

МАТЕМАТИКА

НОВОЕ В ЗАРУБЕЖНОЙ НАУКЕ

РЕДАКТОРЫ СЕРИИ: Ю. И. МАНИН, С. П. НОВИКОВ

50

ТРУДЫ СЕМИНАРА
Н. БУРБАКИ
за 1992 г.

СБОРНИК СТАТЕЙ

Перевод с английского
и французского
под редакцией
В. Л. ПОПОВА



МОСКВА «МИР» 2001

УДК 517+514
ББК 22.16+22.151
Т77

Т77 Труды семинара Н. Бурбаки за 1992 г.: Сб. статей: Пер. с англ. и франц.— М.: Мир, 2001. — 509 с., ил.

ISBN 5-03-003326-2

Продолжение публикации трудов известного семинара Н. Бурбаки, начатой издательством «Мир» в 1990 г. В очередной выпуск включены доклады, посвященные новейшим достижениям в различных областях математики: алгебраической геометрии, современной математической физики, теории динамических систем и др. Среди авторов такие известные французские математики, как Э. Гис, Ж.-К. Йоккоз, Ж.-П. Серр, Ж.-М. Фонтен и др.

Для математиков разных специальностей, аспирантов и студентов.

ББК 22.16+22.151



Издание осуществлено при поддержке
Российского фонда фундаментальных исследований
по проекту № 97-01-14014

Редакция литературы по математическим наукам

© Société Mathématique de France, 1992
Tous droits réservés
© состав, В. Л. Попов, 2001
© перевод на русский язык, «Мир», 2001

ISBN 5-03-003326-2 (русск.)

Настоящий сборник продолжает традицию изданий на русском языке трудов семинара Н. Бурбаки (сборники «Труды семинара Н. Бурбаки за 1988 г.», «Труды семинара Н. Бурбаки за 1989 г.», «Труды семинара Н. Бурбаки за 1990 г.» и «Труды семинара Н. Бурбаки за 1991 г.» были выпущены в 1990 г., 1991 г., 1996 г. и 1998 г. соответственно).

В этот сборник включены доклады, сделанные на семинаре в ноябре 1991 г.— июне 1992 г. Они посвящены новейшим ярким достижениям в следующих областях: динамика потоков на однородных пространствах, гипергеометрические функции, гиперкэлеровы многообразия, дифференциальная теория Галуа, алгебраическая K -теория, арифметическая алгебраическая геометрия, дифференциальные уравнения, вопросы математической физики, минимальные поверхности в евклидовом пространстве, случайные операторы и матрицы, динамические системы.

Издание этого сборника, без сомнения, поможет российским специалистам ознакомиться с достижениями мировой науки.

Пользуясь случаем, выражаю признательность авторам, любезно прочитавшим рукописи переводов своих статей и приславшим замечания. Хочется также поблагодарить всех, кто оказал нам помощь при подготовке настоящего издания.

В. Л. Попов

БАБОЧКА ХОФСТАДТЕРА

(по Б. Хельферу и Ж. Шёстранду)

Жан Беллиссар

1. ЭЛЕКТРОНЫ БЛОХА В МАГНИТНОМ ПОЛЕ

Бабочка Хофстадтера, которой посвящена эта публикация, возникает при изучении движения электрона в однородном магнитном поле. Это одна из тех проблем, которые привлекали в этом веке интерес наибольшего числа специалистов в области физики твердого тела. Ей посвящено более 200 публикаций ведущих специалистов в этой области. Мы не можем их все здесь перечислить, хотя история этого вопроса, без сомнения, заслуживает освещения. Однако на некоторых моментах полезно остановиться.

Первая публикация по этой теме принадлежит Л. Д. Ландау [LA], который в 1930 г. впервые вычислил спектр энергии свободного электрона в однородном магнитном поле. Пайерлс [PE] в 1933 г. заложил основы теории электронов в кристалле для равномерно намагниченного металла, которая показывала, что вклад электронов проводимости в магнитную чувствительность является диамагнитным. В кристалле размерности D при отсутствии магнитного поля и без учета кулоновских взаимодействий электроны проводимости описываются как независимые частицы, энергию которых можно представить одной или несколькими функциями $E(\vec{k})$ вектора «квазиимпульса» $\vec{k}$, периодического относительно группы симметрии обратной решетки рассматриваемого кристалла. Квазиимпульс представляет, с точностью до некоторой физической константы, импульс частицы. Эти функции называются функциями зоны. Рассмотрения в рамках статистики Ферми—Дирака показывают, что только носители заряда с энергией, близкой к уровню Ферми, дают вклад в проводимость. Таким образом, в простейшем случае достаточно одной функции зоны, тогда как в общем случае требуется конечное число таких функций, но редко более трех. Зная, что лабораторные магнитные поля всегда слишком слабы, чтобы давать заметный эффект, Пайерлс предложил описывать движения электрона в однородном магнитном поле оператором Га-

мильтона H , полученным подстановкой вместо квазиимпульса $\vec{k}$ в функцию зоны оператора $\vec{K} = (\vec{P} - e\vec{A})/\hbar$, где $\vec{P} = -\frac{i}{\hbar}\vec{\nabla}$ — оператор импульса, а $\vec{A}$ — векторный магнитный потенциал, связанный с магнитным полем $\vec{B}$ соотношением $\text{rot}(\vec{A}) = \vec{B}$.

Первая проверка этой подстановки стала предметом статьи Латтингера [LU] в 1950 г., который применил к этой проблеме метод, разработанный Слэтером для исследования примесей в кристалле. Эта статья послужила началом огромного потока работ, написанных в пятидесятые годы и посвященных конструкциям эффективных гамильтонианов, более или менее реально отражающих влияние магнитного поля. В одной из этих работ, статье Харпера [Ha] 1955 г., предлагалась упрощенная модель, способная дать качественно корректное описание движения носителей заряда. Зная, что нетривиально только движение, перпендикулярное магнитному полю, Харпер ограничился рассмотрением двумерной проблемы, в которой $\vec{K} = (K_1, K_2)$. Кроме того, в разложении в ряд Фурье функции зоны он оставил только младшие члены. Для квадратной ячейки он получил следующий гамильтониан:

$$H = 2t (\cos(K_1) + \cos(K_2)), \quad (1)$$

где величина t (названная трансфертом или обменным интегралом) представляет энергию, необходимую электрону для перехода с одного энергетического уровня кристалла на другой. Замечательной особенностью этого выражения является то, что этот оператор, как, впрочем, и другие, сконструированные по той же методике, зависит только от «магнитных трансляций» [ZA], а именно от двух унитарных операторов $U_k = e^{iK_i}$ ($k = 1, 2$), удовлетворяющих следующим условиям коммутирования:

$$U_1 U_2 = e^{2i\pi\alpha} U_2 U_1, \quad \alpha = \frac{\phi}{\phi_0}; \quad (2)$$

здесь $\phi_0 = h/e$ — квант магнитного потока, h — постоянная Планка и e — заряд электрона. Эти соотношения являются каноническими соотношениями коммутирования при условии, что постоянная Планка $\hbar$ заменяется нормированным потоком $2\pi\alpha$. Однако для реального кристалла с элементарной ячейкой размером в несколько Å, помещенного в поле величиной в несколько тесл, этот нормированный поток имеет порядок от 10^{-7} до 10^{-9} . В этом случае хорошо работает квазиклассическое приближение.

Математическая проверка подстановки Пайерлса была проделана в работах [BE1, HS0], где было доказано, что электроны с энер-

гией, близкой к уровню Ферми, точно описываются эффективным гамильтонианом, выраженным в виде конечной матрицы, элементами которой являются сходящиеся ряды мономов от магнитных трансляций. Таким образом, модель Харпера, по-видимому, дает аппроксимацию в простейшей зоне эффективного гамильтониана, описывающего электроны проводимости.

Вычисление спектра модели Харпера имеет долгую историю. Действительно, присутствие магнитного поля нарушает инвариантность при трансляции, требуя обращения к теореме Блоха при вычислениях. Квазиклассический метод использовался с конца пятидесятых годов для описания влияния кристаллической решетки на уровень Ландау. Однако физики заметили, что, когда нормированный поток рационален, гамильтониан становится периодическим, так что в этом случае возможно использование теории Блоха. Таким образом, аппроксимируя α подходящей последовательностью рациональных выражений, можно вычислять спектр с возрастающей точностью. Эта программа была впервые осуществлена Шамберсом [CH] в 1965 г., но он вычислил спектр только для малого числа значений потока, надеясь на то, что нет необходимости идти дальше, при этом упустив важнейший аспект этой проблемы, а именно фрактальный характер спектра как функции потока. Пришлось ждать 1976 г. и диссертации Хофstadтера [HO], чтобы впервые увидеть структуру (рис. 1).... в форме фрактальной... бабочки.

Примечательно, что модель Харпера или ее производные перестраивались и использовались в самых разных физических ситуациях многими другими физиками. В 1978 году Обри [AU] использовал ее для описания физики одномерных спонтанно модулируемых цепей для явления, названного «переходом Пайерлса». В восьмидесятые годы в серии работ Геннеса [GE] и Александера [AL] по решеткам сверхпроводников Раммал и его многочисленные сотрудники теоретически и экспериментально доказали, что модель Харпера позволяет описывать кривую перехода металла из нормального в сверхпроводящее состояние для квадратной решетки из нитей сверхпроводников. В частности, квантование потока рациональными значениями квантов потока было доказано экспериментально [PCR]. Эта модель была также использована для понимания теории квантового эффекта Холла [TKN2], открытого в 1981 г. Клицингом и др. [KL]. Наконец, совсем недавно эта модель и ее производные создали необходимую базу для понимания теории высокотемпературной сверхпроводимости сверхпроводящих оксидов

Спектр квадратной решетки

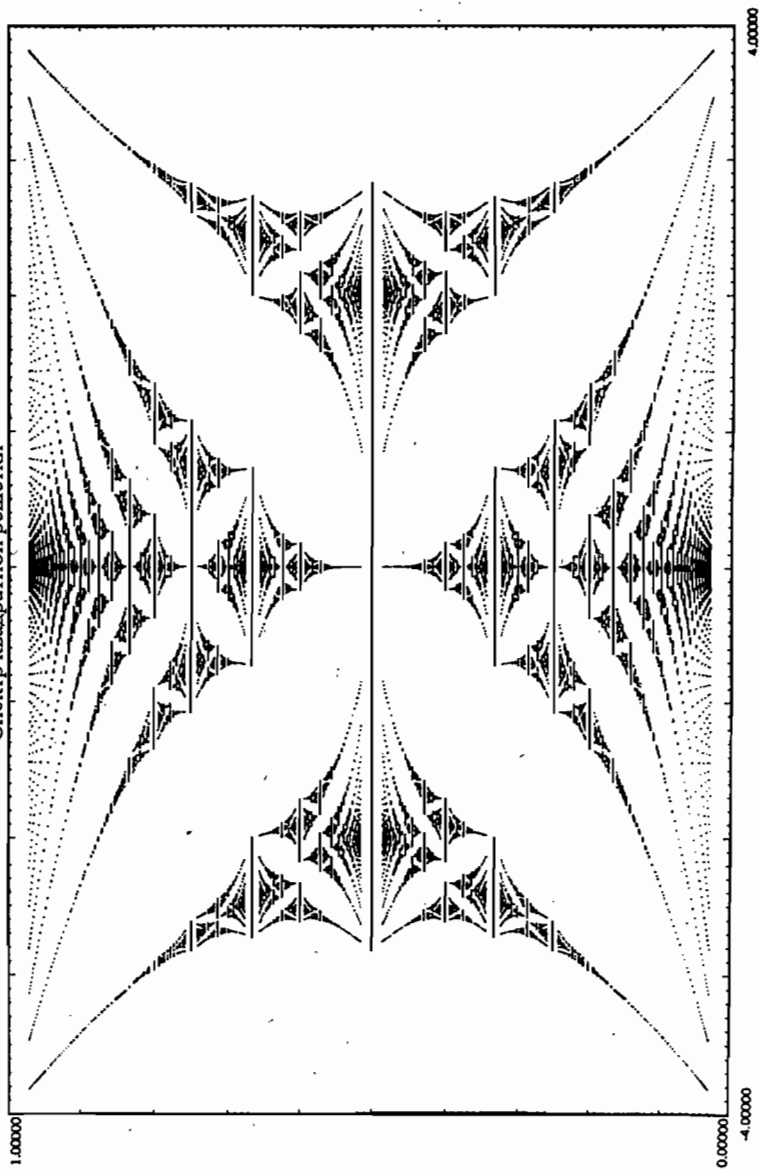


Рис. 1.

меди [RB1]. Заинтересованный читатель может найти дополнительную информацию в [BE2].

2. ЭЛЕМЕНТАРНЫЕ СВОЙСТВА МОДЕЛИ ХАРПЕРА

В этом разделе мы рассмотрим следующую модификацию модели Харпера:

$$H(\alpha, \mu) = U_1 + U_1^{-1} + \mu(U_2 + U_2^{-1}), \quad (3)$$

где операторы U_i унитарны и удовлетворяют уравнению (2) при $\alpha = \phi/\phi_0$, а μ — действительное число. Эта модель в дальнейшем будет называться дискретной моделью Матье. Модель Харпера соответствует случаю $\mu = 1$. Напомним, что спектром оператора H называется множество комплексных чисел z , при которых оператор $z - H$ не имеет ограниченного обратного. Спектр является замкнутым подмножеством комплексной плоскости. Для произвольного (соответственно самосопряженного, унитарного) оператора спектр содержится в $\mathbb{C}$ (соответственно в $\mathbb{R}$, в единичной окружности), и мы будем называть лакунами, или щелями связанные компоненты дополнения к спектру в $\mathbb{C}$ (соответственно в $\mathbb{R}$, на единичной окружности). Напомним также, что канторово множество замкнуто, нигде не плотно и не имеет изолированных точек.

Следующие результаты — краткий перечень простейших свойств этой модели.

Теорема 2.1 (дуальность Обри, см. [AA, AU, BES]). *Спектры операторов $H(\alpha, \mu)$, $\mu H(\alpha, \frac{1}{\mu})$, $H(\alpha, -\mu)$ и $H(1 \pm \alpha, \mu)$ совпадают.*

Теорема 2.2 (см. [BES, MO, CEY]). *Если $\alpha = p/q$ — рациональное число и $\mu \neq 0$, то спектр оператора $H(\alpha, \mu)$ состоит из q зон, разделенных лакунами. Если q нечетно, то лакуны открыты. Если q четно, то все лакуны, кроме центральной, соответствующей нулевой энергии, открыты.*

Теорема 2.3 (см. [BES, MO, CEY]). *На отрезке $[0, 1]$ существует плотное подмножество Ω типа G_δ , такое, что если $\alpha \in \Omega$ и если $\mu \neq 0$, то спектр оператора $H(\alpha, \mu)$ — канторово множество.*

Теорема 2.4 (формула Обри, см. [AA, TH, AMS]). *Если число α иррационально и $\mu \neq 1$, то мера Лебега оператора $H(\alpha, \mu)$ равна $4|1 - \mu|$.*

Теорема 2.5 (см. [SI]). *Существует число $\mu_c > 0$, такое, что для $\mu \geq \mu_c$ спектр оператора $H(\alpha, \mu)$ является канторовым множеством для почти всех α .*

3. АЛГЕБРАИЧЕСКИЕ АСПЕКТЫ

3.1. Введение унитарных операторов U_1 и U_2 в разд. 1 позволяет дать описание оператора Харпера и его производных в чисто алгебраических терминах. Действительно, большинство результатов относительно спектра как множества зависит только от соотношения коммутирования (2) и совершенно не зависит от рассматриваемой модели. Эти соображения систематически развивались в статьях [RB2, BE3].

Обозначим через $\mathcal{P}_\alpha$ комплексную алгебру, порожденную двумя унитарными операторами U_1 и U_2 , удовлетворяющими соотношению коммутирования (2). Введем обозначение

$$W(m) = U_1^{m_1} U_2^{m_2} e^{-i\pi \alpha m_1 m_2} \quad (4)$$

для $m = (m_1, m_2) \in \mathbb{Z}^2$. Тогда эта алгебра состоит из полиномов вида

$$a = \sum_{m \in \mathbb{Z}^2} a_m W(m), \quad (5)$$

где $a_m \in \mathbb{C}$ мы будем называть коэффициентами Фурье оператора a и где почти все a_m равны нулю. Коэффициенты Фурье произведения ab даются формулой

$$(ab)_m = \sum_{m' \in \mathbb{Z}^2} a_{m'} b_{m-m'} e^{i\pi \alpha m' \wedge m}, \quad (6)$$

где $m' \wedge m = m'_1 m_2 - m'_2 m_1$, а коэффициенты Фурье для a^* , сопряженного к a , — формулой

$$(a^*)_m = \overline{a_{-m}}. \quad (7)$$

Единичный оператор тогда задается рядом Фурье $1_m = \delta_{m,0}$, в то время как унитарные операторы U_1 и U_2 задаются рядами $\delta_{m,(1,0)}$ и $\delta_{m,(0,1)}$ соответственно.

Конструкция алгебры $\mathcal{P}_\alpha$ показывает, что эта алгебра изоморфна алгебре $\mathcal{P}_{\alpha+1}$, что позволяет ограничить множество значений α интервалом $[0,1]$.

Мы можем считать α переменной, рассматривая для каждого компакта I из $\mathbb{R}$ алгебру $\mathcal{P}(I)$, определенную по аналогии с тем,

как это было сделано выше, с помощью конечных рядов Фурье; коэффициенты которых являются непрерывными функциями переменной α на I , а произведения и сопряженные для которых задаются формулами (6) и (7). Алгебра $\mathcal{P}_\alpha$ получается, если положить $I = \{\alpha\}$.

Тогда существует естественный гомоморфизм $\eta_\alpha : \mathcal{P}(I) \rightarrow \mathcal{P}_\alpha$, получающийся вычислением коэффициентов Фурье в точке α .

Определим линейную форму τ следующей формулой:

$$\tau(\alpha) = a_0. \quad (8)$$

Она принимает значения в пространстве $C(I)$ непрерывных функций на I . Эта форма является «следом», т. е. удовлетворяет следующим условиям:

$$\tau(a^*a) \geq 0, \quad \tau(ab) = \tau(ba), \quad a, b \in \mathcal{P}(I). \quad (9)$$

3.2. Представление π алгебры $\mathcal{P}(I)$ определяется заданием гильбертова пространства $\mathcal{H}_\pi$ и гомоморфизма $\pi : \mathcal{P}(I) \rightarrow \mathcal{B}(\mathcal{H}_\pi)$, такого, что $\pi(a^*) = \pi(a)^*$. Здесь $\mathcal{B}(\mathcal{H}_\pi)$ — алгебра ограниченных операторов на $\mathcal{H}_\pi$. Мы не будем различать унитарно эквивалентные представления. В этом случае классы унитарно эквивалентных представлений образуют множество, обозначаемое $\text{Rep}(\mathcal{P}(I))$. Норма в смысле C^* -алгебры определяется на $\mathcal{P}(I)$ формулой

$$\|a\| = \sup_{\pi \in \text{Rep}(\mathcal{P}(I))} \|\pi(a)\|. \quad (10)$$

Алгебра $\mathcal{A}(I)$ получается пополнением алгебры $\mathcal{P}(I)$ по этой норме. Легко доказать, что след, гомоморфизм η_α и все представления алгебры $\mathcal{P}(I)$ единственным образом продолжаются по непрерывности на алгебру $\mathcal{A}(I)$ [ВЕЗ].

Семейство представлений π называется точным, если пересечение ядер его элементов равно нулю. Если семейство состоит из одного элемента π , то мы будем говорить, что представление π точно. Интерес к подобным семействам объясняется следующим их свойством:

Лемма 3.1 (см. [PED]). Пусть $\{\pi_s\}_{s \in S}$ — точное семейство представлений алгебры $\mathcal{A}(I)$, индексированное множеством S . Тогда спектр любого элемента $a \in \mathcal{A}(I)$ является замыканием объединения по S спектров операторов $\pi_s(a)$, т. е.

$$\text{Sp}(a) = \overline{\bigcup_{s \in S} \text{Sp}(\pi_s(a))}. \quad (11)$$

Кроме того, норма элемента a задается формулой

$$\|a\| = \sup_{s \in S} \|\pi_s(a)\|. \quad (12)$$

Чаще всего используются следующие представления алгебры $\mathcal{A}_\alpha$.

(i) Определенное в разд. 1 представление магнитного поля $\pi_{M,\alpha}$ действует на пространстве $L^2(\mathbb{R}^2)$ следующим образом:

$$\pi_{M,\alpha}(U_i) = e^{iK_i}, \quad K_i = \frac{\delta}{h} \left(\frac{\hbar}{i} \frac{\partial}{\partial x_i} - eA_i \right), \quad (13)$$

где e — заряд электрона, $\mathbf{A} = (A_1, A_2)$ — векторный потенциал, задаваемый магнитным полем B посредством соотношения $\partial_1 A_2 - \partial_2 A_1 = B = \text{const}$, а δ — период рассматриваемой решетки. В данном случае α — отношение потока в квадратной ячейке со стороной δ к кванту потока h/e (см. (2)). Это представление точное [ВЕЗ].

(ii) Представление $\pi_{\text{ГНС}}$, определенное на пространстве $l^2(\mathbb{Z}^2)$ соотношением

$$\pi_{\text{ГНС}}(U_i)\psi(m) = e^{i\pi\alpha e_i \wedge m} \psi(m - r_i), \quad m \in \mathbb{Z}^2, \quad (14)$$

где $\psi \in l^2(\mathbb{Z}^2)$, а r_i ($i = 1, 2$) — канонические базисные векторы в $\mathbb{Z}^2$. Представление $\pi_{\text{ГНС}}$ тоже точное. ГНС — аббревиатура от Гельфанд—Наймарк—Сигал, которые сопоставляли каждому положительному линейному функционалу на C^* -алгебре естественное представление [РЕД]. В нашем случае таким функционалом является след. Кроме того, это представление соответствует тому, что физики называют «представлением сильной связи», где носители заряда рассматриваются как частицы на решетке $\mathbb{Z}^2$, а не непрерывно размазанными в пространстве.

(iii) Семейство $\{\pi_{\alpha,x}; x \in \mathbb{R}/\mathbb{Z}\}$ действует на пространстве $l^2(\mathbb{Z})$ следующим образом (здесь $\varphi \in l^2(\mathbb{Z}^2)$):

$$\pi_{\alpha,x}(U_1)\varphi(n) = \varphi(n-1), \quad \pi_{\alpha,x}(U_2)\varphi(n) = e^{2i\pi(x-n\alpha)}\varphi(n), \quad n \in \mathbb{Z}. \quad (15)$$

В частности, дискретный оператор Матье $H(\mu)$ (см. (3)) действует на этом представлении следующим образом:

$$\pi_{\alpha,x}(H(\mu))\varphi(n) = \varphi(n+1) + \varphi(n-1) + 2\mu \cos 2\pi(x-n\alpha)\varphi(n). \quad (16)$$

Именно в этой форме Харпер [НА] первоначально задал свое уравнение в случае $\mu = 1$. Это семейство точно для всех α . Кроме того, если α иррационально, то каждое из представлений семейства точно, что неверно в случае, если α рационально.

(iv) Представление Вейля $\pi_{W,\alpha}$ определено на пространстве $L^2(\mathbf{R})$ следующим образом (здесь $f \in L^2(\mathbf{R})$):

$$\pi_{W,\alpha}(U_1)f(x) = f(x + 2\pi\alpha), \quad \pi_{W,\alpha}(U_2)f(x) = e^{ix}f(x), \quad x \in \mathbf{R}. \quad (17)$$

Это представление точно. На самом деле именно оно использовалось такими специалистами в псевдодифференциальном исчислении, как Хельфер и Шёстранд.

Предыдущие представления в композиции с гомоморфизмом η_α дают также представления алгебры $\mathcal{A}(I)$. Получаемые путем варьирования α в I семейства во всех случаях оказываются точными семействами представлений алгебры $\mathcal{A}(I)$.

3.3. Алгебра $\mathcal{A}_\alpha$ обладает достаточно простой структурой, что делает ее особенно интересной в приложениях. В случае иррационального α мы имеем

Предложение 3.2 (см. [BE3]). *Если число α иррационально, то алгебра $\mathcal{A}_\alpha$ проста. В частности, все ее представления точны.*

В случае когда $\alpha = p/q \in \mathbf{Q}$, где p и q — взаимно простые целые числа, получается следующее представление:

Предложение 3.3 (см. [BE3]). *Если $\alpha = p/q \in \mathbf{Q}$, где p и q — взаимно простые целые числа, то алгебра $\mathcal{A}_\alpha$ изоморфна подалгебре непрерывных функций, определенных на двумерном торе $\mathbf{T}^2 = (\mathbf{R}/\mathbf{Z})^2$ и принимающих значения в пространстве комплексных $q \times q$ -матриц, порожденной следующими двумя функциями:*

$$U_i(k) = e^{ik_i} u_i, \quad k = (k_1, k_2) \in \mathbf{T}^2, \quad i = 1, 2, \quad (18)$$

где u_i — унитарные матрицы размера $q \times q$, определенные соотношениями

$$u_i^q = 1, \quad u_1 u_2 = e^{2i\pi p/q} u_2 u_1. \quad (19)$$

Например, мы можем выбрать в качестве u_i следующие матрицы:

$$u_1 = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \quad (20a)$$

$$u_2 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & e^{2i\pi p/q} & 0 & & 0 & 0 \\ 0 & 0 & e^{2i\pi 2p/q} & & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & e^{2i\pi (q-2)p/q} & 0 \\ 0 & 0 & 0 & \dots & 0 & e^{2i\pi (q-1)p/q} \end{pmatrix}. \quad (20b)$$

Это представление, в частности, удобно для численного определения спектра элемента H алгебры $\mathcal{A}_{p/q}$. Действительно, H становится функцией от переменной k с матричными значениями. Таким образом, для каждого значения k достаточно диагонализировать конечную матрицу, что сравнительно просто сделать с помощью компьютера. Каждое собственное значение также определяет функцию $E_r(k)$ от k ($r \in [0, q-1]$), и спектр получается объединением всех возможных ее значений, когда k пробегает 2-тор. В действительности образ этого тора при E_r определяет то, что называют «зоной», для которой достаточно знать границы. В большинстве случаев, известных из литературы, это замечание позволяет свести вычисления к диагонализации относительно небольшого числа матриц. Так, например, обстоит дело в случае модели Харпера на треугольнике (см. рис. 2).

3.4. Примечательно, что рис. 1 и 2, демонстрирующие спектры, полученные предыдущим методом для рациональных значений α , при изменении α принимают фрактальную форму, и это позволяет думать, что при иррациональном α должно получаться канторово множество. Однако для подтверждения этого нужно убедиться, что иррациональный случай хорошо аппроксимируется рациональным. Алгебраический подход действительно позволяет доказать, что спектр элемента алгебры $\mathcal{A}(I)$ непрерывен по α :

Предложение 3.4 (см. [BE3]). *Если H — нормальный элемент алгебры $\mathcal{A}(I)$ (т. е. $HH^* = H^*H$), то границы лакун спектра оператора $\eta_\alpha(H)$ являются непрерывными функциями от α .*

В следующем разделе мы увидим, что при значениях α , близких к нулю, спектр можно изучать с помощью квазиклассического анализа. Как заметил Соколов [SO] в 1981 г., отсюда можно получить форму спектра в окрестности произвольного рационального числа. Действительно, использование подстановки Пайерлса и представления алгебры в рациональном случае матрицами, зависящими от $k \in T^2$, приводит к следующему представлению алгебры $\mathcal{A}_{p/q+\alpha}$:

Спектр треугольной решетки с $fr = 0,500$

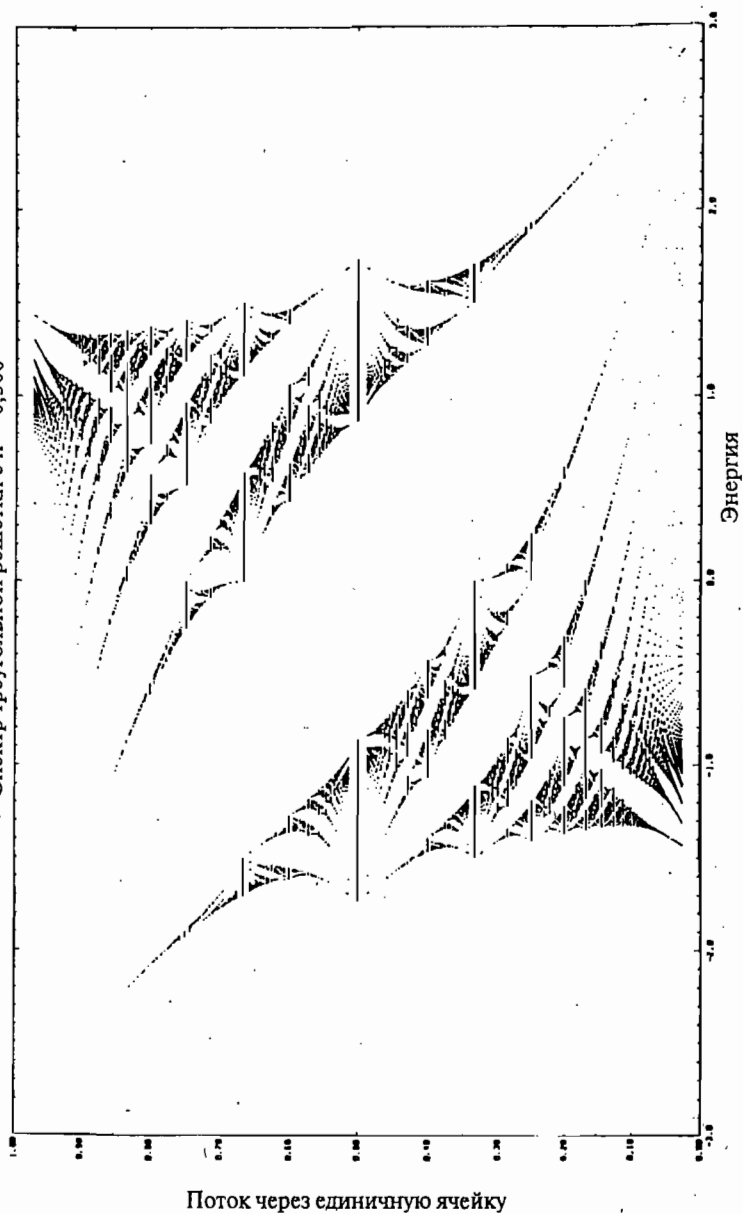


Рис. 2.

Предложение 3.5. Алгебра $A_{p/q+\alpha}$ изоморфна подалгебре алгебры $A_\alpha \otimes M_q$, порожденной элементами

$$\hat{U}_i = U_i \otimes u_i, \quad i = 1, 2. \quad (21)$$

Этот последний результат позволяет использовать квазиклассический анализ в окрестности рационального числа, но за это придется расплачиваться работой с матричными операторами.

4. КВАЗИКЛАССИЧЕСКИЙ АНАЛИЗ

4.1. Ренорм-группа по Вилкинсону. В своей диссертации [W1] Вилкинсон систематически развил метод, основанный на квазиклассическом анализе, который позволил понять, почему спектр Хофstadтера распространяется на всю шкалу. В действительности соотношение коммутирования (2) есть не что иное, как частный случай соотношений Вейля, при условии замены постоянной Планка $\hbar$ на $2\pi\alpha$. Аналогично, предел при $\alpha \rightarrow 0$ есть не что иное, как квазиклассический предел. Рассмотрим сначала ситуацию, когда α мало.

Итак, следуя методу Бора и Зоммерфельда, мы сначала рассмотрим соответствующую классическую проблему. Пусть I — интервал вида $[0, \epsilon]$, и пусть $H \in A(I)$ — самосопряженный элемент. Для простоты будем предполагать, что ряд Фурье для H убывает так быстро, как это нам нужно, и что коэффициенты Фурье — настолько регулярные функции от α , насколько нужно. Поставим в соответствие элементу H классическую функцию Гамильтона $\mathcal{H}$, определенную на двумерном торе $\mathbb{T}$:

$$\mathcal{H}(k) = \sum_{m \in \mathbb{Z}^2} h_m e^{ik \cdot m}, \quad (22)$$

где компоненты вектора k будут записываться в круглых скобках, $k = (x, \xi)$, чтобы согласовать обозначения с обозначениями Хёрмандера, используемыми в микролокальном анализе, тогда как $k \cdot m = x m_1 + \xi m_2$ будет обозначать скалярное произведение.

Получается задача классической механики с одной степенью свободы. Фазовое пространство здесь $\mathbb{R}^2$, но периодичность функции $\mathcal{H}$ по обоим переменным (x, ξ) позволяет свести изучение к одной элементарной ячейке. Следовательно, классические орбиты в фазовом пространстве являются связными компонентами линий уровня для $\mathcal{H}$, которые образуют подмножества, инвариантные от-

носителем трансляций, кратных 2π . В случае модели Харпера (следовательно, при $\mu = 1$) $\mathcal{H}(x, \xi) = \cos x + \cos \xi$ и линии уровня могут быть двух видов (см. рис. 3):

(i) Орбиты, гомотопные окружности с центром в точке, получаемой трансляцией из точки $(0, 0)$ или (π, π) . Эти орбиты окружают максимум или минимум соответственно.

(ii) Сепаратрисы, т. е. прямые вида $x \pm \xi = \pi \pmod{2\pi}$. Эти прямые пересекаются в седловых точках функции $\mathcal{H}$, находящихся в точках $(0, \pi)$ и $(\pi, 0) \pmod{2\pi}$. Тогда орбиты являются различными сегментами, связывающими две соседние седловые точки на одной из этих прямых.

Заметим, что эта структура не изменится с точностью до гомотопии, если рассматривать элемент H_1 , близкий к H , при условии, что симметрии элемента H сохраняются, например, симметрия относительно поворота на $\pi/2$ в фазовом пространстве. Для другой модели структура орбит может оказаться более сложной. Мы не будем здесь этим заниматься.

Квантование по Бору—Зоммерфельду ограничивает рассмотрение только орбитами вида Γ , удовлетворяющими условию

$$\oint_{\Gamma} \xi dx = \left(n + \frac{1}{2}\right) 2\pi\alpha. \quad (23)$$

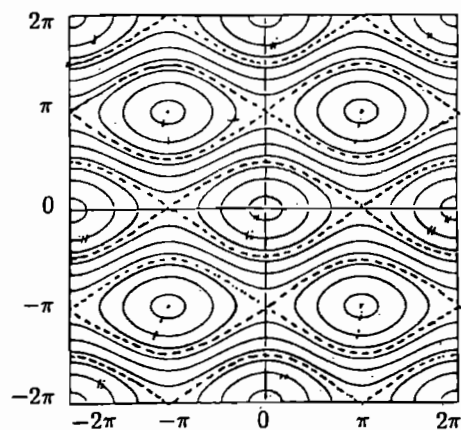
Значение функции H на такой орбите зависит от n и обозначается E_n . Мы назовем E_n собственной энергией. Она не зависит от компонент связности ввиду инвариантности относительно трансляций. Таким образом, эти значения являются собственными значениями оператора H наиболее низкого порядка по степеням α . Обычный метод ВКБ¹⁾ дает возможность вычислить асимптотическое поведение этих собственных значений по степеням α , что позволяет их определить по модулю $O(\alpha^\infty)$.

Интересен частный случай орбит, близких к максимумам и минимумам (донья колодцев). В действительности они близки к окружностям, что позволяет явно вычислить собственную энергию в виде

$$E_n = \mathcal{H}(k_0) \mp 2\pi\alpha \left(n + \frac{1}{2}\right) \sqrt{\det \partial_i \partial_j \mathcal{H}(k_0)} + O(\alpha^2), \quad (24)$$

где знак $+$ выбирается для минимума, а знак $-$ для максимума, k_0 соответствует положению выбранного экстремума и n — квантовый уровень.

¹⁾ Вентцель—Крамер—Бриллюэн.

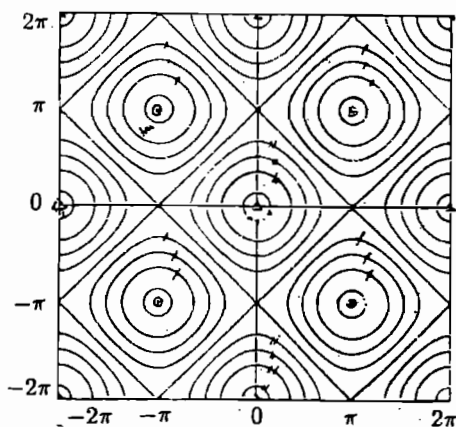


$$- : -(1 - \mu) < E < (1 - \mu)$$

$$+ : -(1 + \mu) < E < -(1 - \mu)$$

$$\# : 1 - \mu \leq E \leq 1 + \mu$$

Для $\mu = 1$



$$- : E = 0 \quad \Theta : E = -2 \quad + : -2 < E < 0 \quad \Delta : E = 2 \quad \# : 0 < E < 2$$

Рис. 3.

В представлении Вейля каждая из этих орбит определяет собственную функцию $\phi_{m,n}$ в пространстве $L^2(\mathbf{R})$, где $m \in \mathbf{Z}^2$ нумеруют центры элементарных ячеек, в которых находятся рассматриваемые орбиты, так, что собственные значения E_n бесконечно убывают. Эти собственные функции тоже могут быть вычислены методом ВКБ. Вследствие туннельного эффекта интеграл перекрытия $(\phi_{n,m}|\phi_{n,m'})$ убывает как $e^{-\text{const}|m-n|/\alpha}$, поскольку семейство $\{\phi_{n,m}; m \in \mathbf{Z}^2\}$ ортонормировано с точностью до экспоненциально малой относительно α ошибки.

Вследствие того же самого туннельного эффекта между орбитами одинаковой энергии, находящимися в различных элементарных ячейках, вырождение собственных значений E_n усиливается, что порождает зону $\mathcal{B}_n$, ширина которой имеет порядок возмущений, создаваемых туннельным эффектом. Каждая из этих зон позволяет с помощью формулы Коши определить спектральный проектор для H

$$P_n = \frac{1}{2i\pi} \oint_C \frac{dz}{z - H}, \quad (25)$$

где C — окружность на комплексной плоскости с центром в E_n и с радиусом, достаточно большим, чтобы охватывать зону $\mathcal{B}_n$, и достаточно малым, чтобы не включать другие зоны. Этот проектор можно вычислить с точностью до экспоненциально малой относительно α ошибки как проектор на пространство, порожденное семейством $\{\phi_{n,m}; m \in \mathbf{Z}^2\}$. Отсюда также следует, что можно вычислить ограничение оператора H на эту зону, рассматривая его матрицу в базисе $\{\phi_{n,m}; m \in \mathbf{Z}^2\}$. Предположим, помимо этого, что мы можем контролировать ошибки, возникающие на каждом этапе этих рассуждений. Тогда мы получим описание каждой зоны, определенной таким же образом эффективным оператором H_n , задаваемым бесконечной матрицей, которая индексирована точками решетки $\mathbf{Z}^2$, соответствующими различным элементарным ячейкам.

Одно из основных замечаний Вилкинсона состоит в доказательстве того, что эта матрица имеет вид

$$H_{n;m,m'} = \tilde{E}_n(\alpha) \delta_{m,m'} + C(\alpha) e^{i\pi\alpha' m \wedge m'} \tilde{H}_{n;m'-m}, \quad (26)$$

где $\tilde{E}_n(\alpha) = E_n + O(\alpha^2)$, α' — дробная часть от $1/\alpha$, $C(\alpha)$ — константа, экспоненциально малая относительно α , и $\tilde{H}_m$ — последовательность, экспоненциально убывающая относительно m/α . Мы узнаём здесь выражение для матрицы элемента $\tilde{H}$ алгебры $\mathcal{A}'_\alpha$ в представлении ГНС. Кроме того, гамильтониан подзоны наследует

симметрию оператора H , и если α мало, то $\tilde{H}$ можно аппроксимировать по модулю экспоненциально малой относительно α ошибки элементами матриц ближайших соседей, так что и после аппроксимации $\tilde{H}$ остается моделью Харпера. Таким образом, гамильтониан каждой такой подзоны задается формулой

$$\tilde{H}_n = H(\alpha', \mu = 1) + O(e^{-\text{const}/\alpha}), \quad (27)$$

показывающей, что спектр оператора H в такой подзоне вычисляется как спектр оператора H при условии «перенормировки» параметров трансферта (здесь — константы $C(\alpha)$) и нормированного потока α и модификации этого оператора посредством экспоненциально малого относительно α возмущения.

Как мы отмечали в предыдущих разделах, эту стратегию можно обобщить на анализ спектра в окрестности числа p/q . В этом случае эффективная постоянная Планка равна разности $\delta = \alpha - p/q$, и ассоциированный классический гамильтониан становится функцией от $\mathcal{H}(x, \xi)$ со значениями в пространстве матриц размера $q \times q$. В случае модели Харпера эта матрица трехдиагональна, и ее диагонализация знакома любому физики. Собственные значения $e_i(x, \xi)$ этой матрицы, т. е. функции зон, играют роль, аналогичную роли собственных значений классической функции Гамильтона, рассматривавшейся выше. В частности, «донья колодцев» для этих функций соответствуют границам рассматриваемой зоны. Однако нужна корректировка, называемая «искривлением», поскольку собственные векторы этой матрицы определяют линейное периодическое расслоение над фазовым пространством или, если угодно, расслоение над двумерным тором. Это расслоение в общем случае нетривиально, и его классы Черна могут быть связаны с квантованием проводимости Холла [TKN2]. Эти вычисления ведут к явному представлению для собственных энергий вблизи границ зон, аналогичному формуле (24) [BE1, BE2, BE3, HS3, RB2]:

$$E_{i,n} = e_i(k_0) \mp 2\pi \left| \alpha - \frac{p}{q} \right| \left(n + \frac{1}{2} \right) \sqrt{\det \partial_i \partial_j e_i(k_0)} + \sigma 2\pi \left(\alpha - \frac{p}{q} \right) + O \left(\left(\alpha - \frac{p}{q} \right)^2 \right),$$

где знак $+$ выбирается для минимума, а знак $-$ для максимума, k_0 — положение рассматриваемого экстремума, n — квантовый уровень, а σ — следующее выражение:

$$\sigma = \text{tr} (\mathcal{P}_i(k_0) \{ \partial_z \mathcal{P}_i(k_0) \partial_\xi \mathcal{H}(k_0) - \partial_\xi \mathcal{P}_i(k_0) \partial_x \mathcal{H}(k_0) \}), \quad (29)$$

где $P_i(x, \xi)$ — спектральный проектор, ассоциированный с собственным значением $e_i(x, \xi)$ матрицы $\mathcal{H}(x, \xi)$. Эта формула, называемая формулой Вилкинсона—Раммаля была доказана и обобщена в [BE1, BE2, HS2, RB2] для произвольного гамильтониана в алгебре $\mathcal{A}(I)$.

4.2. Вклад Хельфера и Шёстранда. Приведенные выше соображения Вилкинсона образуют каркас стратегии Хельфера и Шёстранда, позволяющий дать строгое математическое доказательство. Однако в этих соображениях негласно присутствуют две трудности: первая связана с контролем за ошибками в терминах эффективной постоянной Планка, вторая — с вычислением спектра вблизи сепаратрис. Вклад Хельфера и Шёстранда состоит в заполнении этих лакун.

Микролокальный анализ является математическим инструментом, позволяющим решить первую проблему. В этом состоит главный методологический вклад Хельфера и Шёстранда. В их первой работе [HS1] они исследуют случай малых значений α . Их вторая статья [HS2] содержит анализ в окрестности рационального числа.

Вторая трудность на самом деле непреодолима. Действительно, различные вычислительные работы позволяют убедиться, что главная часть спектра сосредоточена вблизи нулевой энергии, т. е. энергии, которая соответствует в точности классическим орбитам, близким к сепаратрисам. В первых двух статьях о бабочке Хофстадтера [HS1, HS2] Хельфер и Шёстранд уклонились от обсуждения этого деликатного вопроса. Только в третьей работе они дают решение этой проблемы, используя для этого старую работу Азбеля [AZ]. Этот результат является, без сомнения, наиболее оригинальным как по сравнению с численными, так и теоретическими работами физиков. Действительно, ВКБ-анализ вблизи сепаратрис отсутствовал почти во всех предыдущих работах. Даже результаты Азбеля не свободны от неточностей в этом вопросе.

Резюмируя, сформулируем наиболее важный результат, полученный с помощью этого метода:

Теорема 4.1 (см. [HS3]). *Существует целое число $C \geq 1$, такое, что если α — иррациональное число, допускающее разложение в цепную дробь $[a_1, a_2, \dots, a_n, \dots]$, в которой $a_n \geq C$ для любых $n \geq 1$, то спектр оператора $H(\alpha, \mu = 1)$ — канторово множество меры нуль.*

Этой теоремой не исчерпываются все результаты. В действительности из этой статьи можно извлечь много другой информации

о структуре спектра. Разложение в цепную дробь числа α позволяет определить следующим образом иерархию зон: пусть дана зона в окрестности собственного значения, близкого к типу Бора—Зомерфельда; тогда с ней можно связать перенормированные операторы двух типов: либо «возмущенный оператор Харпера вблизи $\alpha = 0$ », либо «возмущенный оператор Харпера вблизи $\alpha = 1/2$ », в котором нужно заменить α на $\alpha' = \{1/\alpha\}$ (здесь $\{z\}$ — дробная часть числа z). Этот оператор обладает тем же спектром, что и предыдущий гамильтониан в этой зоне. Итерируя это правило, можно определить иерархию. Кроме того, каждая подзона Бора—Зомерфельда отделена от соседней расстоянием порядка $O(\alpha)$.

Заметим, однако, что проблема показательна в том отношении, что позволяет четко очертить пределы применимости метода. Действительно, ошибки контролируются только неявными выражениями, в которых известен лишь асимптотический вклад в малую постоянную Планка. Для того чтобы сделать вывод, необходимо предполагать, что эта константа «достаточно мала». Действие ренормгруппы рекурсивно, в частности, при переходе от одного этапа рекурсии к другому поток перенормируется, меняя α на $\alpha' = \{1/\alpha\}$, где $\{x\}$ — дробная часть x . Итерируя это отношение, мы приходим к разложению α в цепную дробь,

$$\alpha = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_n + \dots}}}$$

Зная, что на каждом этапе необходимо предполагать, что α «достаточно мало», чтобы перейти к следующему этапу, можно получить желаемый результат для тех α , цепные дроби которых удовлетворяют условиям

$$\exists C \geq 0, \quad a_n \geq C, \quad \forall n \geq 1. \quad (31)$$

Итак, если $C > 1$, то это условие ограничивает множество значений α такими, при которых результат применяется к множеству меры Лебега нуль! Кроме того, исключаются многие интересные для физиков числа, например золотое сечение. Другими словами, микролокальный анализ, безусловно, обладает качественно корректным и мощным подходом, но не позволяет получить качественно оптимальный результат.

Среди следствий этого анализа имеются два показательных результата, присутствующих как в оригинальных статьях [HS3], так

и в более поздних статьях [GH], которые позволяют проиллюстрировать мощь метода. Первое — это разрешение парадокса, вытекающего, с одной стороны, из теоремы 2.2, а с другой — из предсказания Вилкинсона, согласно которому эффективный гамильтониан, описывающий каждую подзону, является гамильтонианом Харпера. Действительно, теорема 2.2 предсказывает, что если число α рационально, то все лакуны являются открытыми множествами, кроме центральной с нулевой энергией. Пусть тогда p/q — рациональное число с цепной дробью $[a_1, a_2, \dots, a_n]$. Обозначим через p'/q' близкое рациональное число с цепной дробью $[a_1, a_2, \dots, a_n, 2b]$, где $b \geq 1$. Рассмотрим зону B , не содержащую нулевой энергии, в спектре модели Харпера, соответствующей $\alpha = p/q$, и пусть $B(\alpha)$ — зона, полученная деформацией зоны B при изменении α в малой окрестности числа p/q . Соответствующий нормированный поток, появляющийся в эффективном гамильтониане, равен положительному действительному числу α' , такому, что $\alpha = [a_1, a_2, \dots, a_n + \alpha']$. Таким образом, для числа p'/q' мы получаем $\alpha' = 1/2b$. Зона $B' = B(p'/q')$, согласно теореме 2.2, распадается на $2b$ непересекающихся подзон. Поскольку это рациональное число с четным знаменателем, то центральная лакуна существует и является открытым множеством. С другой стороны, если эффективный гамильтониан в точности равен гамильтониану Харпера, то в силу той же теоремы центральная лакуна должна быть замкнутой. Разрешение этого парадокса заключено в том факте, что эффективный гамильтониан на B' отличается от модели Харпера на возмущение, экспоненциально малое относительно α' . Таким образом, несмотря на то, что центральная лакуна — открытое множество, ширина ее имеет порядок $O(e^{-\text{const} \cdot 2b})$! То же самое показывают вычисления, иллюстрируемые рис. 4 (см. [GH]). Заметим, однако, что Эллиот и др. [CEY] доказали алгебраическими методами, что ширина лакуны в случае $\alpha = p/q$ минорируется числом $C \cdot 8^{-q}$, где C — явная числовая константа.

Другой результат относится к центральной лакуне. Согласно теореме 2.2, центральная лакуна для $\alpha = p/q$ при четном q замкнута. Это означает, что две зоны касаются друг друга, и более детальный анализ показывает, что этот контакт в модели Харпера происходит в единственной точке элементарной ячейки и что эта точка коническая. Квазиклассический анализ обязывает учитывать это случайное вырождение, формулируя проблему в терминах 2×2 -матриц. Оказывается, что нормальная форма этих матриц является опера-

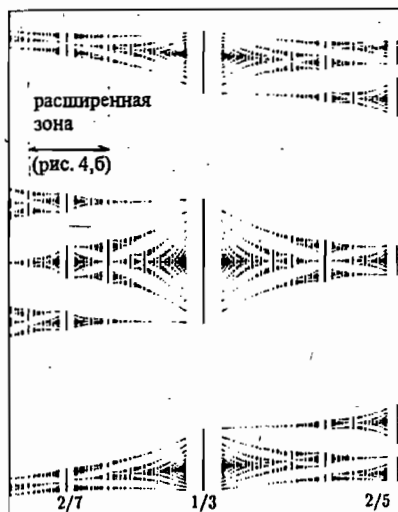


Рис. 4,а.

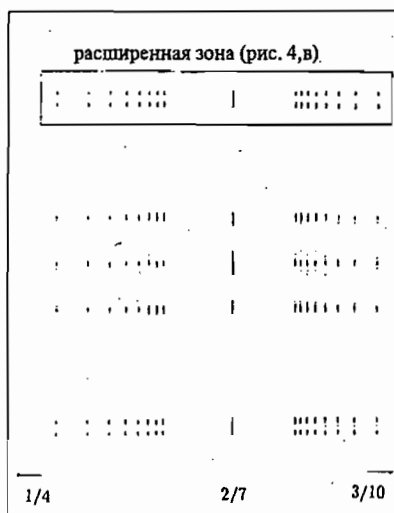


Рис. 4,б.

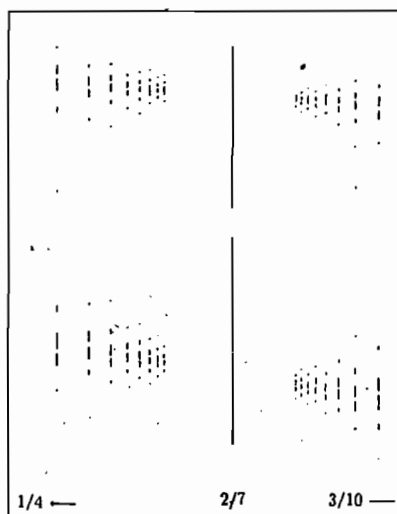


Рис. 4,с.

Спектр для α около $1/(3+1/2)$, большой масштаб,
частичная последовательность $\alpha = 1/(3+1/(2+1/q))$.

тором Дирака, собственные значения которого можно явно вычислить (см. [HS2], а также [RB2]). Соответствующий уровень энергии имеет порядок $\sqrt{|\alpha - p/q|}$.

5. КОНСТРУКЦИЯ ПЕРЕНОРМИРОВАННОГО ГАМИЛЬТониАНА

5.1. Проблема колодца. Первая проблема — изолировать каждый из колодцев. Мы рассмотрим, например, случай максимума (в случае минимума задача решается аналогично).

Напомним, что если $f(x, \xi)$ — регулярная ограниченная функция на $\mathbf{R}^2$, то на $L^2(\mathbf{R})$ с помощью квантования по Вейлю можно следующим образом определить оператор $\hat{f}$:

$$\hat{f}\psi(x) = \int_{\mathbf{R}^2} \frac{dy d\xi}{4\pi^2\alpha} f\left(\frac{x+y}{2}, \xi\right) e^{\frac{i\xi(x-y)}{2\pi\alpha}} \psi(y), \quad \psi \in L^2(\mathbf{R}). \quad (32)$$

Кроме того, определим гауссовское квантование формулой

$$(\tilde{f}\psi)(x) = \int_{\mathbf{R}^2} \tilde{f}(x, y) \psi(y) dy, \quad \psi \in L^2(\mathbf{R}), \quad (33a)$$

где

$$\tilde{f}(x, y) = \frac{1}{\sqrt{\pi}} \int_{\mathbf{R}^2} \frac{dx' d\xi}{4\pi^2\alpha} f(x', \xi) e^{-\frac{(x-x')^2 + (y-y')^2}{2}} e^{\frac{i\xi(x-y)}{2\pi\alpha}}. \quad (33b)$$

При гауссовском квантовании положительной функции f соответствует положительный оператор, локализованный на носителе функции f .

Итак, пусть $E > 0$ — энергия, в окрестности которой мы собираемся вычислить спектр, и пусть число $\epsilon > 0$ таково, что $E - 2\epsilon > 0$. Пусть $m \in \mathbf{Z}^2$ — индекс рассматриваемого колодца. Обозначим через χ регулярную функцию от (x, ξ) , обращающуюся в нуль, если $(x, \xi) \notin [-\pi, \pi]^2$ или если $(x, \xi) \in [-\pi, \pi]^2$ и $\mathcal{H}(x, \xi) \leq E - 2\epsilon$, равную 1 на множестве точек $(x, \xi) \in [-\pi, \pi]^2$, для которых $\mathcal{H}(x, \xi) \geq E - \epsilon$, и такую, что $0 \leq \chi(x, \xi) \leq 1$. Положим $\chi_m(x, \xi) = \chi(x - 2\pi m_1, \xi - 2\pi m_2)$. Оператор $\tilde{\chi}_m$ обозначает тогда гауссовское квантование χ_m . Заметим, что унитарные операторы $T(m)$ ($m \in \mathbf{Z}^2$); задаваемые формулами

$$T(m)\psi(x) = e^{im_2(x - 2\pi \frac{m_1}{2})/\alpha} \psi(x - 2\pi m_1), \quad (34)$$

удовлетворяют соотношениям $T(M)\tilde{\chi}_0 T(m)^{-1} = \tilde{\chi}_m$ и коммутируют с операторами U_i , $i = 1, 2$, в представлении Вейля. Следовательно, они квантуются трансляциями, соответствующими целым числам, умноженным на 2π .

Эта конструкция приводит нас к определению гамильтониана колодца H_m формулой

$$H_m = H - (E - \epsilon) \sum_{m'; m' \neq m} \tilde{\chi}_{m'}. \quad (35)$$

Этот оператор обладает символом, не превосходящим $E - \epsilon$ везде, кроме колодца m . Более того, используя гауссовское квантование функции χ , получаем, что $H_m \leq (E - \epsilon) 1$ по модулю компактного оператора. Кроме того, часть спектра оператора H_m , содержащаяся в интервале $[E - \epsilon/2, E + \epsilon/2]$, состоит из конечного числа собственных значений конечной кратности. Способ вычисления этих собственных значений состоит в аппроксимации ВКБ, но, чтобы доказать, что получается весь спектр в рассматриваемой области, Хельфер и Роберт [HR] вынуждены были сконструировать, используя символическое исчисление, функцию $f(E, \alpha)$, такую, что в интервале $[E - \epsilon/2, E + \epsilon/2]$ и для достаточно малого α оператор $f(H_m, \alpha)$ имеет такой же спектр, как оператор гармонического осциллятора. Заметим, что аналогичный результат был найден в [BV] в рамках алгебраического формализма, где был сконструирован путем возмущений, причём равномерно по α , унитарный оператор, сопряженный к этим двум операторам. Заметим также, что это неверно, если число степеней свободы не равно единице. Преимущества этого метода заключаются в том, что он дает явную конструкцию соответствующей квазимоды с помощью собственных векторов гармонического осциллятора.

5.2. Изучение туннельного эффекта между колодцами.

Приведенная выше техника «заполнения колодца» была развита ранее для изучения туннельного эффекта в уравнении Шрёдингера $H = P^2/2 + V$, где V — потенциал двух вырожденных колодцев, которым мы приписываем индексы $\mp$. В этом случае туннельный эффект задает уровни энергии в виде $E_{n,\pm} = E_n \pm \Delta E$, где E_n — энергия, получающаяся в том случае, когда колодцы отделены друг от друга, а $\Delta E \approx e^{-S/\alpha}$, где S — туннельное взаимодействие между двумя классическими орбитами с энергией E_n . Собственные функции двух гамильтонианов $H_{\pm}$, которые получаются заполнением колодцев, соответствующих собственным значениям, находящимся

в рассматриваемом интервале энергии, равны для каждого колодца собственным функциям гамильтониана H с точностью до ошибки порядка $O(e^{-d/\alpha})$, где $d \approx S$ — «расстояние Агмона» между орбитами с энергией $E + O(\epsilon)$ двух соседних колодцев (расстояние Агмона определяется метрикой $ds^2 = \sqrt{(V(x) - E)_+} dx^2$). Вычисление скачка между двумя подуровнями, создаваемого туннельным эффектом, требует точного знания собственных функций только на середине пути между колодцами, т.е. когда расстояние Агмона равно $d/2$. Следовательно, вычисление этого скачка с помощью собственных функций гамильтонианов H_+ , H_- дает точный результат по модулю ошибки порядка $O(e^{-2d/\alpha})$. Таким образом, используя этот метод, мы можем быть уверены, что находим правильный главный член скачка энергии, возникающего в силу туннельного эффекта.

Оказывается, что в случае гамильтониана, близкого к гамильтониалу Харпера, можно перенести понятие расстояния Агмона в фазовое пространство. Это было технически осуществлено в работе Хельфера и Шёстранда [HS1]. Эта методика показывает, что каждый m -й колодец обладает приближенными собственными функциями $\varphi_{n,m}$ с соответствующими собственными значениями $E_n \in [E - \epsilon/2, E + \epsilon/2]$, не зависящими от m в силу инвариантности H относительно трансляций, кратных 2π . Кроме того, мы знаем, что расстояние между этими собственными значениями равно $O(\alpha)$ и что спектр оператора H отличается от этого множества собственных значений только на величину порядка $O(e^{-d/\alpha})$, вызванную туннельным эффектом между колодцами, где d — расстояние Агмона между двумя соседними колодцами. Более того, скалярные произведения $(\varphi_{n,m} | \varphi_{n,m'})$ не превосходят $O(e^{-d(m,m')/\alpha})$, где $d(m, m')$ — расстояние Агмона между m -м и m' -м колодцами. Для каждого n семейство $\{\varphi_{n,m}; m \in \mathbb{Z}^2\}$, таким образом, почти ортонормированно. Метод Шмидта позволяет построить, исходя из этого семейства, ортонормированный базис $\{\check{\varphi}_{n,m}; m \in \mathbb{Z}^2\}$, который отличается от предыдущего на величину, экспоненциально малую относительно α .

Следовательно, можно построить новый оператор $H_n(E)$, аналитически зависящий от E , в окрестности $O(\alpha)$ точки E_n , действующий на подпространстве, порожденном этим семейством, и такой, что спектр оператора H в этой окрестности будет задаваться множеством E' , таким, что $E' \in Sp(H_n(E'))$. Для этого Хельфер и Шёстранд использовали «метод Грюэна», который состо-

ит в рассмотрении следующей новой проблемы: на пространстве $L^2(\mathbf{R}) \oplus l^2(\mathbf{Z}^2)$ рассматривается оператор $\tilde{H}$, задаваемый выражением

$$\tilde{H} = \begin{pmatrix} H - E & R_+ \\ R_- & 0 \end{pmatrix}, \quad (36a)$$

где

$$R_- \psi_m = (\hat{\varphi}_{n,m} | \psi), \quad \psi \in L^2(R), \quad (36b)$$

и

$$R_+ u = \sum_{m \in \mathbf{Z}^2} u_m \hat{\varphi}_{n,m}. \quad (36c)$$

Если обратную матрицу к $\tilde{H}$ обозначить через

$$\tilde{H}^{-1} = \begin{pmatrix} Q & Q_+ \\ Q_- & Q_{-+} \end{pmatrix}, \quad (37)$$

то $H - E$ имеет ограниченную обратную тогда и только тогда, когда этим свойством обладает Q_{-+} , и в этом случае мы получаем

$$(H - E)^{-1} = Q - Q_+(Q_{-+})^{-1}Q_-, \quad (Q_{-+})^{-1} = -R_+(H - E)^{-1}R_-. \quad (38)$$

Таким образом, эффективный гамильтониан задан бесконечной матрицей Q_{-+} . Так же, как в рассуждениях Вилкинсона, анализ симметрий показывает, что эта матрица может рассматриваться как матрица элемента $H_{eff,n}(E)$ алгебры $\mathcal{A}_\alpha$ в представлении ГНС, которому мы сопоставляем его представитель $H_n(E)$ в представлении Вейля таким образом, чтобы иметь возможность снова использовать символическое исчисление на следующем этапе.

Как мы уже отметили, техника «заполнения колодцев» позволяет точно вычислить главный член туннельного эффекта. Технически это происходит так: учитываются только ближайшие соседние члены в матрице Q_{-+} , которые записываются так же, как в формуле (26), и приводят к новому гамильтониану Харпера, в котором просто перенормируется параметр трансферта t , определяемый главным членом туннельного эффекта. Таким образом, $t = O(e^{-S/\alpha})$, где S — туннельный эффект. Остальные члены получают поправку этого главного члена на величину, экспоненциально малую относительно α .

5.3. Исследование центральной зоны. Предыдущая конструкция приемлема, только если классические орбиты не слишком близки к сепаратрисе. Если мы удовлетворимся рассмотрением спектра

только в областях энергии, удаленных от центральной полосы, то схема перенормировки обязывает нас исключать на каждом этапе малый интервал в центре каждой подзоны. Подзоны в этой части спектра имеют ширину, экспоненциально малую относительно α , из чего по индукции заключаем, что это спектральное множество имеет фрактальную размерность нуль. Однако численные расчеты показывают, что фрактальная размерность спектра должна быть равна $1/2$. Отсюда следует, что спектральная плотность сконцентрирована в основном в исключаемой области. Разумеется, необходимо понять, что здесь происходит.

Чтобы достичь этого, Хельфер и Шёстранд сконструировали, как и в предыдущем случае, квазимоды, микролокализованные вблизи сепаратрис. Для этого нужно взять оператор, сопряженный к гамильтониану (в смысле символического исчисления) в нормальной форме в окрестности седловой точки. Эта форма может быть выбрана в виде $\xi^2 - x^2$ или в виде $x\xi$ на уровне символов. Это отражено в приводимой ниже лемме, в которой $\hbar$ — малый параметр, играющий роль постоянной Планка, и $D = -i\frac{\partial}{\partial x}$:

Лемма 5.1 (см. [HS3]). Пусть $H(x, \hbar D, \hbar)$ — формально аналитический формально самосопряженный псевдодифференциальный оператор порядка 0. Предположим, что его главный символ $\mathcal{H}$ имеет невырожденную седловую точку с координатами $(0, 0)$, такую, что $\mathcal{H}(0, 0) = 0$. Тогда существует действительный аналитический символ $F(t, \hbar) \sim \sum_0^\infty f_j(t)\hbar^j$, определенный для малых $|t|$, а также интегральный аналитический формально унитарный оператор Фурье U , происходящий из классического канонического преобразования, определенного в окрестности точки $(0, 0)$, преобразующий эту точку в себя, такой, что

$$U^*F(H, \hbar)U = H_0 = \frac{1}{2}(x\hbar D + \hbar D x). \quad (39)$$

Уравнение $(H_0 - \alpha)u = 0$ имеет четыре частных решения, а именно функции $u_\pm = \Theta(\pm x)|x|^{-1/2+i\alpha}$, где Θ — функция, равная 0 на отрицательных и 1 на положительных действительных числах, а также преобразования Фурье $v_\pm$ функций $u_\pm$. Таким образом, всякое решение является линейной комбинацией либо решений $u_\pm$, либо решений $v_\pm$, и, следовательно, существует матрица «монодромии», позволяющая переходить от одного представления к другому.

Возвращаясь к изучаемому оператору H , равному модели Харпера по модулю экспоненциально малой относительно α ошибки,

мы видим, что из инвариантности относительно вращения на угол $\pi/2$ его символа автоматически вытекает, что седловые точки расположены в $(0, \pi)$ и $(\pi, 0)$ (по модулю 2π). К каждой из этих точек подходят два участка сепаратрис, соответствующие устойчивому многообразию гамильтонова потока, порожденного символом $\mathcal{H}$ гамильтониана H , а два других участка соответствуют неустойчивому многообразию. Первые два соответствуют решениям $u_{\pm}$, а вторые — решениям $v_{\pm}$ уравнения $(H - E)\psi = 0$, получаемым с помощью леммы 5.1 и предыдущих результатов. Заметим, что эти четыре решения получаются друг из друга поворотом на угол $\frac{\pi}{2}$ в фазовом пространстве. Пусть $\chi \in C_0^{\infty}$ равна 1 в окрестности седловой точки $(0, \pi)$ и 0 вне диска радиуса $3\sqrt{2}\pi/4$ с центром в этой точке. Тогда можно так нормировать решения, что $(i[H, \chi]u_+ | u_+) = 1$, и это позволяет ввести функцию $f = i[H, \chi]u_+$, а также семейство $f_{m,j}$, поворачивая функцию f на углы $j\pi/2$ (что делается с помощью преобразования Фурье) и затем сдвигая ее на $2\pi m$ (с помощью оператора $T(m)$, определенного формулой (34)) в фазовом пространстве.

Рассмотрим проблему Грушина, ассоциированную с операторами $R_{\pm}$, где $R_+ : L^2(\mathbf{R}) \rightarrow l^2(\mathbf{Z}^2 \times \{1, 3\})$ определяется формулой

$$R_+ \psi_{m,j} = (f_{m,j} | \psi), \quad m \in \mathbf{Z}^2, j \in \{1, 3\}, \quad \psi \in L^2(\mathbf{R}), \quad (40a)$$

а $R_- : l^2(\mathbf{Z}^2 \times \{2, 4\}) \rightarrow L^2(\mathbf{R})$ определяется формулой

$$R_- u = \sum_{\mathbf{Z}^2 \times \{2, 4\}} u_{m,j} f_{m,j}, \quad u \in l^2(\mathbf{Z}^2 \times \{2, 4\}), \quad (40b)$$

которая дает 2×2 -матрицу эффективных операторов $H_{j,j'}$, $j \in \{1, 3\}$, $j' \in \{2, 4\}$. Как и в предыдущем случае, мы можем интерпретировать эту матрицу операторов как элемент алгебры $\mathcal{A}_{\alpha'} \otimes M_2(\mathcal{C})$ в представлении ГНС, позволяющем при переходе к представлению Вейля определить перенормированный оператор как 2×2 -матрицу, коэффициенты которой — псевдодифференциальные операторы, происходящие из нашей алгебры.

И так же, как в предыдущем случае, члены, которые не соответствуют ближайшим соседям, будут экспоненциально малы относительно остальных; поэтому мы получаем эффективный гамильтониан в виде малых возмущений ближайших соседних членов.

Нам остается только выяснить, какую каноническую форму принимает такой оператор. Заметим, что символ $\mathcal{H}_{eff}$ теперь явля-

ется функцией на фазовом пространстве со значениями в пространстве 2×2 -матриц. Перенесемся в аналогичную ситуацию квазиклассического анализа в окрестности рациональной точки. Здесь возникают два случая в зависимости от того, будут образы функций зоны этой матрицы разделены или нет. В первом случае получаются два гамильтониана, близкие к гамильтонианам Харпера в алгебре A_α . Во втором случае симметрия задачи и тот факт, что первоначальный гамильтониан близок модели Харпера, влекут за собой касание полос в конической точке. Следовательно, можно перейти к модели, близкой к модели Харпера для $\alpha = 1/2$.

Итак, мы видим, завершая рекурсию, что еще нужно изучить случай, когда исходный гамильтониан сам является 2×2 -матрицей и его символ обладает двумя функциями зон, пересекающимися в конической точке. Это то, что Хельфер и Шёстранд называют «вполне вырожденным случаем». Изучение этого случая требует несколько больше комбинаторики, но при этом не возникает новых трудностей, кроме технических.

Мы не будем останавливаться на деталях, но можно убедиться, что переход от этапа ренорм-группы к следующему требует только анализа небольшого числа типичных случаев.

ЛИТЕРАТУРА

- [AL] Alexander S. Superconductivity on networks. A percolation approach to the effect of disorder, *Phys. Rev.*, **B27** (1983), 1541–1557.
- [AA] Aubry S., André G. Analyticity breaking and the Anderson localization in incommensurate lattices, in: *Ann. Israeli Phys. Soc.*, vol. 3, Hilger, Bristol, 1980, 133–164.
- [AU] Aubry S. The new concept of transition by breaking of analyticity in crystallographic models, *Solid State Sci.*, **8** (1978), 264.
- [AMS] Avron J., van Mouche P. H. M., Simon B. On the measure of the spectrum for the almost Mathieu operator, *Preprint CALTECH*, 1990, а также: *Comm. Math. Phys.*, **132** (1990), No. 1, 103–118.
- [AZ] Азбель М. Я. Энергетический спектр электрона проводимости в магнитном поле. — *ЖЭТФ*, т. 46, 1964, № 3, 929–946.
- [BE1] Bellissard J. C^* -Algebras in solid state physics: 2D electrons in a uniform magnetic field, in: *Operator Algebras and Application*, Vol. 2, D. E. Evans, M. Takesaki (eds.), Cambridge University Press, 1988.
- [BE2] Bellissard J. Almost periodicity in solid state physics and C^* -algebras, in: *The Harald Bohr Centenary*, C. Berg, B. Fuglede (eds.), The Danish Royal Acad. Sci., **42** (1989), No. 3, 35–75.
- [BE3] Bellissard J. The Rotation Algebra, *Lecture donnée à l'école d'été de Nantes*, Juin 1991.

- [BKS] Bellissard J., Kreft C., Seiler R. Analysis of the spectrum of a particle on a triangular lattice with two magnetic fluxes by algebraic and numerical methods, *J. Phys.*, **A24** (1991), 2329-2353.
- [BES] Bellissard J., Simon B. Cantor Spectrum for the Almost Mathieu Equation, *J. Funct. Anal.*, **48** (1982), 408-419.
- [BV] Bellissard J., Vittot M. Heisenberg's picture and non commutative geometry in the semi-classical limit in quantum Mechanics, *Ann. Inst. Henri Poincaré*, **52** (1990), 175-235.
- [CH] Chambers W. G. *Phys. Rev.* **A140** (1965), 135-143.
- [CEY] Choi M. D., Elliott G., Yui N. Gauss polynomials and the rotation algebra, *Invent. Math.* **99** (1990), No. 2, 225-246.
- [GE] de Gennes P. G. Diamagnétisme de grains supraconducteurs près d'un seuil de percolation, *C.R. Acad. Sci.*, **B292** (1981), 9-12; Champ critique d'une boucle supraconductrice ramifiée, *C. R. Acad. Sci.*, **B292** (1981), 279-282.
- [GH] Guillement J. P., Helffer B., Treton P. Walk inside Hofstadter's butterfly, *J. Phys. France*, **50** (1989), 2019-2058.
- [HA] Harper P. G. Single band motion of conduction electrons in a uniform magnetic field, *Proc. Phys. Soc. London*, **A68** (1955), 874.
- [HR] Helffer B., Robert D. Calcul fonctionnel de la transformée de Mellin et applications, *J. Funct. Anal.*, **53** (1983), No. 3, 246-268.
- [HS0] Helffer B., Sjöstrand J. Microlocal Analysis for the Periodic Magnetic Schrödinger Equation and Related Questions, *C.I.M.E. Lectures*, 1989.
- [HS1] Helffer B., Sjöstrand J. Analyse semi-classique pour l'équation de Harper (avec application à l'équation de Schrödinger avec champ magnétique), *Mém. Soc. Math. France* (1988), No. 34, 1-113.
- [HS2] Helffer B., Sjöstrand J. Analyse semi-classique pour l'équation de Harper II. Comportement semi-classique près d'un rationnel, *Mém. Soc. Math. France* (1990), No. 40, 1-139.
- [HS3] Helffer B., Sjöstrand J. Semi-classical analysis for Harper's equation III, *Mém. Soc. Math. France* (1989), No. 39, 1-124.
- [HO] Hofstadter D. R. Energy levels and wave functions of Bloch electrons in a rational or irrational magnetic field, *Phys. Rev.*, **B14** (1976), 2239.
- [KL] von Klitzing K., Dorda G., Pepper M. Realization of a resistance standard based on fundamental constants, *Phys. Rev. Letters*, **45** (1980), 494-497.
- [LA] Landau L. Diamagnetismus der Metalle, *Z. für Phys.*, **64** (1930), 629-637.
- [LU] Luttinger J. M. The effect of a magnetic field on electrons in a periodic potential, *Phys. Rev.*, **84** (1951), 814.
- [MO] van Mouche P. The coexistence problem for the discrete Mathieu operator, *Comm. Math. Phys.*, **122** (1989), 23-34.

- [PCR] Pannetier B., Chaussy J., Rammal R. Experimental determination of the (H, T) phase diagram of a superconducting network, J. Phys. Lettres, **44** (1983), L853-L858.
- [PCRV] Pannetier B., Chaussy J., Rammal R., Villegier J. C. Experimental fine tuning of the frustration: 2D superconducting network in a magnetic field, Phys. Rev. Letters, **53** (1984), 1845-1848.
- [PED] Pedersen G. C^* -Algebras and their automorphism groups, Academic Press, London, New-York, 1979.
- [PE] Peierls R. Zur Theorie des Diamagnetismus von Leitungselektronen, Z. für Phys., **80** (1933), 763-791.
- [RB1] Rammal R., Bellissard J. Ground state of the Fermi gas in 2D lattices with a magnetic field: new exact results, Europhys. Letters, **13** (1990), 205-210.
- [RB2] Rammal R., Bellissard J. An algebraic semi-classical approach to Bloch electrons in a magnetic field, J. Phys. France, **51** (1990), 1803-1830.
- [SI] Sinai Ya. G. Anderson localization for the one dimensional difference Schrödinger operator with quasiperiodic potentials, in: Proc. Int. Congress Math. Phys. Marseille, 1986, World Scientific Pub. Co., Singapore, 1987, 870-903.
- [SO] Sokoloff J. B. Unusual band structure, wave functions and electrical conductance in crystals with incommensurable periodic potentials, Phys. Reports, **126** (1985), 189-244.
- [TH] Thouless D. Bandwidths for a quasiperiodic tight-binding model, Phys. Rev., **B28** (1983), 4272-4276.
- [TKN2] Thouless D., Kohmoto M., Nightingale M., den Nijs M. Quantized Hall conductance in two dimensional periodic potential, Phys. Rev. Letters, **49** (1982), 405.
- [WI] Wilkinson M. Critical properties of electron eigenstates in incommensurate systems, Proc. Roy. Soc. London, **A391** (1984), 305-350.
- [ZA] Zak J. Magnetic translation group, Phys. Rev., **A134** (1964), 1602-1607; Magnetic translation group II: Irreducible representations, Phys. Rev., **A134** (1964), 1607-1611.

JEAN BELLISSARD

Université Paul Sabatier
 Laboratoire de Physique Quantique — IRSAMC
 U.R.A. 505 du C.N.R.S.
 118, route de Narbonne
 F-31062 TOULOUSE CEDEX, FRANCE

«АВТОМАТИЧЕСКОЕ» ДОКАЗАТЕЛЬСТВО ТОЖДЕСТВ И ГИПЕРГЕОМЕТРИЧЕСКИЕ ФУНКЦИИ (по Д. Цайльбергеру)

Пьер Картье

ВВЕДЕНИЕ

Рассмотрим следующий набор тождеств между биномиальными коэффициентами¹⁾:

$$\sum_k \binom{n}{k} = 2^n, \quad \sum_k (-1)^k \binom{n}{k} = 0, \quad (1)$$

$$\sum_k \binom{n}{k}^2 = \binom{2n}{n}, \quad \sum_k (-1)^k \binom{2n}{k}^2 = (-1)^n \binom{2n}{n}, \quad (2)$$

$$\sum_k (-1)^k \binom{2n}{k}^3 = (-1)^n \binom{3n}{n} \binom{2n}{n}. \quad (\text{Диксон, 1890}). \quad (3)$$

Два первых тождества стандартны — они доказываются специализацией *формулы бинома*

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}. \quad (4)$$

Учитывая закон симметрии $\binom{n}{k} = \binom{n}{n-k}$, обе формулы (2) можно получить сравнением коэффициентов при подходящих мономах в тождествах

$$(x + y)^n (x + y)^n = (x + y)^{2n} \quad \text{и} \quad (x + y)^{2n} (x - y)^{2n} = (x^2 - y^2)^{2n}.$$

Тождество Диксона (3) не поддается такого рода методам; впрочем, не существует соответствующей формулы ни для $\sum_k \binom{2n}{k}^3$, ни для более высоких степеней биномиальных коэффициентов.

¹⁾ В соответствии с обычными соглашениями $\binom{n}{k} = 0$, если $n \geq 0$ и k не лежит между 0 и n . Следовательно, все приведенные ряды имеют лишь конечное число ненулевых членов и точные границы суммирования можно не указывать.

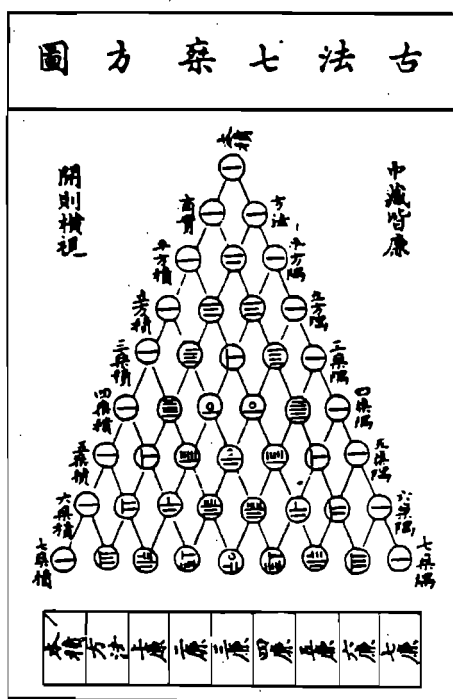


Рис. 1.

Треугольник Паскаля из труда Чжу Ши-цзе (Китай, 1303).

[Иллюстрация из: Collette J.-P. Histoire des mathématiques, Editions Vuibert.]

Во многих рассуждениях формулу бинома (4) можно использовать как *определение* биномиальных коэффициентов $\binom{n}{k}$. Поскольку степени вычисляются одна за другой по правилу

$$(x + y)^n = (x + y)^{n-1}(x + y),$$

биномиальные коэффициенты удовлетворяют *основной формуле сложения*,

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}. \quad (5)$$

Это позволяет производить быстрые вычисления при помощи таблицы, известной под названием «треугольник Паскаля» (и знакомой китайскому автору 1300 г., см. рисунок).

Определение при помощи формулы бинома приводит к *комбинаторной интерпретации*:

$\binom{n}{k}$ равно числу слов из n букв, в которых k раз встречается x и $n - k$ раз y .

Эквивалентным образом, $\binom{n}{k}$ равно числу k -элементных подмножеств в множестве из n элементов. Таким образом, формулы, связывающие биномиальные коэффициенты, интерпретируются в терминах перечисления комбинаторных структур (графов, деревьев, таблиц, путей, слов, ...). Во многих случаях эти формулы можно доказать построением явных *биекций* между различными типами структур. Именно таким способом Фоата и я установили в [Е 1] формулу Диксона и дали различные ее обобщения. Это направление сейчас активно развивается в лотарингской школе Фоата в Страсбурге и квебекской школе братьев Лабеллей.

