Robinson. The $C^1$ closing lemma, including Hamiltonians // Ergod. Theory and Dyn. Systems **3** (1983), №2. P. 261–313.

[80] M. M. Peixoto. Acceptance speech for the TWAS 1986 award in mathematics // The future of science in China and the third world. Proc. of the second general conf. organized by the Third World Ac. Sci / Eds. A. M. Faruqui, M. H. A. Hassan. Singapore: World Scientific, 1989. P. 600–614.

[81] R. Mañé. On the creation of homoclinic points // Publ. Math. IHES **66** (1988). P. 139–159.

[82] R. Mañé. A proof of the $C^1$ stability conjecture // Publ. Math. IHES **66** (1988). P. 161–210.

[83] J. Palis. On the $C^1$ $\Omega$-stability conjecture // Publ. Math. IHES **66** (1988). P. 211–215.

[84] Sh. Hayashi. Connecting invariant manifolds and the solution of the $C^1$ stability and $\Omega$-stability conjectures for flows // Ann. Math. **145** (1997), №1. P. 81–137; *Correction*. **150** (1999), №1. P. 353–356.

[85] Ch. C. Pugh. Against the $C^2$-closing lemma // J. of Diff. Equations **17** (1975), №2. P. 435–443.

[86] O. S. Kozlovski. Structural stability in one-dimensional dynamics. Univ. of Amsterdam, 1997. Препринт.

[87] В. И. Арнольд, Ю. С. Ильяшенко. Обыкновенные дифференциальные уравнения // Итоги науки и техн. Соврем. пробл. матем. Фунд. напр. **1**: Динамические системы — 1. М.: ВИНИТИ, 1985. С. 7–149.

[88] А. А. Болибрух. Проблема Римана—Гильберта на комплексной проективной прямой // Матем. заметки **46** (1989), №3. С. 118–120.

[89] D. V. Anosov, A. A. Bolibruch. The Riemann—Hilbert problem. Wiesbaden, Braunschweig: Vieweg, 1994.

[90] Дифференциальные уравнения с вещественным и комплексным временем / Под ред. Ю. С. Ильяшенко. Тр. МИАН **213**. М., 1997.

[91] Ю. С. Ильяшенко. Мемуар Дюлака «О предельных циклах» и смежные вопросы локальной теории дифференциальных уравнений // Успехи матем. наук **40** (1985), №6. С. 41–78.

[92] Yu. S. Ilyashenko. Finiteness theorems for limit cycles. Providence, R.I.: Amer. Math. Soc., 1991.

[93] J. Ecalle. Fonctions analysable et solution constructive du probleme de Dulac. Paris: Hermann, 1992.

[94] Ж.-К. Йоккоз. Ненакопление предельных циклов // Труды семинара Н. Бурбаки за 1988 г. / Под ред. А. Н. Варченко. М: Мир, 1990.

[95] W. Balser. From divergent power series to analytic functions. Berlin: Springer, 1994.

[96] И. Г. Петровский, Е. М. Ландис. О числе предельных циклов уравнения $\frac{dy}{dx}=\frac{P(x,y)}{Q(x,y)}$, где $P$ и $Q$ — многочлены 2-й степени // Матем. сб. **37** (1955), №2. С. 209–250.

[97] Е. М. Ландис, И. Г. Петровский. О числе предельных циклов уравнения $\frac{dy}{dx}=\frac{P(x,y)}{Q(x,y)}$, где $P$ и $Q$ — полиномы // Матем. сб. **43** (1957), №2. С. 149–168.

[98] И. Г. Петровский, Е. М. Ландис. Поправки к статьям «О числе предельных циклов уравнения $\frac{dy}{dx}=\frac{P(x,y)}{Q(x,y)}$, где $P$ и $Q$ — многочлены 2-й степени» и «О числе предельных циклов уравнения $\frac{dy}{dx}=\frac{P(x,y)}{Q(x,y)}$, где $P$ и $Q$ — полиномы» // Матем. сб. **48** (1959), №2. С. 253–255.

[99] В. В. Амелькин, Н. А. Лукашевич, А. П. Садовский. Нелинейные колебания в системах второго порядка. Минск: Изд-во БГУ, 1982.

[100] Е. М. Ландис, И. Г. Петровский. Письмо в редакцию // Матем. сб. **73** (1967), №1. С. 160.

[101] A. Kotova, V. Stantso. On few-parameter generic families of vector fields on the two-dimensional sphere // Concerning Hilbert 16th problem / Eds. Yu. Ilyashenko, S. Yakovenko. Amer. Math. Soc. Transl., Ser. 2 **165**. Providence, RI: Amer. Math. Soc., 1995. P. 155–201.

[102] A. Starkov. Fuchsian groups from the dynamical viewpoint // J. of Dynam. and Control Syst. **1** (1995), №3. P. 427–445.

[103] М. Ратнер. Взаимодействие эргодической теории, групп Ли и теории чисел // Международный конгресс математиков в Цюрихе, 1994 г. М.: Мир, 1999. С. 222–258.

[104] А. В. Сафонов, А. Н. Старков, А. М. Стёпин. Динамические системы с транзитивной группой симметрий. Геометрические и статистические свойства // Итоги науки и техн. Соврем. пробл. матем. Фунд. напр. **66**: Динамические системы — 9. М.: ВИНИТИ, 1991. С. 187–242.

[105] А. Н. Старков. Новый прогресс в теории однородных потоков // Успехи матем. наук **52** (1997), №4. С. 87–192.

[106] G. Margulis. Oppenheim conjecture // Fields Medalists' Lectures. Rover Edge, N.Y.: World Scientific Publ., 1997.

[107] А. Н. Старков. Динамические системы на однородных пространствах. М: Фазис, 1999.

[108] K. Kuperberg. A smooth counterexample to the Seifert conjecture // Ann. Math. **140** (1994), №3. P. 723–732.

[109] G. Kuperberg, K. Kuperberg. Generalized counterexamples to the Seifert conjecture // Ann. Math. **144** (1996), №2. P. 239–268.

[110] И. Тамура. Топология слоений. М: Мир, 1979.

[111] H. Hofer. Pseudoholomorphic curves in symplectization with applications to the Weinstein conjecture in dimension three // Invent. Math. **114** (1993), №3. P. 515–563.

[112] S. R. Fenley. The structure of branching in Anosov flows of 3-manifolds // Comment. Math. Helv. **73** (1997). P. 259–297.

[113] П. Пансю. Геодезический поток на римановых многообразиях отрицательной кривизны // Труды семинара Н. Бурбаки за 1991 г. / Под ред. В. А. Васильева и М. И. Монастырского. М.: Мир, 1998. С. 226–250.

[114] G. Knieper. The uniqueness of the measure of maximal entropy for geodesic flows on rank 1 manifolds // Ann. Math. **148** (1998). P. 291–314.

[115] G. Besson, G. Courtois, S. Gallot. Les variétés hyperboliques sont des minima locaux de l'entropie topologique // Invent. Math. **117** (1994), №3. P. 403–445.

[116] L. Barreira, Ya. Pesin, J. Schmeling. Dimension and product structure of hyperbolic measures // Ann. Math. **149** (1999), №3. P. 755–783.

[117] М. Л. Бялый, Л. В. Полтерович. Геодезические потоки на двумерном торе и фазовые переходы «соизмеримость—несоизмеримость» // Функц. анализ и его прилож. **20** (1986), №4. С. 9–16.

[118] Л. В. Полтерович. Геодезические на двумерном торе с двумя числами вращения // Изв. АН СССР. Сер. матем. **52** (1988), №4. С. 774–787.

[119] V. Bangert. Mather sets for twist maps and geodesics on tori // Dynamics Reported **1** / Eds. U. Kirchgraber, H. O. Walter. Chichester—Stuttgart: John Wiley and B. G. Teubner, 1988. P. 1–56.

[120] A. B. Katok. Some remarks on Birkhoff and Mather twist map theorems // Ergod. Theory and Dyn. Systems **2** (1982), №2. P. 185–194.

[121] A. Fathi. Theoreme KAM faible et theorie de Mather sur les systemes lagrangiens // C. R. Acad. Sci. Paris, Ser. I Math. **324** (1997), №9. P. 1043–1046.

[122] R.Mañé. Lagrangian flows: the dynamics of globally minimizing orbits // Bol. Soc. Brasil. Mat. (N.S.) **28** (1997), №2. P. 141–153.

[123] G. Contreras, R. Iturriaga, G. P. Paternain, M. Paternain. Lagrangian graphs, minimizing measures and Mane's critical values // Geometric and Functional Analysis **8** (1998), №5. P. 788–809.

[124] А. Д. Брюно. Степенная геометрия в алгебраических и дифференциальных уравнениях. М.: Наука—Физматлит, 1998.

[125] Е. Ф. Мищенко, Н. Х. Розов. Дифференциальные уравнения с малым параметром при старшей производной и релаксационные колебания. М.: Наука, 1975.

[126] Е. Ф. Мищенко, Ю. С. Колесов, А. Ю. Колесов, Н. Х. Розов, Периодические движения и бифуркационные процессы в сингулярно возмущённых системах. М.: Наука—Физматлит, 1995.

[127] П. Картье. Сингулярные возмущения обыкновенных дифференциальных уравнений и нестандартный анализ // Успехи матем. наук **39** (1984), №2. С. 57–76.

[128] А. К. Звонкин, М. А. Шубин. Нестандартный анализ и сингулярные возмущения обыкновенных дифференциальных уравнений // Успехи матем. наук **39** (1984), №2. С. 77–127.

[129] W. Eckhaus. Relaxation oscillations including a standard chase on french ducks // Asymptotic Analysis II. Lect. Notes in Math. **985**. Berlin: Springer, 1983. P. 449–494.

[130] Dynamic bifurcations / Ed. E. Benoit. Lect. Notes in Math. **1493**. Berlin: Springer, 1991.

[131] А. И. Нейштадт. О затягивании потери устойчивости при динамических бифуркациях I // Дифференц. уравнения **23** (1987), №12. С. 2060–2067.

[132] А. И. Нейштадт. О затягивании потери устойчивости при динамических бифуркациях II // Дифференц. уравнения **24** (1988), №2. С. 226–233.

[133] A. I. Neishtadt, C. Simo, D. V. Treschev. On stability loss delay for a periodic trajectory // Nonlinear dynamical systems and chaos / Eds. H. Broer et al. Progress in Nonlinear Differential Equations Appl. **19**. Basel: Birkhäuser, 1996. P. 253–278.

[134] В. И. Бахтин. Об усреднении в многочастотных системах // Функц. анализ и его прилож. **20** (1986), №2. С. 1–7.

[135] В. И. Арнольд, В. В. Козлов, А. И. Нейштадт. Математические аспекты классической и небесной механики // Итоги науки и техн. Соврем. пробл. матем. Фунд. напр. **3**: Динамические системы — 3. М.: ВИНИТИ, 1985.

[136] Д. В. Трещёв. Введение в теорию возмущений гамильтоновых систем. М.: Фазис, 1998.

[137] C. Simo. Averaging under fast quasiperiodic forcing // Hamiltonian mechanics, integrability and chaotic behaviour / Ed. J. Seimenis. NATO ASI Inst., Ser. B. Phys. **331**. N.Y.: Plenum Press, 1994. P. 11–34.

[138] В. Ф. Лазуткин. Аналитические интегралы полустандартного отображения и расщепление сепаратрис // Алгебра и анализ **1** (1989), №2. С. 116–131.

[139] V. F. Lazutkin. An analitic integral along the separatrix of the semistandard map: existence and an exponential estimate for the distance betwen the stable and unstable separatrices // Алгебра и анализ **4** (1992), №4. С. 110–142.