В тождествах (1)–(3) результат суммирования — это «замкнутое» выражение, представленное как произведение степеней a^n и различных биномиальных коэффициентов. Результат не всегда можно выразить так просто: вот, прежде всего, элементарная формула

$$\sum_k \binom{n-k}{k} = F_n, \quad (6)$$

где F_n — это n -е число Фибоначчи, т.е. решение *рекуррентного уравнения*

$$F_n = F_{n-1} + F_{n-2} \quad \text{при} \quad n \geq 2, \quad F_0 = F_1 = 1. \quad (7)$$

Для доказательства иррациональности $\zeta(3)$ Апери [Е 2] должен был просуммировать следующее выражение:

$$u_n := \sum_k \binom{n}{k}^2 \binom{n+k}{k}^2; \quad (8)$$

ему удалось установить *рекуррентное соотношение*

$$n^3 u_n - (34n^3 - 51n^2 + 27n - 5)u_{n-1} + (n-1)^3 u_{n-2} = 0 \quad (n \geq 2) \quad (9)$$

с начальными условиями $u_0 = 1$, $u_1 = 5$. Апери не приводит доказательства формулы (9), оно было предложено Загиром и Коэном

[Е 3]. Идея следующая: положим $F(n, k) = \binom{n}{k}^2 \binom{n+k}{k}^2$; тогда существует некоторая функция $G(n, k)$, удовлетворяющая тождеству

$$G(n, k) - G(n, k-1) = n^3 F(n, k) - (34n^3 - 51n^2 + 27n - 5)F(n-1, k) + (n-1)^3 F(n-2, k). \quad (10)$$

Очевидно, что

$$u_n = \sum_k F(n, k) \quad (11)$$

и, кроме того, при фиксированном n «телескопическая сумма»

$$\sum_k (G(n, k) - G(n, k-1))$$

равна нулю. Соотношение (9) вытекает из (10) после суммирования по k . Этот метод по-английски называется «creative telescoping».

Аналогичным образом можно доказать классическое интегральное тождество

$$\int_{-\infty}^{+\infty} e^{2\pi i x y} e^{-\pi y^2} dy = e^{-\pi x^2}. \quad (12)$$

Обозначим подинтегральную функцию через $F(x, y)$, а затем положим

$$u(x) := \int F(x, y) dy. \quad (13)$$

Нужно доказать равенство $u(x) = e^{-\pi x^2}$, которое равносильно тому, что¹⁾

$$D_x u(x) + 2\pi x u(x) = 0, \quad u(0) = 1. \quad (14)$$

Определим некоторую функцию $G(x, y)$, для которой

$$D_x F(x, y) + 2\pi x F(x, y) = D_y G(x, y). \quad (15)$$

Уравнение (14) выводится из (15) интегрированием по y с учетом двух основных правил — *дифференцирования под знаком интеграла* и *интегрирования по частям*:

$$D_x \int_{-\infty}^{+\infty} F(x, y) dy = \int_{-\infty}^{+\infty} D_x F(x, y) dy, \quad (16)$$

¹⁾ $D_x, D_y, \dots$ обозначают производные по $x, y, \dots$.

$$\int_{-\infty}^{+\infty} D_y G(x, y) dy = 0, \quad \text{если } G(x, \pm\infty) = 0. \quad (17)$$

Следовательно, интегральная формула (12) восстанавливается по дифференциальному соотношению (15) ценой нахождения функции $G(x, y)$.

В чем причина успеха в обоих случаях? Сначала напомним, что биномиальные коэффициенты можно выразить через факториалы по формуле

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \quad (18)$$

Поскольку факториал $n!$ определяется условиями

$$n! = n \cdot (n-1)! \quad \text{при } n \geq 1, \quad 0! = 1, \quad (19)$$

формула (18) непосредственно следует из фундаментального открытия Паскаля, которое можно выразить так:

$$\binom{n-1}{k} = \frac{n-k}{k} \binom{n}{k}, \quad \binom{n-1}{k-1} = \frac{k}{n} \binom{n}{k} \quad (20)$$

(формула сложения является его очевидным следствием!). В примере Апери, следовательно,

$$\frac{F(n+1, k)}{F(n, k)} = \left(\frac{n+k+1}{n-k+1} \right)^2, \quad (21)$$

$$\frac{F(n, k+1)}{F(n, k)} = \frac{1}{(k+1)^4} \left(\frac{n+k+1}{n-k} \right)^2. \quad (22)$$

Оказывается также, что и отношение $G(n, k)/F(n, k)$ — рациональная функция,

$$\frac{G(n, k)}{F(n, k)} = 4 \left(\frac{n-k}{n+k} \right)^2 (2n-1) \{k(2k+1) - (2n-1)^2\}. \quad (23)$$

Для доказательства тождества (10) Загира и Коэна можно разделить обе его части на $F(n, k)$, и останется простое, но непонятное вычисление с рациональными функциями от n и k .

Аналогично, в интегральном примере, рассмотренном выше, логарифмические производные функции $F(x, y) = e^{2\pi i x y} \cdot e^{-\pi y^2}$ — рациональные функции:

$$\frac{D_x F(x, y)}{F(x, y)} = 2\pi i y, \quad \frac{D_y F(x, y)}{F(x, y)} = 2\pi i (x + i y), \quad (24)$$

и для функции $G(x, y)$, удовлетворяющей дифференциальному уравнению (15), справедливо равенство

$$\frac{G(x, y)}{F(x, y)} = -i. \quad (25)$$

Доказательство дифференциального соотношения (15) в результате сводится (после деления на $F(x, y)$) к доказательству совсем элементарного тождества между рациональными функциями от x и y . Для завершения доказательства интегрального соотношения (12) достаточно его проверить при «начальном» условии $x = 0$, т. е. установить соотношение $\int_{-\infty}^{+\infty} e^{-\pi y^2} dy = 1$.

Все это относится к *стадии проверки*, но как обстоит дело со *стадией индукции*? Исходя из вида функции $F(n, k) = \binom{n}{k}^2 \binom{n+k}{k}^2$, можем ли мы утверждать, что существует рекуррентное уравнение вида (9), которому удовлетворяет сумма $u_n = \sum_k F(n, k)$, и, более того, что существует функция $G(n, k)$, удовлетворяющая соотношению (10)? Как явно найти рекуррентное уравнение и функцию $G(n, k)$?

В замечательной серии статей Цайльбергер смог дать два ответа на эти вопросы:

- *теоретический* ответ, использующий теорию дифференциальных голономных систем Бернштейна,
- *практический* ответ, позволяющий находить алгоритмы, которые дают рекуррентное соотношение и $G(n, k)$.

Все приведенные выше соотношения, включая необычную рекуррентную формулу Апери, получаются заново систематическим и автоматическим способом, и предлагается техника, которая позволяет *открывать* и *доказывать* тождества некоторого типа. Вероятно, близок день, когда классические сборники специальных функций будут вытеснены экспертными системами, например расширением MAPLEa.

1. ГОЛОНОМНЫЕ МОДУЛИ [С 3]

1.1. Обозначим через $\mathcal{A}_r$ кольцо *дифференциальных операторов с полиномиальными коэффициентами*, действующих на функции от r переменных; константами являются комплексные числа. Всякий оператор единственным образом записывается в виде

$$P = \sum_{\alpha, \beta} c_{\alpha, \beta} x^\alpha D^\beta. \quad (26)$$

Обозначения классические:

- $x = (x_1, \dots, x_r)$ обозначает систему из r переменных, D_i — частная производная по x_i ;
- $\alpha, \beta, \dots$ обозначают элементы из Z_+^r , т. е. векторы $\alpha = (\alpha_1, \dots, \alpha_r)$ с целыми неотрицательными координатами;
- кроме того,

$$|\alpha| = \alpha_1 + \dots + \alpha_r, \quad x^\alpha = x_1^{\alpha_1} \dots x_r^{\alpha_r}, \quad D^\beta = D_1^{\beta_1} \dots D_r^{\beta_r}.$$

В формуле (26) лишь конечное число коэффициентов $c_{\alpha, \beta}$ отлично от 0.

Порядок m оператора P равен максимальному из чисел $|\beta|$ для таких β , что $c_{\alpha, \beta} \neq 0$; главный символ $\sigma(P)$ оператора P — это полином $\sum_{|\beta|=m} c_{\alpha, \beta} x^\alpha \xi^\beta$ от переменных $(x, \xi) = (x_1, \dots, x_r, \xi_1, \dots, \xi_r)$; он однороден степени m по ξ . Вот основные правила:

- если P имеет порядок m , а Q — порядок n , то порядок PQ равен $m+n$, а порядок коммутатора $[P, Q] = PQ - QP$ не превосходит $m+n-1$;
- имеют место равенства¹⁾

$$\sigma(PQ) = \sigma(P)\sigma(Q), \quad \sigma([P, Q]) = \{\sigma(P), \sigma(Q)\}. \quad (27)$$

В последней формуле мы используем скобку Пуассона двух полиномов от x, ξ , определяемую следующим образом:

$$\{F, G\} = \sum_{j=1}^r \left[\frac{\partial F}{\partial \xi_j} \cdot \frac{\partial G}{\partial x_j} - \frac{\partial G}{\partial \xi_j} \cdot \frac{\partial F}{\partial x_j} \right]. \quad (28)$$

Все это можно представить и более алгебраично. Как алгебра над полем $\mathbb{C}$ комплексных чисел $\mathcal{A}_r$ определяется $2r$ образующими x_i и D_j и соотношениями

$$[x_i, x_j] = 0, \quad [D_i, D_j] = 0, \quad [D_i, x_j] = \delta_{ij}. \quad (29)$$

Обозначим через $F^m = F^m \mathcal{A}_r$ множество операторов порядка $\leq m$. Это векторное пространство над $\mathbb{C}$, последовательность (F^m) является возрастающей ($F^0 \subset F^1 \subset F^2 \subset \dots$), $F^0 = \mathbb{C}[x_1, \dots, x_r]$ и приведенные выше правила записываются в виде

$$F^m \cdot F^n \subset F^{m+n}, \quad [F^m, F^n] \subset F^{m+n-1}. \quad (30)$$

¹⁾ Во второй формуле предполагается, что порядок коммутатора $[P, Q]$ равен $m+n-1$, что эквивалентно соотношению $\{\sigma(P), \sigma(Q)\} \neq 0$.

Благодаря этим свойствам можно определить структуру алгебры в ассоциированном градуированном пространстве

$$\text{gr}^F A_r = \bigoplus_{m \geq 0} F^m / F^{m-1};$$

определим изоморфизм этой алгебры на алгебру полиномов $C[x, \xi]$, отображая в $\sigma(P)$ класс $P \bmod F^{m-1}$ (для всех P из F^m / F^{m-1}). Кольцо A_r не коммутативно, но поскольку $\text{gr}^F A_r$ коммутативно, можно считать A_r почти коммутативным и многие свойства кольца полиномов распространяются на A_r .

1.2. Мы можем действовать дифференциальными операторами из A_r на полиномы от x , на функции класса C^∞ , определенные в $\mathbf{R}^r$ (или в области из $\mathbf{R}^r$), на функции, аналитические в $\mathbf{R}^r$, или функции, голоморфные в $\mathbf{C}^r$, а также на различные пространства обобщенных функций в $\mathbf{R}^r$.

Пусть, например, f — функция класса C^∞ в $\mathbf{R}^r$. Введем множество $I_f \subset A_r$, образованное дифференциальными операторами $P \in A_r$, такими, что $Pf = 0$; это *левый идеал* в A_r . Главные символы $\sigma(P)$, соответствующие операторам P из I_f , являются ненулевыми однородными элементами идеала J_f в $C[x, \xi]$. С этим полиномиальным идеалом связано алгебраическое многообразие в $\mathbf{C}^{2r}$, определенное уравнениями

$$\sigma(P)(x, \xi) = 0 \quad \text{для всех } P \neq 0 \text{ из } I_f.$$

Это *характеристическое многообразие* V_f функции f .

Глубокая теорема Касивары—Кавайи—Сато утверждает, что многообразие V_f *инволютивно*; доказательство можно найти в докладе [С. 11] Мальгранжа на этом семинаре. Смысл теоремы заключен в следующем: пусть J_f^0 — идеал в $C[x, \xi]$, образованный полиномами, множество нулей которых содержит V_f . По теореме Гильберта о нулях полином F содержится в J_f^0 тогда и только тогда, когда некоторая его степень содержится в J_f . Тогда J_f^0 *устойчиво относительно скобки Пуассона* (соответствующее утверждение о J_f тотчас же следует из второй формулы (27)).

Из классического результата симплектической геометрии следует, что *размерность* многообразия V_f *не меньше* r : если $p = (x^{(0)}, \xi^{(0)})$ — гладкая точка многообразия V_f , то касательное пространство к V_f в точке p ортогонально дифференциалам $d_p F$

для F , пробегающего J_f^0 . Но существует билинейная симплектическая форма J_p на кокасательном пространстве в точке p , такая, что

$$\{F, G\}(p) = J_p(d_p F, d_p G). \quad (31)$$

Если F и G лежат в J_f^0 , то и $\{F, G\}$ лежит в J_f^0 по инволютивности, откуда следует, что $J_p(d_p F, d_p G) = 0$. Отсюда вытекает, что множество дифференциалов $d_p F$ для F из J_f^0 образует векторное пространство размерности $\leq r$, а по двойственности касательное пространство к V_f в точке p имеет размерность $\geq 2r - r = r$.

1.3. Говорят, что функция (или обобщенная функция) f голономна¹⁾, если характеристическое многообразие V_f имеет размерность r (наименьшую возможную).

На практике трудно явно найти левый идеал I_f в $\mathcal{A}_r$, а тем более идеалы J_f и J_f^0 в $\mathbb{C}[x, \xi]$. Для проверки того, что функция f голономна, выписывают некоторое количество дифференциальных уравнений

$$P_1 f = \dots = P_m f = 0.$$

Пусть σ_j — главный символ оператора P_j . На многообразии V_f равны нулю полиномы σ_j , а также скобки Пуассона $\{\sigma_j, \sigma_k\}$, $\{\sigma_j, \{\sigma_k, \sigma_l\}\}$ и так далее, поскольку идеал J_f устойчив относительно скобки Пуассона. Если можно найти достаточно уравнений вида $\sigma_j = 0$, $\{\sigma_j, \sigma_k\} = 0$, ... для того, чтобы множество их решений W в $\mathbb{C}^{2r}$ имело размерность $\leq r$, то мы получим $\dim V_f \leq \dim W \leq r$ и f голономна.

Примеры. 1) Если $r = 1$, то функция, голономная в $\mathbb{C}$, — это функция, которая удовлетворяет дифференциальному уравнению

$$\sum_{j=0}^m p_j(z) D^j f(z) = 0, \quad (32)$$

где $p_0(z), \dots, p_m(z)$ — полиномы и $p_m(z) \neq 0$. Соответствующий символ равен $p_m(z)\zeta^m$ и характеристическое многообразие содержится в

$$\{\zeta = 0\} \cup \{z = z_{(1)}\} \cup \dots \cup \{z = z_{(d)}\},$$

¹⁾ Бьёрк в [С 3] называет их *функции класса Бернштейна*.

где $z_{(1)}, \dots, z_{(d)}$ — корни полинома $p_m(z)$. Например, *гипергеометрическая функция Гаусса* — это решение уравнения

$$z(1-z) \frac{d^2 F}{dz^2} + \{c - (a+b+1)z\} \frac{dF}{dz} - abF = 0;$$

она голономна, и ее характеристическое многообразие — это

$$\{z = 0\} \cup \{z = 1\} \cup \{\zeta = 0\}$$

в $\mathbb{C}^2$.

2) Полином $f(z)$ в $\mathbb{C}^r$ удовлетворяет уравнениям

$$D_1^m f = \dots = D_r^m f = 0$$

при достаточно больших m . Следовательно, его характеристическое многообразие содержится в линейном многообразии $\zeta_1 = \dots = \zeta_r = 0$ размерности r , а значит, f — голономная функция.

3) *Обобщенная функция Дирака* δ в $\mathbb{R}^r$ удовлетворяет уравнениям

$$x_1 \delta = \dots = x_r \delta = 0;$$

она голономна, и $x_1 = \dots = x_r = 0$ — ее характеристическое многообразие.

1.4. Если I — левый идеал в $\mathcal{A}_r$, то назовем его *характеристическим многообразием* алгебраическое многообразие $V(I)$ в $\mathbb{C}^{2r}$, пересечение нулей полиномов $\sigma(P)$, где P пробегает I (и $P \neq 0$). Это еще и инволютивное многообразие, а значит, его размерность $\geq r$. Говорят, что *идеал I голономен*, если размерность многообразия $V(I)$ равна r . Следовательно, функция f голономна тогда и только тогда, когда идеал I_f голономен.

Пусть S — кольцо полиномов $\mathbb{C}[u_1, \dots, u_s]$. Известно, что S имеет *гомологическую размерность* s , иначе говоря, всякий S -модуль M конечного типа обладает резольвентой

$$0 \rightarrow L_s \rightarrow L_{s-1} \rightarrow \dots \rightarrow L_0 \rightarrow M \rightarrow 0,$$

где L_i — проективные S -модули конечного типа, и существует S -модуль M , который не обладает более короткой резольвентой. Это то же самое, что сказать, что $\text{Ext}_S^i(M, N) = 0$ при $i > s$ для любых S -модулей конечного типа M и N и что существуют M и N , такие, что $\text{Ext}_S^s(M, N) \neq 0$.

Пусть M есть S -модуль конечного типа. Пусть J — идеал в S , образованный такими p , что $pM = 0$, а W — алгебраическое многообразие в $\mathbb{C}^s$, определенное уравнениями $p(u) = 0$, где p пробегает J (его можно назвать *характеристическим многообразием модуля M*). Обозначим через d или $d(M)$ размерность многообразия W . Классическая теорема утверждает, что

$$\begin{cases} \text{Ext}_S^j(M, S) = 0 & \text{при } 0 \leq j < s - d, \\ \text{Ext}_S^{s-d}(M, S) \neq 0. \end{cases} \quad (33)$$

Вернемся к кольцу A_r с фильтрацией $(F^m)_{m \geq 0}$; для удобства положим $F^m = 0$ при $m < 0$. Далее, пусть M — левый A_r -модуль конечного типа, и пусть $u_1, \dots, u_t$ — система его образующих; мы выберем также целые положительные числа $d_1, \dots, d_t$ и положим

$$\Gamma^m M = F^{m-d_1} \cdot u_1 + \dots + F^{m-d_t} \cdot u_t. \quad (34)$$

Возрастающая последовательность $\Gamma^0 M \subset \Gamma^1 M \subset \dots$, получающаяся таким образом, называется «хорошей» *фильтрацией модуля M* . Поскольку $F^m \cdot \Gamma^n M \subset \Gamma^{n+m} M$, можно определить ассоциированный градуированный объект $\text{gr}^\Gamma M = \bigoplus_{m \geq 0} \Gamma^m M / \Gamma^{m-1} M$, который является модулем над кольцом $\text{gr}^F A_r$. Этот модуль имеет конечный тип¹⁾. Так как кольцо $\text{gr}^F A_r$ является кольцом полиномов $\mathbb{C}[x_1, \dots, x_r, \xi_1, \dots, \xi_r]$, к нему применимо изложенное выше. Введем характеристическое многообразие $W(\text{gr}^\Gamma M)$; учитывая определение (34), легко показать, что характеристическое многообразие $W(\text{gr}^\Gamma M)$ не зависит от Γ , а значит, его можно обозначить через $W(M)$. Оно инволютивно, а следовательно, имеет размерность $\geq r$.

Применяя приведенные выше результаты к $S = \mathbb{C}[x_1, \dots, x_r, \xi_1, \dots, \xi_r]$ и используя спектральную последовательность, связанную с фильтрацией $(\Gamma^m M)_{m \geq 0}$, можно показать, что

$$\begin{cases} \text{Ext}_{A_r}^j(M, A_r) = 0 & \text{при } 0 \leq j < 2r - d, \\ \text{Ext}_{A_r}^{2r-d}(M, A_r) \neq 0, \end{cases} \quad (35)$$

где $d = d(M)$ — размерность характеристического многообразия $W(M) \subset \mathbb{C}^{2r}$. Это дает описание $d(M)$, не зависящее ни от фильтрации алгебры A_r , ни от фильтрации A_r -модуля M . Справедливо также соотношение

$$\text{Ext}_{A_r}^j(M, A_r) = 0 \quad \text{при } j > r,$$

¹⁾ Это свойство дает внутреннюю характеристику «хороших» фильтраций.

и отсюда достаточно просто выводится, что *гомологическая размерность кольца A_r равна r* (а не $2r$, как у $\text{gr}^F A_r$).

Разумеется, говорят, что A_r -модуль M голономен¹⁾, если $d(M) = r$; это то же самое, что предположить, что

$$\text{Ext}_{A_r}^j(M, A_r) = 0 \quad \text{при } j \neq r. \quad (36)$$

Правый A_r -модуль $\text{Ext}_{A_r}^r(M, A_r)$ играет тогда роль двойственного к M ; он также голономен.

1.5. Главное новшество, предложенное Бернштейном, заключается во введении другой фильтрации: $B^m = B^m A_r$ состоит из дифференциальных операторов вида

$$P = \sum_{|\alpha|+|\beta| \leq m} c_{\alpha\beta} x^\alpha D^\beta. \quad (37)$$

Иначе говоря, учитывается полная степень по x_i и D_i , а не только частичная степень по D^i . Можно повторить почти дословно то, что было изложено для фильтрации (F^m) :

- *ассоциированное градуированное кольцо $\text{gr}^B A_r$, также изоморфное²⁾ кольцу полиномов $C[x_1, \dots, x_r, \xi_1, \dots, \xi_r]$: B -символ оператора P равен*

$$\sigma_B(P) = \sum_{|\alpha|+|\beta|=m} c_{\alpha\beta} x^\alpha \xi^\beta \quad (38)$$

в обозначениях (37);

- *хорошая B -фильтрация A_r -модуля M конечного типа, определяемая формулой*

$$\Delta^m M = B^{m-d_1} \cdot u_1 + \dots + B^{m-d_t} \cdot u_t; \quad (39)$$

- *ассоциированный градуированный модуль $\text{gr}^\Delta M$; это модуль конечного типа над $\text{gr}^B A_r$;*

- *B -характеристическое многообразие модуля $\text{gr}^\Delta M$; оно не зависит от выбора хорошей B -фильтрации модуля M ;*

¹⁾ Левый идеал I в A_r голономен тогда и только тогда, когда левый A_r -модуль A_r/I голономен. Если f — функция в $\mathbb{R}^r$, то A_r -модуль A_r/I_f изоморфен $A_r \cdot f$ (множество операторов $P \cdot f$, где P пробегает A_r).

²⁾ Используемая в $C[x, \xi]$ градуировка задается *полной степенью* по x, ξ .

- размерность $\delta = \delta(M)$ B -характеристического многообразия A_r -модуля M ;

- описание $\delta(M)$: наименьшее целое j , такое, что $\text{Ext}_{A_r}^j(M, A_r) \neq 0$, равно $2r - \delta(M)$.

Из последнего утверждения выводится, что размерность $d(M)$ характеристического многообразия модуля M равна размерности $\delta(M)$ его B -характеристического многообразия (сами многообразия различны).

Размерность $\delta = \delta(M)$ можно вычислять так: если A_r -модуль конечного типа M снабжен хорошей B -фильтрацией $(\Delta^m M)_{m \geq 0}$, причем все векторные пространства $\Delta^m M / \Delta^{m-1} M$ конечномерны над C , то можно ввести ряд Пуанкаре

$$\chi_\Delta(t) = \sum_{m \geq 0} t^m \cdot \dim \Delta^m M / \Delta^{m-1} M = (1-t) \sum_{m \geq 0} t^m \cdot \dim \Delta^m M. \quad (40)$$

Так как $\text{gr}^\Delta M$ является градуированным модулем конечного типа над кольцом полиномов, хорошо известно (ряд Гильберта—Самуэля), что $\chi_\Delta(t)$ — рациональная функция вида $Q(t)/(1-t)^\delta$, где полином $Q(t)$ удовлетворяет условию $Q(1) \neq 0$. Из формулы (40) выводится существование целого m_0 и полинома $X(u)$ с рациональными коэффициентами, таких, что

$$\dim \Delta^m M = X(m) \quad \text{при} \quad m \geq m_0.$$

Кроме того, δ равна степени полинома $X(u)$.

Все результаты из пп. 1.4 и 1.5 доказываются независимо от теоремы инволютивности Касивары—Кавайи—Сато. Более того, существует очень простое непосредственное доказательство неравенства $\delta \geq r$ (см. [С 6, р. 178]), откуда получается новое доказательство неравенства $d \geq r$ (где d — размерность характеристического многообразия модуля M).

1.6. Приведем *практическое руководство* по голономным модулям.

A_r -модуль M — это модуль над кольцом полиномов $C[x_1, \dots, x_r]$, снабженный C -линейными операторами дифференцирования $D_1, \dots, D_r$, коммутирующими друг с другом и удовлетворяющими правилу Лейбница

$$D_i(F \cdot \Phi) = D_i F \cdot \Phi + F \cdot D_i \Phi \quad (41)$$

при $1 \leq i \leq r$, где F лежит в $\mathbb{C}[x_1, \dots, x_r]$, а Φ — в M . Например, всякое пространство функций или обобщенных функций от r переменных, которое устойчиво относительно умножения на x_i и дифференцирования по x_i (при $1 \leq i \leq r$), является A_r -модулем.

Элемент Φ из A_r -модуля M называется голономным, если существует константа C , удовлетворяющая следующему условию:

(Н) для любого целого $m \geq 0$ векторное подпространство в M , порожденное элементами $x^\alpha D^\beta \Phi$ при $|\alpha| + |\beta| \leq m$, имеет размерность $\leq C m^r$.

Голономные элементы A_r -модуля образуют в нем A_r -подмодуль. Голономный модуль — это A_r -модуль конечного типа, все элементы которого голономны.

Рациональный A_r -модуль — это векторное пространство V над полем $\mathbb{C}(x_1, \dots, x_r)$ рациональных функций, снабженное $\mathbb{C}$ -линейными операторами $D_1, \dots, D_r$, коммутирующими между собой и удовлетворяющими правилу Лейбница. Пусть Φ — голономный элемент из V ; тогда существуют ненулевые дифференциальные операторы $P_1, \dots, P_r$, аннулирующие Φ и имеющие следующий вид:

$$P_i = \sum_{j=0}^{s_i} p_{ij}(x_1, \dots, x_r) D_i^j \quad (42)$$

(p_{ij} — полиномы). Векторное подпространство в V , порожденное над $\mathbb{C}(x_1, \dots, x_r)$ конечным числом производных $D^\alpha \Phi$ при $0 \leq \alpha_1 \leq s_1, \dots, 0 \leq \alpha_r \leq s_r$, содержит Φ и инвариантно относительно дифференцирований $D_1, \dots, D_r$.

Говорят, что элемент $\Phi \in V$ рационально голономен¹⁾, если он содержится в конечномерном над $\mathbb{C}(x_1, \dots, x_r)$ векторном подпространстве W , инвариантном относительно $D_1, \dots, D_r$ (иначе говоря, W — рациональный A_r -подмодуль). Определим аннулятор I_Φ элемента Φ : это левый идеал в A_r , с которым ассоциировано характеристическое многообразие $V(I_\Phi)$ в $\mathbb{C}^{2r}$ (ср. п. 1.4). Тогда Φ рационально голономен в том и только том случае, когда существует ненулевой полином $H(x_1, \dots, x_r)$, такой, что $V(I_\Phi)$ содержится в объединении гиперповерхности $H = 0$ в $\mathbb{C}^{2r}$ и некоторого многообразия размерности r . Согласно обсуждению из предыдущего абзаца, всякий голономный элемент рационально голономен.

¹⁾ D -конечен в терминологии Цайльбергера.

2. ИЗУЧЕНИЕ ГОЛОНОМНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ И ФУНКЦИЙ

2.1. *Гипергеометрическим рядом* по традиции называется всякий ряд вида $\sum_{n \geq 0} R(n)$, где $R(0) = 1$ и $R(n+1)/R(n)$ является рациональной функцией от n . Разлагая рациональную функцию в произведение линейных сомножителей, можно записать $R(n)$ в виде

$$R(n) = \frac{(a_1)_n \dots (a_p)_n}{(b_1)_n \dots (b_q)_n} \frac{z^n}{n!}, \quad (43)$$

где используется следующее (классическое!) обозначение:

$$(a)_n = a(a+1) \dots (a+n-1), \quad (44)$$

т. е.

$$(a)_0 = 1, \quad (a)_{n+1} = (a)_n(a+n). \quad (45)$$

Сумма ряда $\sum_{n \geq 0} R(n)$ тогда обозначается через ${}_pF_q \left(\begin{smallmatrix} a_1 \dots a_p \\ b_1 \dots b_q \end{smallmatrix} \middle| z \right)$ в соответствии с классическим определением *гипергеометрической функции* [А 2]; функция Гаусса — это функция

$${}_2F_1 \left(\begin{smallmatrix} a \ b \\ c \end{smallmatrix} \middle| z \right) = \sum_{n \geq 0} \frac{(a)_n (b)_n}{(c)_n n!} z^n. \quad (46)$$

На пространстве функций целочисленной переменной n определено действие разностных операторов вида¹⁾

$$(Pu)(n) = \sum_{\alpha \geq 0} \sum_{\beta} c_{\alpha\beta} n^{\alpha} u(n+\beta). \quad (47)$$

В частности, по формуле

$$(Nu)(n) = u(n+1) \quad (48)$$

вводится оператор сдвига N и предыдущий оператор записывается в виде $P(n, N) = \sum_{\alpha, \beta} c_{\alpha\beta} n^{\alpha} N^{\beta}$. Имеется коммутационное соотношение

$$N \cdot n = (n+1) \cdot N. \quad (49)$$

¹⁾ β пробегает все целые числа.

Говорят, что последовательность *u* голономна, если существует ненулевой оператор $P(n, N)$, такой, что $P(n, N)u = 0$. Это так для гипергеометрической последовательности (43) (здесь

$$P(n, N) = (n+1)(n+b_1)\dots(n+b_q)N - z(n+a_1)\dots(n+a_p))$$

и для последовательности Аперри $u_n = \sum_{k=0}^n \binom{n}{k}^2 \binom{n+k}{k}^2$ (где

$$P(n, N) = n^3 - (34n^3 - 51n^2 + 27n - 5)N^{-1} + (n-1)^3 N^{-2}).$$

2.2. Все это обобщается на мультипоследовательности, т.е. на функции *u* с комплексными значениями, определенные в $\mathbf{Z}^s$ (или на подходящем подмножестве из $\mathbf{Z}^s$). С помощью формул

$$(K_i u)(k_1, \dots, k_s) = u(k_1, \dots, k_i + 1, \dots, k_s) \quad (50)$$

вводятся операторы сдвига $K_1, \dots, K_s$, а затем операторы

$$P(k, K) = \sum_{\alpha \in \mathbf{Z}_+^s} \sum_{\beta \in \mathbf{Z}^s} c_{\alpha\beta} k^\alpha K^\beta, \quad (51)$$

преобразующие функцию $u(k_1, \dots, k_s)$ в функцию

$$\sum_{\alpha} \sum_{\beta} c_{\alpha\beta} k_1^{\alpha_1} \dots k_s^{\alpha_s} u(k_1 + \beta_1, \dots, k_s + \beta_s).$$

Операторы вида (51) образуют алгебру $\mathcal{B}_s$, которая определяется образующими $k_1, \dots, k_s, K_1, \dots, K_s, K_1^{-1}, \dots, K_s^{-1}$ и соотношениями

$$k_i k_j = k_j k_i, \quad (52)$$

$$K_i K_j = K_j K_i, \quad (53)$$

$$k_i K_j = K_j k_i \quad \text{при} \quad i \neq j, \quad (54)$$

$$K_i k_i = (k_i + 1) K_i, \quad (55)$$

$$K_i K_i^{-1} = K_i^{-1} K_i = 1. \quad (56)$$

Эта алгебра хорошо известна: сопоставим функции $u : \mathbf{Z}^s \rightarrow \mathbf{C}$ формальный ряд Лорана

$$U(z_1, \dots, z_s) = \sum_{k_1, \dots, k_s} u(k_1, \dots, k_s) z_1^{k_1} \dots z_s^{k_s}. \quad (57)$$

Функции $k_i u$ соответствует ряд $z_i \partial U / \partial z_i$, а функции $K_i u$ — ряд $z_i^{-1} U$. Отождествление

$$k_i \leftrightarrow z_i \frac{\partial}{\partial z_i}, \quad K_i \leftrightarrow z_i^{-1},$$

определяет изоморфизм между B_s и локализацией кольца A_s , получаемой обращением элементов $z_1, \dots, z_s$.

Чтобы определить символ оператора $P(k, K)$, оставим члены наибольшей степени m по k :

$$\sigma(P) = \sum_{|\alpha|=m} \sum_{\beta} c_{\alpha\beta} \zeta^{\alpha} z^{\alpha-\beta} \quad (58)$$

(полином от $z_1, \dots, z_s, z_1^{-1}, \dots, z_s^{-1}, \zeta_1, \dots, \zeta_s$); функции $u(k_1, \dots, k_s)$ ставится в соответствие *характеристическое многообразие* V_u в $(\mathbb{C}^*)^s \times \mathbb{C}^s$ (с координатами $z_1, \dots, z_s, \zeta_1, \dots, \zeta_s$), определенное уравнениями $\sigma(P)(z, \zeta) = 0$, где P пробегает множество таких операторов, что $P \cdot u = 0$. Наконец, говорят, что u голономна, если размерность многообразия V_u равна s .

Примеры. 1) Говорят, что функция u имеет *гипергеометрический тип*, если $K_i u / u$ — рациональная функция от $k_1, \dots, k_s$ при $i = 1, \dots, s$. Она (рационально) голономна.

2) Говорят, что u *замкнута*, если она равна произведению сомножителей вида $(a)_{c_1 k_1 + \dots + c_s k_s}^{\pm 1}$, где $c_1, \dots, c_s$ — данные целые числа, и рациональной функции. Это частный случай последовательности гипергеометрического типа.

3) Рассмотрим функции $u(k)$, определенные в $\mathbb{Z}^s$, со значениями в $\mathbb{C}^N$ и удовлетворяющие уравнениям

$$K_i u(k) = A_i(k) \cdot u(k) \quad (1 \leq i \leq s). \quad (59)$$

В этой формуле $A_1(k), \dots, A_s(k)$ — матрицы размера $N \times N$, рационально зависящие от k и удовлетворяющие условиям согласованности¹⁾

$$K_i A_j \cdot A_i = K_j A_i \cdot A_j. \quad (60)$$

Компоненты функции $u(k)$ — рационально голономные последовательности, чаще всего голономные.

¹⁾ Эта система является дискретным аналогом дифференциальной системы $D_i F = A_i F$, где матрицы A_i удовлетворяют условию *вполне интегрируемости* $D_i A_j - D_j A_i = [A_i, A_j]$.

2.3. Можно также рассматривать *смешанный случай* функций $u(x, k)$ на $\mathbb{C}^r \times \mathbb{Z}^s$. Такой функции сопоставляется формальный ряд Лорана

$$U(x, z) = \sum_{k_1, \dots, k_s} u(x, k) z_1^{k_1} \dots z_s^{k_s}, \quad (61)$$

на который действуют дифференциальные операторы с коэффициентами, полиномиальными по $x_1, \dots, x_r, z_1, \dots, z_s, z_1^{-1}, \dots, z_s^{-1}$, откуда возникает еще одна теория голономных функций.

Пример-иллюстрация. Положим $u(x, k) = P_k(x)$ (k -й многочлен Лежандра). Используя обозначение D для d/dx и K для сдвига по k , запишем классические соотношения следующим образом:

$$\begin{cases} Pu = 0, & \text{дифференциальное уравнение,} \\ Qu = 0, & \text{рекуррентное соотношение,} \end{cases} \quad (62)$$

где

$$P = (1 - x^2)D^2 - 2xD + k(k+1), \quad (63)$$

$$Q = (k+2)K^2 - (2k+3)xK + k+1. \quad (64)$$

Тогда

$$\sigma(P) = (1 - x^2)\xi^2 + z^2\zeta^2, \quad (65)$$

$$\sigma(Q) = \zeta(z + z^{-1} - 2x) \quad (66)$$

(указание: заменить D на ξ , k на $z\xi$, K на z^{-1} и оставить члены наибольшей степени по (ξ, ζ)). Характеристическое многообразие определяется уравнениями $\sigma(P) = \sigma(Q) = \{\sigma(P), \sigma(Q)\} = 0$; оно имеет 4 компоненты, все размерности 2 (в пространстве $\mathbb{C}^3 \times \mathbb{C}^*$ с координатами ξ, ζ, x, z):

$$V: \quad 2x = z + z^{-1}, \quad 2\zeta = \xi(z^{-2} - 1),$$

$$V': \quad \xi = 0, \quad \zeta = 0,$$

$$V_+'' : \quad x = 1, \quad \zeta = 0,$$

$$V_-'' : \quad x = -1, \quad \zeta = 0.$$

Это голономный случай. Как следует из общей теоремы о голономных системах, решения системы (62) образуют пространство конечной размерности ν над $\mathbb{C}$ (здесь $\nu = 4$ соответствует начальным условиям $u(0, 0)$, $u(0, 1)$, $Du(0, 0)$, $Du(0, 1)$).

2.4. Согласно общей теории голономных систем, голономные функции на $C^r \times Z^s$ образуют векторное пространство над C , инвариантное относительно операторов x_j , D_j , k_i , $K_i^{\pm 1}$. Нетрудно также видеть, что если функции $u(x, k)$ и $v(y, m)$ голономны, то голономно и произведение $u(x, k)v(y, m)$ с разделенными переменными (тензорное произведение). Можно также производить линейные замены координат непрерывных переменных x_i и дискретных переменных k_i , а также сдвиги.

Пусть $\Phi(x)$ — голономная функция на C^r ; ее преобразование Лапласа задается формулой

$$L(\xi) = \int_c e^{-x \cdot \xi} \Phi(x) d^r x; \quad (67)$$

интеграл берется по подходящему вещественному многообразию размерности r в C^r , а $d^r x = dx_1 \wedge \dots \wedge dx_r$ и $x \cdot \xi = x_1 \xi_1 + \dots + x_r \xi_r$. Известно, что если $P(x, D_x)$ — дифференциальный оператор с полиномиальными коэффициентами, то преобразование Лапласа функции $P(x, D_x)\Phi(x)$ равно $Q(\xi, D_\xi)L(\xi)$, где оператор $Q(\xi, D_\xi)$ получается из $P(x, D_x)$ заменой x_j на $-\partial/\partial \xi_j$ и $\partial/\partial x_j$ на ξ_j . Если мы воспользуемся практическим руководством (см. п. 1.6), то убедимся, что это преобразование Лапласа — голономная функция¹⁾.

Рассмотрим теперь специализацию $x_r = 0$, которая функции $u(x_1, \dots, x_r)$ сопоставляет функцию

$$S_r u(x_1, \dots, x_r) = u(x_1, \dots, x_{r-1}, 0). \quad (68)$$

Предположим, что u голономна, и пусть M — голономный A_r -модуль, образованный функциями Pu , где P пробегает A_r . Отождествим A_{r-1} с подкольцом в A_r . Тогда ясно, что S_r есть A_{r-1} -линейное отображение и что $S_r(x_r v) = 0$, а значит, S_r определяет A_{r-1} -линейное отображение $\bar{S}_r$ из A_{r-1} -модуля $M/x_r M$ в пространство функций от $r-1$ переменных. Поскольку A_{r-1} -модуль $M/x_r M$ голономен (см. [С 6, р. 193]) и, кроме того, ясно, что фактормодуль голономного A_{r-1} -модуля голономен, получаем, что образ отображения S_r — голономный A_{r-1} -модуль, содержащий $S_r u$. Вывод: функция $S_r u$ голономна.

Комбинируя этот результат с линейными преобразованиями, получаем, например, что если $u(x, y, z)$ — голономная функция от трех

¹⁾ Более алгебраично, существует автоморфизм σ алгебры A_r , определенный формулами $\sigma(x_j) = -D_j$, $\sigma(D_j) = x_j$ при $1 \leq j \leq r$, и σ сохраняет фильтрацию Бернштейна в A_r .

переменных, то $u(x, x, y)$ — голономная функция от двух переменных. Из результата о тензорных произведениях выводим специализацией к диагонали, что обычное произведение двух голономных функций от r переменных голомно.

Все эти результаты распространяются на смешанные функции вида $u(x, k)$.

2.5. Скажем несколько слов о q -аналогах. Введем переменную q , которую можно специализировать к комплексному числу $q \neq 0$. На функции от s обратимых переменных $z_1, \dots, z_s$ действуют операторы гомотетии T_i :

$$T_i f(z_1, \dots, z_s) = f(z_1, \dots, qz_i, \dots, z_s). \quad (69)$$

Рассмотрим алгебру операторов вида

$$P = \sum_{\alpha \in \mathbb{Z}^s} \sum_{\beta \in \mathbb{Z}^s} c_{\alpha\beta} z^\alpha T^\beta, \quad (70)$$

т. е.

$$Pf(z_1, \dots, z_s) = \sum_{\alpha, \beta} c_{\alpha\beta} z^\alpha f(q^{\beta_1} z_1, \dots, q^{\beta_s} z_s). \quad (71)$$

Алгебра ${}_q C_s$ таких операторов определяется образующими $z_1, \dots, z_s, T_1, \dots, T_s$ и их обратными, подчиненными соотношениям

$$z_i z_j = z_j z_i, \quad T_i T_j = T_j T_i, \quad z_i T_j = T_j z_i \quad \text{при } i \neq j, \quad (72)$$

$$T_i z_i = q z_i T_i. \quad (73)$$

Сабба [D 4] недавно изложил основы теории таких алгебр по аналогии с результатами Бернштейна. Здесь также существует фильтрация по полной степени по z_j и T_j , определяются хорошие фильтрации ${}_q C_s$ -модулей конечного типа, вводится размерность δ и показывается, что она удовлетворяет условию $\delta \geq s$, а также условием $\delta = s$ определяются голономные ${}_q C_s$ -модули.

Тогда можно развивать теорию q -голономных функций. Например, функция q -гипергеометрического типа — это функция $F(z_1, \dots, z_s | q)$, для которой отношения $T_i F / F$ являются рациональными функциями от $z_1, \dots, z_s$ и q (при $1 \leq i \leq s$). Для одной переменной бесконечное произведение

$$(z; q)_\infty = \prod_{i \geq 0} (1 - q^i z) \quad (74)$$

является решением уравнения

$$\frac{(qz; q)_{\infty}}{(z; q)_{\infty}} = \frac{1}{1-z}. \quad (75)$$

Положим

$$(z; q)_n = \frac{(z; q)_{\infty}}{(q^n z; q)_{\infty}} = \prod_{i=0}^{n-1} (1 - q^i z); \quad (76)$$

аналог биномиальных коэффициентов

$$\left[\begin{matrix} n \\ k \end{matrix} \right]_q = \frac{(q; q)_n}{(q; q)_k (q; q)_{n-k}} \quad (77)$$

является полиномом от q степени $k(n-k)$, который при $q=1$ равен биномиальному коэффициенту $\binom{n}{k}$. Формулой сложения становится

$$\left[\begin{matrix} n \\ k \end{matrix} \right]_q = \left[\begin{matrix} n-1 \\ k \end{matrix} \right]_q + q^{n-k} \left[\begin{matrix} n-1 \\ k-1 \end{matrix} \right]_q, \quad (78)$$

и аналогом гипергеометрической функции Гаусса является «базисная гипергеометрическая» функция Гейне [А 20]

$${}_2\Phi_1 \left(\begin{matrix} a & b \\ c \end{matrix}; q; z \right) = \sum_{n \geq 0} \frac{(a; q)_n (b; q)_n}{(c; q)_n (q; q)_n} z^n. \quad (79)$$

Если заменить a на q^{α} , b на q^{β} и c на q^{γ} и положить $q=1$, то снова получим ${}_2F_1 \left(\begin{matrix} \alpha & \beta \\ \gamma \end{matrix} \middle| z \right)$.

3. ИНТЕГРИРОВАНИЕ И СУММИРОВАНИЕ ГОЛОНОМНЫХ ФУНКЦИЙ

3.1. Пусть $x = (x_1, \dots, x_r)$ и $x' = (x_1, \dots, x_{r-1})$, так что $x = (x', x_r)$. Исходя из функции $\Phi(x)$, определим интегрированием функцию

$$H(x') = \int \Phi(x', x_r) dx_r. \quad (80)$$

Можно, например, предположить, что x_r — вещественная переменная и что $\Phi(x', x_r)$ при фиксированном x' равна нулю вне компактного интервала. Можно также рассматривать интеграл Коши по

комплексной переменной x_r , где конечные части интегралов расходятся в смысле Адамара. Во всех этих примерах законны правила дифференцирования под знаком интеграла и интегрирования по частям.

Предположим, что $\Phi(x)$ — голономная функция; решение голономной системы

$$P_1(x, D)\Phi = \dots = P_N(x, D)\Phi = 0; \quad (81)$$

иначе говоря, левый идеал I в A_r , порожденный $P_1, \dots, P_N$, голономен и левый A_r -модуль $M := A_r/I$ голономен. По [С 6, р. 193] A_{r-1} -модуль $M/D_r M = A_r/(I + D_r A_r)$ голономен. Если положить

$$J = (I + D_r A_r) \cap A_{r-1}, \quad (82)$$

то A_{r-1} -модуль A_{r-1}/J голономен, потому что он изоморфен подмодулю в $M/D_r M$, и левый идеал J в A_{r-1} голономен. Однако J состоит из дифференциальных операторов вида

$$Q(x', D') = \sum_{j=1}^N A_j(x, D)P_j(x, D) + D_r B(x, D) \quad (83)$$

(здесь $D' = (D_1, \dots, D_{r-1})$). Согласно уравнениям (81),

$$Q(x', D')\Phi(x', x_r) = \partial\Phi_1(x', x_r)/\partial x_r, \quad (84)$$

где

$$\Phi_1(x) = B(x, D)\Phi(x). \quad (85)$$

Тогда

$$\begin{aligned} Q(x', D')H(x') &= Q(x', D') \int \Phi(x', x_r) dx_r \\ &= \int Q(x', D')\Phi(x', x_r) dx_r \\ &= \int \frac{\partial}{\partial x_r} \Phi_1(x', x_r) dx_r \\ &= 0 \end{aligned}$$

по правилам вычисления интегралов, упомянутым выше. Иначе говоря, интеграл $H(x')$ является решением голономной системы

$$Q_1(x', D')H = \dots = Q_M(x', D')H = 0, \quad (86)$$

где всякий оператор $Q_j(x', D')$ имеет вид (83).

3.2. Дискретный аналог описывается следующим образом: рассмотрим решение голономной системы

$$A_1(k, K)u(k) = \dots = A_N(k, K)u(k) = 0 \quad (87)$$

рекуррентных уравнений. Положим

$$v(k_1, \dots, k_{r-1}) = \sum_{k_r} u(k_1, \dots, k_{r-1}, k_r), \quad (88)$$

где, например, суммирование производится по конечному числу ненулевых значений. Тогда $v(k')$ удовлетворяет голономной системе

$$B_1(k', K')v(k') = \dots = B_M(k', K')v(k') = 0, \quad (89)$$

где все операторы $B_i(k', K')$ имеют вид

$$\sum_{j=1}^N C_j(k, K)A_j(k, K) + (K_r - 1)D(k, K). \quad (90)$$

Проблема состоит в том, чтобы произвести это вычисление явно.

Например, функция $F(n, k) = \binom{n}{k}$ является решением голономной системы $PF = P'F = 0$ с

$$P = (n - k + 1)N - (n + 1), \quad P' = (k + 1)K - (n - k) \quad (91)$$

(мы полагаем $NF(n, k) = F(n + 1, k)$ и $KF(n, k) = F(n, k + 1)$).

Простым вычислением получаем

$$(n + 1)(N - 2) = (K + 1)P + NP' - (K - 1)(Nn - n - 1).$$

Иначе говоря, функция $v(n) = \sum_k \binom{n}{k}$ удовлетворяет рекуррентному уравнению

$$(n + 1)(v(n + 1) - 2v(n)) = 0. \quad (92)$$

С учетом начального условия $v(0) = 1$ это приводит к желаемому ответу $v(n) = 2^n$ при $n \geq 0$. Предложенный метод не является самым простым для доказательства тождества $\sum_k \binom{n}{k} = 2^n$, однако имеет то преимущество, что приводит к программируемым алгоритмам.

3.3. Сначала мы опишем «медленный» алгоритм Цайльбергера.

а) Используя коммутационные соотношения

$$kn = nk, \quad KN = NK, \quad kN = Nk, \quad nK = Kn$$

и соотношения

$$Nn = (n+1)N, \quad Kk = (k+1)K, \quad (93)$$

можно показать, что всякий оператор $A(n, k, N, K)$ единственным образом записывается в виде

$$A(n, k, N, K) = \sum_{i=1}^b \sum_{j=0}^c a_{ij}(N, K) k^i n^j + R(n, N, K). \quad (94)$$

б) Рассмотрим два оператора P и P' с нормальными формами, указанными выше:

$$P = \sum_{i=1}^b \sum_{j=0}^c a_{ij}(N, K) k^i n^j + R(n, N, K), \quad (95)$$

$$P' = \sum_{i=1}^{b'} \sum_{j=0}^c a'_{ij}(N, K) k^i n^j + R'(n, N, K). \quad (96)$$

с) Рассмотрим $\mu = (b + b')[c(b + b' - 1) + 1]$ операторов

$$\begin{aligned} k^{\alpha'} n^{\beta} P & \quad \text{при } 0 \leq \alpha' < b', \quad 0 \leq \beta \leq c(b + b' - 1), \\ k^{\alpha} n^{\beta} P' & \quad \text{при } 0 \leq \alpha < b, \quad 0 \leq \beta \leq c(b + b' - 1). \end{aligned}$$

Запишем все эти μ операторов в нормальной форме (94): каждый из них будет суммой оператора, зависящего только от n, N и K , и линейной комбинации мономов $k^i n^j$ с коэффициентами из коммутативного кольца полиномов от N и K . Входящие в это представление мономы $k^i n^j$ удовлетворяют неравенствам

$$1 \leq i \leq b + b' - 1, \quad 0 \leq j \leq c(b + b'),$$

и их число равно $(b + b' - 1)[c(b + b') + 1]$, что совпадает с $\mu - 1$. Так как мы имеем μ линейных комбинаций с коэффициентами из

коммутативного кольца $C[N, K]$ от $\mu - 1$ мономов $k^i n^j$, из обычной линейной алгебры следует существование ненулевой линейной комбинации вида

$$\sum_{\alpha'=0}^{b'-1} \sum_{\beta=0}^{c(b+b'-1)} A_{\alpha'\beta}(N, K) k^{\alpha'} n^{\beta} P + \sum_{\alpha=0}^{b-1} \sum_{\beta=0}^{c(b+b'-1)} A'_{\alpha\beta} k^{\alpha} n^{\beta} P',$$

которая больше не содержит мономов $k^i n^j$, а следовательно, является оператором, зависящим только от n , N и K .

д) Итак, найдутся два оператора A и A' , такие, что $R = AP + A'P'$ больше не зависит от k . Тот факт, что это ненулевой оператор, проверяется при помощи предположения, что система $PF = P'F = 0$ голономна, т.е. символы $\sigma(P)$ и $\sigma(P')$ «независимы» в подходящем смысле.

е) Запишем R в виде $(K - 1)^{\lambda} R'(K, N, n)$. Поскольку $PF = P'F = 0$,

$$(K - 1)^{\lambda} R'(K, N, n)F = 0. \quad (97)$$

Это соотношение показывает, что функция $F' = R'(K, N, n)F$ является полиномом от k . Предположим, что $F(n, k) = 0$ при фиксированном n для достаточно большого $|k|$. Тогда $F'(n, k)$ тоже обладает этим свойством, а поскольку это полином от k , то $F'(n, k) = 0$.

ф) Запишем оператор R' в виде

$$R'(K, N, n) = \sum_{j=0}^d K^j R_j(N, n) \quad (98)$$

и положим

$$S(N, n) = \sum_{j=0}^d R_j(N, n). \quad (99)$$

Отсюда вытекает соотношение

$$R'(K, N, n) = S(N, n) + (K - 1)U(K, N, n). \quad (100)$$

В результате из соотношения $R'F = 0$ следует, что функция $a(n) = \sum_k F(n, k)$ удовлетворяет рекуррентному уравнению

$$S(N, n)a = 0. \quad (101)$$

3.4. Предыдущий алгоритм требует большого времени вычисления и особенно большой памяти. Гораздо более удобен алгоритм, использующий прием суммирования Госпера [С 8], который мы сейчас опишем.

Пусть $(S(k))_{k \geq 0}$ — последовательность гипергеометрического типа, такая, что

$$\frac{S(k)}{S(k-1)} = s(k) \quad \text{при} \quad k \geq 1, \quad (102)$$

где $s(k)$ — рациональная функция от k . Если положить

$$A(k) = S(k) - S(k-1), \quad (103)$$

то $(A(k))_{k \geq 0}$ будет последовательностью гипергеометрического типа. В самом деле, отношение $a(k) = A(k)/A(k-1)$ задается формулой

$$a(k) = s(k-1) \frac{s(k) - 1}{s(k-1) - 1}. \quad (104)$$

Аналогично, отношение $S(k)/A(k)$ равно рациональной функции $s(k)/(s(k) - 1)$.

Задача неопределенного гипергеометрического суммирования ставится так: для данной последовательности гипергеометрического типа $(A(k))_{k \geq 0}$ выяснить, будет ли последовательность $(S(k))_{k \geq 0}$, определяемая формулой

$$S(k) = A(0) + \dots + A(k), \quad (105)$$

последовательностью гипергеометрического типа. Как свидетельствует соотношение (104), эту задачу можно переформулировать так:

Для данной рациональной функции $a(k)$ выяснить, имеет ли уравнение (104) решение $s(k)$, которое является рациональной функцией, и найти его, если это возможно.

Запишем рациональную функцию $a(k) = A(k)/A(k-1)$ в виде

$$a(k) = \frac{p(k)}{p(k-1)} \frac{q(k)}{r(k)}, \quad (106)$$

где полиномы $p(k)$, $q(k)$, $r(k)$ таковы, что $q(k)$ не имеет непостоянного общего множителя ни с одним из полиномов $r(k)$, $r(k+1)$, $r(k+2)$, ... [если $q(k)$ и $r(k+j)$ имеют непостоянный общий множитель $u(k)$, сделаем замену

$$p(k) \leftarrow p(k)u(k)u(k-1) \dots u(k-j+1),$$

$$q(k) \leftarrow q(k)/u(k),$$

$$r(k) \leftarrow r(k)/u(k-j),$$

а затем повторим эту процедуру, если необходимо].

Введем теперь новую последовательность $(\varphi(k))_{k \geq 0}$, такую, что

$$S(k) = \frac{q(k+1)}{p(k)} \varphi(k) A(k). \quad (107)$$

Соотношение $A(k) = S(k) - S(k-1)$ приводит тогда к уравнению

$$q(k+1)\varphi(k) - r(k)\varphi(k-1) = p(k). \quad (108)$$

Если $(S(k))$ — последовательность гипергеометрического типа, $\varphi(k)$ будет рациональной функцией от k , поскольку $S(k)/A(k)$ — рациональная функция от k . Чудо вот в чем: *если рациональная функция $\varphi(k)$ является решением уравнения (108), то она будет полиномом, степень которого можно явно мажорировать целым числом J , вычисляемым исходя из $p(k)$, $q(k)$ и $r(k)$* . Если $J < 0$, то решения нет и последовательность $(S(k))$ не гипергеометрического типа; в противном случае положим

$$\varphi(k) = f_0 + f_1 k + \dots + f_J k^J \quad (109)$$

и уравнение (108) сведется к системе линейных уравнений на коэффициенты $f_0, \dots, f_J$.

3.5. Поставим задачу определенного гипергеометрического суммирования. Задана функция $F(n, k)$ гипергеометрического типа, удовлетворяющая двум соотношениям

$$\frac{F(n+1, k)}{F(n, k)} = A(n, k), \quad \frac{F(n, k+1)}{F(n, k)} = B(n, k), \quad (110)$$

где $A(n, k)$ и $B(n, k)$ — две рациональные функции. Хотелось бы найти полиномы $s_0(n), \dots, s_I(n)$ и рациональную функцию $R(n, k)$, удовлетворяющие соотношению

$$s(n, N)F(n, k) = G(n, k) - G(n, k-1), \quad (111)$$

где

$$s(n, N) = \sum_{i=0}^I s_i(n) N^i, \quad (112)$$

$$G(n, k) = R(n, k)F(n, k). \quad (113)$$

С учетом соотношения (110) после деления на $F(n, k)$ соотношение (111) становится тождеством между рациональными функциями

$$\sum_{i=0}^I s_i(n) \prod_{j=0}^{i-1} A(n+j, k) = R(n, k) - R(n, k-1)/B(n, k-1). \quad (114)$$

Вообще говоря, проверка этого соотношения, когда уже найдены явно $s_0(n), \dots, s_I(n)$ и $R(n, k)$, совсем элементарна; она влечет за собой (в предположении, например, что $F(n, k)$ равна нулю при достаточно больших $|k|$, когда n фиксировано), что для суммы $a(n) = \sum_k F(n, k)$ выполняется рекуррентное соотношение

$$\sum_{i=0}^I s_i(n) a(n+i) = 0. \quad (115)$$

Цайльбергер говорит, что $R(n, k)$ «удостоверяет» выполнение уравнения (115).

Для вычисления $s_0(n), \dots, s_I(n)$, $R(n, k)$ следует действовать, как в алгоритме Госпера, обращаясь с n как с параметром, т. е. заменяя основное поле $\mathbb{C}$ на поле рациональных функций $\mathbb{C}(n)$. Мы просто приведем основные формулы. Как и в методе Госпера, рассмотрим разложение на множители

$$\frac{F(n+i, k)}{F(n+i, k-1)} = \frac{p_i(n, k)}{p_i(n, k-1)} \frac{q(n, k)}{r(n, k)} \quad (116)$$

с полиномами $p_i(n, k)$, $q(n, k)$ и $r(n, k)$, такими, что $q(n, k)$ не имеет общего множителя ни с одним из полиномов $r(n, k+j)$ при $j = 0, 1, 2, \dots$ [определим сначала $p_0(n, k)$, $q(n, k)$ и $r(n, k)$, что всегда возможно, если числитель и знаменатель функции $B(n, k)$ разлагаются в произведение линейных множителей; далее используем то, что $F(n+i, k)/F(n, k)$ — рациональная функция в соответствии с (110)]. Потом определим рациональные функции $s_0(n), \dots, s_I(n)$ и $f_0(n), \dots, f_J(n)$, удовлетворяющие системе линейных уравнений (отождествляя коэффициенты при мономах от k)

$$\sum_{j=0}^J \{q(n, k+1)k^j - r(n, k)(k-1)^j\} \cdot f_j(n) = \sum_{i=0}^I p_i(n, k) \cdot s_i(n). \quad (117)$$

Наконец, полагаем

$$G(n, k) = q(n, k+1) \sum_{j=0}^J f_j(n) k^j \cdot \frac{\sum_{i=0}^I s_i(n) F(n+i, k)}{\sum_{i=0}^I s_i(n) p_i(n, k)}. \quad (118)$$

Поскольку $F(n+i, k)/F(n, k)$ — рациональная функция от n и k , то же самое верно для $G(n, k)/F(n, k)$.

Пример. $F(n, k) = 1/k!(n-k)!$.

Пусть $I = 1$. Алгоритм Госпера гарантирует, что $J = 0$ подходит. Тогда

$$\begin{cases} p_0(n, k) = n - k + 1, \\ p_1(n, k) = 1, \\ q(n, k) = n - k + 2, \\ r(n, k) = k. \end{cases} \quad (119)$$

Неизвестными являются s_0, s_1, f_0 , и система (117) записывается в виде

$$(n - 2k + 1)f_0 = (n - k + 1)s_0 + s_1. \quad (120)$$

Разделяя различные степени k , приходим к системе

$$\begin{cases} (n + 1)f_0 = (n + 1)s_0 + s_1, \\ 2f_0 = s_0. \end{cases} \quad (121)$$

Решением является

$$f_0 = 1, \quad s_0 = 2, \quad s_1 = -(n + 1),$$

откуда следует, что

$$G(n, k) = (n - k + 1) \frac{2F(n, k) - (n + 1)F(n + 1, k)}{2(n - k + 1) - (n + 1)}.$$

Поскольку

$$\frac{F(n + 1, k)}{F(n, k)} = \frac{1}{n - k + 1}, \quad (122)$$

закключаем с помощью вычислений с рациональными функциями, что

$$\frac{G(n, k)}{F(n, k)} = 1. \quad (123)$$

Уравнение (111) приводится в этом случае к виду

$$2F(n, k) - (n + 1)F(n + 1, k) = F(n, k) - F(n, k - 1). \quad (124)$$

Суммируя по k , видим, что $a(n) = \sum_k F(n, k)$ удовлетворяет рекуррентному уравнению

$$2a(n) - (n + 1)a(n + 1) = 0; \quad (125)$$

это вместе с начальным условием $a(0) = 1$ дает решение $a(n) = 2^n/n!$, откуда получаем формулу суммирования

$$\sum_{k=0}^n \frac{1}{k!(n-k)!} = \frac{2^n}{n!}. \quad (126)$$

Мы снова передоказали наш любимый пример $\sum_{k=0}^n \binom{n}{k} = 2^n$.

3.6. Цайльбергер и Вильф с успехом развивают похожие алгоритмы для изучения q -аналогов функций гипергеометрического типа. Совместно с Алмквистом Цайльбергер также смог получить аналогичные алгоритмы для нахождения дифференциальных уравнений, которым удовлетворяют интегралы вида $u(x) = \int v(x, y) dy$, где v — функция *гиперэкспоненциального типа*¹⁾. Наконец, опираясь на предварительные результаты Галлиго [С 7], Такаяма [С 14] смог развить теорию исключения в идеалах дифференциальных операторов, напоминающую технику базисов Грёбнера [С 5] полиномиальных идеалов. Это указывает иные возможности, помимо методов Цайльбергера.

4. ГИПЕРГЕОМЕТРИЧЕСКИЕ РЯДЫ И ФУНКЦИИ²⁾

4.1. Мы уже напоминали определение гипергеометрического ряда Гаусса $F = {}_2F_1$:

$$F \left(\begin{matrix} a & b \\ c \end{matrix} \middle| z \right) = \sum_{k \geq 0} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}; \quad (127)$$

он имеет смысл, когда c не является отрицательным целым числом, и ряд сходится при $|z| < 1$. Он удовлетворяет дифференциальному уравнению

$$z(1-z) \frac{d^2 F}{dz^2} + \{c - (a+b+1)z\} \frac{dF}{dz} - abF = 0. \quad (128)$$

В некоторых частных случаях получаются элементарные функции, например

$$F \left(\begin{matrix} a & b \\ b \end{matrix} \middle| z \right) = (1-z)^{-a}, \quad (129)$$

¹⁾ Это означает, что $D_x v/v$ и $D_y v/v$ — рациональные функции.

²⁾ См. [А 16] и [А 2] по поводу основных свойств гипергеометрических функций и [А 11] (гл. 5), где достаточно детально изучаются суммы, связанные с биномиальными коэффициентами.

$$F\left(\begin{matrix} a & a+1/2 \\ 1/2 \end{matrix} \middle| z\right) = \frac{1}{2}(1+z^{1/2})^{-2a} + \frac{1}{2}(1-z^{1/2})^{-2a}. \quad (130)$$

Функции, отвечающие различным параметрам a, b, c , не независимы; вот, например, соотношение Эйлера:

$$F\left(\begin{matrix} c-a & c-b \\ c \end{matrix} \middle| z\right) = (1-z)^{a+b-c} F\left(\begin{matrix} a & b \\ c \end{matrix} \middle| z\right). \quad (131)$$

В некоторых случаях ряд (127) можно суммировать точно:

$$F\left(\begin{matrix} a & b \\ c \end{matrix} \middle| 1\right) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \quad (\text{Гаусс}), \quad (132)$$

$$F\left(\begin{matrix} a & b \\ 1+a-b \end{matrix} \middle| -1\right) = 2^a \frac{\Gamma(1+a-b)\Gamma(\frac{1}{2})}{\Gamma(1-b+\frac{a}{2})\Gamma(\frac{a+1}{2})} \quad (\text{Куммер}); \quad (133)$$

$$F\left(\begin{matrix} a & 1-a \\ b \end{matrix} \middle| \frac{1}{2}\right) = 2^{1-b} \frac{\Gamma(\frac{1}{2})\Gamma(b)}{\Gamma(\frac{a+b}{2})\Gamma(\frac{b-a+1}{2})}; \quad (134)$$

$$F\left(\begin{matrix} 2a & 2b \\ a+b+\frac{1}{2} \end{matrix} \middle| \frac{1}{2}\right) = \frac{\Gamma(a+b+\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(a+\frac{1}{2})\Gamma(b+\frac{1}{2})}. \quad (135)$$

Имеется два классических доказательства формулы Гаусса (132). Прежде всего рассмотрим интегральное представление Эйлера

$$F\left(\begin{matrix} a & b \\ c \end{matrix} \middle| z\right) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a} dt, \quad (136)$$

которое осуществляет аналитическое продолжение гипергеометрической функции в плоскости, разрезанной вдоль вещественной оси от 1 до $+\infty$. При $z=1$ этот интеграл сводится к бета-функции Эйлера

$$\int_0^1 t^{b-1}(1-t)^{c-b-a-1} dt,$$

а значит, к $\Gamma(b)\Gamma(c-b-a)/\Gamma(c-a)$. Этот метод позволяет установить соотношения (133)–(135).

Второе доказательство опирается на рекуррентное уравнение

$$\begin{aligned} (c-a)(c-b)zF\left(\begin{matrix} a & b \\ c+1 \end{matrix} \middle| z\right) &= c[(2c-a-b-1)z \\ &- c+1]F\left(\begin{matrix} a & b \\ c \end{matrix} \middle| z\right) + c(c-1)(1-z)F\left(\begin{matrix} a & b \\ c-1 \end{matrix} \middle| z\right) \end{aligned} \quad (137)$$

(непосредственная проверка сравнением коэффициентов при z^k). При переходе к пределу при z , стремящемся к 1, последний член исчезает, откуда

$$(c-a)(c-b)F\left(\begin{matrix} a & b \\ c+1 \end{matrix} \middle| 1\right) = c(c-a-b)F\left(\begin{matrix} a & b \\ c \end{matrix} \middle| 1\right). \quad (138)$$

Индукцией по целому $m \geq 0$ выводим

$$F\left(\begin{matrix} a & b \\ c \end{matrix} \middle| 1\right) = \frac{(c-a)_m(c-b)_m}{(c)_m(c-a-b)_m} F\left(\begin{matrix} a & b \\ c+m \end{matrix} \middle| 1\right). \quad (139)$$

Затем перейдем к пределу по m ; нетрудно заметить, что $F\left(\begin{matrix} a & b \\ c+m \end{matrix} \middle| 1\right)$ стремится к 1 при m , стремящемся к бесконечности, а разложение гамма-функции в бесконечное произведение дает соотношение

$$\lim_{m \rightarrow \infty} \frac{(u_1)_m \dots (u_p)_m}{(v_1)_m \dots (v_p)_m} = \frac{\Gamma(v_1) \dots \Gamma(v_p)}{\Gamma(u_1) \dots \Gamma(u_p)}, \quad (140)$$

когда $u_1 + \dots + u_p$ равно $v_1 + \dots + v_p$.

4.2. Формула Гаусса особенно интересна, когда b — целое отрицательное число. С учетом рекуррентного соотношения

$$\Gamma(s+n) = \Gamma(s)(s)_n \quad (141)$$

формула Гаусса после специализации дает

$$F\left(\begin{matrix} a & -n \\ c \end{matrix} \middle| 1\right) = \frac{(c-a)_n}{(c)_n}. \quad (142)$$

Но гипергеометрический ряд сводится к конечной сумме, и после некоторых манипуляций мы получаем соотношение

$$\sum_{k=0}^n (-1)^k \binom{n}{k} \frac{(a)_k}{(c)_k} = \frac{(c-a)_n}{(c)_n}. \quad (143)$$

Остается сделать подстановку $a \leftarrow -u$, $c \leftarrow v-n-1$, чтобы получить формулу Чу–Вандермонда

$$\binom{u+v}{n} = \sum_{k=0}^n \binom{u}{k} \binom{v}{n-k}. \quad (144)$$

Вот другие специализации формулы (142):

$$\sum_{k=0}^n \binom{r+k}{k} = \binom{r+n+1}{n} \quad (a \leftarrow 1, c \leftarrow -n-r), \quad (145)$$

$$\sum_{k=0}^n (-1)^k \binom{r}{k} = (-1)^n \binom{r-1}{n} \quad (a \leftarrow 1, c \leftarrow -n+r+1). \quad (146)$$

В свою очередь, формула Чу—Вандермонда имеет многочисленные специализации, например¹⁾,

$$\sum_k \binom{u}{m+k} \binom{v}{n-k} = \binom{u+v}{m+n}. \quad (147)$$

Аналогичным образом формула Куммера (133) приходит после специализации к виду

$$F \left(\begin{matrix} 1-r-2n \\ r \end{matrix} \middle| -1 \right) = (-1)^n \frac{(2n)!}{n!} \frac{(r-1)!}{(r+n-1)!}, \quad (148)$$

и после нескольких простых манипуляций мы получаем обобщение второй формулы из (2)

$$\sum_{k=0}^{2n} (-1)^k \binom{r}{k} \binom{r}{2n-k} = (-1)^n \binom{r}{n}. \quad (149)$$

Не будем продолжать эту игру: практически всякая формула, представляющая сумму произведений двух биномиальных коэффициентов, получается специализацией из формул (132)–(135).

4.3. Чтобы двигаться дальше, нужно ввести общие гипергеометрические ряды (см. п. 2.1)

$${}_pF_q \left(\begin{matrix} a_1 \dots a_p \\ b_1 \dots b_q \end{matrix} \middle| z \right) = \sum_{k=0}^{\infty} \frac{(a_1)_k \dots (a_p)_k}{(b_1)_k \dots (b_q)_k} \frac{z^k}{k!}. \quad (150)$$

Дифференциальное уравнение (128) также обобщается (полагаем $D = \frac{d}{dz}$ и $\vartheta = z \frac{d}{dz}$):

$$D(\vartheta + b_1 - 1) \dots (\vartheta + b_q - 1)F = (\vartheta + a_1) \dots (\vartheta + a_p)F. \quad (151)$$

¹⁾ Более обширный список см. в [А 11, р. 169].

Это непосредственно выводится из трех формул дифференцирования

$$DF \left(\begin{matrix} a_1 \dots a_p \\ b_1 \dots b_q \end{matrix} \middle| z \right) = \frac{a_1 \dots a_p}{b_1 \dots b_q} F \left(\begin{matrix} a_1 + 1 \dots a_p + 1 \\ b_1 + 1 \dots b_q + 1 \end{matrix} \middle| z \right), \quad (152)$$

$$(\vartheta + a_1) F \left(\begin{matrix} a_1 \dots a_p \\ b_1 \dots b_q \end{matrix} \middle| z \right) = a_1 F \left(\begin{matrix} a_1 + 1 & a_2 \dots a_p \\ b_1 & b_2 \dots b_q \end{matrix} \middle| z \right), \quad (153)$$

$$(\vartheta + b_1 - 1) F \left(\begin{matrix} a_1 \dots a_p \\ b_1 \dots b_q \end{matrix} \middle| z \right) = (b_1 - 1) F \left(\begin{matrix} a_1 & a_2 \dots a_p \\ b_1 - 1 & b_2 \dots b_q \end{matrix} \middle| z \right). \quad (154)$$

Сравним коэффициенты при z^k в двух членах формулы Эйлера (131). Получается соотношение, принадлежащее Пфаффу и Заалшютцу, которое имеет вид

$${}_3F_2 \left(\begin{matrix} a & b & -n \\ c & a+b-c-n+1 \end{matrix} \middle| 1 \right) = \frac{(c-a)_n (c-b)_n}{(c)_n (c-a-b)_n}. \quad (155)$$

Нетрудно восстановить теорему Гаусса о ${}_2F_1 \left(\begin{matrix} a & b \\ c \end{matrix} \middle| 1 \right)$, переходя в формуле Пфаффа—Заалшютца к пределу при $n \rightarrow \infty$.

Мы выпишем наиболее замечательные соотношения, которые получаются при суммировании общих гипергеометрических рядов:

$$\begin{aligned} & {}_3F_2 \left(\begin{matrix} a & b & c \\ a-b+1 & a-c+1 \end{matrix} \middle| 1 \right) \\ &= \frac{\Gamma(\frac{a}{2}+1)\Gamma(a-b+1)\Gamma(a-c+1)\Gamma(\frac{a}{2}-b-c+1)}{\Gamma(a+1)\Gamma(\frac{a}{2}-b+1)\Gamma(\frac{a}{2}-c+1)\Gamma(a-b-c+1)} \quad (\text{Диксон}), \end{aligned}$$

$$\begin{aligned} & {}_3F_2 \left(\begin{matrix} a & b & c \\ \frac{1}{2}(a+b+1) & 2c \end{matrix} \middle| 1 \right) \\ &= \frac{\Gamma(\frac{1}{2})\Gamma(c+\frac{1}{2})\Gamma(\frac{a+b+1}{2})\Gamma(\frac{1}{2}-\frac{a}{2}-\frac{b}{2}+c)}{\Gamma(\frac{a+1}{2})\Gamma(\frac{b+1}{2})\Gamma(c-\frac{a}{2}+\frac{1}{2})\Gamma(c-\frac{b}{2}+\frac{1}{2})} \quad (\text{Ватсон}), \end{aligned}$$

$${}_3F_2 \left(\begin{matrix} a & b & c \\ e & f \end{matrix} \middle| 1 \right) = \frac{\pi \Gamma(e) \Gamma(f)}{2^{2c-1} \Gamma(\frac{a+e}{2}) \Gamma(\frac{a+f}{2}) \Gamma(\frac{b+e}{2}) \Gamma(\frac{b+f}{2})} \quad (\text{Випплъ})$$

(при условии, что $a+b=1$ и $e+f=2c+1$).

Еще более впечатляющее тождество принадлежит Дуголлу:

$${}_7F_6 \left(a \frac{a}{2} + 1, b, c, d, e, -m \middle| \frac{a}{2}, a-b+1, a-c+1, a-d+1, a-e+1, a+m+1 \right) = \frac{(a+1)_m (a-b-c+1)_m (a-b-d+1)_m (a-c-d+1)_m}{(a-b+1)_m (a-c+1)_m (a-d+1)_m (a-b-c-d+1)_m},$$

в предположении, что m — положительное целое число и что $2a+1 = b+c+d+e-m$. Специалисты могут при помощи специализации вывести из этих тождеств огромное количество соотношений между биномиальными коэффициентами.

4.4. Все эти соотношения имеют q -аналоги¹⁾. Мы уже ввели бесконечное произведение $(a; q)_\infty = \prod_{m \geq 0} (1 - aq^m)$, которое приходится двоюродным братом гамма-функции, и частичное произведение

$$(a; q)_n = \frac{(a; q)_\infty}{(aq^n; q)_\infty} = \prod_{m=0}^{n-1} (1 - aq^m). \quad (156)$$

q -аналогом факториала служит

$$[n]_q! = \frac{(q; q)_n}{(1-q)^n}; \quad (157)$$

это полином от q степени $n(n-1)/2$, и его значение при $q = 1$ равно $n!$.

q -аналог гипергеометрической функции Гаусса задается формулой

$${}_2\Phi_1 \left(\begin{matrix} a & b \\ c \end{matrix}; q; z \right) = \sum_{k \geq 0} \frac{(a; q)_k (b; q)_k}{(c; q)_k (q; q)_k} z^k \quad (158)$$

с очевидным обобщением

$${}_r\Phi_s \left(\begin{matrix} a_1 \dots a_r \\ b_1 \dots b_s \end{matrix}; q; z \right) = \sum_{k \geq 0} \frac{(a_1; q)_k \dots (a_r; q)_k}{(b_1; q)_k \dots (b_s; q)_k} \frac{z^k}{(q; q)_k}. \quad (159)$$

По аналогии с тем, что биномиальный ряд является частным случаем гипергеометрической функции

$$(1-z)^{-a} = {}_1F_0 \left(\begin{matrix} a \\ - \end{matrix} \middle| z \right), \quad (160)$$

¹⁾ См. книги Экстона [А 17] и Гаспера—Рахмана [А 19].

а экспонента выражается формулой

$$e^z = {}_0F_0 \left(\begin{matrix} - \\ - \end{matrix} \middle| z \right), \quad (161)$$

мы введем q -биномиальный ряд

$${}_1\Phi_0 \left(\begin{matrix} a \\ - \end{matrix} ; q; z \right) = \sum_{k \geq 0} \frac{(a; q)_k}{(q; q)_k} z^k = \frac{(az; q)_\infty}{(z; q)_\infty} \quad (162)$$

и q -экспоненциальный ряд

$${}_0\Phi_0 \left(\begin{matrix} - \\ - \end{matrix} ; q; z \right) = \sum_{k \geq 0} \frac{z^k}{(q; q)_k} = \frac{1}{(z; q)_\infty} \quad (163)$$

(чтобы вернуться к классическому случаю, нужно заменить z на $(1 - q)x$ и устремить q к 1).

Если в ряде ${}_r\Phi_s$ заменить z на 1 и если один из параметров a_i имеет вид q^{-m} (с целым $m \geq 0$), то мы получим сумму q -биномиальных коэффициентов (см. п. 2.5). Формулы гипергеометрического суммирования (Гаусс, Куммер, ...) имеют q -аналоги, например,

$${}_2\Phi_1 \left(\begin{matrix} a & b \\ c \end{matrix} ; q; c/ab \right) = {}_2\Pi_2 \left(\begin{matrix} c/a & c/b \\ c & c/ab \end{matrix} \middle| q \right) \quad (\text{Гаусс}),$$

$$\begin{aligned} & {}_2\Phi_1 \left(\begin{matrix} a & b \\ aq/b \end{matrix} ; q; -q/b \right) \\ &= {}_1\Pi_2 \left(\begin{matrix} -q & -q \\ -q/b & aq/b \end{matrix} \middle| q \right) {}_2\Pi_0 \left(\begin{matrix} aq & aq^2/b^2 \\ - & - \end{matrix} \middle| q^2 \right) \quad (\text{Куммер}), \end{aligned}$$

$${}_3\Phi_2 \left(\begin{matrix} a & b & q^{-n} \\ c & abc^{-1} \end{matrix} ; q; q \right) = \frac{(c/a; q)_n (c/b; q)_n}{(c; q)_n (c/ab; q)_n} \quad (\text{Пфаффе-Заалшютц}),$$

$$\begin{aligned} & {}_3\Phi_2 \left(\begin{matrix} a^2 & b & c \\ a^2q/b & a^2q/c \end{matrix} ; q; q^2a/bc \right) \\ &= {}_4\Pi_4 \left(\begin{matrix} b/a^2 & c/a^2 & q/a & bc/a \\ b/a & c/a & q/a^2 & bc/a^2 \end{matrix} \middle| q \right) \quad (\text{Диксон}). \end{aligned}$$

Мы принимаем такое соглашение:

$${}_r\Pi_s \left(\begin{matrix} a_1 \dots a_r \\ b_1 \dots b_s \end{matrix} \middle| q \right) = \frac{(a_1; q)_\infty \dots (a_r; q)_\infty}{(b_1; q)_\infty \dots (b_s; q)_\infty}. \quad (164)$$

Эти формулы можно специализировать многими способами; приведем два знаменитых примера:

$$\sum_{n=-\infty}^{+\infty} q^{n^2} z^n = \prod_{m=1}^{\infty} \frac{1}{(1 + q^{2m-1}z)(1 + q^{2m-1}z^{-1})(1 - q^{2m})} \quad (\text{Якоби}),$$

$$\sum_{k=0}^{\infty} \frac{q^{k^2}}{(1-q)(1-q^2) \dots (1-q^k)} = \prod_{n=0}^{\infty} \frac{1}{(1-q^{5n+1})(1-q^{5n+4})} \quad (\text{Роджерс—Рамануджан}).$$

4.5. В чем состоит вклад Цайльбергера? Я приведу типичный пример, а большое количество подобных приложений можно найти в работах [В 16] и [В 18], включая формулу Дуголла [В 14].

Прежде всего заметим, что формула Диксона (3) является частным случаем при $a = b = c = n$ другой формулы Диксона (1903), а именно

$$\sum_k (-1)^k \binom{a+b}{a+k} \binom{b+c}{b+k} \binom{c+a}{c+k} = \frac{(a+b+c)!}{a!b!c!} \quad (165)$$

(a, b, c — положительные целые числа). Это не что иное, как специализация гипергеометрической формулы Диксона

$${}_3F_2 \left(\begin{matrix} 1-a-2n & 1-b-2n & -2n \\ a & b \end{matrix} \middle| 1 \right) = (-1)^n \frac{(2n)!}{n!} \frac{(a+b+2n-2)_n}{(a)_n(b)_n}. \quad (166)$$

Вот доказательство формулы (165), полученное Цайльбергером совместно со своим верным электронным помощником Shalosh Ekhad. Положим

$$F(n, k) = (-1)^k \frac{(n+b)!(n+c)!(b+c)!}{(n+k)!(n-k)!(b+k)!(b-k)!(c+k)!(c-k)!}, \quad (167)$$

$$R(n) = \frac{(n+b+c)!}{n!b!c!}. \quad (168)$$

Формула, которую нужно доказать, — это случай $n = a$ формулы

$$\sum_k F(n, k) = R(n). \quad (169)$$

Случай $n = 0$ очевиден, и мы имеем рекуррентное уравнение

$$(n+1)R(n+1) - (n+b+c+1)R(n) = 0. \quad (170)$$

Нам нужно найти для $F(n, k)$ «компаньона» $G(n, k)$ (вместе они образуют то, что соавторы Вильф и Цайльбергер скромно называют ВЦ-парой), удовлетворяющего соотношению

$$(n+1)F(n+1, k) - (n+b+c+1)F(n, k) = G(n, k) - G(n, k-1). \quad (171)$$

Компьютер Shalosh Ekhad, используя алгоритмы, описанные в п. 3.5, ищет решение в виде

$$G(n, k) = R(n, k)F(n, k), \quad (172)$$

где $R(n, k)$ — рациональная функция. Кроме того, $F(n, k)$ удовлетворяет двум рекуррентным уравнениям

$$F(n+1, k) = A(n, k)F(n, k), \quad F(n, k+1) = B(n, k)F(n, k), \quad (173)$$

и после деления на $F(n, k)$ уравнение (171) записывается так:

$$(n+1)A(n, k) - (n+b+c+1) = R(n, k) - \frac{R(n, k-1)}{B(n, k-1)}. \quad (174)$$

Функции $A(n, k)$ и $B(n, k)$, разумеется, известны:

$$A(n, k) = \frac{(n+b+1)(n+c+1)}{(n+k+1)(n-k+1)}, \quad (175)$$

$$B(n, k) = -\frac{(n-k)(b-k)(c-k)}{(n+k+1)(b+k+1)(c+k+1)}, \quad (176)$$

и компьютер нашел подходящее решение уравнения (174):

$$R(n, k) = \frac{(b-k)(c-k)}{2(n+k+1)}. \quad (177)$$

Когда функция $R(n, k)$ найдена, проверка уравнения (174) тривиальна.

5. ПРЕПЯТСТВИЯ И НОВЫЕ ПУТИ

Теперь мы располагаем точным словарем, сводящим доказательство гипергеометрических тождеств (и, в частности, соотношений между биномиальными коэффициентами) к тождествам между рациональными функциями. Но остается еще привести в систему массу полученных результатов и угадать, какие структуры лежат в основе. Как видно из некоторых примеров, ключ, вероятно, следует искать в когомологиях де Рама дифференциальных форм с коэффициентами в поле рациональных функций, их аналогов на решетке $\mathbb{Z}^s$ и в q -технике. Голономные модули тесно связаны с векторными расслоениями с интегрируемой связностью, и нужно изучать что-то вроде дифференциальных групп Галуа.

В рекуррентных соотношениях с одной переменной k приходится порознь исследовать *асимптотическое поведение* при k , стремящемся к $+\infty$, и при k , стремящемся к $-\infty$. Например, уравнение

$$f(k+1) = (k+1)f(k)$$

имеет два решения

$$f_1(k) = k! \quad \text{при } k \geq 0,$$

$$f_2(k) = \begin{cases} (-1)^{k+1}/(-k-1)! & \text{при } k < 0, \\ 0 & \text{при } k \geq 0. \end{cases}$$

Аомото [D 1, D 2, D 3] начал изучать подобное асимптотическое поведение также в случае уравнений вида $F(qz)/F(z) = a(q; z)$ и, главным образом, в многомерном случае.

Формула Диксона (165) — это частный случай при $n = 3$ формулы Дайсона, утверждающей, что если разложить в ряд Лорана произведение $\prod_{i \neq j} (1 - x_i/x_j)^{a_i}$ (где i, j пробегает целые числа от 1 до n), то постоянный член будет равен

$$\frac{(a_1 + \dots + a_n)!}{a_1! \dots a_n!}.$$

Это источник многочисленных тождеств такого же типа, связанных с простыми алгебрами Ли: гипотезы Макдональда и Морриса—Макдональда. Почти все случаи теперь исследованы, но некоторые (такие, как случай G_2 и F_4) поддаются только компьютерам, в которых используются методы Цайльбергера. Здесь мы сталкиваемся

с тонкой проблемой: если $x, x^2, x^3, x_1 + x_2, x_1 + x_2 + x_3$ — полиномы, то x^n или $x_1 + \dots + x_n$ с алгоритмической точки зрения (при переменном n) полиномами не являются.

Мы спотыкаемся о другое препятствие, когда пытаемся доказать тождества, не содержащие никакого параметра; например, для доказательства с помощью компьютера формулы Роджерса—Рамануджана необходимо сначала получить промежуточный результат

$$\sum_k \frac{q^{k^2}}{(q; q)_k (q; q)_{n-k}} = \sum_k \frac{(-1)^k q^{(5k^2-k)/2}}{(q; q)_{n-k} (q; q)_{n+k}}, \quad (178)$$

который содержит свободный параметр n , а затем устремить n к бесконечности.

С аналогичной трудностью мы сталкиваемся, когда пытаемся доказать чисто числовые соотношения. Как доказать, что в следующих формулах речь идет об одном и том же числе π :

$$\frac{\pi}{4} = \sum_{n=1}^{\infty} \frac{(-1)^n}{2n+1}, \quad \frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2}, \quad \int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi} ?$$

В сущности, что такое вещественное число с точки зрения *алгебраического анализа* в смысле Микио Сато и его соперников? Здесь, возможно, существу дела отвечает замечание Метрополиса и Роты: пусть $(a_n)_{n \geq 0}$ — полиномиально растущая последовательность целых чисел

$$|a_n| \leq Cn^k \quad (n \geq 1) \quad (179)$$

с двумя константами $C > 0$ и $k \geq 1$. Пусть $b \geq 2$ — целое число. Тогда ряд $\sum_{n=0}^{\infty} a_n/b^n$ сходится (тривиально); его сумма равна 0 тогда и только тогда, когда существует некоторая *полиномиально растущая* последовательность (c_n) , такая, что

$$a_n = c_n - bc_{n-1}. \quad (180)$$

Это новый пример телескопической суммы.

Наконец, хорошо известна аналогия между гауссовыми суммами и гамма-функцией. Тождество для гауссовых сумм, предложенное Анной Хелверсен-Пасотто [Е 8], на самом деле является аналогом тождества Гаусса для ${}_2F_1 \left(\begin{smallmatrix} a & b \\ c \end{smallmatrix} \middle| 1 \right)$. Как обобщить это на соотношения Диксона, Заалшютца, ...? Каков соответствующий алгоритмический метод? Существует ли связь с характеристическими пучками Люстига или с квантовыми группами? Есть о чем помечтать!

БИБЛИОГРАФИЯ С КОММЕНТАРИЯМИ

А. Основные работы

Классический репертуар о рядах, интегралах, специальных функциях:

- [1] Abramowitz M., Stegun I. *Handbook of Mathematical Functions*, Dover, New York, 1965. См. также: Справочник по специальным функциям с формулами, графиками и математическими таблицами (ред. Абрамовиц М., Стиган И.). — М.: Наука, 1979.
- [2] Erdelyi A. (editor), *Higher Transcendental Functions*, 3 volumes, McGraw-Hill, New York, 1953. См. также: Бейтмен Г., Эрдейн А. Высшие трансцендентные функции. Вып. 1-3. — М.: Наука, 1965, 1966, 1967.
- [3] Градштейн Н. С., Рыжик И. М. *Таблицы интегралов, сумм, рядов и произведений*. — М.: Наука, 1971.

Несколько классических работ, в основном о гипергеометрических функциях:

- [4] Appell P., Kampé de Fériet J. *Fonctions hypergéométriques et hypersphériques*, Gauthier-Villars, Paris, 1926.
- [5] Rainville J. E. *Special Functions*, MacMillan, New York, 1960.
- [6] Slater L. *Confluent Hypergeometric Functions*, Cambridge University Press, Cambridge, 1960.
- [7] Whittaker E., Watson G. *Modern Analysis*, Cambridge University Press, Cambridge, 1946. [Имеется перевод первого изд.: Уиттекер Э. Т., Ватсон Г. Н. Курс современного анализа, ч. I, II — М.: ГИТТЛ, 1933, 1934.]

Несколько хороших ссылок по комбинаторике:

- [8] Aigner M. *Combinatorial Theory*, Springer-Verlag, Berlin, 1979 (особенно гл. 3). [Имеется перевод: Айгнер М. Комбинаторная теория. — М.: Мир, 1982].
- [9] Andrews G. *The Theory of Partitions*, Addison-Wesley, Reading, 1976.
- [10] Comtet L. *Advanced Combinatorics*, Reidel, Dordrecht, 1974.
- [11] Graham R., Knuth D., Patashnik O. *Concrete Mathematics*, Addison-Wesley, Reading, 1989. [Имеется перевод: Грехэм Р., Кнут Д., Паташник О. Конкретная математика. Основание информатики. — М.: Мир, 1998.]
- [12] Knuth D. *The Art of Computer Programming*, 3 vol., Addison-Wesley, Reading, 1968-1973. [Имеется перевод: Кнут Д. Искусство программирования для ЭВМ, т. 1-3. — М.: Мир, 1976, 1977, 1978.]
- [13] Macmahon P. *Combinatory Analysis*, 2 vol., Chelsea, New York, 1960 (переиздание).
- [14] Riordan J. *An Introduction to Combinatorial Analysis*, John Wiley, New York, 1958. [Имеется перевод: Риордан Дж. Введение в комбинаторный анализ. — М.: ИЛ, 1963.]

Наконец, несколько обзоров по q -гипергеометрическим функциям (называемым также базисными):

- [15] Andrews G. q -Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics and Computer Algebra, CBMS Regional Conference Lecture Series, 66, Amer. Math. Soc., Providence, 1986.
- [16] Bailey W. Generalized Hypergeometric Series, Cambridge University Press, Cambridge, 1935 (переиздание: Stechert-Hafner, New York, 1964).
- [17] Exton H. q -Hypergeometric Functions and Applications, Ellis Horwood/John Wiley, New York, 1983.
- [18] Fine N. Basic Hypergeometric Series and Applications, Math. Surv., vol. 27, Amer. Math. Soc., Providence, 1988.
- [19] Gasper G., Rahman M. Basic Hypergeometric Series, Cambridge University Press, Cambridge, 1990. [Имеется перевод: Гаспер Дж., Рахман М. Базисные гипергеометрические ряды. — М.: Мир, 1993.]
- [20] Heine E. Handbuch der Kugelfunktionen. Theorie und Anwendung, 2 volumes, Springer-Verlag, 1898 (= Physica Verlag, Würzburg, 1961).
- [21] Slater L. Generalized Hypergeometric Functions, Cambridge University Press, Cambridge, 1966.

В. Статьи Цайльбергера и соавторов

- [1] Zeilberger D. Sister Celine's technique and its generalizations, J. Math. Anal. Appl., 85 (1982), 114–145.
- [2] Zeilberger D. A holonomic systems approach to special functions identities, J. Comput. Appl. Math., 32 (1990), 321–368.
- [3] Zeilberger D. A fast algorithm for proving terminating hypergeometric identities, Discrete Math., 80 (1990), 207–211.
- [4] Zeilberger D. The method of creative telescoping, J. Symbolic Computation, 11 (1991), 195–204.
- [5] Zeilberger D. Closed form (pun intended!), in: A tribute to Emil Grosswald: Number Theory and Related Analysis, Contemp. Math. vol. 143, Amer. Math. Soc., Providence, RI, 1993, 579–607.
- [6] Zeilberger D. Three recitations on holonomic systems and hypergeometric series, Proceedings of the Séminaire Lotharingien de combinatoire, 24, IRMA, Strasbourg; J. Symbolic Comput., 20 (1995), No. 5–6, 699–724.
- [7] Zeilberger D. Plain (Lagrange interpolation) proofs of fancy (representation theory) formulas, готовится к печати.
- [8] Ekhad S. B. Short proofs of two hypergeometric summation formulas of Karlsson, Proc. Amer. Math. Soc., 107 (1989), 1143–1144.
- [9] Ekhad S. B. A very short proof of Dixon's theorem, J. Combin. Theory, Series A, 54 (1990), 141–142.
- [10] Ekhad S. B. A one-line proof of the Habsieger-Zeilberger G_2 constant term identity, J. Comput. Appl. Math., 34 (1991), 133–134.

- [11] Ekhad S.B. A short proof of a «strange» combinatorial identity conjectured by Gosper, *Discrete Math.*, **90** (1991), No. 3, 319–320.
- [12] Ekhad S.B. A short, elementary and easy, WZ proof of the Askey-Gasper inequality that was used by de Branges in his proof of the Bieberbach conjecture, *Theoret. Comput. Sci.*, **117** (1993), No. 1–2, 199–202.
- [13] Ekhad S.B., Tre S. A purely verification proof of the first Rogers-Ramanujan identity, *J. Combin. Theory Ser. A*, **54** (1990), 309–311.
- [14] Ekhad S.B., Zeilberger D. A 21st century proof of Dougail's hypergeometric sum identity, *J. Math. Anal. Appl.*, **147** (1990), 610–611.
- [15] Wilf H.S., Zeilberger D. Towards computerized proofs of identities, *Bull. Amer. Math. Soc.*, **23** (1990), 77–83.
- [16] Wilf H.S., Zeilberger D. Rational functions certify combinatorial identities, *J. Amer. Math. Soc.*, **3** (1990), 147–158.
- [17] Wilf H.S., Zeilberger D. A general theory of multi-variate hypergeometric identities, *готовится к печати*.
- [18] Wilf H.S. 54 computer-generated proofs of binomial coefficient identities, *готовится к печати*.

Наконец, вот три статьи с детальными программами для MAPLE, реализующие теоретические алгоритмы:

- [19] Almkvist G., Zeilberger D. The method of differentiating under the integral sign, *Journ. Symb. Computation*, **10** (1990), 571–591.
- [20] Almkvist G., Zeilberger D. A MAPLE program that finds, and proves, recurrences and differential equations satisfied by hyperexponential definite integrals, *SIGSAM Bulletin*, **25** (1991).
- [21] Zeilberger D. A MAPLE program for proving hypergeometric series, *SIGSAM Bulletin*, **25** (1991).

С. Алгебра дифференциальных операторов и теория исключения

- [1] Бернштейн И. Н. Модули над кольцом дифференциальных операторов. Изучение фундаментальных решений уравнений с постоянными коэффициентами. — *Функц. анализ и прил.*, т. 5, 1971, вып. 2, с. 1–16.
- [2] Бернштейн И. Н. Аналитическое продолжение обобщенных функций по параметру. — *Функц. анализ и прил.*, т. 6, 1972, вып. 4, с. 26–40.
- [3] Björk J.-E. Rings of Differential Operators, North Holland, Amsterdam, 1979.
- [4] Borel A. et al. Algebraic D-modules, *Perspectives in Math.*, vol. 2, Academic Press, Boston, 1987.
- [5] Buchberger B. An algorithmic method in polynomial ideal theory, N. K. Bose ed., *Recent Trends in Multidimensional Systems Theory*, D. Reidel Publishing Corp., 1985.
- [6] Ehlers F. The Weyl Algebra, гл. V из [4], 173–205.

- [7] Galligo A. Some algorithmic questions on ideals of differential operators, *Lect. Note in Comp. Sci.*, **204** (1985), 413–421.
- [8] Gosper R. Decision procedure for indefinite hypergeometric summation, *Proc. Nat. Acad. Sci. USA*, **75** (1978), 40–42.
- [9] Kashiwara M. B -functions and holonomic systems, *Invent. Math.*, **38** (1976), 33–35.
- [10] Kashiwara M. Vanishing cycles sheaves and holonomic systems of differential equations, *Lect. Notes in Math.*, vol. 1016, Springer-Verlag, 1983, 134–142.
- [11] Malgrange B. L'involutivité des caractéristiques des systèmes différentiels et microdifférentiels, *Sém. Bourbaki*, 1977–1978, exp. 522. *Lect. Notes in Math.*, vol. 710, Springer-Verlag, Berlin, 1979, 277–289.
- [12] Risch R. The solution of the problem in integrating in finite terms, *Bull. Amer. Math. Soc.*, **76** (1970), 605–608.
- [13] Takayama N. Gröbner basis and the problem of contiguous relations, *Japan J. Appl. Math.*, **6** (1989), 147–160.
- [14] Takayama N. An algorithm for constructing the integral of a module — an infinite dimensional analog of Gröbner basis, *Proc. of ISSAC'90*, A.C.M. Press.
- [15] Takayama N. An approach to the zero recognition problem by Buchberger algorithm, *J. Symbolic. Comput.*, **14** (1992), No. 2–3, 265–282.

D. Голономные q -разностные системы

- [1] Aomoto K. A note on holonomic q -difference systems, in: *Algebraic Analysis (in honor of M. Sato)*, M. Kashiwara and T. Kawai (eds.), Academic Press, 1988, 25–28.
- [2] Aomoto K. q -analogue of de Rham cohomology associated with Jackson integrals, *Proc. Japan Acad.*, **66** (1990), 161–164.
- [3] Aomoto K. Finiteness of a cohomology associated with certain Jackson integrals, *Tohoku J. Math.*, **43** (1991), 75–101.
- [4] Sabbah C. Systèmes holonomes d'équations aux q -différences, *Pré-publication Centre de Math. Ecole Polytechnique, Palaiseau*, 1991.

E. Различные аспекты

О формуле Диксона:

- [1] Cartier P., Foata D. Problèmes combinatoires de commutation et réarrangements, *Lect. Notes in Math.*, vol. 85, Springer-Verlag, 1969.

Об иррациональности $\zeta(3)$:

- [2] Apéry R. Irrationalité de $\zeta(2)$ et $\zeta(3)$, *Astérisque*, **61** (1979), 11–13.
- [3] van des Poorten A. A proof that Euler missed... Apéry's proof of the irrationality of $\zeta(3)$, *Math. Intelligencer*, **1** (1979), 195–203.

О гипотезах Макдональда:

- [4] MacDonald I. G. Affine root systems and Dedekind η -function, *Invent. Math.*, **15** (1972), 91–143.
- [5] MacDonald I. G. Some conjectures for root systems, *SIAM Journ. Math. Anal.*, **13** (1982), 988–1007.
- [6] Zeilberger D. Unified approach to Macdonald's root system conjectures, *SIAM Journ. Math. Anal.*, **19** (1988), 987–1013.
- [7] Garvan F., Bonnet G.. Macdonald's constant term conjectures for exceptional root systems, *Bull. Amer. Math. Soc.*, **24** (1991), 343–347.

О суммах Гаусса:

- [8] Helversen-Pasotto A. L'identité de Barnes pour les corps finis, *C.R. Acad. Sci. Paris, Série A*, **286** (1978), 297–300.
- [9] Greene J., Stanton D. A character sum evaluation and Gaussian hypergeometric series, *Journ. Number Theory*, **23** (1986), 136–148.

PIERRE CARTIER

École Normale Supérieure
Département de Mathématiques
et Applications
45, rue d'Ulm
F-75230 Paris Cedex 05

ДИНАМИКА УНИПОТЕНТНЫХ ПОТОКОВ НА ОДНОРОДНЫХ ПРОСТРАНСТВАХ

Этьен Гис

ВВЕДЕНИЕ

Рассмотрим поток ϕ^t на торе $T^n = \mathbb{R}^n/\mathbb{Z}^n$, определенный формулой $\phi^t(x \bmod \mathbb{Z}^n) = (x + tw) \bmod \mathbb{Z}^n$, где $w = (w_1, \dots, w_n)$ — вектор из $\mathbb{R}^n$. Со времен Л. Кронекера известно, что если w_i линейно независимы над $\mathbb{Q}$, то все орбиты потока ϕ^t плотны в торе. Вообще, замыкание орбиты есть подтор тора T^n размерности, равной рангу системы $\{w_i\}$ над $\mathbb{Q}$.

Пусть теперь G — связная группа Ли и Γ — решетка в G , т. е. дискретная подгруппа, такая, что объем (Хаара) пространства правых смежных классов $\Gamma \backslash G$ конечен. Произвольная подгруппа H группы G естественно действует справа на $\Gamma \backslash G$. При каких условиях справедлив результат, аналогичный результату Кронекера, гарантирующий однородность замыканий орбит?

В серии из четырех статей М. Ратнер получила весьма удовлетворительный ответ на этот вопрос, доказав тем самым гипотезу М.С. Рагунатана [71–74]. Чтобы сформулировать его, условимся говорить, что элемент g группы Ли G является *унипотентным*, если присоединенный автоморфизм $\text{Ad}(g)$ алгебры Ли $\mathfrak{g}$ унипотентен, т. е. 1 является его единственным собственным значением. Подгруппа H группы G унипотентна, если все ее элементы унипотентны.

Теорема (Ратнер, 1990). Пусть H — унипотентная подгруппа группы Ли G и Γ — решетки в G . Для любого x из $\Gamma \backslash G$ существует замкнутая подгруппа $H(x)$ группы G , такая, что замыкание орбиты xH точки x относительно H в $\Gamma \backslash G$ совпадет с орбитой $xH(x)$ точки x относительно $H(x)$.

Эта формулировка — не самая лучшая из возможных, далее мы дадим другие варианты.

Неудивительно, что этот результат имеет следствия, касающиеся диофантовых приближений. До того как предыдущая теорема была доказана, Г. А. Маргулис развил методы, которые позволили ему в 1987 г. доказать гипотезу А. Оппенгейма [50].

Теорема (Маргулис, 1987). Пусть Q — неопределенная невырожденная квадратичная форма на $\mathbf{R}^n$ ($n \geq 3$). Предположим, что Q не кратна никакой рациональной форме. Тогда для любого $\varepsilon > 0$ существует вектор v из $\mathbf{Z}^n$, такой, что $0 < Q(v) < \varepsilon$.

Как мы увидим далее, впоследствии С. Г. Дани и Г. А. Маргулис значительно улучшили эту теорему.

Доказательства теорем Дани—Маргулиса и Ратнер длинные и трудные, и нет возможности их здесь привести. В этом элементарном докладе мы поставили себе очень скромную цель. Мы предполагаем проиллюстрировать примененные там методы в простом случае, когда группа G есть $SL(2, \mathbf{R})$: в этом случае речь идет об изучении орициклического потока на поверхности кривизны -1 . Этот случай уже включает значительную часть всех трудностей. Мы надеемся, что такое предварительное изучение позволит читателю легче понять полные доказательства, которые уже опубликованы или находятся в процессе публикации. Вместе с тем отметим статью [27] Дани и Маргулиса, в которой предлагается элементарное введение в вопросы диофантовых приближений.

Этот доклад построен следующим образом. Первый раздел посвящен мотивировкам: в нем описывается история исследований орициклического потока и даются более точные формулировки теорем Ратнер. В разд. 2 описываются некоторые методы, использованные Дани и Маргулисом для подхода к гипотезе Рагунатана. В разд. 3 даются некоторые части доказательства Ратнер, которые все время иллюстрируются случаем $SL(2, \mathbf{R})$. Наконец, в разд. 4 мы обсуждаем три приложения: значения квадратичных форм в целочисленных точках, явления эргодической жесткости и существования геодезических ламинаций на гиперболических многообразиях.

1. ОРИЦИКЛИЧЕСКИЙ ПОТОК

1.1. Геодезический и орициклический потоки на пространствах отрицательной кривизны являются очень богатыми примерами динамических систем. Параллельно с развитием эргодической теории производилось все более и более глубокое исследование этих потоков. Несмотря на то что они имеют очень разное топологическое и эргодическое поведения, их нельзя изучать один без другого.

Сначала напомним определения. Пусть H^2 — полуплоскость Пуанкаре $\{x + iy \in \mathbf{C} \mid y > 0\}$, снабженная метрикой кривизны -1 , определенной формулой $ds^2 = (dx^2 + dy^2)/y^2$. Геодезическая в H^2 —

это кривая $\gamma : \mathbf{R} \rightarrow H^2$, которая является изометрическим погружением. Образ геодезической есть полуокружность (или полупрямая), ортогональная к оси x . Любая геодезическая полностью определяется своим касательным вектором в точке $t = 0$, элементом единичного касательного расслоения $T_1 H^2$ над H^2 . Очевидно, что если γ — геодезическая и $s \in \mathbf{R}$, то отображение $\gamma^s : t \in \mathbf{R} \mapsto \gamma(t + s) \in H^2$ также является геодезической. Следовательно, однопараметрическая группа преобразований $\gamma \mapsto \gamma^s$ определяет поток на пространстве геодезических, т. е. на $T_1 H^2$. Это *геодезический поток* на H^2 .

Группа собственных изометрий полуплоскости H^2 — это группа $\text{PSL}(2, \mathbf{R})$, действующая на H^2 по формуле $\pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} (z) = \frac{az+b}{cz+d}$. Она действует свободно и транзитивно слева на $T_1 H^2$ и, очевидно, коммутирует с геодезическим потоком. Отсюда следует, что геодезический поток может также рассматриваться как однопараметрическая группа g^s правых сдвигов на $\text{PSL}(2, \mathbf{R})$. Легко убедиться, что этот поток отождествляется со следующей группой сдвигов:

$$g^s : \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{PSL}(2, \mathbf{R}) \\ \mapsto \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \exp(s/2) & 0 \\ 0 & \exp(-s/2) \end{pmatrix} \in \text{PSL}(2, \mathbf{R}).$$

Пусть $v \in T_1 H^2$ и $\gamma : \mathbf{R} \rightarrow H^2$ — геодезическая, заданная вектором v . Если T стремится к $+\infty$ (соответственно к $-\infty$), то (неевклидова) окружность с центром $\gamma(T)$, проходящая через $\gamma(0)$, «сходится» к положительному (соответственно отрицательному) *орициклу*, определенному вектором v (см. рис 1). В модели Пуанкаре это окружность, касательная к оси x (с выколотой точкой касания), или прямая, параллельная оси. С внутренней точки зрения это кривые кривизны ± 1 .

Так же, как геодезические, орициклы позволяют определить некоторый поток. Если v — единичный касательный вектор к H^2 в точке z , а t — действительное число, то обозначим через $v \cdot h_+^t$ единичный вектор, ортогональный к положительному орициклу, заданному вектором v , в точке, расположенной на расстоянии t от z вдоль орицикла (ориентации указаны на рис. 1). По той же причине, что и выше, этот поток отождествляется с некоторой однопараметрической подгруппой h_+^t правых сдвигов на $\text{PSL}(2, \mathbf{R})$.

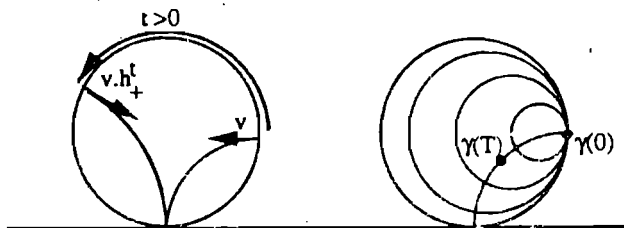


Рис. 1.

Нетрудно проверить, что речь идет о таких сдвигах:

$$\pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot h_{\pm}^t = \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Используя отрицательные орициклы, можно построить также третий поток h_{-}^t , который, очевидно, соответствует правым сдвигам на $\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$. Два потока h_{+}^t и h_{-}^t сопряжены относительно симметрии, отображающей вектор v расслоения $T_1 H^2$ в противоположный ему. Поэтому мы ограничимся изучением h_{+}^t .

Отметим, что все эти потоки сохраняют (биливариантную) меру Хаара на $\text{PSL}(2, \mathbb{R})$. С другой стороны, левоинвариантные векторные поля, соответствующие потокам g^s , h_{+}^t и h_{-}^t , образуют базис алгебры Ли группы $\text{PSL}(2, \mathbb{R})$.

Связь между геодезическими и орициклическими потоками заключается в следующей формуле, очень важной для дальнейшего:

$$g^s h_{\pm}^t g^{-s} = h_{\pm}^{t \exp(\pm s)}. \quad (*)$$

Ее очень легко проверить, используя матричное исчисление. Следовательно, геодезический поток нормализует орициклические потоки. Он сжимает положительные орициклы и растягивает отрицательные. Это наблюдение является истоком теории потоков Аносова [1] (см. рис. 2).

Пусть теперь Γ — дискретная подгруппа конечного кообъема группы $\text{PSL}(2, \mathbb{R})$. Если действие группы Γ на H^2 свободно, мы получаем поверхность $\Sigma = \Gamma \backslash H^2$, площадь которой конечна, а ее единичное касательное расслоение отождествляется с $\Gamma \backslash \text{PSL}(2, \mathbb{R})$. В общем случае действие группы Γ на H^2 может не быть свободным, но стабилизаторы конечны и факторпространство $\Sigma = \Gamma \backslash H^2$ является поверхностью с особенностями, которую лучше рассматривать

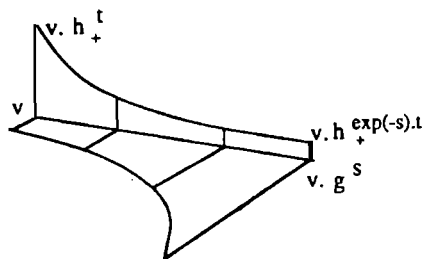


Рис. 2.

как «орбиобразие» (в смысле Тёрстона). Пример $\Gamma = \text{PSL}(2, \mathbb{Z})$ очень важен: мы получаем модулярное орбиобразие, изучение которого является основополагающим в теории чисел.

Потоки $h_{\pm}^t$ и g^s , конечно, переносятся на фактор $\Gamma \backslash \text{PSL}(2, \mathbb{R})$: получаются орициклические и геодезические потоки на поверхности (соответственно орбиобразии) $\Gamma \backslash H^2$. Такие потоки не имеют неподвижных точек. Они сохраняют форму объема с конечной полной массой, которая может называться мерой Хаара или мерой Лиувилля в зависимости от принятой точки зрения.

1.2. Для иллюстрации того, что геодезические и орициклические потоки изучаются параллельно, мы прежде всего рассмотрим несколько простых и довольно стандартных примеров.

Предложение. Если $\Gamma \backslash \text{PSL}(2, \mathbb{R})$ компактно, то орициклический поток $h_{\pm}^t$ не имеет периодических орбит.

Доказательство. Соотношение (*) показывает, что если γ — периодическая орбита потока h_{+}^t с периодом T , то образ $\gamma \cdot g^s$ является другой периодической орбитой с периодом $\exp(-s) \cdot T$, стремящимся к 0, когда s стремится к $+\infty$. Предложение вытекает из того факта, что поток без неподвижных точек на компактном многообразии не может иметь периодических орбит произвольно малых периодов. $\square$

В случае когда $\Gamma \backslash \text{PSL}(2, \mathbb{R})$ имеет конечный объем, но некомпактно, поверхность $\Sigma = \Gamma \backslash H^2$ (для простоты предполагаемая неособой) разлагается на компактное подмножество и конечное число каспов (см. рис. 3). Каждый из этих каспов изометричен фактору множества $\{x + iy \in \mathbb{C}, y \geq y_0 > 0\}$ по целочисленным сдвигам

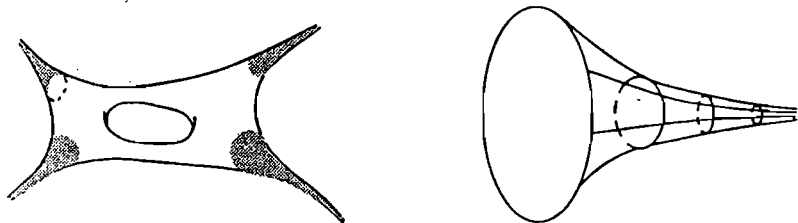


Рис. 3.

$z \mapsto z + n$. Кривая $y = y_0$ отображается в замкнутый орицикл поверхности Σ , т. е. замкнутую орбиту γ потока h_+^t . Очевидно, для любого действительного числа s кривая $\gamma \cdot g^s$ также является периодической орбитой. При s , стремящемся к $+\infty$, такая орбита $\gamma \cdot g^s$ стремится к концу поверхности Σ (см. рис. 3). Нетрудно убедиться, что таким образом мы получаем все периодические орбиты потока h_+^t , которые образуют конечное число семейств, параметризованных множеством $\mathbf{R}$. При s , стремящемся к $-\infty$, оказывается, что орбиты $\gamma \cdot g^s$ становятся все более и более плотными в $T_1 \Sigma$. В [19, 77] и [85] есть описание этого явления и интересная переформулировка гипотезы Римана в терминах замкнутых орициклов.

Вот пример, когда орициклический поток дает возможность понять геодезический поток. Это известная теорема Э. Хопфа [41–42]:

Теорема. Если Γ — решетка группы $\mathrm{PSL}(2, \mathbf{R})$, то геодезический поток на $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$ является эргодическим по отношению к мере Хаара, т. е. измеримая функция, постоянная на орбитах потока, постоянна почти всюду.

Доказательство. Это пример применения «феномена Маутнера» (см. [55]). Пусть λ — мера Хаара на $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$, нормированная так, что она является вероятностной. Пусть F — функция с λ -интегрируемым квадратом на $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$, инвариантная относительно геодезического потока g^s . Сейчас мы покажем, что F также инвариантна (λ -почти всюду) относительно орициклического потока h_+^t . В самом деле, использование формулы (*) и того факта, что g^s сохраняет λ , показывает, что

$$\begin{aligned} \int F(v) \cdot F(v \cdot h_+^t) d\lambda &= \int F(v \cdot g^s) \cdot F(v \cdot g^s h_+^t) d\lambda = \\ &= \int F(v) \cdot F(v \cdot h_+^{\exp(-s) \cdot t}) d\lambda \xrightarrow{s \rightarrow +\infty} \int F(v) d\lambda. \end{aligned}$$

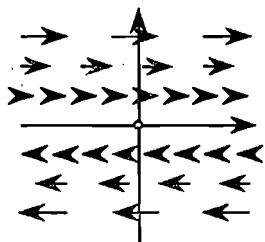


Рис. 4.

Таким образом, $F(v \cdot h_+^t) = F(v)$ для λ -почти всякого v (случай равенства в неравенстве Коши—Шварца). Аналогично рассуждаем с h_-^t (устремляя s к $-\infty$). Следовательно, F инвариантна (λ -почти всюду) относительно g^s , h_+^t и h_-^t , а значит, относительно правого действия группы $\text{PSL}(2, \mathbb{R})$, которое транзитивно. Этим показано, что F постоянна λ -почти всюду. $\square$

Обратно, геодезический поток позволит нам доказать следующий результат, принадлежащий Хедлунду [39–40]:

Теорема. Если Γ — решетка группы $\text{PSL}(2, \mathbb{R})$, то орициклический поток h_+^t на $\Gamma \backslash \text{PSL}(2, \mathbb{R})$ эргодичен по отношению к мере Хаара.

Доказательство. Мы приведем некоторые детали, так как идея доказательства лежит в основе многих примеров, с которыми мы познакомимся далее.

Стабилизатор точки $(1, 0)$ из $\mathbb{R}^2$ при линейном действии группы $\text{SL}(2, \mathbb{R})$ имеет вид $H = \{h_+^t\} = \left\{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \right\}$. Следовательно, однородное пространство $\text{SL}(2, \mathbb{R})/H$ отождествляется с $\mathbb{R}^2 - \{(0, 0)\}$, а левое действие подгруппы H на $\text{SL}(2, \mathbb{R})/H$ — это, очевидно, действие трансвекциями (см. рис. 4):

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + ty \\ y \end{pmatrix}.$$

Ясно, что непрерывная функция на $\mathbb{R}^2 - \{(0, 0)\}$, инвариантная относительно этих трансвекций, является функцией от y . В частности, такая функция постоянна на оси x и, следовательно, на орбите точки $(1, 0)$ относительно потока $g^s = \begin{pmatrix} \exp(s/2) & 0 \\ 0 & \exp(-s/2) \end{pmatrix}$.

Итак, мы получили следующую лемму. Если $\Phi : \mathrm{SL}(2, \mathbf{R}) \rightarrow \mathbf{R}$ — непрерывная функция, постоянная на двусторонних смежных классах по подгруппе H , то Φ постоянна на однопараметрической подгруппе g^s .

Прежде чем применять этот факт для нашей цели, заметим, что однородное пространство группы $\mathrm{PSL}(2, \mathbf{R})$ также является однородным пространством группы $\mathrm{SL}(2, \mathbf{R})$; поэтому достаточно рассмотреть решетку Γ группы $\mathrm{SL}(2, \mathbf{R})$ (расширяя ее, если нужно, при помощи $\mathbf{Z}/2\mathbf{Z}$).

Итак, предположим, что F — функция с λ -интегрируемым квадратом на $\Gamma \backslash \mathrm{SL}(2, \mathbf{R})$, инвариантная относительно потока h_+^t . Функция Φ , определенная на $\mathrm{SL}(2, \mathbf{R})$ формулой

$$\Phi(g) = \int F(v) \cdot F(v \cdot g) d\lambda,$$

непрерывна и удовлетворяет равенствам

$$\Phi(g) = \Phi(h_+^t g) \quad \text{и} \quad \Phi(g) = \Phi(g h_+^t),$$

так как F инвариантна относительно h_+^t и h_+^t сохраняет λ .

Применяя предыдущее замечание, получаем

$$\int F^2(v) d\lambda = \int F(v) \cdot F(v \cdot g^s) d\lambda.$$

Следовательно, F также инвариантна (λ -почти всюду) относительно геодезического потока g^s . Как мы знаем, из этого следует, что F постоянна λ -почти всюду. $\square$

За дальнейшей информацией о геодезическом потоке мы отсылаем к докладу Пансю на этом семинаре [62]. Отметим, однако, что эта динамика очень сложна. Например:

(i) объединение периодических орбит является плотным подмножеством [1, 6].

(ii) некоторые орбиты нерекурсивны: орбита может накручиваться, например, на периодическую орбиту (рис. 5) или на более сложное предельное множество.

(iii) существуют инвариантные замкнутые множества, не содержащие никакой периодической орбиты [56].

(iv) существует несчетное множество инвариантных и эргодических вероятностных мер [81].

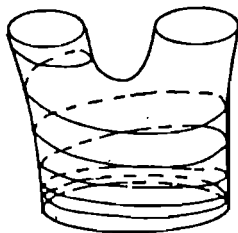


Рис. 5.

Для орициклических потоков ситуация совсем другая и в некотором смысле значительно более простая. Следующая теорема принадлежит Хедлунду [38]; она будет доказана в разд. 3.

Теорема. Пусть $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$ имеет конечный объем; тогда любая орбита орициклического потока h_+^t является периодической или плотной.

Естественно попытаться описать асимптотическое распределение орбит в $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$. Пусть ϕ^t — поток на многообразии V , сохраняющий вероятностную меру μ . Точка v из V равномерно распределена по отношению к μ , если для любой непрерывной функции $F : V \rightarrow \mathbf{R}$ с компактным носителем

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T F(v \cdot h_+^t) dt = \int_V F d\mu.$$

Эргодическая теорема Биркгофа гарантирует, что если μ — эргодическая мера, то μ -почти любая точка v равномерно распределена по отношению к μ .

Следующая теорема принадлежит Дани и Смилли [28]; ей предшествовало множество работ (в частности [7, 8, 32, 35, 37, 46, 84]). набросок доказательства более общего результата мы приведем в разд. 3.

Теорема. Если $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$ имеет конечный объем, то любая непериодическая относительно потока h_+^t точка является равномерно распределенной по отношению к мере Хаара.

1.3. Конечно, невозможно долго ограничиваться изучением геодезических и орициклических потоков. Естественно изучать, более общим образом, правое действие подгруппы H группы Ли G на факторпространстве $G \backslash G$, имеющем конечный объем [2, 9]. Для этого

имеется много причин, и некоторые из них мы приведем в разд. 4. Мы отсылаем читателя к статье Дани [19], содержащей прекрасный обзор соответствующих работ. Особо отметим очень полное исследование случая, когда G является нильпотентной, в работах В. Парри [63, 64] (и в более новых работах [44, 45]), а также исследование действия орисферических групп, сделанное Дани [21]. Именно на результатах, полученных этими авторами, основывалась гипотеза Рагунатана о том, что если H унипотентна, то замыкание орбиты в $\Gamma \backslash G$ совпадает с орбитой некоторой подгруппы Ли группы G , содержащей H . В частности, из этой гипотезы следует, что замыкания орбит являются подмногообразиями, — ситуация, существенно отличная от случая геодезического потока...

Решительный прогресс в отношении этой гипотезы был достигнут Маргулисом и Дани, которые доказали ее для случая $G = \mathrm{SL}(3, \mathbb{R})$ [26]. Мы кратко опишем их методы в разд. 2. Фактически Маргулис в [50–52] сформулировал более сильное предположение, чем гипотеза Рагунатана, которое было доказано Ратнер [74, 75].

Теорема. Пусть H — подгруппа связной группы Ли G , порожденная унипотентными элементами. Пусть Γ — решетка группы G . Тогда замыкание любой орбиты подгруппы H на $\Gamma \backslash G$ совпадает с орбитой некоторой подгруппы Ли группы G , содержащей H .

Подход Ратнер состоит в том, чтобы сначала изучить природу мер, инвариантных относительно H , и затем, на основании этого, — строение замыканий орбит.

Фиксируем по-прежнему решетку Γ связной группы Ли G , и пусть x — элемент из $\Gamma \backslash G$. Рассмотрим замкнутую подгруппу L группы G , такую, что $x \cdot L$ замкнуто в $\Gamma \backslash G$. Кроме того, предположим, что орбита $x \cdot L \simeq L/x\Gamma x^{-1} \cap L$ имеет конечный объем по отношению к мере Хаара подгруппы L . Поскольку xL содержится в $\Gamma \backslash G$, мы получаем конечную меру на $\Gamma \backslash G$. Полученные этим методом меры будут называться *однородными мерами* на $\Gamma \backslash G$. Ратнер получила следующий результат [70–73]:

Теорема. Пусть H — подгруппа группы Ли G , порожденная унипотентными элементами. Пусть Γ — решетка группы G . Любая мера конечной полной массы на $\Gamma \backslash G$, инвариантная и эргодическая относительно действия подгруппы H , является однородной мерой.

Для случая, когда H — однопараметрическая группа ϕ^t , Ратнер получила обобщение теоремы Дани и Смилли.

Теорема. Пусть ϕ^t — унитарная однопараметрическая подгруппа связной группы Ли G , и пусть Γ — решетка группы G . Тогда любая точка из $\Gamma \backslash G$ равномерно распределена по отношению к некоторой однородной мере.

Закончим этот раздел тремя замечаниями.

1. Результаты Ратнер в действительности являются более общими и покрывают некоторые случаи, когда Γ не является решеткой. Мы не даем здесь формулировки, но в п. 3.4 приведем пример.

2. Класс групп, порожденных унитарными элементами, гораздо шире, чем класс унитарных подгрупп. Например, любая подгруппа Ли группы $SL(n, \mathbb{R})$, изоморфная $SL(2, \mathbb{R})$, порождается унитарными элементами. Типичным примером группы, не порождаемой унитарными элементами, является разрешимая подгруппа H группы $GL(n, \mathbb{R})$, образованная всеми верхнетреугольными матрицами, так как унитарные элементы из H — это строго верхнетреугольные матрицы.

3. Предположение « H порождена унитарными элементами» почти оптимально. Старков показал, что если H не порождается такими элементами h , что спектр оператора $\text{Ad}(h)$ лежит на единичной окружности, то существуют замыкания орбит в $\Gamma \backslash G$, которые не являются подмногообразиями. Он также показал (используя результаты Ратнер), что, напротив, если H порождается такими элементами, то все замыкания орбит являются подмногообразиями (но не обязательно орбитами какой-либо подгруппы Ли L) [83].

2. ТОПОЛОГИЧЕСКАЯ ДИНАМИКА УНИТАРНЫХ ПОТОКОВ

2.1. В этом разделе мы предлагаем элементарное введение в методы, развитые Маргулисом. Начнем с нескольких общих замечаний, касающихся динамики унитарных потоков. Фиксируем связную группу Ли G , решетку Γ и однопараметрическую унитарную подгруппу ϕ^t . Для простоты мы временно предположим, что $\Gamma \backslash G$ компактно.

Левые сдвиги группы G позволяют отождествить касательные пространства в различных точках группы G с алгеброй Ли $\mathcal{G}$ (левинвариантных полей). Поток ϕ^t действует справа на $\Gamma \backslash G$, а его дифференциал, вычисленный в $\mathcal{G}$, задается присоединенным оператором $\text{Ad}(\phi^t)$. Налицо замечательное свойство: дифференциал потока, вычисленный на тривиализации касательного расслоения, не

зависит от точки, в которой он рассматривается. Можно сравнить это свойство с понятием «автономности», введенным Зегибом [94].

Поскольку ϕ^t унипотентен, $\text{Ad}(\phi^t)$ является многочленом от t . Если фиксировать на $\mathcal{G}$ евклидову структуру $\| \cdot \|$ и если $w \in \mathcal{G}$ — ненулевой вектор, то функция $t \mapsto \| \text{Ad}(\phi^t)(w) \|^2$ есть многочлен, который принимает только строго положительные значения. В частности,

$$\inf_{t \in \mathbb{R}} \| \text{Ad}(\phi^t)(w) \| > 0.$$

Это свойство *инфинитезимальной дистальности* потока ϕ^t . Оно лежит в основе большей части доказательств, касающихся унипотентных потоков, позволяя исключить существование устойчивых многообразий и, следовательно, явлений накручивания. Общее изучение инфинитезимально дистальных потоков было предпринято в [76], но еще многое остается неясным. Например, вытекает ли из этого свойства, что замыкания орбит являются подмногообразиями?

Сформулируем аналогичное неинфинитезимальное свойство. Если x и y — достаточно близкие точки из $\Gamma \backslash G$, то существует единственный малый элемент $w(x, y)$ алгебры $\mathcal{G}$, такой, что $y = x \cdot \exp w(x, y)$. Очевидно, $y \cdot \phi^t = x \cdot \phi^t \cdot \phi^{-t} \cdot \exp w(x, y) \cdot \phi^t$, так что если x и y таковы, что $x \cdot \phi^t$ и $y \cdot \phi^t$ остаются близкими для $t \in [0, T]$, то

$$w(x \cdot \phi^t, y \cdot \phi^t) = \text{Ad}(\phi^t)w(x, y).$$

Итак, имеется свойство *полиномиальной расходимости* орбит. Отметим тот факт, что унипотентный поток не обязательно является дистальным, т. е. что расстояние между $x \cdot \phi^t$ и $y \cdot \phi^t$ в $\Gamma \backslash G$ может не ограничиваться снизу строго положительным числом [36]. Например, орициклический поток — не дистальный.

2.2. Плотность орициклов в кокомпактном случае. Мы хотим доказать теорему Хедлунда, касающуюся замыканий орбит. Это доказательство стимулировано работами [50, 52, 54] и напоминает работы [6] и [65].

Пусть $G = \text{SL}(2, \mathbb{R})$. Положим $H = \{h_+^t\} = \left\{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \right\}$, и пусть B — подгруппа матриц $\begin{pmatrix} \exp(s) & t \\ 0 & \exp(-s) \end{pmatrix}$, нормализующая H в группе G . Пусть Γ — решетка в G . Начнем со следующего классического факта:

Предложение. *Все орбиты подгруппы B в $\Gamma \backslash G$ плотны.*

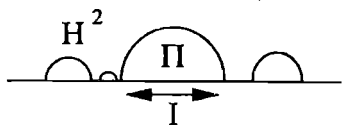


Рис. 6.

Доказательство. Речь идет о доказательстве того, что все двойные смежные классы $\Gamma \cdot x \cdot B$ плотны в G или, равносильно, что все орбиты решетки Γ в $G \setminus B$ плотны. Предположим для простоты, что Γ содержит $-\text{id}$, так что достаточно доказать плотность орбит решетки Γ в $\text{PSL}(2, \mathbb{R})/B$, т. е. на границе $\mathbb{R} \cup \{\infty\}$ полуплоскости Пуанкаре. (При $-\text{id} \notin \Gamma$ доказательство аналогично.) Напротив, пусть существует орбита решетки Γ , которая неплотна, и пусть I — связная компонента дополнения к замыканию этой орбиты (см. рис. 6). Геодезическая плоскости H^2 , соединяющая концы интервала I , определяет вместе с I некоторую полуплоскость Π в H^2 . Стабилизатор Γ_Π полуплоскости Π является циклической подгруппой и, следовательно, факторпространство $\Gamma_\Pi \backslash \Pi$ имеет бесконечную площадь. Это противоречит тому, что Σ имеет конечную площадь. $\square$

Предположим теперь, что Γ кокомпактна, т. е. что $\Gamma \backslash G$ компактно. Подмножество M пространства $\Gamma \backslash G$ называется *минимальным множеством*, если это замкнутое непустое подмножество, инвариантное относительно H и минимальное среди подмножеств с этими свойствами. Такое минимальное множество существует в силу леммы Цорна и компактности пространства $\Gamma \backslash G$. Любая орбита точки из M относительно группы H плотна в M . Мы покажем, что $M = \Gamma \backslash G$, т. е. что любая орбита орициклического потока плотна.

Используя правое действие группы G на $\Gamma \backslash G$, положим

$$M = \{g \in G \mid M \cdot g \cap M \neq \emptyset\}.$$

Ясно, что

- i) M замкнуто и содержит нейтральный элемент id ;
- ii) M является объединением двойных смежных классов из $\Gamma \backslash G/H$, поскольку M инвариантно относительно H .

Пусть x — точка из M . Мы знаем, что орбита точки x относительно H не является периодической и что она плотна в M . Следовательно, существует последовательность g_n элементов из G , сходящаяся к единице и такая, что $g_n \notin H$ и $x \cdot g_n \in M$. В частности,

iii) id — точка прикосновения для $M - H$.

Пусть $g \in M \cap B$. Ввиду того, что элементы подгруппы B нормализуют H , образ $M \cdot g$ есть непустое замкнутое множество, инвариантное относительно H и имеющее непустое пересечение с M . Следовательно, $M \cdot g = M$. Таким образом,

iv) $M \cap B$ — замкнутая подгруппа, образованная элементами, сохраняющими M .

Мы уже изучали пространство $H \backslash G / H$ в п. 1.2. Проекция M' множества M на $G / H \simeq \mathbb{R}^2 - \{(0, 0)\}$ есть замкнутое множество, накапливающееся к $(1, 0)$ и инвариантное относительно трансвекций $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$. Отсюда непосредственно выводится, что M' содержит последовательность точек вида $(\exp(s_n), 0)$, где $s_n \neq 0$ сходятся к 0.

Если s — действительное число, то $(\exp(s), 0)$ принадлежит M' тогда и только тогда, когда $\begin{pmatrix} \exp(s) & t \\ 0 & \exp(-s) \end{pmatrix}$ принадлежит M (для произвольных значений t). Таким образом, замечание iv) показывает, что множество таких действительных чисел s является замкнутой аддитивной подгруппой в $\mathbb{R}$, и, согласно предыдущему, она недискретна. Отсюда вытекает, что M содержит B . Используя iv), получаем, что M инвариантно относительно группы B . Так как мы знаем, что орбиты группы B плотны в $\Gamma \backslash G$, то мы показали, что $M = \Gamma \backslash G$. Этим заканчивается доказательство теоремы Хедлунда в случае, когда $\Gamma \backslash G$ компактно.

2.3. Случай некокомпактных решеток: отсутствие сходимости к бесконечности. Рассмотрим теперь случай, когда $\Gamma \backslash G$ имеет конечный объем, но не компактно. В предыдущем доказательстве компактность позволяла гарантировать существование минимального множества. Периодическая орбита, разумеется, есть минимальное множество, но в нашем случае, когда $\Gamma \backslash G$ имеет конечный объем, неясно, любое ли инвариантное непустое замкнутое множество содержит минимальное множество. Это первое свойство, которое нужно проверить.

Предложение. *Любое непустое замкнутое подмножество в $\Gamma \backslash G$, инвариантное относительно орициклического потока, содержит периодическую орбиту.*

Прежде чем доказывать это предложение, мы должны сформулировать две леммы.

Мы уже видели, что поверхность (орбитообразия) $\Sigma = \Gamma \backslash H^2$ может быть разложена на компактную часть K и конечное число кас-

пов, изометричных фактору множества $\{x + iy \in \mathbb{C} \mid y > y_0 > 0\}$ по целочисленным сдвигам $z \mapsto z + n$.

Лемма. Пусть $\gamma : \mathbb{R} \rightarrow \Sigma$ — непериодический орицикл поверхности Σ . Тогда множества $\{t \geq 0 \mid \gamma(t) \in K\}$ и $\{t \leq 0 \mid \gamma(t) \in K\}$ неограниченны.

Доказательство. Достаточно констатировать, что полуорицикл плоскости H^2 , содержащийся в $\{x + iy \mid y > y_0\}$, является горизонтальной полупрямой и, следовательно, порождает периодический орицикл в Σ . $\square$

Поскольку всякое замкнутое непустое подмножество в $\Gamma \backslash G$, инвариантное относительно H и не содержащее периодических орбит, пересекается с компактом $\bar{K}$ единичных касательных векторов к Σ в точках из K , применима лемма Цорна. Из нее вытекает, что любое инвариантное непустое замкнутое множество содержит минимальное множество. Следующая лемма, (легкое) доказательство которой мы оставляем читателю, показывает, что любое минимальное множество обязательно компактно. (Более общие формулировки см. в [53].)

Лемма. Пусть ϕ^t — поток на локально компактном пространстве X . Предположим, что существует компакт $\bar{K}$ в X , такой, что для любого x множества $\{t \geq 0 \mid x \cdot \phi^t \in \bar{K}\}$ и $\{t \leq 0 \mid x \cdot \phi^t \in \bar{K}\}$ в $\mathbb{R}$ неограниченны. Тогда X компактно.

Теперь мы можем доказать предложение. Мы видели, что любое замкнутое непустое инвариантное множество содержит компактное минимальное множество M . Если бы множество M не сводилось к периодической орбите, то рассуждение из п. 2.2 было бы применимо без изменения и мы имели бы $M = \Gamma \backslash G$, что неверно, так как мы предположили, что $\Gamma \backslash G$ некомпактно. $\square$

Предложение, которое мы сейчас доказали, очевидно, слабее, чем теорема Хедлунда, согласно которой непериодическая орбита плотна. Для ее доказательства мы можем теперь применить тот же метод, что и в п. 2.2. Пусть x — непериодическая точка пространства $\Gamma \backslash G$ и F — замыкание орбиты точки x в $\Gamma \backslash G$. Мы знаем, что F содержит периодическую точку y с периодом $T > 0$. Положим

$$M = \{g \in G \mid y \cdot g \in F\}.$$

Тогда M — замкнутое множество, являющееся объединением правых смежных классов пространства G/H , которое инвариантно относительно левого действия потока h_+^T и содержит последовательность g_n из $G - H$, сходящуюся к id . Соответствующее замкнутое множество M' в $\mathbb{R}^2 - \{(0, 0)\}$ инвариантно относительно $\begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix}$ и накапливается к $(1, 0)$. Мы знаем, что точки $y \cdot g$, где $g \in B$, являются периодическими для орициклического потока и, с другой стороны, непериодическая орбита точки x накапливается к точке y . Отсюда вытекает, что $(1, 0)$ является пределом последовательности точек множества M' , лежащих вне $\mathbb{R} \times \{0\}$.

Все эти свойства доказывают, что M' содержит $\mathbb{R}^* \times \{0\}$. Иначе говоря, для любого g из B имеем $y \cdot g \in F$. Так как все орбиты подгруппы B плотны в $\Gamma \backslash G$, то мы установили, что $F = \Gamma \backslash G$, т. е. что непериодическая орбита точки x плотна в $\Gamma \backslash G$. $\square$

2.4. Некоторые указания для общего случая. Бóльшая часть техники, описанной в пп. 2.2. и 2.3, распространяется на случай унитарной подгруппы группы Ли, отличной от $\text{SL}(2, \mathbb{R})$. Разработка этих идей позволила Дани и Маргулису доказать гипотезу Рагунатана в случае, когда $G = \text{SL}(3, \mathbb{R})$, $\Gamma = \text{SL}(3, \mathbb{Z})$ и ϕ^t — однопараметрическая унитарная группа общего положения (т. е. такая, что $\phi^t - \text{id}$ имеет ранг 2 для любого $t \neq 0$). Мы не будем вдаваться в детали доказательства, так как в [27] дается превосходное элементарное введение.

В доказательстве теоремы Хедлунда, которое мы привели, важно установить, что непериодический орицикл может только конечное время пребывать в каспе, не выходя из него. Аналогичная формулировка справедлива в общем случае, но ее доказательство значительно сложнее. Вот два результата в этом направлении.

Теорема. Пусть G — связная полупростая группа Ли и Γ — решетка. Пусть задано $\epsilon > 0$; тогда существует компакт K пространства $\Gamma \backslash G$, такой, что для любой точки x из $\Gamma \backslash G$ и любой однопараметрической унитарной подгруппы ϕ^t имеются две возможности:

- (i) существует собственная замкнутая подгруппа L группы G , содержащая ϕ^t , такая, что орбита $x \cdot L$ замкнута и имеет конечную меру (относительно меры Хаара на L);
- (ii) если T достаточно велико, то $m(\{t \in [0, T] \mid x\phi^t \in K\}) \geq (1 - \epsilon)T$, где m обозначает меру Лебега на $\mathbb{R}$.

Теорема. Пусть G — связная группа Ли, Γ — решетка, ϕ^t — однопараметрическая унитарная подгруппа, x — точка пространства $\Gamma \backslash G$ и $\varepsilon > 0$. Существует компакт K в $\Gamma \backslash G$, такой, что $m(\{t \in [0, T] \mid x \cdot \phi^t \in K\}) \geq (1 - \varepsilon)T$ для достаточно большого T .

Первая версия этих результатов принадлежат Маргулису, который показал, что орбита унитарного потока в $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R})$ не стремится к бесконечности [48]. Дани доказал улучшенный вариант этого результата в виде первой из двух теорем, цитированных выше [13, 14, 16, 18, 22]. Доказательство Дани разделено на случай, когда $\mathbf{R}$ -ранг группы G больше или равен 2 (тогда используется теорема арифметичности Маргулиса), и случай, когда этот ранг равен 1 (для которого доказательство другое и более легкое). Формулировка второй из приведенных теорем взята из [74]. Для введения в методы доказательства мы отсылаем к приложению из [27]. Было бы, впрочем, приятно найти общий подход к «несходимости к бесконечности», не опирающийся, например, на трудные теоремы арифметичности.

3. ЭРГОДИЧЕСКОЕ ПОВЕДЕНИЕ УНИПОТЕНТНЫХ ПОТОКОВ

В этом разделе мы опишем некоторые из методов, используемых Ратнер.

3.1. Дрейф в направлении централизатора. Алгебра Ли $\mathfrak{sl}(2, \mathbf{R})$ группы $SL(2, \mathbf{R})$ состоит из (2×2) -матриц с нулевым следом. Базис имеет вид

$$g = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad h_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad h_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Имеем

$$\begin{aligned} \text{Ad}(h_+^t) \begin{pmatrix} a & c \\ b & -a \end{pmatrix} &= \begin{pmatrix} 1 & -t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & c \\ b & -a \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} a - tb & t^2b + 2at + c \\ b & -a + tb \end{pmatrix}. \end{aligned}$$

Снабдим $\mathfrak{sl}(2, \mathbf{R})$ нормой $\sup(|a|, |b|, |c|)$ и фиксируем $C > 0$ и $\varepsilon > 0$ (очень малое по сравнению с C). Пусть $w = \begin{pmatrix} a & c \\ b & -a \end{pmatrix} \in \mathfrak{sl}(2, \mathbf{R})$

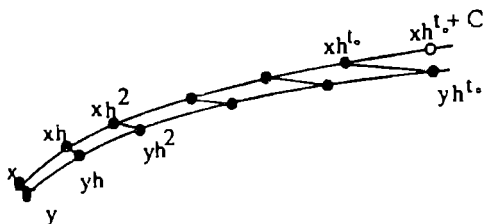


Рис. 7.

имеет норму $\rho < C$. Тогда $\text{Ad}(h_+^t)(w)$ стремится к бесконечности, кроме случая, когда $a = b = 0$, т. е. когда w направлен в сторону h_+ . Пусть t_0 — первый положительный момент времени, когда $\|\text{Ad}(h_+^{t_0})(w)\| = C$. Несложно показать, что если ρ достаточно мало, то именно многочлен второй степени первым достигает значения $\pm C$, т. е. $|-t_0^2 b + 2at_0 + c| = C$, и, кроме того, $|-a + t_0 b|$ (и $|b|$) меньше ε . Иначе говоря, $\|\text{Ad}(h_+^t)(w) \pm Ch_+\| < \varepsilon$ при $t = t_0$. Можно даже предположить, что если ρ всегда брать меньшим, чем некоторое действительное число ρ_0 (зависящее только от ε и C), то последнее неравенство остается справедливым для $t \in [t_0, (1+\alpha)t_0]$, где $\alpha > 0$ зависит только от ε и C .

Предположим теперь, что Γ — кокомпактная решетка в $G = \text{SL}(2, \mathbb{R})$. Выбранная на $\mathfrak{sl}(2, \mathbb{R})$ норма снабжает $\Gamma \backslash G$ метрикой. Предыдущие рассуждения показывают, что для достаточно малого C справедливо следующее утверждение.

Для любого $\varepsilon > 0$ существуют $\rho_0 > 0$ и $\alpha > 0$, такие, что если две точки x, y пространства $\Gamma \backslash G$ находятся на расстоянии, меньшем, чем ρ_0 , и не удовлетворяют равенству $y = x \cdot h_+^t$ при $|t| < 2\rho_0$, то существует $t_0 > 0$, такое, что расстояние между $x \cdot h_+^t$ и $y \cdot h_+^{t+C}$ или между $x \cdot h_+^t$ и $y \cdot h_+^{t-C}$ меньше ε при любом t из $[t_0, (1+\alpha)t_0]$ (см. рис. 7).

Читатель догадывается, что речь идет об общем явлении для унитарных потоков ϕ^t . Если w — маленький вектор, а t_0 обозначает первый момент, когда норма элемента $w_t = \text{Ad}(h_+^t)(w)$ достигает ощутимого значения C , то w_{t_0} очень близко к централизатору потока ϕ^t в G и это сохраняется во всем интервале времени $[t_0, (1+\alpha)t_0]$. Это утверждение было введено Ратнер под названием свойства H [66, 67, 68]. Чтобы показать его полезность, мы опишем классификацию Ратнер факторов орициклического потока.

Пусть ψ^t — измеримый поток, сохраняющий вероятность μ на измеримом пространстве $(E, \mathcal{A})$ (изоморфном алгебре борелевских множеств сепарабельного метрического пространства). Говорят, что ψ^t — измеримый фактор орициклического потока на пространстве $\Gamma \backslash G$, если существует измеримое отображение $F : \Gamma \backslash G \rightarrow E$, переводящее меру Хаара λ (нормализованную таким образом, что она является вероятностной) в μ и такое, что $F(x \cdot h_+^t) = F(x) \cdot \psi^t$ для λ -почти любого x и почти любого t . Мы всегда будем предполагать, что F нетривиально, т. е. что F не является постоянным почти всюду.

Пусть $\Gamma \subset \Gamma'$ — две вложенные одна в другую решетки группы $G = \mathrm{SL}(2, \mathbb{R})$. Тогда ясно, что Γ имеет конечный индекс в Γ' , а конечное накрытие $\Gamma \backslash G \rightarrow \Gamma' \backslash G$ показывает, что орициклический поток на $\Gamma' \backslash G$ является фактором орициклического потока на $\Gamma \backslash G$. Отметим тот факт, что для решетки Γ «общего положения» не существует решетки Γ' , строго содержащей Γ .

Теорема. Если (E, ψ^t) — измеримый фактор орициклического потока на пространстве $\Gamma \backslash G$, то (E, ψ^t) изоморфно орициклическому потоку на пространстве $\Gamma' \backslash G$, где $\Gamma \subset \Gamma'$.

Набросок доказательства. Сущность теоремы состоит в том, что если $F : \Gamma \backslash G \rightarrow E$ — измеримый фактор, то почти все слои отображения F конечны (и, следовательно, имеют одну и ту же мощность в силу эргодичности). Заметим, что слои отображения F разбивают $\Gamma \backslash G$ и переставляются потоком h_+^t . Слой, который глобально сохраняется потоком h_+^t , порождает периодическую орбиту потока ψ^t с периодом τ . Так как поток h_+^t эргодический, то это верно и для ψ^t . Следовательно, объединение периодических орбит потока ψ^t имеет нулевую μ -меру, если только μ не сосредоточена на единственной периодической орбите (которая не является неподвижной точкой, так как F непостоянна). В двух случаях можно утверждать, что для λ -почти любого x из $\Gamma \backslash G$ точки $F(x)$ и $F(x \cdot h_+^t)$ различны для достаточно малых t .

Рассмотрим сначала случай, когда $\Gamma \backslash G$ компактно, E — метрическое пространство и $F : \Gamma \backslash G \rightarrow E$ непрерывно.

Если слой $F^{-1}(e)$ бесконечен, то он содержит две сколь угодно близкие различные точки x, y . Тогда к ним можно применить свойство H , для того чтобы убедиться, что y не имеет вид $x \cdot h_+^t$ при малом t . Следовательно, существует $t_0 > 0$, такое, что если $t \in [t_0, (1 + \alpha)t_0]$, то точки $x \cdot h_+^t$ и $y \cdot h_+^{t \pm C}$ очень близки. Тогда от-

сюда следует в силу непрерывности F , что точки $F(x \cdot h_+^t) = e \cdot \psi^t$ и $F(y \cdot h_+^{t \pm C}) = e \cdot \psi^{t \pm C}$ являются близкими при $t \in [t_0, (1 + \alpha)t_0]$. Беря в $F^{-1}(e)$ пары все более и более близких различных точек x, y , находим куски орбиты точки e вида $\{e \cdot \psi^t \mid t \in [t_0, (1 + \alpha)t_0]\}$, где t_0 постоянно увеличивается, состоящие из точек, все более и более близких к неподвижным относительно $\psi^{\pm C}$.

Выберем такое C , что μ -мера множества неподвижных точек для $\psi^{\pm C}$ равна нулю. Следовательно, μ -мера множества почти неподвижных точек для $\psi^{\pm C}$ мала. Тогда эргодическая теорема показывает, что для μ -почти любого e плотность точек, почти неподвижных относительно $\psi^{\pm C}$, на части орбиты $\{e \cdot \psi^t \mid t \in [0, T]\}$, где T велико, мала. Следовательно, для μ -почти любого e слой $F^{-1}(e)$ не может быть бесконечным, что и доказывает теорему.

В случае когда $F : \Gamma \backslash G \rightarrow E$ не предполагается непрерывным, используем теорему Лузина, которая позволяет найти компакт K пространства $\Gamma \backslash G$ с большой мерой, на котором F непрерывна. Эргодическая теорема показывает, что орбита общего положения потока h_+^t «находится большую часть времени» в K , что позволяет приспособить доказательство к этому случаю.

Если $\Gamma \backslash G$ не предполагается компактным, то используется теорема, цитированная в п. 2.4, согласно которой орбита группы h_+^t «находится большую часть времени» в некотором компакте пространства $\Gamma \backslash G$. Это позволяет применить свойство H и обобщить предыдущее доказательство. $\square$

В общем случае унипотентного потока ϕ^t на $\Gamma \backslash G$, где G — произвольная связная группа Ли, измеримые факторы были классифицированы Виттом [92] с помощью результатов и методов Ратнер. Получен следующий результат: эти факторы изоморфны действию потока ϕ^t на $L_1 \backslash G / L_2$, где L_1 — некоторая замкнутая подгруппа, содержащая Γ , а L_2 — некоторая замкнутая подгруппа аффинных отображений пространства $L_1 \backslash G$, коммутирующих с ϕ^t .

3.2. Дрейф в направлении нормализатора. Фиксируем дискретную подгруппу Γ группы $G = \text{SL}(2, \mathbb{R})$. Впервые в этом докладе мы не будем предполагать, что Γ — решетка. В предыдущем пункте мы изучали отделимость двух точек $x \cdot h_+^t$ и $y \cdot h_+^t$. Теперь мы изучим отделимость двух орбит в трансверсальном направлении. Обозначим через $h_+^\perp$ подпространство алгебры $\mathfrak{sl}(2, \mathbb{R})$, порожденное h_- и g , а через $p_\perp : \mathfrak{sl}(2, \mathbb{R}) \rightarrow h_+^\perp$ проекцию параллельно h_+ . Снабдим, как и раньше, $\mathfrak{sl}(2, \mathbb{R})$ нормой $\sup$ относительно ба-

зиса h_+ , h_- , g . Существует константа $C_0 > 0$, обладающая следующим свойством. Возьмем x и $y = x \cdot \exp w$, где $\|w\| < C_0 n$, и предположим, что $\|p_\perp \text{Ad}(h_+^t)(w)\| < C_0$ для любого $t \in [0, T]$. Тогда существуют функция $\tau : [0, 2T] \rightarrow \mathbb{R}$, где $\tau(0) \leq 2C_0$ и $0,99 \leq \frac{d\tau}{dt} \leq 1,01$, и кривая $t \in [0, 2T] \mapsto w_t \in h_+^\perp$, где $\|w_0\| < 2C_0$, такие, что $y \cdot h_+^t = x \cdot h_+^{\tau(t)} \exp w_t$. Если $\exp w = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$, то элементарное вычисление показывает, что $\exp w_t = \begin{pmatrix} (bt+d)^{-1} & 0 \\ b & bt+d \end{pmatrix}$. Если $\|w\|$ мало, т. е. если малы $|b|$ и $|d-1|$, то w_t очень близко к $bh_- - (bt + (d-1))g$. Выберем достаточно маленькое C_0 , такое, что если v_1 и v_2 — два вектора подпространства $h_+^\perp$ с нормой, меньшей, чем $2C_0$, и $\exp v_i = \begin{pmatrix} a_i & c_i \\ b_i & d_i \end{pmatrix}$ ($i = 1, 2$), то $0,99\|v_1 - v_2\| \leq \sup(|a_1 - a_2|, |b_1 - b_2|, |c_1 - c_2|, |d_1 - d_2|) \leq 1,01\|v_1 - v_2\|$.

Фиксируем константу $C < C_0$ и вектор w , где $\|w\| = \rho$ мало по сравнению с C . Формула, задающая w_t , показывает теперь, что w_t постоянно при $b = 0$. Это неудивительно и соответствует тому факту, что геодезический поток g^s нормирует h_+^t . Напомним, что мы обозначаем через B подгруппу группы G , порожденную h_+^t и g^s . Если $b \neq 0$, то пусть t_0 — первый положительный момент, когда $\|w_{t_0}\| = C$. Это время t_0 имеет порядок $\frac{C}{|b|}$, и $w_{t_0} - Cg$ или $w_{t_0} + Cg$ имеют норму, меньшую, чем 2ρ . Кроме того, $\|w_t - w_{t_0}\| \leq 2\eta C$, если $t \in [(1-\eta)t_0, (1+\eta)t_0]$.

Следующее утверждение представлено в более неопределенных терминах. Если две точки x, y очень близки и не удовлетворяют равенству $y = x \cdot g$, где g — малый элемент группы B , то их орбиты должны трансверсально отделяться. В этом случае, если в момент времени t_0 такое разделение w_{t_0} имеет ощутимую норму C , то точка $y \cdot h_+^{t_0}$ очень близка к точке вида $z = (x \cdot h_+^\tau)g^{\pm C}$. Наконец, орбиты точек $y \cdot h_+^{t_0}$ и z остаются на расстоянии C в течение линейного промежутка времени, зависящего от t_0 .

Здесь мы опять приводим формулировку, обобщающуюся на любой унипотентный поток. Тогда «трансверсальный дрейф» между двумя орбитами происходит в направлении *нормализатора* потока. Это свойство R , одна из основных идей работ Ратнер. Мы проиллюстрируем его использование в теореме, которую уточним далее и доказательство которой является частным случаем из [71].

Теорема. Пусть Γ — дискретная подгруппа группы $SL(2; \mathbb{R})$ (которая не обязательно является решеткой). Пусть μ — веро-

ятностная мера на $\Gamma \backslash \text{SL}(2, \mathbb{R})$, инвариантная и эргодическая относительно потока h_+^t . Если μ не сосредоточена на периодической орбите, то она также инвариантна относительно геодезического потока g^s .

Доказательство. Пусть $C > 0$ мало, и предположим, что g^C не сохраняет μ . Поскольку g^C нормализует поток h_+^t , вероятность $g_*^C \mu$ инвариантна также относительно h_+^t , отлична от μ и эргодична. Отсюда следует, что μ и $g_*^C \mu$ взаимно просты, т. е. существует борелевское множество X , такое, что $\mu(X) = 1$ и $X \cdot g^C \cap X = \emptyset$. Пусть $K \subset X$ — компакт, имеющий μ -меру, большую, чем $1 - \varepsilon$ (при произвольно малом $\varepsilon > 0$). Два компакта K и $K \cdot g^C$ не пересекаются, и, следовательно, существует $\delta > 0$, такое, что если $x = y \cdot \exp w$ при $\|w \pm Cg\| \leq \delta$, то одновременно не могут иметь место включения $x \in K$ и $y \in K$. Выберем η достаточно малым для выполнения неравенства $2\eta C \leq \delta/2$.

Для μ -почти любой точки ее орбита находится в K большую, чем $1 - \varepsilon$, часть времени. Тогда можно найти множество $A_1 \subset \Gamma \backslash G$, имеющее μ -меру, большую, чем $1 - \eta/10$, и $T_1 > 0$, такие, что если $x \in A_1$ и $T > T_1$, то

$$\frac{1}{T} m(\{t \in [0, T] \mid x \cdot h_+^t \in K\}) \geq 1 - 2\varepsilon$$

(где через m обозначается, как обычно, мера Лебега).

Затем можно найти $A \subset \Gamma \backslash G$, имеющее меру, произвольно близкую к 1, и $T_2 > 0$, такие, что если $x \in A$ и $T > T_2$, то

$$\frac{1}{T} m(\{t \in [0, T] \mid x \cdot h_+^t \in A_1\}) \geq 1 - \frac{\eta}{5}.$$

Рассмотрим теперь две точки x и $y = x \cdot \exp w$ из A , где $\|r - w\| = \rho$. Сначала мы покажем, что если ρ достаточно мало, то $\exp r - w$ принадлежит B , т. е. что $y = x \cdot h_+^t g^s$, где s и t малы. Предположим противное, и пусть t_0 — первый положительный момент, когда $\|w_{t_0}\| = C$. Имеем $t_0 \geq 2T_1$ и $t_0 \geq 2T_2/\eta$ при достаточно малом ρ .

Согласно определению A , интервал $[0, t_0]$ содержит больше, чем $(1 - \eta/5)$ -ю часть, действительных чисел t , таких, что $x \cdot h_+^t \in A_1$. Пусть τ — это введенная выше функция, такая, что $y \cdot h_+^t = x \cdot h_+^{\tau(t)} \exp w_t$. Поскольку $0,99 \leq \frac{d\tau}{dt} \leq 1,01$, то интервал $[0, t_0]$

содержит большую, чем $1 - 1,01\frac{\eta}{5} \geq 1 - \frac{\eta}{4}$, часть действительных чисел t , таких, что $x \cdot h_+^{\tau(t)} \in A_1$. Следовательно, существует $t_1 \in [(1 - \eta)t_0, t_0]$, такое, что одновременно $x \cdot h_+^{t_1} \in A_1$ и $x \cdot h_+^{\tau(t_1)} \in A_1$.

В свою очередь, определение A_1 показывает, что существует $t_2 \in [(1 - \eta)t_0, (1 + \eta)t_0]$, такое, что как $\bar{x} = x \cdot h_+^{t_2}$, так и $\bar{y} = y \cdot h_+^{\tau(t_2)}$ принадлежат K . Вместе с тем мы знаем, что если $t \in [(1 - \eta)t_0, (1 + \eta)t_0]$, то $\|w_t - w_{t_0}\| \leq 2\eta C \leq \delta/2$ и $\|w_{t_0} \pm Cg\| \leq 2\rho$. Значит, если $\rho < \frac{\delta}{4}$, то $\|w_{t_2} \pm Cg\| \leq \delta$.

Но это противоречит определению δ , в соответствии с которым $\|w_{t_2} \pm Cg\| > \delta$, если $\bar{x}$ и $\bar{y} = \exp w_{t_2}$ принадлежат K . Следовательно, мы установили, что если точки x и y близки в A , то $y = x \cdot \exp w$, где $\exp w \in B$ близко к тождественному отображению. В частности, одна из орбит группы B имеет ненулевую μ -меру, а эргодичность фактически показывает, что μ сосредоточена на некоторой орбите группы B в $\Gamma \backslash G$.

Остается показать, что μ сосредоточена на периодической орбите потока h_+^t . Орбита $x \cdot B$ имеет вид B/Λ , где Λ — дискретная подгруппа группы B . Эти дискретные подгруппы группы B легко описать. Возможны три случая:

(i) Λ тривиальна. Поток h_+^t , действующий на B , не может сохранять вероятность, так как все его орбиты стремятся к бесконечности в B .

(ii) Λ порождается матрицами $\begin{pmatrix} \exp(s_0) & t_0 \\ 0 & \exp(s_0) \end{pmatrix}$, где $s_0 \neq 0$. Однородное пространство B/Λ является гиперболическим цилиндром. Здесь опять все орбиты потока h_+^t в B/Λ стремятся к бесконечности, и, следовательно, не существует никаких инвариантных вероятностей.

(iii) Λ порождается матрицами $\begin{pmatrix} 1 & t_0 \\ 0 & 1 \end{pmatrix}$, где $t_0 \neq 0$. Однородное пространство B/Λ является каспом, где все орбиты потока h_+^t периодические. Следовательно, инвариантная и эргодическая вероятность сосредоточена на некоторой периодической орбите.

Этим доказательство теоремы завершается. $\square$

3.3. Классификация инвариантных вероятностей: случай решеток. Следующая теорема принадлежит Дани [13, 16, 17], но доказательство, которое мы представляем, подсказано более общим результатом Ратнер [71].

Теорема. Пусть Γ — решетка группы $G = \mathrm{SL}(2, \mathbf{R})$. Тогда (с точностью до множителя) единственными конечными мерами на $\Gamma \backslash G$, инвариантными и эргодическими относительно орициклического потока, являются мера Хаара на $\Gamma \backslash G$ и меры, концентрирующиеся на периодических орбитах.

Набросок доказательства. Сначала предположим, что $\Gamma \backslash G$ компактно. Пусть μ — вероятностная мера на $\Gamma \backslash G$, инвариантная и эргодическая относительно h_+^t . Пусть λ — мера Хаара, нормализованная так, что она является вероятностной на $\Gamma \backslash G$. Если τ_+ , τ_- и σ — три действительных положительных числа, то положим $Q_{\tau_+, \tau_-, \sigma} = \{h_+^{t_+} h_-^{t_-} g^s \in \mathrm{SL}(2, \mathbf{R}) \mid |t_+| \leq \tau_+, |t_-| \leq \tau_-, |s| \leq \sigma\}$. Компактность пространства $\Gamma \backslash G$ позволяет гарантировать существование такого $\rho > 0$, что если τ_+ , τ_- и σ меньше ρ , то множество $x \cdot Q_{\tau_+, \tau_-, \sigma}$ инъективно вложено в $\Gamma \backslash G$ для любого x из $\Gamma \backslash G$. Для простоты записи предположим, что $\rho = 2$, так что вся дуга орбиты $\{x \cdot h_+^t \mid t \in [0, 1]\}$ содержится во внутренней части множества $x \cdot Q_{2, 2, 2}$.

Пусть $Q_0 = x_0 \cdot Q_{\tau_+, \tau_-, \sigma}$ — одно из таких множеств (τ_+ , τ_- и $\sigma \leq 2$). Мы покажем, что $\mu(Q_0) = \lambda(Q_0)$, откуда будет следовать, что $\lambda = \mu$.

Поскольку поток h_+^t эргодический относительно λ , для λ -почти любого x имеем

$$\lim_{T \rightarrow \infty} \frac{1}{T} m(\{t \in [0, T] \mid x \cdot h_+^t \in Q_0\}) = \lambda(Q_0).$$

Эта сходимостъ равномерна на множестве почти полной меры. Следовательно, существуют $X \subset \Gamma \backslash G$, где $\lambda(X) \geq 1 - \varepsilon$ ($\varepsilon > 0$ произвольно), и T_0 , такие, что если $x \in X$ и $T > T_0$, то

$$(1 - \varepsilon)\lambda(Q_0) \leq \frac{1}{T} m(\{t \in [0, T] \mid x \cdot h_+^t \in Q_0\}) \leq (1 + \varepsilon)\lambda(Q_0).$$

Рассмотрим теперь точку y общего положения для другой меры μ , т. е. такую, что

$$\lim_{T \rightarrow \infty} \frac{1}{T} m(\{t \in [0, T] \mid y \cdot h_+^t \in Q_0\}) = \mu(Q_0).$$

Для анализа прохождения потока h_+^t в Q_0 возьмем образ относительно геодезического потока $g^{\mathrm{Log} T}$ для большого T и положим $y_T = y \cdot g^{\mathrm{Log} T}$. В результате дуга $\{y \cdot h_+^t \mid t \in [0, T]\}$ сжимается

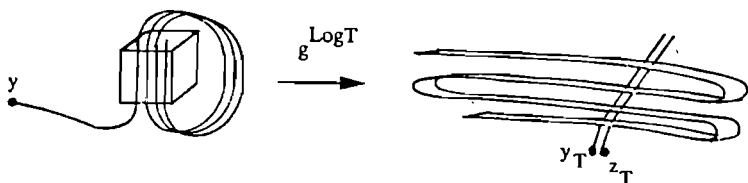


Рис. 8.