[140] A. Delshams, V. Gelfreich, A. Jorba, T. M. Seara. Exponentially small splitting of separatrices under fast quasiperiodic forcing // Comm. Math. Phys. **189** (1997), №1. P. 35–71.

[141] J. A. Ellison, M. Kummer, A. W. Saenz. Transcendentally small transversality in the rapidly forced pendulum // J. Dyn. Diff. Equations **5** (1993). P. 241–277.

[142] J. Pöschel. Nekhoroshev estimates for quasi-convex Hamiltonian systems // Math. Z. **213** (1993), №2. P. 187–216.

[143] П. Лошак. Каноническая теория возмущений: подход, основанный на совместных приближениях // Успехи матем. наук **47** (1992), №6. P. 59–140.

[144] P. Lochak, A. I. Neishtadt. Estimates of stability time for nearly integrable systems with a quasiconvex Hamiltonian // Chaos **2** (1992), №4. P. 495–499.

[145] А. Б. Каток. Гипотеза об энтропии // Гладкие динамические системы / Под ред. Д. В. Аносова. М.: Мир, 1977. С. 181–203.

[146] Y. Yomdin. Volume growth and entropy // Israel J. Math. **57** (1987), №3. P. 285–300.

[147] Y. Yomdin. $C^k$-resolution of semialgebrai mappings. Addendum to "Volume growth and entropy" // Israel J. Math. **57** (1987), №3. P. 301–317.

[148] М. Громов. Энтропия, гомологии и полуалгебраическая геометрия // Математический анализ и геометрия. Избранные труды семинара Н. Бурбаки / Под ред. А. Н. Варченко. М.: Мир, 1990. С. 207–223.

[149] М. Оден. Вращающиеся волчки: курс интегрируемых систем. Ижевск: Удмурдский университет, 1999.

[150] Ю. Мозер. Интегрируемые гамильтоновы системы и спектральная теория. Ижевск: Удмурдский университет, 1999.

[151] В. В. Козлов, Интегрируемость и неинтегрируемость в гамильтоновой динамике, Успехи матем. наук **38** (1983), №1, 3–67.

[152] В. В. Козлов, Симметрия, топология и резонансы в гамильтоновой механике, Ижевск: Изд-во УдГУ, 1995.

[153] И. А. Тайманов. Топология римановых многообразий с интегрируемыми геодезическими потоками // Новые результаты в теории топологической классификации интегрируемых систем. Труды МИАН **205**. М.: Наука, 1994. С. 150–163.

[154] L. Butler. A new class of homogeneous manifolds with Liouville-integrable geodesic flows. Queen's Univ. at Kingstone. Math. preprint №1998–8.

[155] A. V. Bolsinov, I. A. Taimanov. Integrable geodesic flows on the suspension of toric automorphisms. Preprint Sfb 288 №426 (1999).

[156] Ch. Conley. Isolated invariant sets and the Morse index. Conf. board math. sci. Regional conf. ser. in math. **38**, Providence, RI: Amer. Math. Soc., 1978.

[157] K. Mischaikow. Conley index theory // Dynamical systems (Montecatini Terme, 1994). Lecture Notes in Math. **1609**. Berlin et al.: Springer, 1995. P. 119–207.

[158] J. N. Mather, R. McGehee. Solutions of the collinear four-body problem which become unbounded in finite time // Lect. Notes in Phys. **38**. Berlin et al.: Springer, 1975. P. 573–597.

[159] Zh. Xia. The existence of noncollision singularities in newtonian systems // Ann. Math. **135** (1992), №3. P. 411–468.

[160] M. Grayson, C. Pugh, M. Shub. Stably ergodic diffeomorphisms // Ann. Math. **140** (1994), №2. P. 295–329.

[161] C. Pugh, M. Shub. Stably ergodic dynamical systems and partial hyperbolicity // J. of Complexity **13** (1997) №1. P. 125–179.

[162] C. Pugh, M. Shub. Stable ergodicity and partial hyperbolicity // Internat. conf. on dynamical systems, Montevideo 1995, a tribute to R. Mañé / Eds. F. Ledrappier et al. Pitman Research Notes in Math. **362**. Longman, Harlow, 1996. P. 182–187.

[163] M. Shub, A. Wilkinson. Pathological foliations and removable zero exponents // Invent. Math. **139** (2000), №3. P. 495–508.

[164] R. Adler, B. Kitchens, M. Shub. Stably ergodic skew products // Discr. and Contin. Dynam. Systems **2** (1996), №3. P. 349–350.

[165] W. Parry, M. Pollicott. Stability of mixing for toral extensions of hyperbolic systems // Динамические системы и смежные вопросы. Труды МИАН **216**. М.: Наука, 1997. С. 354–363.

[166] А. Б. Каток, Я. Г. Синай, А. М. Стёпин. Теория динамических систем и общих групп преобразований с инвариантной мерой // Итоги науки и техн. Математический анализ **13**. М.: ВИНИТИ, 1975. С. 129–262.

[167] H. Furstenberg, Y. Katznelson, D. Ornstein. The ergodic theoretical proof of Szemerédy's theorem // Bull. Amer. Math. Soc. **7** (1982), №3. P. 527–552.

[168] H. Furstenberg. Recurrence in ergodic theory and combinatorial number theory. Princeton, N.J.: Princeton Univ. Press, 1981.

[169] R. J. Zimmer. Extensions of ergodic group actions // Illinois J. of Math. **20** (1976), №3. P. 373–409.

[170] R. J. Zimmer. Ergodic actions with generalized discrete spectrum // Illinois J. of Math. **20** (1976), №4. P. 555–588.

[171] H. Furstenberg. Poincáre recurrence and number theory // Bull. Amer. Math. Soc. **5** (1981), №3. P. 211–234.

[172] G. R. Goodson. A survey of recent results in the spectral theory of ergodic dynamical systems // J. of Dynam. and Control Syst. **5** (1999), №2. P. 173–226.

[173] О. Н. Агеев. Функция кратности спектра и геометрические представления перекладывания // Матем. сб. **190** (1999), №1. С. 3–28.

[174] О. Н. Агеев. О функции кратности спектра динамических систем // Матем. заметки **65** (1999), №4. С. 619–621.

[175] И. П. Корнфельдь, Я. Г. Синай. Энтропийная теория динамических систем // Итоги науки и техн. Соврем. пробл. матем. Фундам. направления **2**, Динамические системы — 2. М.: ВИНИТИ, 1985. С. 44–70.

[176] S. Ferenczi. Systems of finite rank // Colloq. Math. **73** (1997), №1. P. 35–65.

[177] J.-P. Thouvenot. Some properties and applications of joinings in ergodic theory // Proc. of the 1993 Alexandria Conference. Ergodic theory and its connections with harmonic analysis. Cambridge Univ. Press, 1995. P. 207–235.

[178] V. V. Ryzhikov. Around simple dynamical systems. Induced joinings and multiple mixing // J. of Dynam. and Control Syst. **3** (1997), №1. P. 111–127.

*А. А. Разборов*

# Основы теории сложности вычислений

*Лекция 23 апреля 1998 года*

Эта лекция рассчитана на тех, кто не знаком с теорией сложности вычислений. Поэтому здесь будет рассказано только об основах этой теории и самых первых её результатах. При этом мы постараемся передать основные идеи, которыми руководствуются исследователи в этой области науки.

Сюжетной канвой дальнейшего рассказа послужат истории, происходящие с некоторым персонажем — назовём его М (от слова «математик»). И начнём рассказ со следующей истории.

**1. Предыстория.** Однажды М сидел дома и пытался доказать некоторую (возможно важную, возможно нет) теорему $T$. Он пытался доказать эту теорему в течение недели, двух, месяца, ..., но ничего у него не получалось. В конце концов он не выдержал и задал вполне естественный вопрос:

*А можно ли в принципе доказать эту теорему?*

Вопрос был обращён неизвестно к кому, и, скорее всего, улетел бы в пространство, если бы мимо не проходил другой наш персонаж, которого мы обозначим буквой Л, от слова «логик».

Л услышал вопрос, зашёл в комнату и объяснил, что такими вопросами математики начали интересоваться примерно с начала 20-го века. В общем виде этот вопрос входит в знаменитую «программу Гильберта», посвящённую понятию математического доказательства.

Эта программа, в частности, включала три следующих пункта.

**Формализация понятия доказательства.** Перед тем как задавать вопросы о том, можно ли доказать то или иное утверждение, необходимо дать строгое математическое определение доказуемости.

**Полнота.** После формализации понятия доказательства нужно установить *полноту* построенной *формальной теории*. Это означает, что любое истинное утверждение $T$ должно быть доказуемым в нашей формализации. В частности (в силу некоторых исторических причин), самого Гильберта преимущественно интересовал вопрос о доказуемости утверждения о непротиворечивости математики, формализованного подходящим образом.

**Разрешимость.** Следующей целью программы было построение такого вычислительного устройства, которое по внешнему виду теоремы $T$, записанной в некотором формальном языке, определяло бы, является ли эта теорема доказуемой (предполагалось, что в силу пункта 2 программы это будет эквивалентно истинности).

Хорошо известно, что первый пункт программы был успешно выполнен. В настоящее время под словом «теорема» большинство математиков (хотя не все это признают) понимает то, что можно доказать в теории множеств Цермело—Френкеля.

Хуже дело обстояло с последующими двумя пунктами. Первым ударом, потрясшим в свое время всё математическое сообщество, был полученный в 30-е годы результат Курта Гёделя, утверждающий, что никакая достаточно сильная теория с задаваемым явным списком множеством аксиом не может быть полной. Более точно, если наша теория непротиворечива и в ней можно формализовать рассуждения о натуральных числах, то существует утверждение, которое нельзя ни доказать, ни опровергнуть, более того, в качестве примера можно взять интересовавшее Гильберта утверждение о непротиворечивости рассматриваемой теории.

Мы сегодня занимаемся вычислениями, поэтому для нас более интересна теорема Чёрча (1936). Он доказал, что не существует никакого алгоритма, который по утверждению автоматически проверял бы, является это утверждение доказуемым или нет.

Итак, наш математик М узнал от логика Л, что в полной общности ответить на вопрос о доказуемости математического утверждения невозможно.

Впрочем, к этому моменту у него уже начали появляться сомнения в том, что теорема $T$ истинна, и он решил заняться поиском контрпримера. М пошел в компьютерную комнату и написал программу $П$, проверяющую последовательно все слова до тех пор, пока она не найдет данные, на которых $T$ не верна. Он запустил свою программу и стал ждать окончания её работы.

Программа работала час, два, день, неделю, ... И снова у математика возник вопрос, на этот раз:

*Остановится ли П когда-нибудь?*

На этот раз математик уже знал, к кому обратиться с таким вопросом. Он нашёл логика и спросил существует ли способ узнать, прекратит ли данная программа своё выполнение или же будет работать до бесконечности. И снова Л дал вполне квалифицированный ответ. По теореме, доказанной Тьюрингом в 30-е годы, не существует алгоритма, определяющего, остановится ли когда-нибудь данная программа или нет, т. е. проблема остановки неразрешима.

**2. Отправная точка.** После этого наш математик отправился домой. Как это иногда бывает, его ребёнок, изучающий геометрию в седьмом или восьмом классе, попросил папу помочь сделать домашнее задание. В народе бытует мнение, что люди, занимающиеся математикой, умеют хорошо считать, решать квадратные уравнения и разные задачки из элементарной геометрии, что не всегда соответствует действительности.

Итак, наш математик начал решать задачу, предложенную своим чадом, и понял, что напрочь забыл геометрию, которую учил в школе. Он однако помнил, что любую задачу геометрии можно записать в координатах на языке действительных чисел.