до дуги $\{y_T \cdot h_+^t \mid t \in [0, 1]\}$ длины 1 и множество Q_0 переходит в $\{x_0 \cdot g^s h_-^t \mid |s| \leq \sigma, |t_-| \leq T\tau_-, |t_+| \leq \frac{T}{\tau_+}\}$ (см. рис. 8).

Геодезический поток сохраняет λ , и, значит, $\lambda(Xg^{\text{Log } T}) \geq 1 - \varepsilon$. Следовательно, если ε достаточно мало, то маленький шар с центром в y_T не может не пересекать $Xg^{\text{Log } T}$. Пусть тогда точка $z \in X$ такова, что $z_T = z \cdot g^{\text{Log } T}$ и y_T близки. Тогда отрезки орбит $\{y_T \cdot h_+^t \mid t \in [0, 1]\}$ и $\{z_T \cdot h_+^t \mid t \in [0, 1]\}$ очень близки и содержатся в одном и том же множестве $y_T \cdot Q_{2,2,2}$. Рисунок 8 показывает, что время, проведенное этими двумя кусками орбит в $Q_0 \cdot g^{\text{Log } T}$, почти одинаково (и это нетрудно обосновать). По определению точек z и y первый из этих двух отрезков времени близок к $\lambda(Q_0)$, а второй — к $\mu(Q_0)$. Устремляя T к $+\infty$, а ε к 0, устанавливаем, что $\lambda(Q_0) = \mu(Q_0)$. Это завершает доказательство в случае, когда $\Gamma \backslash G$ компактно.

Если $\Gamma \backslash G$ имеет конечный объем, но некомпактно, то существование $\rho (=2!)$ больше не гарантировано. Однако рассмотренное доказательство можно применять, если существует последовательность T_n , стремящаяся к бесконечности, такая, что y_{T_n} остается в некотором компакте пространства $\Gamma \backslash G$. В противном случае $y_T = y \cdot g^{\text{Log } T}$ покидает любой компакт, т. е. соответствующая геодезическая стремится к каспу. Мы видели, что это явление возникает в точности тогда, когда орицикл $y \cdot h_+^t$ периодичен, т. е. когда μ сосредоточена на периодической орбите. $\square$

3.4. Классификация инвариантных вероятностей: случай бесконечного кообъема. Как мы уже указывали, Ратнер получила результаты, годящиеся для дискретных подгрупп общего вида [72], частным случаем которых является следующая

Теорема. Пусть Γ — дискретная подгруппа группы $G = \text{SL}(2, \mathbb{R})$, имеющая бесконечный кообъем. Любая вероятностная мера на

$\Gamma \backslash G$, инвариантная и эргодическая относительно орициклического потока, сосредоточена на некоторой периодической орбите.

Доказательство. Предположим, что существует вероятностная мера μ на $\Gamma \backslash G$, инвариантная и эргодическая относительно h_+^t , не сосредоточенная на периодической орбите. Сейчас мы покажем, что это противоречит тому факту, что объем $\Gamma \backslash G$ бесконечен.

В п. 3.2 мы видели, что μ инвариантна и относительно геодезического потока g^s . Сначала убедимся, что μ также эргодична относительно g^s . Это следствие феномена Маутнера, описанного в п. 1.2: любая функция с μ -интегрируемым квадратом, инвариантная относительно g^s , инвариантна также относительно h_+^t и, следовательно, постоянна μ -почти всюду.

Инвариантность меры μ относительно группы Ли B влечет за собой следующий факт. Если $X \subset \Gamma \backslash G$ имеет полную μ -меру, то для μ -почти любой точки x мера Хаара (на группе B) множества $\{y \in B \mid xg \in X\}$ является полной.

Фиксируем теперь непрерывную положительную функцию $F : \Gamma \backslash G \rightarrow \mathbf{R}$ с компактным носителем, интеграл от которой по мере μ равен 1. Положим для $x \in \Gamma \backslash G$

$$\bar{F}(x) = \liminf_{S \rightarrow \infty} \frac{1}{S} \int_0^S F(x \cdot g^{-s}) ds.$$

Согласно эргодической теореме, $\bar{F}(x) = \int F d\mu = 1$ для x из некоторого множества X μ -меры 1. Впрочем, очевидно, что $\bar{F}$ постоянна на орбитах отрицательного орициклического потока h_-^t . Из предыдущего замечания вытекает, что объединение множеств $X \cdot h_-^t$, где $t \in \mathbf{R}$, является множеством полной λ -меры Хаара в $\Gamma \backslash G$. Таким образом, функция $\bar{F}$ λ -почти всюду постоянна и равна 1. Тогда лемма Фату дает

$$\begin{aligned} \lambda(\Gamma \backslash G) &= \int_{\Gamma \backslash G} 1 d\lambda \leq \liminf \int_{\Gamma \backslash G} \left(\frac{1}{S} \int_0^S F(x \cdot g^{-s}) ds \right) d\lambda(g) \\ &= \int F d\lambda < \infty. \end{aligned}$$

Это — искомое противоречие, завершающее доказательство теоремы. $\square$

3.5. Равномерная распределенность орбит. Следующая теорема в случае решеток сводится к теореме Дани и Смилли.

Теорема. Пусть Γ — дискретная подгруппа группы $G = \text{SL}(2, \mathbf{R})$, x — непериодическая точка для орициклического потока h_+^t и $F: \Gamma \backslash G \rightarrow \mathbf{R}$ — непрерывная функция с компактным носителем. Если T стремится к $+\infty$, то среднее значение $\frac{1}{T} \int_0^T F(x \cdot h_+^t) dt$ сходится

- (i) к $\int_{\Gamma \backslash G} F d\lambda$, если $\Gamma \backslash G$ имеет конечный объем (где λ всегда обозначает меру Хаара, нормированную так, что она является вероятностью).
- (ii) к 0, если $\Gamma \backslash G$ имеет бесконечный объем.

Набросок доказательства. Пусть $\hat{V} = V \cup \{\infty\}$ — компактификация Александера пространства $\Gamma \backslash G$ и $\hat{h}_+^t$ — продолжение потока h_+^t в $\hat{V}$. Пусть μ_T — вероятность на $\hat{V}$, определенная формулой $\int F d\mu_T = \frac{1}{T} \int_0^T F(x \cdot \hat{h}_+^t) dt$. В силу компактности пространства вероятностей на $\hat{V}$ существует слабый предел ν последовательности μ_{T_n} (где T_n стремится к $+\infty$).

Сначала рассмотрим случай, когда Γ — решетка. Тогда речь идет о доказательстве того, что $\nu = \lambda$, так как это влечет за собой слабую сходимость меры μ_T к λ . Мы утверждаем, что $\nu(\{\infty\}) = 0$. Это следует из теоремы (приведенной в п. 2.4), согласно которой орицикл находится малую часть времени вне компакта. Таким образом, ν сосредоточена на $\Gamma \backslash G$ и, очевидно, является вероятностью, инвариантной относительно h_+^t .

В п. 3.3 мы классифицировали инвариантные эргодические вероятности. Чтобы закончить доказательство в случае решеток, нужно доказать, что ν -мера объединения периодических орбит равна нулю. Мы даем только набросок этого доказательства, которое сильнее всего использует свойства трансвекций. Пусть γ — периодическая орбита, $A_{\alpha, \beta}$ — кольцо, являющееся объединением орбит $\gamma \cdot g^s$, где $\alpha \leq s \leq \beta$, $B_{\alpha, \beta}^r$ — насыщение $\{y h_-^t \mid y \in A_{\alpha, \beta}, |t| \leq \tau\}$ и, наконец, $S_{\alpha, \beta}^r = \{y_0 \cdot g^s h_-^t, \alpha \leq s \leq \beta, |t| \leq \tau\}$, где y_0 — одна из точек орбиты γ . Отображение первого возвращения потока h_+^t на трансверсаль $S_{\alpha, \beta}^r$ является трансвекцией. Пусть $[\alpha_1, \beta_1]$ и $[\alpha_2, \beta_2]$ — два непересекающихся интервала одинаковой длины в $[\alpha, \beta]$. Очевидно, что если τ мало, то большой кусок орбиты находится приблизительно одинаковое время в B_{α_1, β_1}^r и B_{α_2, β_2}^r (см. рис. 9). Устремляя τ к 0, получим, что $\nu(A_{\alpha_1, \beta_1}) = \nu(A_{\alpha_2, \beta_2})$. Следовательно, все кольца $A_{n, n+1}$ ($n \in \mathbf{Z}$) имеют одну и ту же ν -меру и попарно

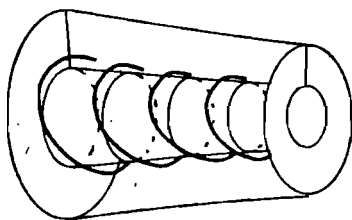


Рис. 9.

не пересекаются. Поскольку ν имеет конечную полную массу, мы получим, что ν -мера объединения периодических орбит потока h_t^+ равна нулю.

Когда объем пространства $\Gamma \backslash G$ конечен, речь идет, напротив, о доказательстве того, что $\nu(\{\infty\}) = 1$. Предыдущее доказательство показывает, что ν -мера объединения периодических орбит равна нулю. Классификация инвариантных вероятностей, данная в п. 3.4, показывает, что ν обязательно сосредоточена в точке ∞ . $\square$

3.6. Некоторые указания для общего случая. Классификация инвариантных мер на однородном пространстве конечного объема $\Gamma \backslash G$ с помощью унитарной подгруппы H использует все методы, описанные выше. Как указано во введении, мы не можем привести здесь детали ввиду сложности, которую представляет это доказательство. Мы ограничимся некоторыми замечаниями.

Множество элементов g из G , правые сдвиги на которые сохраняют μ , является замкнутой подгруппой Λ группы G , и надо доказать, что μ сосредоточена на орбите подгруппы Λ .

Если G разрешима, то $\Lambda \cap [G, G]$ унитарна. Это следует из применения к этой подгруппе свойства R , которое Ратнер использовала для доказательства своей теоремы в разрешимом случае [71].

Если G полупроста и ϕ^t — унитарная однопараметрическая подгруппа, то можно вложить экземпляр алгебры $\mathfrak{sl}(2, \mathbf{R})$ в алгебру Ли $\mathfrak{g}$, отображая орициклический поток на ϕ^t . Таким образом, мы располагаем «геодезическим потоком» g^s в G , что позволяет обобщить большую часть техники, описанной выше. Конечно, ситуация усложняется тем, что g^s не поток типа Аносова, так как он может обладать нетривиальным централизатором [72–73].

Для вывода своего результата о равномерной распределенности и о замыканиях орбит Ратнер использовала метод типа приведенного в п. 3.5, но в существенно более развитом виде [74].

4. ТРИ ПРИЛОЖЕНИЯ

В этом разделе мы приводим три примера применения результатов Ратнер.

4.1. Жесткость унипотентных преобразований. Пусть ϕ^t — унипотентная однопараметрическая подгруппа простой группы Ли G и Γ — решетка группы G . Классические эргодические инварианты потока ϕ^t на пространстве $\Gamma \backslash G$ (оснащенном мерой Хаара) не очень интересны: нулевая энтропия и счетный спектр Лебега. Примечательно, что теоремы Ратнер позволяют классифицировать все измеримые изоморфизмы между этими потоками. Следующая теорема принадлежит Д. Витту [89]; она обобщает частный случай группы $SL(2, \mathbb{R})$, рассмотренный Ратнер [67, 68]. Для простоты мы ограничимся случаем простых групп Ли, отсылая к [89] и [73] за общими формулировками.

Теорема. Пусть ϕ_i^t — унипотентная однопараметрическая подгруппа односвязной группы Ли G_i ($i = 1, 2$) и Γ_i — решетка в G_i . Пусть $F : \Gamma_1 \backslash G_1 \rightarrow \Gamma_2 \backslash G_2$ — измеримое сопряжение между ϕ_1^t и ϕ_2^t , отображающее меру Хаара λ_1 на λ_2 . Тогда существует изоморфизм α группы G_1 на G_2 , отображающий Γ_1 на Γ_2 и ϕ_1^t на ϕ_2^t , такой, что $F(\Gamma_1 \cdot g_1) = \Gamma_2 \cdot \alpha(g_2)$ для λ_1 -почти любого g_1 .

На самом деле можно сказать значительно больше. Джойнинг (joining) пары ϕ_1^t, ϕ_2^t — это по определению вероятностная мера μ на произведении $\Gamma_1 \backslash G_1 \times \Gamma_2 \backslash G_2$, инвариантная относительно диагонального потока $(\phi_1^t \times \phi_2^t)$ и такая, что $\mu(\pi_i^{-1}(X)) = \lambda_i(X)$, где π_i — проекция на i -й сомножитель, а X — борелевское множество пространства $\Gamma_i \backslash G_i$ ($i = 1, 2$). Произведение двух мер Хаара $\lambda_1 \otimes \lambda_2$ называется тривиальным джойнингом. Если $F : \Gamma_1 \backslash G_1 \rightarrow \Gamma_2 \backslash G_2$ — измеримое сопряжение между ϕ_1^t и ϕ_2^t , отображающее λ_1 на λ_2 , то ясно, что образ меры λ_1 при отображении $\text{id} \times F : \Gamma_2 \backslash G_2 \rightarrow \Gamma_1 \backslash G_1 \times \Gamma_2 \backslash G_2$ является нетривиальным джойнингом. Нетрудно вывести отсюда, что предыдущая теорема является лишь частным случаем следующей теоремы (мы ограничимся случаем простых групп).

Теорема. Сохраним обозначения предыдущей теоремы. Пусть μ — нетривиальный эргодический джойнинг пары ϕ_1^t, ϕ_2^t . Тогда существуют изоморфизм $\alpha : G_1 \rightarrow G_2$ и решетка Γ группы G_1 , содержащаяся (с конечным индексом) в Γ_1 и $\alpha^{-1}(\Gamma_2)$, такие, что μ — образ меры Хаара на пространстве $\Gamma \backslash G_1$ при отображении $\text{id} \times \alpha : \Gamma \backslash G_1 \rightarrow \Gamma_1 \backslash G_1 \times \Gamma_2 \backslash G_2$.

Доказательство. Согласно теореме Ратнер, эргодический джойнинг является однородной мерой на $\Gamma_1 \times \Gamma_2 \backslash G_1 \times G_2$, полученной из меры Хаара на некоторой замкнутой подгруппе H группы $G_1 \times G_2$. Так как речь идет о джойнинге, то проекция подгруппы H на каждый сомножитель G_1 и G_2 сюръективна. В силу простоты G_1 и G_2 можно утверждать, что H является графиком изоморфизма между G_1 и G_2 , если только $H = G_1 \times G_2$. Конечно, последний случай исключается, если μ нетривиальна. $\square$

Отметим два обобщения этих теорем. Прежде всего, измеримые сопряжения между орициклическими потоками в случае переменной отрицательной кривизны были классифицированы в статье [33] (которая дополняется работой [61]). С другой стороны, в статье [47] (также дополняемой работой [61]) изучается случай топологических эквивалентностей между орициклическими потоками, т. е. гомеоморфизмов, отображающих орбиты на орбиты без учета параметризаций. За аналогичными результатами в высших размерностях мы отсылаем к [34] и [91].

4.2. Значения квадратичных форм в целочисленных точках. Приведенная во введении теорема о диофантовых приближениях была сформулирована в качестве гипотезы в 1929 г. Оппенгеймом для $n \geq 5$ [57, 58], а затем в 1946 г. Давенпортом для $n \geq 3$. Для $n \geq 21$ она была доказана Бёрчем, Давенпортом и Риду. В [3, 29, 43, 58, 59, 87, 88] можно найти другие предварительные результаты, все они получены методами, отличными от тех, которые мы опишем. Мы покажем здесь, каким образом результаты Ратнер позволяют получить более сильное утверждение. Вывод этой более сильной формулировки из гипотезы Рагунатана принадлежит Дани и Маргулису. Впрочем, последние доказывают в [24–26] частный случай $n = 3$ следующей теоремы. По-видимому, гипотеза Оппенгейма была одной из мотивировок, которые привели Рагунатана к его гипотезе.

Теорема. Пусть Q — невырожденная неопределенная квадратичная форма на $\mathbb{R}^n$ ($n \geq 3$) и $f(x, y) = \frac{1}{4}(Q(x + y) - Q(x - y))$ — соответствующая билинейная форма. Предположим, что Q не кратна никакой рациональной форме. Пусть α_{ij} , $1 \leq i \leq n - 1$, $1 \leq j \leq n - 1$, — действительные числа, выбранные так, что существуют векторы $w_1, \dots, w_{n-1}$ в $\mathbb{R}^n$, такие, что $\alpha_{ij} = f(w_i, w_j)$. Тогда для любого $\varepsilon > 0$ существуют $v_1, \dots, v_{n-1}$ из $\mathbb{Z}^n$, такие, что $|f(v_i, v_j) - \alpha_{ij}| < \varepsilon$ для любых i, j . Можно также считать, что

набор $(v_1, \dots, v_{n-1})$ примитивен в том смысле, что его можно дополнить до базиса группы $\mathbf{Z}^n$.

Начнем с выражения диофантовых задач в терминах динамики на однородных пространствах.

Фиксируем два целых положительных числа p, q , такие, что $p+q=n$, и обозначим через $\Phi_{p,q}$ множество квадратичных форм на $\mathbf{R}^n$ сигнатуры (p, q) . Пусть Q_0 — форма $x_1^2 + \dots + x_p^2 - x_{p+1}^2 - \dots - x_n^2$ и $O(p, q)$ — ее ортогональная группа. Для простоты записи предположим, что n — такое нечетно число, что любая невырожденная форма кратна единственной форме с дискриминантом $+1$. Группа $GL(n, \mathbf{R})$ транзитивно действует справа на $\Phi_{p,q}$ по формуле $Q \cdot g(w) = Q(g^{-1}(w))$, и, следовательно, $\Phi_{p,q}$ отождествляется с $GL(n, \mathbf{R})/O(p, q)$.

Две формы Q и Q' из $\Phi_{p,q}$ эквивалентны, если существуют $c \in \mathbf{R}$ и $g \in SL(n, \mathbf{Z})$, такие, что $Q'(w) = c^2 Q(g(w))$ для любого w из $\mathbf{R}^n$. Ясно, что Q и Q' принимают, с точностью до множителя, одни и те же значения на $\mathbf{Z}^n$.

Таким образом, мы получаем естественную биекцию между

(1) классами эквивалентности форм из $\Phi_{p,q}$ и

(2) смежными классами группы $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R})/SO(p, q)$

или, эквивалентным образом, орбитами группы $SO(p, q)$ на $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R})$.

Теорема Ратнер позволит описать замыкания орбит группы $SO(p, q)$. Достаточно сформулировать эти результаты в терминах квадратичных форм.

Если $n \geq 3$ и $p, q \geq 1$, то группа $SO(p, q)$ содержит группу $SO(2, 1)$, локально изоморфную $SL(2, \mathbf{R})$, и, значит, содержит унипотентные элементы. Эти унипотентные элементы порождают компоненту единицы $SO_0(p, q)$ группы $SO(p, q)$, и, следовательно, к $SO_0(p, q)$ можно применить теорему Ратнер. Поскольку $SO_0(p, q)$ имеет индекс 2 в $SO(p, q)$, то отсюда легко вытекает, что замыкания орбит группы $SO(p, q)$ в $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R})$ являются орбитами относительно замкнутой подгруппы, лежащей между $SO(p, q)$ и $SL(n, \mathbf{R})$. Следующую лемму мы оставляем в качестве упражнения. Аналогичные, но значительно более общие формулировки можно найти в [30, 31].

Лемма. *Не существует никакой замкнутой подгруппы, строго лежащей между $SO(p, q)$ и $SL(n, \mathbf{R})$.*

Отсюда сразу получается

Предложение. Орбита группы $SO(p, q)$ в $SL(n, \mathbb{Z}) \backslash SL(n, \mathbb{R})$ замкнута или плотна.

Теперь речь пойдет об интерпретации в терминах квадратичных форм понятий плотности и замыкания соответствующих орбит. Это цель двух следующих предложений, которые завершают доказательство теоремы.

Предложение. Пусть Q — квадратичная форма, такая, что связанная с ней орбита плотна. Тогда Q удовлетворяет заключению предыдущей теоремы.

Доказательство. Плотность орбиты сводится к плотности подмножества $SL(n, \mathbb{Z}) \cdot SO(Q)$ в $SL(n, \mathbb{R})$ (где $SO(Q)$ обозначает, конечно, специальную ортогональную группу формы Q). Пусть $(e_1, \dots, e_n)$ — канонический базис пространства $\mathbb{R}^n$ и g — элемент группы $SL(n, \mathbb{R})$, такой, что $g(e_i) = w_i$ для $i = 1, \dots, n-1$. Пусть $h_\alpha \in SO(Q)$ и $\gamma_\alpha \in SL(n, \mathbb{Z})$ — две последовательности, такие, что $\gamma_\alpha h_\alpha$ сходится к g^{-1} при α , стремящемся к $+\infty$. Пусть $v_i^\alpha = \gamma_\alpha^{-1}(e_i)$ ($i = 1, \dots, n-1$). Очевидно, что $(v_1^\alpha, \dots, v_{n-1}^\alpha)$ является набором из $n-1$ примитивных векторов в $\mathbb{Z}^n$ и что

$$f(v_i^\alpha, v_j^\alpha) = f(h_\alpha^{-1} \gamma_\alpha^{-1}(e_i), h_\alpha^{-1} \gamma_\alpha^{-1}(e_j)) \xrightarrow{\alpha \rightarrow +\infty} f(g(e_i), g(e_j)) = a_{ij}. \quad \square$$

Второе предложение утверждает больше, чем нам необходимо. Мы предположим теперь, что p и q не равны нулю.

Предложение. Орбита группы $SO(p, q)$ замкнута тогда и только тогда, когда соответствующая квадратичная форма (точнее, ее класс эквивалентности) кратна некоторой рациональной форме. Она компактна тогда и только тогда, когда кратна некоторой рациональной форме, не представляющей нуль (в $\mathbb{Q}^n$).

Доказательство. Одно из направлений (которое мы не использовали) — частный случай классической теоремы Бореля и Харпш-Чандры [4]. Если Q — рациональная форма, то $SL(n, \mathbb{Z}) \backslash SO(Q)$ замкнуто в $SL(n, \mathbb{Z}) \backslash SL(n, \mathbb{R})$ и компактно, если Q не представляет нуль.

Пусть Q такова, что $SL(n, \mathbb{Z}) \cdot SO(Q)$ замкнуто в $SL(n, \mathbb{R})$. Мы покажем, что Q кратна некоторой рациональной форме, следуя [27], где часть доказательства приписывается Борелю. Пусть $\Gamma = SL(n, \mathbb{Z}) \cap SO(Q)$. Метод состоит в доказательстве того, что Q

и ее кратные — единственные формы, ортогональная группа которых содержит Γ . Отсюда будет следовать, что прямая, порожденная формой Q в пространстве квадратичных форм, задается целочисленным уравнением: $\{Q' | Q \cdot \gamma = Q' \text{ для любого } \gamma \text{ из } \Gamma\}$, и, следовательно, Q кратна некоторой рациональной форме. Фиксируем теперь форму Q' , такую, что $\Gamma \subset SO(Q')$. Пусть ϕ^t — однопараметрическая унитарная подгруппа группы $SO_0(Q)$ (где $SO_0(Q)$ — компонента единицы группы $SO(Q)$). Если мы покажем, что ϕ^t образована изометриями формы Q' , мы выведем отсюда, что $SO_0(Q) \subset SO(Q')$, так как эти подгруппы порождают $SO_0(Q)$. Непосредственно из этого будет следовать, что $SO(Q) = SO(Q')$ и что Q и Q' пропорциональны.

Пусть $w \in \mathbb{R}^n$. Функция $F : g \in SO(Q) \mapsto Q'(g(w)) \in \mathbb{R}$ после факторизации дает функцию $\Gamma \backslash SO(Q) \rightarrow \mathbb{R}$. Мы уже приводили теорему Маргулиса, в соответствии с которой орбита группы ϕ^t не может стремиться к бесконечности в $SL(n, \mathbb{Z}) \backslash SL(n, \mathbb{R})$. Значит, существует последовательность t_n , стремящаяся к бесконечности, такая, что $SL(n, \mathbb{Z}) \cdot \phi^{t_n}$ остается в некотором компакте K группы $SL(n, \mathbb{Z}) \backslash SL(n, \mathbb{R})$. Поскольку мы предположили, что орбита формы Q , изоморфная $\Gamma \backslash SO(Q)$, замкнута в $SL(n, \mathbb{Z}) \backslash SL(n, \mathbb{R})$, то ее пересечение с компактом K является компактом. Значит, функция $t \mapsto F(\phi^t) = Q'(\phi^t(w))$ принимает ограниченные значения в точках последовательности t_n . Так как, с другой стороны, эта функция полиномиальна, она постоянна. Таким образом, $Q'(\phi^t(w)) = Q'(w)$ при любом t , т. е. ϕ^t содержится в $SO(Q')$, что мы и утверждали. Этим завершается доказательство предложения и, следовательно, теоремы. $\square$

В случае $n = 2$ теорема неприменима, так как $SO(1, 1) \subset SL(2, \mathbb{R})$ содержит только полупростые элементы. Динамика группы $SO(1, 1)$ на $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$ является динамикой геодезического потока на модулярном орбиобразии, для которого, как мы уже говорили, существуют орбиты, которые не являются ни периодическими, ни плотными. Конкретно, пусть θ — число постоянного типа, т. е. удовлетворяющее следующим неравенствам для некоторого $C > 0$:

$$\left| \theta - \frac{y}{x} \right| \geq Cx^{-2} \text{ для любых } (x, y) \in \mathbb{Z}^2 - (0, 0).$$

Например, квадратичные иррациональные числа над $\mathbb{Q}$ относятся к постоянному типу. Неопределенная и невырожденная квадратичная форма Q на $\mathbb{R}^2$, заданная формулой $Q(x, y) = y^2 - \theta^2 x^2$, не

кратна никакой рациональной форме, если θ^2 иррационально. Очевидно, однако, что $Q(x, y) = x^2(y/x - \theta)(y/x + \theta)$ не может быть произвольно малым, если x и y — ненулевые целые числа.

Мы следуем теперь [54], чтобы более точно описать значения неопределенных и невырожденных квадратичных форм Q в целочисленных точках из $\mathbb{Z}^n$. Для такой формы положим $\nu(Q) = \inf\{|Q(v)| \mid v \in \mathbb{Z}^n - \{0\}\}$. Пусть $\text{disc}(Q)$ — дискриминант формы Q и $M(Q) = \nu(Q)|\text{disc}(Q)|^{-1/n}$, так что $M(Q)$ инвариантно относительно эквивалентности форм. Пусть M_n — множество значений, принимаемых $M(Q)$, когда Q пробегает невырожденные неопределенные формы на $\mathbb{R}^n$. Это компакт на прямой $\mathbb{R}$ для любого $n \geq 2$.

Изучение множества M_2 — увлекательное занятие. В частности, см. [79] и главу из [10], посвященную пересечению M_2 с $]4/9, 4/5]$. Известно также, что M_2 содержит интервал $[0, \varepsilon]$ (см. [10, 78]).

Из теоремы Маргулиса следует, что если Q не кратна рациональной форме, то $M(Q) = 0$. Если Q рациональна и $n \geq 5$, то знаменитая теорема Мейера утверждает, что Q нетривиально представляет нуль и, следовательно, $M(Q) = 0$ [11]. Таким образом, $M_n = \{0\}$, если $n \geq 5$. Следующая теорема была доказана Касселем и Свиннертоном-Дайером в предположении справедливости гипотезы Оппенгейма [12] и, в более сильной форме, Вулахом [86]. Здесь мы представляем доказательство, несколько отличающееся от доказательства Маргулиса [54].

Теорема. *Если $n = 3$ или 4 и $\varepsilon > 0$, то существует только конечное число классов эквивалентности невырожденных и неопределенных квадратичных форм Q на $\mathbb{R}^n$, таких, что $M(Q) \geq \varepsilon$.*

Набросок доказательства. Мы знаем, что неопределенная форма Q , такая, что $M(Q) > 0$, кратна рациональной форме, не представляющей 0, и, следовательно, соответствует компактной орбите относительно действия группы $\text{SO}(p, q)$ на $\text{SL}(n, \mathbb{Z}) \backslash \text{SL}(n, \mathbb{R})$. Предположим противное, т. е. что существует бесконечное множество попарно неэквивалентных форм Q_i с одинаковой сигнатурой (p, q) , таких, что $M(Q_i) \geq \varepsilon$. Таким образом, получаем бесконечное число компактных связных орбит O_i для действия компоненты единицы $\text{SO}_0(p, q)$ группы $\text{SO}(p, q)$ (имеющей индекс 2). Проверим (используя критерий Малера), что условие $M(Q_i) \geq \varepsilon$ означает, что все компакты O_i содержатся в одном и том же компакте пространства $\text{SL}(n, \mathbb{Z}) \backslash \text{SL}(n, \mathbb{R})$. Пусть O_∞ — предел (в смысле Хаусдорфа) компактов O_i . Очевидно, что O_∞ — связный компакт, объединение

орбит группы $SO_0(p, q)$. Поскольку каждая орбита плотна или замкнута, то O_∞ обязательно является объединением счетного числа компактных орбит, так как существует только счетное число рациональных форм. Тогда легко убедиться, что в силу связности множества O_∞ оно сводится к единственной компактной орбите. Мы, следовательно, должны показать, что последовательность различных компактных орбит не может сходиться к компактной орбите O_∞ .

Пусть $\mathcal{N}$ — дополнение алгебры Ли $\mathfrak{so}(p, q)$ в $\mathfrak{sl}(n, \mathbf{R})$, инвариантное относительно присоединенного представления группы $SO_0(p, q)$. Это определяет $SO_0(p, q)$ -эквивариантную тривиализацию нормального расслоения над орбитами группы $SO_0(p, q)$. Если i достаточно велико, то O_i остается в трубчатой окрестности множества O_∞ и, следовательно, существует малая функция $u : O_i \rightarrow \mathcal{N}$, такая, что $O_\infty = \{x \cdot \exp u(x) \mid x \in O_i\}$. Пусть $t \in [0, 1]$. Тогда очевидно, что $O_\infty^t = \{x \cdot \exp tu(x) \mid x \in O_i\}$ — также компактная орбита группы $SO_0(p, q)$. Это противоречит счетности множества компактных орбит и завершает доказательство теоремы. $\square$

4.3. Геодезические ламинации. Пусть H^n — гиперболическое пространство размерности $n \geq 2$. Известно, что полные вполне геодезические подмногообразия изометричны пространству H^p для $0 \leq p \leq n$. Пусть $\text{Rep}(H^n)$ — расслоение положительно ориентированных ортонормированных реперов касательного расслоения к H^n ; это главное однородное пространство относительно действия группы собственных изометрий пространства H^n , изоморфной группе $SO_0(n, 1)$.

Пусть теперь V — полное риманово многообразие кривизны -1 и конечного объема. Тогда оно изометрично фактору пространства H^n по решетке Γ группы $SO_0(n, 1)$. Факторпространство $\Gamma \backslash SO_0(n, 1)$ также отождествляется с расслоением $\text{Rep}(V)$ реперов пространства V . Орбитами действия группы $SO_0(p, 1) \times SO(n-p) \subset SO_0(n, 1)$ на $\Gamma \backslash SO_0(n, 1) \simeq \text{Rep}(V)$ являются подмногообразия, проекции которых в V в точности являются образами изометрических погружений пространства H^p в V . Следовательно, выполняются условия для применения теоремы Ратнер, как только $p > 1$. Мы дадим результат, полученный Шахом при помощи этого метода [80]:

Теорема. Пусть V — полное риманово многообразие размерности n конечного объема и кривизны -1 . Пусть $i : H^p \rightarrow V$ — изометрическое погружение, где $p > 1$. Тогда замыкание множества

$i(H^p)$ в V является вполне геодезическим подмногообразием W многообразия V размерности $\geq p$. Кроме того, ортонормированные p -реперы, касательные к $i(H^p)$, плотны в ортонормированных p -реперах многообразия V , касательных к W .

Напомним, что геодезическая ламинация $\mathcal{L}$ размерности p многообразия V — это замкнутое множество, которое является дизъюнктивным объединением полных и вполне геодезических подмногообразий размерности p , инъективно погруженных в V , называемых листами. Предположим, что p -реперы, касательные к листам ламинации $\mathcal{L}$, образуют замкнутое множество в расслоении p -реперов многообразия V . Ламинация тривиальна, если все ее листы замкнуты (тогда их конечное число). Фундаментальная роль геодезических ламинаций размерности 1 была раскрыта В. Тёрстоном. Она контрастирует со следующим утверждением:

Следствие. Любая геодезическая ламинация размерности $p \geq 2$ гиперболического многообразия V конечного объема тривиальна. В частности, для $1 < p < \dim V$ не существует слоения класса C^0 , все листы которого были бы вполне геодезическими.

В [93] можно найти частичные обобщения на случай переменной кривизны. Вопрос о существовании слоений размерности 1 и класса C^0 , листы которых являются геодезическими многообразия V , остается открытым, хотя доказано, что такие слоения не могут быть класса C^1 .

ЛИТЕРАТУРА

- [1] Аносов Д. В., Геодезические потоки на замкнутых римановых многообразиях отрицательной кривизны. — Труды МИАН им. В. А. Стеклова, т. 90, 1967.
- [2] Auslander L., Green L., Hahn F. Flows on homogeneous spaces, Ann. of Math. Studies, Princeton, **53** (1963). [Имеется перевод: Ауслендер Л., Грин Л., Хан Ф., Потоки на однородных пространствах. — М.: Мир, 1966.]
- [3] Baker R. C., Schlickewei H. P. Indefinite quadratic forms, Proc. London Math. Soc., **54** (1987), No. 3, 385–411.
- [4] Borel A. Introduction aux groupes arithmétiques, Hermann, Paris, 1969.
- [5] Borel A., Prasad G. Valeurs de formes quadratiques aux points entiers, C. R. Acad. Sci. Paris, Série I, **307** (1988), 217–220.
- [6] Bowen R. Periodic orbits for hyperbolic flows, Amer. J. Math., **94** (1972), 1–30.

- [7] Bowen R. Weak mixing and unique ergodicity on homogeneous spaces, *Israel J. Math.*, **23** (1976), 267–273.
- [8] Bowen R., Marcus B. Unique ergodicity of horospherical foliations, *Israel J. Math.*, **26** (1977), 43–67.
- [9] Brezin J., Moore C. Flows on homogeneous spaces: a new look, *Amer. J. Math.*, **103** (1981), 571–613.
- [10] Cassels J. W. S. *An Introduction to Diophantine Approximations*, Cambridge University Press, 1957. [Имеется перевод: Касселс Дж. В. С., Введение в теорию диофантовых приближений. — М.: ИЛ, 1961.]
- [11] Cassels J. W. S. *Rational Quadratic Forms*, Academic Press, London–New York, 1978. [Имеется перевод: Касселс Дж., Рациональные квадратичные формы. — М.: Мир, 1982.]
- [12] Cassels J. W. S., Swinnerton-Dyer H. P. F. On the product of three homogeneous linear forms and indefinite ternary quadratic forms, *Philos. Trans. Roy. Soc. London, Ser. A*, **248** (1955), 73–96.
- [13] Dani S. G. Invariant measures of horospherical flows on non compact homogeneous spaces, *Invent. Math.*, **47** (1978), 101–138.
- [14] Dani S. G. On invariant measures, minimal sets and a lemma of Margulis, *Invent. Math.*, **51** (1979), 239–260.
- [15] Dani S. G. Dynamics of the horocycle flow, *Bull. Amer. Math. Soc.*, **3** (1980), 1037–1039.
- [16] Dani S. G. Invariant measures and minimal sets of horospherical flows, *Invent. Math.*, **64** (1981), 357–385.
- [17] Dani S. G. On uniformly distributed orbits of certain horocyclic flows, *Erg. Theory and Dyn. Syst.*, **2** (1982), 139–158.
- [18] Dani S. G. On orbits of unipotent flows on homogeneous spaces, *Erg. Theory and Dyn. Syst.*, **4** (1984), 25–34.
- [19] Dani S. G. Dynamics of flows on homogeneous spaces: a survey, *Proceed. of Colloquio de sistemas dinamicos (Guanajuato, 1983)*, *Aportaciones Mat.*, **1** (1985), 1–30.
- [20] Dani S. G. Divergent trajectories of flows on homogeneous spaces and diophantine approximation, *J. Reine Angew. Math.*, **359** (1985), 55–89.
- [21] Dani S. G. Orbits of horospherical flows, *Duke Math. J.*, **53** (1986), 177–188.
- [22] Dani S. G. On orbits of unipotent flows on homogeneous spaces II, *Erg. Theory and Dyn. Syst.*, **6** (1986), 167–182.
- [23] Dani S. G. Dense orbits of horospherical flows, dynamical systems and ergodic theory, *Banach center publications*, vol. 23, Polish Scientific Publishers, Warsaw, 1989.
- [24] Dani S. G., Margulis G. A. Values of quadratic forms at primitive integral points, *C.R. Acad. Sci. Paris, Série I*, **308** (1989), 199–203.
- [25] Dani S. G., Margulis G. A. Values of quadratic forms at primitive integral points, *Invent. Math.*, **98** (1989), 405–424.
- [26] Dani S. G., Margulis G. A. Orbit closures of generic unipotent flows on homogeneous spaces of $SL(3, \mathbb{R})$, *Math. Ann.*, **286** (1990), 101–128.

- [27] Dani S. G., Margulis G. A. Values of quadratic forms at integral points: an elementary approach, *L'enseignement mathématique*, **36** (1990), 143–174.
- [28] Dani S. G., Smillie J. Uniform distribution of horocyclic orbits for Fuchsian groups, *Duke Math. J.*, **51** (1984), 185–194.
- [29] Davenport H., Heilbronn H. On indefinite quadratic in five variables, *J. London Math. Soc.*, **21** (1946), 185–193.
- [30] Дынкин Е. В. Полупростые подалгебры полупростых алгебр Ли. — Матем. сб., 1952, т. 30, вып. 2, с. 349–462.
- [31] Дынкин Е. В. Максимальные подгруппы классических групп, Труды ММО, т. 1, 1952, с. 39–166.
- [32] Ellis R., Perrizo W. Unique ergodicity of flows on homogeneous spaces, *Israel J. Math.*, **29** (1978), 276–284.
- [33] Feldman J., Ornstein O. Semirigidity of horocycle flows over compact surfaces of variable negative curvature, *Erg. Theory and Dyn. Syst.*, **7** (1987), 49–72.
- [34] Flaminio L. An extension of Ratner's rigidity theorem to n -dimensional hyperbolic space, *Erg. Theory and Dyn. Syst.*, **7** (1987), 73–92.
- [35] Furstenberg H. Strict ergodicity and transformations of the torus, *Amer. J. Math.*, **83** (1961), 573–601.
- [36] Furstenberg H. The structure of distal flows, *Amer. J. Math.*, **85** (1963), 477–515.
- [37] Furstenberg H. The unique ergodicity of the horocyclic flow, in: *Recent Advances in Topical Dynamics*, Springer-Verlag, 1972, 95–115.
- [38] Hedlund G. Fuchsian groups and transitive horocycles, *Duke Math. J.*, **2** (1936), 530–542.
- [39] Hedlund G. Dynamics of geodesic flows, *Bull. Amer. Math. Soc.*, **45** (1939), 241–260.
- [40] Hedlund G. Fuchsian groups and mixtures, *Ann. of Math.*, **40** (1939), 370–383.
- [41] Hopf E. Fuchsian group and ergodic theory, *Trans. Amer. Math. Soc.*, **39** (1936), 299–314.
- [42] Hopf E. Statistik der Geodätische Linien Manigfaltigkeiten Negativer Krümmung, *Ber. Voch. Sachs. Akad. Wiss. Leipzig*, **91** (1939), 261–304.
- [43] Iwaniec H. On indefinite quadratic forms in four variables, *Acta Arith.*, **33** (1977), 209–229.
- [44] Lesigne E. Théorèmes ergodiques pour une translation sur une nil-variété, *Erg. Theory and Dyn. Syst.*, **9** (1989), 115–126.
- [45] Lesigne E. Sur une nil-variété, les parties minimales associées à une translation sont uniquement ergodiques, *Erg. Theory and Dyn. Syst.*, **11** (1991), 379–391.
- [46] Marcus B. Unique ergodicity of the horocyclic flow: variable curvature case, *Israel J. Math.*, **21** (1975), 133–144.
- [47] Marcus B. Topological conjugacy of horocycle flows, *Amer. J. Math.*, **105** (1983), 623–632.

- [48] Margulis G. A. On the action of unipotent groups in the space of lattices, in: Proc. of the summer school on group representations, Bolyai Janos Math. Soc., Budapest, 1971, 365–370.
- [49] Margulis G. A. Lie groups and ergodic theory, in: Avramov L. L. (ed.) Algebra, Some Current Trends, Varna 1986, Lect. Notes in Math., vol. 1352, Springer-Verlag, 1988, 130–146.
- [50] Margulis G. A. Formes quadratiques indéfinies et flots unipotents sur les espaces homogènes, C. R. Acad. Sci., Série I, 304 (1987), 249–253.
- [51] Margulis G. A. Indefinite quadratic forms and unipotent flows on homogeneous spaces, Banach center publications, vol. 23, Polish Scientific Publishers, Warsaw, 1989.
- [52] Margulis G. A. Discrete subgroups and ergodic theory, in: Symp. in honour of A. Selberg, Number Theory, Trace Formulas and Discrete Groups, Academic Press, 1989, 377–398.
- [53] Margulis G. A. Compactness of minimal closed invariant sets of actions of unipotent groups, Geometriae Dedicata, 37 (1991), 1–9.
- [54] Margulis G. A. Orbits of group actions and values of quadratic forms at integral points, Preprint IHES.
- [55] Moore C., The Mautner phenomenon for general unitary representations, Pacific J. Math., 86 (1980), 155–169.
- [56] Morse M. Recurrent geodesics on a surface of negative curvature, Trans. Amer. Math. Soc., 22 (1921), 84–100.
- [57] Oppenheim A. The minima of indefinite quaternary quadratic forms of signature 0, Proc. Nat. Acad. Sci. USA, 15 (1929), 724–727.
- [58] Oppenheim A. The minima of indefinite quaternary quadratic forms, Ann. of Math., 32 (1931), 271–298.
- [59] Oppenheim A. Values of quadratic forms, I, II, Quart. J. Math., Oxford Ser. (2), 4 (1953), 54–59, 60–66.
- [60] Oppenheim A. Values of quadratic forms, III, Monatsh. Math. Phys., 57 (1953), 97–101.
- [61] Otal J.-P. Le spectre marqué des longueurs des surfaces à courbure négative, Ann. of Math., 131 (1990), 151–162.
- [62] Pansu P. Le flot géodésique des variétés riemanniennes à courbure négative, Séminaire Bourbaki, exp. 738, février 1991. [Имеется перевод: Пансю П. Геодезический поток на римановых многообразиях отрицательной кривизны. — В кн.: Труды семинара Н. Бурбаки за 1991 г. — М.: Мир, 1996, 226–250.]
- [63] Parry W. Ergodic properties of affine transformations and flows on nil-manifolds, Amer. J. Math., 91 (1969), 757–771.
- [64] Parry W. Metric classification of ergodic nilflows, Amer. J. Math., 93 (1971), 819–828.
- [65] Plante J. Anosov flows, Amer. J. Math., 94 (1972), 729–754.
- [66] Ratner M. Factors of horocyclic flows, Erg. Theory and Dyn. Syst., 2 (1982), 465–489.
- [67] Ratner M. Rigidity of horocycle flows, Ann. of Math., 115 (1982), 597–614.

- [68] Ratner M. Horocyclic flows: joinings and rigidity of products, *Ann. of Math.*, **118** (1983), 277–313.
- [69] Ratner M. Ergodic theory in hyperbolic space, in: *Contemp. Math.*, vol. 26, 1984, 309–334.
- [70] Ratner M. Invariant measures for unipotent translations on homogeneous spaces, *Proc. Nat. Acad. Sci. USA*, **87** (1990), 4309–4311.
- [71] Ratner M. Strict measure rigidity for unipotent subgroups of solvable groups, *Invent. Math.*, **101** (1990), 449–482.
- [72] Ratner M. On measure rigidity of unipotent subgroups of semisimple groups, *Acta Math.*, **165** (1990), 229–309.
- [73] Ratner M. On Raghunathan's measure conjecture, *Ann. of Math.*, **134** No 3, (1991), 445–607.
- [74] Ratner M. Raghunathan's topological conjecture and distributions of unipotent flows, *Duke Math. J.*, **63** (1991), 235–280.
- [75] Ratner M. Distribution rigidity for unipotent actions on homogeneous spaces, *Bull. Amer. Math. Soc.*, **24** (1991), 321–325.
- [76] Rees M. Tangentially distal flows, *Israel J. of Math.*, **15** (1980), 9–31.
- [77] Sarnak P. Asymptotic behaviour of periodic orbits of the horocycle flow and Eisenstein series, *Comm. Pure Applied Math.*, **34** (1981), 719–739.
- [78] Schmidt W. M. Diophantine approximations, *Lect. Notes in Math.*, vol. 785, Springer-Verlag, Berlin, 1980. [Имеется перевод: Шмидт В. Диофантовы приближения. — М.: Мир, 1983.]
- [79] Series C. The geometry of Markov numbers, *Math. Intelligencer*, **7** (1985), 20–29.
- [80] Shah N. A. Closures of totally geodesic immersions in manifolds of constant negative curvature, in: *Group theory from a geometrical viewpoint*, Trieste 1990, É. Ghys, A. Haefliger, A. Verjovsky (eds.), World Sci. Publishing, River Edge, 1991.
- [81] Sigmund K. On the space of invariant measures for hyperbolic flows, *Amer. J. Math.*, **94** (1972), 31–37.
- [82] Sikorav J.-C. Valeurs des formes quadratiques indéfinies irrationnelles (d'après G.A. Margulis), in: *Séminaire de théorie des nombres*, Paris, 1987–88, C. Goldstein (ed.), *Prog. in Math.*, vol. 81, Birkhäuser, 1990 307–315.
- [83] Старков А. Н. Строение орбит однородных потоков и гипотеза Рагунатана. — УМН, т. 45, 1990, вып. 2, с. 219–220.
- [84] Veech W. Unique ergodicity of horospherical flows, *Amer. J. Math.*, **99** (1977), 827–859.
- [85] Verjovsky A. Arithmetic, geometry and dynamics in unit tangent bundle of the modular orbifold, preprint I.C.T.P.
- [86] Vulakh L. Ya. On minima of rational indefinite quadratic forms, *J. Number Theory*, **21** (1985), 275–285.
- [87] Watson G. L. On indefinite quadratic forms in five variables, *Proc. London Math. Soc.*, **3** (1953), No. 3, 170–181.
- [88] Watson G. L. On indefinite quadratic forms in three or four variables, *J. London Math. Soc.*, **28** (1953), 239–242.

- [89] Witte D. Rigidity of some translations on homogeneous spaces, *Invent. Math.*, **81** (1985), 1–27.
- [90] Witte D. Zero entropy affine maps on homogeneous spaces, *Amer. J. Math.*, **109** (1987), 927–961.
- [91] Witte D. Rigidity of horospherical foliations, *Erg. Theory and Dyn. Syst.*, **9** (1989), 191–205.
- [92] Witte D. Measurable quotients of unipotent translations on homogeneous spaces, *Trans. Amer. Math. Soc.*, **345** (1994), No. 2, 577–594.
- [93] Zeghib A. Laminations et hypersurfaces géodésiques des variétés hyperboliques, *Ann. Scient. Éc. Norm. Sup.*, **24** (1991), 171–188.
- [94] Zeghib A. Systèmes dynamiques autonomes. Partie I : Ensembles invariants des flots d'Anosov algébriques, Prépublication E.N.S., Lyon, n° 47, 1991.

ÉTIENNE GHYS

École Normale Supérieure de Lyon
 Unité de Mathématiques Pures et Appliquées
 U.M.R. 128 du C.N.R.S.
 46, allée d'Italie
 F-69364 LYON CEDEX 07
 E-mail: ghys@umpa.ens-lyon.fr

1. ВВЕДЕНИЕ

1.1. «То тут, то там я ощущал присутствие гальванической цепи, и искрами, которые посыпались из нее, были фундаментальные равенства, связывающие i , j и k , заданные ровно в том виде, в каком я их всегда впоследствии использовал» [Н].

Уверенность Гамильтона в том, что кватернионы должны играть в математике и физике роль столь же значительную, как комплексные числа, так и не нашла подтверждения в его дни. Теперь же появление богатой теории многообразий, связанных с алгеброй кватернионов, отчасти доказывает его правоту. Эти многообразия, кроме того, естественным образом возникают в теории уравнений математической физики. Это *гиперкэлеровы многообразия*.

«Фундаментальные равенства, связывающие i , j и k », которые Гамильтон с таким воодушевлением вырезал на Бругхэмском мосту в октябре 1843 г., — это, конечно же,

$$i^2 = j^2 = k^2 = ijk = -1.$$

Гиперкэлерово многообразие — это многообразие, которое допускает действие кватернионов i , j и k на касательных векторах, согласованное с метрикой (размерность такого многообразия обязательно кратна 4). Приведем точное определение.

Определение. *Гиперкэлерово многообразие* — это риманово многообразие с тройкой плоских (относительно связности Леви-Чевиты) ортогональных автоморфизмов I , J и K касательного расслоения, удовлетворяющих кватернионным тождествам $I^2 = J^2 = K^2 = IJK = -1$.

1.2. Напомним, что риманово многообразие с одним таким автоморфизмом называется кэлеровым. Таким образом, термин «гиперкэлерово», принадлежащий Э. Калаби [Ca], совершенно справедлив — метрика является кэлеровой по отношению к нескольким

комплексным структурам, — хотя напоминает он в большей степени «гиперкомплексные числа» Грассмана, чем гамильтоновы кватернионы. Однако между кэлеровыми и гиперкэлеровыми многообразиями имеется существенное различие. Добавление эрмитовой формы $\partial\bar{\partial}f$ с произвольной достаточно малой C^∞ -функцией f не влияет на кэлеровость; поэтому пространство кэлеровых метрик бесконечномерно. Кроме того, построение кэлеровых многообразий не представляет никаких затруднений. Каждое комплексное подмногообразие в CP^n наследует кэлерову метрику, и, таким образом, проективные многообразия дают обширный запас примеров кэлеровых многообразий.

Гиперкэлерова метрика, напротив, — объект очень жесткий. Пространство таких метрик с точностью до изометрии на компактном многообразии не более чем конечномерно, а построение примеров — задача непростая. В частности, искать их как кватернионные подмногообразия в кватернионном проективном пространстве HP^n бесполезно [Gr].

1.3. Понятие гиперкэлерова многообразия впервые появилось в 1955 г. в связи с классификацией М. Берже групп голономии римановых многообразий. На гиперкэлеровом многообразии параллельные переносы сохраняют автоморфизмы I , J и K (поскольку они плоские); следовательно, группа голономии содержится как в ортогональной группе $O(4n)$, так и в группе $GL(n, \mathbb{H})$ обратимых кватернионных матриц (т. е. линейных преобразований, коммутирующих с правым умножением на i , j и k). Пересечением этих групп является $Sp(n)$ — группа унитарных кватернионных $n \times n$ -матриц. Эта группа встречается в списке Берже.

1.4. Заметим, что группа $Sp(n)$ также является пересечением группы $U(2n)$ и группы $Sp(2n, \mathbb{C})$ линейных преобразований пространства $\mathbb{C}^{2n}$, сохраняющих некоторую невырожденную кососимметричную форму. Таким образом, гиперкэлерово многообразие естественным образом является комплексным многообразием с голоморфной симплектической формой. Это можно увидеть и непосредственно. Рассмотрим тройку кэлеровых 2-форм

$$\omega_1(X, Y) = g(IX, Y), \quad \omega_2(X, Y) = g(JX, Y), \quad \omega_3(X, Y) = g(KX, Y),$$

связанных с комплексными структурами I , J и K . Форма $\omega_c = \omega_2 + i\omega_3$ является невырожденной и плоской по отношению к комплексной структуре I ; поэтому она замкнута и голоморфна.

Такая точка зрения дает способ построения примеров гиперкэлеровых многообразий и показывает, какого сорта дифференциальные уравнения управляют гиперкэлеровостью. Во-первых, голоморфная форма объема ω^n на гиперкэлеровом многообразии задает плоскую тривиализацию канонического линейного расслоения. Кривизна канонического расслоения относительно любой кэлеровой метрики — это форма Риччи; поэтому наличие гиперкэлеровой метрики влечет за собой обращение в нуль тензора Риччи. В случае наименьшей размерности четыре это означает, что такая метрика удовлетворяет вакуумному уравнению Эйнштейна.

Теорема Калаби-Яу [Y] утверждает, что на компактном кэлеровом многообразии с голоморфно тривиальным каноническим расслоением существует кэлерова метрика с нулевым тензором Риччи. Кроме того, гораздо более ранняя теорема С. Бохнера [Bo] доказывает, что любая голоморфная дифференциальная форма на компактном кэлеровом многообразии с нулевым тензором Риччи является плоской. Из этих двух теорем следует, что на всяком компактном кэлеровом многообразии с голоморфной симплектической формой имеется гиперкэлерова метрика. Таким образом можно доказать существование гиперкэлеровой метрики на многих комплексных многообразиях. В четырехмерном (вещественном) случае единственный (но очень важный) нетривиальный пример дает $K3$ -поверхность. В высших размерностях естественный класс голоморфно-симплектических, а стало быть, и гиперкэлеровых многообразий составляют схемы Гильберта 0-мерных циклов на $K3$ -поверхности и на 2-мерном комплексном торе [Bea].

1.5. В нашем обзоре мы постараемся не только доказать существование метрики. Мы будем *строить* решения более явным образом, чтобы лучше понять геометрию гиперкэлеровых многообразий и ощутить ее богатство. Такая возможность появилась совсем недавно — каких-то двадцать лет назад было практически невозможно придумать нетривиальную риманову метрику с нулевым тензором Риччи.

Имеются два основных подхода к построению гиперкэлеровых метрик: (а) теория твисторов и (b) гиперкэлерова редукция.

Твисторный подход базируется на основополагающей работе Р. Пенроуза по теории относительности [P]. Он позволяет *закодировать* данные гиперкэлеровой метрики в терминах голоморфной геометрии. Можно сказать, что таким образом все дифференци-

альные уравнения сводятся к одному — уравнению Коши—Римана. Раскодировать информацию для написания метрики зачастую непросто, вывести же глобальные свойства метрики, такие, как полнота, — практически невозможно. С другой стороны, гиперкэлерова редукция позволяет выводить такие свойства очень легко, хотя и уступает твисторному методу в универсальности.

Гиперкэлерова редукция возникает также в ряде вопросов математической физики, например, в связи с суперсимметрией [HKLR]. Фактически имеется два способа ее применения. Первый — это конечномерная редукция. В этом случае отыскание метрики сводится к решению алгебраических уравнений. Второй способ использует бесконечномерные методы, хотя результат может быть и конечномерным. В этом случае для нахождения метрики приходится решать дифференциальные уравнения. Однако для этих уравнений часто известны способы решения, и потому *в принципе* мы получаем больше информации, чем просто теорему существования.

1.6. Мы проиллюстрируем эти два способа на многочисленных примерах, которые выбраны согласно нашему основному принципу: искать комплексные симплектические многообразия. Все следующие гиперкэлеровы многообразия *априори* являются комплексными многообразиями с голоморфными симплектическими формами:

- (i) разрешения рациональных особенностей поверхностей;
- (ii) орбиты коприсоединенного действия комплексной группы Ли;
- (iii) пространства представлений фундаментальной группы поверхности в комплексной группе Ли;
- (iv) пространство базированных (переводящих отмеченную точку в отмеченную точку) рациональных отображений $f : \mathbb{CP}^1 \rightarrow \mathbb{CP}^1$ степени k ;
- (v) пространство базированных петель в комплексной группе Ли.

Конструкции гиперкэлеровых метрик на этих пространствах содержатся в работах П. Б. Кронхаймера, С. К. Дональдсона и других авторов. Особенно замечательным кажется то, что на почти всех этих столь различных пространствах гиперкэлерова метрика описывается в терминах специальных решений уравнений антиавтотодуальности Янга—Миллса в $\mathbb{R}^4$. При этом сами эти уравнения, имеющие физическое происхождение, существенно используют отождествление $\mathbb{R}^4$ с пространством кватернионов. Душа Гамильтона может, наконец, обрести покой.

2. ТВИСТОРНАЯ КОНСТРУКЦИЯ

2.1. Гиперкэлерово многообразие M^{4n} по определению снабжено тройкой комплексных структур I, J и K . В действительности, если $u = (a, b, c) \in \mathbb{R}^3$, то

$$(aI + bJ + cK)^2 = -(a^2 + b^2 + c^2);$$

поэтому при $\|u\| = 1$ мы получаем еще одну плоскую (и, следовательно, интегрируемую) комплексную структуру I_u . Гиперкэлерова метрика является кэлеровой по отношению ко всем этим комплексным структурам. *Твисторное пространство* многообразия M — это произведение $Z = M \times S^2$. Касательное пространство к двумерной сфере S^2 , рассматриваемой как сфера Римана, обладает естественной комплексной структурой I_0 . Для $X \in T_m M, Y \in T_u S^2$ формула

$$I(X, Y) = (I_u X, I_0 Y)$$

определяет комплексную структуру в касательном пространстве $T_m M \oplus T_u S^2$ к твисторному пространству Z . Имеется теорема [HKLR, S, AHS], утверждающая, что эта почти комплексная структура *интегрируема*, т.е. Z является комплексным многообразием комплексной размерности $2n + 1$.

2.2. Твисторное пространство обладает следующими свойствами. Во-первых, проекция на $S^2 \cong \mathbb{C}P^1$ голоморфна. Во-вторых, антиподальное отображение σ на единичной сфере переводит комплексную структуру I_u в $-I_u$, а I_0 — в $-I_0$. Это позволяет рассматривать Z как *вещественное* комплексное многообразие (комплексное многообразие с антиголоморфной инволюцией — что-то вроде проективного многообразия, заданного уравнениями с вещественными коэффициентами).

Тройка кэлеровых форм ω_1, ω_2 и ω_3 на M определяется выбором ортонормированного базиса в $\mathbb{R}^3$. Поэтому на слоях проекции $M \times SO(3) \rightarrow SO(3)$ определена 2-форма ω_c , комплексная относительно структуры I_u , соответствующей первому элементу u ортонормированного базиса. При естественной проекции $SO(3) \rightarrow S^2$ форма ω_c опускается до подкрученной на касательное расслоение $O(2)$ к $S^2 \cong \mathbb{C}P^1$ голоморфной формы на Z .

Наконец, каждая точка $m \in M$ определяет сечение $\{m\} \times S^2$ проекции $p : Z = M \times S^2 \rightarrow \mathbb{C}P^1$, являющееся одновременно голоморфным и вещественным (т.е. σ -инвариантным). Нормальное

расслоение этого сечения изоморфно расслоению $\mathbb{C}^{2n} \otimes \mathcal{O}(1)$. Мы получаем, таким образом, основную теорему [HKLR].

Теорема 1. Если Z — твисторное пространство гиперкэлера многообразия M^{4n} , то

- (1) Z является голоморфным расслоением $p: Z \rightarrow \mathbb{C}P^1$ над проективной прямой;
- (2) морфизм p обладает семейством сечений, нормальное расслоение к каждому из которых изоморфно $\mathbb{C}^{2n} \otimes \mathcal{O}(1)$;
- (3) расслоение $\Lambda^2 T_F^* \otimes \mathcal{O}(2)$ обладает голоморфным сечением ω , индуцирующим симплектическую форму на каждом слое;
- (4) Z обладает вещественной структурой σ , согласованной с (1)–(3) и индуцирующей антиподальное отображение на $\mathbb{C}P^1$.

Обратно, пространство параметров семейства вещественных сечений любого комплексного многообразия Z^{2n+1} , удовлетворяющего условиям (1)–(4), является $4n$ -мерным многообразием с гиперкэлеровой метрикой, для которого Z служит твисторным пространством.

(Здесь $\mathcal{O}(k)$ обозначает обратный образ относительно морфизма p соответствующего линейного расслоения на $\mathbb{C}P^1$, а T_F^* — кокасательное расслоение вдоль слоев этого морфизма. Если E — векторное расслоение, то через $E(k)$ обозначается тензорное произведение $E \otimes \mathcal{O}(k)$.)

Ключевым моментом в доказательстве обратного утверждения является тот факт, что касательное пространство к пространству сечений отождествляется с

$$H^0(\mathbb{C}P^1; T_F) \cong H^0(\mathbb{C}P^1; T_F(-1)) \otimes H^0(\mathbb{C}P^1; \mathcal{O}(1)).$$

Кососимметрическая форма ω индуцирует симплектическую форму на первом сомножителе, которая вместе с канонической кососимметрической формой на втором сомножителе задает симметрическую форму на тензорном произведении, т. е. метрику.

2.3. Таким образом, гиперкэлера метрика «закодирована» в геометрии твисторного пространства Пенроуза. Дифференциальные уравнения, описывающие метрику, проявляются в нелинейности геометрической конструкции пространства Z . Кроме того, Z может рассматриваться как самостоятельный объект голоморфной симплектической геометрии. Можно рассматривать проекцию p и

форму ω как задание симплектического многообразия, определенного над полем функций от одной переменной. Тогда сечения становятся рациональными точками этого многообразия. В отдельных примерах такой «диофантов» аспект становится естественным, и нахождение сечений сводится к решению алгебраических уравнений с полиномиальными коэффициентами [Н1, Н4].

2.4. Пример. Пусть Z — тотальное пространство двумерного векторного расслоения $E = \mathbb{C}^2(1)$ над CP^1 . Тогда Z является твисторным пространством для плоской гиперкэлеровой структуры на $\mathbb{R}^4$.

Рассмотрим теперь не обращающееся в нуль сечение s расслоения E . Оно порождает тривиальное линейное подрасслоение, а факторрасслоение изоморфно $O(2)$. Сдвиги на кратное сечению s превращают E в главное S -расслоение над тотальным пространством T расслоения $O(2)$. Для каждого $\lambda \in \mathbb{R}$ мы определяем линейное расслоение L^λ на T формулой

$$L^\lambda = E \times_{\mathbb{C}} \mathbb{C},$$

где $u \in \mathbb{C}$ действует по правилу $w \mapsto e^{\lambda u} w$.

Обозначим через z тавтологическое сечение обратного образа расслоения $O(2)$ в его тотальном пространстве T . Тогда

$$Z = \{(x, y) \in L^\lambda(1) \oplus L^{-\lambda}(1) : xy = z\}$$

определяет твисторное пространство для полной гиперкэлеровой метрики на $\mathbb{R}^4$ — метрики *Taub-NUT*. Явное вычисление метрики, исходя из твисторной конструкции, содержится в [Be].

3. ГИПЕРКЭЛЕРОВА РЕДУКЦИЯ

3.1. Отправной точкой твисторной конструкции для гиперкэлеровых метрик является тройка комплексных структур I, J и K . На самом деле естественнее было бы рассматривать эти структуры на многообразии и без существования согласованной с ними метрики. Это более общая теория *гиперкомплексных многообразий*. Соответствующая твисторная теория отличается от изложенной выше отсутствием п. (3) в теореме 1.

Гиперкэлерова редукция, напротив, исходит не из комплексных структур, а из соответствующих кэлеровых форм ω_1, ω_2 и ω_3 . Сворачивая эти формы с обратными формами, мы можем восстановить как автоморфизмы I, J и K , так и саму метрику. Другими словами, стабилизатор форм $\omega_1, \omega_2, \omega_3$ равен $Sp(n)$, в то время как стабилизатор автоморфизмов I, J, K равен $GL(n, \mathbb{H})$.

Имеется очень простое описание гиперкэлеровых многообразий в терминах этих форм [HKLR]:

Теорема 2. Пусть M^{4n} — многообразие с тройкой 2-форм $\omega_1, \omega_2, \omega_3$, таких, что их стабилизатор в группе $GL(4n, \mathbb{R})$ в каждой точке $m \in M$ сопряжен с $Sp(n)$. Тогда эти формы определяют гиперкэлерову структуру в том и только в том случае, когда они замкнуты.

Эта теорема, являющаяся прямым следствием теоремы Ньюлендера—Ниренберга, прочно закрепляет теорию гиперкэлеровых многообразий в царстве симплектической геометрии.

3.2. Конструкция гиперкэлеровой редукции повторяет конструкцию редукции Марсдена—Вайнштейна в симплектической геометрии.

Если (M, ω) — симплектическое многообразие с симплектическим действием группы Ли G , то при слабых ограничениях можно определить эквивариантное отображение моментов $\mu : M \rightarrow \mathfrak{g}^*$, принимающее значения в двойственном пространстве алгебры Ли группы G , для которого выполнено следующее условие. Для любого $\xi \in \mathfrak{g}$ имеем $d\mu(\xi) = i(X_\xi)\omega$, где X_ξ — векторное поле на M , порожденное действием группы G . Отображение μ определено с точностью до аддитивной константы $\zeta \in \mathfrak{z} \subseteq \mathfrak{g}^*$, где $\mathfrak{z}$ — подпространство G -инвариантных элементов в $\mathfrak{g}^*$.

Симплектическая редукция заключается в следующем. Выберем регулярное для μ значение $\zeta \in \mathfrak{z}$. Тогда ограничение формы ω на подмногообразии $\mu^{-1}(\zeta)$ инвариантно и вырождено в направлении G -орбит, а следовательно, опускается до формы $\bar{\omega}$ на фактормногообразии $\mu^{-1}(\zeta)/G$. Можно доказать, что форма $\bar{\omega}$ симплектическая.

3.3. Пусть M — гиперкэлерово многообразие с действием группы Ли G , сохраняющим три кэлеровы формы ω_1, ω_2 и ω_3 . Мы получаем три отображения моментов μ_1, μ_2 и μ_3 или, что то же самое, векторнозначное отображение моментов

$$\mu : M \rightarrow \mathfrak{g}^* \otimes \mathbb{R}^3.$$

Имеем следующую теорему [HKLR]:

Теорема 3. Если $\zeta \in \mathfrak{z} \otimes \mathbb{R}^3$ — регулярное значение отображения моментов μ , то $\mu^{-1}(\zeta)/G$ — гиперкэлерово многообразие.

Доказательство основано на теореме 2. Каждая из форм ω_i , как и в симплектическом случае, опускается до формы $\bar{\omega}_i$. Остается проверить, что формы $\bar{\omega}_1$, $\bar{\omega}_2$ и $\bar{\omega}_3$ все еще удовлетворяют алгебраическим кватернионным соотношениям.

Заметим, что

$$\dim \mu^{-1}(\zeta)/G = \dim M - 4 \dim G.$$

3.4. Симплектическая редукция играет важную роль в *кэлеровой геометрии*, поскольку если M — кэлерово многообразие и группа G сохраняет не только симплектическую форму, но и комплексную структуру, то результат симплектической редукции — тоже кэлерово многообразие [Ki, HKLR]. Возникающая на нем комплексная структура отождествляет его с комплексным фактором M^s/G^c некоторого открытого подмножества M^s *стабильных* точек в M относительно действия комплексной группы G^c , голоморфно продолжающего действие группы G . Точка является стабильной, если ее G^c -орбита пересекает $\mu^{-1}(\zeta)$. Эта симплектическая точка зрения очень хорошо работает для проективных многообразий и эффективно соотносится с геометрической теорией инвариантов [Ki].

3.5. Пример. Вот простейший пример кэлеровой симплектической редукции. Пусть $M = \mathbb{C}^n$ со стандартной эрмитовой структурой и действием окружности $G = S^1$ скалярными умножениями. Отображение моментов $\mu : \mathbb{C}^n \rightarrow i\mathbb{R}$ определяется формулой

$$\mu(z) = i\|z\|^2,$$

и $\mu^{-1}(i) = S^{2n-1}$. Симплектическая редукция равна

$$S^{2n-1}/S^1 = \mathbb{C}P^{n-1} = \mathbb{C}^n \setminus \{0\}/\mathbb{C}^*,$$

так что $M^s = \mathbb{C}^n \setminus \{0\}$ и $\mathbb{C}P^{n-1}$ наследует естественную кэлерову метрику — метрику Фубини—Штуди.

3.6. В гиперкэлеровой ситуации формы ω_1 , ω_2 и ω_3 задают автоморфизмы I , J и K ; поэтому действие группы G , сохраняющее эти симплектические формы, автоматически сохраняет комплексные структуры. Зафиксируем одну из комплексных структур I . Тогда функция $\mu_c = \mu_2 + i\mu_3$ голоморфна. На самом деле μ_c является отображением моментов действия комплексной группы G^c по отношению к голоморфной симплектической форме ω_c . Таким образом,

$$\mu^{-1}(\zeta) = \mu_1^{-1}(a) \cap \mu_c^{-1}(\alpha)$$

для некоторых $a \in \mathfrak{z}$ и $\alpha \in \mathfrak{z} \otimes \mathbb{C}$.

С этой точки зрения гиперкэлерова редукция совпадает с симплектической редукцией кэлерова многообразия $\mu_c^{-1}(\alpha)$. Следовательно, индуцированная комплексная структура I на гиперкэлеровой редукции совпадает с комплексной структурой фактормногообразия $\mu_c^{-1}(\alpha)/G^c$ (фактор берется с учетом стабильности). Таким образом, гиперкэлерова редукция является голоморфным аналогом симплектической редукции.

3.7. Позволим теперь комплексной структуре I меняться (иначе говоря, будем рассматривать различные слои твисторного пространства $p : Z \rightarrow \mathbb{C}P^1$). Твисторное пространство гиперкэлеровой редукции по существу совпадает с послойной симплектической редукцией пространства Z по голоморфному действию группы G^c . Для каждой из комплексных структур соответствующее комплексное отображение моментов μ_c является ограничением голоморфного сечения $\tilde{\mu}$ расслоения $(g^c)^* \otimes \mathcal{O}(2)$ (подкрутка появляется из-за скрученности формы ω из теоремы 1). Таким образом, твисторное пространство гиперкэлеровой редукции равно $\tilde{\mu}^{-1}(\zeta)^s/G^c$, где $\zeta \in \mathfrak{g}^* \otimes \mathbb{R}^3 \subset (g^c)^* \otimes H^0(\mathbb{C}P^1; \mathcal{O}(2)) \subset H^0(Z; (g^c)^* \otimes \mathcal{O}(2))$.

На языке п. 2.3 это означает, что твисторное пространство гиперкэлеровой редукции является симплектической редукцией твисторного пространства над полем рациональных функций.

3.8. Замечательно, что можно получить интересные примеры, начиная с плоского гиперкэлерова многообразия $M^{4n} = \mathbb{H}^n$ и производя редукцию по линейному действию группы. Вот, например, ближайший аналог конструкции проективного пространства из п. 3.5.

Пусть V есть n -мерное эрмитово векторное пространство, а V^* — его двойственное. Тогда

$$M^{4n} = V \oplus V^*$$

является плоским гиперкэлеровым многообразием. Пусть $G = S^1$ действует на M по правилу

$$u(x, \xi) = (ux, u^{-1}\xi).$$

Действие сохраняет гиперкэлеровы 2-формы. Комплексное и вещественное отображения моментов находятся легко:

$$\mu_c(x, \xi) = \xi(x), \quad \mu_1(x, \xi) = i(\|x\|^2 - \|\xi\|^2).$$

Взяв $\zeta = (i, 0, 0) \in \mathfrak{g}^* \otimes \mathbb{R}^3$, мы получаем

$$\mu^{-1}(\zeta) = \{(x, \xi) : \xi(x) = 0, \|x\|^2 - \|\xi\|^2 = 1\}.$$

Второе условие следующим образом определяет множество стабильных точек: $\mu_c^{-1}(0)^s = \{(x, \xi) : \xi(x) = 0, x \neq 0\}$. Значит, $\mu_c^{-1}(0)^s/\mathbb{C}^*$ является *кокасательным расслоением* проективного пространства $P(V) = \mathbb{C}P^{n-1}$. Однако это лишь одна точка на двумерной сфере комплексных структур на этом гиперкэлеровом многообразии. Вследствие симметрий остальные комплексные структуры отождествляются с факторами $\mu_c^{-1}(\alpha)/\mathbb{C}^*$ при $\alpha \neq 0$ и являются аффинными расслоениями над $\mathbb{C}P^{n-1}$, ассоциированными с кокасательным расслоением. Заметим, что комплексное кокасательное расслоение естественно наделяется структурой комплексного симплектического многообразия.

Метрика, возникающая в частном случае $n = 2$, называется метрикой *Егучи—Хансона* [ЕН]. Общая комплексная структура совпадает со структурой аффинной квадррики в $\mathbb{C}^3$. Первые примеры в высших размерностях были получены Э. Калаби [Ca].

3.9. Довольно широкий класс гиперкэлеровых отображений моментов, с которыми мы будем сталкиваться, основан на следующем примере.

Пусть G — группа Ли с биинвариантной метрикой. Положим

$$M = \mathfrak{g} \otimes \mathbb{H},$$

где $\mathfrak{g}$ — соответствующая алгебра Ли. Это плоское гиперкэлерово многообразие. Присоединенное действие группы G сохраняет симплектические формы. Гиперкэлеровы отображения моментов определяются формулами

$$\begin{aligned}\mu_1(A) &= [A_0, A_1] + [A_2, A_3], \\ \mu_2(A) &= [A_0, A_2] + [A_3, A_1], \\ \mu_3(A) &= [A_0, A_3] + [A_1, A_2],\end{aligned}$$

где элемент $A \in \mathfrak{g} \otimes \mathbb{H}$ представлен в виде $A = A_0 + A_1 i + A_2 j + A_3 k$. Заметим, что комплексное отображение моментов задается формулой

$$\mu_c(A) = [A_0 + iA_1, A_2 + iA_3].$$

Полагая $\alpha = A_0 + iA_1$ и $\beta = A_2 + iA_3$, записываем три отображения моментов как комплексное и вещественное отображения моментов

$$\begin{aligned}\mu_c &= [\alpha, \beta], \\ \mu_r &= [\alpha, \alpha^*] + [\beta, \beta^*].\end{aligned}$$

Стоит указать, что, подставляя вместо A_i ковариантные производные ∇_i , мы получаем запись системы уравнений антиавтотуальности Янга—Миллса в виде одного уравнения $\mu = 0$. Мы встретимся с этими уравнениями в последующих примерах.

4. РАЦИОНАЛЬНЫЕ ДВОЙНЫЕ ТОЧКИ

4.1. При изучении особенностей поверхностей *рациональные двойные точки* характеризуются тривиальностью канонического расслоения их минимального разрешения. Таким образом, минимальные разрешения рациональных двойных точек локально являются двумерными комплексными многообразиями с голоморфной симплектической формой. Основным примером является пространство $\mathbb{C}^2/\Gamma$, где Γ — конечная подгруппа в $SU(2)$. Мы опишем конструкцию гиперкэлеровой метрики на минимальном разрешении этой особенности, принадлежащую Кронхаймеру [K1, K2]. В п. 3.8 мы уже встречались с примером такой метрики — метрикой Егучи—Хансона. Последняя была определена на кокасательном расслоении к $\mathbb{C}P^1$. Но нулевое сечение кокасательного расслоения имеет индекс самопересечения -2 и, следовательно, может быть стянуто в особую точку — обычную двойную точку $\mathbb{C}^2/\pm 1$. Конструкция Кронхаймера опирается на использование гиперкэлеровой редукции (как в п. 3.8).

4.2. Пусть Γ — конечная подгруппа в $SU(2)$. Пусть $V = L^2(\Gamma)$ — конечномерное гильбертово пространство функций на Γ , $U(V)$ — унитарная группа пространства V и $\mathfrak{u}(V)$ — ее алгебра Ли. Тогда, как и в п. 3.9, $\mathfrak{u}(V) \otimes \mathbb{H}$ является плоским гиперкэлеровым многообразием — кватернионным векторным пространством с инвариантным скалярным произведением.

Поскольку $SU(2)$ изоморфна группе единичных кватернионов, то группа Γ действует слева как на кватернионах, так и на $L^2(\Gamma) = V$, порождая левое действие на $\mathfrak{u}(V) \otimes \mathbb{H}$, коммутирующее с умножением справа на кватернионы. Поэтому подпространство инвариантов

$$M = (\mathfrak{u}(V) \otimes \mathbb{H})^\Gamma$$

также является плоским гиперкэлеровым многообразием.

Проективная группа $PU(V)$ действует сопряжениями на $\mathfrak{u}(V) \otimes \mathbb{H}$, сохраняя гиперкэлерову структуру; поэтому группа $G = PU(V)^\Gamma$ действует на M таким же образом. Это как раз та ситуация, в которой можно применять гиперкэлерову редукцию.

4.3. Чтобы получить больше информации, разложим представление V в сумму неприводимых,

$$V = L^2(\Gamma) = \bigoplus_{i=0}^l (V_i^L \otimes V_i^R).$$

Суммирование ведется по всем неприводимым представлениям относительно левого и правого действия, и V_0 — тривиальное представление. Отсюда получаем

$$U(\hat{V})^\Gamma = U(V_0^R) \times \dots \times U(V_l^R).$$

Полагая $\dim V_i^R = \dim V_i^L = n_i$, получаем

$$\dim G = \sum_{i=0}^l n_i^2 - 1.$$

Из приведенного выше разложения сразу видны центр группы G и подпространство инвариантов $\mathfrak{z} \subset \mathfrak{g}^*$. Имеем

$$\dim \mathfrak{z} = l.$$

4.4. Структура неприводимых представлений группы Γ удобным образом кодируется с помощью соответствия Маккея [Мак]. Каждой подгруппе (циклической, бинарной диэдральной, тетраэдральной, октаэдральной или икосаэдральной) соответствует диаграмма Дынкина типа A_l , D_l , E_6 , E_7 или E_8 соответственно. Каждая вершина диаграммы соответствует нетривиальному неприводимому представлению группы Γ .

Напомним, что диаграммы Дынкина классифицируют простые комплексные алгебры Ли. Вершинам диаграммы отвечают простые корни α_i соответствующей алгебры Ли. Старший корень выражается через простые корни как

$$\sum_{i=1}^l n_i \alpha_i.$$

Согласно Маккею, целое число n_i в точности равно размерности представления, соответствующего i -й вершине. Ситуацию можно несколько упростить с помощью *расширенной* диаграммы Дынкина, включающего в себя старший корень. Соотношение

$$\sum_{i=0}^l n_i \alpha_i = 0,$$

где $n_0 = 1$ и α_0 — старший корень со знаком минус, делает тривиальное представление равноправным по отношению к остальным.

4.5. Для нахождения размерности пространства M из п. 4.2 необходимо знать, как устроено тензорное произведение тавтологического двумерного представления группы $\Gamma \subset SU(2)$ и представлений V_i . Здесь снова ответ дается соответствием Маккея:

$$\mathbb{C}^2 \otimes V_i = \bigoplus_j a_{ij} V_j,$$

где a_{ij} равно единице, если вершины i и j расширенной диаграммы Дынкина соединены, и нулю в противном случае.

Как комплексное векторное пространство M равно $(\text{Hom}(V, V) \otimes \mathbb{C}^2)^\Gamma$ и потому

$$\begin{aligned} \dim_{\mathbb{C}} M &= \dim \bigoplus_{i,j} (\text{Hom}(V_j^L \otimes V_j^R, V_i^L \otimes V_i^R \otimes \mathbb{C}^2))^\Gamma \\ &= \dim \bigoplus_{i,j} a_{ij} \text{Hom}(V_i^L, V_j^L) \\ &= \sum_{i,j} a_{ij} n_i n_j. \end{aligned}$$

Применяя формулу Картана к соотношению из п. 4.4, мы получаем

$$-2 \sum_{i=0}^l n_i^2 + \sum_{i,j} a_{ij} n_i n_j = 0;$$

следовательно,

$$\dim_{\mathbb{R}} M = 4 \sum_{i=0}^l n_i^2 = 4(\dim G + 1)$$

ввиду п. 4.3.

Таким образом, если группа G действует свободно, то мы получим гиперкэлерову редукцию размерности $\dim M - 4 \dim G = 4$. Заметим, что $\dim \mathfrak{z} = l$, а следовательно, у нас есть возможность выбора отображения моментов, которая должна быть использована для получения свободного действия.