И ему в голову пришёл вопрос, а существует ли универсальный алгоритм, решающий школьные задачи по геометрии. После разговора с логиком, он знал, что если в теории выразимы натуральные числа, то она неразрешима. Интуиция подсказывала, что теория, работающая с действительными числами, которых гораздо больше, должна быть уж тем более неразрешима. Всё же, для очистки совести, он решил позвонить логику и удостовериться в этом. Как ни странно, выяснилось противоположное. Классический результат Тарского, доказанный в 1948 году, утверждал существование алгоритма, проверяющего доказуемость утверждений элементарной геометрии.[1)]

Математик обрадовался. Ему уже начинало казаться, что от логики нет никакой пользы, а тут оказалось, что логика имеет весьма практичное применение в реальной жизни.

Итак, М попросил ребёнка подождать, а сам отправился к торговцам программным обеспечением. Там он обнаружил два компакт-диска, посвящённых решению задач по геометрии, которые назывались соответственно

Tarsky for Windows 95 и Collins for Windows 95.

Первый диск стоил 30 у.е., второй — 300 у.е. Естественно, М попытался понять, чем же вызвана такая разница в цене. Никаких объяснений он не получил, поэтому он купил более дешёвый диск Tarsky for Windows 95.

М пришёл домой, вставил CD в компьютер и решил проверить программу на задачке из школьного учебника. Однако стала повторяться всё та же история. Программа работала час, два, ... но не проявляла никакого желания выдать решение. Математик прервал выполнение программы и попросил её доказать какую-нибудь элементарную теорему, например о том, что сумма углов в треугольнике равна 180 градусам. Результат не сильно изменился, и программа продолжала думать час, второй, сутки, другие,... Тогда М позвонил логику и несколько повышенным тоном спросил, что же происходит. Л ответил, что это не его проблема. Тарский доказал теорему о существовании

---

[1)] И вообще любых утверждений о вещественных числах, использующих арифметические операции, элементарные логические связки (отрицание, конъюнкция, дизъюнкция, импликация) и кванторы по множеству всех вещественных чисел: «для всех (вещественных чисел)», «существует (вещественное число), такое что...»; совокупность таких утверждений называется алгеброй Тарского.

алгоритма, он совершенно уверен, что авторы этого диска запрограммировали алгоритм правильно, а что происходит дальше — не имеет никакого отношения ни к математике, ни к логике.

И это как раз то место, где начинается теория сложности вычислений. Нас интересует не просто существование алгоритмов для нашей задачи, а то, насколько они эффективны.

Конечно, рассказанная история несколько стилизована, но, в общем-то, так и происходило развитие исследований, которое привело к современному состоянию. В частности, тот этап, когда стало выкристаллизовываться понимание того, что одни алгоритмы могут быть лучше других и это существенно, пришёлся на 60-е годы. Тогда стали появляться настоящие компьютеры (они назывались таким страшным словосочетанием: «электронные вычислительные машины»), например, БЭСМ (многие в аудитории и не знают, наверное, что это такое). Стало понятно, что необходимо построение некоторой математической теории.

**3. Начала теории: основные понятия.** Давайте пока оставим в стороне истории из жизни математика M, потом мы к нему ещё вернемся не раз.

Попробуем дать несколько определений.

Первое наблюдение, которое сделал M, когда пытался разобраться в этой теории, состоит в том, что подавляющее большинство алгоритмических задач можно представить в подходящей кодировке как задачи вычисления некоторого отображения

$$f\colon\ \{0,1\}^* \to \{0,1\}^*$$

множества конечных двоичных слов в себя. Именно такими отображениями мы и будем заниматься.

Второй вопрос: на каких устройствах мы собираемся вычислять наше отображение? Есть огромное разнообразие и алгоритмических языков, и компьютерных архитектур, но насколько нам это важно? Именно ответ на этот вопрос отличает теорию сложности вычислений от различных смежных областей. Поэтому рассмотрим этот вопрос более подробно.

В качестве примера возьмём программу, с которой чаще всего и сталкивается большинство математиков, которые не программируют. Это TEX процессор. Выбор TEX здесь никакого особого значения не имеет, то же самое можно сказать и о других программах.

Итак, имеется некоторое отображение $f$, которое в данном случае представляет собой отображение

$$\boxed{\texttt{paper.tex}} \longrightarrow \boxed{\texttt{paper.dvi}},$$

преобразовывающее файл `paper.tex` в файл `paper.dvi` (оба файла мы рассматриваем как два длинных двоичных слова). Это преобразование осуществляется некоторым алгоритмом `texdvi.exe`, закодированным для выполнения на 286-м процессоре (те, кто пытался использовать TEX на 286-м процессоре, помнят, как это было).

Нас сейчас интересует время (быстродействие) работы алгоритма. Именно такую, как говорят в этой науке, *меру сложности* мы будем рассматривать сегодня чаще всего. Бывают и другие характеристики (память — вторая по значению мера сложности), но за недостатком времени мы их не сможем подробно обсудить.

Развитие техники, относящееся к рассматриваемой нами задаче, шло по двум направлениям. Во-первых, и это хорошо все знают, улучшались процессоры:

$$\text{Intel 286},\ \text{Intel 386},\ \text{Intel 486},\ \ldots$$

появлялись один за другим всё более мощные модели, возникали разнообразные конструкторские ухищрения, ускоряющие их работу, и т. д. Во-вторых, улучшался сам алгоритм, появлялись его версии

`texdvi1.exe`, ... `texdvi10.exe`, ...

Справедлива, хотя и с некоторыми оговорками, такая формула

$$\left\langle \begin{matrix}\text{общее время}\\ \text{работы}\end{matrix} \right\rangle = \left\langle \begin{matrix}\text{число операций}\\ \text{алгоритма}\end{matrix} \right\rangle \times \left\langle \begin{matrix}\text{время на одну}\\ \text{операцию}\end{matrix} \right\rangle.$$

Улучшение обоих членов в этой простой формуле — вещи более-менее независимые. Грубо говоря, программы (software) отвечают за первые сомножитель, а процессоры (hardware) — за второй.

Что здесь происходит с математической точки зрения? Если у нас имеется какой-то алгоритм, который использует $t$ операций, то от улучшения быстродействия нашего процессора время решения нашей задачи умножается на некоторую константу.

Поэтому, и это очень важно, в теории сложности вычислений сложилась традиция измерять время работы алгоритма с точностью до мультипликативной константы («с точностью до $O(\cdot)$», как говорят математики). Такой подход позволяет нам абстрагироваться от выбора конкретной модели вычислительного устройства, от того, сколько времени занимает выполнение одной операции, какую систему команд имеет используемая нами машина и т. п. Именно этот подход и позволяет нам построить довольно красивую математическую теорию.

Теория сложности вычислений несколько отличается от всеобъемлющей (по определению) теории марксизма-ленинизма (не все в аудитории видимо знают, что это такое, и это хорошо): если в формуле есть два члена, и за второй наша теория никакой ответственности не несёт, то об этом заявляется явно. Улучшением процессоров занимаются другие люди, мы занимаемся улучшением алгоритмов с точностью до мультипликативной константы.

Продолжим построение строгой теории. Одним из непосредственных преимуществ введённого выше соглашения является то, что нас не очень заботит выбор точной модели. Как правило, от смены модели (сейчас я имею в виду уже абстрактные математические модели) улучшение или ухудшение происходят с точностью до мультипликативной константы, а мы договорились этого не замечать.

Стандартный способ определения вычислительной модели состоит в том, что для исполнения алгоритма у нас имеется некоторая машина с адресуемой памятью, ячейки которой индексируются натуральными числами; в каждой ячейке можно хранить натуральные числа; можно выполнять арифметические действия и т. п. Детали здесь мало интересны, потому что наше $O(\cdot)$-соглашение позволяет о них заботиться не слишком. Если же заменить «с точностью до $O(\cdot)$» на слегка более грубое «с точностью до полинома», то вообще все реалистические вычислительные устройства оказываются эквивалентными.

Итак, фиксируем некоторую вычислительную модель. Если есть машина $M$, которая вычисляет функцию $f$, и определены входные данные $x$ — некоторое двоичное слово, тогда можно определить нашу основную функцию $T(M, x)$ — количество операций (тактов), которые требуются машине $M$ для работы на входном слове $x$. Функция $T(M, x)$ называется *сигнализирующей функцией по времени.*

**4. Теорема об ускорении.** Давайте теперь изучать эту функцию $T(M, x)$. Первое, что приходит в голову: у нас имеется алгоритмическая задача $f$; давайте выберем самый хороший алгоритм для решения этой задачи и назовём сложностью задачи сложность этого самого хорошего алгоритма.

Оказывается, что такой интуитивно очевидный способ действий, к сожалению, неосуществим и на этот счёт имеется *теорема Блюма об ускорении.* В вольной формулировке эта теорема утверждает, что как бы мы ни пытались определять понятие «наилучшая для данной задачи машина», у нас ничего не получится (по крайней мере для некоторых задач).

Теорема Блюма открывает серию теорем, лежащих в основе современной теории сложности вычислений. Все эти теоремы были доказаны в 70-е годы, например, теорема Блюма об ускорении — это 1971 год. Примерно в то же время появилась концепция NP-полноты, изложением которой и завершится наш вводный рассказ о теории сложности вычислений.

Для дальнейшего нам потребуется ещё одно важное определение. Функция $T(M, x)$ устроена очень нерегулярно. Рассмотрим пример TEX процессора. Для подавляющего большинства файлов произойдёт остановка в самом начале работы из-за несоответствия входному формату TEX, а на каких-то файлах это время возможно окажется бесконечным из-за зацикливания. В общем случае эту функцию изучать никакой возможности не представляется, она слишком рыхлая. Мы хотим извлечь из неё функцию натурального аргумента, т. е. получить функцию из натуральных чисел в натуральные числа, которая отражала бы поведение $T(M, x)$. Есть несколько способов подойти к этой задаче. Мы здесь ограничимся самым распространённым, который называется *сложность в наихудшем случае*. Она определяется следующей формулой

$$t_M(n) = \max_{|x| \leqslant n} T(M, x).$$

Из всех слов битовой длины, не превышающей $n$, выбираем то, на котором наша машина работает хуже всего (т. е. дольше всего). Время работы на таком слове и называем сложностью $t_M(n)$. За время $t_M(n)$ машина гарантированно закончит работу на любом входном слове длины, не превышающей $n$. Может, конечно, случиться, что на каких-то словах работа завершится раньше.

Мы приводим упрощённый вариант теоремы Блюма; на самом деле, вместо $\log t$ в ней можно взять любую «разумную» стремящуюся к бесконечности функцию.

**Теорема 1** (Блюм, 1971). *Существует такая вычислимая*[2] *функция $f$, что любую машину $M$, вычисляющую $f$, можно ускорить следующим образом: существует другая машина $M'$, также вычисляющая $f$ и такая, что*

$$t_{M'}(n) \leqslant \log t_M(n)$$

*для почти всех $n$.*

Функция из теоремы Блюма — это некоторая экзотика (как следует из формулировки теоремы, время её вычисления в наихудшем случае растёт очень быстро для любой машины). Одна из «идеологических» задач в теории сложности вычислений — это строить теорию так, чтобы такие патологические явления в ней по возможности не возникали. Теорема Блюма доказывается построением с помощью техники диагонализации, и получающаяся функция не имеет никакого отношения к реальной практике вычислений или к остальной математике. Но тем не менее, поскольку мы строим математическую теорию, ничего не поделаешь — мы вынуждены признать, что выбранный подход не годится для её построения. Нужно двигаться дальше.

**5. Сложностные классы.** Итак, мы не можем надеяться на построение для каждой функции самой лучшей машины, вычисляющей эту функцию. Альтернативой является понятие *сложностного класса* — одно из центральных понятий для теории сложности вычислений.

---

[2] Вычислимая функция — это такая функция, которую можно вычислить хотя бы одним алгоритмом.

Несколько вольная аналогия может быть здесь проведена с определениями интеграла по Риману и по Лебегу. Если мы не можем проинтегрировать по Риману, то мы меняем ось и начинаем суммировать по другой оси. В нашей ситуации, если мы не можем для функции сказать, что такое самая хорошая машина, то давайте также сменим ось. Рассмотрим множество всех машин, которые нас устраивают, и назовём класс всех функций, вычислимых такими машинами, *классом сложности*. Пожалуй, проще всего дать сразу конкретный пример, чем долго рассуждать на эту тему:

$$\mathrm{DTIME}(t(n)) := \{f \mid \exists M : (M \text{ вычисляет } f)\,\&\,(t_M(n) = O(t(n)))\}.$$

Это одно из центральных определений в теории сложности. Буква D означает детерминированные алгоритмы (бывают и другие), TIME означает ровно то, о чём вы подумали. Если у нас есть произвольная функция $t(n)$ от натурального аргумента, то мы образуем класс сложности, состоящий из таких функций, для которых существует вычисляющая $f$ машина $M$, такая что сигнализирующая функция по времени ограничена исходной функцией $t(n)$ с точностью до мультипликативной константы. Приведённая выше теорема Блюма справедлива только для некоторых специальных функций. Но если мы хотим достичь ускорения, скажем, в 10 раз, то это можно сделать с любой функцией — например, путём увеличения количества машинных команд. Именно поэтому в правой части определения стоит $O$-большое.

Теперь определим один из самых важных сложностных классов

$$\mathrm{P} = \bigcup_{k\geqslant 0} \mathrm{DTIME}(n^k).$$

Класс P — это те функции, которые можно вычислить на наших машинах, и время их вычисления растёт полиномиально с ростом длины слова. Он очень удобен и с практической, и с теоретической точки зрения. С практической точки зрения это достаточно хорошая аппроксимация (бывают исключения, о которых будет сказано ниже) класса тех функций, которые поддаются вычислению за реальное время на реальных компьютерах. С математической точки зрения этот класс бесконечно удобен тем, что он замкнут относительно суперпозиции. Как мы увидим дальше, именно это обстоятельство позволяет построить теорию вычислимости для этого класса.

Бывают аналогичные классы языков, которые можно распознать за экспоненциальное время

$$\mathrm{EXPTIME} = \bigcup_{k\geqslant 0} \mathrm{DTIME}\left(2^{n^k}\right),$$

также можно определить двойное экспоненциальное время

$$\mathrm{DOUBLEEXPTIME} = \bigcup_{k\geqslant 0} \mathrm{DTIME}\left(2^{2^{n^k}}\right)$$

и так далее.

**6. Теорема об иерархии и сложность элементарной геометрии.** Теперь давайте вернёмся к нашему математику M. Оказывается, время работы алгоритма Тарского, с помощью которого M пытался решать задачи элементарной геометрии, не лежит ни в одном из приведённых выше классов. Для него есть такая нижняя оценка

$$t_{\mathsf{Tarsky}}(n) \geqslant 2^{2^{\cdot^{\cdot^{2}}}}\Big\}\,\varepsilon n$$

Как вы помните, для разрешимости алгебры Тарского продавался ещё и диск Collins for Windows 95.

**Теорема** (Коллинз). *Алгебра Тарского принадлежит классу сложности* DOUBLEEXPTIME.

Теперь M смог понять разницу между этими CD-ROM: время работы алгоритма Коллинза неизмеримо меньше времени работы алгоритма Тарского (хотя оно всё равно может быть очень велико — оценка времени работы алгоритма двойной экспонентой не гарантирует нам, что даже теорема о сумме углов треугольника будет доказана за время существования Вселенной).

Возникает вопрос: а нельзя ли ещё улучшить алгоритм разрешимости для алгебры Тарского? И более общий вопрос: а нельзя ли то же самое сделать вообще с любым алгоритмом? Вдруг, например, всякая вычислимая функция лежит в классе P. Или, по крайней мере, всякая функция из DOUBLEEXPTIME лежит в P.

Другими словами, вопрос в том, есть ли у нас предмет исследования, или, возможно, теорема Блюма об ускорении применима вообще ко *всем* функциям.

Второй краеугольный камень теории сложности вычислений — это *теорема об иерархии*.

Как и в случае с теоремой об ускорении, мы приводим её далеко не в самой общей форме.

**Теорема** (Хартманис (1965)). P $\neq$ EXPTIME.

Таким образом, не все сложностные классы совпадают, так что предмет исследования есть.

Я не могу отказать себе в удовольствии привести почти полное доказательство этой теоремы. Если вы ещё помните нашего математика M, второй момент его злоключений состоял в том, что он спрашивал, остановится ли его программа *когда-нибудь*. Сейчас, наученный горьким опытом, он задал такой вопрос:

*остановится ли эта программа до Нового Года?*

Выясняется, что если до Нового Года осталось в точности экспоненциальное время, то эта задача и отделяет EXPTIME от P. Потому что есть очень простой алгоритм, позволяющий проверить, остановится ли программа до Нового Года, а именно, нужно просто подождать до него, и всё само собой решится. Теорема об иерархии утверждает, что эту задачу нельзя решить существенно быстрее, чем описанным выше способом.

Я, конечно, слегка утрирую, есть там и кое-какие технические детали, но смысл состоит именно в этом.

И теперь наш математик оказался вполне подготовлен к восприятию следующей теоремы.

**Теорема** (Фишер—Рабин (1974)). *Алгебра Тарского не принадлежит к классу* P. *Время работы любого разрешающего алгоритма для алгебры Тарского не меньше, чем* $2^{\varepsilon n}$, *где* $\varepsilon$ — *некоторая константа.*

Такая высокая *нижняя оценка* объясняет причины неудачи нашего математика с практическим решением задач в алгебре Тарского.

Наиболее сложная и, по-видимому, наиболее важная область теории вычислений, связана как раз с доказательством нижних оценок. В англоязычной литературе тот раздел теории сложности, который занимается конструированием алгоритмов, так и называется — “Theory of Algorithms”, а собственно под “Complexity Theory” подразумевается как раз построение нижних оценок. Таким образом, теория сложности пытается доказать, что не существует эффективных алгоритмов.

Уже на этом примере видны две трудности, стоящие перед людьми, которые пытаются доказывать нижние оценки. Смотрите, была теорема Тарского, её пытались улучшить, но ничего не получалось Естественно было предположить, что эта оценка оптимальна. После этого придумали алгоритм, основанный на совсем других идеях. Алгоритмов бывает много, алгоритмы бывают сложные, алгоритмы бывают разные. А ведь мы пытаемся доказать, что более эффективный алгоритм не появится никогда, и при этом пытаемся анализировать свойства алгоритмов из достаточно широкого класса.

**7. Сводимость и полнота.** Давайте продвинемся дальше в построении нашей теории. Понятие сложностного класса было введено вовсе не для того, чтобы экономить на формулировках теорем (в конце концов, теорему Фишера—Рабина можно сформулировать и без упоминания каких бы то ни было сложностных классов — для этого достаточно оставить в её формулировке лишь вторую фразу). Понятие сложностного класса становится важным в тот момент, когда у нас появляется понятие *сводимости*, и это второе центральное понятие современной теории сложности вычислений. Есть несколько вариантов определения сводимости, я расскажу лишь о наиболее важном из них.

**Сводимость по Карпу.** Эта сводимость устроена очень просто. Для начала заметим, что мы занимаемся вычислением функций, отображающих конечные слова в конечные слова. Но во многих случаях оказывается гораздо удобнее, и это, как правило, не приводит к потере общности, рассматривать так называемые *языки*. Язык $L$ можно рассматривать как множество слов $L \subseteq \{0,1\}^*$ или интерпретировать его как отображение вида $\varphi\colon \{0,1\}^* \to \{0,1\}$, тогда $L = \varphi^{-1}(1)$. Переход к языкам не слишком нас ограничивает: если есть произвольная функция из слов в слова, то с ней можно связать такой язык — множество пар $(x, i)$ таких, что $i$-й бит $f(x)$ равен 1.

**Определение.** Язык $L_1$ сводится к языку $L_2$ по Карпу (обозначается это таким образом $L_1 \preceq_p L_2$), если существует такая функция $f$ из P, что

$$\forall\, x : (x \in L_1 \equiv f(x) \in L_2)\,.$$

Сводимость можно рассматривать как использование некоторой подпрограммы, перерабатывающей исходное слово $x$ в $f(x)$, к которому затем нужно применить алгоритм распознавания принадлежности языку $L_2$. В этом варианте понятия сводимости обращение к программе распознавания $L_2$ осуществляется ровно один раз, если же разрешено использовать в качестве подпрограммы распознавание языка $L_2$ сколько угодно раз, то получается другая сводимость (по Тьюрингу). Сейчас это различие нам неважно.

Отношение сводимости — это отношение предпорядка. Оно рефлексивно и транзитивно. Самое главное свойство состоит в том, что если $L_2 \in$ P, то и $L_1 \in$ P. Здесь важно, что класс полиномов замкнут относительно операции суперпозиции.

Например, EXP-сводимость не получится. Не будет замкнутости относительно такой сводимости. Замкнется всё лишь на конечных башнях экспонент — классе *элементарных рекурсивных функций*.

Если говорить только о естественных классах, в которых реально есть хотя бы одна разумная функция и которые замкнуты относительно суперпозиции, то есть ещё класс квазиполиномов $2^{(\log n)^{O(1)}}$ и класс квазилинейных функций $n \log^{O(1)} n$. Квазилинейными функциями сейчас начинают интенсивно интересоваться, потому что считается, что по-настоящему хороший алгоритм — это алгоритм, который работает как раз квазилинейное время.

В любом случае, класс P является центральным в теории, и мы сейчас будем рассматривать именно полиномиальную сводимость.

Легко видеть, что класс EXPTIME замкнут относительно этой сводимости. Поэтому естественно поставить вопрос, а существуют ли в этом классе самые сложные языки, т. е. такие, что любой другой язык из этого класса к ним сводится. В этом случае можно решить любую задачу из EXPTIME, используя произвольный алгоритм распознавания для такого самого сложного языка и полиномиальную сводимость. Языки из некоторого сложностного класса, к которым сводится любой язык из этого класса, называются *полными* (относительно данного класса и данного типа сводимости). А если мы опустим требование принадлежности нашему классу самого языка, получим определение *трудного* языка.

Теорема Фишера—Рабина доказывается именно так. Вместо того, чтобы доказывать напрямую, что не существует полиномиального алгоритма для алгебры Тарского, доказывается, что алгебра Тарского трудна для класса EXPTIME. Поэтому полиномиальный алгоритм для алгебры Тарского давал бы полиномиальный алгоритм и для всех остальных задач из этого класса. А мы знаем, что в EXPTIME есть задачи, которые нельзя решить за полиномиальное время («остановка программы до Нового Года»).

Такой способ рассуждений типичен для теории сложности. Мы не пытаемся действовать «в лоб», а сводим одну задачу к другой. И, естественно, ситуация наиболее хороша, когда к данной задаче сводится много других задач из данного класса.

Успех этой науки определяется тем практическим обстоятельством, что полных задач оказывается много в самых разных ситуациях и они более естественны, чем проблема остановки.