4.6. Для того чтобы выявить связь с разрешением особенности C^2/Γ , рассмотрим точку $\zeta = (\zeta_1, 0, 0) \in \mathfrak{z} \otimes \mathbf{R}^3$, где $\mathfrak{z} \subset (\mathfrak{u}(V)^\Gamma)^*$ — подпространство инвариантов. Поскольку фактически мы находимся в ситуации п. 3.9, то уравнения гиперкэлерова отображения моментов можно записать как уравнения на пару комплексных матриц

$$\begin{aligned} [\alpha, \beta] &= 0, \\ [\alpha, \alpha^*] + [\beta, \beta^*] &= \zeta_1, \end{aligned}$$

где $\alpha, \beta \in \text{Hom}(V, V)$ определяют Γ -инвариантный элемент пространства $\text{Hom}(V, V) \otimes C^2$.

Инвариантность означает следующее. Пусть $e \in V$ — общий собственный вектор операторов α и β (он существует, так как α и β коммутируют) с собственными значениями a, b , т. е. $(\alpha, \beta)e = (a, b)e$. Тогда

$$(\alpha, \beta)e^\gamma = (a, b)^\gamma e^\gamma,$$

где $(a, b) \mapsto (a, b)^\gamma$ — тавтологическое действие группы Γ на C^2 .

Группа Γ на $C^2 \setminus \{0\} = \mathbf{H} \setminus \{0\}$ действует свободно (умножением на ненулевой кватернион). Поэтому если одна пара собственных значений отлична от нуля, то имеется $|\Gamma| = \dim V$ различных пар собственных значений. В частности, паре (α, β) можно сопоставить Γ -орбиту пары (a, b) . Такое сопоставление определяет голоморфное по отношению к комплексной структуре I отображение p из гиперкэлеровой редукции в C^2/Γ . При подходящем выборе ζ_1 гиперкэлерова редукция является неособой; поэтому, проверив, что отображение p биголоморфно вне особой точки в C^2/Γ , мы докажем, что оно является разрешением особенности.

Четырехмерные метрики, получающиеся таким образом, являются не только полными гиперкэлеровыми метриками, но и *асимптотически локально евклидовыми* (ALE), т. е. при приближении к бесконечности хорошо аппроксимируются евклидовой метрикой на $\mathbf{R}^4/\Gamma$.

5. КОПРИСОЕДИНЕННЫЕ ОРБИТЫ

5.1. Орбиты коприсоединенного действия группы Ли в двойственном пространстве ее алгебры Ли являются симплектическими многообразиями относительно канонической симплектической структуры Костанта—Кириллова. Если, кроме того, группа G компактна,

то орбиты являются кэлеровыми многообразиями. Этот факт можно использовать для получения всех неприводимых представлений группы G в ситуации геометрического квантования (теорема Бореля—Вейля).

Существование симплектической структуры — факт совершенно общий. В частности, коприсоединенные орбиты комплексной группы Ли G^c являются голоморфными симплектическими многообразиями. Во многих случаях, как было показано в работе Кронхаймера [К3], эти многообразия обладают естественной *гиперкэлеровой* метрикой. Первый пример такой метрики — это метрика Егучи—Хансона; аффинная квадрика в $\mathbb{C}^3$ (общая комплексная структура для этой метрики) является орбитой полупростого элемента в $\mathfrak{g} = \mathfrak{g}^*$ относительно действия группы $G = SL(2, \mathbb{C})$. Мы воспроизведем общую конструкцию для *регулярной* полупростой орбиты — орбиты вида G^c/T^c , где G^c — комплексная полупростая группа Ли и T^c — максимальный комплексный тор. Компактным аналогом этого многообразия является многообразие флагов G/T — тип орбиты, который выделяется в теореме Бореля—Вейля.

5.2. Для того чтобы получить метрику, мы применим гиперкэлерову редукцию к плоскому бесконечномерному гиперкэлерову многообразию. Рассмотрим компактную полупростую группу Ли G и тривиальное главное G -расслоение P над $\mathbb{R}^4 \setminus \{0\}$. Рассмотрим пространство связностей в P . Это бесконечномерное аффинное пространство, группа переносов которого является векторным пространством $\Omega^1(\mathbb{R}^4 \setminus \{0\}; \text{ad } P)$ дифференциальных 1-форм с коэффициентами в расслоении алгебр Ли. Отождествление $\mathbb{R}^4$ с $\mathbb{H}$ индуцирует в пространстве $\Omega^1(\mathbb{R}^4 \setminus \{0\}; \text{ad } P)$ структуру кватернионного пространства. Левое действие группы $Sp(1)$ единичных кватернионов коммутирует с правым действием; поэтому пространство $Sp(1)$ -инвариантных связностей также является кватернионным пространством. На этом пространстве действует группа $Sp(1)$ -инвариантных калибровочных преобразований.

Для определения гиперкэлеровой метрики необходимо ввести дополнительные граничные условия. Затем можно будет рассмотреть гиперкэлерову редукцию по действию группы калибровочных преобразований.

5.3. Условие $Sp(1)$ -инвариантности выделяет в $\mathbb{R}^4 \setminus \{0\}$ радиальную координату $s \in (-\infty, 0]$. Уравнения отображения моментов

приводятся к системе нелинейных дифференциальных уравнений

$$\begin{aligned} \left[\frac{d}{ds} + B_0, B_1 \right] + [B_2, B_3] &= 0, \\ \left[\frac{d}{ds} + B_0, B_2 \right] + [B_3, B_1] &= 0, \\ \left[\frac{d}{ds} + B_0, B_3 \right] + [B_1, B_2] &= 0 \end{aligned}$$

на функции $B_i(s)$, принимающие значения в алгебре Ли $\mathfrak{g}$. Интересно, что замена $A_0 = d/ds + B_0$ и $A_i = B_i$ при $i > 0$ превращает эти уравнения в уравнения п. 3.9.

Теперь опишем граничные условия. Пусть τ_1, τ_2, τ_3 лежат в фиксированной картановской подалгебре $\mathfrak{h} \subset \mathfrak{g}$, которая является их общим централизатором. Тогда набор $B_0 = 0, B_i = \tau_i$ является решением выписанной системы. Пространство операторов $d/ds + B_0, B_1, B_2, B_3$, близких к указанному выше решению в некоторой C^1 -норме с экспоненциальным весом [КЗ], обладает корректно определенным эрмитовым произведением (в приведенной выше четырехмерной интерпретации это просто эрмитово произведение пространства L^2).

Присоединенное действие группы G -значных гладких функций $g(s)$ на четверках операторов $d/ds + B_0, B_1, B_2, B_3$ сохраняет гиперкэлерову структуру (в четырехмерной интерпретации это группа $Sp(1)$ -инвариантных калибровочных преобразований).

Легко видеть, что в этом случае гиперкэлерова редукция дает конечномерное многообразие. Действительно, калибровочным преобразованием оператор $d/ds + B_0$ можно привести к виду d/ds , после чего останется система обыкновенных дифференциальных уравнений относительно B_1, B_2, B_3 с действием конечномерной группы G .

5.4. Вычисление комплексной структуры является отдельной задачей. Для этого необходимо выбрать комплексную структуру в пространстве связностей и переписать уравнения отображения моментов в виде комплексной и вещественной части, как это было сделано в п. 3.9:

$$\begin{aligned} \frac{d\beta}{ds} + [\alpha, \beta] &= 0, \\ \frac{d}{ds}(\alpha - \alpha^*) + [\alpha, \alpha^*] + [\beta, \beta^*] &= 0, \end{aligned}$$

где $\alpha = (B_0 + iB_1)(s)$ и $\beta = (B_2 + iB_3)(s)$.

Затем следует убедиться, что при выполнении граничных условий отображение

$$(\alpha, \beta) \mapsto \beta(0) \in \mathfrak{g}^c$$

отождествляет пространство классов эквивалентности решений указанной системы по модулю действия группы G с орбитой элемента $\tau_2 + i\tau_3 \in \mathfrak{g}^c$ относительно присоединенного действия. Если группа G^c полупроста, то эта орбита изоморфна коприсоединенной орбите. Само доказательство основано на теореме Дональдсона [D1].

5.5. Заметим, что произвол в выборе отображения моментов на этот раз проявился в граничных условиях. Частично он состоит в выборе элемента $\tau_2 + i\tau_3$, т. е. в выборе конкретной коприсоединенной орбиты. Дополнительный произвол в выборе τ_1 в картановской подалгебре дает нам целое семейство гиперкэлэровых метрик.

Следует также заметить, что хотя для явного вычисления метрики необходимо решать нелинейные уравнения п. 5.3, все же эти уравнения являются разновидностью *уравнений Нама*, общее решение которых описывается в терминах геометрии якобиана алгебраической кривой [H3]. Ясно, что это дает значительно больше, чем просто теорема существования.

6. ПРЕДСТАВЛЕНИЯ ФУНДАМЕНТАЛЬНОЙ ГРУППЫ ПОВЕРХНОСТИ

6.1. Если Σ — компактная ориентируемая поверхность, то ее фундаментальная группа обладает внутренней симплектической природой [G]. В частности, если группа Ли G обладает инвариантным скалярным произведением на своей алгебре Ли $\mathfrak{g}$, то многообразие модулей неприводимых представлений группы $\pi_1(\Sigma)$ в группе G является симплектическим многообразием. Многообразие модулей описывается как $\text{Hom}^{\text{irr}}(\pi_1(\Sigma), G)/G$, где G действует сопряжениями. Касательное пространство к данному представлению отождествляется с группой когомологий $H^1(\pi_1(\Sigma); \mathfrak{g})$. Билинейная форма на $\mathfrak{g}$ индуцирует кососимметрическое спаривание со значениями в пространстве $H^2(\pi_1(\Sigma))$, которое порождается фундаментальным классом.

Если группа G компактна, то выбор комплексной структуры на поверхности Σ превращает многообразие модулей представлений в кэлэрово многообразие. Согласно теореме Нарасимхана и Шарди [NS], его голоморфная структура совпадает со структурой

пространства стабильных голоморфных расслоений на римановой поверхности Σ .

В случае комплексной полупростой группы G^c многообразие модулей

$$\text{Hom}^{\text{irr}}(\pi_1(\Sigma), G^c)/G^c$$

имеет естественную структуру комплексного симплектического многообразия и в действительности обладает гиперкэлеровой метрикой, также определяемой выбором комплексной структуры на поверхности Σ .

6.2. Эта метрика возникает в результате бесконечномерной гиперкэлеровой редукции.

Рассмотрим аффинное пространство $\mathcal{A}^c$ всех G^c -связностей в главном G^c -расслоении P с фиксированной редукцией к максимальной компактной группе G над поверхностью Σ . Тогда на каждом ассоциированном векторном расслоении возникает метрика. Касательное пространство к $\mathcal{A}^c$ отождествляется с пространством 1-форм $\alpha \in \Omega^1(\Sigma; \text{ad } P \otimes \mathbb{C})$, где $\text{ad } P$ обозначает расслоение, ассоциированное с присоединенным представлением группы G . Если обозначить через $\text{Tr}(AB)$ скалярное произведение на $\mathfrak{g}$, то на пространстве $\mathcal{A}^c$ имеется комплексная симплектическая форма

$$\omega(\alpha, \beta) = \int_{\Sigma} \text{Tr}(\alpha \wedge \beta).$$

Зафиксировав комплексную структуру на Σ , мы можем разложить форму $\alpha = \alpha^{1,0} + \alpha^{0,1}$ в сумму форм различного типа и ввести вещественную симплектическую форму — кэлерову форму метрики

$$\|\alpha\|^2 = \int_{\Sigma} \text{Tr}(\alpha^{1,0} \wedge \alpha^{1,0*}) - \text{Tr}(\alpha^{0,1} \wedge \alpha^{0,1*}).$$

Таким образом, пространство $\mathcal{A}^c$ превращается в бесконечномерное гиперкэлерово многообразие, а группа $\mathcal{G}$ калибровочных преобразований действует на $\mathcal{A}^c$, сохраняя гиперкэлерову структуру.

6.3. Согласно основополагающему наблюдению Атьи и Ботта [AB], отображение моментов этого гиперкэлерового действия выражается через кривизну связности. Если мы запишем его в виде вещественного и комплексного отображений моментов, то получим

$$\begin{aligned}\mu_c(A) &= F_A \in \Omega^2(\Sigma; \text{ad } P \otimes \mathbb{C}), \\ \mu_r(A) &= F' - F'' \in \Omega^2(\Sigma; \text{ad } P),\end{aligned}$$

где F' и F'' обозначают кривизны единственных G -связностей ∇' и ∇'' , таких, что

$$(\nabla')^{1,0} = \nabla_A^{1,0}, \quad (\nabla'')^{0,1} = \nabla_A^{0,1}.$$

Следовательно, $\mu_c^{-1}(0)$ является пространством плоских связностей в главном G^c -расслоении.

Теорема Дональдсона [D2] и Корлетта [Co] утверждает, что каждая неприводимая G^c -орбита содержит решение уравнения $\mu_r = 0$, единственное по модулю действия группы G ; поэтому комплексная структура на гиперкэлеровой редукции отождествляется с комплексной структурой пространства плоских неприводимых связностей по модулю комплексной калибровочной эквивалентности. Но последнее пространство очевидным образом отождествляется с многообразием модулей представлений $\text{Hom}^{\text{irr}}(\pi_1(\Sigma); G^c)/G^c$.

6.4. Доказательство теоремы Дональдсона и Корлетта основано на иной интерпретации уравнений отображения моментов. Для того чтобы взять гиперкэлерову редукцию, надо зафиксировать метрику на главном расслоении, после чего изучать классы эквивалентности плоских связностей. С другой стороны, мы можем зафиксировать плоскую связность, после чего искать такую метрику, чтобы выполнялось равенство $F' = F''$. Метрика, согласованная с G -структурой, — это сечение ассоциированного плоского G^c/G -расслоения, и уравнение, которое нам необходимо решить, равносильно тому, что это сечение *гармоническое*. Неположительность кривизны G^c/G влечет за собой в силу теоремы Илса—Сэмпсона [ES] существование решения.

Техника, необходимая для строгого построения многообразия модулей и метрики на нем, включает в себя применение соболевских пространств для компактных многообразий и является совершенно стандартной (см. [H5]).

6.5. Построенное $4(g-1) \dim G$ -мерное гиперкэлерово многообразие, где g обозначает род римановой поверхности Σ , обладает многими свойствами, характерными для случая простой метрики Егучи—Хансона. Комплексная структура на многообразии модулей представлений — структура аффинного многообразия, точно так же как для метрики Егучи—Хансона общая комплексная структура есть структура аффинной квадратики. С другой стороны, одна структура из двумерной сферы комплексных структур, порожденной

структурами I, J, K , не является аффинной, но содержит в качестве плотного открытого множества кокасательное расслоение к многообразию модулей стабильных G^c -расслоений аналогично тому, как кокасательное расслоение к CP^1 является одной из комплексных структур метрики Егучи—Хансона. Гиперкэлерово многообразие с этой последней комплексной структурой отождествляется с многообразием модулей стабильных расслоений Хиггса (или стабильных пар) на римановой поверхности Σ [Н5]. В этом случае компактность поверхности налагает граничные условия на уравнения отображения моментов, найти явное решение которых не удастся.

7. РАЦИОНАЛЬНЫЕ ОТОБРАЖЕНИЯ

7.1. Конструкции гиперкэлеровых многообразий, предложенные в качестве примеров в разд. 4, 5 и 6, основаны на уравнениях в размерностях 0, 1 и 2. В $\mathbf{R}^3$ также имеется система уравнений, которая задает гиперкэлерово многообразие модулей. Это уравнения Богомольного для магнитных монополей.

Рассмотрим тривиальное главное G -расслоение P над $\mathbf{R}^3$, где G — компактная группа Ли. Пусть $\mathcal{A}$ — пространство связностей в P . Положим

$$M = \mathcal{A} \times \Omega^0(\mathbf{R}^3; \text{ad } P).$$

Точка пространства M — это пара (A, ϕ) , где A — связность, а ϕ — сечение присоединенного расслоения, т. е. поле Хиггса.

Касательное пространство отождествляется с множеством пар

$$(\alpha, \phi) \in \Omega^1(\mathbf{R}^3; \text{ad } P) \times \Omega^0(\mathbf{R}^3; \text{ad } P),$$

обладающим очевидной гиперкэлеровой структурой — касательный вектор можно записать в виде

$$\phi + \alpha_1 i + \alpha_2 j + \alpha_3 k.$$

На пространстве M действует группа $\mathcal{G}$ калибровочных преобразований.

Если наложить граничные условия (условия близости с модельной конфигурацией, как в разд. 5) на (A, ϕ) и на $\mathcal{G}$, то эрмитово произведение пространства L^2 определяет гиперкэлерову метрику на M , инвариантную относительно действия группы $\mathcal{G}$. Теперь можно применять гиперкэлерову редукцию.

7.2. Обозначим через ∇_i ковариантную производную относительно связности A в i -м направлении. Тогда уравнения отображения моментов примут вид (ср. с п. 3.9)

$$[\phi, \nabla_1] + [\nabla_2, \nabla_3] = 0,$$

$$[\phi, \nabla_2] + [\nabla_3, \nabla_1] = 0,$$

$$[\phi, \nabla_3] + [\nabla_1, \nabla_2] = 0,$$

или, в более компактном виде, с использованием хodgeва оператора $*$,

$$F_A = *\nabla_A \phi,$$

т. е. получится уравнение Богомольного.

В простейшем случае $G = SU(2)$ граничные условия обычно выбирают в виде $\|\phi\| \rightarrow 1$ при $r \rightarrow \infty$. Здесь как параметр решения возникает целое число k — степень отображения сферы большого радиуса в $\mathbb{R}^3$ в единичную сферу в $\mathfrak{g} \cong \mathbb{R}^3$. Это число называется *зарядом* решения.

7.3. В случае $G = SU(2)$ дополнительная теорема, доказанная Дональдсоном [D1], отождествляет полученное многообразие модулей с пространством базированных рациональных отображений $f : \mathbb{CP}^1 \rightarrow \mathbb{CP}^1$ степени k . В действительности для отождествления необходимо выбрать направление u и решать проблему рассеяния для обыкновенного дифференциального уравнения $(\nabla_u + i\phi)s = 0$.

Рациональное отображение f степени k , переводящее ∞ в 0 , можно записать в виде

$$f(z) = \frac{a_0 + \dots + a_{k-1}z^{k-1}}{b_0 + b_1z + \dots + z^k} = \frac{p(z)}{q(z)},$$

где многочлены $p(z)$ и $q(z)$ не имеют общих делителей. Пространство таких отображений, очевидно, является комплексным многообразием вещественной размерности $4k$. Как и на всяком гиперкэлеровом многообразии, на нем должна быть комплексная симплектическая форма. Она строится следующим образом.

Прежде всего разложим знаменатель в произведение: $q(z) = (z - \beta_1) \dots (z - \beta_k)$. Поскольку $p(z)$ и $q(z)$ не имеют общих делителей, $p(\beta_i) \neq 0$. Следовательно, форма

$$\omega = \sum_{i=1}^k \frac{dp(\beta_i) \wedge d\beta_i}{p(\beta_i)}$$

является голоморфной симплектической формой на пространстве рациональных отображений. Она плоская по отношению к гиперкэлеровой метрике.

7.4. Строгая конструкция этих метрик требует привлечения дополнительной техники, которая была разработана К. Х. Таубсом. Сводка его результатов и свойств полученных метрик содержится в [АН]. Одно из свойств позволяет отличить это семейство от предыдущих. Отождествление с пространством рациональных отображений требует выбора направления в $\mathbf{R}^3$. С другой стороны, сами уравнения Богомольного $SO(3)$ -инвариантны. Таким образом, группа $SO(3)$ транзитивно действует на двумерной сфере комплексных структур на многообразии модулей, которые, следовательно, все эквивалентны. Этим свойством также обладает метрика Taub-NUT, твисторное пространство которой, описанное в п. 2.4, очевидным образом наследует действие группы $SU(2)$ с $\mathbf{CP}^1$.

Твисторные пространства для монопольных метрик могут быть точно описаны (см. [АН]). Кроме того, в случаях $k = 1$ и 2 на них явно вычислены метрики: в первом случае это плоская метрика на $S^1 \times \mathbf{R}^3$, а во втором случае метрика описывается в терминах эллиптических интегралов [АН].

8. ГРУППЫ ПЕТЕЛЬ

8.1. Пусть G — компактная группа Ли. Обозначим через $LG = \text{Map}(S^1; G)$ пространство гладких отображений из окружности в группу G . Хорошо известно, что фактор LG/G по подгруппе постоянных отображений обладает структурой бесконечномерного симплектического многообразия [PS]. Кроме того, он обладает кэлеровой метрикой. Симплектическая форма получается сдвигом косимметрической формы, определенной на касательном пространстве к единичному отображению,

$$\omega(f, g) = \int_{S^1} \text{Tr}(fg') d\theta,$$

где $f, g : S^1 \rightarrow \mathfrak{g}$.

Ясно, что то же определение работает и в случае комплексной полупростой группы G^c , т.е. комплексное многообразие LG^c/G^c обладает естественной голоморфной симплектической формой. Дональдсон [D3] показал, как можно снабдить это пространство естественной гиперкэлеровой метрикой.

8.2. Рассмотрим тривиальное главное G -расслоение над единичным диском $D \subset \mathbb{C}$ и пространство A^c связностей на ассоциированном G^c -расслоении P^c , гладких вплоть до границы. Таким образом, ситуация сходна с рассматривавшейся в разд. 6, но теперь имеется нетривиальная граница — окружность S^1 . Однако если мы рассмотрим группу калибровочных преобразований $\mathcal{G}$, тождественных на границе, то гиперкэлеровы отображения моментов окажутся такими же, как и в п. 6.2:

$$\begin{aligned}\mu_c(A) &= F_A \in \Omega^2(D; \text{ad } P \otimes \mathbb{C}), \\ \mu_\tau(A) &= F' - F'' \in \Omega^2(D; \text{ad } P).\end{aligned}$$

8.3. Для решения уравнений отображения моментов мы опять воспользуемся формулировкой в терминах гармонических отображений. В нашем случае все плоские связности тривиальны, так как диск односвязен; поэтому решения уравнений — это в точности гармонические отображения $f : D \rightarrow G^c/G$. Результаты Р. Гамильтона [На] доказывают разрешимость задачи Дирихле. Таким образом, отображение f , заданное на границе, однозначно продолжается до гармонического отображения всего диска.

На языке связностей это означает, что при данном граничном значении существует единственная по модулю калибровочных преобразований, сохраняющих граничные значения, плоская G^c -связность на D , являющаяся решением уравнений отображения моментов. Однако все такие связности на окружности калибровочно эквивалентны тривиальной связности, группа автоморфизмов которой равна G^c . Поэтому гиперкэлерова редукция совпадает с

$$\text{Map}(S^1; G^c)/G^c.$$

8.4. Голоморфная форма ω из разд. 6 при ограничении на пространство плоских связностей $\mu_c^{-1}(0)$ дает

$$\omega(\alpha, \beta) = \int_D \text{Tr}(\alpha \wedge \beta).$$

Но если α и β — касательные векторы к пространству плоских связностей в точке A , то $\alpha = d_A \varphi$ и $\beta = d_A \psi$ для некоторых $\varphi, \psi \in \Omega^0(D; \text{ad } P^c)$, а следовательно,

$$\begin{aligned}\omega(\alpha, \beta) &= \int_D \text{Tr}(d_A \varphi \wedge d_A \psi) = \\ &= \int_{S^1} \text{Tr}(\varphi d_A \psi),\end{aligned}$$

что, очевидно, совпадает с определенной выше кососимметрической формой на граничных значениях отображений $f : D \rightarrow G^c$.

8.5. В завершение заметим, что в этом примере уравнения отображения моментов можно записать в более привычной форме п. 3.9 с помощью подстановки. Полагая

$$\nabla = (\nabla' + \nabla'')/2, \quad \phi = (\nabla' - \nabla'')/2$$

и записывая $\phi = \phi_1 dx_1 + \phi_2 dx_2$, переписываем эти уравнения так:

$$[\nabla_1, \nabla_2] + [\phi_1, \phi_2] = 0,$$

$$[\nabla_1, \phi_2] + [\nabla_2, \phi_1] = 0,$$

$$[\nabla_1, \phi_1] + [\phi_2, \nabla_2] = 0.$$

9. ЗАКЛЮЧЕНИЕ

9.1. Гиперкэлерова редукция позволяет построить огромное количество гиперкэлеровых многообразий, в действительности столь огромное, что возникает новая задача — навести в них порядок. Это очень важно, ведь даже среди приведенных здесь примеров много соотношений — гиперкэлеровы метрики могут появляться на одном и том же пространстве с помощью различных конструкций.

Коприсоединенные орбиты Кронхаймера — один из таких примеров. Конструкция, приведенная в разд. 6, — редукция бесконечномерного пространства — основана на обыкновенных дифференциальных уравнениях. С другой стороны, Д. Бернс [Bu] предложил алгебраическую твисторную конструкцию (на примере которой очень хорошо видны трудности доказательства полноты метрики) для метрик на тех же самых пространствах. Помимо этого, например, компоненты множества неподвижных точек относительно действия окружности на комплексной группе петель из разд. 8 совпадают с коприсоединенными орбитами. Наконец, некоторые из этих пространств (как видно на примере метрики Егучи—Хансона) могут быть получены конечномерной гиперкэлеровой редукцией. Доказать изометрию оказывается не всегда просто.

Другие примеры дают многообразия модулей монополей. С одной стороны, гиперкэлерова метрика на них вводится с помощью бесконечномерной редукции, основанной на уравнениях Богомольного в размерности 3. С другой стороны, преобразование Нама позволяет определить метрику с помощью уравнений Нама в размерности 1. Согласно результатам типа результатов Накадзимы [N],

две полученные метрики совпадают. Аналогична ситуация с многообразиями модулей инстантонов: здесь результаты работ [BvB] и [Ma] доказывают совпадение метрик, полученных двумя различными способами.

9.2. Операция редукции может быть определена и для более широкого класса многообразий, связанных с кватернионами. Мы уже упоминали *гиперкомплексные* многообразия. Помимо них есть еще класс *кватернионных кэлеровых* многообразий, т.е. римановых многообразий с группой голономии $Sp(n).Sp(1)$, например, HP^n или класс *кватернионных* многообразий, обладающих $GL(n, \mathbb{H}).Sp(1)$ -связностью без кручения [S]. Для всех этих многообразий определена операция редукции $[GL, J]$, однако количество известных примеров здесь меньше, чем в гиперкэлеровом случае.

9.3. Последний вопрос, которого мы коснемся, относится к компактным гиперкэлеровым многообразиям, таким, как $K3$ -поверхности, в случае которых пока можно лишь сослаться на теорему существования. Возможно ли их получить гиперкэлеровой редукцией? Например, как гиперкэлерову редукцию конечномерного векторного пространства?

Ответ на этот вопрос отрицательный. Существуют компактные кэлеровы многообразия, являющиеся кэлеровой симплектической редукцией, например проективное пространство (см. п. 3.5), однако специфические свойства кривизны гиперкэлерова многообразия являются препятствием к существованию компактной гиперкэлеровой редукции. Мы покажем это в простейшем, четырехмерном случае.

Предположим, что компактное гиперкэлерово многообразие M^4 является гиперкэлеровой редукцией линейного гиперкэлерова действия группы G в пространстве $\mathbb{H}^n$. Пусть μ — гиперкэлерово отображение моментов. Тогда $\mu^{-1}(\zeta)$ является главным G -расслоением P на M . Плоская метрика на $\mathbb{H}^n$ индуцирует G -инвариантную метрику на P , причем ортогональные дополнения к направлениям орбит определяют связность на P . Для каждой комплексной структуры на $\mathbb{H}^n$ (и, следовательно, на M) ассоциированное главное G^c -расслоение изоморфно $\mu_c^{-1}(\alpha)$, т.е. голоморфно. Это означает, что связность согласована со всеми тремя комплексными структурами I, J и K , а значит (см. [AHS]), является решением уравнения антиавтодуальности Янга—Миллса на M . Согласно конструкции Атьи—Уорда, такому решению соответствует голоморфное главное рассло-

ские на твисторном пространстве Z многообразия M , которое, как и следовало ожидать, совпадает с $\tilde{\mu}^{-1}(\zeta)$.

Рассмотрим вложение пространства $\tilde{\mu}^{-1}(\zeta)$ в твисторное пространство $\mathbb{C}^{2n}(1)$, соответствующее вложению $P = \mu^{-1}(\zeta) \subset \mathbb{H}^n$. Это вложение задает эквивариантное сечение расслоения $\mathbb{C}^{2n}(1)$ на главном G^c -расслоении над Z , которое [H2] посредством преобразования Пенроуза интерпретируется как решение ψ некоторого дифференциального уравнения на M — твисторного уравнения, спаренного с антиавтотдуальной связностью Янга—Миллса. К такому объекту можно применить формулу Вайтценбёка, из которой на компактном многообразии следует справедливость некоторой теоремы об обращении в нуль. В нашем случае, поскольку тензор Риччи, а следовательно, и скалярная кривизна гиперкэлэровой многообразия равны нулю, эта теорема об обращении в нуль [H2] означает, что сечение ψ плоское. Это, в свою очередь, означает, что отображение $P \subset \mathbb{H}^n$ постоянно в горизонтальном направлении, что противоречит его инъективности.

9.4. Остается надежда на бесконечномерную редукцию. Так, например, многообразия модулей инстантонов наделяются гиперкэлэровой структурой с помощью интерпретации уравнений Янга—Миллса как уравнений отображения моментов, но эти многообразия либо некомпактны, либо имеют особенности, соответствующие приводимым связностям. Однако возможно, что, допуская связности с особенностями, мы получим более регулярные многообразия модулей. Таким образом, остается вероятность того, что удастся добавить КЗ-поверхности к списку гиперкэлэровых редукций¹⁾. Это не означает, что нам удастся в явном виде отыскать на них метрику (работа [AH] показывает, что явный вид дается большим трудом), но мы получим более глубокое понимание универсальности геометрии, связанной с кватернионами.

ЛИТЕРАТУРА

- [AHS] Atiyah M.F., Hitchin N.J., Singer I.M. Self-duality in four-dimensional Riemannian geometry, Proc. Roy. Soc. Lond., A362 (1978), 425–461.

¹⁾ Согласно теореме Мукая, каждую КЗ-поверхность можно представить как многообразие модулей антиинстантонов на ней самой. При этом метрика Вейля—Петерсона является гиперкэлэровой и получается гиперкэлэровой редукцией из пространства связностей.

- [AB] Atiyah M. F., Bott R. The Yang-Mills equations over Riemannian surfaces, *Phil. Trans. Roy. Soc. Lond.*, **A308** (1982), 523-615.
- [AH] Atiyah M. F., Hitchin N. J. *The Geometry and Dynamics of Magnetic Monopoles*, Princeton University Press, Princeton, 1988.
- [Bea] Beauville A. Variétés Kählériennes dont la 1ère classe de Chern est nulle, *J. Diff. Geometry*, **18** (1983), 755-782.
- [Be] Besse A. *Einstein Manifolds*, Springer-Verlag, Berlin, 1987.
- [Bo] Bochner S. Curvature in Hermitian metric, *Bull. Amer. Math. Soc.*, **53** (1947), 179-195.
- [Bu] Burns D. Some examples of the twistor construction, in: *Contributions to several complex variables: in honor of Wilhelm Stoll* (eds. A. Howard and P.-M. Wong), Vieweg, Braunschweig, 1986.
- [BvB] Braam P. J., van Baal P. Nahm's transformation for instantons, *Comm. Math. Phys.*, **122** (1989), 267-280.
- [Ca] Calabi E. Métriques kählériennes et fibrés holomorphes, *Ann. Éc. Norm. Sup.*, **12** (1979), 269-294.
- [Co] Corlette K. Flat G -bundles with canonical metrics, *J. Diff. Geometry*, **28** (1988), 351-382.
- [D1] Donaldson S. K. Nahm's equations and the classification of monopoles, *Comm. Math. Phys.*, **96** (1984), 387-407.
- [D2] Donaldson S. K. Twisted harmonic maps and the self-duality equations, *Proc. Lond. Math. Soc.*, **55** (1987), 127-131.
- [D3] Donaldson S. K. Boundary value problems for Yang-Mills fields, Preprint (Oxford), 1991.
- [EH] Eguchi T., Hanson A. J. Asymptotically flat self-dual solutions to Euclidean gravity, *Phys. Lett.*, **74B** (1978), 249-251.
- [ES] Eells J., Sampson J. H. Harmonic mappings of Riemannian manifolds, *Amer. J. Math.*, **86** (1964), 109-160.
- [GL] Galicki K., Lawson H. B. Quaternionic reduction and quaternionic orbifolds, *Math. Ann.*, **282** (1988), 1-21.
- [G] Goldman W. M. The symplectic nature of fundamental groups of surfaces, *Adv. in Math.*, **54** (1984), 200-225.
- [Gr] Gray A. A note on manifolds whose holonomy group is a subgroup of $Sp(n).Sp(1)$, *Michigan Math. J.*, **16** (1969), 125-128.
- [Ha] Hamilton R. S. *Harmonic Maps of Manifolds with Boundary*, *Lect. Notes in Math.*, vol. 471, Springer-Verlag, Berlin, 1975.
- [H] Hamilton W. R. Letter to P. G. Tait, October 15th, 1958.
- [H1] Hitchin N. J. Polygons and gravitons, *Math. Proc. Camb. Phil. Soc.*, **85** (1979), 465-476.
- [H2] Hitchin N. J. Linear field equations on self-dual spaces, *Proc. Roy. Soc. Lond.*, **A370** (1980), 173-191.
- [H3] Hitchin N. J. On the construction of monopoles, *Comm. Math. Phys.*, **89** (1983), 145-190.
- [H4] Hitchin N. J. Twistor construction of Einstein metrics, in: *Global Riemannian geometry* (eds. T. J. Willmore and N. J. Hitchin), Ellis Horwood, Chichester, 1984.

- [H5] Hitchin N. J. The self-duality equations on a Riemannian surface, *Proc. Lond. Math. Soc.*, **55** (1987), 59–126.
- [HKLR] Hitchin N. J., Karlhede A., Lindström U., Roček M. Hyperkähler metrics and supersymmetry, *Comm. Math. Phys.*, **108** (1987), 535–559.
- [J] Joyce D. The hypercomplex quotient and the quaternionic quotient, *Math. Ann.*, **290** (1991), 323–340.
- [Ki] Kirwan F. *Cohomology of Quotients in Symplectic and Algebraic Geometry*, Princeton Univ. Press, Princeton, 1985.
- [K1] Kronheimer P. B. Instantons gravitationnelles et singularités de Klein, *C. R. Acad. Sci. Paris Sér. 1 Math.*, **303** (1986), 53–55.
- [K2] Kronheimer P. B. The construction of ALE spaces as hyperkähler quotients, *J. Diff. Geometry*, **29** (1989), 665–683.
- [K3] Kronheimer P. B. A hyperkählerian structure on coadjoint orbits of a semisimple Lie group, *J. Lond. Math. Soc.*, **42** (1990), 193–208.
- [Ma] Maciocia A. Metrics on the moduli spaces of instantons over Euclidean 4-space, *Comm. Math. Phys.*, **135** (1991), 467–482.
- [McK] McKay J. Graphs, singularities and finite groups, in: *Proc. Sympos. Pure Math.*, Vol. 37, Amer. Math. Soc., Providence, 1980, 183–186.
- [N] Nakajima H. Monopoles and Nahm's equations, Preprint (Tokyo), 1991.
- [NS] Narasimhan M. S., Seshadri C. S. Stable and unitary vector bundles on a compact Riemann surface, *Ann. of Math.*, **82** (1965), 540–567.
- [P] Penrose R. Nonlinear gravitons and curved twistor theory, *Gen. Relativ. Grav.*, **7** (1976), 31–52. [Имеется перевод в сб.: *Твисторы и калибровочные поля*. — М.: Мир, 1983.]
- [PS] Pressley A., Segal G. B. *Loop Groups*, Oxford University Press, Oxford, 1986. [Имеется перевод: Пресли А., Сигал Г., *Группы петель*. — М.: Мир, 1990].
- [S] Salamon S. Differential geometry of quaternionic manifolds, *Ann. Sci. Éc. Norm. Sup. Paris*, **19** (1986), 31–55.
- [Y] Yau S. T. On Calabi's conjecture and some new results in algebraic geometry, *Proc. Natl. Acad. Sci. USA*, **74** (1977), 1798–1799.

NIGEL HITCHIN

Math. Institute
24–29 St. Giles
GB–OXFORD OX1 3LB
(England)

E-mail: hitchin@maths.ox.ac.uk

НАКРЫТИЯ АЛГЕБРАИЧЕСКИХ КРИВЫХ

Жан-Пьер Серр

1. ВВЕДЕНИЕ

1.1. Постановка задачи. Пусть k — алгебраически замкнутое поле. Обозначим через C алгебраическую кривую над k . Предположим, что она неприводимая и гладкая. Тогда

$$C = \overline{C} - S,$$

где $\overline{C}$ является гладкой проективной кривой, а S — конечное подмножество в $\overline{C}(k)$.

Каковы (неразветвленные) накрытия Галуа кривой C ?

Этот вопрос можно уточнить двумя способами:

- (a) Пусть дана конечная группа G . При каком условии существует связное накрытие Галуа $C' \rightarrow C$ с группой Галуа G ?
- (b) Пусть $\pi_C = \pi_1^{\text{alg}}(C, x)$ — (алгебраическая) фундаментальная группа кривой C относительно базовой точки x . Эта группа проконечна. Требуется определить ее структуру.

Замечания. 1) Заметим, что вопрос (b) — более конкретный, чем вопрос (a): конечная группа G удовлетворяет условию п. (a) тогда и только тогда, когда она изоморфна факторгруппе группы π_C .

2) При замене основного поля k его алгебраически замкнутым расширением k' группа π_C не изменяется, если характеристика равна 0. Это не так, если характеристика p больше 0: в этом случае существуют непостоянные семейства накрытий. В любом случае, однако, рассуждение со специализацией показывает, что конечные факторгруппы группы π_C над k и над k' одинаковы. Вопрос (a) тем самым «не зависит» от выбора основного поля (разумеется, если предполагается, что оно алгебраически замкнуто).

1.2. Комплексный случай. Если $k = \mathbb{C}$, то теорема существования Римана утверждает, что каждое конечнолистное (в топологическом смысле) накрытие поверхности $C(\mathbb{C})$ обладает ровно одной алгебраической структурой, совместимой с проекцией на $C(\mathbb{C})$.

(см., например, [5, ch. VI]). Это позволяет ответить на сформулированные выше вопросы (а) и (б). Ответ на вопрос (б) следующий: пусть g — род кривой $\bar{C}$ и $s = |S|$. Тогда группа π_C есть *проконечное пополнение* (т. е. пополнение в топологии подгрупп конечного индекса) группы, заданной $2g + s$ образующими a_i, b_i ($i = 1, \dots, g$) и c_j ($j = 1, \dots, s$), связанными соотношением

$$a_1 b_1 a_1^{-1} b_1^{-1} \dots a_g b_g a_g^{-1} b_g^{-1} c_1 \dots c_s = 1.$$

Если $S \neq \emptyset$ (т. е. если C аффинна), эта группа является свободной группой ранга $2g + s - 1$. В этом случае ответ на вопрос (а) п. 1.1 оказывается очень простым: конечная группа G удовлетворяет рассматриваемому условию тогда и только тогда, когда ее можно задать $2g + s - 1$ образующими.

(Теория Римана на самом деле дает более точный результат: элементы $c_1, \dots, c_s$ могут быть выбраны таким образом, что они порождают группы инерции точек множества S .)

1.3. Случай характеристики 0. Если характеристика поля k равна 0, то принцип Лефшеца и рассуждения со специализацией показывают, что *результаты п. 1.2 остаются без изменений*.

Заметим, что, хотя результаты выглядят алгебраическими, единственное имеющееся доказательство опирается на трансцендентную теорию для случая $k = \mathbb{C}$ (существенным моментом является доказательство того, что на компактной римановой поверхности имеется достаточно много мероморфных функций).

Попытка алгебраического подхода к классификации накрытий была сделана Вейлем в 1938 г. (см. [21, р. 84–86], а также Nori V. *Comp. Math.*, **33** (1976), 29–41). Его отправной точкой было следующее соображение: каждому накрытию Галуа $C' \rightarrow C$ с группой Галуа G соответствуют (через линейные представления группы G) *векторные расслоения* над C , дискретные суммы и тензорные произведения которых связаны алгебраическими уравнениями. Если бы удалось достаточно точно описать структуру *пространств модулей* векторных расслоений, то можно было бы выделить в нем алгебраические элементы и восстановить по ним накрытия Галуа. Этот подход «по Танаке» интересен, но до их пор не привел к конкретным результатам; даже элементарный случай $g = 1$, $s = 0$ представляется слишком сложным для исследования этим методом.

1.4. Случай характеристики $p > 0$. Этот случай долгое время оставался неизученным. Вот что писал в 1946 г. Вейль в работе [22]:

«... прежде чем приступить к описанию расширений числовых полей через их локальные свойства, было бы удобно решить аналогичную проблему, уже достаточно трудную, относительно алгебраических функций одной переменной над конечным основным полем, а именно, распространить на эти функции теоремы существования Римана. В частности, играет ли модулярная группа, структурой которой определяются поля функций одной комплексной переменной, разветвленных в трех точках, ту же роль в случае конечного основного поля, по крайней мере для расширений степени, взаимно простой с характеристикой? Возможно, все такие вопросы можно изучать с помощью единого метода, который позволил бы для каждого установленного (например, топологическими методами) результата в характеристике 0 вывести соответствующий результат в характеристике p . Открытие такого метода означало бы очень важное продвижение ... »

В этом отрывке утверждается, что:

- (i) для групп порядка, взаимно простого с p , теория такая же, как для случая характеристики 0;
- (ii) должна существовать возможность перехода от характеристики 0 к характеристике p (и обратно).

Утверждение (i) было уточнено в 1956 г. Абъянкарм [1] в виде следующей гипотезы:

- (i') Пусть π'_C — наибольшая факторгруппа группы π_C , порядок которой (как проконечной группы) взаимно прост с p . Тогда π'_C изоморфна группе, соответствующей кривой над C с теми же инвариантами g и s .

2. ТЕОРЕМЫ ГРОТЕНДИКА И ГИПОТЕЗЫ АБЪЯНКРА

2.1. Теоремы Гротендика. Одним из первых достижений теории схем Гротендика было доказательство в 1958 г. гипотезы (i') п. 1.4 (см. [7, р. 182–27], а также [8, р. 392]); метод состоит в переходе к характеристике 0, как и было предсказано Вейлем.

Точнее, выберем в поле вычетов k полное кольцо дискретного нормирования A , поле частных K которого имеет характеристику 0. Доказательство происходит в три этапа:

(2.1.1) Можно «поднять» проективную кривую $\bar{C}$ над k до гладкой проективной схемы $\bar{C}_A$ над A (которая дает кривую $\bar{C}_K$ над K , обладающую хорошей редукцией).

Этот результат почти очевиден в случае $g < 5$, но неочевиден для больших значений g . Гротендик доказал его, перейдя от $\bar{C}$ к формальной схеме и показав, что эта схема алгебраическая, поскольку на ней имеется обильное расслоение ранга 1 (а именно, расслоение, определенное над точкой). В доказательстве используются теоремы сравнения типа «формальная теория $\Leftrightarrow$ алгебраическая теория», т. е. «формальный вариант GAGA».

(2.1.2) После того, как $\bar{C}_A$ выбрана, пусть S_K — накрытие множества S над $\bar{C}_A(A) = \bar{C}_K(K)$. Если $C' \rightarrow C$ — накрытие Галуа кривой C с группой Галуа G , которое является «ручным»¹⁾ во всех точках множества S , то можно поднять C' единственным образом до накрытия схемы $C_K = S_K$: это опять-таки доказывается применением «формального варианта GAGA».

(2.1.3) Обратно, пусть $C'_K \rightarrow C_K = S_K$ — абсолютно неприводимое накрытие Галуа с группой Галуа G . Предположим, что порядок группы G взаимно прост с p . В этом случае можно доказать, что после замены A на более разветвленное кольцо накрытие C'_K «хорошо приводится», т. е. продолжается до накрытия схемы $\bar{C}_A = S_K$, которое определяет накрытие кривой C (в доказательстве используется лемма Аббяжара и теорема о чистоте накрытия).

Из этих результатов следует, что:

(*) Если p не делит $|G|$, то G -накрытия в характеристике p и в характеристике 0 одни и те же, см. (2.1.2) и (2.1.3).

Это очень неплохой результат, несмотря на то, что условие « p не делит $|G|$ » является весьма ограничительным: при $p = 2$ оно выполнено только для разрешимых групп, согласно Фейту и Томпсону.

(**) Если p делит $|G|$, но мы ограничиваемся накрытиями, «ручными» во всех точках множества S , то накрытий в характеристике p существует не больше (как правилó, строго меньше), чем в характеристике 0.

(Точную формулировку результатов (*) и (**)) в терминах специализации фундаментальной группы можно найти в [8].)

2.2. Случай полной кривой C . В этом случае $S = \emptyset$. Все накрытия C «ручные». Согласно утверждению (**), группа π_C является

¹⁾ В оригинале *modéré*, что соответствует термину *tame* в англоязычной литературе.

факторгруппой соответствующей фундаментальной группы в характеристике 0. *Какова эта факторгруппа?* Ее удастся определить только в случае $g = 0$ или 1, когда она коммутативна:

для $g = 0$ группа π_C равна $\{1\}$;

для $g = 1$ группа π_C изоморфна $\prod_{\ell \neq p} (\mathbb{Z}_\ell \times \mathbb{Z}_\ell)$ или $\mathbb{Z}_p \times \prod_{\ell \neq p} (\mathbb{Z}_\ell \times \mathbb{Z}_\ell)$ в зависимости от того, является кривая C суперсингулярной или обычной.

В случае $g \geq 2$ структурной теоремы не существует даже в виде гипотезы. Имеются только сведения о когомологиях. В частности, когомологии группы π_C изоморфны этальным когомологиям кривой C , совсем как в случае, когда «универсальное накрытие» кривой C стягиваемо (это доказывается с помощью спектральной последовательности Картана–Лере для накрытий, см. [14, с. 134, теорема 2.20]). Из этого, согласно [4, ехр. IX, X], вытекает, что

$$\mathrm{cd}_p(\pi_C) = 1 \text{ и } \dim H^1(\pi_C, \mathbb{Z}/p\mathbb{Z}) \leq g;$$

$$\mathrm{cd}_\ell(\pi_C) = 2, \dim H^1(\pi_C, \mathbb{Z}/\ell\mathbb{Z}) = 2g \text{ и } \dim H^2(\pi_C, \mathbb{Z}/\ell\mathbb{Z}) = 1 \text{ при } \ell \neq p.$$

(Напомним, что $\mathrm{cd}_p(\pi)$ означает *когомологическую p -размерность* проконечной группы π , см. [18, с. 23]. Утверждение $\mathrm{cd}_p(\pi_C) \leq 1$ эквивалентно тому, что силовские p -подгруппы группы π_C являются *свободными про- p -группами*; эти группы нетривиальны, согласно Рено [16, сог. 4.3.2]).

2.3. Случай аффинной кривой C : гипотеза Абъянкара.

В этом случае $S \neq \emptyset$. Фундаментальная группа над C является свободной группой ранга $2g + s - 1$ (см. п. 1.2).

Пусть группа G конечна. Обозначим через $p(G) = O^{p'}(G)$ подгруппу группы G , порожденную ее силовскими p -подгруппами. Группа $G/p(G)$ — наибольшая факторгруппа группы G , порядок которой взаимно прост с p . Если G является факторгруппой группы π_C , то, согласно Гротендику (см. (2.1.2)), $G/p(G)$ является факторгруппой группы, аналогичной группе π_C в характеристике 0. Иными словами, эта группа может быть задана $2g + s - 1$ образующими. В работе [1] Абъянкар предположил, что это условие является достаточным:

Гипотеза 2.3.1. *Конечная группа G является факторгруппой группы π_C тогда и только тогда, когда факторгруппа $G/p(G)$ может быть порождена $2g + s - 1$ элементами.*

Если эта гипотеза верна, то она дает ответ на вопрос (а) п. 1.1 (но не дает ответа на вопрос (b)).

Замечание. В этом случае когомологии группы π_C также совпадают с этальными когомологиями кривой C , откуда выводится, что

$$\mathrm{cd}_\ell(\pi_C) = 1 \text{ для всех } \ell,$$

$$\dim H^1(\pi_C, \mathbf{Z}/\ell\mathbf{Z}) = \begin{cases} \mathrm{Card}(k), & \text{если } \ell = p, \\ 2g + s - 1 & \text{в противном случае.} \end{cases}$$

В частности (см. [18, с. 79]) группа π_C обладает *свойством под-нятия*: если $f: \tilde{G} \rightarrow G$ — сюръективный гомоморфизм конечных групп и ϕ — гомоморфизм группы π_C в G , то существует гомоморфизм $\tilde{\phi}: \pi_C \rightarrow \tilde{G}$, такой, что $f \circ \tilde{\phi} = \phi$ (но если ϕ сюръективен, то не всегда можно выбрать $\tilde{\phi}$ сюръективным, даже если $\tilde{G}$ удовлетворяет условию гипотезы 2.3.1).

3. ГИПОТЕЗА АБЪЯНКАРА ДЛЯ АФФИННОЙ ПРЯМОЙ

3.1. Гипотеза и результаты. Мы сейчас ограничиваемся случаем, когда C является аффинной прямой $D = \mathrm{Spec} k[X]$, причем поле k алгебраически замкнуто и имеет характеристику $p > 0$.

В дальнейшем $g = 0$ и $s = 1$, откуда $2g + s - 1 = 0$, и гипотеза 2.3.1 принимает следующий вид:

Гипотеза 3.1.1. *Конечная группа G является факторгруппой группы π_D тогда и только тогда, когда $G = p(G)$, иными словами, когда G порождается своими силовскими p -подгруппами.*

Группа G , для которой $G = p(G)$, называется *квази- p -группой*.

Пример. Простая группа, порядок которой делится на p , является квази- p -группой. Из гипотезы 3.1.1 следует, что при $p = 2$ каждая простая неабелева группа является факторгруппой группы π_D .

Позднее мы увидим, что гипотеза 3.1.1 доказана во многих частных случаях, а именно, когда

G — группа рациональных точек полупростой алгебраической группы, односвязной над конечным полем характеристики p (Нори [10]);

G — знакопеременная группа A_n , $n \geq p > 2$, или симметрическая группа S_n , $n \geq p = 2$ (Абъянкар [2]);

G разрешима [19];

G порождена своими подгруппами, удовлетворяющими некоторым условиям (Харбатер [9], Рено [17]).

Замечание. Прямая D играет до некоторой степени универсальную роль для накрытий аффинных многообразий. Точнее, пусть

V — неприводимое аффинное многообразие над k размерности > 0 , и пусть G — конечная группа, являющаяся факторгруппой группы π_D . В этом случае существует связное накрытие Галуа $V' \rightarrow V$ с группой Галуа G , см. [19, п° 2].

3.2. Примеры Нори. Пусть q — степень числа p , и пусть Σ — полупростая алгебраическая группа, односвязная над F_q . Пусть $G = \Sigma(F_q)$ — группа F_q -точек группы Σ .

Группа G является квази- p -группой.

Теорема 3.2.1 (Нори). *Группа G является факторгруппой группы π_D .*

Следствие. *Гипотеза 3.1.1 верна для групп $SL_n(F_q), Sp_{2n}(F_q), \dots, E_8(F_q)$.*

Доказательство (Ж. Ломон). Пусть B^+ и B^- — противоположные борелевские подгруппы группы Σ , а U^+ и U^- — их унитарные радикалы. Отображение умножения $U^+ \times U^- \rightarrow \Sigma$ является изоморфизмом на замкнутое подмногообразие $V \subset \Sigma$ (замкнутость V в дальнейшем не будет использоваться). Обозначим через $f: \Sigma \rightarrow \Sigma$ изогению Ленга, определенную формулой

$$f(x) = x^{-1}F(x),$$

где F — эндоморфизм Фробениуса группы Σ по отношению к F_q . Эта изогения определяет связное накрытие Галуа группы Σ , группа Галуа которого есть G . Пусть $W = f^{-1}(V)$. Многообразие W связно. Действительно, пусть W^0 — его компонента связности, содержащая 1, и пусть G^0 — нормализатор компоненты W^0 в группе G . Поскольку W^0 содержит U^+ , группа G^0 содержит $U^+(F_q)$; аналогично, она содержит $U^-(F_q)$. Известно, однако (см. [20, лемма 64]), что группа G порождается $U^+(F_q)$ и $U^-(F_q)$. Поэтому $G = G^0$, откуда следует, что W связно. Таким образом получается связное накрытие Галуа $W \rightarrow V$, группа Галуа которого есть G . Но V изоморфно $U^+ \times U^-$ и, следовательно, является аффинным пространством размерности > 0 (за исключением случая $\Sigma = \{1\}$, но в этом случае доказывать нечего). Ограничивая накрытие $W \rightarrow V$ на прямую Δ в V , получаем накрытие прямой Δ с группой Галуа G . Согласно одному из вариантов теоремы Бертини, это накрытие является связным, если прямая Δ достаточно общая. Отсюда получается требуемый результат, поскольку Δ изоморфна D .

Замечание. Это доказательство приложимо также к «подкрученным» группам типа Судзуки и Ри в характеристике 2 и 3.

3.3. Примеры Абъянкара.

Теорема 3.3.1 (Абъянкар [2]). *Следующие группы являются факторгруппами группы π_D :*

знакопеременная группа A_n при $p \neq 2$ и $n \geq p$;

симметрическая группа S_n при $p = 2$.

Метод доказательства состоит в том, чтобы явно выпписать уравнения n -й степени, определяющие этальные накрытия прямой D степени n , и показать, что группы Галуа замыканий Галуа этих накрытий являются подгруппами группы S_n , изоморфными A_n или S_n , как указано в теореме. Для этого показывается, что исследуемая группа в достаточной степени транзитивна (в некоторых случаях Абъянкар вынужден использовать «классификационную теорему» для конечных простых групп).

Одно из используемых уравнений таково:

$$Y^n - X \cdot Y^t + 1 = 0, \quad (I)$$

где $n = p + t$, $t \geq 1$ и $t \not\equiv 0 \pmod{p}$.

Абъянкар (см. [2, 3]) доказывает, что группа Галуа этого уравнения есть

$$\begin{array}{ll} \text{PSL}_2(\mathbf{F}_p), & \text{если } t = 1; \\ \text{SL}_2(\mathbf{F}_8), & \text{если } t = 2, p = 7; \\ A_n, & \text{если } t = 2, p \neq 2, 7; \\ A_n, & \text{если } t > 2, p \neq 2; \\ S_n, & \text{если } p = 2. \end{array}$$

Приведем доказательство в самом простом случае — это случай, когда $t \geq 3$. Пусть G — группа Галуа уравнения (I). Это подгруппа группы S_n , обладающая следующими свойствами:

она транзитивна;

она содержит цикл с порядка p (из-за группы инерции точки $X = \infty$, $Y = \infty$);

она содержит элемент порядка, взаимно простого с p , который циклически переставляет t неподвижных точек цикла s (группа инерции точки $X = \infty$, $Y = 0$).

Поскольку $n = p + t$ и $(p, t) = 1$, легко проверяется, что из указанных свойств следует примитивность группы G . Поскольку

G содержит цикл простого порядка, не превосходящего $n - 3$, по теореме Жордана (см. [23, р. 39]) $G = S_n$ или A_n . Если $p = 2$, то из того, что G содержит транспозицию, следует, что $G = S_n$. Если $p \neq 2$, то, поскольку G является квази- p -группой, $G \neq S_n$. Следовательно, $G = A_n$, откуда и вытекает требуемый результат.

Замечания. 1) При $p = 2$ использованный метод, по-видимому, неприменим к случаю A_n с $n \geq 5$. Тем не менее, Рено мне сообщил, что можно исследовать этот случай с помощью теоремы 1.1 работы [17] и процедуры «переклеивания» группы A_5 (см. п. 3.5).

2) С помощью редукции $\text{mod } p$ некоторых известных уравнений в характеристике нуль можно получать явные уравнения, имеющие интересные группы Галуа. Например:

а) Согласно Фрике (см. [6, 24]), существует накрытие Галуа проективной прямой с группой Галуа $G = \text{SL}_2(\mathbb{F}_8)$, разветвленное в трех точках с порядками ветвления 2, 3 и 7, определенное над полем $\mathbb{Q}(\cos(2\pi/7))$. Соответствующая кривая имеет род 7. Если представить G как транзитивную подгруппу группы A_9 , то при этом возникает накрытие степени 9 проективной прямой, уравнение которого выписал Гурса [6]. Приводя это уравнение в характеристике 7 (после соответствующей замены переменных), получаем уравнение Абъянкара $Y^9 - X \cdot Y^2 + 1 = 0$, что еще раз приводит к основному результату работы [3].

б) Приводя в характеристике 11 уравнение Мацата [13], получаем уравнение

$$Y^{11} + 2Y^9 + 3Y^8 - X^8 = 0 \quad (p = 11),$$

которое определяет этальное накрытие прямой D степени 11, группа Галуа которого есть группа Матье M_{11} .

в) В характеристике 23 уравнение

$$Y(Y - 1)^2(Y + 1)^4(Y^2 + 17Y + 4)^8 - X^8 = 0 \quad (p = 23)$$

определяет этальное накрытие прямой D степени 23. По-видимому, группа Галуа этого накрытия должна быть группой Матье M_{23} ; было бы интересно доказать это.

3.4. Расширения разрешимых групп. Рассмотрим точную последовательность конечных групп $1 \rightarrow N \rightarrow \tilde{G} \rightarrow G \rightarrow 1$. Предположим, что $\tilde{G}$ является квази- p -группой, а N разрешима. Тогда,

если гипотеза 3.1.1 верна для G , то она верна и для $\tilde{G}$. Иными словами, справедлива

Теорема 3.4.1 (см. [19]). При указанных выше предположениях, если G является факторгруппой группы π_D , то это верно и для $\tilde{G}$.

Следствие 1. Гипотеза 3.1.1 верна для разрешимых групп: всякая разрешимая квази- r -группа является факторгруппой группы π_D .

Это — частный случай с $G = \{1\}$.

Следствие 2. Если G является факторгруппой группы π_D , то факторгруппой группы π_D является также и всякое расширение r -группы с помощью G .

Действительно, такое расширение является квази- r -группой, что позволяет применить теорему 3.4.1.

Доказательство теоремы 3.4.1. Ограничимся для простоты случаям, когда N — элементарная абелева группа типа $(\ell, \dots, \ell)$, где ℓ — простое число, и действие группы G на N неприводимо.

По условию теоремы существует сюръективный морфизм $\phi : \pi_D \rightarrow G$. Согласно сказанному в п. 2.3, его можно поднять до морфизма $\tilde{\phi} : \pi_D \rightarrow \tilde{G}$. Если $\tilde{G}$ — нерасщепленное расширение с помощью G , то $\tilde{\phi}$ сюръективен и теорема доказана. Случай, когда $\tilde{G}$ является полупрямым произведением группы G на N , более тонок. Если $\ell = p$, то доказывается, что можно выбрать $\tilde{\phi}$ сюръективным. Если $\ell \neq p$, то это в общем случае невозможно. Тем не менее в этом случае можно изменить ϕ так, чтобы это стало возможным. Точнее, выберем целое число $m > 1$, взаимно простое с p . Морфизм $D \rightarrow D$, заданный формулой $X \mapsto X^m$, индуцирует эндоморфизм $f_m : \pi_D \rightarrow \pi_D$ (в качестве базисной точки выбирается 0). Этот эндоморфизм сюръективен. Его композиция с ϕ дает гомоморфизм $\phi_m : \pi_D \rightarrow G$, который, подобно ϕ , сюръективен, но «более разветвлен» (его инвариант Свана в бесконечности в m раз больше). Доказывается, что можно поднять ϕ_m до сюръективного морфизма $\tilde{\phi}_m : \pi_D \rightarrow \tilde{G}$ (доказательство основывается на изучении модулей Галуа накрытий, снабженных гомологиями (mod ℓ), см. [19, th. 2]). Отсюда следует утверждение теоремы.

3.5. Конструкции Харбатера и Рено. Сначала заметим, что гипотеза Абъянкара 3.1.1 приводит (с помощью индукции по $|G|$ и

сведения к случаю циклической группы G) к следующему утверждению:

(3.5.1?) Если G порождена двумя подгруппами G_1 и G_2 , которые являются факторгруппами группы π_D , то G является факторгруппой группы π_D .

Харбатер [9] и Рено [17] доказали это утверждение при некоторых ограничительных предположениях. Например:

Теорема 3.5.2 [9, th. 4(i)]. Утверждение 3.5.1? верно, если G_1 является p -группой, содержащей силовскую p -подгруппу группы G_2 .

В следующем утверждении Рено участвуют группы инерции в бесконечности рассмотренных накрытий (определенные с точностью до сопряжения):

Теорема 3.5.3 [17, th. 1.1]. Пусть Q есть p -подгруппа группы G . Предположим, что для $i = 1, 2$ существует связное накрытие Галуа $\widetilde{D}_i \rightarrow D_i$ с группой Галуа G_i , группа инерции которого в бесконечности есть подгруппа Q_i группы Q . Тогда существует связное накрытие Галуа прямой D с группой Галуа G , группой инерции которого в бесконечности является Q .

Доказательства теорем 3.5.2 и 3.5.3 используют соответственно «формальный вариант GAGA» и «жесткий вариант GAGA», а в остальном практически совпадают (см. [15]).

Вот изложение (очень неполное) метода, примененного Рено в работе [17]:

Используется жесткая геометрия над локальным полем $K = k((T))$. На проективной прямой P_1 над K выбираются два замкнутых диска Δ_1 и Δ_2 , не содержащих ∞ , на положительном расстоянии друг от друга (например, диски $|z| \leq 1/p$ и $|z - 1| \leq 1/p$). Затем выбираются содержащие их чуть большие диски Δ_i^+ , удовлетворяющие тем же условиям. После этого определяются три накрытия Галуа (в жестком смысле) с группой Галуа G :

- (a₁) накрытие диска Δ_1^+ ;
- (a₂) накрытие диска Δ_2^+ ;
- (a₃) накрытие множества $P_1 - \Delta_1 - \Delta_2 - \{\infty\}$, которое продолжается до накрытия, разветвленного в бесконечности, с группой инерции Q .

Конструкция накрытий (a_1) и (a_2) такова: диск Δ_i , «приведенный $\text{mod } T$ », можно отождествить с аффинной k -прямой D_i . G_i -накрытие $\widetilde{D}_i \rightarrow D_i$ определяет путем поднятия на K жесткое G_i -накрытие диска Δ_i , и можно показать, что это накрытие продолжается на чуть больший диск. Тем самым получается G_i -накрытие, и индуцирование с G_i на G дает искомое накрытие (a_i) (заметьте, что в общем случае это накрытие не является неприводимым: число его неприводимых компонент равняется индексу подгруппы G_i в G).

Что касается (a_3) , то оно получается при вложении Q в G из (алгебраического) Q -накрытия пространства $P_1 - \{\infty\}$. Последнее нужно выбрать так, чтобы выполнялось следующее свойство:

(b) Для $i = 1, 2$ ограничение накрытия (a_3) на открытую корону $\Delta_i^+ - \Delta_i$ изоморфно ограничению накрытия (a_i) на эту корону (по крайней мере, если Δ_i^+ достаточно малы).

Это свойство позволяет *склеить* (в смысле жесткой геометрии) три накрытия (a_i) и получить накрытие Галуа пространства P_1 с группой Галуа G , разветвленное только в бесконечности, группой инерции которого является Q . По «жесткому варианту GAGA» (см. [11, 12]) это накрытие является алгебраическим. Кроме того, конструкция показывает, что это накрытие является геометрически неприводимым (именно здесь используется предположение, что G порождена подгруппами G_1 и G_2). Априори это накрытие определено над K ; путем специализации (см. п. 1.1) из него получаем накрытие над k , обладающее требуемыми свойствами.

Следствие 3.5.4. *Если G — группа Галуа связного накрытия прямой D , то это накрытие можно выбрать так, чтобы группы инерции в бесконечности были силовскими p -подгруппами группы G .*

Пусть I — группа инерции в бесконечности. Применяя лемму Абъянкара (т. е. замену переменных $X \mapsto X^n$ с соответствующим n), можно предположить, что I является p -группой. Пусть P — силовская p -подгруппа группы G , содержащая I . Применим теорему 3.5.3 с $G_1 = G$, $G_2 = P$, $Q_1 = I$, $Q_2 = P$, $Q = P$ (это законно, поскольку известно, что каждая конечная p -группа является факторгруппой группы π_D ; см., например, [19]).

Из этого результата следует, что теорема 3.5.2 является частным случаем теоремы 3.5.3.

Дополнения. Из теоремы 3.5.3 Рено вывел следующий результат:

Пусть G — конечная квази- p -группа и P — ее силовская p -подгруппа. Тогда существует подгруппа H группы G , содержащая подгруппу P , являющаяся факторгруппой группы π_D и *максимальная* среди всех подгрупп, обладающих этими свойствами. В частности, H сохраняется любым автоморфизмом группы G , сохраняющим P . Кроме того, всякая подгруппа группы G , являющаяся факторгруппой группы π_D , сопряжена с некоторой подгруппой группы H . Наконец, Рено показал [17, th. 1.5], что если $P \neq G$, то $H \neq P$ (это зачастую позволяет доказать, что $H = G$, т. е., иными словами, что гипотеза Абъянкара выполнена для G).

Приложение. Покажем, например, как можно вывести из теоремы 3.5.3, что *знакопеременная группа A_n , $n \geq 5$, является при $p = 2$ факторгруппой группы π_D .*

Пусть P — силовская 2-подгруппа группы A_4 . Покажем индукцией по $n \geq 5$, что существует связное накрытие Галуа прямой D с группой Галуа A_n , группы инерции которого в бесконечности сопряжены P . Это верно для $n = 5$, поскольку $A_5 = \text{SL}_2(\mathbb{F}_4)$ (применяем теорему 3.2.1 и следствие 3.5.4 или, что проще, используем уравнение $Y^5 + X^3Y + 1 = 0$). Если $n \geq 6$, применяем теорему 3.5.3 к группе $G = A_n$ с

$$G_1 = A_{n-1}, \text{ стабилизатор числа } n \text{ в } A_n;$$

$$Q_1 = P;$$

$$G_2 = \text{стабилизатор числа } n-1 \text{ в } A_n$$

$$(\text{следовательно, } G_2 \simeq A_{n-1});$$

$$Q_2 = P;$$

$$Q = P.$$

По предположению индукции все необходимые условия выполняются.

ЛИТЕРАТУРА

- [1] Abhyankar S. Coverings of algebraic curves, Amer. J. Math., **79** (1957), 825–856.
- [2] Abhyankar S. Galois theory on the line in nonzero characteristic, Bull. Amer. Math. Soc., **27** (1992), 68–133.
- [3] Abhyankar S. Square-root parametrization of plane curves, Purdue Univ., 1991.

- [4] Artin M., Grothendieck A., Verdier J.-L. Théorie des Topos et Cohomologie Étale des Schémas (SGA 4), vol. 3, Lect. Notes in Math., vol. 305, Springer-Verlag, 1973.
- [5] Douady R., Douady A. Algèbre et théories galoisiennes. 2: Théories galoisiennes. CEDIC, F. Nathan, Paris, 1979.
- [6] Fricke R. Über eine einfache Gruppe von 504 Operationen, Math. Ann., 52 (1899), 321–339.
- [7] Grothendieck A. Géométrie formelle et géométrie algébrique, Sémin. Bourbaki, exp. 182, 1958/1959, Collection Hors Série Astérisque, 5 (1995), 193–200.
- [8] Grothendieck A. Revêtements étales et groupe fondamental (SGA 1), Lect. Notes in Math., vol. 224, Springer-Verlag, 1971.
- [9] Harbater D. Formal patching and adding branch points, Preprint, 1991.
- [10] Kambayashi T. Nori's construction of Galois coverings in positive characteristics, in: Algebraic and Topological Theories, Tokyo, 1985, 640–647.
- [11] Kiehl R. Der Endlichkeitssatz für eigentliche Abbildungen in der nichtarchimedischen Funktionentheorie, Invent. Math., 2 (1967), 191–214.
- [12] Köpf U. Über eigentliche Familien algebraischer Varietäten über affinen Räumen, Schriftenreihe Univ. Münster, 2 Serie, Heft 7, 1974.
- [13] Matzat B. H. Konstruktion von Zahlkörpern mit der Galoisgruppe M_{11} über $\mathbb{Q}(\sqrt{-11})$, Man. Math., 27 (1979), 103–111.
- [14] Milne J. Etale Cohomology, Princeton Math. Series, vol. 33, Princeton, 1980. [Имеется перевод: Милн Дж. Эталонные когомологии. — М.: Мир, 1983.]
- [15] Raynaud M. Géométrie analytique rigide, d'après Tate, Kiehl; ..., Bull. Soc. Math. France, Mémoire No 39–40 (Table ronde anal. non-archim. 1972), 1974, 319–327.
- [16] Raynaud M. Sections des fibrés vectoriels sur une courbe, Bull. Soc. Math. France, 110 (1982), 103–125.
- [17] Raynaud M. Autour d'une conjecture d'Abhyankar, неопубликовано, 1991.
- [18] Serre J.-P. Cohomologie galoisienne, 4ème édit., Lect. Notes in Math., vol. 5, Springer-Verlag, 1973. [Имеется перевод издания 1964 г.: Серр Ж.-П. Когомологии Галуа. — М.: Мир, 1968.]
- [19] Serre J.-P. Construction de revêtements étales de la droite affine en caractéristique p , C. R. Acad. Sci. Paris, 311 (1990), série I, 341–346.
- [20] Steinberg R. Lectures on Chevalley groups, Notes prepared by John Faulkner and Robert Wilson, Yale University, 1967. [Имеется перевод: Стейнберг Р. Лекции о группах Шевалле. — М.: Мир, 1975.]
- [21] Weil A. Généralisation des fonctions abéliennes, J. Liouville, 17 (1938), 47–87 (=Oe [1938a]).
- [22] Weil A. L'avenir des mathématiques, Les Grands Courants de la Pensée Mathématique, éd. F. Le Lionnais, Cahiers du Sud, 1947, 307–320 (=Oe [1947a]).

- [23] Wielandt H. Finite Permutation Groups, Acad. Press, New York, 1964.
[24] Wohlfahrt K. Macbeath's curve and the modular group, Glasgow Math. J., 27 (1985), 239–247; Corrigendum, *ibid.*, 28 (1986), 241.

Добавление, сделанное в сентябре 1992 г. Гипотеза 3.1.1 Абъянкара для аффинной прямой доказана М. Рено. См. Raynaud M. Revêtements de la droite affine en caractéristique $p > 0$ et conjecture d'Abhyankar, *Invent. Math.*, 116 (1994), no. 1–3, 425–462

JEAN-PIERRE SERRE

Collège de France
3, rue d'Ulm
F-75005 PARIS.

E-mail: serre@ens.fr

АЛГЕБРАИЧЕСКИЕ ГРУППЫ И ЛИНЕЙНЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

Даниель Бертран

Приводя частичное описание списка Шварца [Sc], полученное с помощью теоремы Дирихле об арифметических прогрессиях, Ландау в 1904 г. отмечает: «Замечательно, что данная теоретико-функциональная проблема, которая связывает воедино различные геометрические, аналитические и алгебраические результаты, дает повод к применению теоремы чистой теории чисел ...» [La].

Проблема, о которой идет речь, состоит в том, чтобы составить список гипергеометрических уравнений Гаусса, которые допускают базис из алгебраических решений, т.е. группа монодромии которых конечна. Более общо, для данного линейного дифференциального уравнения с рациональными коэффициентами можно пытаться определить степень трансцендентности поля, порожденного его решениями. При наличии иррегулярных особенностей группа монодромии уже не контролирует эту степень, но дифференциальная теория Галуа позволяет преодолеть эту трудность. (В силу своей алгебраической природы эта теория позволяет упростить вычисления и в фуксовом случае.)

Известно [Bl], какого развития достигла эта теория в работах Колчина. Мы не будем здесь заниматься ее обобщениями (см. [Ko1, Po]) или приложениями (см., например, [Ne, Sn1, Um]). Предметом нашего изложения являются недавно разработанные, в частности Ж.-П. Рамисом и Н. Кацем, методы явного *вычисления* дифференциальных групп Галуа классических уравнений. Переходя к их обзору, мы будем придерживаться порядка, предложенного Ландау: геометрические методы рассматриваются в разд. 2, аналитические — в разд. 3, алгебраические — в разд. 4 и, наконец, арифметические — в разд. 5 (а в разд. 1 собраны основные понятия).

1. ОПРЕДЕЛЕНИЯ (см. [De])

Пусть C — алгебраически замкнутое поле характеристики нуль, (K, ∂) — дифференциальное поле с полем констант C и V — векторное пространство конечной размерности n над K , снабженное

связностью^(1) 1). Свертывая ее с ∂ , мы получаем дифференциальный оператор D на V , задаваемый в K -базисе пространства V формулой $Y \rightarrow DY = \partial Y + AY$, где A — квадратная матрица порядка n с элементами из поля K . Для любого дифференциального расширения (L, ∂) поля (K, ∂) решения в L^n системы дифференциальных уравнений $DY = 0$ представляют элементы пространства $(V \otimes L)^\partial$ горизонтальных векторов для связности $(V \otimes_K L, D \otimes \partial)$. Поле L называется *расширением Пикара—Вессю* поля K (ассоциированным с V), если выполнены следующие три условия: поля констант у L и K совпадают, $(V \otimes L)^\partial \otimes_C L = V \otimes_K L$ (другими словами, уравнение $DY = 0$ допускает фундаментальную матрицу $U(D)$ решений в L^n) и, наконец, L порождено пространством $(V \otimes L)^\partial$ (т. е. коэффициентами матрицы $U(D)$) над K . В этом случае группа $\text{Gal}_\partial(L/K)$ (или $\text{Gal}(D)$) автоморфизмов расширения L/K , перестановочных с ∂ , называется *дифференциальной группой Галуа* поля L над K (или пространства V/K ; корректность такой терминологии будет обоснована в следующем абзаце). Поскольку D — линейный оператор, векторное C -пространство $V^\partial = (V \otimes L)^\partial$ является точным линейным представлением группы $\text{Gal}(D)$. В духе Галуа можно интерпретировать элементы из $\text{Gal}(D)$ как те автоморфизмы пространства V^∂ , которые оставляют инвариантными пространства горизонтальных векторов в различных подпространствах всевозможных тензорных конструкций⁽²⁾ W над V , инвариантных относительно соответствующих тензорных связностей (см. [Ka2, De]). В частности, $\text{Gal}(D)$ является алгебраической подгруппой алгебраической группы $\text{Aut}_C(V^\partial)$.

Отправной точкой теории Колчина является тот факт, что при сделанных предположениях относительно поля C существует единственное с точностью до дифференциального изоморфизма расширение Пикара—Вессю, ассоциированное с V . За исчерпывающим доказательством этого факта мы отсылаем читателя к статье [Le]. Доказательство существования есть простое упражнение по дифференциальной алгебре. Более тонкий факт, что любые два расширения Пикара—Вессю L, L' связаны (неединственным) изоморфизмом $i(L, L')$, вытекает из интерпретации таких изоморфизмов как C -точек главного однородного пространства (torseur) алгебраической C -группы $\text{Gal}_\partial(L/K)$. В [De] получено обобщение (и новое доказательство) этих результатов в терминах таннакиевых категорий. С этой точки зрения отметим (тривиальное) существование

¹⁾ Цифры в скобках отсылают к примечаниям в конце статьи

функтора слоя над K , который позволяет рассматривать $\text{Gal}(D)$ как C -форму алгебраической подгруппы $\text{Gal}(V/K)$ алгебраической K -группы $\text{Aut}_K(V)$, определенной «теми же» условиями, что и выше, тогда как изоморфизмы $i(L, L')$, равно как и матрица $U(D)$, имеют трансцендентную природу. Вычисление этих изоморфизмов, или, на языке групп монодромии, отыскание «формул связности», является, таким образом, сложной задачей, но иногда это дает ключ к вычислению группы Галуа $\text{Gal}(D)$. Мы увидим это на примерах в разд. 2 и 3.

Указанные выше результаты собраны — и уточнены — в следующей теореме.

Теорема 1. (i) Обычное соответствие Галуа устанавливает биекцию между алгебраическими подгруппами группы $\text{Gal}(D)$ и промежуточными дифференциальными расширениями в L/K , причем замкнутые нормальные подгруппы соответствуют расширениям Пикара—Вессии поля K .

(ii) Категория рациональных конечномерных C -представлений группы $\text{Gal}(D)$ эквивалентна категории связностей на подфакторах тензорных конструкций над V . В частности, представление V^D группы $\text{Gal}(D)$ неприводимо тогда и только тогда, когда связность пространства V определена над K .

(iii) Степень трансцендентности поля L над K равна размерности группы $\text{Gal}(D)$. Более точно, $U(D)$ есть общая точка главного однородного (K -неприводимого) K -пространства алгебраической группы $\text{Gal}(V/K)$.

2, ВОКРУГ МОНОДРОМИИ

Здесь мы предполагаем, что пространство V происходит из расслоения с голоморфной связностью над дополнением к конечному множеству Σ точек на сфере Римана. Для любого элемента $\sigma \in \Sigma$ обозначим через $K_{\sigma, \text{an}}$ (соответственно $K_{\sigma, f} = \hat{K}$) поле частных кольца ростков аналитических функций (соответственно формальных степенных рядов) в точке σ . Пусть $G_{\sigma, \text{an}}$ и $G_{\sigma, f}$ — аналитическая и формальная группы Галуа пространства V в точке σ , т. е. группы Галуа пространств со связностями $V_{\sigma, \text{an}}$ и $\hat{V}$, полученных из V путем расширения скаляров с $K = \mathbb{C}(P_1)$ до $K_{\sigma, \text{an}}$ и $\hat{K}$ соответственно. Они отождествляются (см. [Ка3, 2.7]) с алгебраическими подгруппами в $\text{Gal}(D)$ (вообще говоря, не обязательно нормальными), причем $G_{\sigma, \text{an}}$ содержит замыкание по Зарисскому $M_{\sigma, \text{an}}$ ло-

кальной группы монодромии оператора D в точке σ . Из теоремы 1, совпадения классов мероморфных и рациональных функций на $\mathbf{P}_1$ и теории Фукса выводится

Предложение 1. В случае $K = \mathbf{C}(z)$ группа Галуа $\text{Gal}(D)$ порождена⁽³⁾ (в топологии Зарисского) своими локальными аналитическими подгруппами. Кроме того, группы $G_{\sigma, \text{an}}$, $G_{\sigma, f}$ и $M_{\sigma, \text{an}}$ совпадают, если σ — регулярная особая точка оператора D .

(Напомним, что особенность σ называется регулярной, если любое горизонтальное сечение имеет в ее окрестности не более чем полиномиальный рост.) В частности, если Σ содержит только регулярные особые точки, то $\text{Gal}(D)$ есть замыкание по Зарисскому группы монодромии расслоения со связностью, которое порождает V .

Пример 1. Гипергеометрическое уравнение Гаусса $H = H(a, b, c)$:

$$z(1-z)\partial^2 y + (c - (a+b+1)z)\partial y - aby = 0 \quad (\partial = d/dz)$$

[или, в другой форме, $(\theta+a)(\theta+b)y - z^{-1}\theta(\theta+c-1)y = 0$ ($\theta = z\partial$)].

Все три его особые точки $0, \infty, 1$ регулярны. Метод Фробениуса позволяет вычислить здесь локальные аналитические подгруппы. Между тем известна восходящая к Риману «формула связности», которая позволяет их сравнивать в рамках одной и той же реализации группы $\text{Gal}(H)$ (ср. [Va]). Для рациональных чисел a, b, c таким образом можно проверить путем трудных вычислений, что условие чередования Каца [Ka1] (см. ниже пример 3) необходимо и достаточно для того, чтобы уравнение $H(a, b, c)$ попадало в список Шварца.