**8. Всякий ли полиномиальный алгоритм хорош?** Теперь самое время вернуться к нашему бедному математику M и поговорить об исключениях из правила «класс P = класс эффективно вычислимых функций». Как-то M понадобилось для каких-то своих целей решать системы линейных неравенств:

$$\sum_{j=1}^{n} a_{ij}x_j \leqslant b_i \qquad (i = 1, \ldots, m).$$

Другими словами, ему потребовался пакет программ для задач линейного программирования. Специалисты по теории сложности вычислений перед покупкой программного обеспечения изучают литературу, и M, наученный предыдущим горьким опытом, стал следовать этому правилу. Он, конечно же, слышал про симплекс-метод, которым пользуются для решения задачи линейного программирования почти всюду, включая военную кафедру Московского университета. Но M обнаружил статью, в которой доказывается, что симплекс-метод не полиномиален. А после этого наш M обнаружил и статью Хачияна (1979), в которой был построен полиномиальный алгоритм для решения задачи линейного программирования. Поэтому, приехав на Митинский радиорынок, он попытался найти что-то вроде CD Khachiyan for Windows 95. К его удивлению, ничего похожего не было. Всё, что ему предлагали, было основано на симплекс-методе и его вариациях. Оказывается, что хотя теоретически алгоритм решения задачи линейного программирования симплекс-методом экспоненциальный, а алгоритм Хачияна полиномиальный, на практике первый работает быстрее второго. Чтобы построить пример задачи линейного программирования, на которой симплекс-метод будет работать долго, нужно очень и очень постараться, а полиномиальный алгоритм Хачияна работает примерно одинаково на любом входе (показатель экспоненты 6, что весьма плохо).

Это самое известное исключение из того правила, что полиномиальные алгоритмы хороши, а экспоненциальные — плохи. Но это исключение, на самом деле, подтверждает правило, потому что, хотя алгоритмом Хачияна никто и не пользуется для решения задач линейного программирования, выяснилось, что с помощью этого алгоритма можно решать такие задачи, к которым с симплекс-методом подступиться в принципе невозможно. Например, пусть у нас есть произвольное выпуклое тело $K$ и задано некоторое направление. Нужно максимизировать значение соответствующей линейной формы на $K$. Про выпуклое тело вы ничего не знаете: оно задано с помощью чёрного ящика (или, как иногда говорят, *оракула*). То есть, если вы указываете точку $p$, то с некоторой погрешностью $\varepsilon$ можно сказать, лежит ли точка $p$ в теле $K$, а если не лежит — получить некоторую отделяющую $p$ от $K$ гиперплоскость. Ответ нужно получить с той же точностью. Никакой симплекс-метод здесь, понятное дело, не работает — вообще нет никаких вершин, перебором которых занимается симплекс-метод. А алгоритм Хачияна и та наука, которая из него выросла, замечательно (т. е. полиномиально) с такими задачами справляется. Роль параметра $n$, описывающего размер входа, здесь играет $d \cdot (\log \varepsilon^{-1})$, где $d$ — размерность, а $\varepsilon$ — точность вычисления.

Таких исключений известно очень мало. Второй известный пример — проверка простоты числа. Как правило, подтверждается тезис, что если у вас имеется алгоритм, который теоретически работает хорошо, и это — алгоритм для нормальной задачи, которая и в самом деле откуда-то возникла, а не сконструирована с хулиганскими намерениями, то этот алгоритм будет хорошо работать и на практике. В частности, показатель $k$ в оценке скорости работы алгоритма $n^k$ обычно оказывается небольшим, а когда он поначалу велик, его можно уменьшить различными ухищрениями. Для подавляющего большинства естественных задач показатель не превосходит 3.

**9. Недетерминированные вычисления.** По-видимому, те, кто пришел на эту лекцию, хотели узнать что-нибудь про самую известную открытую проблему в этой области $\mathrm{P} \overset{?}{\neq} \mathrm{NP}$. Что такое P, я уже немножко объяснил. Теперь давайте займёмся NP.

Для этого опять вернёмся к **M**. Пока он разбирался в теории сложности, его сын поступил в университет на первый курс и стал изучать математическую логику. Изучение математической логики, как известно, начинается с такой знаменитой науки, как исчисление высказываний. У вас имеется некоторая пропозициональная формула $\Phi(p_1, \ldots, p_n)$, она называется тавтологией, если она истинна независимо от того, что мы в неё подставим. И одно из неприятных упражнений при занятиях этой наукой состоит в том, что нужно выяснить по выписанной пропозициональной формуле, является она тавтологией или нет. Именно с таким вопросом и обратился сын нашего математика к своему папе. M, естественно, уже никуда не поехал, а стал, как и поступают всегда специалисты по теории сложности, пытаться поместить задачу в один из уже известных сложностных классов. Чтобы использовать стандартные обозначения, будем говорить о двойственной ей задаче $SAT$ — выполнимости[3]: есть ли хотя бы один набор переменных, при котором формула истинна.

Накладывая эту задачу на нашу картину сложностных классов, M увидел первым делом, что $SAT \in \mathrm{EXPTIME}$. Алгоритм решения задачи $SAT$ за экспоненциальное время очевиден — всего возможных наборов значений переменных $2^n$, а время вычисления значения формулы при заданных значениях переменных полиномиально. Следующий шаг — классифицировать эту задачу: лежит ли она в P, или полна в EXPTIME? Этот вопрос, возникший в начале 70-х годов, до сих пор открыт. Почему у специалистов, потративших на решение этой задачи почти 30 лет, не получается построение полиномиального алгоритма, объяснить сложно (я, во всяком случае, никак это объяснять не берусь). Но тот факт, что не получается доказательство полноты этой задачи в EXPTIME, некоторому объяснению поддаётся. Если мы посмотрим на настоящие экспоненциальные алгоритмы, например, на алгоритм для аналога алгебры Тарского над комплексными числами, и сравним с ними то детское рассуждение, которое мы привели выше, то сразу невооружённым глазом видна разница — приведенный алгоритм для $SAT$ использует экспоненциальное время очень слабо — только для *перебора* экспоненциального числа возможностей, с каждой из которых он справляется за полиномиальное время.

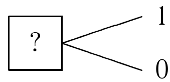
Взрыв в теории сложности вычислений начался с того, что было сформулировано определение класса NP, как класса языков, которые распознаются переборными алгоритмами. Более научно, переборные алгоритмы называют недетерминированными алгоритмами. NP — это класс языков, которые можно распознать за недетерминированное полиномиальное время.

Теперь дадим определение этого класса. В этом определении, как и в определении класса P, будет фигурировать слово «машина», сокращенно мы будем обозначать её НМТ («недетерминированная машина Тьюринга»); впрочем, использование слова «машина» в этом контексте может вызывать у людей, имеющих отношение к реальным машинам, некоторое чувство неудовольствия. Под недетерминированной машиной мы понимаем такую машину, которая работает как обычно, но в какой-то момент она может нарисовать в какой-то ячейке знак вопроса и после этого её работа разделяется на две

---

[3] $SAT$ — от слова satisfiability — выполнимость.

ветви 0 или 1:




(в клетке окажется записанным либо 0, либо 1). После этого машина продолжает работу. В какой-то момент она может раздвоиться ещё раз. Появляется дерево вычислений. Вдоль каждой ветви дерева вычислений НМТ работает как обычное вычислительное устройство, но результат работы НМТ зависит от результатов работы вдоль всех ветвей. Определим результат работы НМТ в случае проверки принадлежности слова $x$ к языку $L$. На каждой из ветвей вычисления получается один из двух возможных ответов: «да» или «нет». НМТ распознаёт язык $L$, если всякое слово $x$ принадлежит $L$ тогда и только тогда, когда *хотя бы на одной* ветке вычислений получен ответ «да».

Проиллюстрируем возможности НМТ на примере с упражнениями на проверку выполнимости пропозициональных формул. Частенько, когда студент (полиномиальная детерминированная машина Тьюринга) проваливается и не может решить задачу, преподаватель выступает в роли недетерминированной машины — он показывает (загодя приготовленный) ответ, который легко может быть проверен. Другими словами можно сказать так: НМТ «стремится» доказать утверждение $x \in L$, а в момент раздвоения она обладает неограниченными интеллектуальными возможностями и выбирает наилучший вариант. Если существует ветка вычислений, при которых ответ «да», то $x \in L$. В противном случае никакое ветвление не приведёт к положительному результату (слово не принадлежит языку лишь тогда, когда у НМТ нет никакой возможности доказать обратное).

Не спрашивайте меня о физике процесса: как происходит такое раздвоение, где находится такая машина, можно ли её посмотреть... Таких машин в реальности нет. Одним из самых значительных достижений в теории сложности за последние годы стало то, что была сформулирована квантовая модель вычислений. Квантовые компьютеры являются одним из кандидатов на роль такой машины в реальном мире. По крайней мере, существование квантовых компьютеров не вызывает принципиальных возражений у физиков, а надежда на имитацию недетерминированных машин квантовыми компьютерами пока не вызывает категорических возражений со стороны специалистов в теории сложности. Более того, абстрактные квантовые компьютеры уже умеют решать некоторые важнейшие переборные задачи, такие, как факторизация чисел[4].

---

[4] На последнем Международном Конгрессе (Берлин, август 1998г.) американский математик П. Шор был удостоен за эти исследования премии им. Неванлинны.

Подчеркнем ещё раз, что недетерминированная машина является чисто теоретическим понятием. Удобства от введения такого понятия получаются совершенно фантастические.

Это самая первая вымышленная модель, появившаяся в теории сложности вычислений. В настоящий момент есть масса таких моделей, гораздо более сложных: интерактивные доказательства и т. п. Эти модели появляются не сами по себе, а с целью определения сложностных классов и для классификации естественных задач.

Интуитивно ясно, что понятие НМТ прекрасно приспособлено к моделированию переборных алгоритмов. Собственно, уже показано, как это делать. Если у нас есть алгоритм, который пытается перебрать некоторое количество возможностей, то наша машина может угадать, какая из возможностей является хорошей, а потом имитировать вторую часть (полиномиальную проверку конкретного варианта). В обратную сторону это несколько сложнее, но всё равно достаточно просто. Если у нас есть НМТ, то она порождает дерево вычисления, и перебор нужно производить по всем возможным ветвям.

Современная теория сложности вычислений началась с результатов Кука, Карпа, Левина (который получил их независимо) в начале 70-х годов (1970–1972). Эта серия теорем состоит вот в чём.

1. Задача *выполнимость* ($SAT$) полна для класса NP (теорема Кука). Таким образом, задача нашего математика **M** (можно ли придумать полиномиальный алгоритм для $SAT$) эквивалентна вопросу: совпадают ли классы P и NP ($\mathrm{P} \overset{?}{\neq} \mathrm{NP}$). Если полиномиального алгоритма нет, то эти классы не совпадают — их отделяет задача $SAT$, если же он есть, то можно эффективно решить любую задачу из класса NP. Задача выполнимости отвечает за класс NP. Грубо говоря, NP — это не что иное, как задачи, которые могут быть сведены к выполнимости пропозициональных формул.

2. Естественно, что такой результат может вызвать некоторую реакцию отторжения — не настолько уж важна задача выполнимости, чтобы строить целую теорию её решения. Но следующим шагом была статья Карпа (1971), в которой уже была указана 21 полная задача для класса NP. Все они между собой эквивалентны.

В настоящее время список NP-полных задач, которые возникают буквально во всех областях математики, содержит тысячи задач. Везде, где возникают алгоритмы, возникают и переборные задачи. Это неудивительно, потому что бóльшая часть программистской работы и состоит в выборе варианта получше. Удивительно то, что часто, а практически — всегда (с некоторыми исключениями) — если у вас есть конкретная переборная задача, то её можно сравнительно легко классифицировать — либо для неё есть полиномиальный алгоритм, либо она полна.