Совершенно иная картина складывается в случае иррегулярной особенности, который будет служить нам в качестве модельной ситуации вплоть до конца статьи. Напомним, что $\hat{V} = V \otimes \hat{K}$. Согласно фундаментальной теореме Хукухары—Турритина—Левельта (см. [Ka3]), существует расширение $\hat{L}$ поля $\hat{K}$ степени $N \leq n!$, для которого $\hat{V} \otimes \hat{L}$ может быть получено в результате цепочки последовательных расширений пространств со связностями с одномерными факторами. (Имеются многочисленные доказательства этой теоремы. Последнее по времени, принадлежащее Бэббиту и Варадараяну [BV1], основано на двойной индукции по порядку и размерности орбиты наиболее полярной части матрицы A , задающей оператор D , под действием калибровочных преобразований из

$\text{Aut}(\widehat{V})$; недавно это доказательство распространено Сабба [Sa] на уравнения в частных производных). Другими словами, существует определенный над $\widehat{L}$ горизонтальный изоморфизм $\widehat{\Phi}$ пространства $\widehat{V}$ на пространство со связностью $(\widehat{V}_0 = \widehat{K}^n, \widehat{D}_0)$, где

$$\widehat{D}_0 = t(d/dt) - \Lambda + t^{-1}[(d/dt)Q](t^{-1}), \text{ причем } U(\widehat{D}_0) = t^\Lambda \exp(Q(t^{-1})),$$

для униформизирующей $t \in \widehat{L}$, диагональной матрицы Q , элементы которой — полиномы без свободного члена (называемые детерминантными множителями и переставляемые группой инерции $\text{Gal}(\widehat{L}/\widehat{K})$), и жордановой матрицы Λ , перестановочной с Q (по поводу $\widehat{K}$ -рациональной формы изоморфизма $\widehat{\Phi}$ см. [Ju, p. 32]). Экспоненты детерминантных множителей порождают расширение Пикара—Вессии поля $\widehat{L}$, дифференциальной группой Галуа которого является $\text{tor } T_\sigma(D)$ (называемый экспоненциальным), изоморфный группе характеров $\mathbb{Z}$ -подмодуля в $\mathbb{C}[t^{-1}]$, порожденного этими множителями. Зафиксируем, кроме того, «оператор формальной монодромии» $t \mapsto t \cdot \exp(2\pi i/N)$. Он действует на $T_\sigma(D)$ посредством своего образа в $\text{Gal}(\widehat{L}/\widehat{K})$. Обозначим также через $M_{\sigma,f}(D)$ (формальная монодромия пространства V) замыкание по Зарискому подгруппы в $\text{Aut}(\widehat{V}_0^\partial)$, порожденной этим оператором. При этих условиях справедливо следующее утверждение ([Ra2, Mi]; см. также [Ka3, 1.4.5]):

Предложение 2. *Формальная дифференциальная группа Галуа $G_{\sigma,f}(D)$ порождена $M_{\sigma,f}(D)$ и $T_\sigma(D)$.*

Так как изоморфизм $\widehat{\Phi}$ задается, вообще говоря, расходящимся рядом, группы $M_{\sigma,f}(D)$ и $M_{\sigma,an}(D)$ никак не связаны друг с другом, и у нас нет никакого естественного способа свести вычисление группы $G_{\sigma,f}(D)$ к $G_{\sigma,an}(D)$. Предположим, далее, что множество Σ сводится к двум точкам $\{0, \infty\}$, так что глобальная группа монодромии оператора D абелева, и что 0 является регулярной особенностью. (Это первая нетривиальная ситуация, в которой описанные ниже методы приводят к успеху. Следует признать, что на настоящий момент это по существу единственный пример успешного приложения указанных методов, в котором оператор D не является фуксовым. См. [Ka4, Ka2 (10.3), Mi].) В этом случае $G_{0,f} = G_{0,an} = M_{0,f} = M_{0,an}$ вкладывается в $G_{\infty,an} = \text{Gal}(D)$ в качестве подгруппы, замкнутой в $M_{\infty,an}$, но уяснить это совсем не просто:

Пример 2. Уравнение Бесселя $J = J(v)$:

$$\partial^2 y + z^{-1} \partial y + (1 - v^2 z^{-2}) y = 0 \quad (\partial = d/dz).$$

[Оно получается из гипергеометрического уравнения при вырождении двух особенностей в одну; к сожалению, сегодняшний уровень знаний о поведении дифференциальных групп Галуа при специализации ([Ka2 (2.4), Sn3]) не позволяет извлечь пользу из этого замечания.] С помощью подконструкции «максимальной внешней степени» (вронскиан) мы можем вложить $\text{Gal}(J)$ в $SL(2)$. Метод Фробениуса позволяет указать образующую группы $G_{0,an}$ — нетривиальный унипотентный элемент при целых v и полупростой элемент в противном случае. Вместе с тем детерминантные множители для уравнения J в ∞ равны $\pm iz$ (лемма Гензеля, см. [Po]), так что $T_\infty(J)$ — одномерный тор, и группа формальной монодромии $M_{\infty,f}$ равна $\mathbb{Z}/2\mathbb{Z}$ (т. е. уже не содержит унипотентных элементов!) Одной этой информации, вообще говоря, недостаточно, чтобы определить $\text{Gal}(J)$ (окончательный ответ см. в [Ko2] или ниже).

Разумеется, мы располагаем принадлежащим Ганкелю выражением для стандартного базиса решений уравнения J :

$$J_{\pm v}(z) = z^{-1/2} (c_{\pm v} e^{-iz} H_v(z) + c_{\pm v}^{-1} e^{iz} H_v(-z)),$$

$$Y_v(z) = z^{-1/2} (c'_v e^{iz} H_v(-z) + c'_v{}^{-1} e^{-iz} H_v(-z)) \quad (\text{для целых } v)$$

в терминах функции H_v , аналитической при $|\text{Arg}(z)| < \pi$ и имеющей в окрестности ∞ асимптотическое разложение $\hat{H}_v \in K_{\infty,f}$. Но гомоморфизм «перехода к асимптотическому разложению» из $K(H_v)$ в $K_{\infty,f}$ не продолжается, вообще говоря, до дифференциального гомоморфизма из $K(J_v, J'_v, J_{-v} \text{ (или } Y_v))$ в $K_{\infty,f}(z^{1/2}, e^{iz})$. Так, $2i \sin(z)$ равно $S(z)e^{iz}$, где S имеет на бесконечности в области $\text{Im}(z) < 0$ асимптотическое разложение $\hat{S} = 1$, но не существует дифференциального гомоморфизма из $K(\sin(z), \cos(z))$ в $K_{\infty,f}(e^{iz})$, переводящего Se^{iz} в $\hat{S}e^{iz} = e^{iz}$. Следовательно, мы пока не можем переносить информацию между $M_{\infty,an}$ и $G_{\infty,f}$.

Теория Рамиса, к изложению которой мы переходим, дает решение этих проблем, обращая направление искомым изоморфизмов.

3. ГРУППА СТОКСА (по Рамису)

В этом разделе будут рассматриваться локальные вопросы; поэтому мы будем считать, что основное поле $K = K_{0,an}$ есть поле

ростков мероморфных функций в окрестности точки $\sigma = 0$. Обозначения, относящиеся к его формальному пополнению $\widehat{K} = \mathbb{C}((z))$, — такие же, как в разд. 2. Предположим для простоты, что детерминантные множители $\{Q_1, \dots, Q_n\}$ пространства $V = V_{0,an}$ не разветвляются (т. е. что $N = 1$), и обозначим через $k(V) = \sup_i (d^0(Q_i))$ ранг Каца пространства V . Что более важно, мы ограничимся случаем общего положения (с диаграммой Ньютона пространства $\text{Epd}(V)$, состоящей из одного отрезка), в котором все ненулевые разности $Q_i - Q_j$ имеют степень, равную $k = k(\text{End}(V))$.

Сначала мы дадим обзор метода Рамиса в его аналитической форме (см. [MR1, MR2, RS, Si]), а затем опишем его когомологическую версию, предложенную Делинем и Мальгранжем (см. [Ma2, MR2, BV2, LR]). В следующем утверждении собраны в более классической форме ключевые факты, необходимые для этих двух подходов (см. [M-R, RS]). В дальнейшем мы фиксируем изоморфизм $\widehat{V}$ на нормальную форму $\widehat{V}_0$, матрицу которого обозначим через $\widehat{\Phi}$.

Предложение 3. Пусть B — росток открытого сектора в окрестности точки 0.

(i) (Пуанкаре—Хукухара—Турритин) Если $\text{Ouv}(B) < \pi/k$, то существует голоморфная матрица Φ_B , асимптотически равная $\widehat{\Phi}$ в секторе B , для которой матрица

$$U_B(D)(z) = \Phi_B(z) z^\Lambda \exp(Q(z^{-1}))$$

является фундаментальной матрицей решений уравнения $DY = 0$.

(ii) (Ватсон—Неванлинна) Если $\text{Ouv}(B) > \pi/k$, то любая функция, голоморфная в B и имеющая в точке 0 экспоненциальное убывание порядка k , тождественно равна 0.

[Здесь $\text{Ouv}(B)$ обозначает угол раствора сектора B . Напомним также, что голоморфная в B функция f допускает асимптотическое разложение в 0, если существует (обязательно единственный) формальный ряд Лорана $\widehat{f} = T_B(f) = \sum_{v(\widehat{f}) \leq n} f_n z^n$, такой, что для любого целого $m \geq 0$ функция

$$R_m f(z) = z^{-m} \left(f(z) - \sum_{v(\widehat{f}) \leq m-1} f_n z^n \right)$$

ограничена на любом относительно компактном подсекторе в B .]

Согласно Коши, удовлетворяющие утверждению (i) матричные функции $\Phi_B(z) = U_B(z) \exp(-Q(z^{-1}))z^{-\Lambda}$ аналитически продолжаются в любой сектор с углом раствора $< 2\pi$, содержащий B . В общем случае для этого продолжения $\hat{\Phi}$ более не будет асимптотическим разложением, оно вообще может не иметь такого разложения (так что мы не можем использовать п. (ii)), но, учитывая указанное выше соотношение, мы можем утверждать, что такие разрывы могут возникнуть только вдоль направлений с аргументом θ , для которых в окрестности $z = 0$ при переходе через линию $\text{Arg}(z) = \theta$ выражение $\text{Re}(Q_i - Q_j)(z^{-1})$ меняет знак хотя бы для одной пары (i, j) . Эти направления называются *линиями Стокса*. Для того чтобы оказаться в ситуации утверждения (ii), поставим им в соответствие

- *сингулярные линии*, т.е. такие направления с аргументом α , что хотя бы одно из выражений $\exp(Q_i - Q_j)(z^{-1})$ убывает с максимальной скоростью (экспоненциально порядка k) при $z \rightarrow 0$, $\text{Arg}(z) = \alpha$;

- «*хорошие*» *покрытия* окружности S направлений из 0. В случае общего положения, которым мы занимаемся (по поводу общего случая см. [MR2, LR]), мы можем, например, взять объединение по различным сингулярным линиям (с аргументами α) интервалов вида $B_\alpha =]\alpha - (\frac{\pi}{2k}), \alpha + (\frac{\pi}{2k}) + \delta[$, $B'_\alpha =]\alpha - (\frac{\pi}{2k}) - \delta, \alpha + (\frac{\pi}{2k})[$, ограниченных линиями Стокса и не содержащих других линий Стокса вне $b_\alpha = B'_\alpha \cap B_\alpha$ (см. [Mi, 2.3.6]). Наконец, перенесем эти обозначения на универсальное накрытие $\tilde{S}$ окружности S и отождествим соответствующие интервалы на $\tilde{S}$ с открытыми секторами с центром в 0, которые ими задаются.

3.1. Аналитический подход. Отправной точкой этого подхода является то обстоятельство, что коэффициенты матрицы $\hat{\Phi}$ принадлежат (собственной) дифференциальной подалгебре поля $\hat{K}$, над которой отображение $T_{B(d)}$ допускает каноническое сечение для любого сектора $B(d)$ с несингулярной биссектрисой d и углом раствора немного больше π/k (см. [Si, A.4.1.2]). Дело в том, что плоские функции из образа $T_{B(d)}$ имеют экспоненциальное убывание порядка k , а следовательно, равны 0 согласно предположению 3(ii). Рамис описывает это сечение, используя операцию повторного суммирования $\Sigma_d = \Sigma_{k,d}$, которая есть композиция формального преобразования Бореля и преобразования Лапласа (см. [MR2, § 1]) и тем самым перестановочна с алгебраически-

ми операциями и дифференцированием. Следовательно, Σ_d есть изоморфизм дифференциальных алгебр $K[\hat{\Phi}]$ и $K[\Sigma_d \hat{\Phi}]$, который легко продолжается до дифференциального изоморфизма алгебры $K[\hat{\Phi}, z^\lambda, \exp Q_i(z^{-1}), \text{Log}(z)]$ на $K[\Sigma_d \hat{\Phi}, z^\lambda, \exp Q_i(z^{-1}), \text{Log}(z)]$ (см. [MR2, Prop. 9]). Следовательно, для любого несингулярного направления d матрица $\Sigma_d \hat{\Phi} = \Phi_{B(d)}$ удовлетворяет условиям предложения 3(i) в $B(d)$ (в случае общего положения можно прямо из утверждения 3(i) вывести существование и единственность такого решения; см. [IK]⁽⁴⁾). Теперь сопоставим каждому сингулярному направлению, или, более точно (с учетом матрицы Λ), каждому из соответствующих $\alpha \in \tilde{S}$ фундаментальные матрицы

$$U_\alpha(D) = \Sigma_{\alpha+(\varepsilon/2)}(\hat{\Phi} z^\Lambda \exp Q(z^{-1})),$$

$$U'_\alpha(D) = \Sigma_{\alpha-(\varepsilon/2)}(\hat{\Phi} z^\Lambda \exp Q(z^{-1}))$$

решений уравнения $DY = 0$ в B_α, B'_α . В секторе b_α они связаны соотношением $U'_\alpha(D) = U_\alpha(D)C_\alpha$, где унитарная матрица C_α представляет канонический S -эндоморфизм st_α пространства V^∂ решений уравнения $DY = 0$, голоморфных в b_α . Но st_α есть композиция дифференциальных изоморфизмов $\Sigma_{\alpha+(\varepsilon/2)}^{-1}$ и $\Sigma_{\alpha-(\varepsilon/2)}$, и, следовательно, принадлежит дифференциальной группе Галуа пространства V над K . Назовем *группой Стокса* (в иррегулярной особенности $\sigma = 0$) алгебраическую подгруппу $St_\sigma(D) = St_0(D)$ в $G_{0,an}(D)$, порожденную⁽⁵⁾ элементами st_α , где α пробегает множество всех прообразов в $\tilde{S}$ различных сингулярных направлений для V .

Уже из данного определения видно, что группа Стокса, вообще говоря, не является нормальной подгруппой в $G_{0,an}$, но $M_{0,f}$ ее нормализует. С другой стороны, каждый изоморфизм st_α^{-1} вычисляет *аналитическое продолжение* решения $U'_\alpha(D)$ с B'_α на B_α (ср. [DM, Prop. 6.1]). Взяв их композицию вдоль S и применив оператор формальной монодромии, чтобы вернуться к начальному значению аргумента, мы, таким образом, получим образующий элемент группы аналитической монодромии $M_{0,an}(D)$. Следовательно, природа отмеченной выше несогласованности между $M_{0,f}$ и $M_{0,an}$ прояснена. Именно отсюда Стокс начал путь к своему феномену, как он это описал в одном из писем к своей невесте [St, p. 62], — женитьба, тем не менее, состоялась. Более общим образом, Рамис установил следующее:

Теорема 2. Аналитическая локальная группа Галуа $G_{0,an}(D)$ порождена своими замкнутыми подгруппами $G_{0,f}(D)$ и $St_0(D)$ (или, если угодно, подгруппами $T_{0,f}(D)$, $M_{0,an}(D)$ и $St_0(D)$).

Доказательство этой теоремы (см. [MR2, th. 6]) основано на теореме 1 и том факте (эквивалентном предыдущему замечанию), что решение, инвариантное относительно группы Стокса и формальной монодромии, равномерно в окрестности точки 0. Доказательство в духе таннакиевых категорий предложил Делинь.

3.2. Когомологический подход. Этот подход опирается на классификацию с точностью до K -изоморфизма векторных K -пространств V со связностью, снабженных, как и выше, формальным изоморфизмом $\hat{\Phi}$ пространства $\hat{V}$ на фиксированную нормальную форму $\hat{V}_0$ (теорема Сибуи—Мальгранжа; см. [Ma2, Si, BV2]). Пусть A — пучок ростков над S мероморфных функций, допускающих асимптотическое разложение в 0, а $ST(V)$ — пучок (неабелевых) групп горизонтальных сечений пучка $A \otimes \text{End}(V)$, асимптотически равных тождественному отображению. Применяя предложение 3(i) (3(ii)) в ситуации, когда $k = \infty$ (соответственно $k = 1/2$ или $k = 0$ для $\tilde{S}$), мы получаем, что $ST(V)$ локально изоморфен $ST(V_0)^{(6)}$ и что искомое множество классов изоморфных пар $(V, \hat{\Phi})$ с естественной отмеченной точкой $(V_0, \hat{id})$ (это с точностью до изоморфизма единственная пара, для которой не возникает феномен Стокса) отождествляется с множеством одномерных когомологий $H^1(S, ST(V_0))$. Но $ST(V_0)$ легко описать! Его сечения над сектором B (поднятым в $\tilde{S}$) — это голоморфные в B матрицы P_B , асимптотически равные I , для которых $P_B(z)z^\Lambda \exp(Q(z^{-1})) = z^\Lambda \exp(Q(z^{-1}))C_B$, где C_B — постоянная матрица. В частности, матрица $\exp(Q(z^{-1}))C_B \exp(-Q(z^{-1}))$ должна быть асимптотически равна I в B , и пучок $ST(V_0)$ кусочно постоянен, причем разрывы могут возникать лишь на линиях Стокса пространства V . Таким образом, в случае общего положения (по поводу общего случая см. [BV2, MR2, LR]) классы когомологий пучка $ST(V_0)$ задаются набором постоянных унитарных матриц C_α , взятых по всем секторам b_α , все недиагональные элементы которых равны 0, кроме может быть стоящих на (i, j) -м месте, где i, j — индексы детерминантных факторов, связанных с сингулярным направлением α .

Пусть $\{C_\alpha\}$ — такое семейство матриц. Ему соответствует единственная с точностью до K -изоморфизма пара $(V, \hat{\Phi})$, принадле-

жащая $H^1(S, ST(V_0))$. Поскольку «хорошие» покрытия достаточно малы (см. [BV2, 3.4.2]), каждая матрица $P_\alpha(z)$, соответствующая C_α (на $\tilde{S}$), может быть записана в виде частного $(\Phi_\alpha)^{-1}(\Phi'_\alpha)$ двух горизонтальных сечений пучка $A \otimes \text{Hom}(V, V_0)$ над B_α, B'_α , асимптотически равных $\hat{\Phi}$, и притом единственным образом (поскольку тождественный оператор является единственным глобальным сечением пучка $ST(V)$). Другими словами, с $\hat{\Phi}$ и с каждым сингулярным направлением α (точнее, с его прообразом в $\tilde{S}$) канонически связаны две фундаментальные матрицы $U_\alpha(D) = \Phi_\alpha(z)z^\Lambda \exp(Q(z^{-1}))$, $U'_\alpha(D)$ решений уравнения $DY = 0$, такие, что $U'_\alpha(D) = U_\alpha(D)C_\alpha$ в b_α . Тем самым мы получаем канонические элементы $st_\alpha \in \text{Aut}_{\mathbb{C}}(V^\partial)$ (из однозначности их «аналитического» описания следует, что это те же элементы, что и в п. 3.1). Но нормальные формы W_0 различных тензорных конструкций W над V согласованы с соответствующими подсвязностями. Следовательно, автоморфизмы пространств W^∂ , определяемые элементами st_α , тоже согласованы с подсвязностями (по крайней мере, если W — достаточно общего положения). Таким образом, мы снова видим, что st_α лежат в группе Галуа $\text{Gal}(D)$. Эта точка зрения позволяет также доказать теорему 3: для любой конструкции W над V $G_{0,f}(D)$ -инвариантное $\mathbb{C}$ -подпространство $X' \subset W^\partial$ задает подсвязность X_0 на соответствующей конструкции W_0 над V_0 ; если к тому же X' инвариантно относительно $ST_0(D)$, то ограничения матриц C_α на X определяют элемент пространства $H^1(S, ST(X_0))$, который с помощью изоморфизма Мальгранжа—Сибуи можно интерпретировать как подсвязность X в W с пространством горизонтальных векторов $X^\partial = X'$ (см. [LR]⁽⁷⁾).

3.3. Вычисления и замечания. Остается научиться вычислять матрицы Стокса—Рамиса C_α . При аналитическом подходе операторы повторного суммирования позволяют выразить коэффициенты этих матриц в интегральной форме, и теорема о вычетах сразу дает ответ (см. [MR1]). Что касается кохомологического подхода, одно из его преимуществ состоит в обосновании предельных переходов: пространство $H^1(S, ST(V_0))$ снабжено естественной структурой аффинного многообразия, будучи пространством модулей (классов изоморфизма) пар $(V, \hat{\Phi})$ (см. [BV2, LR]).

Пример 2 (продолжение). Уравнение Бесселя $J(v)$. Этот пример относится к случаю общего положения (в $\sigma = \infty$) при $k = 1$. Сингулярные линии (соответственно линии Стокса) — это две мни-

мые полуоси $\alpha_1 = -i$, $\alpha_2 = i$ (соответственно две вещественные полуоси), а хорошее покрытие состоит из интервалов

$$B_1 =] - 2\pi, 0[, \quad B'_1 =] - \pi, \pi[, \quad B_2 =] - \pi, \pi[, \quad B'_2 =]0, 2\pi[,$$

где $b_1 =] - \pi, 0[, \quad b_2 =]0, \pi[$. Матрицы Стокса—Рамиса равны

$$C_i = \begin{bmatrix} 1 & 0 \\ c_i(v) & 1 \end{bmatrix}, \quad C_{-i} = \begin{bmatrix} 1 & c_{-i}(v) \\ 0 & 1 \end{bmatrix},$$

где

$$c_{\pm i}(v) = \mp 2 \exp(\mp 2i\pi v) \cos \pi v, \quad \text{согласно [MR1, 3.4.12];}$$

$$c_{\pm i}(v) = \mp 2 \cos \pi v, \quad \text{согласно [BV2, p. 94].}$$

Причина несовпадения — не «ошибка эксперимента», а различный выбор базисных решений (или изоморфизма $\hat{\Phi}$). (Матрицы в [MR1] выписаны в базисе $z^{-1/2}e^{-iz}H_v(z)$, $z^{-1/2}e^{iz}H_v(-z)$; на самом деле $H_v = \Sigma_0(\hat{H}_v)$, и мы можем забыть о предосторожностях, принятых в примере 2 в разд. 2...). Но это несущественно. Нам достаточно знать, что $c_i(v)$ и $c_{-i}(v)$ не равны 0, чтобы из полученных выражений немедленно вывести, что при v , не сравнимом с $\frac{1}{2} \bmod \mathbf{Z}$, группа Стокса, а значит и $\text{Gal}(J(v))$, совпадает со всей $SL(2)$. В этом смысле наш метод, можно сказать, избыточен: он дает точные «формулы связности», тогда как достаточно знать местоположение ненулевых коэффициентов у некоторых унипотентных матриц (см. [MR1]).

Другие примеры, относящиеся к случаю общего положения, можно найти в [MR1] (вырожденные гипергеометрические уравнения порядка 2) и в [DM, Mi] (обширный класс подобных уравнений высших порядков; см. ниже пример 3). За рассмотрением общего случая мы отсылаем читателя к [Ra2] и [LR], а здесь лишь отметим, что с теоретической точки зрения этот случай можно исследовать, отправляясь от канонического разложения изоморфизма $\hat{\Phi}$, индексированного различными звеньями диаграммы Ньютона пространства $\text{End}(V)$. В этом случае аналитический подход основан на фундаментальном понятии «мультисуммируемости» [MR2, BBRs], и для него, в частности, необходима соответствующая версия предложения 3(ii) (см. [M-R, Ma2]). Отсюда можно получить «фильтрацию Жевре» на дифференциальной группе Галуа [Ra2], которой при когомологическом подходе соответствует фильтрация того же типа на пучке $ST(V_0)$ [MR2, LR].

Наконец, один из наиболее интересных аспектов теории Рамиса связан с *дикой группой монодромии* $\pi_{1,s}$ роста открытой пунктированной окрестности начала координат [MR2, § 4, pp. 381, 388]. Эта группа есть полупрямое произведение обычной группы монодромии и группы, являющейся полупрямым произведением группы характеров $\mathbb{Z}$ -модуля конечного типа в алгебре разветвленных полиномов от z^{-1} и «группы перестроек». Последняя есть группа, ассоциированная со свободной алгеброй Ли, порожденной различными сингулярными линиями. Можно рассматривать образ $\text{Gal}(D)$ в $\text{Aut}(V^\partial)$ как представление группы $\pi_{1,s}$ и вывести отсюда [MR2, th. 18] обобщенное соответствие Римана—Гильберта между категориями векторных $K_{0,\text{an}}$ -пространств со связностью и конечномерных представлений группы $\pi_{1,s}$. (См. также [MR2] по поводу связи этого аспекта с работами д'Экаля.)

4. РЕДУКТИВНЫЕ ГРУППЫ (ср. [Ka4, Be])

Результаты этого раздела основаны на весьма жестких свойствах комплексных полупростых алгебр Ли. Идея их использования является классической (см. [Кар] и [Ko2] для $\mathfrak{sl}(2)$ в применении к уравнениям Эйри и Бесселя). Последние исследования Бёккера, Браунавелла, Хекмана [BBH, BH] и Каца и Габбера [Ka3, Ka4] дали новое развитие этому направлению. Поскольку этому сюжету уже посвящена половина книги [Ka4] (содержание второй ее половины мы тоже рекомендуем читателю) и обзор [Be], наше изложение будет кратким.

Предположим, что векторное K -пространство V со связностью неприводимо или, более общо, вполне приводимо. По теореме 1 точное представление V^∂ группы $G = \text{Gal}(D)$ в этом случае вполне приводимо, и G является редуктивной группой, разлагающейся в почти прямое произведение $Z \cdot DG$ своего центра Z и коммутанта DG . Поскольку характеристика поля равна 0, все представления группы G и, следовательно, все тензорные конструкции над V вполне приводимы, и можно надеяться, что это облегчит нахождение подсвязностей, определяющих G . Например (см. [BBH, Be]), если DG бесконечна, то неприводимости симметрического квадрата $S^2(V)$ достаточно для того, чтобы все симметрические степени пространства V были неприводимы, и тогда DG есть вся группа $SL(n)$ или симплектическая группа $Sp(n)$.

Следовательно, важно иметь в распоряжении критерии неприводимости пространства V . Более конкретно, речь идет о том, чтобы выяснить для пространства V , определенного дифференциаль-

ным уравнением $Hu = 0$, неприводим ли H в кольце $K[\partial]$ дифференциальных операторов с коэффициентами из K . Это трудная проблема⁽⁸⁾ (например, мы не располагаем аналогом леммы Гаусса для кольца $\mathbb{C}[z, d/dz]$), но иногда мы можем преодолеть эту трудность с помощью локального анализа. Так, в обозначениях разд. 2, если все детерминантные множители пространства $\hat{V}$ имеют один и тот же порядок, равный $-n$, то $\hat{V}$ неприводимо над $\hat{K}$ и, следовательно, V неприводимо над K (о применении к обобщенным уравнениям Эйри см. [Ка3, 2.2.8, 4]).

Пример 2 (окончание). Уравнение Бесселя $J(v)$. Если параметр v целый, присутствие логарифма в решении Y_v вместе с иррациональностью $2\frac{J'_v}{J_v}$ показывает, что симметрический квадрат пространства V неприводим над $K = \mathbb{C}(z)$. Так как $\dim \text{Gal}(J(v)) = \text{tr.deg } K(J, J', Y)/K \geq 2$, группа DG не может быть конечной, и, как следует из изложенного выше, $\text{Gal}(J(v))$ совпадает с $SL(2)$. То же самое верно, если v не полуцелое (см. ниже). Напротив, $\text{Gal}(J(\frac{1}{2}))$ есть тор (функции Эрмита).

Вместе с тем эти методы позволяют описать скорее алгебру Ли группы G , что приводит к изучению ограничения представления V^∂ на связную компоненту G^0 . Пространство V называется *неприводимым в смысле Ли*, если это представление группы G^0 неприводимо (иными словами, если V «абсолютно» неприводимо). Приведем типичный результат, взятый из [Ка4] и [Be], который можно вывести из классификации полупростых алгебр Ли и результата Габбера о подгруппах в их нормализаторах:

Предложение 4. Пусть G — связная подгруппа в $GL(n)$, действующая неприводимым образом в $\mathbb{C}^n$, нормализатор которой содержит диагональную матрицу μ .

(i) Если $\mu = (x, 1, \dots, 1)$ имеет порядок > 2 (соответственно $= 2$), то $DG = SL(n)$ (соответственно $SL(n)$ или $SO(n)$).

(ii) Если $n = 7$ и μ топологически порождает алгебраический тор $(x, y, xy, x^{-1}, y^{-1}, (xy)^{-1}, 1)$, то $DG = SL(7)$, $SO(7)$ или G_2 .

Вот пример, иллюстрирующий это предложение:

Пример 3 [BBH, BH, Ка4]. Обобщенное гипергеометрическое уравнение H :

$$(\theta + a_1) \dots (\theta + a_p) y - z^{-1} (\theta + c_1 - 1) \dots (\theta + c_q - 1) y = 0 \quad (\theta = z d/dz).$$

Оно неприводимо тогда и только тогда, когда разности $a_i - c_j$ не являются целыми. Если $q > p$, то его особенности суть $\{0, \infty\}$, и уравнение может стать приводимым лишь над циклическим расширением $C(z^{1/N})$ поля $C(z)$, и это условие легко выразить в терминах параметров уравнения (по поводу универсального метода вычисления группы $\text{Gal}(D)/\text{Gal}(D)^0$ см. [Ka3, 1.2.5]); более тонкий анализ требуется в фуксовом случае $q = p$, где $\Sigma = \{0, 1, \infty\}$. Предположим к тому же, что H неприводимо в смысле Ли. При $q = p$, применяя предложение 4(i) к G^0 и образующей μ группы $M_{1,an}$, получаем, что $G = \text{Gal}(H)$ содержит $SO(n)$ ($SL(n)$), если число $2(\sum_i(a_i) - \sum_j(c_j))$ не является четным (соответственно целым). При $q > p$, применяя обобщение предложения 4 к G^0 и топологической образующей μ тора $T_{\infty,f}$, можно свести изучение группы G к исследованию линейных зависимостей между корнями $(q-p)$ -й степени из 1. Утверждение (ii) даст в этом случае, что $\text{Gal}(H)^0 = G_2$ для $(p, q; \{a\}, \{c\}) = (1, 7; \{\frac{1}{2}\}, \{\frac{1}{7}; 0 \leq j < 6\})$. Заметим, что часть этих результатов также была получена методами разд. 3; по поводу скрученной версии этого примера см. [DM] и [Mi, 4.8.2].

Ограничимся теперь фуксовым случаем $q = p = n$ и предположим, напротив, что H не является неприводимым в смысле Ли. В общем случае представление V^0 группы G индуцировано с представления подгруппы конечного индекса d , делящего n , и изучение H сводится с помощью теоремы типа Римана—Гильберта, принадлежащей Левельту (см. [Ka4, 3.5.4]), к изучению гипергеометрического уравнения порядка n/d . Но в противоположном случае, когда ограничение представления на G^0 является изотипическим и приводимым, изучение группы $M_{1,an}$ (см. [Ka4, 3.5.7]) показывает, что представление является мономиальным! Если вронскиан уравнения H является алгебраической функцией, то G^0 тривиальна и H допускает алгебраический базис решений. Бёкерс и Хекман, которым принадлежит данное решение обобщенной проблемы Шварца, показали в [BH], что этот случай возникает тогда и только тогда, когда параметры a_i, c_j рациональны и для любого автоморфизма ξ порожденного ими кругового расширения поля $\mathbb{Q}$ корни из единицы $\xi(e(a_1)), \dots, \xi(e(a_n))$ и $\xi(e(c_1)), \dots, \xi(e(c_n))$ чередуются на единичной окружности.

Наконец, случай, когда само H является приводимым, сводится к изложенным выше с помощью техники когомологий Галуа, подсазанной теорией Куммера (см. [Bo]). (Этот случай может быть исследован также методами разд. 3 [DM].)

Отметим в заключение, что Кац ввел в [КаЗ] фильтрацию на формальной группе Галуа в иррегулярной особенности, что позволило уточнить информацию, заключенную в экспоненциальном торе $T_{\sigma,f}(D)$. Для сохранения равновесия мы не будем говорить о ней более подробно, также как и о фильтрации Жевре, введенной Рамисом.

5. ГИПОТЕЗА ГРОТЕНДИКА (по Кацу)

Метод, который мы изложим в заключение нашего обзора, описывает, в обозначениях разд. 1, « K -форму» $\text{Gal}(V/K)$ дифференциальной группы Галуа $\text{Gal}(D)$. Для этого, в частности, не требуется никакой информации о функторах слоя решений типа V^∂ . Принципиальный недостаток этого метода (или принципиальное достоинство, смотря на чей вкус) состоит в его на настоящий момент гипотетическом характере.

Хорошо известно, что коэффициенты всякого интересного дифференциального уравнения принадлежат полю $\mathbf{Q}(z)$ (с помощью стандартной техники [Ка2] можно всегда свести дело к случаю, когда скаляры являются алгебраическими числами над $\mathbf{Q}$). Итак, пусть V есть n -мерное векторное пространство над полем $K = \mathbf{Q}(z)$ со связностью ∇ (в обозначениях разд. 1 $\nabla(\partial) = D$, где, скажем, $\partial = d/dz$). Для любого достаточно большого простого числа p рассмотрим векторное пространство со связностью (V_p, ∇_p) над дифференциальным полем $(\mathbf{F}_p(z), \partial_p = d/dz)$, полученное из V редукцией $\text{mod } p$. Тогда (подробнее об этом см. [Ка1]) $(\partial_p)^p$ является дифференцированием поля $\mathbf{F}_p(z)$ и $\psi_p(\nabla(\partial)) = (\nabla_p(\partial_p))^p - \nabla_p((\partial_p)^p) \in \text{End}_{\mathbf{F}_p}(V_p)$ есть $\mathbf{F}_p(z)$ -линейное отображение. Оно (а точнее, оператор ψ_p) называется *оператором p -кривизны* связности ∇ . Если мы определим по индукции последовательность матриц A_k , так что $A_1 = A(z)$ — матрица линейной части оператора D в базисе V над $\mathbf{Q}(z)$ и $A_{k+1} = \partial A_k + A A_k$, то p -кривизна задается в соответствующем базисе V_p над $\mathbf{F}_p(z)$ редукцией $\text{mod } p$ матрицы $A_p(z)$.

В этих условиях горизонтальные векторы формального расширения $\hat{V}_p$ пространства V_p , полученного расширением скаляров до $\mathbf{F}_p((z))$, образуют векторное пространство размерности n_p над полем констант $\mathbf{F}_p((z^p))$, получаемое расширением скаляров из $\mathbf{F}_p(z^p)$ -пространства горизонтальных векторов пространства V_p (теорема Хонды), и $n_p = n$ (т. е. связность «интегрируема») тогда и только тогда, когда p -кривизна $\psi_p(\nabla(\partial))$ связности ∇ равна нулю (теорема Картье). Следовательно, по теореме Эйзенштейна дифференциальная группа Галуа $\text{Gal}(V/K)$ может быть конечной, а ее

алгебра Ли $\mathrm{gal}(V/K)$ — нулевой, только если p -кривизны $\psi_p(\nabla(\partial))$ связности ∇ равны нулю для всех достаточно больших p (см. [La]). Более общим образом, по соображениям нётеровости можно ввести в рассмотрение наименьшую алгебраическую K -подалгебру Ли $\mathrm{countb}(V/K) \subset \mathfrak{gl}(n, K)$, редукции которой $\bmod p$ содержат $\psi_p(\nabla(\partial))$ для всех достаточно больших простых p . Из определения p -кривизн следует, что $\mathrm{countb}(V/K)$ содержится в $\mathrm{gal}(V/K)$. Следующая гипотеза Каца [Ka2] обобщает гипотезу Гротендика, относящуюся к случаю $\mathrm{countb}(V/K) = 0$:

Гипотеза. *Алгебры Ли $\mathrm{countb}(V/K)$ и $\mathrm{gal}(V/K)$ совпадают.*

Отметим формальную аналогию этой гипотезы с теоремой Амброуза—Зингера [AS], согласно которой алгебра Ли группы голономии риманова многообразия M порождается операторами кривизны во всех точках этого многообразия (точнее, сопряженными к ним при помощи параллельного переноса). Кроме того, Андре [An2] сопоставил любой связности над $\mathbb{F}_p(z)$ «схему дифференциальных групп Галуа» (не обязательно приведенную), алгебра Ли которой порождена p -кривизной этой связности. Таким образом, гипотеза Каца эквивалентна тому, что $\mathrm{Gal}(V/K)^0$ является наименьшей алгебраической подгруппой в $\mathrm{Aut}_K(V)$, редукции которой $\bmod p$ содержат эти подсхемы почти для всех p .

Рассмотрим нашу «тестовую» ситуацию с двумя особенностями $\Sigma = \{0, \infty\}$. В этой ситуации гипотеза доказана в [Ka2, 10.3] редукцией к случаю $\mathrm{countb}(V/K) = 0$, в котором она вытекает из [Ka1]. Ее нелегко проверить непосредственно. Во всех случаях обнаруживается таинственная связь между группами Стокса и p -кривизнами, поскольку (даже если рассматривать только примеры 2 и 3 при $q > p$) формальные группы Галуа сами по себе не порождают $\mathrm{Gal}(V/K)^0$. Отметим, однако, что логарифмы автоморфизмов Стокса нильпотентны, в то время как при наличии иррегулярных особенностей большая часть p -кривизн этим свойством не обладает [Ka2, 8.1(1)]. К тому же расходящиеся ряды, возникающие в формальных изоморфизмах $\hat{\Phi}$ (и ответственные за феномен Стокса), при редукции $\bmod p$ часто сводятся к полиномам (типичный пример — ряд Эйлера $\sum_{n \geq 0} (n!)z^n$); лишь p -адический анализ позволяет просчитать все до конца (см. [Dw]).

На самом деле, за исключением некоторых уравнений с унипотентными группами Галуа [An1, 8.1.3; BR], во всех случаях, когда гипотеза Каца проверена, она сводится к гипотезе Гротендика (см. [Ka3, 10.2]), где, собственно и не возникает необходимости в вы-

числении $\text{contb}(V/K)$, ввиду теоремы Картье... Помимо рассмотренного выше случая, имеются еще уравнения первого порядка на P_1 [Ho] и гипергеометрические уравнения ($[Ka1, BH]$; см. выше пример 3, а также [Dw]). Тут новый подход, основанный на более тонкой версии критерия трансцендентности Шнайдера—Ленга, предложен Чудновскими [CC]. Он позволяет проверить гипотезу Гротендика для всех дифференциальных уравнений первого порядка на римановой поверхности (произвольного рода) и для уравнений Ламе (которые, как подчеркнул Кац, являются уравнениями Эйри на эллиптической кривой — по поводу нахождения соответствующих списков Шварца см. [BD, Ch]).

С точки зрения «вычисления $\text{Gal}(D)$ », которая нас здесь интересует, поразительным в доказательствах из [CC] является то, что нет необходимости располагать всей локальной информацией, нужна лишь информация, обеспечиваемая некоторым конечным набором *эффективно вычислимых* простых чисел p . Ввиду эффективной версии теоремы Чеботарева [Se, th. 5], то же замечание применимо и к доказательству Хонды [Ho]. Это приводит к следующему вопросу, уточняющему гипотезу Каца, который мы обсудим в заключение. Назовем высотой (V, ∇) максимум степеней и логарифмических высот коэффициентов матрицы A , задающей оператор D , и их общего знаменателя T . Будем говорить, что ∇ имеет хорошую редукцию в p , если $p > n$ и p не делит ни один из коэффициентов T . Сделанный выбор базиса пространства V позволяет определить (логарифмическую) высоту $h(X)$ векторного K -подпространства $X \subset V$. Вопрос состоит в следующем:

(?) *Существуют ли эффективно вычислимые постоянные $c(n)$, $\kappa(n)$, удовлетворяющие следующему условию: пусть $h \in \mathbf{R}$, $h \geq 1$, ∇ — связность высоты $\leq h$ на V и X есть K -прямая в V , такая, что при любом простом $p \leq c(n)(h + h(X))^{\kappa(n)}$, дающем хорошую редукцию для ∇ , редукция X_p подпространства $X \bmod p$ инвариантна относительно $\psi_p(\nabla(\partial))$; тогда X инвариантно относительно $\text{Gal}(V/K)^0$.*

ЛИТЕРАТУРА

(С точностью до перестановки букв, сокращения L.O.D.E., S.E.D., ... обозначают: линейные обыкновенные дифференциальные уравнения (linear ordinary differential equations), системы дифференциальных уравнений (systèmes des équations différentielles), ...)

- [An1] André Y. *G*-functions and geometry, Aspects of Math., vol. 13, 1989.
- [An2] André Y. Notes sur la théorie de Galois différentielle, Prépubl. IHES, 1989.
- [AS] Ambrose W., Singer I. A theorem on holonomy, Trans. Amer. Math. Soc., **75** (1953), 428–443. (См. также: Bourguignon J-P. Astérisque, **126** (1985), 169–183.)
- [BV1] Babbitt D., Varadarajan V. Formal reduction of meromorphic D.E.: a group theoretic view, Pacific J. Math., **108** (1983), 1–80.
- [BV2] Babbitt D., Varadarajan V. Local moduli for meromorphic D.E., Astérisque, **169–170** (1989).
- [BD] Baldassarri F., Dwork B. On second order D.E. with algebraic solutions, Amer. J. Math., **101** (1979), 42–76.
- [BBRS] Balser W., Braaksma B., Rainis J.-P., Sibuya Y. Multisummability of formal power series solutions of L.O.D.E., Asympt. Anal., **5** (1991), 27–45.
- [Be] Beukers F. Differential Galois theory, in: From Number Theory to Physics, Springer-Verlag, 1992, 413–419.
- [BBH] Beukers F., Brownawell D., Heckman G. Siegel normality, Ann. of Math., **127** (1988), 279–308.
- [BH] Beukers F., Heckman G. Monodromy for the hypergeometric function ${}_nF_{n-1}$, Invent. Math., **95** (1989), 325–354.
- [BR] Beziniv J.-P., Robba P. Rational solutions of L.D.E., J. Austr. Math. Soc., **46** (1989), 184–196.
- [Bl] Blanchard A. Groupes algébriques et équations différentielles linéaires [d'après E. Kolchin], Sem. Bourbaki, n° 17 (1949/50), Collection Hors Série Asterisque, **1** (1995), 103–109.
- [Bo] Boussel K. Groupes de Galois et équations hypergéométriques..., C.R. Acad. Sci., Paris, **309** (1989), 587–589.
- [Ch] Chiarellotto B. On Lamé operators which are pull-backs of hypergeometric ones, Prépubl. Univ. Padova, 1991.
- [CC] Chudnovsky D., Chudnovsky G. Applications of Padé approximations to the Grothendieck conjecture on L.D.E., Lect. Notes in Math., vol. 1135, Springer-Verlag, 1984, 52–100.
- [De] Deligne P. Catégories tannakiennes, in: Grothendieck Festschrift II, Prog. in Math., vol. 87, Birkhäuser, 1990, 111–195.
- [DM] Duval A., Mitschi C. Matrices de Stokes et groupes de Galois des équations hypergéométriques..., Pacific J. Math., **138** (1989), 25–56.
- [Dw] Dwork B. Hypergeometric functions, Oxford Univ. Press, 1990.
- [Ho] Honda T. Algebraic differential equations, Symp. Math., **24** (1981), 169–204.
- [IK] Ильяшенко Ю. С., Хованский А. Г. Группы Галуа, операторы Стокса и теорема Рамиса. — Функц. анализ и его прил., т. 24, 1990, вып. 4, 31–42.
- [Ju] Jurkat W. Meromorphe Differentialgleichungen, Lect. Notes in Math., vol. 637, Springer-Verlag, 1978.
- [Kap] Kaplansky I. An introduction to differential algebra, Hermann, 1957.

- [Ka1] Katz N. Algebraic solutions of differential equations; p -curvature and the Hodge filtration, *Invent. Math.*, **18** (1972), 1–118.
- [Ka2] Katz N. A conjecture in the arithmetic theory of differential equations, *Bull. Soc. Math., France*, **110** (1982), 203–239, 347–348.
- [Ka3] Katz N. On the calculation of some differential Galois groups, *Invent. Math.*, **87** (1987), 13–61.
- [Ka4] Katz N. Exponential sums and differential equations, Princeton Univ. Press, 1990.
- [Ko1] Kolchin E. Differential algebra and algebraic groups, Academic Press, 1973.
- [Ko2] Kolchin E. Algebraic groups and algebraic dependence, *Amer. J. Math.*, **90** (1968), 1151–1164.
- [La] Landau E. Eine Anwendung des Eisensteinsches Satz auf die Theorie der Gaussche Differentialgleichung, *Crelle*, **127** (1904), 92–102.
- [Le] Levelt A. Differential Galois theory and tensor products, *Indag. Math.*, **1** (1990), 439–449.
- [LR] Loday-Richaud M. Classification méromorphe locale des S.D.L.... : phénomène de Stokes et applications, Thèse, Univ. Orsay, 1991.
- [Ma1] Malgrange B. La classification des connexions irrégulières à une variable, in: *Prog. in Math.*, vol. 37, Birkhäuser, 1983, 381–399.
- [Ma2] Malgrange B. Équations différentielles à coefficients polynomiaux, *Prog. in Math.*, vol. 96, Birkhäuser, 1991.
- [M-R] Malgrange B., Ramis J.-P. Fonctions multisommables, *Ann. Inst. Fourier*, **42** (1992), No. 1–2, 353–368.
- [MR1] Martinet J., Ramis J.-P. Théorie de Galois différentielle et resommation, in: *Computer Algebra and Differential Equations*, E. Tournier (ed.), Academic Press, 1989, 117–214.
- [MR2] Martinet J., Ramis J.-P. Elementary acceleration and multisummability, *Ann. I.H.P., Phys. Th.*, **54** (1991), 331–401.
- [Mi] Mitschi C. Differential Galois groups and G -functions, in: *Differential Equations and Computer Algebra*, M. Singer (ed.), Academic Press, 1991, 149–180.
- [Ne] Нестеренко Ю. В. Об алгебраической зависимости компонент решений системы линейных дифференциальных уравнений. — *Изв. АН СССР, сер. матем.*, т. 38, 1974, № 3, с. 495–512.
- [Po] Pommaret J.-F. Differential Galois Theory, Gordon & Breach, 1982.
- [Ra1] Ramis J.-P. Théorèmes d'indice Gevrey pour les E.D.O., *Mem. Amer. Math. Soc.*, **296** (1984).
- [Ra2] Ramis J.-P. Filtration Gevrey sur le groupe de Picard–Vessiot d'une E.D. irrégulière, *Prépubl. IMPA*, **45** (1985), 1–38.
- [RS] Ramis J.-P., Sibuya Y. Hukuhara's domains and fundamental existence and uniqueness theorems for asymptotic solutions of Gevrey type, *Asymp. Anal.*, **2** (1989), 39–94.
- [Sa] Sabbah C. Connexions méromorphes formelles des deux variables, *Prépubl. Éc. Polytechnique*, 1991.

- [Sc] Schwarz H. Ueber diejenige Fälle in welchen die Gaussche hypergeometrische Reihe einer algebraische Funktion ihres 4. en Elementes darstellt, Crelle, **75** (1873), 292–335.
- [Se] Serre J.-P. Quelques applications du théorème de densité de Chebotarev, Publ. Math. I.H.E.S., **54** (1981), 123–201.
- [Si] Sibuya Y. L.D.E. in the complex domain: problems of analytic continuation, Transl. Math. Monographs, vol. 82, Amer. Math. Soc., 1990.
- [Sn1] Singer M. Algebraic relations among solutions of L.D.E.: Fano's theorem, Amer. J. Math., **110** (1988), 115–144.
- [Sn2] Singer M. An outline of differential Galois theory, in: Computer Algebra and Differential Equations, Academic Press, 1990, 3–57.
- [Sn3] Singer M. Moduli of L.D.E. on the Riemann sphere with fixed Galois group, Pacific. J. Math., **106** (1993), 343–395.
- [St] Stokes G. Early letters to Lady Stokes, Cambridge Univ. Press., 1907. Прочитано в: Ramis J.-P. Séries divergentes et théories asymptotiques, Panoramas et Synthèses, Soc. Math. de France, 1993.
- [Um] Umemura H. Second proof of the irreducibility of the first D.E. of Painlevé, Nagoya Math. J., **117** (1990), 125–171.
- [Va] Varadarajan V. Meromorphic differential equations, Expos. Math., **9** (1991), 97–188.

[Добавлено в октябре 1992]

- (1) То есть C -линейным отображением из V в $V \otimes \Omega_{K/C}^1$, удовлетворяющим формуле Лейбница.
- (2) Под «тензорными конструкциями» понимаются прямые суммы тензорных произведений пространств V и V^* , снабженные естественной связностью. Помимо [De] см. об этом также: Breen L. Tannakian categories, in: Motives (Seattle, 1991), Proc. Symp. Pure Math., vol. 55, Part 1, Amer. Math. Soc., 1994, 337–376 (в оригинале ссылка на еще не вышедшее издание — *Ред.*).
- (3) В этом утверждении мы предполагаем фиксированным согласованный выбор локальных групп в своих классах сопряженности в $\text{Gal}(D)$: например, если мы рассматриваем представление группы $\text{Gal}(D)$ (соответственно $G_{\sigma, an}$), связанное с простой точкой s относительно V (соответственно с точкой σ' , близкой к σ), то этот выбор может быть задан любой системой путей из s в каждую из точек σ' .
- (4) И проверить, как в [IK], что это поднятие согласовано с тензорными конструкциями и, следовательно, с группой Галуа. Этот метод в общем случае получил развитие у П. Дели-

- ния (письмо Ж.-П. Рамису, февраль 1986) в терминах локальных I -фильтрованных систем, ассоциированных с парой $\{V, \widehat{\Phi}\}$ (ср. [Ma1, BV2]).
- (5) Здесь также удобно зафиксировать согласованный выбор представителей st_α в своих классах сопряженности: для представления группы $G_{0,an}$, связанного с точкой s в окрестности 0, мы можем, например, фиксировать набор точек σ , аргументы которых приближаются снизу к прообразам в $\tilde{S}$ сингулярных направлений α , и набор путей, соединяющих s и σ . Кроме того, сопряжение с $\Sigma_{\arg(s)}$ переводит $G_{0,f}(D)$ в подгруппу данного представления $G_{0,an}$; именно эта подгруппа фигурирует в утверждении теоремы 2.
- (6) $(\widehat{V}_0, \widehat{D}_0)$ получено пополнением пространства со связностью (V_0, D_0) , снабженного дифференциальным оператором, таким, что $D_0 \ll \widehat{D}_0$ на $V_0 = K^n$. Заметим, что каноническое вложение K в $\widehat{K}$ продолжается до инъективного гомоморфизма расширения Пикара—Вессю V_0/K в $\widehat{V}_0/\widehat{K}$. На протяжении этого раздела мы фиксируем один из таких гомоморфизмов.
- (7) См. также Loday-Richaud M. Stokes phenomenon, multisummability and differential Galois groups, Prépubl. Univ. Orsay, n° 1371, 1992; Ann. Inst. Fourier, 44 (1994), No. 3, 849–906.
- (8) См. однако работы М. Зингера (Singer M., J. Symb. Comp., 11 (1991), 251–273), Ф. Ульмера (Ulmer F. Appl. Algebra Engrg. Comm. Comput., 2 (1992), No. 3, 171–193) и А. Дюваля и М. Лоде-Ришо (Duval A., Loday-Richaud M., Appl. Algebra Engrg. Comm. Comput., 3 (1992), No 3, 211–246) по поводу алгоритма Ковачича.

DANIEL BERTRAND

Université de Paris VI
U.F.R. de Mathématiques
Case 247
4, place Jussieu
F-75252 PARIS CEDEX 05

E-mail: bertrand@math.jussieu.fr

ЗНАЧЕНИЯ L -ФУНКЦИЙ МОТИВОВ В ЦЕЛЫХ ТОЧКАХ

Жан-Марк Фонтен

Прошло ровно семь лет с тех пор, как Суле сделал на семинаре Бурбаки доклад [So85] о гипотезах Делиня и Бейлинсона¹⁾, дающих, среди прочего, формулу для $L(H^m(X), r)$, где X — гладкое проективное многообразие над полем F и $r \in \mathbb{Z}$, с *точностью до умножения на рациональное число*. Центральной темой данного доклада являются гипотезы Блоха и Като [BK90], позволяющие предсказать значение этого рационального числа.

Я сделал спорный выбор, предпочтя гипотезы результатам. Последние являются весьма частными (см. по этому поводу доклады Б. Перрин-Риу [Pe89], Гросса [Gr91] и Рубина [Ru91] о гипотезе Бёрча и Свиннертона-Дайера, книгу [RSS88] и доклады Рамакришнана [Ra89] и Денингера—Шолля [DS91] о гипотезах Делиня и Бейлинсона), но достаточно обнадеживающими для того, чтобы считать, что ситуация будет быстро меняться.

1. ВВЕДЕНИЕ

В дальнейшем E и F будут обозначать конечные расширения поля $\mathbb{Q}$. Можно надеяться, что со временем появится хорошее определение категории $MM_F(E)$ *смешанных мотивов над F с коэффициентами в E* . Каждому смешанному мотиву M можно сопоставить *функцию* $L(M, s)$ со значениями в $\mathbb{C} \otimes E$, голоморфную при $\operatorname{Re} s \gg 0$. Ожидается, что эта функция допускает мероморфное продолжение на всю комплексную плоскость, и хотелось бы дать интерпретацию, по крайней мере гипотетическую, ее значений в целых точках; иначе говоря, хотелось бы обобщить гипотезу Бёрча и Свиннертона-Дайера, дающую значение $L(M, 0)$ в случае, когда M — чистый мотив веса -1 , ассоциированный с абелевым многообразием, определенным над F (где $E = \mathbb{Q}$). Примерами рассматриваемых нами L -функций являются L -функции Артина, ассоциированные с конечномерными представлениями группы $\operatorname{Gal}(\bar{F}/F)$

¹⁾ См. также оригинальные статьи Блоха [Bl84], Делиня [De79] и Бейлинсона [Be85].

с коэффициентами в E , L -функции, ассоциированные с «новыми» модулярными формами, и функции $L(H^m(X), s)$, где X — проективное неособое многообразие над F и $m \in \mathbb{N}$.

Заменяя M на его «скручивание по Тейту» $M(i)$, можно ограничиться изучением значения L -функции в точке $s = 0$. Пусть r_M — единственное целое число, такое, что $s^{-r_M} L(M, s)$ стремится к конечному ненулевому пределу $L^*(M, 0)$ при $s \rightarrow 0$.

Обозначим через $1_{F,E}$ «единичный объект» категории $\mathcal{MM}_F(E)$ и положим для любого $i \in \mathbb{N}$

$$H^i(F, M) = \text{Ext}_{\mathcal{MM}_F(E)}^i(1_{F,E}, M).$$

Можно определить (см. разд. 6 ниже) E -векторное подпространство $H_f^1(F, M)$ пространства $H^1(F, M)$, классифицирующее расширения $1_{F,E}$ с помощью M , имеющие «хорошую редукцию во всех конечных точках поля F ».

«Гипотеза» $C_r(M)$ ¹⁾. Пусть $M^*(1)$ — «двойственный мотив к M , скрученный по Тейту»; тогда

$$r_M = \dim_E H_f^1(F, M^*(1)) - \dim_E H^0(F, M^*(1)).$$

Определена реализация де Рама M_{dR} мотива M , являющаяся $E \otimes_{\mathbb{Q}} F$ -модулем конечного типа, снабженным некоторой убывающей фильтрацией $(\text{Fil}^i M_{dR})_{i \in \mathbb{Z}}$ своими $E \otimes_{\mathbb{Q}} F$ -подмодулями — фильтрацией Ходжа. Обозначим через t_M и назовем касательным пространством к M фактор $M_{dR}/\text{Fil}^0 M_{dR}$. Пусть p принадлежит $S_{\infty}(F)$, множеству бесконечных точек поля F . Тогда для нее определена реализация Бетти $M_{B,p}$ мотива M , являющаяся конечномерным E -векторным пространством, снабженным, если точка p вещественна, действием комплексного сопряжения c . Положим $M_{B,p}^+ = (M_{B,p})^c$, если p вещественна, и $M_{B,p}^+ = M_{B,p}$ в противном случае.

Для любого поля K и любого K -векторного пространства V конечной размерности обозначим через $\det_K V = \Lambda_K^{\dim V} V$ детерминант пространства V и через $\det_K^* V$ двойственное пространство к $\det_K V$. Положим

$$L_f(M) = \det_E H^0(F, M) \otimes \det_E^* H_f^1(F, M),$$

¹⁾ Поскольку на практике не всегда известно, что функция L допускает аналитическое продолжение на $\mathbb{C}$, эту гипотезу нужно понимать следующим образом: утверждается, что функция $L(M, s)$, определенная при $\text{Re}(s) \gg 0$, допускает мероморфное продолжение в некоторую окрестность нуля, и ее порядок в точке $s = 0$ определяется данной формулой.

$$L_f(M^*(1)) = \det_E H^0(F, M^*(1)) \otimes \det_E^* H_f^1(F, M^*(1))$$

и введем, наконец, *фундаментальную прямую мотива* M

$$\Delta_f(M) = L_f(M) \otimes L_f(M^*(1)) \otimes \det_E^* \left(\bigoplus_{p \in S_\infty(F)} M_{B,p}^+ \right) \otimes \det_E t_M.$$

Мы увидим, что, по модулю некоторых «гипотетических» свойств категории $\mathcal{MM}_F(E)$, можно

а) определить канонический изоморфизм

$$\iota_M : \mathbf{R} \otimes_{\mathbf{Q}} \Delta_f(M) \rightarrow \mathbf{R} \otimes_{\mathbf{Q}} E;$$

б) для любой конечной точки λ поля E , используя λ -адическую реализацию M_λ мотива M , снабдить $E_\lambda \otimes_E \Delta_f(M)$ канонической нормой $|\cdot|_{\lambda,EP}$.

«Гипотеза» $C_{DB}(M)$. Существует (обязательно единственный) базис b в $\Delta_f(M)$, такой, что

$$\iota_M(1 \otimes b) \cdot L^*(M, 0) = 1.$$

«Гипотеза» $C_{BK}(M)$. «Гипотеза» $C_{DB}(M)$ верна, и для любой конечной точки λ поля E выполнено равенство

$$|1 \otimes b|_{\lambda,EP} = 1.$$

Очевидно, что $C_{DB}(M)$ (соотв. $C_{BK}(M)$) определяет $L^*(M, 0)$ с точностью до умножения на элемент из E^* (соотв. из $\mathcal{O}_E^*$). Следовательно, когда $E = \mathbf{Q}$, гипотеза $C_{BK}(M)$ определяет вещественное число $L^*(M, 0)$ с точностью до знака (который «легко» определить: ожидается, что нули и полюсы функции $L(M, s)$, лежащие на вещественной прямой, являются целыми числами; имеем $r_{M(i)} = 0$ для $i \gg 0$, и знак должен быть равен четности числа $\sum_{i>0} r_{M(i)}$).

Я взял здесь слово «гипотеза» в кавычки, поскольку в этих утверждениях рассматриваются объекты категории $\mathcal{MM}_F(E)$, точного определения которой я не знаю. Тем не менее известно много объектов этой категории.

Пусть M — «известный мотив». По крайней мере в двух случаях можно получить «корректную» гипотезу:

- когда M является f -замкнутым, т.е. таким, что $H_f^1(F, M) = H_f^1(F, M^*(1)) = 0^{1)}$ (если M является чистым веса $\neq -1$, то это все равно, что сказать, что M является *критическим*);
- когда M — чистый мотив веса -1 , связанный с некоторым абелевым многообразием A над F , снабженным вложением E в $Q \otimes \text{End}(A)$. Теория 1-мотивов Делиня дает тогда некоторую корректно определенную полную подкатегорию в $MM_F(E)$, которая является достаточно большой для всех необходимых вычислений (см. разд. 8).

Кроме того, и это «классическая» точка зрения Бейлинсона, можно обойтись без использования категории $MM_F(E)$, введя «мотивные когомологи», т.е. некоторые E -векторные пространства, определенные с помощью алгебраической K -теории, о которых предполагается, что их можно отождествить с $H_f^1(F, M)$ и $H_f^1(F, M^*(1))$. Кроме того, требуется определить некоторые E -линейные отображения из этих двух векторных пространств, позволяющие построить «регуляторы» и в некоторых случаях аналог *высоты* Нерона–Тейта для абелевых многообразий. Наиболее важным является случай $M = H^m(X)(r)$, где X — гладкое проективное многообразие над F , $m \in \mathbb{N}$, $r \in \mathbb{Z}$ (и $E = Q$); см. разд. 9.

После введения некоторых обозначений и соглашений (разд. 2), мы переходим к *первой части*, посвященной *общему формализму*, где приводим обзор результатов о представлениях Галуа с коэффициентами в E_λ (разд. 3); затем объясняем (разд. 4), как определить каноническую норму $|\cdot|_{\lambda, E, P}$, исходя из λ -адической реализации; в разд. 5 обсуждаются смешанные структуры Ходжа; в некотором смысле именно они играют роль λ -адических реализаций в архимедовых точках; далее (разд. 6) описываются формальные свойства, которым должна удовлетворять категория $MM_F(E)$, определяется $H_f^1(F, M)$ и строится отображение ι_M , используемое в гипотезе $C_{BD}(M)$; в разд. 7 обсуждается, как гипотезы упрощаются при некоторых ограничениях на веса мотива M или если M есть f -замкнутый мотив.

Во *второй части* обсуждаются различные *частные случаи*: 1-мотивы Делиня (разд. 8), мотивы вида $H^m(X)$, где X — гладкое

¹⁾ Если гипотетический формализм, связанный с категорией $MM_F(E)$, верен, то существует универсальный способ сопоставления каждому мотиву M некоторого f -замкнутого мотива N , и наши гипотезы верны для M тогда и только тогда, когда они верны для N .

проективное многообразие над F (разд. 9), смешанные мотивы Тейта, скрученные мотивы Артина, ... (разд. 10).

Третья часть содержит различные дополнения: в разд. 11 объясняется, каким образом при вычислении чисел, входящих в $С_{BK}(M)$, появляются группы Шафаревича и числа Тамагавы. Это приводит к другой формулировке гипотезы $С_{BK}(M)$, совпадающей (по существу) с формулировкой гипотезы Бёрча—Свиннертона-Дайсера в случае абелевых многообразий и с оригинальной формулировкой Блоха—Като [BK90] для тех случаев, которые они рассматривали. Наконец, в разд. 12 обсуждается понятие *полустабильности* E_X -представлений и мотивов (ожидается, что любой мотив над F становится полустабильным над некоторым конечным расширением поля F). Именно для полустабильных мотивов можно надеяться развить технику, которая, возможно, позволит достичь значительного прогресса в данной теории. В этом контексте также можно интерпретировать значения неполных L -функций, понять, почему обсуждаемые здесь гипотезы ведут себя хорошо относительно коротких точных последовательностей и являются совместимыми с функциональным уравнением и каким образом они отражают некоторые свойства категории смешанных мотивов (когомологическая размерность 1, F -полупростота, ...).

Эти «гипотезы» и/или их различные варианты являются результатом (если так можно сказать) усилий многих математиков. После пионерских работ Бёрча и Свиннертона-Дайсера [BS63, BS65], расширенных и углубленных Тейтом [Ta66], наиболее существенный вклад, без сомнения, внесли Блох [Bl84], Бейлинисон [Be84] и Делинь [De79] для $С_{BD}(M)$ и Блох и Като [BK90] для $С_{BK}(M)$. С развиваемой здесь точки зрения некоторые идеи Блоха [Bl80], Делиня [De85], Янсена [Ja88, Ja89, Ja90], Като [Ka91], Лихтенбаума [Li72], Перрин-Риу [FP91] и Шолля [Sc91] также сыграли большую роль¹⁾.

Наконец, я хочу поблагодарить Бернадетт Перрин-Риу, которой этот текст столь многим обязан: в течение долгого времени мы обсуждали вместе данный предмет, и именно наша совместная работа послужила для меня стимулом к этому докладу.

¹⁾ Наше изложение в значительной степени обязано работам [FP91, FP92] и [Ka91]. Точка зрения Блоха и Като излагается в разд. 11. Оригинальные гипотезы рассматривали менее общую ситуацию, и докладчик несет ответственность за всю ту бессмыслицу, которая могла возникнуть в результате изменения точки зрения и обобщений, время для которых еще не пришло.

2. ОБОЗНАЧЕНИЯ И СОГЛАШЕНИЯ

Если G — проконечная группа и λ — конечная точка поля E , то λ -адическое представление группы G — это E_λ -векторное пространство конечной размерности, снабженное линейным непрерывным действием группы G . Имеем $H^0(G, V) = V^G$, в то время как $H^1(G, V)$ классифицирует расширения тривиального представления E_λ с помощью V . В случае, когда $G = \text{Gal}(\bar{K}/K)$, где $\bar{K}$ — сепарабельное замыкание поля K , мы пишем также $H^i(K, V)$ вместо $H^i(G, V)$.

Для любого числового поля K обозначим через $S(K)$ множество всех его точек. Отождествим $S(\mathbb{Q})$ с множеством {простые числа} $\cup \infty$. Для любого $p \in S(\mathbb{Q})$ обозначим через $S_p(K)$ множество точек K , лежащих над p .

Выберем алгебраическое замыкание $\bar{F}$ поля F и для всякой точки $p \in S(F)$ алгебраическое замыкание $\bar{F}_p$ полного поля F_p и вложение $\bar{F}$ в $\bar{F}_p$. Пусть $G_F = \text{Gal}(\bar{F}/F)$ и $G_p = \text{Gal}(\bar{F}_p/F_p) \subset G_F$. Если p — бесконечная точка, то выберем отождествление замыкания $\bar{F}_p$ с $\mathbb{C}$. Если p — конечная точка, лежащая над простым числом p , то обозначим через I_p подгруппу инерции в G_p ; для всякого расширения L поля F_p , содержащегося в $\bar{F}_p$, обозначим через L_0 максимальное неразветвленное расширение поля $\mathbb{Q}_p$, содержащееся в L ; поле вычетов $\bar{k}_p$ поля $\bar{F}_p$ совпадает с полем вычетов поля $(\bar{F}_p)_0$ и является алгебраическим замыканием поля вычетов k_p поля F_p ; обозначим через $\sigma_p : \bar{k}_p \rightarrow \bar{k}_p$ отображение $x \mapsto x^p$, а также лежащий над ним непрерывный автоморфизм поля $(\bar{F}_p)_0$. Группа G_p/I_p отождествляется с $\text{Gal}((\bar{F}_p)_0/(F_p)_0) = \text{Gal}(\bar{k}_p/k_p)$ и топологически порождена геометрическим морфизмом Фробениуса $f_p = \sigma_p^{-r_p}$, где $r_p = [k_p : \mathbb{F}_p]$.

А. ОБЩИЙ ФОРМАЛИЗМ

3. L-ФУНКЦИЯ И ФУНДАМЕНТАЛЬНАЯ ПРЯМАЯ λ -АДИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ГРУППЫ G_F

В дальнейшем λ обозначает конечную точку поля F , лежащую над простым числом l .

3.1. Пусть p — конечная точка поля F , лежащая над простым числом p . Отсылаем читателя к [FI90, II90] или к [Bures] за определением поля $B_{dK,p}$ p -адических периодов и его подкольца $B_{\text{cris},p}$, соответствующих расширению $\bar{F}_p/F_p$. Поле $B_{dK,p}$ является $\bar{F}_p$ -алгеброй,

снабженной убывающей фильтрацией $(\text{Fil}^i B_{dR,p})_{i \in \mathbb{Z}}$ и действием группы G_p . Кольцо $B_{\text{cris},p}$ является $(\overline{F}_p)_0$ -подалгеброй в $B_{dR,p}$, инвариантной относительно действия группы G_p и снабженной эндоморфизмом ϕ , коммутирующим с этим действием и являющимся σ_p -полулинейным. Имеем $(B_{dR,p})^{G_p} = F_p$, тогда как $(B_{\text{cris},p})^{G_p} = (F_p)_0$.

3.2. Положим $E_{\lambda,p} = E_\lambda$, если $l \neq p$, и $E_{\lambda,p} = (F_p)_0 \otimes_{\mathbb{Q}_p} E_\lambda$ в противном случае.

Пусть V есть λ -адическое представление группы G_p размерности $b(V)$. Положим

$$\underline{D}_p(V) = V^{I_p}, \text{ если } l \neq p, \text{ и } \underline{D}_p(V) = (B_{\text{cris},p} \otimes_{\mathbb{Q}_p} V)^{G_p}, \text{ если } l = p.$$

Это свободный $E_{\lambda,p}$ -модуль конечного ранга $b_p(V) \leq b(V)$, снабженный линейным действием f_p : если $l \neq p$, то это действие индуцировано действием $f_p \in G_p/I_p$ на $\underline{D}_p$; если $l = p$, то ϕ действует на $B_{\text{cris},p} \otimes V$ по формуле $b \otimes v \mapsto \phi b \otimes v$, $\underline{D}_p$ инвариантно относительно ϕ и f_p действует как ϕ^{r_p-1} .

Говорят, что V имеет хорошую редукцию, если $b_p(V) = b(V)$. Если $l \neq p$, то это значит, что V неразветвлено, в то время как если $l = p$, то это означает, что V является кристаллическим. Обозначим через $H_f^1(F_p, V)$ линейное E_λ -подпространство в $H^1(F_p, V)$, образованное теми классами W расширений поля E_λ с помощью V , для которых естественное отображение $\underline{D}_p(W) \rightarrow \underline{D}_p(E_\lambda) = E_{\lambda,p}$ является сюръективным; в том случае, когда V имеет хорошую редукцию, это условие эквивалентно тому, что W тоже имеет хорошую редукцию.

С другой стороны, если $l = p$, то положим $\underline{D}_{dR,p}(V) = (B_{dR,p} \otimes_{\mathbb{Q}_p} V)^{G_p}$. Это $F_p \otimes_{\mathbb{Q}_p} E_\lambda$ -модуль, обладающий убывающей фильтрацией $F_p \otimes E_\lambda$ -подмодулями $\text{Fil}^i \underline{D}_{dR,p}(V) = (\text{Fil}^i B_{dR,p} \otimes_{\mathbb{Q}_p} V)^{G_p}$ (где $i \in \mathbb{Z}$). Размерность $\underline{D}_{dR,p}(V)$ как векторного пространства над F_p не превосходит $[E_\lambda : \mathbb{Q}_p] \cdot b(V)$. Говорят, что V имеет тип де Рама, если имеет место равенство; в этом случае $\underline{D}_{dR,p}(V)$ является свободным модулем ранга $b(V)$ над $F_p \otimes E_\lambda$ и естественное отображение

$$B_{dR,p} \otimes_{F_p} \underline{D}_{dR,p}(V) \rightarrow B_{dR,p} \otimes_{\mathbb{Q}_p} V$$

есть изоморфизм.

¹⁾ Обратите внимание: мы берем f_p равным ϕ^{r_p} , а не его обратному!

Наконец, в случае $l = p$ назовем касательным пространством к V $F_p \otimes E_\lambda$ -модуль $t_{V,p} = ((B_{dR,p}/\text{Fil}^0 B_{dR,p}) \otimes_{Q_p} V)^{G_p}$. Если V имеет тип де Рама, то $t_{V,p}$ можно отождествить с $\underline{D}_{dR,p}(V)/\text{Fil}^0 \underline{D}_{dR,p}(V)$.

3.3. Пусть теперь V — некоторое λ -адическое представление группы G_F . Для каждой конечной точки p поля F положим

$$P_p(V, t) = \det_{E_{\lambda,p}}(1 - f_p t | \underline{D}_p(V));$$

это полином с коэффициентами в E_λ (включая случай $l = p$); имеем $\deg P_p \leq b(V)$, и равенство выполняется тогда и только тогда, когда V имеет хорошую редукцию в p .

Далее нам потребуются три следующих определения:

(i) Пусть S — конечное множество точек поля F , содержащее $S_\infty(F)$. Скажем, что V допускает L_S -функцию над E , если для всякой точки $p \notin S$ многочлен $P_p(V, t)$ принадлежит $E[t]$ и если бесконечное произведение

$$L_S(V, s) = \prod_{p \notin S} P_p(V, N(p)^{-s})^{-1}$$

сходится в $\mathbb{C} \otimes_{\mathbb{Q}} E$ для $\text{Re } s \gg 0$. В случае $S = S_\infty(F)$ будем говорить, что V допускает L -функцию над E , и положим $L(V, s) = L_{S_\infty(F)}(V, s)$.

(ii) Скажем, что V является псевдогеометрическим, если оно неразветвлено в дополнении к конечному множеству конечных точек поля F и имеет тип де Рама во всех точках, делящих l .

(iii) Обозначим через $H_f^1(F, V)$ E_λ -векторное подпространство в $H^1(F, V)$, состоящее из элементов, образ которых в $H^1(F_p, V)$ лежит в $H_f^1(F_p, V)$ при любом p . Это конечномерное векторное пространство над E_λ .

Тривиальное представление E_λ является псевдогеометрическим и допускает L -функцию над E , которая является обычной ζ -функцией Римана¹⁾, и $H_f^1(F, E_\lambda) = 0$. Свойство представлений быть псевдогеометрическими сохраняется при переходе к подобъектам, факторобъектам, прямым суммам, тензорным произведениям и двойственным объектам. Если V — псевдогеометрическое

¹⁾ Автор имеет в виду ζ -функцию Дедекинда поля F .

представление, то для любого $r \in \mathbb{Z}$ представление $V(r)^{1)}$ тоже является псевдогеометрическим, и $L(V(r), s) = L(V, r + s)$.

3.4. Для всякого псевдогеометрического λ -адического представления V группы G_F положим $L_f(V) = \det_{E_\lambda} H^0(F, V) \otimes \det_{E_\lambda}^* H_f^1(F, V)$, $V^+ = \bigoplus_{p \in S_\infty(F)} H^0(G_p, V)$ и $t_V = \bigoplus_{p \in S_l(F)} t_{V,p}$. Введем *фундаментальную прямую* представления V

$$\Delta_f(V) = L_f(V) \otimes L_f(V^*(1)) \otimes \det_{E_\lambda}^* V^+ \otimes \det_{E_\lambda} t_V.$$

Цель следующего раздела — ввести на этой прямой некоторую каноническую норму.

4. НОРМА ЭЙЛЕРА–ПУАНКАРЕ ПСЕВДОГЕОМЕТРИЧЕСКОГО λ -АДИЧЕСКОГО ПРЕДСТАВЛЕНИЯ²⁾

4.1. Пусть $\mathcal{O}_\lambda$ — кольцо целых поля E_λ . Обозначим через $D_{\text{par}}(\mathcal{O}_\lambda)$ (соотв. $D_{\text{par}}(E_\lambda)$) производную категорию категории совершенных комплексов $\mathcal{O}_\lambda$ -модулей (соотв. E_λ -модулей).³⁾

Если $d \in \mathbb{N}$, $i \in \mathbb{Z}$ и M — свободный $\mathcal{O}_\lambda$ -модуль (E_λ -модуль) ранга d , то обозначим через $\det_\lambda^i M$ внешнюю d -степень модуля M , если i четно, и двойственный к ней модуль, если i нечетно. Если

$$C : \dots \rightarrow C^{i-1} \rightarrow C^i \rightarrow C^{i+1} \rightarrow \dots$$

— совершенный комплекс $\mathcal{O}_\lambda$ -модулей (E_λ -модулей), то положим $\det_\lambda(C) = \bigotimes_{i \in \mathbb{Z}} \det_\lambda^i C^i$ и назовем его *детерминантом* комплекса C . Любой квазиизоморфизм индуцирует изоморфизм детерминантов, что позволяет говорить о *детерминанте* объекта категории $D_{\text{par}}(\mathcal{O}_\lambda)$ (или $D_{\text{par}}(E_\lambda)$). В частности, можно говорить о детерминанте $\mathcal{O}_\lambda$ -модуля конечного типа (если $M = M_0/M_{-1}$, где M_0 и M_{-1} — свободные модули конечного типа, то $\det_\lambda M = \det_\lambda^0 M_0 \otimes \det_\lambda^{-1} M_{-1}$). Если

$$C' \rightarrow C \rightarrow C''$$

¹⁾ Как обычно, $Z_l(1) = \varprojlim_n \mu_n(\overline{\mathbb{Q}})$, и если $r \in \mathbb{Z}$, то $Z_l(r) = \text{Sym}_{Z_l}^r Z_l(1)$, и

$Z_l(-r)$ — двойственное представление; наконец, $V(r) = V \otimes_{Z_l} Z_l(r)$ для всякого $r \in \mathbb{Z}$.

²⁾ См. в [FP91, note I] или в [Ka91, § 3] более элементарное изложение в терминах когомологий Галуа. Я благодарен Делинню, познакоившему меня с точкой зрения, изложенной здесь, которая сразу показалась мне более простой и естественной.

³⁾ Совершенный комплекс — это комплекс конечной длины, состоящий из конечно порожденных проективных модулей.

— выделенный треугольник в $D_{\text{par}}(\mathcal{O}_\lambda)$ (или в $D_{\text{par}}(E_\lambda)$), то можно отождествить $\det_\lambda C$ с $\det_\lambda C' \otimes \det_\lambda C''$ ¹⁾. Наконец, расширение скаляров определяет функтор из $D_{\text{par}}(\mathcal{O}_\lambda)$ в $D_{\text{par}}(E_\lambda)$, и для любого C из $D_{\text{par}}(\mathcal{O}_\lambda)$ модуль $\det_\lambda(E_\lambda \otimes C)$ совпадает с $E_\lambda \otimes \det_\lambda C$, что позволяет отождествить $\det_\lambda C$ с решеткой в $\det_\lambda(E_\lambda \otimes C)$.

4.2. В дальнейшем S будет конечным множеством точек поля F , содержащим все бесконечные точки и все точки, делящие l . Обозначим через S_f множество конечных точек в S , через $\mathcal{O}_F$ — кольцо целых поля F и через U_S открытую подсхему схемы $\text{Spec } \mathcal{O}_F$, являющуюся дополнением к S_f . Выражение «пучок» (соотв. $\mathcal{O}_\lambda$ -пучок, E_λ -пучок) означает пучок (соотв. $\mathcal{O}_\lambda$ -пучок, E_λ -пучок), конструктивный в этальной топологии²⁾.

Для всякого $\mathcal{O}_\lambda$ -пучка (соотв. E_λ -пучка) T на U_S $R\Gamma_{\text{ét}}(U_S, T)$ является объектом категории³⁾ $D^+(\mathcal{O}_\lambda)$ ($D_{\text{par}}(E_\lambda)$)⁴⁾. Для всякой точки $p \in S_f$ это же верно и для $R\Gamma(F_p, T)$ (определенного как $R\Gamma_{\text{ét}}(\text{Spec } F_p, \nu_p^* j_* T)$, где $\nu_p: \text{Spec } F_p \rightarrow \text{Spec } \mathcal{O}_F$ и $j: U_S \rightarrow \text{Spec } \mathcal{O}_F$ — естественные морфизмы). По определению комплекс, вычисляющий относительные когомологии $R\Gamma_c(U_S, T)$ — это $C[-1]$, где C — конус морфизма $R\Gamma_{\text{ét}}(U_S, T) \rightarrow \bigoplus_{p \in S} R\Gamma(F_p, T)$. Имеем $H_c^i(U_S, T) = 0$ при $i \neq 1, 2, 3$, и $R\Gamma_c(U_S, T)$ является объектом категории $D_{\text{par}}(\mathcal{O}_\lambda)$ (соотв. $D_{\text{par}}(E_\lambda)$).

Теперь можно определить *прямую Эйлера–Пуанкаре* T ,

$$\Delta_{\lambda, EP}(T) = \det_\lambda R\Gamma_c(U_S, T).$$

¹⁾ Существуют проблемы со знаками при определении детерминанта (если $V = \bigoplus_{i \in I} V_i$ — разложение в конечную прямую сумму, то отождествление $\det_\lambda V$ с тензорным произведением модулей $\det_\lambda V_i$ зависит с точностью до знака от порядка V_i), которые не имеют для нас значения и которыми мы будем пренебрегать.

²⁾ См. в [SGA4, exp. IV; SGA5, app. á l'exp. I, V, VI; SGA4.5] общие сведения о пучках кручения, $\mathcal{O}_\lambda$ -пучках и E_λ -пучках, конструктивных в этальной топологии.

³⁾ И даже $D_{\text{par}}(\mathcal{O}_\lambda)$ при $l \neq 2$.

⁴⁾ Назовем $\mathcal{O}_\lambda$ -представлением проконечной группы G любой $\mathcal{O}_\lambda$ -модуль конечного типа с заданным на нем линейным и непрерывным действием группы G . Пусть G_S — группа Галуа максимального неразветвленного вне S расширения Галуа поля F , содержащегося в $\bar{F}$. Если T есть $\mathcal{O}_\lambda$ -пучок (соотв. E_λ -пучок), то его общий слой является $\mathcal{O}_\lambda$ -представлением (соотв. λ -адическим представлением) группы G_S . Обратно, всякое $\mathcal{O}_\lambda$ -представление (соотв. λ -адическое представление) T группы G_S является общим слоем единственного локально постоянного $\mathcal{O}_\lambda$ -пучка (соотв. E_λ -пучка), который мы снова обозначим через T , и $R\Gamma_{\text{ét}}(U_S, T)$ отождествляется с комплексом, вычисляющим непрерывные когомологии Галуа $H_{\text{cont}}^*(G_S, T)$.

Нетрудно видеть, что этот детерминант зависит только от общего слоя пучка T . Если T — пучок кручения, то $\prod_i \# H_c^i(U_S, T)^{(-1)^i} = 1$ и, следовательно, $\Delta_{\lambda, EP}(V) = \mathcal{O}_\lambda$. Отсюда вытекает, что если V есть E_λ -пучок на U_S и если выбран $\mathcal{O}_\lambda$ -пучок T , такой, что $E_\lambda \otimes T = V$, то получающаяся решетка $\Delta_{\lambda, EP}(\tilde{E})$ в $\Delta_{\lambda, EP}(V)$ не зависит от выбора T . Обозначим через $|\cdot|_{\lambda, EP}$ единственную норму на $\Delta_{\lambda, EP}(V)$, такую, что если δ — некоторый базис в $\Delta_{\lambda, EP}(T)$, то $|\delta|_{\lambda, EP} = 1$.

Нетрудно также видеть, что если S' — конечное множество точек поля F , содержащее S , и V' есть E_λ -пучок на $U_{S'}$ с тем же общим слоем, что и у V , то $\Delta_{\lambda, EP}(V')$ и $\Delta_{\lambda, EP}(V)$ изоморфны как нормированные прямые над E_λ . Это позволяет, в частности, определить нормированную E_λ -прямую $\Delta_{\lambda, EP}(V)$ для любого λ -адического представления G_F , неразветвленного в дополнении к конечному множеству точек: достаточно выбрать множество S так же, как и выше, и E_λ -пучок на U_S с общим слоем V ; полученный результат не зависит от выбора S и V .

4.3. Пусть теперь p — конечная точка поля F , делящая простое число p , и пусть V — некоторое λ -адическое представление группы G_p . «Разрежем» комплекс $R\Gamma(F_p, V)$ на два комплекса, чтобы получить выделенный треугольник

$$R\Gamma_f(F_p, V) \longrightarrow R\Gamma(F_p, V) \longrightarrow R\Gamma_{/f}(F_p, V),$$

характеризующийся тем, что $H^1(R\Gamma_f(F_p, V))$ есть E_λ -векторное подпространство $H_f^1(F_p, V)$ в $H^1(F_p, V)$, определенное в п. 3.2, $H^0(R\Gamma_f(F_p, V)) = H^0(F_p, V)$ и $H^i(R\Gamma_f(F_p, V)) = 0$ при $i \neq 0, 1$.

Пусть $\mathcal{O}_p$ — кольцо целых поля F_p , $j_p: \text{Spec } F_p \rightarrow \text{Spec } \mathcal{O}_p$ и $i_p: \text{Spec } k_p \rightarrow \text{Spec } \mathcal{O}_p$ — естественные морфизмы. Если $l \neq p$, то $R\Gamma_f(F_p, V) = R\Gamma(k_p, i_p^* j_{p*} V) = R\Gamma(k_p, \underline{D}_p(V))$. Этот комплекс можно отождествить с комплексом, сосредоточенным в степенях 0 и 1,

$$\underline{D}_p(V) \longrightarrow \underline{D}_p(V)$$

(это отображение переводит d в $(f_p - 1)(d)$). В частности, $\det_\lambda R\Gamma_f(F_p, V) = E_\lambda$.

Если же $l = p$ и если π обозначает проекцию пространства $B_{\text{cris}, p} \subset B_{dR, p}$ на $B_{dR, p} / \text{Fil}^0 B_{dR, p}$, то можно показать (см., например, [BK90, проп. 1.7]), что последовательность

$$0 \rightarrow \mathbb{Q}_p \rightarrow B_{\text{cris}, p} \rightarrow B_{\text{cris}, p} \oplus (B_{dR, p} / \text{Fil}^0 B_{dR, p}) \rightarrow 0 \quad (*)$$

(где $\mathbb{Q}_p \rightarrow B_{\text{cris},p}$ — естественное вложение, а следующее отображение переводит b в $((\phi - 1)b, \pi b)$) является точной. Если ее умножить тензорно на V и взять G_p -инварианты, то получим¹⁾ точную последовательность

$$0 \rightarrow H^0(F_p, V) \rightarrow \underline{D}_p(V) \rightarrow \underline{D}_p(V) \oplus t_{V,p} \rightarrow H_f^1(F_p, V) \rightarrow 0,$$

которая позволяет отождествить $R\Gamma_f(F_p, V)$ с комплексом, сосредоточенным в степенях 0 и 1,

$$\underline{D}_p(V) \rightarrow \underline{D}_p(V) \oplus t_{V,p}.$$

В частности, $\det_\lambda R\Gamma_f(F_p, V)$ совпадает с $\det_\lambda^1 t_{V,p}$.

4.4. Рассмотрим снова E_λ -пучок V на U_S . Определим комплекс $R\Gamma_f(U_S, V)$, вычисляющий f -когомологии пучка V как $C[-1]$, где C — конус композиции морфизмов

$$R\Gamma_{\text{ét}}(U_S, V) \rightarrow \bigoplus_{p \in S_f} R\Gamma(F_p, V) \rightarrow \bigoplus_{p \in S_f} R\Gamma_{/f}(F_p, V).$$

Из двух выделенных треугольников

$$R\Gamma_c(U_S, V) \rightarrow R\Gamma_{\text{ét}}(U_S, V) \rightarrow \bigoplus_{p \in S} R\Gamma(F_p, V)$$

и

$$R\Gamma_f(U_S, V) \rightarrow R\Gamma_{\text{ét}}(U_S, V) \rightarrow \bigoplus_{p \in S_f} R\Gamma_{/f}(F_p, V)$$

можно построить третий²⁾

$$\begin{aligned} R\Gamma_c(U_S, V) &\rightarrow R\Gamma_f(U_S, V) \\ &\rightarrow \left(\bigoplus_{p \in S_f} R\Gamma_f(F_p, V) \right) \oplus \left(\bigoplus_{p \in S_\infty(F)} R\Gamma(F_p, V) \right). \end{aligned}$$

1) Поскольку $B_{\text{cris},p}$ и $B_{dR,p}$ имеют бесконечную размерность над $\mathbb{Q}_p$, то нужно или показать, что последовательность (*) расщепляется как последовательность топологических векторных пространств над $\mathbb{Q}_p$, или заметить, что можно заменить $B_{\text{cris},p}$ и $B_{dR,p}$ на объединение их $\mathbb{Q}_p$ -векторных подпространств конечной размерности, инвариантных относительно действия G_p .

2) Для этого нужно использовать аксиому октаэдра из определения триангулированных категорий.

Из него мы получаем изоморфизм

$$\det_{\lambda} R\Gamma_f(U_S, V) = \Delta_{\lambda, EP}(V) \otimes \det_{\lambda}^1 t_V \otimes \det_{\lambda} V^+. \quad (a)$$

С другой стороны, вычислить $H_f^i(U_S, V) := H^i(R\Gamma_f(U_S, V))$ легко: при $i \neq 0, 1, 2, 3$ они равны 0. Имеем $H_f^0(U_S, V) = H^0(F, V)$ и $H_f^1(U_S, V) = H_f^1(F, V)$. Также ясно, что $H_f^3(U_S, V)$ можно отождествить с двойственным пространством к $H^0(F, V^*(1))$; для этого сначала заметим, что оба этих пространства естественным образом являются факторами пространства $\bigoplus_{p \in S_f} H^2(F_p, V)$ (первое по определению, второе в силу локальной двойственности (см., например, [Mi86, Чар. 1, § 2]), отождествляющей каждое $H^2(F_p, V)$ с двойственным к $H^0(F_p, V^*(1))$ и позволяющей использовать двойственное отображение к естественной инъекции $H^0(F, V^*(1)) \rightarrow \bigoplus_{p \in S_f} H^0(F_p, V^*(1))$, и можно проверить, что эти два фактора совпадают.

Наконец, если $H_0^2(U_S, V)$ обозначает ядро естественного отображения

$$H^2(U_S, V) \rightarrow \bigoplus_{p \in S_f} H^2(F_p, V),$$

то $H_f^2(U_S, V)$ есть прямая сумма $H_0^2(U_S, V)$ и коядра $\tilde{H}_f^2(U_S, V)$ естественного отображения

$$H^1(U_S, V) \oplus \left(\bigoplus_{p \in S_f} H_f^1(F_p, V) \right) \rightarrow \bigoplus_{p \in S_f} H^1(F_p, V).$$

Локальная двойственность [loc.cit] отождествляет каждое пространство $H^1(F_p, V)$ с двойственным к $H^1(F_p, V^*(1))$, откуда получаем отображение $\bigoplus_{p \in S_f} H^1(F_p, V) \rightarrow H_f^1(U_S, V^*(1))^*$, а теоремы глобальной двойственности Пуату—Тейта (см. например, [Mi86, чар. II, § 1]) позволяют отождествить его коядро с двойственным к $H_0^2(U_S, V)$ пространством. Если V является псевдогеометрическим, то можно проверить, что последовательность

$$\begin{aligned} H^1(U_S, V) \oplus \left(\bigoplus_{p \in S_f} H_f^1(F_p, V) \right) &\rightarrow \bigoplus_{p \in S_f} H^1(F_p, V) \rightarrow \\ &\rightarrow H_f^1(U_S, V^*(1))^* \rightarrow H_0^2(U_S, V) \rightarrow 0 \end{aligned}$$

точна. Следовательно, имеем короткую точную последовательность

$$0 \rightarrow \tilde{H}_f^2(U_S, V) \rightarrow H_f^1(U_S, V^*(1))^* \rightarrow H_0^2(U_S, V) \rightarrow 0.$$

Так как $H_f^2(U_S, V) = \tilde{H}_f^2(U_S, V) \oplus H_0^2(U_S, V)$, то $\det_\lambda^1 H_f^1(U_S, V^*(1))$ отождествляется с $\det_\lambda H_f^2(U_S, V)$. Собирая все вместе, получаем канонический изоморфизм

$$\det_\lambda R\Gamma_f(U_S, V) = L_f(V) \otimes L_f(V^*(1));$$

вместе с (а) это дает канонический изоморфизм между $\Delta_f(V)$ и $\Delta_{\lambda, EP}(V)$. Мы используем этот изоморфизм для того, чтобы перенести на $\Delta_f(V)$ норму $|\cdot|_{\lambda, EP}$, определенную на $\Delta_{\lambda, EP}$.

5. СМЕШАННЫЕ СТРУКТУРЫ ХОДЖА

5.1. В случае $l = \infty$ роль l -адических представлений играют смешанные структуры Ходжа.

Если V есть $\mathbf{R}$ -векторное пространство, то положим $V_{\mathbf{C}} = \mathbf{C} \otimes_{\mathbf{R}} V$ и обозначим через $\tau : V_{\mathbf{C}} \rightarrow V_{\mathbf{C}}$ отображение, переводящее $z \otimes v$ в $\bar{z} \otimes v$. Напомним (см. [De71]), что *смешанная $\mathbf{R}$ -структура Ходжа над $\mathbf{C}$* — это конечномерное векторное пространство V над $\mathbf{R}$ с заданными на нем возрастающей фильтрацией (*весовой фильтрацией*) $(W_m V)_{m \in \mathbf{Z}}$ $\mathbf{R}$ -линейными подпространствами, такой, что $W_i V = 0$ при $i \ll 0$ и $W_i V = V$ при $i \gg 0$, и убывающей фильтрацией (*фильтрацией Ходжа*) $(Fil^i V_{\mathbf{C}})_{i \in \mathbf{Z}}$ пространства $V_{\mathbf{C}}$ $\mathbf{C}$ -линейными подпространствами, причем если $\overline{Fil}^i$ обозначает фильтрацию, сопряженную с Fil^i (т. е. фильтрацию в $V_{\mathbf{C}}$, для которой $\overline{Fil}^i V_{\mathbf{C}} = \tau(Fil^i V_{\mathbf{C}})$), то фильтрации W , Fil и $\overline{Fil}$ являются дополнительными (т. е. если для любого $m \in \mathbf{Z}$ фильтрации, индуцированные на $gr_m^W V_{\mathbf{C}}$, снова обозначить через Fil и $\overline{Fil}$, то естественное отображение

$$Fil^r gr_m^W V_{\mathbf{C}} \oplus \overline{Fil}^{m+1-r} gr_m^W V_{\mathbf{C}} \rightarrow gr_m^W V_{\mathbf{C}}$$

будет изоморфизмом для любого $r \in \mathbf{Z}$).

Напомним, что целые m , такие, что $gr_m^W V_{\mathbf{C}} \neq 0$, называются *весами* структуры V и что V называется *чистой*, если у нее есть один и только один вес.

Смешанные $\mathbf{R}$ -структуры Ходжа над $\mathbf{C}$ образуют очевидным образом $\mathbf{R}$ -линейную абелеву категорию $\underline{SH}_{\mathbf{C}}(\mathbf{R})$. Обозначим через $\underline{SH}_{\mathbf{C}}(\mathbf{C})$ категорию *смешанных $\mathbf{C}$ -структур Ходжа над $\mathbf{C}$* , т. е.

$\mathcal{C}$ -линейную категорию, получающуюся из $\underline{SH}_{\mathcal{C}}(\mathbf{R})$ расширением скаляров. Объект V этой категории является, таким образом, смешанной $\mathbf{R}$ -структурой Ходжа над $\mathcal{C}$, снабженной $\mathbf{R}$ -вложением поля $\mathcal{C}$ в ее кольцо эндоморфизмов (обратите внимание, что хотя на соответствующем V вещественном векторном пространстве задана структура векторного пространства над $\mathcal{C}$, фильтрация Ходжа определена только на $V_{\mathcal{C}} = \mathcal{C} \otimes_{\mathbf{R}} V$).

5.2. *Смешанная структура Ходжа над $\mathbf{R}$* может быть определена как смешанная $\mathbf{R}$ -структура Ходжа над $\mathcal{C}$, на которой задано действие $\mathrm{Gal}(\mathcal{C}/\mathbf{R})$. Это равносильно заданию разложения в прямую сумму

$$V = V^+ \oplus V^-,$$

совместимого с весовой фильтрацией и такого, что фильтрация Ходжа происходит из некоторой фильтрации, определенной на $V_{dR} = (V_{\mathcal{C}})^{\mathrm{Gal}(\mathcal{C}/\mathbf{R})} = V^+ \oplus iV^-$. Получаем абелеву категорию $\underline{SH}_{\mathbf{R}}(\mathbf{R})$ и обозначаем через $\underline{SH}_{\mathbf{R}}(\mathcal{C})$ категорию смешанных $\mathcal{C}$ -структур Ходжа над $\mathbf{R}$, т.е. категорию, получающуюся из $\underline{SH}_{\mathbf{R}}(\mathbf{R})$ с помощью расширения скаляров $\mathbf{R} \rightarrow \mathcal{C}$.

5.3. Далее на протяжении этого раздела $K, L \in \{\mathbf{R}, \mathcal{C}\}$, $\bar{K} = \mathcal{C}$, $G_K = \mathrm{Gal}(\mathcal{C}/K)$ и V — смешанная L -структура Ходжа над K^1 .

Если $K = \mathcal{C}$, то положим $V^+ = V$ и $V_{dR} = V_{\mathcal{C}}$. Таким образом, в любом случае имеем $V^+ = V^{G_K}$ и $V_{dR} = (V_{\mathcal{C}})^{G_K}$.

Обозначим через $1_{K,L}$ единичный объект категории $\underline{SH}_K(L)$, т.е. единственный чистый объект веса 0, такой, что $1_{K,L} = 1_{K,L}^+ = L$. Мы предлагаем читателю дать определения тензорного произведения двух объектов и двойственного объекта²). Объект Тейта $1_{K,L}(1)$ — это единственный чистый объект веса -2 , такой, что $1_{K,L}(1) = L \cdot (2\pi i)$, причем $1_{K,L}(1)^- = 1_{K,L}(1)$, если $K = \mathbf{R}$. Мы будем обозначать через $V^*(1)$ тензорное произведение двойственного объекта к V на $1_{K,L}(1)$.

Положим $t_V = V_{dR}/\mathrm{Fil}^0 V_{dR}$. Вложение V в $V_{\mathcal{C}}$ индуцирует после взятия инвариантов относительно G_K и перехода к фактору L -линейное отображение

$$\alpha_V : V^+ \longrightarrow t_V.$$

¹) Мы будем применять этот формализм в случае $K = F_p$ и $L = E_\lambda$, где $p \in S_\infty(F)$ и $\lambda \in S_\infty(E)$ (напомним, что $\bar{F}_p$ отождествлено с $\mathcal{C}$).

²) Категория $\underline{SH}_K(L)$ является нейтральной категорией Таннаки [DM82, def. 2.19] над L .

5.4. Для всякого $i \in \mathbb{N}$ положим $H^i(K, V) = \text{Ext}_{\underline{SH}_K(L)}^i(1_{K,L}, V)$. Это L -векторные пространства конечной размерности, равные нулю при $i \neq 0, 1$. В дальнейшем ядро и коядро отображения α_V будут играть большую роль, хотя казалось бы более естественным рассматривать $H^i(K, V)^1$. Эти пространства тесно связаны друг с другом: $H^i(K, V) = H^i(K, W_0 V)$, и если $W_0 V = V$, то $H^0(K, V) = \text{Ker } \alpha_V$ и $H^1(K, V) = \text{Coker } \alpha_V$; отсюда получаем точную последовательность L -векторных пространств

$$0 \rightarrow H^0(K, V) \rightarrow \text{Ker } \alpha_V \rightarrow \text{Ker } \alpha_{V/W_0 V} \rightarrow H^1(K, V) \rightarrow \text{Coker } \alpha_V \rightarrow 0,$$

и можно показать, что $\text{Ker } \alpha_{V/W_0 V}$ отождествляется с двойственным пространством к $H^1(K, (V/W_0 V)^*(1))$.

Легко также проверить, что существует совершенное спаривание

$$\text{Coker } \alpha_V \times \text{Ker } \alpha_{V^*(1)} \rightarrow L.$$

6. МОТИВЫ НАД F И ИХ РЕАЛИЗАЦИИ

Обозначим через $A_f(Q)$ (соотв. $A_f(E)$) кольцо конечных аделей поля Q (соотв. поля E). Положим $Q_\infty = \mathbb{R}$ и $B_{dR,p} = \overline{F_p} = \mathbb{C}$ для любого $p \in S_\infty(F)$.

6.1. Поскольку точное определение категории $\mathcal{MM}_F(E)$ неизвестно, обычно вводят некоторую вполне корректно определенную «достаточно вместительную» категорию, в которой должна содержаться $\mathcal{MM}_F(E)$ и внутри которой пытаются работать (см., например, [De90, § 1; Ja90, § 2; BK90, def. 5.5; FP91, note II, § 1; ...]). Та категория, которая используется в настоящее время, обозначается через $\mathcal{C}_F(E)$ и будет определена в следующем пункте. Сначала мы введем категорию $\mathcal{C}_{F,0}$:

• Объект M категории $\mathcal{C}_{F,0}$ определяется следующим набором данных:

¹⁾ Если «забыть» о весовой фильтрации, то можно рассматривать V как объект некоторой аддитивной неабелевой L -линейной категории; тогда $\text{Ker } \alpha_V$ отождествляется с L -векторным пространством морфизмов из $1_{K,L}$ в V в этой категории, тогда как $\text{Coker } \alpha_V$ классифицирует «расширения» $1_{K,L}$ с помощью V . В частности, любая короткая точная последовательность

$$0 \rightarrow V' \rightarrow V \rightarrow V'' \rightarrow 0$$

в категории $\underline{SH}_K(L)$ порождает длинную точную последовательность

$$0 \rightarrow \text{Ker } \alpha_{V'} \rightarrow \text{Ker } \alpha_V \rightarrow \text{Ker } \alpha_{V''} \rightarrow \text{Coker } \alpha_{V'} \rightarrow \text{Coker } \alpha_V \rightarrow \text{Coker } \alpha_{V''} \rightarrow 0.$$

- (i) F -векторное пространство конечной размерности M_{dR} , снабженное убывающей фильтрацией $(\text{Fil}^i M_{dR})_{i \in \mathbb{Z}}$ F -векторными подпространствами, причем $\text{Fil}^i M_{dR} = M_{dR}$ при $i \ll 0$ и $\text{Fil}^i M_{dR} = 0$ при $i \gg 0$;
- (ii) для любого $p \in S_\infty(F)$ $\mathbb{Q}$ -векторное пространство $M_{B,p}$, снабженное действием G_p (положим $M_{B,p}^+ = (M_{B,p})^{G_p}$);
- (iii) свободный $A_f(\mathbb{Q})$ -модуль конечного ранга M_f , на котором линейно и непрерывно действует G_F ; если для произвольного простого l обозначить через M_l его l -компоненту, то требуется, чтобы она была неразветвленной в дополнении к конечному множеству точек;
- (iv) для любого $p \in S_\infty(F)$ изоморфизм $A_f(\mathbb{Q})[\text{Gal}(\overline{F_p}/F_p)]$ -модулей

$$i_{M,p}^B : A_f(\mathbb{Q}) \otimes_E M_{B,p} \rightarrow M_f;$$

- (v) для каждой точки $p \in S_p(F)$ (p конечно или нет)

$$i_{M,p} : B_{dR,p} \otimes_{\mathbb{Q}_p} M_p \longrightarrow B_{dR,p} \otimes_F M_{dR}^{1)}$$

(где через M_p обозначен $\mathbb{Q}_p[G_p]$ -модуль, соответствующий M_p при конечном p и для $\mathbb{R} \otimes_{\mathbb{Q}} M_{B,p}$ при бесконечном p), изоморфизм $B_{dR,p}$ -векторных пространств, совместимый с действием группы G_p и такой, что для любого $r \in \mathbb{Z}$

$$i_{M,p}(\text{Fil}^r B_{dR,p} \otimes_{\mathbb{Q}_p} M_p) = \sum_{s+t=r} (\text{Fil}^s B_{dR,p} \otimes_F \text{Fil}^t M_{dR}),$$

если $p \neq \infty$.

• Если $M = (M_{dR}, (M_{B,p})_{p \in S_\infty(F)}, M_f)$, $N = (N_{dR}, (N_{B,p})_{p \in S_\infty(F)}, N_f)$ — два объекта в $\mathcal{C}_{F,0}$, то морфизм $\eta : M \rightarrow N$ — это набор отображений $\eta = (\eta_{dR}, (\eta_{B,p})_{p \in S_\infty(F)}, \eta_f)$, удовлетворяющих всем естественным соотношениям совместимости.

6.2. Категория $\mathcal{C}_{F,0}$ является абелевой $\mathbb{Q}$ -линейной категорией. Обозначим через $\mathcal{C}_F(E)_0$ категорию, получающуюся из $\mathcal{C}_{F,0}$ расширением скаляров $\mathbb{Q} \rightarrow E$; таким образом, объект категории

¹⁾ Предположим, что $F = \mathbb{Q}$, и пусть $\overline{\mathbb{Z}}$ — кольцо целых в $\overline{\mathbb{Q}}$. Изменяя соответствующим образом обычную конструкцию $B_{dR,p}$ (фактически заменив в ней $\overline{\mathbb{Z}}_p/p\overline{\mathbb{Z}}_p$ на $[\overline{\mathbb{Z}}/p\overline{\mathbb{Z}}]$), можно определить $A_f(\mathbb{Q})$ -алгебру $B_{dR,f}$, снабженную $\mathbb{Z}$ -фильтрацией и действием группы G_F , и переформулировать (v) как требование задания изоморфизма

$$B_{dR} \otimes_{\mathbb{Q}} M_B = B_{dR} \otimes_{\mathbb{Q}} M_{dR}$$

(где $B_{dR} = \mathbb{C} \times B_{dR,f}$), удовлетворяющего надлежащим условиям.

$C_F(E)_0$ — это объект M категории $C_{F,0}$, снабженный вложением поля E в $\mathbf{Q}$ -алгебру $C_{F,0}$ -эндоморфизмов объекта M , тогда как морфизм в $C_F(E)_0$ — это морфизм соответствующих объектов категории $C_{F,0}$, коммутирующий с действием E .

Если M — объект в $C_F(E)_0$, то M_{dR} является свободным $E \otimes_{\mathbf{Q}} F$ -модулем конечного ранга, который мы обозначим через $b(M)$, каждое $M_{B,p}$ является E -векторным пространством размерности $b(M)$, а M_f — свободным $A_f(E)$ -модулем ранга $b(M)$. Для каждой конечной точки λ поля E E_λ -компонента M_λ модуля M_f является псевдогеометрическим λ -адическим представлением группы G_F размерности $b(M)$. Назовем касательным пространством к M модуль $t_M = M_{dR}/\text{Fil}^0 M_{dR}$ над $E \otimes F$. Пусть p — точка поля F и λ — точка поля E , лежащие над одним и тем же $p \in S(\mathbf{Q})$; тогда

- если p простое, то $\underline{D}_{dR,p}(M_\lambda)$ отождествляется с $(E_\lambda \otimes_{\mathbf{Q}_p} F_p) \otimes_{E \otimes F} M_{dR}$ и $t_{M_\lambda,p} = (E_\lambda \otimes_{\mathbf{Q}_p} F_p) \otimes_{E \otimes F} t_M$;
- если $p = \infty$, то положим $M_{p,\lambda} = E_\lambda \otimes_{\mathbf{R}} M_p$; $\mathbf{C}$ -линейное отображение $i_{M,p}$ отождествляет $\mathbf{C} \otimes_{\mathbf{R}} M_{p,\lambda}$ с $\mathbf{C} \otimes_{E \otimes F} M_{dR}$; в частности, можно говорить о фильтрации Ходжа и сопряженной к ней фильтрации на $\mathbf{C} \otimes_{\mathbf{R}} M_{p,\lambda}$.

Если S — конечное множество точек поля F , содержащее $S_\infty(F)$, то будем говорить, что M допускает L_S -функцию, если все G_F -модули Галуа M_λ допускают L_S -функцию и эти функции совпадают для всех конечных точек λ поля E (обозначим тогда эту общую функцию через $L_S(M, s)$). В случае $S = S_\infty(F)$ будем говорить, что M допускает L -функцию (и обозначать ее через $L(M, s)$).

6.3. Категория $C_F(E)$ определяется следующим образом:

- ее объект — это объект M категории $C_F(E)_0$, допускающий L -функцию и снабженный возрастающей фильтрацией (весовой фильтрацией) подобъектами $(W_i M)_{i \in \mathbf{Z}}$ (такими, что $W_i M = 0$ при $i \ll 0$ и $= M$ при $i \gg 0$), причем если λ (соотв. p) — архимедова точка поля E (соотв. F), то $M_{p,\lambda}$ является смешанной E_λ -структурой Ходжа над F_p (это эквивалентно тому, что на $\mathbf{C} \otimes_{\mathbf{R}} M_{p,\lambda}$ весовая фильтрация, фильтрация Ходжа и сопряженная к ней фильтрация являются дополнительными друг к другу).
- морфизм в $C_F(E)$ — это морфизм соответствующих объектов из $C_F(E)_0$, совместимый с их весовыми фильтрациями.

Нетрудно проверить, что $C_F(E)$ также является абелевой E -линейной категорией. Мы предоставляем читателю возможность

дать определение объектов $M \otimes N$ и $\text{Hom}(M, N)$ для двух объектов M и N категории $\mathcal{C}_F(E)$. Имеется также единичный объект $1_{F,E}$ (это чистый объект веса 0, причем $1_{F,E,dR} = E$, $1_{F,E,B,p} = E \otimes F$ для любого $p \in S_\infty(F)$ и $1_{F,E,f} = A_f(E)$ с тривиальным действием группы G_F); двойственный объект M^* к M есть тогда $\text{Hom}(M, 1_{F,E})$ (и $\text{Hom}(M, N) = M^* \otimes N$). Мы получаем, таким образом, нейтральную категорию Таннаки над E [DM82, § 2]. Разумеется, категория $\mathcal{C}_F(E)$ — это не что иное, как категория Таннаки $\mathcal{C}_F(\mathbb{Q})_{(E)}$, получающаяся из $\mathcal{C}_F(\mathbb{Q})$ расширением скаляров $\mathbb{Q} \subset E$ [DM82, § 3].

6.4. Если F' — конечное расширение поля F , содержащееся в $\bar{F}$, то имеются следующие функторы: функтор расширения скаляров $\text{Ext}_{F/F'} : \mathcal{C}_F(E) \rightarrow \mathcal{C}_{F'}(E)$ и сопряженный к нему функтор ограничения скаляров по Вейлю $\text{Res}_{F'/F} : \mathcal{C}_{F'}(E) \rightarrow \mathcal{C}_F(E)$ (заметим, что если λ — конечная точка, то $\text{Ext}_{F/F'}(M)_\lambda$ есть не что иное, как ограничение представления M_λ на подгруппу $G_{F'} \subset G_F$, тогда как $\text{Res}_{F'/F}(N)_\lambda$ есть представление группы G_F , индуцированное представлением N_λ группы $G_{F'}$). Имеем $L(\text{Res}_{F'/F}(N), s) = L(N, s)$.

Аналогично, если E' — некоторое конечное расширение поля E , то определен функтор расширения поля скаляров с E до E' , переводящий $\mathcal{C}_F(E)$ в $\mathcal{C}_F(E')$, для которого существует сопряженный функтор ограничения поля скаляров с E' до E .

Обозначим через $1(1)$ объект Тейта категории $\mathcal{C}_{\mathbb{Q}}(\mathbb{Q})$ (он является чистым веса -2 ; имеем $1(1)_{dR} = \mathbb{Q}$, $1(1)_B = 2\pi i \cdot \mathbb{Q}$; для любого $n \in \mathbb{N}$ конечная группа $2\pi i \mathbb{Z} / 2\pi i n \mathbb{Z}$ отождествляется с $\mu_n(\bar{\mathbb{Q}})$ с помощью отображения, переводящего $2\pi i$ в $e^{2\pi i/n}$, и это дает действие группы $G_{\mathbb{Q}}$ на

$$1(1)_f = \mathbb{Q} \otimes_{\bar{\mathbb{Z}}} \varprojlim_n (2\pi i \mathbb{Z} / 2\pi i n \mathbb{Z})$$

при очевидном выборе изоморфизмов сравнения. В дальнейшем положим $t = 2\pi i$ для бесконечных p ; тем самым элемент t определен в $B_{dR,p}$ для всех (конечных или бесконечных) точек p поля $\mathbb{Q}$.

Объект Тейта $1_{F,E}(1)$ — это объект категории $\mathcal{C}_F(E)$, получающийся из $1(1)$ при двойном расширении скаляров.

Для любого $r \in \mathbb{N}$ обозначим через $1_{F,E}(r)$ r -ю тензорную степень объекта $1_{F,E}(1)$, а через $1_{F,E}(-r)$ двойственный к ней объект. Для любого объекта M категории $\mathcal{C}_F(E)$ и любого $i \in \mathbb{Z}$ положим $M(i) = M \otimes 1_{F,E}(i)$. Скрученный по Тейту двойственный объект $M^*(1)$ будет играть в дальнейшем важную роль.

6.5. Пусть $\mathcal{M}$ — полная подкатегория категории $\mathcal{C}_F(E)$, устойчивая относительно взятия подобъекта, факторобъекта, прямой суммы, а также перехода $M \mapsto M^*(1)$ и содержащая $1_{F,E}$. Для любого объекта M из $\mathcal{M}$ положим $H^0(F, M) = \text{Hom}_{\mathcal{C}_F(E)}(1_{F,E}, M) (= \text{Hom}_{\mathcal{M}}(1_{F,E}, M))$ и обозначим через $H^1_{\mathcal{M}}(F, M)$ E -векторное пространство классов изоморфизма расширений $1_{F,E}$ с помощью M в категории $\mathcal{M}$. Для любой конечной точки λ поля E и любого $x \in H^1_{\mathcal{M}}(F, M)$ обозначим через x_λ образ x в $H^1(F, M_\lambda)$. Положим

$$H^1_{\mathcal{M},f}(F, M) = \{x \in H^1(F, M) \mid x_\lambda \in H^1_f(F, M_\lambda) \text{ для любой точки } \lambda\}$$

и также $H^0_{\mathcal{M},f}(F, M) = H^0(F, M)$.

Скажем, что категория $\mathcal{M}$ является f -допустимой, если для любого объекта M из $\mathcal{M}$, любой точки λ поля E и любого $i \in \{0, 1\}$ она обладает следующим свойством $C^i_\lambda(M) = C^i_\lambda(M, \mathcal{M})$:

а) если λ конечна, то естественное отображение

$$E_\lambda \otimes_E H^i_{\mathcal{M},f}(F, M) \rightarrow H^i_f(F, M_\lambda)$$

есть изоморфизм;

б) если $\lambda \in S_\infty(E)$ и $H^0_{\mathcal{M},f}(F, M^*(1)) = H^1_{\mathcal{M},f}(F, M^*(1)) = 0$, то естественное отображение

$$E_\lambda \otimes_E H^i_{\mathcal{M},f}(F, M) \rightarrow \bigoplus_{p \in S_\infty(F)} H^i(F_p, M_\lambda)$$

является изоморфизмом.

6.6. Назовем подкатегорией Таннаки в $\mathcal{C}_F(E)$ любую полную подкатегорию, устойчивую относительно взятия подобъекта, факторобъекта, двойственного объекта, прямой суммы и тензорного произведения. Естественно считать, что каждый мотив над F с коэффициентами в E определяет некоторый объект в $\mathcal{C}_F(E)$, его реализацию. В таком случае можно ввести категорию $\underline{SM}_F(E)$ мотивных структур над F с коэффициентами в E как наименьшую подкатегорию Таннаки в $\mathcal{C}_F(E)$, содержащую любой объект, изоморфный реализации некоторого мотива.

Разумеется, когда мы будем располагать подходящим определением категории $\mathcal{MM}_F(E)$, «стандартные гипотезы для смешанных мотивов» превратятся в утверждение о том, что функтор «реализации» индуцирует эквивалентность между категориями $\mathcal{MM}_F(E)$ и $\underline{SM}_F(E)$.

«Гипотеза». Категория $\underline{SM}_F(E)$ является f -допустимой.

6.7. Замечание. Известно (см. [De71, De74, Ja90]), что с каждым гладким квазипроективным многообразием U над F можно связать его реализацию $H(U)$, являющуюся объектом категории $\mathcal{C}_F(\mathbf{Q})$ (например, его реализация де Рама — это гиперкогомологии комплекса де Рама на U), со следующими двумя оговорками:

- нужно определить изоморфизмы сравнения $i_{M,p}$ для всех конечных p : это уже сделал Фалтингс ([Fa89], см. также [Pi90] в случае проективного U ; общий случай должен следовать из работы Фалтингса [Fa90]);
- можно показать, что для каждого U существует конечное множество точек S поля F , содержащее $S_\infty(F)$, такое, что M допускает функцию L_S , но не всегда известно, что можно взять $S = S_\infty(F)$.

Пусть $\underline{SM}_{F,J}(\mathbf{Q})$ — наименьшая подкатегория Таннаки в $\mathcal{C}_F(\mathbf{Q})$, содержащая любой объект, изоморфный $H(U)$ для некоторого гладкого квазипроективного многообразия U над F . Предполагается, что $\underline{SM}_F(\mathbf{Q})$ содержит $\underline{SM}_{F,J}(\mathbf{Q})$. Следуя Янсену [Ja90, § 4], можно спросить, верно ли, что $\underline{SM}_F(\mathbf{Q}) = \underline{SM}_{F,J}(\mathbf{Q})$, т. е. является ли категория $\underline{SM}_{F,J}(\mathbf{Q})$ «достаточно большой», что можно также переформулировать (фиксировав простое число l) как утверждение о том, что для любого объекта M категории $\underline{SM}_{F,J}(\mathbf{Q})$ естественное отображение

$$Q_l \otimes_{\mathbf{Q}} H_{\underline{SM}_{F,J}(\mathbf{Q}),f}^i(F, M) \rightarrow H_f^i(F, M_l)$$

сюръективно. К сожалению, в настоящее время эта задача представляется совершенно неприступной.

6.8. В дальнейшем будем предполагать, что задана некоторая полная подкатегория $\mathcal{M}$ в $\underline{SM}_F(E)$, являющаяся f -допустимой (если «гипотеза», приведенная выше, верна, то это условие эквивалентно требованию, чтобы $\mathcal{M}$ содержала $1_{F,E}$ и являлась устойчивой относительно взятия подобъекта, факторобъекта и прямой суммы, а также удовлетворяла условию, что если она содержит объект M , то содержит и $M^*(1)$, и любое расширение $1_{F,E}$ с помощью объекта M , класс которого лежит в $H_{\underline{SM}_F(E),f}^1(F, M)$). Положим $H_f^i(F, M) = H_{\mathcal{M},f}^i(F, M)$ для $i = 0, 1$.

Будем называть объекты категории $\mathcal{M}$ *мотивными структурами*.

6.9. Для любого E -векторного пространства X конечной размерности и любой точки λ поля E положим $X_\lambda = E_\lambda \otimes_E X$ и обозначим через X_λ^* его E_λ -двойственное векторное пространство.

Пусть M — мотивная структура и $\lambda \in S_\infty(E)$. Если $p \in S_\infty(F)$, то $M_{p,\lambda}$ является смешанной E_λ -структурой Ходжа над F_p . Можно, следовательно, рассмотреть (см. п. 5.3) отображение

$$\alpha_{M_\lambda} = \bigoplus_{p \in S_\infty(F)} \alpha_{M_{p,\lambda}} : \left(\bigoplus_{p \in S_\infty(F)} M_{B,p}^+ \right)_\lambda \longrightarrow t_{M,\lambda}$$

(где $M_{B,p}^+ = (M_{B,p}^{G_p})$), и (loc.cit) E_λ -векторные пространства $\text{Ker } \alpha_{M_\lambda}$ и $\text{Coker } \alpha_{M^*(1)_\lambda}$ являются двойственными друг другу, так же как и $\text{Coker } \alpha_{M_\lambda}$ и $\text{Ker } \alpha_{M^*(1)_\lambda}$. Обозначим через u_{M_λ} композицию естественных отображений

$$H^0(F, M)_\lambda \longrightarrow \bigoplus_p H^0(F_p, M_\lambda) \longrightarrow \bigoplus_p \text{Ker } \alpha_{M_{p,\lambda}} = \text{Ker } \alpha_{M_\lambda},$$

через v_{M_λ} композицию

$$H_f^1(F, M_\lambda) \longrightarrow \bigoplus_p H^1(F_p, M_\lambda) \longrightarrow \bigoplus_p \text{Coker } \alpha_{M_{p,\lambda}} = \text{Coker } \alpha_{M_\lambda},$$

через $v'_{M_\lambda} : \text{Ker } \alpha_{M_\lambda} \rightarrow H_f^1(F, M^*(1))_\lambda^*$ сопряженное отображение к $u_{M^*(1)_\lambda}$ и через $u'_{M_\lambda} : \text{Coker } \alpha_{M_\lambda} \rightarrow H^0(F, M^*(1))_\lambda^*$ сопряженное отображение к $u_{M^*(1)_\lambda}$.

Пусть M — мотивная структура и N — некоторое расширение объекта $1_{F,E}$ с помощью M в категории $\mathcal{M}$, класс которого лежит в $H_f^1(F, M)$. Если $x \in H_f^1(F, M^*(1))_\lambda^*$, то обозначим через x_N его образ в $H_f^1(F, N^*(1))_\lambda^*$, и пусть

$$\Lambda(M, N, \lambda) = \{x \in H_f^1(F, M^*(1))_\lambda^* \mid x_N \in \text{Im } v'_{N_\lambda}\}.$$

Если $x \in \Lambda(M, N, \lambda)$ и $x_N = v'_{N_\lambda}(y)$, то образ y в $\text{Ker } \alpha_{1_{F,E}}$ совпадает с образом некоторого $z \in H^0(F, 1_{F,E})_\lambda$ и образ $\delta_{M,N,\lambda}(x)$ элемента z в $H_f^1(F, M)_\lambda$ не зависит от произвола в выборе y и z .

Нетрудно видеть, что из f -допустимости вытекает существование единственного естественного преобразования функторов $\delta_\lambda = (\delta_{M,\lambda})_{M \in \text{Ob } \mathcal{M}}$, где $\delta_{M,\lambda} : H_f^1(F, M^*(1))_\lambda^* \rightarrow H_f^1(F, M)_\lambda$, такого, что

если M и N означают то же, что и выше, и если $x \in \Lambda(M, N, \lambda)$, то $\delta_{M, \lambda}(x) = \delta_{M, N, \lambda}(x)$.

6.10. Из условия f -допустимости можно также вывести, что для любой мотивной структуры M и любого $\lambda \in S_{\infty}(E)$ последовательность конечномерных E_{λ} -векторных пространств

$$s_{f, \lambda}(M) : 0 \rightarrow H^0(F, M)_{\lambda} \rightarrow \text{Ker } \alpha_{M_{\lambda}} \rightarrow H_f^1(F, M^*(1))_{\lambda}^* \rightarrow \\ \rightarrow H_f^1(F, M)_{\lambda} \rightarrow \text{Coker } \alpha_{M_{\lambda}} \rightarrow H^0(F, M^*(1))_{\lambda}^* \rightarrow 0$$

является точной.

Отсюда получаем канонический изоморфизм из $(L_f(M) \otimes_E L_f(M^*(1)))_{\lambda}$ (здесь мы используем обозначения введения) на $\det_{E_{\lambda}}(\text{Ker } \alpha_{M_{\lambda}}) \otimes_{E_{\lambda}} \det_{E_{\lambda}}^*(\text{Coker } \alpha_{M_{\lambda}})$. С другой стороны, очевидная точная последовательность

$$0 \rightarrow \text{Ker } \alpha_{M_{\lambda}} \rightarrow \left(\bigoplus_{p \in S_{\infty}(F)} M_{B, p}^+ \right)_{\lambda} \rightarrow t_{M, \lambda} \rightarrow \text{Coker } \alpha_{M_{\lambda}} \rightarrow 0$$

дает изоморфизм между $(\det_E^*(\bigoplus_{p \in S_{\infty}(F)} M_{B, p}^+) \otimes_E \det_E t_M)_{\lambda}$ и $\det_{E_{\lambda}}^*(\text{Ker } \alpha_{M_{\lambda}}) \otimes_{E_{\lambda}} \det_{E_{\lambda}}(\text{Coker } \alpha_{M_{\lambda}})$. Тензорное произведение этих двух изоморфизмов дает изоморфизм

$$\imath_{M, \lambda} : \Delta_f(M)_{\lambda} \rightarrow E_{\lambda}.$$

Имеем $\mathbf{R} \otimes_{\mathbf{Q}} \Delta_f(M) = \bigoplus_{\lambda \in S_{\infty}(E)} \Delta_f(M)_{\lambda}$, тогда как $\mathbf{R} \otimes_{\mathbf{Q}} E = \bigoplus_{\lambda \in S_{\infty}(E)} E_{\lambda}$, и изоморфизм $\imath_M$, входящий в формулировку гипотезы $C_{DB}(M)$, определяется как прямая сумма изоморфизмов $\imath_{M, \lambda}$.

Наконец, если λ — конечная точка поля E , то свойство $C_{\lambda}(M)$ позволяет отождествить $\Delta_f(M)_{\lambda}$ и $\Delta_f(M_{\lambda})$ (см. разд. 3 и 4) и, таким образом, определить норму $|\cdot|_{\lambda, E, P}$, что позволяет сформулировать гипотезу $C_{BK}(M)$.

6.11. Замечание. Если M — мотивная структура над F , то $\text{Res}_{F/\mathbf{Q}}(M)$ является объектом категории $C_{\mathbf{Q}}(E)$, допускающим L -функцию, причем она совпадает с $L(M, s)$. Естественно предположить, что $\text{Res}_{F/\mathbf{Q}}(M)$ есть мотивная структура над $\mathbf{Q}$ и что, наоборот, если N — мотивная структура над $\mathbf{Q}$, то $\text{Ext}_{\mathbf{Q}/F}(N)$ является мотивной структурой над F и что функторы $\text{Res}_{F/\mathbf{Q}}$ и $\text{Ext}_{\mathbf{Q}/F}$

являются сопряженными друг к другу¹⁾. Если это так, то гипотезы для M эквивалентны гипотезам для $\text{Res}_{F/\mathbf{Q}}(M)$, что позволяет, не ограничивая общности, считать, что $F = \mathbf{Q}$, что мы и будем иногда делать²⁾.

7. УПРОЩЕНИЯ И РЕДУКЦИИ

7.1. В большинстве случаев, которые поддаются эффективно рассмотрению, по меньшей мере два из четырех пространств $H^0(F, M)$, $H_f^1(F, M)$, $H^0(M^*(1))$ и $H_f^1(F, M^*(1))$ равны нулю, что упрощает структуру точных последовательностей $s_{f,\lambda}$ и формулировку гипотез $C_r(M)$ и $C_{DV}(M)$.

Так обстоит дело, если $W_0 M = 0$ (в частности, если M — чистая структура веса ≥ 1); в этом случае $H^0(F, M) = H_f^1(F, M) = H^0(F, M^*(1)) = 0$ (для H^0 это следует из равенств $gr_0^W M = gr_0^W M^*(1) = 0$; что же касается H_f^1 , то заметим, что любое расширение N объекта $1_{F,E}$ с помощью M расщепляется, поскольку $W_0 N = 1_{F,E}$ является подобъектом в N , откуда вытекает, что $H^1(F, M)$, а значит, и $H_f^1(F, M)$ — нулевые группы). Если $\lambda \in S_\infty(E)$, то последовательность $s_{f,\lambda}$ сводится к одному изоморфизму

$$\text{Ker } \alpha_{M,\lambda} \longrightarrow (H_f^1(F, M^*(1)))_\lambda^*$$

(в частности, $r_M = \dim_E H_f^1(F, M^*(1)) = \dim_{E_\lambda} \text{Ker } \alpha_{M,\lambda}$, и эту размерность можно вычислить, зная лишь смешанные структуры Ходжа $M_{p,\lambda}$ для $p \in S_\infty(F)$).

Обратно, если $W_{-3} M = M$ (в частности, если M — чистая структура веса ≤ -3), то $W_0(M^*(1)) = 0$; тогда r_M должно равняться нулю, и если $\lambda \in S_\infty(F)$, то $s_{f,\lambda}(M_\lambda)$ сводится к изоморфизму

$$H_f^1(F, M)_\lambda \longrightarrow \text{Coker } \alpha_{M,\lambda}.$$

Ожидается (хотя это и не следует из развитого к настоящему времени формализма), что не существует нетривиальных расширений между двумя чистыми мотивами одного и того же веса.

¹⁾ Ожидается, что то же самое верно для функторов $\text{Res}_{F'/F}$ и $\text{Ext}_{F'/F}$ для произвольного подполя F' поля F .

²⁾ Иногда, однако, удобно работать с «полустабильными» мотивами и неполными L -функциями (см. разд. 12). В этом случае уже нельзя, вообще говоря, считать, что $F = \mathbf{Q}$.

Если это утверждение верно, то можно показать, что при условии $W_{-1}M = 0$ (в частности, если M — чистая структура веса 0) $s_{f,\lambda}(M_\lambda)$ сводится к

$$0 \rightarrow H^0(F, M)_\lambda \rightarrow \text{Ker } \alpha_{M_\lambda} \rightarrow (H_f^1(F, M^*(1))_\lambda)^* \rightarrow 0$$

(и, следовательно, снова к одному изоморфизму $\text{Ker } \alpha_{M_\lambda}$ на двойственное пространство к $(H_f^1(F, M^*(1))_\lambda$, если $H^0(F, M) = 0$), тогда как если $W_{-2}M = M$ (в частности, если M — чистая структура веса -2), то r_M должно равняться $-\dim_E H^0(F, M^*(1))$ и $s_{f,\lambda}(M_\lambda)$ сводится к

$$0 \rightarrow H_f^1(F, M)_\lambda \rightarrow \text{Coker } \alpha_{M_\lambda} \rightarrow H^0(F, M^*(1))_\lambda^* \rightarrow 0.$$

Наконец, если M является чистой структурой веса -1 , то $\text{Ker } \alpha_M = \text{Coker } \alpha_{M_\lambda} = H^0(F, M) = H^0(F, M^*(1)) = 0$ и последовательность $s_{f,\lambda}(M_\lambda)$ сводится к изоморфизму

$$H_f^1(F, M^*(1))_\lambda^* \rightarrow H_f^1(F, M)_\lambda$$

или, иначе говоря, к совершенному спариванию

$$H_f^1(F, M)_\lambda \times H_f^1(F, M^*(1))_\lambda \rightarrow E_\lambda.$$

7.2. Ситуация оказывается особенно приятной, если M есть *f -замкнутая структура*, т.е. если $H_f^1(F, M) = H_f^1(F, M^*(1)) = 0$ (для этого достаточно, чтобы $H_f^1(F, M_\lambda) = H_f^1(F, M^*(1)_\lambda) = 0$ хотя бы для одной конечной точки λ). Тогда все отображения, входящие в последовательность $s_{f,\lambda}(M)$, являются тривиальными, что позволяет сформулировать гипотезу $C_{DV}(M)$ особенно простым образом, тем более если $H^0(F, M) = H^0(F, M^*(1)) = 0$: пусть B (соотв. B') — некоторый базис в $\bigoplus_{p \in S_\infty(F)} M_{B,p}^+$ (соотв. в t_M) над E ; для любого $\lambda \in S_\infty(E)$ отображение

$$\alpha_{M_\lambda} : (\text{bigbig}_{p \in S_\infty(F)} M_{B,p}^+)_\lambda \rightarrow t_{M,\lambda}$$

должно быть изоморфизмом, и если b_λ — определитель матрицы α_{M_λ} относительно выбранных базисов, то должен существовать элемент $c \in E^*$, такой, что $L(M, 0) = c \cdot (b_\lambda)_{\lambda \in S_\infty(E)}$.

Назовем *f -эквивалентностью на M* наиболее ограничительное отношение эквивалентности на $\text{Ob}(\mathcal{M})$, такое, что

- (i) если N — расширение $1_{F,E}$ с помощью M , класс которого лежит в $H_f^1(F, M)$, то M и N f -эквивалентны;
- (ii) если M и N являются f -эквивалентными, то это же верно и для $M^*(1)$ и $N^*(1)$.

Если M и N f -эквивалентны, то $C_r(M)$ (соотв. $C_{DV}(M)$, $C_{BK}(M)$) верна тогда и только тогда, когда верна $C_r(N)$ (соотв. $C_{DV}(N)$, $C_{BK}(N)$). Поскольку любой объект категории $\mathcal{M}$ является f -эквивалентным некоторому f -замкнутому объекту, то теоретически эти гипотезы достаточно проверять только для f -замкнутых объектов.

7.3. Чистая мотивная структура M веса w называется *критической*, если существует точка $\lambda \in S_\infty(E)$, такая, что α_{M_λ} — изоморфизм (и тогда α_{M_μ} будет изоморфизмом для любой точки $\mu \in S_\infty(E)$). Если вес структуры M не равен -1 , то отсюда следует ее f -замкнутость. Если же вес M равен -1 и M является критической, то гипотетически $r_M \geq 0$ и M является f -замкнутой тогда и только тогда, когда $r_M = 0$, т. е. когда $L(M, 0) \neq 0$.

В. ЧАСТНЫЕ СЛУЧАИ

8. СЛУЧАЙ 1-МОТИВОВ

8.1. Напомним [De74, §10], что полуабелево многообразие над F — это расширение абелева многообразия с помощью тора и что *гладкий 1-мотив над F* — это комплекс коммутативных групповых схем над F , сосредоточенный в степенях -1 и 0 ,

$$\mathcal{X} \xrightarrow{U} S,$$

где S — полуабелево многообразие, а $\mathcal{X}$ локальна постоянна над F в этальной топологии, т. е. $\mathcal{X}(\bar{F})$ является свободным $\mathbf{Z}$ -модулем конечного ранга.

Гладкие 1-мотивы над F образуют аддитивную категорию $\mathcal{MM}_{F,1}$. Обозначим через $\mathcal{MM}_{F,1}(\mathbf{Q})$ категорию 1-мотивов с точностью до изогении, т. е. следующую категорию:

- ее объекты — это гладкие 1-мотивы,
- $\text{Hom}_{\mathcal{MM}_{F,1}(\mathbf{Q})}(M_1, M_2) = \mathbf{Q} \otimes_{\mathbf{Z}} \text{Hom}_{\mathcal{MM}_{F,1}}(M_1, M_2)$.

Нетрудно видеть, что $\mathbf{Q}$ -линейная категория $\mathcal{MM}_{F,1}(\mathbf{Q})$ является абелевой. Обозначим через $\mathcal{MM}_{F,1}(E)$ категорию 1-мотивов

над F с коэффициентами в E , т. е. E -линейную категорию, получающуюся из $\mathcal{MM}_{F,1}(\mathbf{Q})$ расширением скаляров $\mathbf{Q} \rightarrow E$. Объект M категории $\mathcal{MM}_{F,1}(E)$ удобно рассматривать как комплекс $[X \xrightarrow{u} S]$, где X — конечномерное векторное пространство над E , снабженное дискретным действием группы G_F , S — полуабелево многообразие над F , снабженное вложением E в $\mathbf{Q} \otimes_{\mathbf{Z}} \text{End}_F(S)$, и $u : X \rightarrow \mathbf{Q} \otimes_{\mathbf{Z}} S(\bar{F})$ есть E -линейное G_F -эквивариантное отображение.

Для каждого 1-мотива $M = [X \xrightarrow{u} S]$ над F с коэффициентами в E можно определить (loc.cit) его весовую фильтрацию и связать с ним его реализацию де Рама M_{dR} (это алгебра Ли универсального расширения мотива M с помощью векторной группы, тогда как касательное пространство к M — это алгебра Ли многообразия S), реализацию Бетти $M_{B,p}$ (называемую реализацией Ходжа в loc.cit) для любой бесконечной точки p поля F и реализацию Галуа M_f (имеем $M_f = \varprojlim_n M_n$, где

$$M_n = \{(x, s) \in X \times S(\bar{F}) \mid u(x) = ns\} / \{(nx, u(x)) \mid x \in X\}.$$

Известно также [Bures], как определить изоморфизмы сравнения, и можно показать, что таким образом получается объект категории $\mathcal{C}_F(E)$. Эта конструкция является функториальной. Из гипотезы Тейта для абелевых многообразий (доказанной Фалтингсом, см. [De83]) можно легко вывести, что функтор

$$\underline{\text{Real}} : \mathcal{MM}_{F,1}(E) \longrightarrow \mathcal{C}_F(E)$$

является вполне строгим и потому индуцирует эквивалентность между категорией $\mathcal{MM}_{F,1}$ и ее существенным образом, который мы обозначим через $\underline{SM}_{F,1}(E)$. Этот образ содержит объект $\mathbf{1}_{F,E}$ (являющийся реализацией $[E \rightarrow 0]$) и является устойчивым относительно $M \mapsto M^*(1)$.

Если $M = \underline{\text{Real}}([X \xrightarrow{u} S])$, то всякий элемент $s \in \mathbf{Q} \otimes S(F)$ определяет некоторое расширение $\mathbf{1}_{F,E} = \underline{\text{Real}}([E \rightarrow 0])$ с помощью M , а именно $\underline{\text{Real}}([X \oplus E \xrightarrow{(u,v)} S])$, где $v(1) = s$. Это расширение расщепляется тогда и только тогда, когда $s \in u(X)$. Отсюда получаем точную последовательность

$$X \longrightarrow \mathbf{Q} \otimes S(F) \longrightarrow H^1_{\underline{SM}_{F,1}(E)}(F, M) \longrightarrow 0.$$

8.2. Если M — объект категории $\underline{SM}_{F,1}(E)$, то $h^{r,s}(M) = 0^1$ для всех r, s , кроме, возможно, $r, s \in \{-1, 0\}$. Кажется разумным предположить, что $\underline{SM}_{F,1}(E)$ — это в точности подкатегория в $\underline{SM}_F(E)$, образованная теми объектами M , для которых $h^{r,s} = 0$, если $r \notin \{0, -1\}$ или $s \notin \{0, -1\}$.

В частности, $\underline{SM}_{F,1}(E)$ должна быть подкатегорией в $\underline{SM}_F(E)$, устойчивой относительно взятия подобъекта, факторобъекта и прямой суммы и такой, что если она содержит M , то содержит и любое расширение $1_{F,E}$ с помощью M . Таким образом, она должна быть f -допустимой²), т.е. любой объект $M = \text{Real}([X \xrightarrow{\sim} S])$ должен удовлетворять гипотезе $C_\lambda^i(M)$ для каждой точки λ поля E . Нетрудно видеть, что это действительно так, если точка λ архимедова. Если же λ конечна, то можно проверить, что отображение $E_\lambda \otimes_E H^0(F, M) \rightarrow H^0(F, M_\lambda)$ является изоморфизмом, в то время как отображение $E_\lambda \otimes_E H_f^1(F, M) \rightarrow H_f^1(F, M_\lambda)$ инъективно. Пусть S — расширение абелева многообразия A с помощью тора T ; тогда сюръективность последнего отображения для всех точек λ , делящих данное простое l , эквивалентна (см. [BK90, Prop. 5.4]) конечности l -примарной компоненты $\text{Ш}(A)(l)$ группы Шафаревича абелева многообразия A ,

$$\text{Ш}(A) = \text{Ker}(H^1(F, A(\overline{F})) \rightarrow \prod_{p \in S(F)} H^1(F_p, A(\overline{F}_p))).$$

Следовательно, справедливо

Предложение. Категория $\underline{SM}_{F,1}(\mathbb{Q})$ является f -допустимой тогда и только тогда, когда для любого абелева многообразия A над F и любого простого числа l группа $\text{Ш}(A)(l)$ конечна. Если это так, то $\underline{SM}_{F,1}(E)$ также является f -допустимой.

Кроме того, уже установленных свойств категории $\underline{SM}_{F,1}(E)$ нам достаточно для того, чтобы построить точную последовательность $s_{f,\lambda(M)}$ для любого объекта M категории $\underline{SM}_{F,1}(E)$ и любой точки $\lambda \in S_\infty(E)$, даже и не предполагая конечности групп Шафаревича.

¹) Можно определить числа Ходжа произвольной мотивной структуры M , положив $h^{r,s}(M) = \dim_E(\text{Fil}^r(M)/\text{Fil}^{r+1}(M))$, если M — чистая структура веса $r + s$, и $h^{r,s}(M) = h^{r,s}(gr_{r+s}^W M)$ в общем случае.

²) Мы будем писать $H^1(F, M) = H_{\underline{SM}_{F,1}(E)}^1(F, M)$ и $H_f^1(F, M) = H_{\underline{SM}_{F,1}(E), f}^1(M)$.

8.3. Положим $E = \mathbf{Q}$ и рассмотрим некоторые частные случаи.

а) Случай $M = \mathbf{1}_{F, \mathbf{Q}} = \text{Real}([\mathbf{Q} \rightarrow 0])$. Тогда $M^*(1) = \mathbf{1}_{F, \mathbf{Q}}(1) = \text{Real}([0 \rightarrow \mathbb{G}_m])$. Имеем $H^0(F, M) = \mathbf{Q}$, $H_f^1(F, M) = H^1(F, M) = 0$, $H^0(F, M^*(1)) = 0$, $H^1(F, M^*(1)) = \mathbf{Q} \otimes F^*$, тогда как $H_f^1(F, M^*(1))$ отождествляется с $\mathbf{Q}$ -векторным подпространством $\mathbf{Q} \otimes U_F$ в $\mathbf{Q} \otimes F^*$, где U_F обозначает группу единиц числового поля F . Для любой точки $p \in S_\infty(F)$ имеем $M_{B, p} = \mathbf{Q}$ с тривиальным действием группы G_F и $t_M = 0$, так что α_{M_∞} является нулевым отображением, $\text{Ker } \alpha_{M_\infty} = \bigoplus_{p \in S_\infty(F)} \mathbf{R}$ и $\text{Coker } \alpha_{M_\infty} = 0$. Точная последовательность $s_{f, \infty}(M)$ сводится к последовательности

$$0 \rightarrow \mathbf{R} \rightarrow \bigoplus_{p \in S_\infty(F)} \mathbf{R} \rightarrow (\mathbf{R} \otimes_{\mathbf{Z}} U_F)^* \rightarrow 0,$$

где $\mathbf{R} \rightarrow \bigoplus_{p \in S_\infty(F)} \mathbf{R}$ — диагональное отображение, а следующее отображение является сопряженным к $\mathbf{R}$ -линейному отображению, получающемуся расширением скаляров из отображения

$$U_F \rightarrow \bigoplus_{p \in S_\infty(F)} \mathbf{R},$$

которое переводит u в $(\log \|u\|_p)_{p \in S_\infty(F)}^{(1)}$.

Функция $L(M, s)$ совпадает с функцией $\zeta(F, s)$ числового поля F . Гипотеза $C_r(M)$ утверждает, что если $r = \#S_\infty(F)$, то $\zeta(F, s)$ имеет в точке $s = 0$ нуль порядка $r - 1$; гипотеза $C_{DB}(M)$ утверждает, что если $u_1, u_2, \dots, u_{r-1}$ — линейно независимая над $\mathbf{Z}$ система единиц поля F , $p_1, p_2, \dots, p_{r-1}$ — различные элементы из $S_\infty(F)$ и $R = \det(\log \|u_i\|_{p_j})$, то $\zeta^*(F, 0)/R^* \in \mathbf{Q}^*$, тогда как техника, кратко описанная ниже в разд. 11, позволяет интерпретировать гипотезу $C_{BK}(M)$ как утверждение о том, что если e — индекс в U_F подгруппы, порожденной набором $\{u_i\}$, и h — число классов поля F , то

$$\zeta^*(F, 0) = \pm Rh/e,$$

что хорошо известно (см., например, [Та84, Chap. I]).

б) Случай $\mathbf{1}_{F, \mathbf{Q}}(1) = \text{Real}[0 \rightarrow \mathbb{G}_m]$. Этот случай является «двойственным к предыдущему». Если положить $t = 2\pi i \in \mathbf{C}$, то $M_{B, p} = \mathbf{Q} \cdot t$ для любой точки $p \in S_\infty(F)$; следовательно, $M_{B, p}^+ = 0$,

¹⁾ Где $\| \cdot \|_p$ — обычное абсолютное значение, если точка p вещественная, и его квадрат, если p комплексная.

если p вещественная, и $M_{M,p}^+ = \mathbb{Q} \cdot t$ в противном случае. Имеем $t_M = F$. Если обозначить множество вещественных (соотв. комплексных) точек поля F через $S_R(F)$ (соотв. $S_C(F)$), то отображение

$$\alpha_{1_F, \mathbb{Q}(1)_\infty} : \bigoplus_{p \in S_C(F)} \mathbb{R} \cdot t \longrightarrow \mathbb{R} \otimes_{\mathbb{Q}} F = \left(\bigoplus_{p \in S_R(F)} \mathbb{R} \right) \oplus \left(\bigoplus_{p \in S_C(F)} \mathbb{C} \right)$$

переводит $(u_p \cdot t)_{p \in S_C(F)}$ в $(0, (u_p \cdot 2\pi i)_{p \in S_C(F)})$. Имеем $\text{Ker } \alpha_{M_\infty} = 0$ и $\text{Coker } \alpha_{M_\infty} = \bigoplus_{p \in S_\infty(F)} \mathbb{R}$, что канонически отождествляется со своим двойственным и, следовательно, с $\text{Ker } \alpha_{(1_F, \mathbb{Q})_\infty}$. Точная последовательность $s_{F, \infty}(M)$ является двойственной к $s_{F, \infty}(1_F, \mathbb{Q})$.

Имеем $L(M, s) = \zeta(F, s+1)$. Гипотеза $C_r(M)$ утверждает, что $\zeta(F, s)$ имеет простой полюс в точке $s=1$, $C_{DB}(M)$ утверждает, что если d — дискриминант поля F , то $\zeta^*(F, 1) \cdot \sqrt{d}/R \in \mathbb{Q}^*$, и $C_{BK}(M)$ может быть переформулирована с помощью техники разд. 11 как утверждение о том, что

$$\zeta^*(F, 1) = \pm 2^{r_1} \cdot (2\pi)^{r_2} \cdot d^{-1/2} \cdot hR/e,$$

что снова хорошо известно (loc.cit.).

с) *Случай нетривиального расширения $1_{F, \mathbb{Q}}$ с помощью $1_{F, \mathbb{Q}(1)}$.* Мы будем в дальнейшем предполагать для простоты, что $F = \mathbb{Q}$. Тогда существует натуральное число $q \geq 2$, такое, что $M = \text{Real}([\mathbb{Q} \xrightarrow{u} \mathbb{G}_m])$, где $u(1) = q$. Имеем $W_{-1}M = 1_{\mathbb{Q}, \mathbb{Q}(1)}$. Можно записать $M_{B, \infty} = \mathbb{Q} \cdot t \oplus \mathbb{Q}u$; тогда $M_{B, \infty}^+ = \mathbb{Q}u$, $t_M = t_{W_{-1}M} = \mathbb{Q}$. Отображение α_{M_∞} переводит u в $\log(q)$ и является изоморфизмом. Мотивная структура $M^*(1)$ канонически отождествляется с M . Имеем $H^0(\mathbb{Q}, M) = 0$, а $\mathbb{Q}$ -векторное пространство $H^1(\mathbb{Q}, M)$ отождествляется с фактором $\mathbb{Q} \otimes \mathbb{Q}^*$ по прямой, порожденной $1 \otimes q$. Положим $q = p_1^{r_1} \cdot p_2^{r_2} \cdot \dots \cdot p_a^{r_a}$, где p_i — различные простые числа, а r_i — целые числа ≥ 1 . Можно проверить, что $H_f^1(\mathbb{Q}, M)$ отождествляется с $\mathbb{Q}$ -векторным подпространством в $H^1(\mathbb{Q}, M)$, порожденным образами элементов $1 \otimes p_i$. Точная последовательность $s_{f, \infty}(M)$ сводится к изоморфизму $H_f^1(\mathbb{Q}, M)_\infty^*$ на $H_f^1(\mathbb{Q}, M)_\infty$, что эквивалентно заданию билинейной невырожденной формы

$$h : H_f^1(\mathbb{Q}, M)_\infty \times H_f^1(\mathbb{Q}, M)_\infty \longrightarrow \mathbb{R};$$

если положить $\varepsilon_{i,i} = \log(p_i)/r_i$ и $\varepsilon_{i,j} = 0$ при $i \neq j$, то форма h задается соотношениями

$$h(p_i, p_j) = (\log(p_i) \cdot \log(p_j)) / \log(q) - \varepsilon_{i,j}.$$

Имеем $L(M, s) = \zeta(s) \cdot \zeta(s+1) \cdot \prod_{1 \leq i \leq a} (1 - p_i^{-s})$. Гипотезы $C_r(M)$, $C_{DB}(M)$ и $C_{BK}(M)$ тогда являются теоремами, которые легко выводятся из результатов о значениях дзета-функции Римана при $s=0$ и $s=1$.

d) *Случай абелева многообразия*, т. е. случай, когда $M = \text{Real}[0 \rightarrow A]$, где A — абелево многообразие над F . Применяя к A функтор ограничения скаляров по Вейлю, можно предполагать, что $F = \mathbf{Q}$. Легко проверить, что $M^*(1)$ отождествляется с M , что α_{M_∞} является изоморфизмом, что $H^0(\mathbf{Q}, M) = 0$ и что $H_f^1(\mathbf{Q}, M) = H^1(\mathbf{Q}, M) = \mathbf{Q} \otimes A(\mathbf{Q})$, так что точная последовательность $s_{f,\infty}(M)$ снова сводится к изоморфизму $H_f^1(\mathbf{Q}, M)_\infty^*$ на $H_f^1(\mathbf{Q}, M)_\infty$, а это эквивалентно заданию невырожденной билинейной формы

$$h : \mathbf{R} \otimes A(\mathbf{Q}) \times \mathbf{R} \otimes A(\mathbf{Q}) \rightarrow \mathbf{R}.$$

Можно показать (см. [Bl80]), что эта форма является не чем иным, как *билинейной формой, отвечающей высоте Нерона—Тейта*.

Имеем $L(M, s) = L_A(s+1)$, где L_A есть L -функция Хассе—Вейля абелева многообразия A . Техника разд. 11 позволяет показать, что объединение гипотез $C_r(M)$ и $C_{BK}(M)$ эквивалентно гипотезе Бёрча—Свиннертона-Дайера¹⁾. В последнее время в доказательстве этой гипотезы достигнут значительный прогресс (см. [Ko90, Pe89, Gr91, Ru91]).

9. МОТИВЫ ГЛАДКИХ ПРОЕКТИВНЫХ МНОГООБРАЗИЙ

9.1. Пусть X — гладкое проективное многообразие над $\mathbf{Q}$ ²⁾. Будем изучать мотивные структуры³⁾ $H^i(X)(m)$, где $i, m \in \mathbf{Z}$ и $i \geq 1$. Если $M = H^i(X)(m)$, то двойственность Пуанкаре вместе с «трудной»

¹⁾ В предположении конечности группы $\text{Ш}(A)$, которая необходима для придания смысла как гипотезе $C_{BK}(M)$, так и гипотезе Бёрча—Свиннертона-Дайера.

²⁾ Случай гладкого проективного многообразия Y над F сводится к данному, если взять за X ограничение скаляров по Вейлю для Y .

³⁾ Я слегка лукавлю: при сопоставлении $H^i(X)(m)$ объекта категории $C_F(\mathbf{Q})$ могут возникнуть проблемы в точках плохой редукции.

теоремой Лефшеца позволяют отождествить $M^*(1)$ с $H^i(X)(i+1-m)$. Мы будем изучать одновременно гипотезы для M и $M^*(1)$, т. е. значения $L(H^i(X), s)$ в точках $s = m$ и $s = n$, где $n = i+1-m$, так что можно считать, что $m \leq (i+1)/2 \leq n$. Заметим, что M является чистой структурой веса $i-2m$, тогда как $M^*(1)$ — чистой структурой веса $i-2n$.

9.2. Предположим сначала, что $m < i/2$, откуда вытекает, что M имеет вес ≥ 1 и $M^*(1)$ имеет вес ≤ -3 . Получаем, следовательно (п. 7.1), что $H_f^1(\mathbf{Q}, M) = 0$, $r_{M^*(1)} = 0$ и $r_M = \dim_{\mathbf{Q}} H_f^1(\mathbf{Q}, M^*(1)) = \dim_{\mathbf{R}} \text{Ker } \alpha_{M_\infty} = \dim_{\mathbf{R}} \text{Coker } \alpha_{M^*(1)_\infty}$. Можно проверить, что $\text{Coker } \alpha_{M^*(1)_\infty} = H_{dR}^i(X)_{\mathbf{R}} / (F\text{il}^n H_{dR}^i(X)_{\mathbf{R}} + H_B^i(X(\mathbf{C}), \mathbf{Q}(n))_{\mathbf{R}})$ отождествляется с группой когомологий Делиня $H_D^{i+1}(X/\mathbf{R}, \mathbf{R}(n))$ (см., например, [Sc88; Ja88, §4]). По модулю обычных гипотез о функциональном уравнении для функции $L(H^i(X), s)$ (ор. cit., §1) гипотезы $C_r(M)$ и $C_r(M^*(1))$ являются теоремами, а гипотезы $C_{DB}(M)$ и $C_{DB}(M^*(1))$ эквивалентны (ситуация со связью между $C_{BK}(M)$ и $C_{BK}(M^*(1))$ является несколько более деликатной, см. разд. 12).

Легко также проверить, что M является критическим тогда и только тогда, когда $H_D^{i+1}(X/\mathbf{R}, \mathbf{R}(n)) = 0$, и что в этом случае $C_{DB}(M)$ и $C_{DB}(M^*(1))$ являются частными случаями гипотезы Делиня [De79, conj. 1.8].

9.3. Чтобы прояснить связь с гипотезой Бейлинсона, напомним (см., например, [So85]), что для любого $r \in \mathbf{N}$ на $K_r(X) \otimes \mathbf{Q}$ определены операции Адамса $(\psi^a)_{a \in \mathbf{N}^*}$ и $K_r(X) \otimes \mathbf{Q} = \bigoplus_{j \in \mathbf{N}} K_r(X)^{(j)} \otimes \mathbf{Q}$, где $K_r(X)^{(j)}$ обозначает $\mathbf{Q}$ -векторное подпространство в $K_r(X) \otimes \mathbf{Q}$, на котором ψ^a действует умножением на a^j для любого a . Напомним также, что Бейлинсон ввел группу мотивных когомологий $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n)) = K_{n-m}(X)^{(n)}$ и построил некоторое отображение

$$r_{i,n,\infty} : H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n)) \longrightarrow H_D^{i+1}(X/\mathbf{R}, \mathbf{R}(n)).$$

Известно также (по крайней мере, если X имеет регулярную модель над $\mathbf{Z}$, — в противном случае требуется еще одна гипотеза, см. ор. cit.), что можно определить некоторое $\mathbf{Q}$ -векторное подпространство $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}}$ в $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))$; гипотеза Бейлинсона утверждает, что в данной ситуации $r_{i,n,\infty}$ индуцирует изоморфизм

$$c_{i,n,\infty} : \mathbf{R} \otimes_{\mathbf{Q}} H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}} \longrightarrow H_D^{i+1}(X/\mathbf{R}, \mathbf{R}(n))$$

и, таким образом, изоморфизм $d_{i,n,\infty}$ между детерминантами этих $\mathbf{R}$ -векторных пространств и что если $b_{\mathcal{D}}$ — «хорошо» выбранный базис в¹⁾ $\det_{\mathbf{R}} H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n))$, то существует базис $b_{\mathcal{M}}$ в $\det H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}}$, такой, что $d_{i,n,\infty}(b_{\mathcal{M}}) = L^*(H^i(X), n) \cdot b_{\mathcal{D}}$.

Кажется естественным предположить, что $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}}$ должно отождествляться с $H_f^1(\mathbf{Q}, M^*(1))$, а $c_{i,n,\infty}$ — с естественным отображением из $H_f^1(\mathbf{Q}, M^*(1))_{\mathbf{R}}$ в $\mathrm{Coker} \alpha_{M^*(1)\infty}$. По модулю этого отождествления и функционального уравнения проверка эквивалентности гипотезы $C_{DV}(M)$ гипотезе Бейлинсона является простым упражнением.

9.4. Используя l -адические классы Черна, можно также для любого простого l определить [So81] отображения

$$r_{i,n,l} : H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n)) \longrightarrow H^1(\mathbf{Q}, H_{\text{ét}}^i(X \otimes \overline{\mathbf{Q}}, \mathbf{Q}_l(n)))$$

Предполагается, что образ пространства $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}}$ содержится в $H_f^1(\mathbf{Q}, H_{\text{ét}}^i(X \otimes \overline{\mathbf{Q}}, \mathbf{Q}_l(n)))$ (при достаточно общих предположениях это доказано Гроссом [Gr90])²⁾. Разумеется, на самом деле ожидается, что это отображение индуцирует $\mathbf{Q}_l$ -изоморфизм³⁾

$$c_{n,i,l} : \mathbf{Q}_l \otimes H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}} \longrightarrow H_f^1(\mathbf{Q}, H_{\text{ét}}^i(X \otimes \overline{\mathbf{Q}}, \mathbf{Q}_l(n)))$$

и что последний совпадает с естественным изоморфизмом из $\mathbf{Q}_l \otimes H_f^1(\mathbf{Q}, M^*(1))$ на $H_f^1(\mathbf{Q}, M^*(1)_l)$. Иначе говоря, использование *мотивных групп когомологий* $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}}$ и *регуляторов* $c_{n,i,l}$ для (конечных или бесконечных) точек l поля $\mathbf{Q}$ позволяет нам опустить кавычки при формулировке гипотез $C_{BK}(M)$ и $C_{BK}(M^*(1))$, т. е. переформулировать их таким образом, чтобы в них не упоминалась категория $\mathcal{M}$.

¹⁾ Более точно, $\det_{\mathbf{R}} H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n))$ можно отождествить с $\mathbf{R} \otimes_{\mathbf{Q}} V$, где $V = \det_{\mathbf{Q}} H^i(X(\mathbf{C}), \mathbf{Q}(n-1))^{(-1)^{n-1}} \otimes \det_{\mathbf{Q}} \mathrm{Fil}^n H_{\mathrm{dR}}^i(X)$, и нужно взять $b_{\mathcal{D}} = 1 \otimes b$, где b — некоторый базис в V .

²⁾ В контексте гипотез Блоха–Като кажется более простым определить $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_f$ как подгруппу в $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))$, образованную такими x , что $r_{n,i,l}(x) \in H_f^1(\mathbf{Q}, H_{\text{ét}}^i(X \otimes \overline{\mathbf{Q}}, \mathbf{Q}_l(n)))$ для любого l , и заменить в формулировке гипотез $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}}$ на $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_f$, избежав таким образом необходимости предполагать, что $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}} = H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_f$.

³⁾ Тот факт, что $c_{n,i,l}$ есть изоморфизм, доказан в некоторых частных случаях Янсеном [Ja89, Ja90].

9.5. *Случай $m = i/2$* , где i четно, очень похож на предыдущий, но, поскольку вес структуры M равен 0, то возможно, что $H^0(Q, M) \neq 0$. Функция $L(H^{2m}(X), s)$ снова должна иметь в точке $s = m$ нуль порядка $\dim_{\mathbf{R}} H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n))$, но у нее также должен быть полюс порядка $\dim_{\mathbf{Q}} H^0(Q, H^{2m}(X)(m))$ в точке $s = n = m + 1$. При той же переформулировке, что и в случае $m < i/2$, точная последовательность $s_{f, \infty}(M^*(1))$ записывается в виде

$$0 \rightarrow \mathbf{R} \otimes_{\mathbf{Q}} H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}} \xrightarrow{!} H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n)) \rightarrow \mathbf{R} \otimes_{\mathbf{Q}} H^0(Q, M)^* \rightarrow 0,$$

и если $C_h^m(X)$ обозначает группу классов циклов коразмерности m по модулю гомологической эквивалентности, то отображение, входящее в гипотезу Бейлинсона (предполагается, что оно является изоморфизмом), имеет вид

$$(\mathbf{R} \otimes_{\mathbf{Q}} H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}}) \oplus (\mathbf{R} \otimes C_h^m(X)) \rightarrow H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n)),$$

где $\mathbf{R} \otimes C_h^n(X) \rightarrow H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n))$ индуцировано отображением цикла в когомологиях де Рама. По модулю той же интерпретации, что и при $m < i/2$, и стандартной гипотезы, которая утверждает, что естественное отображение $C_h^m(X)_{\mathbf{Q}} \rightarrow H^0(Q, M)$ есть изоморфизм, можно проверить, что гипотеза $C_{DV}(M)$ снова эквивалентна гипотезе Бейлинсона¹⁾.

9.6. Рассмотрим, наконец, *случай, когда $m = n = (i + 1)/2$* (где i нечетно). Тогда M является чистой структурой веса -1 и отождествляется с $M^*(1)$, откуда следует (см. п. 7.1), что точная последовательность $s_{f, \infty}(M^*(1))$ соответствует наличию невырожденной билинейной формы

$$(\cdot, \cdot)_M : \mathbf{R} \otimes H_f^1(Q, M) \times \mathbf{R} \otimes H_f^1(Q, M) \rightarrow \mathbf{R},$$

и размерность $\mathbf{Q}$ -векторного пространства $H_f^1(Q, M)$ должна равняться порядку нуля функции $L(H^{2m-1}(X), s)$ в точке $s = m$.

¹⁾ Чтобы это доказать, нужно вычислить отображение-композицию $\mathbf{R} \otimes C_h^m(X) \rightarrow H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n)) \rightarrow \mathbf{R} \otimes H^0(Q, M)^*$. Оно пропускается через естественное отображение $\mathbf{R} \otimes C_h^m(X)$ в $\mathbf{R} \otimes H^0(Q, M)$. Возникающее при этом отображение $\mathbf{R} \otimes H^0(Q, M) \rightarrow \mathbf{R} \otimes H^0(Q, M)^*$ получается расширением скаляров из изоморфизма $H^0(Q, M) \rightarrow H^0(Q, M)^*$, который возникает из отождествления M с двойственной ей структурой («трудная» теорема Лефшеца плюс двойственность Пуанкаре).

Напомним [Be87, Bl84b], что в этом случае, если $CH^m(X)^0$ обозначает группу классов циклов коразмерности m , гомологичных нулю, по модулю рациональной эквивалентности и d — размерность X , то существует спаривание

$$\langle , \rangle_m : CH^m(X)^0 \times CH^{d+1-m}(X)^0 \longrightarrow \mathbf{R}.$$

Предполагается (loc.cit), что $\mathbf{Q} \otimes CH^m(X)^0$ имеет конечную размерность над $\mathbf{Q}$, что $\mathbf{R}$ -векторные пространства $\mathbf{R} \otimes CH^m(X)^0$ и $\mathbf{R} \otimes CH^{d+1-m}(X)^0$ являются двойственными друг к другу относительно этого спаривания и что если $L \in CH^1(X)$ — класс некоторого гиперплоского сечения, то (пусть $m \leq (d+1)/2$; если это не так, то нужно поменять местами m и $d+1-m$) естественное отображение

$$L^{d+1-2m} : \mathbf{Q} \otimes CH^m(X)^0 \longrightarrow \mathbf{Q} \otimes CH^{d+1-m}(X)^0$$

есть изоморфизм.