Таким образом строится теория переборных задач. Из-за того, что переборные алгоритмы возникают почти всюду, она приобрела большое значение.

По мнению американского тополога Смейла, вопрос $\mathrm{P} \overset{?}{\neq} \mathrm{NP}$ будет одним из самых важных вопросов математической науки следующего столетия.

Вот и подошёл к концу рассказ о том, что является, так сказать, ядром теории сложности вычислений. Не стоит понимать этот рассказ так, что вся теория сложности занимается только лишь соотношением $\mathrm{P} \overset{?}{\neq} \mathrm{NP}$. Изложенная схема исследований — которая объединяет задачи в сложностные классы и потом исследует эти задачи с помощью их сводимости друг к другу — оказалась удивительно эффективна плодотворна в самых разных ситуациях.

Перечислим коротко несколько самых важных возможностей (помимо уже упомянутых выше квантовых вычислений).

Бывает сложность алгебраическая. Если нас интересует вычисление некоторого полинома, и мы абстрагируемся не только от деталей битового вычисления, но и от того, как производятся арифметические операции, то всё зависит лишь от того, с какими числами производятся арифметические операции. Тогда длина вычисления определяется количеством сложений и умножений.

Бывает сложность геометрическая — диаграммы Вороного и близкие к ним вещи, о которых я говорить не буду.

Бывает сложность булева — она отличается от того, чем мы занимались сегодня, тем, что нас интересуют функции $f\colon \{0,1\}^n \to \{0,1\}$, определённые на словах фиксированной длины.

В любом из этих разделов возникает масса постановок интереснейших задач именно в рамках этой общей идеологии.

**Благодарности.** Я выражаю глубокую благодарность М. Н. Вялому и М. В. Алехновичу за квалифицированную помощь в приведении данной лекции к пригодному для чтения виду и проф. J. Krajíček за ряд ценных замечаний.

*С. П. Новиков*

# Уравнение Шредингера и симплектическая геометрия

*Лекция 25 июня 1998 года*

Эта лекция посвящена довольно элементарным вещам. Я познакомлю вас с некоторыми полезными понятиями математической физики. Прежде чем говорить о том, о чём я собираюсь говорить (о некоторых экзотических операторах на графах), позвольте мне напомнить вам, что такое уравнение Шредингера. Уравнение

$$-\psi'' + u(x)\psi = \lambda\psi$$

мы будем называть уравнением Шредингера, а величину $u(x)$ будем называть потенциалом. Иногда это уравнение рассматривают формально, но если уравнение Шредингера возникает из квантовой механики, то обычно предполагается, что $\psi \in L_2(\mathbb{R})$, т. е.

$$\int_{\mathbb{R}} |\psi(x)|^2 dx < \infty.$$

Пространство $L_2(\mathbb{R})$ гильбертово. Именно для векторов этого пространства рассматривается спектральная задача.

В отличие от традиционных матричных ситуаций, где возникает задача на собственные значения, здесь бывает такое понятие, как «непрерывный спектр».

Будем считать, что при $x \to \infty$ функция $\psi(x)$ достаточно быстро стремится к нулю. Чтобы избежать сложностей с обоснованием сходимости, можно даже считать, что функция $\psi(x)$ финитна, т. е. тождественно равна нулю вне конечной области.

**1. Квантовое рассеяние.** Нас интересует такое понятие, как квантовое рассеяние. Для финитного потенциала естественно предположить, что существуют такие решения $\psi_\pm$, что $\psi_\pm(x) \to e^{\pm ikx}$ при $x \to -\infty$ (здесь $k^2 = \lambda$). Это один базис пространства решений. Второй базис мы получим, если рассмотрим такие решения $\varphi_\pm$, что $\varphi_\pm(x) \to e^{\pm ikx}$ при $x \to +\infty$.

Пространство решений дифференциального уравнения второго порядка при любом фиксированном $\lambda$ является двумерным пространством, поэтому

$$a\psi_+ + b\psi_- = \varphi_+, \qquad c\psi_+ + d\psi_- = \varphi_-. \tag{1}$$

Возникает матрица $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. В некоторых чисто математических книгах эту матрицу неправильно называют матрицей рассеяния. Правильное название — *матрица монодромии*. Матрица монодромии очень часто возникает в разных разделах математики, например, в комплексной теории дифференциальных уравнений. Матрица монодромии — это матрица переноса из $-\infty$ в $+\infty$ вдоль оси $x$, т. е. матрица переноса слева направо. У этой матрицы есть целый ряд очень интересных свойств. В частности, поставим теперь вопрос о рассеянии. Рассмотрим $\lambda \in \mathbb{R}$. Мы знаем, что самосопряжённые матрицы имеют вещественный спектр. Так будет и для операторов, если они самосопряжённые в разумном смысле. Более того, легко видеть, что в данном случае интересующий нас спектр будет только для положительных $\lambda$. Для $k^2 = \lambda$ это соответствует тому, что $k \in \mathbb{R}$. Область $\lambda \geqslant 0$ на прямой назовём *зоной рассеяния*.

Теперь, что такое матрица рассеяния. Для $k \in \mathbb{R}$ выполняются следующие свойства: $\overline{\psi}_+ = \psi_-$ и $\overline{\varphi}_+ = \varphi_-$. В таком случае матрица $T$ имеет вид $T = \begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix}$, причём по некоторым причинам, которые нас как раз особенно интересуют, $\det T = |a|^2 - |b|^2 = 1$. Такие матрицы образуют группу $SU(1,1)$. Это — специальные унитарные матрицы, которые сохраняют индефинитное эрмитово скалярное произведение с одним положительным и одним отрицательным квадратом. Я рекомендую в виде полезного алгебраического упражнения доказать, что $SU(1,1) \cong SL_2(\mathbb{R})$. Дело в том, что мы получили комплексную матрицу для чисто вещественного уравнения: мы требуем, чтобы функция $u(x)$ была вещественной (по смыслу $u$ обычно является электрическим потенциалом). Если бы мы при выборе базиса взяли на $\pm\infty$ не $e^{\pm ikx}$, а $\cos kx$ и $\sin kx$, то матрица монодромии была бы вещественной матрицей из группы $SL_2(\mathbb{R})$.

Мы переходили от базиса $\psi_+, \psi_-$ к базису $\varphi_+, \varphi_-$. Если же перейти от базиса $\varphi_-, \psi_-$ к базису $\varphi_+, \psi_+$, то можно доказать, что матрица перехода будет унитарной (хотя и не любая унитарная матрица так получается). Эту матрицу обозначают $S(\lambda)$ и называют *матрицей рассеяния*.

Это — элементарный материал, который физики учат на втором курсе. Математики не всегда его учат вообще. Но для физиков-теоретиков обыкновенные дифференциальные уравнения излагаются именно так.

Переход от матрицы $T(\lambda) \in SU(1,1)$ к матрице $S(\lambda) \in U(2)$ называют *преобразованием Кэли*.

Вот ещё одно полезное упражнение, взятое из хорошего учебника В. И. Арнольда «Дополнительные главы теории обыкновенных дифференциальных уравнений». Рассмотрим пару векторов $e_1$ и $e_2$ и матрицу $T \in SL_2(\mathbb{R})$, отображающую эту пару векторов в пару $e_1'$ и $e_2'$. Введём четырёхмерное пространство $\mathbb{R}^4$ с базисом $e_1, e_2, e_1', e_2'$ и определим в нём кососимметрическое скалярное произведение так, что $\langle e_1, e_2\rangle = -\langle e_1', e_2'\rangle = 1$, а остальные произведения базисных элементов нулевые (разумеется, $\langle e_2, e_1\rangle = -\langle e_2', e_1'\rangle = -1$). Пространство $\mathbb{R}^4$ с таким кососимметрическим скалярным произведением назовём *пространством симплектической геометрии*, или *пространством симплектической линейной алгебры*. Отметим в нём двумерное подпространство — график отображения $T$ в $\mathbb{R}^4$.

**Упражнение.** График отображения $T$ является лагранжевым подпространством, т. е. для любого вектора $\eta$ выполняется равенство $\langle\eta, \xi\rangle = 0$, где $\xi = T\eta$.

Теперь я могу рассказать об аналогах теории рассеяния, возникающих в теории графов. *Граф* $\Gamma$ — это одномерный симплициальный комплекс; у него есть только рёбра и вершины. Мы требуем, чтобы у графа не было двойных рёбер, т. е. пересечение любых двух рёбер либо пусто, либо состоит из одной вершины. Кроме того, мы требуем, чтобы граф не имел концов. Это означает, что если есть хотя бы одно ребро, пришедшее в вершину, то должно быть ещё одно ребро, выходящее из неё. Для вершины $P$ обозначим через $n_P$ число выходящих из этой вершины рёбер. Мы будем предполагать, что $n_P < \infty$, а сами графы иногда будут и бесконечными.

Простейший пример бесконечного графа — дискретизированная прямая $\mathbb{R}$ с отмеченными целыми точками.

**2. Операторы Шредингера на графах.** Есть два оператора Шредингера на графах. Один действует на функциях от вершин, а другой действует на функциях от рёбер.

**Определение.** Оператор Шредингера действует на пространстве функций, зависящих от вершин графа, следующим образом:

$$(L\psi)_P = \sum_{P'} b_{P,P'}\psi_{P'},$$

причём $b_{P,P'} = b_{P',P} \in \mathbb{R}$ и $b_{P,P'} \neq 0$, только если $P = P'$ или $P \cup P' = \partial R$ — граница ребра $R$.

Число $b_{P,P'}$ называют *потенциалом*. Взаимодействовать могут только соседние вершины. В теории вероятностей и в комбинаторике есть тысячи работ, посвященных изучению оператора Шредингера (или, как его ещё называют, оператора второго порядка), действующего на функциях от вершин графа.

Оператор Шредингера, действующий на функциях от рёбер, определяется аналогично:

$$(L\psi)_R = \sum_{R'} d_{R,R'}\psi_{R'},$$

Требование аналогичное: $d_{R,R'} \neq 0$ только для рёбер, являющихся ближайшими соседями друг друга. Ближайшими соседями мы считаем либо совпадающие рёбра, либо рёбра, имеющие общую вершину.

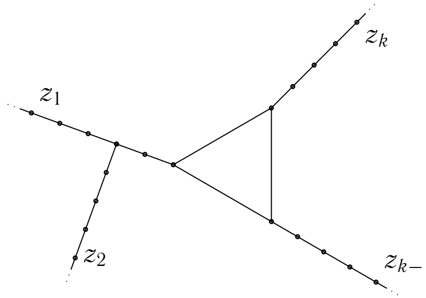



*Рис. 1. Граф с $k$ хвостами*

Рёберные операторы не сводятся к вершинным.

Простейшим примером оператора Шредингера на симплициальных комплексах (и при этом действующих на симплексах любой размерности), являются так называемые операторы Лапласа—Бельтрами, которые интенсивно использовались топологами, начиная со знаменитой работы Зингера с соавторами, посвящённой изучению так называемого кручения Райдемайстера—Рея—Зингера. Так что многомерная ситуация изучалась.

Рассмотрим простейшую ситуацию — дискретизированную прямую. Уравнения Шредингера и Штурма—Лиувилля в вычислительной математике дискретизировали с самого момента её возникновения, и естественно, что дискретное уравнение Шредингера многократно изучалось. Сопоставим ребру $R_n$ и вершине $P_n$ их номер $n$. В случае дискретизированной прямой рёберный и вершинный операторы Шредингера не отличаются друг от друга. Уравнение Шредингера на дискретизированной прямой записывается следующим образом:

$$c_{n-1}\psi_{n-1} + c_{n+1}\psi_{n+1} + v_n\psi_n = \lambda_n\psi_n.$$

В теории солитонов это уравнение интенсивно использовалось в так называемой теории дискретных систем. В классической вычислительной математике предполагалось, что $c_n = 1$ для всех $n$. Это ограничение неудобно. Впоследствии выяснилось, что в теории солитонов и в квантовой физике удобно рассматривать общий класс операторов Шредингера.

На дискретизированной прямой при любом $\lambda$ тоже имеется двумерное пространство решений. Естественно возникает оператор монодромии. Если коэффициенты стремятся к определённым константам на $\pm\infty$, то возникает матрица рассеяния с теми же свойствами. Это мало чем отличается от непрерывной прямой.

Рассмотрим граф $\Gamma$, который имеет $k$ хвостов $z_1, \ldots, z_k$ (рис. 1). Каждый хвост — это дискретизированная полупрямая. Мы потребуем, чтобы вне конечной области $c_n = 1$ и $v_n = 0$, т. е. уравнение Шредингера имеет вид

$$\psi_{n-1} + \psi_{n+1} = \lambda\psi_n. \tag{2}$$

Как устроены решения уравнения (2)? Ответ очень простой: можно взять $\psi_{n,\pm} = a_\pm^n$, где

$$a_\pm = \frac{1}{2}\left(\lambda \pm \sqrt{\lambda^2 - 4}\right).$$

Зона рассеяния (зона непрерывного спектра): $|a_\pm| = 1$ (или, что то же самое, $-2 \leqslant \lambda \leqslant 2$). Именно для этой зоны мы хотим построить рассеяние.

В каждом хвосте $z_j$ выберем решения $\psi_{+,j}$ и $\psi_{-,j}$. Эти решения могут не продолжаться на весь граф. Решения $\psi_\pm$ — это решения типа экспоненты. Ещё можно ввести решения $c_j$ и $s_j$ — аналоги косинуса и синуса. Они получаются так:

$$c_j + a_\pm s_j = \psi_{\pm,j}.$$

Это будет удобный вещественный базис.

Введём пространство асимптотических состояний

$$\mathbb{R}^{2k} = \bigoplus_{j=1}^{k} \mathbb{R}_j^2,$$

где $\mathbb{R}_j^2 = \{c_j, s_j\}$ — пространство решений на $j$-м хвосте.

Рассмотрим следующую величину, которую мы назовём *вронскианом*:

$$W_{\vec{R}}(\varphi, \psi) = b_{P,P'}(\varphi_{P'}\psi_P - \psi_{P'}\varphi_P)$$

для вершинного оператора;

$$W_{\vec{R}}(\varphi, \psi) = \sum_{R' \cap \vec{R} = P} d_{R,R'}(\varphi_{R'}\psi_R - \psi_{R'}\varphi_R)$$

для рёберного оператора. Здесь $\vec{R} = PP'$ — ориентированное ребро.

Что является непрерывным аналогом этой величины? Если имеется оператор Шредингера $-\psi'' + u(x)\psi = \lambda\psi$, то для пары решений $(\varphi, \psi)$ детерминант Вронского $W(\varphi, \psi)$ имеет вид

$$W(\varphi, \psi) = \varphi'\psi - \varphi\psi'.$$

В непрерывном случае основное свойство детерминанта Вронского заключается в том, что $W(\varphi, \psi) = \mathrm{const}$. Для операторов Шредингера на графах есть прямой аналог этой теоремы.

**Теорема 1.** а) $W_{\vec{R}}(\varphi, \psi)$ — *корректно определённая функция пары решений и ориентированного ребра. Эта функция кососимметрическая: она меняет знак при перестановке решений и меняет знак при изменении ориентации ребра*;

б) $\partial W = 0$, *т. е.* $W$ — *цикл.*

Физическая интерпретация того, что $W$ — цикл,— это первый закон Кирхгофа. Если $\Phi = \varphi + i\psi$ интерпретировать как поле, то $-2iW(\varphi, \psi) = j_\Phi$ — ток, и равенство $\partial W = 0$ выражает первый закон Кирхгофа.

Цикл $W$ открытый: если у графа есть бесконечный хвост, то у этого цикла нет конечного носителя. Утверждение о том, что $W$ — цикл, соответствует утверждению о том, что в непрерывном случае вронскиан является константой. Действительно, для прямой любой открытый цикл — это просто прямая с каким-то коэффициентом.

Мы интерпретируем $W$ как кососимметрическое скалярное произведение на парах решений. Значения этого скалярного произведения векторные. Таким образом, мы считаем $W$ элементом одномерной группы гомологий графа $H_1(\Gamma, \mathbb{R})$.

Чтобы это могло быть полезно, нужно, чтобы при данном $\lambda$ было более чем одно решение. Действительно, любая кососимметрическая билинейная форма на одномерном пространстве тождественно равна нулю. Пространство решений очень часто имеет размерность больше 1 в том случае, когда граф обладает симметрией. Второй случай, который я хочу объяснить, это когда у графа никакой симметрии нет, но у него есть $k$ бесконечных хвостов. В этом случае пространство решений не менее чем $k$-мерно.

Я начал этим заниматься, исходя из такого соображения. Классики теории чисел, начиная с Сельберга, и современные геометры, включая Сарнака, рассматривали оператор Лапласа—Бельтрами в разных областях на плоскости Лобачевского. Имеются в виду области, связанные с дискретными группами, у которых фундаментальная область имеет конечную площадь. На плоскости Лобачевского им соответствуют области с конечным числом концов. Ещё И. М. Гельфанд в 50-е годы обращал внимание на результаты Сельберга и говорил, что их полезно перевести на язык теории рассеяния. Следуя идее Гельфанда, это делала ленинградская школа Фаддеева. Но это делалось только для случая, когда хвостов не более двух. М. Громов развивал и рекламировал общую похожую идею, довольно очевидную, впрочем. Если вы смотрите на гиперболическую геометрию из бесконечности, то она предстаёт перед вами как чисто одномерное образование. Чем ближе вершины треугольника к границе, тем меньше его площадь и тем больше он похож на граф. Естественно возникает мысль, что спектральную теорию, в особенности в случае непрерывного спектра операторов Лапласа—Бельтрами, естественно моделировать, используя теорию графов с хвостами.

Пусть имеется граф $\Gamma$ с $k$ хвостами $z_1, \ldots, z_k$. Я только что ввёл пространство асимптотических состояний $\mathbb{R}^{2k} = \bigoplus\limits_{j=1}^{k} \mathbb{R}^2_j$. Это — пространство решений уравнения Шредингера на хвостах в предположении, что коэффициенты оператора стремятся на хвостах к определённым стандартным, которые я только что выписал. А именно,

$$\psi_{\pm,j,n} = c_{j,n} + a_{\pm} s_{j,n}.$$

Здесь $j$ — номер хвоста, $n$ — номер вершины.

С помощью детерминанта Вронского пространство асимптотических состояний можно превратить в симлектическое пространство следующим образом. Мы потребуем, чтобы выполнялось равенство $\langle c_j, s_{j'}\rangle = \delta_{j,j'}$, и введём такое кососимметрическое скалярное произведение. Оно совпадает с вронскианом для двух решений на каждом хвосте.

Теперь обратим внимание на следующее. Пусть есть пара решений $\varphi, \psi$ уравнения $L\psi = \lambda\psi$. Сопоставим (вещественному) решению $\psi$ его асимптотическое значение, являющееся элементом пространства $\mathbb{R}^{2k}$:

$$\psi \mapsto \psi^{as} \in \mathbb{R}^{2k}.$$

Это делается следующим образом. Если есть решение уравнения Шредингера на графе, то оно чему-то равно на каждом хвосте. Мы берём прямую сумму этих состояний. Я сейчас рассматриваю вещественные решения, но можно и всё комплексифицировать и рассматривать комплексные решения тоже.

**Теорема 2.** *Пусть $L\psi = \lambda\psi$ и $L\varphi = \lambda\varphi$. Тогда $\langle \varphi^{as}, \psi^{as}\rangle = 0$.*

Эта теорема означает следующее. Если два набора асимптотических значений, заданных произвольно в каждом хвосте, продолжаются до решений уравнения Шредингера на всем графе, то их скалярное произведение всегда обращается в нуль. В действительности это свойство полностью определяет все свойства унитарности процесса рассеяния. Иными словами, как говорят симплектические геометры, множество всех асимптотических значений реально существующих решений, образует лагранжево подпространство размерности $k$ в пространстве размерности $2k$.

Этот факт, по сути дела, является чисто топологическим. Пусть есть граф $\Gamma$ с $k$ хвостами $z_1, z_2, z_3, \ldots, z_k$. Рассмотрим пару решений $L\psi = \lambda\psi$ и $L\varphi = \lambda\varphi$. Неважно, являются рассматриваемые операторы Шредингера рёберными или вершинными: на хвостах различия между ними нет. Составим вронскиан $W$. Его граница равна нулю.

**Упражнение.** Из того, что вронскиан $W$ — цикл, следует, что он представляется в виде

$$W = \sum_{i=1}^{k} \alpha_i z_i + \text{нечто конечное},$$

т. е. если какое-то ребро хвоста вошло в $W$ с каким-то коэффициентом, то и все остальные рёбра этого хвоста войдут в $W$ с тем же самым коэффициентом.

Заметим также, что один хвост никогда не продолжается до цикла на всём графе. Могут продолжаться только разности хвостов, причём разными способами. Разности этих продолжений являются какими-то конечными циклами в графе.

**Упражнение.** $W = \sum\limits_{j=2}^{k} \beta_j(z_1 - z_j) +$ нечто конечное.

Если сопоставить эти два выражения, то можно увидеть следующее: $\sum \alpha_j = 0$. Действительно, $\alpha_1 = \sum_{j=2}^{k} \beta_j$ и $\alpha_j = -\beta_j$ при $j > 1$. Равенство $\sum \alpha_j = 0$ эквивалентно тому, что $\langle \varphi^{as}, \psi^{as} \rangle = 0$. Таким образом, тот факт, что в окрестности бесконечности асимптотические решения образуют лагранжеву плоскость, является фундаментальным. Дальше легко показать, что эта плоскость задаётся $k$ уравнениями, поэтому её размерность не меньше $k$. С другой стороны, лагранжева плоскость не может иметь размерность больше $k$.

Поэтому оператор, сопоставляющий решениям их асимптотические значения, при любом $\lambda$ задаёт $k$-мерное подпространство. У него может быть ядро: это такие собственные функции, у которых сосредоточены тождественные нули во всех хвостах. Они имеют интересные аналоги в геометрии, связанной со спектром, который изучался, начиная с Сельберга. Для графов общего положения их можно ликвидировать малым возмущением. Но для графов с групповой симметрией их ликвидировать нельзя.

Матрица рассеяния здесь строится таким образом. Выберем базис векторов лагранжевой плоскости в следующем виде:

$$\psi_j = \psi_{+,j} + \sum_q s_{jq}(\lambda)\psi_{-,j}.$$

Тогда $\big(s_{jq}(\lambda)\big)$ — матрица рассеяния (по определению).

Если в лагранжевом пространстве выбран чисто вещественный базис, то матрица рассеяния $S$ унитарная и симметрическая. Это можно пояснить следующим образом. Возьмём унитарную матрицу $A$ и запишем матрицу $S$ в виде $S = AA^t$. Посмотрим, сколько таких матриц. Если $A$ заменить на $AO$, где матрица $O$ принадлежит вещественной ортогональной группе, то $AA^t$ заменится на $AOO^tA^t = AA^t$. Тем самым, $S \in U_k/O_k$. Известно, что множество всех лагранжевых плоскостей изоморфно именно $U_k/O_k$.

Дискретизированная прямая с полным набором коэффициентов впервые появилась при интегрировании так называемой цепочки Тодда, и выяснилось, что такой оператор Шредингера обладает гораздо лучшими свойствами, чем те его частные случаи, которые рассматривались в вычислительной математике.

# Что такое НМУ

## Как всё началось

Летом 1991 года в московской школе № 57 собрались около тридцати математиков и приняли решение об организации Математического колледжа — нового негосударственного высшего учебного заведения, ориентированного на подготовку математиков-исследователей. Инициатива принадлежала Николаю Николаевичу Константинову — знаменитому организатору математических классов и пользовавшихся большой популярностью соревнований. Разработкой программ обучения и определением стратегических направлений развития Колледжа занялся научный совет, куда вошли крупнейшие российские математики: академики РАН В. И. Арнольд, С. П. Новиков, Я. Г. Синай, Л. Д. Фаддеев, профессора А. А. Бейлинсон, Р. Л. Добрушин (ныне покойный), Б. А. Дубровин, А. А. Кириллов, А. Н. Рудаков, В. М. Тихомиров, А. Г. Хованский, М. А. Шубин. Председателем совета стал В. И. Арнольд. Создание колледжа сразу поддержали американские ученые П. Делинь и Р. Макферсон.

Математический колледж и созданный несколько ранее Высший колледж математической физики образовали Независимый московский университет (НМУ). Первым ректором университета стал М. К. Поливанов, а Н. Н. Константинов возглавил Координационный совет. Начинать обучение было решено сразу же, и в сентябре 1991 года в НМУ появились первые студенты.

В настоящее время ректор НМУ — профессор Юлий Сергеевич Ильяшенко. Он же возглавляет Высший Колледж математики (так теперь называется Математический колледж). Деканом Высшего колледжа математической физики является профессор А. И. Кириллов. НМУ расположен в самом центре Москвы — в здании Московского центра непрерывного математического образования по адресу Б. Власьевский пер., д. 11. Тел. (095) 241-40-86.

Университет преследует следующие цели:

- подготовка математиков-исследователей и физиков-теоретиков;
- привлечение активно работающих исследователей к преподаванию;
- сохранение и развитие лучших традиций московской математической школы;
- содействие интеграции отечественной науки в мировое научное сообщество;
- пропаганда естественно-научных знаний.

## Ещё немного истории

Начиная с сентября 1991 года в течение пяти лет занятия НМУ происходили по вечерам в помещениях московских школ в районе Московского государственного университета, большей частью в помещении известной «Второй школы». Почти все студенты были одновременно студентами других вузов (в основном мехмата) — в частности, потому, что НМУ не предоставляет отсрочку от призыва в армию. Это означало для студентов двойную нагрузку. Неудивительно, что далеко не все могли с этим справиться. Скорее стоит удивляться, что хоть кто-то это выдерживает и НМУ продолжает существовать.

## Преподаватели

Научный уровень профессорско-преподавательского состава НМУ не уступает уровню любого из мировых научных центров. Это подтверждается большим количеством исследовательских статей и монографий, опубликованных преподавателями, в том числе в последние годы. Среди пленарных докладчиков на Международных Математических Конгрессах — высших форумах математической науки, проводимых раз в четыре года, — профессора Университета В. И. Арнольд, С. П. Новиков, В. А. Васильев, Б. Л. Фейгин. Профессора Ю. С. Ильяшенко и А. Н. Рудаков были секционными докладчиками на Конгрессах. В. И. Арнольд был также пленарным докладчиком на 1-м Европейском Математическом Конгрессе в 1992 году в Париже. Он является лауреатом Краффордской премии, задуманной как аналог Нобелевской премии для ученых, специальности которых не включены Нобелем в его завещание. Профессора университета занимают лидирующие позиции в мире в таких областях как теория дифференциальных уравнений и динамических систем, теория особенностей (катастроф), теория представлений, топология и другие.

Исследовательские проекты сотрудников университета регулярно получают гранты Российского Фонда Фундаментальных Исследований, международных научных фондов.

Основные математические курсы в НМУ в разное время читали Д. В. Аносов, В .И. Арнольд, А. А. Белавин, В. К. Белошапка, Ю. М. Бурман, В. А. Васильев, Э. Б. Винберг, А. Л. Городенцев, С. М. Гусейн-Заде, Ф. Л. Зак, Ю. С. Ильяшенко, М. Э. Казарян, А. А. Кириллов, С. К. Ландо, С. М. Натанзон, С. П. Новиков, И. М. Парамонова, Г. Л. Рыбников, А. Г. Сергеев, А. Б. Сосинский, В. М. Тихомиров, Б. Л. Фейгин, М. В. Финкельберг, А. Г. Хованский, О. В. Шварцман, О. К. Шейнман и другие.

В традициях московской математической школы научная работа в НМУ концентрируется вокруг исследовательских семинаров. В 1999 году начал работу общеуниверситетский семинар «Глобус», выступления на котором предназначены для широкой аудитории и призваны развивать и укреплять междисциплинарные связи.

## Учебный план

Первые два года обучения включают следующие курсы:

- современный курс анализа, включая анализ на многообразиях (три семестра),
- линейная алгебра, группы, кольца, поля, модули, алгебры, теория Галуа, теория представлений (четыре семестра),
- гиперболическая и проективная геометрия (один семестр),
- дифференциальная геометрия (два семестра),
- теория функций комплексного переменного (два семестра),
- основы топологии (два семестра).

В течение следующих трех лет обучения студенты должны изучить определённое число специальных курсов. В колледже читались курсы алгебраической топологии, алгебраической геометрии, гомологической алгебры, динамических систем, дифференциальных уравнений, теории особенностей, теории представлений, теории чисел, функционального анализа и многие другие. Материалы прочитанных оригинальных курсов, как правило, публикуются.

## Студенты и вольнослушатели

Обучение в НМУ бесплатное. Прием происходит по результатам серьезных вступительных экзаменов, что позволяет отобрать для обучения способную молодежь с установившимся интересом к естественным наукам. В настоящее время на 1—5 курсах обучается около 90 студентов и вольнослушателей (университет открыт для всех желающих заниматься). Начиная с третьего курса, а иногда и раньше, студенты ведут самостоятельную научную работу под руководством преподавателей. Результаты такого подхода налицо: в 1998 году Американское математическое общество опубликовало сборник статей, который составили труды преподавателей, аспирантов и студентов Высшего Колледжа математики.

В 1996 году состоялся первый выпуск студентов, прошедших полный курс обучения. К настоящему времени уже 20 человек получили диплом Высшего коллежда математики.

## Аспиранты

В 1993 году Высший колледж математики НМУ совместно с Московским Математическим институтом начал свою аспирантскую программу. В настоящее время в Колледже 18 аспирантов. Для руководства аспирантами приглашаются не только преподаватели Колледжа, но и другие ведущие ученые-математики. Каждый аспирант имеет опубликованные научные работы. Аспиранты активно участвуют в преподавании. Многие из них читают собственные специальные, а иногда и обязательные курсы. Пятнадцать выпускников аспирантуры Колледжа уже защитили кандидатские диссертации.

## Международные связи

В 1994 году Независимый университет заключил договор о дружбе и сотрудничестве с ведущим высшим учебным заведением Франции — Нормальной школой в Париже. Программа научных обменов студентами и аспирантами, начатая осенью 1994 года, предусматривает ежегодные взаимные поездки групп учащихся.

Университет поддерживает тесные научные контакты с Французским математическим обществом и Американским математическим обществом (AMS). В частности, библиотека университета получает все научные журналы, издаваемые AMS. В 1998 году заключен также договор о сотрудничестве с университетом города Ренна (Франция).

Международные научные контакты сотрудников университета настолько многообразны, что не поддаются даже поверхностному перечислению. Студенты, аспиранты, молодые преподаватели привлекаются к работе школ и конференций, проводимых по всему миру.

## Московский центр непрерывного математического образования

В 1995 году Независимый университет, совместно с Московским Государственным Университетом, Отделением математики РАН, Математическим институтом им. В. А. Стеклова РАН, Префектурой Центрального административного округа г. Москвы, Московским департаментом образования выступил соучредителем Московского Центра непрерывного математического образования (директор Иван Валерьевич Ященко). В Попечительский Совет Центра, возглавляемый В. И. Арнольдом, входят зам. директора Математического института им. В. А. Стеклова РАН академик А. А. Болибрух, проректор зам. министра образования РФ проф. В. В. Козлов, министр правительства Москвы А. И. Музыкантский, академик-секретарь отделения математики РАН Л. Д. Фаддеев и другие. С сентября 1996 года НМУ располагается в здании Центра, где находятся совместная библиотека и компьютерный класс.

Библиотека своевременно получает ряд ведущих математических журналов. Издательство НМУ со времени своего создания в 1991 году опубликовало не менее тридцати изданий курсов лекций, специальных курсов, прочитанных в университете. Оно не только обеспечивает потребности студентов в современной учебной литературе, но и способствует широкому распространению научных знаний.

Учащиеся и сотрудники университета могут пользоваться компьютерным классом Центра, который постоянно подключен к сети Интернет. Они также могут получить домашний адрес электронной почты для частной переписки.

### Независимый университет и среднее образование

Университет имеет самые тесные связи со школьным образованием, специальными школами, школьными олимпиадами в Москве, в России в целом. Он активно участвует в пропаганде естественнонаучных знаний среди учащихся, формируя, в частности, базу своего дальнейшего развития. Работу со школьниками возглавляет Н. Н. Константинов.

---

## Семинар «Глобус»

Проходит (как правило) раз в две недели, по четвергам, в $15^{\underline{40}}$, в конференц-зале НМУ. Более подробно см. www.mccme.ru/ium/globus.html

Темы докладов в весеннем семестре 1999/2000 г.

16 марта *Ю. С. Ильяшенко*
**Столетняя история 16 проблемы Гильберта (о предельных циклах)**

23 марта *О. И. Богоявленский*
**Диофантовы уравнения и задачи анализа, возникающие в теории астрофизических струй**

6 апреля *В. А. Васильев*
**Ветвящиеся интегралы и теория Пикара—Лефшеца**

20 апреля *Б. Л. Фейгин*
**Вертексные алгебры и их применения в алгебраической геометрии**

4 мая *М. А. Цфасман*
**Алгебраическая геометрия и теория чисел в задаче о плотной упаковке шаров в $\mathbb{R}^n$**

18 мая *В. Иврий*
**Все началось с Вейля**

29 мая *Ю. И. Манин*
**Некоммутативная геометрия и квантовые тэта-функции**

7 июня *Я. Г. Синай*
**Динамика адиабатического поршня (нарушение второго закона термодинамики)**

---

# Содержание

Студенческие чтения НМУ
Выпуск 1

Научный редактор *В. Прасолов*
Редактор *Ю. Торхов*
Верстка *В. Радионов*

Издательство Московского центра
непрерывного математического образования

Лицензия ИД №01335 от 24.03.2000 г.
Подписано в печать 25.07.2000 г. Формат $60 \times 90\ 1/16$
Бумага офсетная №1. Печать офсетная. Печ. л. 16,0
Тираж 1000. Заказ №

МЦНМО
121002, Москва, Большой Власьевский пер., 11

Вы можете приобрести книги издательства МЦНМО
в «Математическом библиоклубе» по адресу
Большой Власьевский пер., д. 11.
Тел. (095) 241–72–85. E-mail: mbc@mccme.ru