Если с его помощью отождествить два последних $\mathbf{Q}$ -векторных пространства, то $\langle , \rangle_m$ даст некоторую билинейную форму на $\mathbf{R} \otimes CH^m(X)^0$. Ожидается, что существует естественный изоморфизм $\mathbf{Q} \otimes CH^m(X)^0$ на $H_f^1(\mathbf{Q}, M)$, отождествляющий $\langle , \rangle_m$ с $\langle , \rangle_M$. Если это так, то $C_{DV}(M)$ эквивалентна гипотезе Бейлинсона.

Для каждого простого l можно определить [Ja90] l -адическое отображение Абеля—Якоби

$$CH^m(X)^0 \longrightarrow H^1(\mathbf{Q}, H_{\text{ét}}^{2m-1}(X \otimes \overline{\mathbf{Q}}, \mathbf{Q}_l(m))) = H^1(\mathbf{Q}, M_l).$$

Можно снова ожидать, что его образ содержится в H_f^1 и что это отображение можно отождествить с естественным отображением из $H_f^1(\mathbf{Q}, M) \otimes \mathbf{Q}_l$ в $H_f^1(\mathbf{Q}, M_l)$.

10. СКРУЧЕННЫЕ МОТИВЫ АРТИНА

10.1. Блох и Като [BK90] доказали (по модулю некоторых мелочей) свою гипотезу для значений дзета-функции Римана в целых положительных числах. Их метод дает несколько больше: пусть d — целое ≥ 1 и χ — характер Дирихле, т.е. гомоморфизм из $\text{Gal}(\mathbf{Q}(\mu_d)/\mathbf{Q}) = (\mathbf{Z}/d\mathbf{Z})^*$ в E^* . Он определяет мотив Артина $E(\chi)$ над $\mathbf{Q}$ с коэффициентами в E (который можно рассматривать как реализацию 1-мотива $[E \rightarrow 0]$, где $G_{\mathbf{Q}}$ действует на E умножением

на характер χ). Нас интересует поведение функции $L(\chi, s)$ в точке $s = m$, т. е. поведение $L(M, s)$ при $s = 0$, где $M = E(\chi)(m)$. Скажем, что χ является четным (соотв. нечетным), если $\chi(-1) = 1$ (соотв. $\chi(-1) = -1$).

а) Если $m = 0$ или 1 , то мы имеем дело с 1-мотивами и гипотезы верны.

б) Если $m < 0$ и m и χ имеют различную четность, то $C_r(M)$ и $C_{DB}(M)$ утверждают, что $L(\chi, m)$ является ненулевым рациональным числом, что хорошо известно; в этом случае соответствующая мотивная структура является критической, и если l — простое число, то из «основной гипотезы» теории Ивасабы (для l) можно вывести, что если b такое же, как в формулировке $C_{DB}(M)$, то $|1 \otimes b|_{l,EP} = 1$; согласно Мазуру и Вайлсу [MW84], эта гипотеза является теоремой, если l^2 не делит $d^{(1)}$.

с) Если $m \geq 2$ и m и χ имеют одинаковую четность, то в этом случае наша мотивная структура снова является критической (двойственной к предыдущей, если заменить χ на χ^{-1}); функциональное уравнение показывает, что $C_r(M)$ и $C_{DB}(M)$ верны. Достаточно тонкий локальный анализ позволяет вывести из предыдущего случая, что если b такое, как в $C_{DB}(M)$, то $|1 \otimes b|_{l,EP} = 1$, по крайней мере если l не делит d .

д) Если $m < 0$ и m и χ имеют одинаковую четность, то функция $L(M, s)$ имеет простой нуль в точке $s = 0$ и, следовательно, $\dim_E H_f^1(\mathbf{Q}, M^*(1))$ должна равняться 1, где $M^*(1) = E(\chi^{-1})(1-m)$. Делинь построил некоторое нетривиальное расширение $1_{\mathbf{Q},E}$ с помощью $M^*(1)$ в категории $C_{\mathbf{Q}}(E)$ ([De89, п° 3.18]; для полной строгости нужно еще проверить существование изоморфизмов сравнения для p , делящих d). Ожидается, разумеется, что это расширение является мотивной структурой и что его класс порождает $H_f^1(\mathbf{Q}, M^*(1))$. Если это так, то $C_{DB}(M)$ верна и, в прежних обозначениях, $|1 \otimes b|_{l,EP} = 1$, по крайней мере, если l не делит $\varphi(d)$.

е) Наконец, случай, когда $m \geq 2$ и четности различны, выводится из д) так же, как с) из б): по модулю того, что расширение Делиня является «хорошим», устанавливаем, что $C_{DB}(M)$ верна и что $|1 \otimes b|_{l,EP} = 1$, по крайней мере если l не делит $\varphi(d)$.

10.2. Замечание. Рассмотрим случай е); можно определить группу мотивных когомологий $K_{2m-1}(\chi)^{(m)}$, являющуюся E -векторным пространством размерности 1, и регуляторы [So79], позволяющие,

¹⁾ Иначе говоря, $C_{BK}(M)$ верна с точностью до умножения на рациональное число, являющееся единицей, помимо тех простых l , для которых $l^2 | d$.

как и в пп. 9.3 и 9.4, переформулировать гипотезы без ссылок на расширения мотивов. Похоже, что Бейлинсон доказал [Be90], что эти конструкции совместимы с конструкциями Делиня.

10.3. Предположим, что поле F является вполне вещественным, и рассмотрим функцию $\zeta(F, s)$ в окрестности $s = m$. Конечно, это частный случай ситуации, рассматривавшейся в разд. 9, где $X = \text{Спец } F$ и $i = 0$. Видно, что целые отрицательные нечетные m являются критическими. Гипотеза $C_{DV}(M)$ утверждает, что для любого такого m число $\zeta(F, m)$ рационально и не равно нулю, и можно показать, что $C_{BK}(M)$ эквивалентна гипотезе Лихтенбаума [Li72]. В тех же обозначениях, что и выше, $C_{BK}(M)$ сводится к проверке того, что $|1 \otimes b|_{l, EP} = 1$ для любого l . Известно [BN78], что для фиксированного l это условие следует из соответствующей «основной гипотезы». В частности, если $F/\mathbb{Q}$ абелево, это верно, по крайней мере если l^2 не делит кондуктор F^1 .

10.4. Замечания. (i) Рассмотрим категорию *смешанных мотивов Тейта над F* , т. е. полную подкатегорию категории смешанных мотивов над F с коэффициентами в $\mathbb{Q}$, объекты которой есть последовательные расширения скручиваний по Тейту единичного объекта. Разумеется, единственная интересная L -функция, возникающая из этой категории, — это функция $\zeta(F, s)$. Между тем результаты Делиня о $\mathbb{P}^1 - \{0, 1, \infty\}$ ([De89], см. также [Il89, BGSV90]), по-видимому, дают способ реализации этой категории как полной подкатегории категории $\mathcal{C}_F(\mathbb{Q})$; интересно было бы проверить, так ли это на самом деле и является ли полученная категория f -допустимой. Можно поставить тот же вопрос и для несколько большей категории, образованной мотивами, являющимися последовательными расширениями скрученных мотивов Артина.

(ii) Мы лишь слегка коснулись понятия мотивных когомологий, которые все-таки являются одним из главных инструментов в исследованиях, связанных с данными гипотезами.

(iii) Похоже, что блестящее будущее ожидает *использование многообразий Шимуры для построения смешанных мотивов* (это позволяет, в частности, проверить в некоторых случаях, что если $L(M, 0) = 0$, то существует нетривиальное расширение единичного мотива с помощью $M^*(1)$; см. доклад [Ra90] и недавние работы Хардера ([Ha89] и следующие) и Франке.

¹⁾ Это можно также вывести из п. 10.1, разлагая функцию $\zeta(F, s)$ в произведение L -функций Дирихле.

(iv) Помимо результатов о гипотезе Бёрча—Свиннертона-Дайера и о L -функциях Дирихле имеются некоторые частные результаты для других мотивов. Отметим только следующее:

- а) Блох и Като [BK90, § 7] получили результаты о $L(H^1(X), 2)$, где X — эллиптическая кривая с комплексным умножением, определенным над E ; в общих чертах, они показали, что $C_{DV}(M)$ верна в этом случае, и, используя результаты Колывагина и Рубина об «основной гипотезе» в эллиптическом случае, проверили, что $|1 \otimes b|_{l, EP} = 1$, если l является «регулярным для E » [So87].
- б) Нековар [Ne92] получил некоторые результаты «в духе Колывагина», которые подтверждают гипотезу Блоха—Като для $L(M, m)$, где M — мотив ранга 2, ассоциированный с некоторой новой модулярной формой веса $2m$ относительно $\Gamma_0(N)$, имеющих рациональные коэффициенты.

(v) Общим для всех этих результатов является использование идей Колывагина (в частности, систем Эйлера, см. [Pe89, Ko90]) и то, что в них большую роль играет *теория Ивасава*. В каждом случае используется «основная гипотеза» этой теории для соответствующего мотива, и в качестве первого шага нужно понять, как она формулируется. В «обыкновенном» случае это сделано Гринбергом [Gr89] и Шнейдером [Sc89].

Что же касается общего случая, то Като приступил к его изучению в [Ka91]. В общих чертах, речь идет о том, чтобы понять, как ведет себя мотив M над F при конечном абелевом расширении L поля F , что заключается фактически в переносе известного формализма на случай мотивов с коэффициентами в этальной алгебре $E[\text{Gal}(L/F)]$. Параллельно с этим нужно развивать p -адическую теорию, т. е. строить и изучать p -адические L -функции, стараясь понять, что является p -адическим аналогом гипотез Блоха—Като (см. среди прочего [Co89, Co91, CP89, Gre91, Ne92b, Pa91, Pe92, ...]).

С. ДОПОЛНЕНИЯ

11. ГРУППЫ ШАФАРЕВИЧА И ЧИСЛА ТАМАГАВЫ

В первоначальном виде гипотеза Блоха—Като, так же как и гипотеза Бёрча—Свиннертона-Дайера, включала в свою формулировку группы Шафаревича и числа Тамагавы. Для простоты мы ограничимся случаем $E = \mathbb{Q}$.

11.1. Рассмотрим $\mathbb{Z}_l$ -модуль конечного типа T , снабженный линейным и непрерывным действием группы G_F , неразветвленным в до-

полнении к конечному множеству точек, и такой, что представление $V = \mathbf{Q}_l \otimes_{\mathbf{Z}_l} T$ является псевдогеометрическим. Обозначим через $H_f^1(F, T)$ прообраз $H_f^1(F, V)$ в $H^1(F, T)$. Это $\mathbf{Z}_l$ -модуль конечного типа, и можно говорить о его детерминанте $\det_l H_f^1(F, T)$, являющемся $\mathbf{Z}_l$ -решеткой в $\det_l H_f^1(F, V)$; аналогично, $\det_l H^0(F, T)$ — решетка в $\det_l H^0(F, V)$ и $L_f(T) = \det_l H^0(F, T) \otimes \det_l^* H_f^1(F, T)$ — решетка в $L_f(V)$; если $T^+ = \bigoplus_{p \in S_\infty(F)} H^0(G_p, T)$, то $\det_l T^+$ — решетка в $\det_l V^+$; наконец, если $T^*(1)$ — скрученный по Тейту двойственный модуль к T , то $L_f(T^*(1))$ — решетка в $L_f(V^*(1))$. Следовательно, $\Delta'_f(T) = L_f(T) \otimes L_f(T^*(1)) \otimes \det_l^* T^+$ является решеткой в $\Delta'_f(V) = L_f(V) \otimes L_f(V^*(1)) \otimes \det_l^* V^+$. Выберем базис ω'_f в $\Delta'_f(T)$ и базис ω в $\det_l t_V$. Проблема заключается в вычислении $|\omega'_f \otimes \omega|_{l, EP} \in l^{\mathbf{Z}}$ (что априори зависит от выбора T и ω , но не ω'_f).

11.2. Для любой точки p поля F обозначим через $H_f^1(F_p, T)$ прообраз $H_f^1(F_p, V)$ в $H^1(F_p, T)$. Если p — конечная точка, то естественное отображение

$$\mathbf{Q}_l / \mathbf{Z}_l \otimes_{\mathbf{Z}_l} H_f^1(F_p, T) \longrightarrow H^1(F_p, V/T)$$

инъективно, и мы обозначим через $H_{f/f}^1(F_p, V/T)$ его коядро. Если p бесконечна, то положим $H_{f/f}^1(F_p, V/T) = H^1(F_p, V/T)$. Выберем некоторое конечное множество S точек поля F , содержащее все бесконечные точки, все точки, делящие l , и все точки, в которых представление V разветвлено. Блох и Като определяют тогда группу Шафаревича $\text{Ш}(T)$ как первую группу когомологий комплекса

$$\mathbf{Q}_l / \mathbf{Z}_l \otimes_{\mathbf{Z}_l} H_f^1(F, T) \longrightarrow H^1(U_S, V/T) \longrightarrow \prod_{p \in S} H_{f/f}^1(F_p, V/T).$$

Легко видеть, что эта группа конечна и не зависит от S ¹⁾.

11.3. Для любого конечномерного $\mathbf{Q}_l$ -векторного пространства W и любого базиса ω в $\det_l W$ обозначим через μ_ω единственную меру Хаара на W , такую, что если Λ_0 — решетка в W , удовлетворяющая условию $\det_l \Lambda_0 = \mathbf{Z}_l \cdot \omega$, то $\mu_\omega(\Lambda_0) = 1$. Если Λ есть $\mathbf{Z}_l$ -модуль конечного типа с заданным изоморфизмом $\mathbf{Q}_l \otimes_{\mathbf{Z}_l} \Lambda$ на W и если $\det_l \Lambda = \mathbf{Z}_l \cdot a\omega$, где $a \in \mathbf{Q}_l^*$, то $\mu_\omega(\Lambda) = |a|_l$.

¹⁾ В случае когда $T = T_l(A)$ — модуль Тейта абелева многообразия A , можно проверить, что если группа $\text{Ш}(A)(l)$ конечна, то $\text{Ш}(A)(l) = \text{Ш}(T)$.

Пусть p — конечная точка поля F ; условимся, что $t_{V,p} = 0$, если p не делит l . Имеется (см. п. 4.3) точная последовательность

$$0 \rightarrow H^0(F_p, V) \rightarrow \underline{D}_p(V) \rightarrow \underline{D}_p(V) \oplus t_{V,p} \rightarrow H_f^1(F_p, V) \rightarrow 0,$$

позволяющая отождествить $\det_l H_f^1(F_p, V) \otimes \det_l^* H^0(F_p, V)$ с $\det_l t_{V,p}$. Пусть ω_p — базис в $\det_l t_{V,p}$, и ω_1 и ω_0^* — базисы в $\det_l H_f^1(F_p, V)$ и $\det_l^* H^0(F_p, V)$ соответственно, такие, что $\omega_p = \omega_1 \otimes \omega_0^*$; тогда число $\mu_{\omega_l}(H_f^1(F_p, T)) / \mu_{\omega_0}(H^0(F_p, T)) \in l^{\mathbb{Z}}$ зависит только от ω_p , и мы обозначим его через $\text{Tam}_{p, \omega_p}^0(T)$. Если p не делит l , то $\det_l t_{V,p} = \mathbb{Q}_l$, и положим $\text{Tam}_p^0(T) = \text{Tam}_{p,1}^0(T)$. Можно показать, что $\text{Tam}_p^0(T) = 1$ для почти всех конечных точек p , не делящих l . Таким образом, произведение

$$\prod_{p \nmid l} \text{Tam}_p^0(T) \times \prod_{p \mid l} \text{Tam}_{p, \omega_p}^0(T)$$

имеет смысл и зависит лишь от базиса $\omega = \pm \otimes_{p \mid l} \omega_p$ в $\det_l t_V$. Обозначим его через $\text{Tam}_{\omega}^0(T)$.

11.4. Можно доказать следующее

Предложение. В приведенных выше предположениях и обозначениях имеем $\#\text{Ш}(T) = \#\text{Ш}(T^*(1))$ и

$$(|\omega'_f \otimes \omega|_{l,EP})^{-1} = \text{Tam}_{\omega}^0(T) \times \#\text{Ш}(T^*(1)).$$

11.5. Пусть теперь p — конечная точка поля F , такая, что $P_p(V, 1) \neq 0$. Легко видеть, что тогда $H^0(F_p, V) = 0$ и, кроме того,

- если p не делит l , то $H_f^1(F_p, V) = 0$ и, следовательно, $H_f^1(F_p, T) = H^1(F_p, T)_{\text{tor}}$
- если p делит l , то естественное отображение, определенное в п. 4.3 (экспонента Блоха—Като),

$$t_{V,p} \rightarrow H_f^1(F_p, V)$$

является изоморфизмом; в частности, выбор базиса ω_p в $\det_l t_{V,p}$ определяет на $H_f^1(F_p, V)$ меру Хаара μ_{BK,p,ω_p} . Если p не делит l и $\text{Tam}_p(T) := \#H^1(F_p, T)_{\text{tor}}$, то можно проверить, что $\text{Tam}_p^0(T) = |L_p(V, 0)|_l \cdot \text{Tam}_p(T)$. Аналогично, если p делит l и $\text{Tam}_{p, \omega_p} := \mu_{BK,p,\omega_p}(H_f^1(F_p, T))$, то $\text{Tam}_{p, \omega_p}^0(T) = |L_p(V, 0)|_l \cdot \text{Tam}_{p, \omega_p}(T)$.

11.6. Предположим, что $F = \mathbb{Q}$, и приведем оригинальную гипотезу Блоха и Като. Ее формулировка аналогична формулировкам теорем о числах Тамагавы полупростых групп [C188], и нам придется ввести некоторые дополнительные обозначения.

Пары (M, Θ) , состоящие из мотивной структуры M над $\mathbb{Q}$ с коэффициентами в $\mathbb{Q}$ и $\mathbb{Z}$ -модуля конечного типа Θ , снабженного линейным и непрерывным действием группы $G_{\mathbb{Q}}$ на $\hat{\Theta} = \hat{\mathbb{Z}} \otimes \Theta$ и изоморфизмом $\mathbb{Q} \otimes \Theta$ на $M_{B, \infty}$, и удовлетворяющие очевидным условиям согласования, составляют естественным образом абелеву категорию. Обозначим через $H^1(\mathbb{Q}, \Theta) = H^1(\mathbb{Q}, (M, \Theta))$ группу классов расширений $(1_{\mathbb{Q}, \mathbb{Q}}, \mathbb{Z})$ с помощью (M, Θ) в этой категории, и через $H_f^1(\mathbb{Q}, \Theta) \subset H^1(\mathbb{Q}, \Theta)$ прообраз группы $H_f^1(\mathbb{Q}, M)$.

Рассмотрим такую мотивную структуру M , что (а) $W_{-3}M = M$ и (b) $P_p(M, 1) = P_p(M_l, 1)$ для произвольного $l \neq 0$ для любого простого числа¹⁾ p . Выберем модуль Θ так же, как и выше, без кручения. Положим $A_{\Theta}(\mathbb{Q}) = H_f^1(\mathbb{Q}, \Theta)$, $A_{\Theta}(\mathbb{Q}_p) = H_f^1(\mathbb{Q}_p, \hat{\Theta})$ для простого p , $A_{\Theta}(\mathbb{C}) = \mathbb{C} \otimes t_M / \text{Im}(\Theta)$ и $A_{\Theta}(\mathbb{R}) = (A_{\Theta}(\mathbb{C}))^{\text{Gal}(\mathbb{C}/\mathbb{R}) \times \{2, 3\}}$.

Пусть $\hat{\Theta} = \prod \Theta_l$ и p — простое; тогда $A_{\Theta}(\mathbb{Q}_p) = \prod_l H_f^1(\mathbb{Q}_p, \Theta_l)$, и из (b) вытекает, что группа $H_f^1(\mathbb{Q}_p, \Theta_l)$ конечна при $l \neq p$ и тривиальна для почти всех l . Следовательно, группа $A_{\Theta}(\mathbb{Q}_p)$ компактна, если точка p конечна, локально компактна при $p = \infty$ и фактор

$$\left(\prod_{p \in S(\mathbb{Q})} A_{\Theta}(\mathbb{Q}_p) \right) / A_{\Theta}(\mathbb{Q})$$

компактен.

Выберем в $\det_{\mathbb{Q}} t_M$ некоторый базис ω над $\mathbb{Q}$. Его можно рассматривать также как базис в $\det_{\mathbb{Q}_p} t_{M_p}$ для любой точки $p \in S(\mathbb{Q})$. Если $p = \infty$, то ω определяет меру Хаара $\mu_{BK, \infty, \omega}$ на $A_{\Theta}(\mathbb{R})$. Если p конечно, то меру Хаара $\mu_{BK, p, \omega}$ на $H_f^1(\mathbb{Q}_p, M_p) = \mathbb{Q} \otimes_{\mathbb{Z}} A_{\Theta}(\mathbb{Q}_p)$ можно рассматривать как меру на $A_{\Theta}(\mathbb{Q}_p)$. Используя некоторое

¹⁾ Гипотетически (b) следует из (а), поскольку ожидается, что в кольце $\mathbb{C}[t]$ имеем $P_p(M, t) = \prod (1 - \alpha_i t)$, где $|\alpha_i| \leq p^{w/2}$, если $W_{\text{ш}} M = M$; согласно Делиню (гипотезы Вейля), это верно для почти всех p .

²⁾ Блох и Като используют точку зрения Бейлинсона, т.е. они рассматривают $H_f^1(\mathbb{Q}, M)$ как группу, определенную с помощью мотивных когомологий и регуляторов.

³⁾ Эти обозначения подсказывают, что должен быть способ определить пучок абелевых групп A_{Θ} в некоторой топологии. В случае $M = \underline{\text{Real}}[0 \rightarrow A]$, где A — абелево многообразие над $\mathbb{Q}$, и $\Theta = H_1(A(\mathbb{C}), \mathbb{Z})$ можно показать, что $A_{\Theta}(K) = A(K)$ при $K = \mathbb{Q}, \mathbb{Q}_p, \mathbb{C}$ или $\mathbb{R}$.

техническое предположение о p -адических изоморфизмах сравнения (которое на практике всегда выполнено), Блох и Като показали, что

$$\mu_{BK,p,\omega}(A_{\Theta}(\mathbf{Q}_p)) = P_p(M, 1)$$

для почти всех p и что бесконечное произведение

$$\prod_{p \in S_f(\mathbf{Q})} \mu_{BK,p,\omega}(A_{\Theta}(\mathbf{Q}_p))$$

сходится. Это позволяет ввести меру-произведение $\mu_{BK} = \prod \mu_{BK,p,\omega}$ на локально компактной группе $\prod_{p \in S(\mathbf{Q})} A_{\Theta}(\mathbf{Q}_p)$, которая, согласно формуле произведения, не зависит от ω и которую Блох и Като называли *мерой Тамагавы*. Они положили

$$\text{Tam}(\Theta) = \mu_{BK} \left(\left(\prod_{p \in S(\mathbf{Q})} A_{\Theta}(\mathbf{Q}_p) \right) / A_{\Theta}(\mathbf{Q}) \right)$$

(это также равно $\mu_{BK,\infty,\omega}(A_{\Theta}(\mathbf{R})/A_{\Theta}(\mathbf{Q})) \cdot \prod_{p \neq \infty} \mu_{BK,p,\omega}(A_{\Theta}(\mathbf{Q}_p))$) и *предположили*, что конечная группа $\text{Ш}(\Theta_l)$ равна нулю для почти всех l и что если положить $\text{Ш}(\Theta) = \prod_{l \text{ простые}} \text{Ш}(\Theta_l)$, то

$$\text{Tam}(\Theta) = \#H^0(\mathbf{Q}, M^*(1)/\Theta^*(1)) / \#\text{Ш}(\Theta).$$

Для $p \in S(\mathbf{Q})$ положим $\mu_{BK,p,\omega}^0 = \mu_{BK,p,\omega}$, если $p = \infty$, и $\mu_{BK,p,\omega}^0 = |P_p(M, 1)|^{-1} \cdot \mu_{BK,p,\omega}$ в противном случае (таким образом, это мера, рассматривавшаяся в п. 11.3). Обозначим через μ_{BK}^0 соответствующую меру-произведение на $\prod A_{\Theta}(\mathbf{Q}_p)$ и положим

$$\text{Tam}^0(M) = \mu_{BK}^0 \left(\left(\prod_{p \in S(\mathbf{Q})} A_{\Theta}(\mathbf{Q}_p) \right) / A_{\Theta}(\mathbf{Q}) \right).$$

Гипотеза Блоха и Като эквивалентна следующему утверждению:

$$C_{\text{Tam}}(M) : \quad L(M, 0) = \text{Tam}^0(\Theta) \cdot \#\text{Ш}(\Theta) / \#H^0(\mathbf{Q}, M^*(1)/\Theta^*(1)).$$

Из предложения 11.4 нетрудно вывести, что эта гипотеза эквивалентна $C_{BK}(M)$.

11.7. Замечание. Похоже на то, что гипотезу $C_{BK}(M)$ можно переформулировать в духе $C_{\text{Tam}}(M)$ для любой мотивной структуры M . Но тогда нужно работать с мерой μ_{BK}^0 (в дополнении к области сходимости мера μ_{BK} не определена) и заменить «группу» A_{Θ} на некоторый комплекс.

12. λ -АДИЧЕСКИЕ ПРЕДСТАВЛЕНИЯ И ПОЛУПРОСТЫЕ МОТИВЫ, ФУНКЦИОНАЛЬНЫЕ УРАВНЕНИЯ, НЕПОЛНЫЕ L -ФУНКЦИИ¹⁾

12.1. Положим для любого $s \in \mathbb{C}$ (см. [Se70]) $\Gamma_{\mathbf{R}}(s) = \pi^{-s/2} \Gamma(s/2)$. Пусть $K \in \{\mathbf{R}, \mathbf{C}\}$ и V — смешанная структура Ходжа над K с вещественными коэффициентами. Определим на V убывающую фильтрацию $\mathbf{R}$ -векторными подпространствами, полагая для любого $r \in \mathbb{Z}$

$\gamma^r V =$ наибольшее подпространство в V , такое, что $(\gamma^r V)_{\mathbf{C}} \subset \text{Fil}^r V_{\mathbf{C}}$.

В случае $K = \mathbf{R}$ заметим, что подпространства $\gamma^r(V)$ инвариантны относительно группы $\text{Gal}(\mathbf{C}/\mathbf{R})$, которая, таким образом, действует на $gr_{\gamma}^r V$; положим тогда $n_r^+ = \dim_{\mathbf{R}}(gr_{\gamma}^r V)^+$ и $n_r^- = \dim_{\mathbf{R}}(gr_{\gamma}^r V)^-$. Если же $K = \mathbf{C}$, то положим $n_r^+ = n_r^- = \dim_{\mathbf{R}} gr_{\gamma}^r V$. В любом из этих двух случаев положим²⁾

$$L(V, s) = \prod_{r \in \mathbb{Z}} \Gamma_{\mathbf{R}}(s + \varepsilon_r - r)^{n_r^+} \cdot \Gamma_{\mathbf{R}}(s + 1 - \varepsilon_r - r)^{n_r^-},$$

где $\varepsilon_r \in \{0, 1\}$ определяется из соотношения $\varepsilon_r \equiv r \pmod{2}$.

Аналогичным образом (см. [FP92]) можно определить также «локальную константу» $\varepsilon(V, \psi_0, \mu_0)$.

Заметим, что $(\gamma^0 V)^+$ отождествляется с $\text{Ker } \alpha_V$ и что функция $L(V, s)$ имеет в точке $s = 0$ полюс порядка $\dim_{\mathbf{R}} \text{Ker } \alpha_V$.

12.2. Замечание. «Хорошие свойства» локальных архимедовых факторов не меняются при замене $\Gamma_{\mathbf{R}}(s)$ на $a \cdot \Gamma_{\mathbf{R}}(s)$. Недавние работы Денингера [Den90, Den91] подсказывают, что нужно взять $a = 1/\sqrt{2}$.

12.3. Пусть теперь $\mathfrak{p}$ — конечная точка F и $B_{st, \mathfrak{p}}$ есть $B_{\text{cris}, \mathfrak{p}}$ -алгебра B_{st} , соответствующая расширению $\overline{F}_{\mathfrak{p}}/F_{\mathfrak{p}}$ (см., например, [P90, п. 1.2.3]). Пусть l — произвольное простое число; тогда говорят, что l -адическое представление V группы $G_{\mathfrak{p}}$ является *полустабильным*

- (i) в случае, когда $\mathfrak{p}$ не делит l , если действие группы инерции $I_{\mathfrak{p}}$ унипотентно;

¹⁾ См. [FP92]. Для простоты мы будем предполагать, что $E = \mathbf{Q}$, но обобщение не представляет труда.

²⁾ Легко проверить, что в случае чистой структуры Ходжа получается обычное определение [Se70].

(ii) в случае, когда p делит l , если $\dim_{(F_p)_0}(B_{st} \otimes_{Q_l} V)^{G_p} = \dim_{Q_l} V$.

Говорят, что представление V является *потенциально полустабильным*, или pst -представлением, если существует такое конечное расширение F'_p поля F_p , что V является полустабильным как представление группы $\text{Gal}(\overline{F_p}/F'_p)$. Если p не делит l , то это всегда так (это утверждение теоремы Гротендика об l -адической монодромии, [ST68, Appendix]). Если же p делит l , то это уже не всегда верно¹⁾.

С любым l -адическим pst -представлением V группы G_p можно связать некоторое линейное представление группы W'_p , группы Вейля—Делиня поля F_p , с коэффициентами в Q_l , если p не делит l (см. [De73; Ta79, § 4])²⁾, и в $(\overline{F_p})_0$, если p делит l (см. [Bures]). Это приводит к тому, что V можно сопоставить не только множитель $L_p(V, s) = P_p(V, N(p)^{-s})^{-1}$, определенный, как в п. 3.3, но также кондуктор $a_p(V)$ и «локальную константу» $\varepsilon_p(V, \psi, dx, s)$ [Ta79, § 4].

12.4. Говорят, что l -адическое представление G_F является *геометрическим*, если оно является pst -представлением во всех конечных точках и неразветвлено в дополнении к их конечному числу.

«Гипотеза» (см. [FM92]). Пусть l — простое число. Функтор $M \mapsto M_l$ индуцирует тензорную эквивалентность между Q_l -линейной категорией, получающейся из категории $\underline{SM}_F(Q)$ расширением скаляров $Q \rightarrow Q_l$, и полной подкатегорией категории l -адических представлений группы G_F , объектами которой являются геометрические представления.

В частности, если M — объект категории $\underline{SM}_F(Q)$, то для любого простого числа l его l -адическая реализация является геометрической³⁾. Если это так и если p — точка поля F , лежащая над простым числом p , то имеется занумерованное простыми числами l семейство представлений группы W'_p , о которых предполагается,

1) Любое потенциально полустабильное представление является представлением типа де Рама; я не знаю, верно ли обратное утверждение, представляющее собой что-то вроде теоремы о p -адической монодромии.

2) Конструкция этого представления зависит от выбора некоторых данных, но его класс изоморфизма не зависит.

3) Таким образом, «хорошими» являются *полустабильные* мотивы, т. е. такие M , что для любых l и p представление M_l группы G_p является полустабильным. Должна иметь место *теорема о полустабильной редукции*, утверждающая, что каждая мотивная структура над F становится полустабильной после конечного расширения скаляров. Для 1-мотивов это утверждение без труда выводится из теоремы о полустабильной редукции для абелевых многообразий.

что они определены над $\mathbf{Q}$ и совместимы [De73, п° 8.7]. Это позволяет определить $a_p(M)$ и $\varepsilon_p(M, \psi, dx)$ как $a_p(M_l)$ и $\varepsilon_p(M_l, \psi, dx)$ для любого выбора l , включая $l = p$. Если $p \in S_\infty(F)$, то обозначим через $\varepsilon_p(M, \psi_0, \mu_0, s)$ константу $\varepsilon_p(M, \psi_0, \mu_0)$, где M_p — смешанная структура Ходжа над F , ассоциированная с M . Пусть ψ — нетривиальный аддитивный характер группы $\mathbf{A}_F/F$ и $dx = \otimes dx_p$ — мера Тамагавы на $\mathbf{A}_F$; тогда можно определить множитель $\varepsilon(M, s)$ обычной формулой

$$\varepsilon(M, s) = \prod_p \varepsilon_p(M, \psi_p, dx_p, s).$$

В этой общей ситуации, если $\Lambda(M, s) = \prod_{p \in S(F)} L_p(M, s)$ — «полная» L -функция объекта M , то функциональное уравнение записывается в следующем виде:

«Гипотеза» $C_{EF}(M)$. Функция $\Lambda(M, s)$ допускает мероморфное продолжение на $\mathbf{C}$ и удовлетворяет уравнению $\Lambda(M, s) = \varepsilon(M, s) \cdot \Lambda(M^*(1), -s)$.

12.5. По-видимому, можно проверить, что если

$$0 \rightarrow M' \rightarrow M \rightarrow M'' \rightarrow 0$$

— короткая точная последовательность в $\underline{SM}_F(\mathbf{Q})$ и если выполняются две из трех гипотез $C_{EF}(M')$, $C_{EF}(M)$, $C_{EF}(M'')$, то так же обстоит дело и с третьей.

Если $C_{EF}(M)$ выполнена, то можно проверить, что гипотезы $C_r(M)$ и $C_r(M^*(1))$ эквивалентны, так же как и гипотезы $C_{DV}(M)$ и $C_{DV}(M^*(1))$. С гипотезой C_{BK} ситуация более деликатна: Б. Перрин-Риу анонсировала [FR92] некоторую гипотетическую формулу $C_{EP,p}(V)$, элементарную по сути, но слишком длинную для того, чтобы ее здесь привести, для вычисления «характеристики Эйлера—Пуанкаре» p -адического pst -представления V группы G_p (где p делит p). Можно доказать, что если мотив $\det(M)$ является скрученным по Тейту мотивом Артина, $C_{EP,p}(M)$ верна для любого p и верна $C_{EF}(M)$, то гипотезы $C_{BK}(M)$ и $C_{BK}(M^*(1))$ эквивалентны¹⁾.

¹⁾ Интересно отметить, что в доказательстве Блоха и Като их гипотезы для значений дзета-функции Римана, которую они формулировали лишь для значений в положительных целых числах, им приходится сначала проводить доказательство для целых отрицательных чисел, затем проверять $C_{EP,p}(\mathbf{Q}_p(i))$ и использовать функциональное уравнение.

12.6. Пусть S — конечное множество точек F , содержащее все бесконечные точки. Тогда можно сформулировать гипотезы $C_{r,S}(M)$, $C_{DB,S}(M)$ и $C_{BK,S}(M)$ о поведении функции $L_S(M, s)$ в точке $s = 0$ с помощью формализма, совершенно аналогичного тому, который был развит для $L(M, s)$ и который мы не будем здесь развивать за недостатком места. Скажем только, что группы $H_f^1(F, M)$ и $H_f^1(F, M^*(1))$ следует заменить на $H_{f,S}^1(F, M)$ и $H_{f,S}^1(F, M^*(1))$, где

$$H_{f,S}^1(F, M) = \{x \in H^1(F, M) \mid x_{l,p} \in H_f^1(F_p, M_l)\}$$

для любого простого l и $p \notin S$,

что ожидаемый порядок нуля функции $L_S(M, s)$ в точке $s = 0$ равен $\dim_{\mathbf{Q}} H_{f,S}^1(F, M^*(1)) - \dim_{\mathbf{Q}} H^0(F, M^*(1))$ и что естественное отображение из $H_{f,S}^1(F, M)$ в $H^1(F, M_l)$ должно индуцировать изоморфизм $\mathbf{Q}_l \otimes_{\mathbf{Q}} H_{f,S}^1(M)$ на

$$H_{f,S}^1(F, M_l) = \{x \in H^1(F, M_l) \mid x_p \in \begin{cases} H_f^1(F_p, M_l), & \text{если } p \notin S, \\ H_g^1(F_p, M_l), & \text{если } p \in S, \end{cases}\}$$

где $H_g^1(F_p, M_l)$ обозначает подгруппу в $H^1(F_p, M_l)$, классифицирующую pst -расширения $\mathbf{Q}_l$ с помощью M_l .

Интересно отметить, что единственный разумный способ сделать все эти гипотезы совместимыми друг с другом заключается во введении следующих гипотез:

«Гипотеза». Категория $\underline{SM}_F(\mathbf{Q})$ имеет когомологическую размерность 1.

Гипотеза Для любого простого числа l категория геометрических l -адических представлений группы G_F имеет когомологическую размерность 1. Кроме того, для любого геометрического представления V и любой конечной точки p поля F представление группы W'_p , ассоциированное с V , является F -полупростым (в смысле [De73, déf. 8.6]).

По-видимому, можно проверить, что по модулю этих гипотез $C_{BK}(M)$ и $C_{BK,S}(M)$ эквивалентны. Кроме того, легко видеть, что если

$$0 \rightarrow M' \rightarrow M \rightarrow M'' \rightarrow 0$$

— короткая точная последовательность в $\underline{SM}_F(\mathbf{Q})$, имеющая «хорошую редукцию в дополнении к S » (в этом случае $L_S(M, s) =$

$L_S(M', s) \cdot L_S(M'', s)$), и если верны две из трех гипотез $C_{BK, S}(M)$, $C_{BK, S}(M')$, $C_{BK, S}(M'')$, то верна и третья. Из этих двух «результатов» следует, что гипотеза $C_{BK}(M)$ хорошо ведет себя по отношению к коротким точным последовательностям.

12.7. Другой причиной интереса к неполным L -функциям служит то, что для «известных» (в терминологии п. 6.2) мотивов M не всегда известно, что M допускает функцию $L(M, s)$, но, как правило, можно доказать, что он допускает функцию $L_S(M, s)$, если S достаточно большое.

Доказательство «гипотезы» C_{BK} должно состоять из двух частей:

- a) существование хорошего формализма;
- b) доказательство гипотезы $C_{BK, S}(M)$ для мотива M , имеющего «хорошую редукцию в дополнении к S » (о котором можно также предполагать, что $H_{f, S}^1(F, M) = H_{f, S}^1(F, M^*(1)) = 0$).

Но, может быть, не стоит мечтать о слишком многом!

ЛИТЕРАТУРА

- [BN78] Bayer P., Neukirch J. On values of zeta functions and l -adic Euler characteristics, *Invent. Math.*, **50** (1978), 35–64.
- [Be85] Бейлинсон А. А. Высшие регуляторы и значения L -функций. — В кн.: *Современные проблемы математики. Новейшие достижения*, т. 24, 1984, с. 181–238.
- [Be87] Beilinson A. Height pairing between algebraic cycles, in: *Contemp. Math.*, vol. 67, 1987, 1–24.
- [Be90] Beilinson A. Polylogarithms and cyclotomic elements, *Preprint*, 1990.
- [BGSV90] Beilinson A., Goncharov A., Schechtman V., Varchenko A. Aomoto dilogarithms, mixed Hodge structures and motivic cohomology of pairs of triangles on the plane, in: *The Grothendieck Festschrift*, vol. I, *Prog. in Math.*, vol. 86, Birkhäuser, Boston, 1990, 135–172.
- [BMS87] Beilinson A., MacPherson R., Schechtman V. Notes on motivic cohomology, *Duke Math. J.*, **54** (1987), 679–710.
- [BS63] Birch B., Swinnerton-Dyer H. Notes on elliptic curves I, *J. Reine Angew. Math.*, **212** (1963), 7–25.
- [BS65] Birch B., Swinnerton-Dyer H. Notes on elliptic curves II, *J. Reine Angew. Math.*, **218** (1965), 79–108.
- [Bl80] Bloch S. A note on height pairings, Tamagawa numbers and Swinnerton-Dyer conjecture, *Invent. Math.*, **58** (1980), 65–76.

- [Bl84] Bloch S. Algebraic cycles and values of L -functions, J. Reine Angew. Math., **350** (1984), 94–108.
- [Bl84b] Bloch S. Height pairings for algebraic cycles, J. Pure Applied Algebra, **34** (1984), 119–145.
- [Bl86] Bloch S. Algebraic cycles and the Beilinson conjectures, in: Contemp. Math., vol. 58, 1986, 65–79.
- [BK90] Bloch S., Kato K. L functions and Tamagawa numbers of motives, in: The Grothendieck Festschrift, vol. I, Progr. in Math., Birkhäuser, Boston, 1990, 333–400.
- [CL88] Clozel L. Nombres de Tamagawa des groupes semi-simples (d'après Kottwitz), Séminaire Bourbaki, exp. 702, novembre 1988. [Имеется перевод: Клозель Л. Числа Тамагавы полупростых групп. — в кн.: Труды семинара Бурбаки за 1989 г. — М.: Мир, 1991, с. 57–78.]
- [Co89] Coates J. On p -adic L -functions, Séminaire Bourbaki, exp. 701, novembre 1988. [Имеется перевод: Коутс Дж. p -адические L -функции. — В кн.: Труды семинара Бурбаки за 1989 г. — М.: Мир, 1991, с. 32–56.]
- [Co91] Coates J. Motivic p -adic L -functions, in: L -functions and Arithmetic, Proc. of the Durham Symposium, London Math. Soc. L.N.S., vol. 153, Cambridge University Press, 1991, 141–172.
- [CP89] Coates J., Perrin-Riou B. On p -adic L -functions attached to motives over $\mathbb{Q}$, in: Adv. Stud. Pure Math., vol. 17, 1989, 23–54.
- [De71] Deligne P. Théorie de Hodge II, Publ. Math. I.H.E.S., **40** (1971), 5–57. [Имеется перевод: Делинь П. Теория Ходжа II. Математика, т. 17, 1973, вып. 5, с. 3–56.]
- [De73] Deligne P. Les constantes des équations fonctionnelles des fonctions L , in: Modular Functions of One Variable, II, Lect. Notes in Math., vol. 349, Springer-Verlag, 1973, 501–595.
- [De74] Deligne P. Théorie de Hodge III, Publ. Math. I.H.E.S., **44** (1974), 5–77.
- [De79] Deligne P. Valeurs de fonctions L et périodes d'intégrales, in: Proc. Symp. Pure Math., vol. 33, 1979, 313–346. [Имеется перевод: Делинь П. Значения L -функций и периоды интегралов. — В кн.: Автоморфные формы, представления и L -функции. — М.: Мир, 1984, с. 188–248.]
- [De83] Deligne P. Preuve des conjectures de Tate et Shafarevitch (d'après G. Faltings), Séminaire Bourbaki, exp. 616, novembre 1983. [Имеется перевод: Делинь П. Доказательство гипотез Тейта и Шфаревича. — В кн.: Алгебра и теория чисел. — М.: Мир, 1987, с. 100–124.]
- [De85] Deligne P. Lettre à C.Soulé, 1985.
- [De89] Deligne P. Le groupe fondamental de la droite projective moins trois points, in: Galois Groups Over $\mathbb{Q}$, MSRI publications, vol. 16, Springer-Verlag, 1989, 79–297.

- [DM82] Deligne P., Milne J. Tannakian categories, in: Hodge Cycles, Motives and Shimura Varieties, Lect. Notes in Math., vol. 900, Springer-Verlag, 1982, 101–228. [Имеется перевод: Делинь П., Милн Дж. С. Категории Таннаки. — В кн.: Ходжевы циклы и мотивы. — М.: Мир, 1988, с. 94–201.]
- [Den91] Deninger C. On the Γ -factors attached to motives, Invent. Math., **104** (1991), 245–261.
- [Den92] Deninger C. Local L -factors of motives and regularized determinants, Invent. Math., **107** (1992), 137–150.
- [DS91] Deninger C., Scholl A. The Beilinson conjectures, in: L -functions and Arithmetic, Proc. of Durham Symposium, London Math. Soc. L.N.S., vol. 53, Cambridge University Press, 1991, 173–209.
- [Fa89] Faltings G. Crystalline cohomology and p -adic Galois representations, in: Algebraic Analysis, Geometry and Number Theory, Johns Hopkins University Press, Baltimore, 1989, 25–80.
- [Fa90] Faltings G. F -isocrystals on open varieties: results and conjectures, in: The Grothendieck Festschrift, II, Prog. in Math., vol. 87, Birkhäuser, Boston, 1990, 219–248.
- [FI90] Fontaine J.-M., Illusie L. p -adic periods: a survey, prépublication d'Orsey, 1990; см. также: Proc. of the Indo-French conf. on geometry (Bombay, 1989), Hindustan Book Agency, Delhi, 1993.
- [FM92] Fontaine J.-M., Mazur B., Représentations l -adiques géométriques, in: [Bures].
- [FP91] Fontaine J.-M., Perrin-Riou B. Autour des conjectures de Bloch et Kato: I: Cohomologie galoisienne, II: Structures motiviques f -closes, III: Le cas général, C. R. Acad. Sci. Paris, séries I, **313** (1991), 189–196, 349–356, 421–428.
- [FP92] Fontaine J.-M., Perrin-Riou B. Autour des conjectures de Bloch et Kato: cohomologie galoisienne et valeurs de fonctions L , in: Proc. Symp. Pure Math., vol. 55, part I, 1994, 599–706.
- [Gr90] Gros M. Régulateurs syntomiques et valeurs de fonctions L p -adiques (с приложением: Kurihara M. Computation of the syntomic regulator in the cyclotomic case), Invent. Math., **99** (1990), 293–320.
- [Gr91] Gross B. Kolyvagin's work for modular elliptic curves, in: L -functions and Arithmetic, London Math. Soc. L.N.S., vol. 153, Cambridge University Press, 1991, 235–256.
- [Gre89] Greenberg R. Iwasawa theory for p -adic representations, in: Adv. Stud. Pure Math., vol. 17, 1989, 97–137.
- [Gre91] Greenberg R. Iwasawa theory for motives, in: L -functions and Arithmetic, Proc. of the Durham Symposium, London Math. Soc. L.N.S., vol. 153, 1991, 373–392.
- [Ha89] Harder G., Arithmetische Eigenschaften von Eisenstein Klassen die modulare Konstruktion von gemischten Motiven und von Erweiterungen endlicher Galoismoduln, Preprint, 1989. См. также: Harder G. Eisensteinkohomologie und die Konstruktion

gemischter Motive. Lect. Notes in Math., vol. 1562, Springer-Verlag, 1993.

- [Il90] Illusie L. Cohomologie de de Rham et cohomologie étale p -adique, Séminaire Bourbaki, exp. 726, juin 1990. [Имеется перевод: Иллюзи Л. Когомологии Рама и p -адические когомологии. — В кн.: Труды семинара Н. Бурбаки за 1990 г. — М.: Мир, 1996, с. 243–282.]
- [Ja88] Jannsen U. Deligne homology, Hodge- D -conjecture, and motives, in [RSS88], 305–372.
- [Ja89] Jannsen U. On the l -adic cohomology of varieties over number fields and its Galois cohomology, in: Galois groups over $\mathbf{Q}$, MSRI publications, vol. 16, Springer-Verlag, 1989.
- [Ja90] Jannsen U. Mixed Motives and Algebraic K -Theory, Lect. Notes in Math., vol. 1400, Springer-Verlag, Berlin, 1990.
- [Ka91] Kato K. Iwasawa Theory and p -Adic Hodge Theory, Preprint, 1991; см. также Lect. Notes in Math., vol. 1535, Springer-Verlag, 1993.
- [Ko90] Kolyvagin V. A. Euler systems, in: The Grothendieck Festschrift, II, Prog. in Math., vol. 87, Birkhäuser, Boston, 1990, 435–483.
- [Li72] Lichtenbaum S. On the values of zeta and L -functions, I, Ann. of Math., 96 (1972), 338–360.
- [Li84] Lichtenbaum S. Values of zeta functions at non-negative integers, in: Number theory, Noordwijkerhout, Lect. Notes in Math., vol. 1068, Springer-Verlag, Berlin, 1984, 127–138.
- [Li87] Lichtenbaum S. The construction of weight-two arithmetic cohomology, Invent. Math., 88 (1987), 183–215.
- [Mi86] Milne J.S. Arithmetic Duality Theorems, Perspectives in Mathematics, vol. 1, Academic Press, 1986.
- [MW84] Mazur B., Wiles A., Class fields of abelian extensions of $\mathbf{Q}$, Invent. Math. 76 (1984), 179–330.
- [Ne92] Nekovar J. Kolyvagin's method for Chow groups of Kuga-Sato varieties, Invent. Math., 107 (1992), 99–125.
- [Ne92b] Nekovar J. On p -adic height pairing, Séminaire de Théorie des Nombres de Paris. 1992.
- [Pa91] Panichshkin A. Non-Archimedean L -Functions of Siegel and Hilbert Modular Forms, Lect. Notes in Math., vol. 1471, Springer-Verlag, 1991.
- [Pe89] Perrin-Riou B. Travaux de Kolyvagin et Rubin, Séminaire Bourbaki, exp. 717, novembre 1989. [Имеется перевод: Перрен-Риу Б. Работы Колывагина и Рубина. — В кн.: Труды семинара Н. Бурбаки за 1990 г. — М.: Мир, 1996, с. 50–77.]
- [Pe92] Perrin-Riou B. Théorie d'Iwasawa et hauteur p -adique, Invent. Math., 109 (1992), 137–185.
- [Ra89] Ramakrishnan D. Regulators, algebraic cycles, and values of L -functions, in: Algebraic K -theory and algebraic number theory, Contemp. Math., vol. 83, 1989, 183–310.

- [Ra90] Ramakrishnan D. Problems arising from Tate and Beilinson conjectures in the context of Shimura varieties, in: Automorphic forms, Shimura varieties and L -functions, vol. II, Perspectives in Mathematics, vol. 11, Academic Press, 1990, 227–252.
- [RSS88] Rapoport M., Schappacher N., Schneider P. (eds.) Beilinson's Conjectures on Special Values of L -Functions, Perspectives in Mathematics, vol. 4, Academic Press, 1988.
- [Ru91] Rubin K. The «main conjectures» of Iwasawa theory for imaginary quadratic fields, Invent. Math., 103 (1991), 25–68.
- [Sc88] Schneider P. Introduction to the Beilinson conjectures, in [RSS88], 1–35.
- [Sc89] Schneider P. Motivic Iwasawa theory, in: Adv. Stud. Pure Math., vol. 17, 1989, 421–456.
- [Sc90] Scholl A.J. Motives for modular forms, Invent. Math., 100 (1990), 419–430.
- [Sc91] Scholl A.J. Remarks on special values of L -functions, in: L -Functions and Arithmetic, Proc. of Durham Symposium, London Math. Soc. L.N.S., vol. 153, 1991, 373–392.
- [Se70] Serre J.-P. Facteurs locaux¹ des fonctions zéta des variétés algébriques (définitions et conjectures), Collected papers, vol. 2, Springer-Verlag, Berlin, 1986, 581–592. [Имеется перевод: Серр Ж.-П. Локальные множители дзета-функций алгебраических многообразий (определения и гипотезы). — Математика, т. 15, 1971, вып. 1, с. 3–13.]
- [ST68] Serre J.-P., Tate J. Good reduction of abelian varieties, Ann. of Math., 88 (1968), 492–517 [Имеется перевод: Серр Ж.-П., Тейт Дж. Хорошая редукция абелевых многообразий. — Математика, т. 15, 1971, вып. 5, с. 140–165.]
- [So79] Soulé C. K -théorie des anneaux d'entiers de corps de nombres et cohomologie étale, Invent. Math., 55 (1979), 251–295.
- [So81] Soulé C. On higher p -adic regulators, in: Algebraic K -theory, Lect. Notes in Math., vol. 854, Springer-Verlag, Berlin, 1981, 371–401.
- [So83] Soulé C. K -théorie et zéros aux points entiers de fonctions zéta, Proc. ICM Warszawa 1983, PWN, Warsaw, 1984, 437–445.
- [So85] Soulé C. Régulateurs, Séminaire Bourbaki, exp. 644, février 1985.
- [So87] Soulé C. p -adic K -theory of elliptic curves, Duke Math. J., 54 (1987), 249–269.
- [Ta65] Tate J. Algebraic cycles and poles of zeta-functions, in: Arithmetical algebraic geometry, Harper and Row, New York, 1965. [Имеется перевод: Дж. Тейт. Алгебраические циклы ко-гомологий. — УМН, т. 20, 1965, с. 27–40.]
- [Ta66] Tate J. On the conjectures of Birch and Swinnerton-Dyer and a geometric analog, Séminaire Bourbaki, exp. 306, février 1966. [Имеется перевод: Дж. Тейт. О гипотезе Бёрча—Свиннerton-Дайера и геометрическом аналоге. — Математика, т. 12, 1968, вып. 6; с. 41–55.]

- [Ta79] Tate J. Number theoretic background, in: Proc. Symp. Pure Math., vol. 33, 1979, part 2, 3–26. [Имеется перевод: Тейт Дж. Теоретико-числовое введение. — В кн.: Автоморфные формы, представления и L -функции. — М.: Мир, 1984, с. 71–112.]
- [Ta84] Tate J. Les conjectures de Stark sur les fonctions L d'Artin en $s = 0$, Birkhäuser, Boston, 1984.
- [Bures] Séminaire sur les périodes p -adiques, Asterisque, **223** (1993).
- [SGA4] Théorie des topos et cohomologie étale des schémas, dirigé par M. Artin, A. Grothendieck et J.-L. Verdier, Lect. Notes in Math., vol. 269, 270, 305, Springer-Verlag, Berlin, 1972–73.
- [SGA4.5] Cohomologie étale, dirigé par P. Deligne, Lect. Notes in Math., vol. 569, Springer-Verlag, Berlin, 1977.
- [SGA5] Cohomologie l -adiques et fonctions L , dirigé par A. Grothendieck, Lect. Notes in Math., vol. 589, Springer-Verlag, Berlin-New York, 1977.

JEAN-MARC FONTAINE

Université de Paris-Sud

Arithmétique et Géométrie Algébrique

C.N.R.S. URA D0752

Bâtiment 425

F-91405 ORSAY CEDEX

ВЫРЕЗАНИЕ В АЛГЕБРАИЧЕСКОЙ K -ТЕОРИИ

(по А. Суслину и М. Водзицки)

Жан-Луи Лоде

Пусть A — произвольное кольцо. Тогда группы $K_n(A)$, $n \geq 0$, алгебраической K -теории — абелевы группы, определенные следующим образом. При $n = 0$ $K_0(A)$ является группой Гротендика проективных модулей конечного типа над A . Для $n \geq 1$ $K_n(A) := \pi_n(BGL(A)^+)$, где $GL(A)$ — объединение линейных групп $GL_n(A)$ (обратимых $n \times n$ -матриц) и $BGL(A)$ — классифицирующее пространство дискретной группы $GL(A)$. Плюс-конструкция Квиллена дает пространство $BGL(A)^+$, которое является H -пространством с теми же гомологиями, что и $BGL(A)$. В малых размерностях имеем

$$K_1(A) = H_1(GL(A)) = GL(A)/[GL(A), GL(A)] = GL(A)/E(A),$$

где $E(A)$ — подгруппа в $GL(A)$, порожденная элементарными матрицами. При $n = 2$ имеем $K_2(A) = H_2(E(A))$.

Для вычислений важна возможность сравнивать $K_n(A)$ и $K_n(A/I)$, где I — двусторонний идеал. Определим абстрактно относительные группы K -теории $K_n(A, I)$ (см. п. 1.3) как группы, получающиеся из следующей точной последовательности:

$$\begin{aligned} \cdots \rightarrow K_{n+1}(A/I) \rightarrow K_n(A, I) \rightarrow K_n(A) \\ \rightarrow K_n(A/I) \rightarrow \cdots \rightarrow K_0(A/I). \quad (0) \end{aligned}$$

Тогда возникает естественный вопрос: зависят ли группы $K_n(A, I)$ только от I или они зависят от объемлющего кольца A ? На самом деле для всех псевдоколец (= колец без единицы) I можно определить группы $K_n(I)$ следующим образом.

Закон $(n, u)(m, v) = (nm, nv + mu + uv)$ превращает $\mathbb{Z} \ltimes I$ в кольцо с единицей (единица = $(1, 0)$), которое содержит I в качестве двустороннего идеала. Положим

$$K_n(I) := \text{Ker}(K_n(\mathbb{Z} \ltimes I) \rightarrow K_n(\mathbb{Z})).$$

Заметим, что тогда $K_n(I) = K_n((\mathbb{Z} \ltimes I), I)$. Естественный гомоморфизм колец $\mathbb{Z} \ltimes I \rightarrow A$, тождественный на I , позволяет переформулировать предыдущий вопрос следующим образом:

• является ли естественный гомоморфизм $K_n(I) \rightarrow K_n(A, I)$ изоморфизмом для всех n и всех колец A ?

Если ответ на этот вопрос положителен, говорят, что I является K -вырезаемым.

Х. Басс [В] показал, что ответ положителен для $n = 0$. Однако это неверно для $n > 0$ (см. обсуждение в разд. 6). В [S-W] Суслин и Водзickи дают простое необходимое и достаточное условие для вырезаемости кольца I в рациональной алгебраической K -теории¹⁾.

Определение [W3]. Говорят, что $\mathbb{Q}$ -алгебра I (не обязательно с единицей) H -унитарна, если комплекс

$$B_*(I) : \dots \rightarrow I^{\otimes n+1} \xrightarrow{b'} I^{\otimes n} \rightarrow \dots \rightarrow I \otimes_{\mathbb{Q}} I \xrightarrow{b'} I \rightarrow 0,$$

где $\otimes = \otimes_{\mathbb{Q}}$ и $b'(a_0 \otimes \dots \otimes a_n) = \sum_{i=0}^{n-1} (-1)^i (a_0 \otimes \dots \otimes a_i a_{i+1} \otimes \dots \otimes a_n)$, ациклический (т. е. $H_n(B_*(I)) = 0$, $n \geq 0$).

Заметим, что тогда $I^2 = I$. Все $\mathbb{Q}$ -алгебры с единицей H -унитарны, так как $s : I^{\otimes n-1} \rightarrow I^{\otimes n}$, $s(a_1 \otimes \dots \otimes a_{n-1}) = (1 \otimes a_1 \otimes \dots \otimes a_{n-1})$, удовлетворяет тождеству $sb' + b's = \text{Id}$.

Предположим, что для всех конечных наборов элементов $a_0, \dots, a_n$ из I существует элемент $u \in I$, такой, что $ua_i = a_i u = a_i$ для всех i . Тогда алгебра I , называемая в этом случае *локально унитарной*, является H -унитарной. Другие примеры можно найти в [W2].

Для любой абелевой группы M через $M_{\mathbb{Q}}$ мы будем обозначать локализацию $M \otimes_{\mathbb{Z}} \mathbb{Q}$.

Теорема 1. Для любого псевдокольца I следующие свойства эквивалентны:

- (a) I $K_{\mathbb{Q}}$ -вырезаемо,
- (b) $I_{\mathbb{Q}}$ H -унитарна.

Отсюда легко вывести следующий результат:

Теорема 2. Для любой $\mathbb{Q}$ -алгебры I (не обязательно с единицей) следующие свойства эквивалентны:

- (a) I $K_{\mathbb{Q}}$ -вырезаема,
- (b) $I_{\mathbb{Q}}$ H -унитарна.

А. Суслин известил меня, что в настоящее время он записывает доказательство следующего результата для $\Lambda = \mathbb{Z}$, $\mathbb{Z}/n\mathbb{Z}$ или $\mathbb{Q}$:

¹⁾ Решение проблемы вырезания для K -теории с любыми коэффициентами приведено в статье А. А. Суслина [Su3*].

I является K_Λ -вырезаемым тогда и только тогда,
когда $\text{Tot}_*^{Z \ltimes I}(Z, \Lambda) = 0$.

Цель этого сообщения — обозначить главные этапы доказательства теоремы 1 в случае, когда I есть $\mathbb{Q}$ -алгебра. В разд. 1 K -вырезание выражается в терминах гомологий группы GL . В разд. 2 предположение об H -унитарности связывается сначала с циклическими гомологиями, а затем с гомологиями алгебры Ли gl . Общий план доказательства обсуждается в разд. 3. Разделы 4 и 5 посвящены сравнению гомологий группы GL и алгебры Ли gl .

В последнем разд. 6 затрагивается проблема вычисления препятствия к вырезанию, которая является более общей задачей по сравнению с определением условий, при которых это препятствие обращается в нуль.

Всюду в дальнейшем под алгеброй понимается ассоциативная алгебра, не обязательно с единицей.

Прежде чем перейти к доказательствам, приведем здесь одно обобщение полученных результатов, а также некоторое их приложение, содержащиеся в одной и той же статье [S-W].

Теорема 3. *Предположим, что для любого набора из m элементов $a_1, \dots, a_m$ псевдокольца I существуют элементы $b_1, \dots, b_m, c, d \in I$, такие, что $a_i = b_i c d$, $1 \leq i \leq m$, и левые аннуляторы элементов cd и c совпадают. Тогда I является K -вырезаемым.*

Одним из замечательных приложений этих теорем является доказательство следующего результата, известного как «гипотеза Каруби».

Теорема 4. *Для всех C^* -алгебр A существует канонический изоморфизм*

$$K_n(A \hat{\otimes} K) \cong K_n^{\text{top}}(A \hat{\otimes} K), \quad n \geq 0.$$

В этой формуле K обозначает C^* -алгебру компактных операторов счетномерного гильбертова пространства, $\hat{\otimes}$ — тензорное произведение в категории C^* -алгебр, верхний индекс top в K_n^{top} означает, что речь идет о топологической K -теории, т.е. K_n^{top} — это $\pi_n(BGL(-))$, где для построения классифицирующего пространства учитывается топология группы GL .

А вот еще одно приложение в том же ключе, полученное М. Водзицки:

Теорема 5. Для всех C^* -алгебр A группы $K_n(A \otimes K)$ алгебраической K -теории являются периодическими периода 2.

Я благодарю Андрея Суслина и Мариуса Водзицки за их комментарии в то время, когда эта статья редактировалась.

1. ОБЩАЯ ЛИНЕЙНАЯ ГРУППА И АЛГЕБРАИЧЕСКАЯ K -ТЕОРИЯ

Для всякого кольца A линейная группа $GL_n(A)$ — это группа обратимых $n \times n$ -матриц с коэффициентами в A . Окаймляя матрицу нулями и одной единицей, получаем вложение $i_n: GL_n(A) \hookrightarrow GL_{n+1}(A)$. Индуктивный предел этих вложений есть по определению группа $GL(A)$. Рассматриваемая как дискретная группа, $GL(A)$ допускает построение классифицирующего пространства $BGL(A)$ (пространства Эйленберга—Маклейна типа $K(GL(A), 1)$), гомологии которого обозначаются через $H_*(BGL(A))$ или $H_*(GL(A))$.

Квиллен показал, что можно изменить пространство $BGL(A)$, прибавляя к нему клетки, таким образом, чтобы из нового пространства $BGL(A)^+$ сделать H -пространство без изменения его гомологий (см. [Q] или [L1]). По определению алгебраическая K -теория кольца A — это

$$K_n(A) := \pi_n(BGL(A)^+), \quad n \geq 1. \quad (1.1)$$

Главным свойством этих групп, которое будет нам нужно, является следующее. Градуированное векторное $\mathbb{Q}$ -пространство $H_*(GL(A), \mathbb{Q}) = \bigoplus_{n \geq 0} H_n(GL(A), \mathbb{Q})$ наделено коумножением, расположенным на диагонали, и умножением, сосредоточенным на прямой сумме матриц¹⁾ (так как это внутренняя операция на

¹⁾ Для любых двух матриц $\alpha, \beta \in GL(R)$ определим матрицу $\alpha \oplus \beta \in GL(R)$ формулами: $(\alpha \oplus \beta)_{ij} = 0$, если i, j имеют разную четность, $(\alpha \oplus \beta)_{2i, 2j} = \alpha_{i, j}$, $(\alpha \oplus \beta)_{2i+1, 2j+1} = \beta_{i, j}$. Тем самым возникает гомоморфизм $\varphi: GL(R) \times GL(R) \rightarrow GL(R)$ и соответствующее отображение классифицирующих пространств $B\varphi: BGL(R) \times BGL(R) \rightarrow BGL(R)$. Существует единственное, с точностью до гомотопии, отображение Σ , делающее следующую диаграмму гомотопически коммутативной:

$$\begin{array}{ccc} BGL(R) \times BGL(R) & \xrightarrow{B\varphi} & BGL(R) \\ \downarrow & & \downarrow \\ BGL(R)^+ \times BGL(R)^+ & \xrightarrow{\Sigma} & BGL(R)^+ \end{array}$$

Отображение Σ превращает $BGL(R)^+$ в гомотопически ассоциативную и коммутативную H -группу. (См. [L3].)

$GL(A)$), которые превращают его в градуированную коммутативную и кокоммутативную $\mathbf{Q}$ -алгебру Хопфа. Следовательно, $H_*(GL(A), \mathbf{Q})$ полностью определена своей примитивной частью, которая есть в точности K -теория кольца A . А именно, имеем изоморфизм

$$\Lambda \left(\bigoplus_{n \geq 1} K_n(A) \mathbf{Q} \right) \cong H_*(GL(A), \mathbf{Q}), \quad (1.2)$$

где Λ — функтор «симметричной градуированной алгебры».

Пусть I — двусторонний идеал в A . Сюръективный гомоморфизм $A \rightarrow A/I$ индуцирует гомоморфизм групп $GL(A) \rightarrow GL(A/I)$, образ которого мы обозначим через $\overline{GL}(A/I)$. Индуцированное отображение $BGL(A)^+ \rightarrow \overline{BGL}(A/I)^+$ имеет гомотопически связный слой, обозначаемый $F(A, I)$. Заметим, что $\pi_n(BGL(A/I)^+) = \pi_n(\overline{BGL}(A/I)^+)$ для $n > 1$ и $\pi_1(\overline{BGL}(A/I)^+) = \text{Im}(K_1(A) \rightarrow K_1(A/I))$. По определению *относительные K -группы* пары (A, I) — это группы

$$K_n(A, I) := \pi_n(F(A, I)), \quad n \geq 1. \quad (1.3)$$

Следовательно, точная последовательность (0) из вводной части есть просто длинная точная гомотопическая последовательность слоения.

Отметим здесь, что группы K_n существуют также для $n < 1$. На самом деле они определены, начиная с группы Гротендика K_0 , для которой проблема вырезания решена.

Первый этап в изучении вопроса вырезаемости в алгебраической рациональной K -теории состоит в сравнении гомологий группы

$$GL(I) := \text{Ker}(GL(\mathbf{Z} \ltimes I) \rightarrow GL(\mathbf{Z}))$$

с гомологиями группы $\widetilde{GL}(I) := \lim_n \widetilde{GL}_n(I)$, где

$$\begin{aligned} \widetilde{GL}_n(I) &:= GL_n(I) \ltimes M_{n1}(I) \\ &= \left\{ \begin{pmatrix} \alpha & v \\ 0 & 1 \end{pmatrix} \in GL_{n+1}(I) \mid \alpha \in GL_n(I), v \in M_{n1}(I) \right\}. \end{aligned}$$

Здесь $M_{n1}(I)$ обозначает $n \times 1$ -матрицы с коэффициентами в I . Заметим, что для всякой подгруппы G в $GL_n(I)$ (соотв. $GL(I)$) имеется группа $\tilde{G}$, определенная формулой $\tilde{G} := G \ltimes M_{n1}(I)$. Ана-

логичным образом группа $\widetilde{\widetilde{GL}}_n(I)$ (и, более общим образом, группа $\widetilde{\widetilde{G}}$) определена формулой

$$\widetilde{\widetilde{GL}}_n(I) := M_{1n}(I) \ltimes GL_n(I).$$

1.4. Предложение. *Рассмотрим следующую коммутативную диаграмму гомотопических слоений:*

$$\begin{array}{ccccc} BGL(I) & \longrightarrow & BGL(A) & \longrightarrow & B\overline{GL}(A/I) \\ \downarrow \varphi & & \downarrow & & \downarrow \\ F(A, I) & \longrightarrow & BGL(A)^+ & \longrightarrow & B\overline{GL}(A/I)^+ \end{array}$$

Если вложения $GL(I) \rightarrow \widetilde{\widetilde{GL}}(I)$ и $GL(I) \rightarrow \widetilde{\widetilde{GL}}(I)$ индуцируют изоморфизмы гомологий, то отображение φ индуцирует гомотопическую эквивалентность:

$$\varphi^+ : BGL(I)^+ \longrightarrow F(A, I).$$

Доказательство (набросок) (ср. следствие 1.7 в [S-W]). Два вертикальных отображения справа индуцируют изоморфизмы в гомологиях по основному свойству плюс-конструкции. Чтобы доказать, что φ индуцирует изоморфизм в гомологиях, мы хотим применить теорему сравнения точных спектральных последовательностей Зимана. Для выполнения предположений необходимо, чтобы группа π_1 базы действовала тривиально на гомологиях слоя. Для нижнего слоения это немедленно следует из того, что мы имеем дело с H -пространствами. Для верхнего слоения достаточно показать, что все элементы из $\pi_1(BGL_n(A)) = GL_n(A)$ действуют тривиально (действие задается сопряжением) на $H_*(GL(I))$. С помощью преобразования матриц (следствие 1.6 в loc. cit) можно установить, что для выполнения этого требования достаточно показать, что $GL(\mathbb{Z})$ действует тривиально на $H_*(GL(I))$. Первое предположение (с $\overline{GL}$) позволяет показать, что так обстоит дело для элементарных матриц $E_{i,i+1}(1)$, а второе (с $\widetilde{\widetilde{GL}}$) — для $E_{i+1,i}(1)$. Так как $GL(\mathbb{Z})$ порождена этими элементарными матрицами и -1 (которая действует тривиально), все остальное ясно. $\square$

Таким образом, имеется гомологическое условие на GL , которое влечет за собой K -вырезаемость псевдокольца I . Заметим, что достаточно показать, что это условие выполняется для $\overline{GL}$: условие для $\widetilde{\widetilde{GL}}$ получается тогда по двойственности.

2. АЛГЕБРЫ ЛИ МАТРИЦ И ЦИКЛИЧЕСКИЕ ГОМОЛОГИИ

Если в предыдущем разделе мы заменим группу $GL(A)$ на алгебру Ли $gl(A)$ (где A теперь является ассоциативной алгеброй с единицей над полем k характеристики 0), мы получим также, что гомологии алгебры Ли $gl(A)$ с коэффициентами в k есть градуированная коммутативная и кокоммутативная алгебра Хопфа. Следовательно, она полностью определена своей примитивной частью, которую можно назвать «аддитивной K -теорией» алгебры A .

Огромное отличие от предыдущей ситуации состоит в том, что можно «вычислить» эту примитивную часть в том смысле, что можно предъявить комплекс *без матриц*, гомологии которого являются этой примитивной частью.

2.1. Теорема (Лоде—Квиллен [L-Q], Цыган [Т]). *Для любой ассоциативной алгебры A с единицей над полем k характеристики 0 существует изоморфизм градуированных алгебр Хопфа*

$$H_*(gl(A), k) \cong \Lambda HC_{*-1}(A),$$

где $HC_n(A)$, $n \geq 0$, — циклические гомологии алгебры A .

Здесь циклические гомологии определяются следующим образом. Возьмем комплекс $B_*(A)$ из введения, использовавшийся для определения понятия H -унитарности. Когда A — алгебра с единицей, он является резольвентной для A , состоящей из A -бимодулей (берем $A^{\otimes n+2}$ в степени n). Обозначим через A^{op} кольцо, противоположное A , через A^e — тензорное произведение $A \otimes_k A^{op}$ и через $(a_0, \dots, a_n)$ порождающую $(a_0 \otimes \dots \otimes a_n)$ в $A^{\otimes n+1}$. По определению комплекс Хохшильда $(C_*(A), b)$ — это комплекс $(A \otimes_{A^e} B_*(A), \text{id}_A \otimes b')$. Точнее, имеем $C_n(A) = A^{\otimes n+1}$, так как $a \otimes (a_0, \dots, a_{n+1}) \mapsto (a_{n+1}aa_0, a_1, \dots, a_n)$ задает изоморфизм $A \otimes_{A^e} B_n(A) \cong A^{\otimes n+1}$. Отображение $b : C_n(A) \rightarrow C_{n-1}(A)$ тогда задается формулой

$$b(a_0, \dots, a_n) = \sum_{i=0}^{n-1} (-1)^i (a_0, \dots, a_i a_{i+1}, \dots, a_n) + (-1)^{n+1} (a_n a_0, a_1, \dots, a_{n-1}).$$

Гомологии комплекса $(C_*(A), b)$ обозначаются $HH_n(A)$, $n \geq 0$, и называются *гомологиями Хохшильда* алгебры A .

Введем теперь *циклический оператор* $t : C_n(A) \rightarrow C_n(A)$, определяемый формулой

$$t(a_0, \dots, a_n) = (-1)^n (a_n, a_0, \dots, a_{n-1}).$$

А. Кон [С] заметил, что имеется тождество

$$b(1-t) = (1-t)b'.$$

Следовательно, если мы положим $C_n^\lambda := A^{\otimes n+1}/(1-t)$, то получим новый комплекс $(C_*^\lambda(A), b)$, гомологии которого, обозначаемые $HC_n(A)$, $n \geq 0$, называются *циклическими гомологиями* алгебры A (это подходящее определение, когда k содержит $\mathbb{Q}$, — случай, который нас здесь интересует; другое определение см. в [L-Q, L2]). Одно из главных свойств циклических гомологий — длинная точная последовательность периодичности Кона

$$\cdots \rightarrow HH_n(I) \rightarrow HC_n(I) \rightarrow HC_{n-2}(I) \rightarrow HH_{n-1}(I) \rightarrow \cdots, \quad (2.2)$$

которая справедлива при приведенных здесь определениях для всех H -унитарных алгебр I (ср. [С] для унитарного и рационального случаев и [L-Q] для общего случая).

Хотя нам это не будет нужно в дальнейшем, приведем для закрепления указанных понятий следующее вычисление циклических гомологий гладкой (коммутативной и с единицей) $\mathbb{Q}$ -алгебры:

$$HC_n(A) \cong \Omega_A^n / d\Omega_A^{n-1} \oplus H_{DR}^{n-2}(A) \oplus H_{DR}^{n-4}(A) \oplus \cdots$$

Здесь Ω_A^n обозначает модуль n -форм абсолютных дифференциалов и H_{DR}^n — когомологии де Рама.

Проблема вырезания поставлена также для теорий HH и HC . Для этих теорий ее, очевидно, решать гораздо легче, чем для алгебраической K -теории.

Всякий эпиморфизм алгебр $A \rightarrow A/I$ задает эпиморфизм комплексов $C(A) \rightarrow C(A/I)$ (соотв. $C^\lambda(A) \rightarrow C^\lambda(A/I)$). По определению относительные гомологии Хохшильда $HH_n(A, I)$ (соотв. циклические гомологии $HC_n(A, I)$) — это гомологии ядра этого эпиморфизма. Говорят, что I является *HH -вырезаемой* (соотв. *HC -вырезаемой*), если $HH_n(I) := HH_n(\mathbb{Z} \ltimes I, I) \rightarrow HH_n(A, I)$ (соотв. $HC_n(\mathbb{Z} \ltimes I, I) \rightarrow HC_n(A, I)$) — изоморфизм для всех n .

2.3. Предложение [W1, th. 3]. Для всякой $\mathbb{Q}$ -алгебры I следующие условия эквивалентны:

- (a) I H -унитарна,
- (b) I HH -вырезаема,
- (c) I HC -вырезаема.

Комментарии к доказательству. Эквивалентность между пп. (b) и (c) непосредственно следует из точной последовательности Кона. Чтобы доказать, что из (b) следует (a), достаточно просчитать частный случай, когда алгебра содержит нильпотентный идеал. Чтобы доказать, что из (a) следует (b), Водзицки заметил, что комплекс, ацикличность которого нужно доказать, является на самом деле тотальным комплексом «мультикомплекса». Предположение об H -унитарности влечет за собой ацикличность мультикомплекса, поскольку гомологии берутся в одном направлении. Этого достаточно для того, чтобы отсюда следовала ацикличность тотального комплекса. $\square$

Принимая во внимание все, что сделано для K -теории, и связь между $HC_*(A)$ и $H_*(gl(A))$, совершенно естественно переписать свойство вырезания для HC в терминах гомологий алгебр Ли gl .

2.4. Предложение. Для всякой H -унитарной $\mathbb{Q}$ -алгебры I гомоморфизм

$$H_*(gl(I)) \rightarrow H_*(\tilde{gl}(I))$$

является изоморфизмом.

Доказательство. Те же аргументы, что и в предыдущем доказательстве, позволяют доказать, что алгебра $I_1 = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \mid a, b \in I \right\}$ H -унитарна и что вложение $I \hookrightarrow I_1$ индуцирует изоморфизм гомологий Хохшильда (ср. [W1]). Ввиду точной последовательности Кона это также изоморфизм циклических гомологий. Для того чтобы перенести этот результат на гомологии алгебр Ли gl , используется обобщение теоремы Лоде—Квиллена—Цыгана (ср. п. 2.1) для H -унитарных алгебр, полученное П. Ханлоном [H]. Отсюда получается изоморфизм

$$H_*(gl(I)) \xrightarrow{\sim} H_*(gl(I_1)).$$

Наконец, легко сравнить гомологии алгебр Ли $gl(I_1)$ и $\tilde{gl}(I)$. $\square$

2.5. Замечание. Отметим, что верно и обратное утверждение к предложению 2.4 (М. Водзицки, частное обобщение).

3. ГЛАВНАЯ ТЕОРЕМА

Мы ограничимся следующим случаем теоремы Суслина и Водзицки.

3.1. Теорема. Для всякой $\mathbf{Q}$ -алгебры (не обязательно с единицей) I следующие условия эквивалентны.

- (a) I H -унитарна,
- (b) I HC -вырезаема (с $k = \mathbf{Q}$),
- (c) I вырезаема для $K_{\mathbf{Q}}$.

Импликации (a) $\Leftrightarrow$ (b) и (c) $\Rightarrow$ (a) есть уже в [W1] (ср. разд. 2 и 5.5). Новый результат, который появился в [S-W], — это импликация (a) $\Rightarrow$ (b), которую мы теперь и собираемся разобрать.

Мы хотим выразить свойство K -вырезания (соотв. HC -вырезания) в терминах гомологий групп GL (соотв. алгебр gl), ср. п. 1.4 (соотв. п. 2.4). Следовательно, нам нужно найти способ сравнить эти две группы гомологий. Идея состоит в том, чтобы использовать конструкцию Володина для описания гомологий групп GL . Эта конструкция использует возможность восстановления группы $GL(I)$ по подгруппе $T_n^\sigma(I)$ треугольных матриц. А именно, мы собираемся построить пространство $\underline{T}(I)$ (соотв. $\tilde{T}(I)$), начиная с построения классифицирующих пространств $BT_n^\sigma(I)$ (соотв. $B\tilde{T}_n^\sigma(I)$), а затем применяя непрерывное отображение $\underline{T}(I) \rightarrow \tilde{T}(I)$.

Аналогичным образом, в рамках алгебр Ли, будем ставить в соответствие алгебре Ли $gl(I)$ (соотв. $\tilde{gl}(I)$) цепной комплекс $\underline{t}(I)$ (соотв. $\tilde{t}(I)$), строящийся по алгебрам Ли $t_n^\sigma(I)$, соответствующим $T_n^\sigma(I)$. Затем строится морфизм комплексов $\underline{t}(I) \rightarrow \tilde{t}(I)$.

Ключевой момент, использовавшийся впервые в аналогичном контексте в статье Т. Гудвилли [G], состоит в том, что, поскольку $T_n^\sigma(I)$ — нильпотентная группа, можно сравнить ее рациональные гомологии с рациональными гомологиями алгебры Ли $t_n^\sigma(I)$. Это позволяет сравнить $H_*(\underline{T}(I))$ с $H_*(\underline{t}(I))$ и $H_*(\tilde{T}(I))$ с $H_*(\tilde{t}(I))$.

Сформулируем план доказательства импликации (a) $\Rightarrow$ (c) теоремы 3.1:

I H -унитарна

$\Downarrow$ (0)

I HC -вырезаема $\stackrel{(1)}{\Leftrightarrow} H_*(gl(I)) \cong H_*(\tilde{gl}(I)) \stackrel{(2)}{\Leftrightarrow} H_*(\underline{t}(I)) \cong H_*(\tilde{t}(I))$

$\Downarrow$ (3)

I $K_{\mathbf{Q}}$ -вырезаема $\stackrel{(1')}{\Leftrightarrow} H_*(GL(I)) \cong H_*(\widetilde{GL}(I)) \stackrel{(2')}{\Leftrightarrow} H_*(\underline{T}(I)) \cong H_*(\tilde{T}(I))$.

Этапы (0) и (1) обсуждались в разд. 2. Этап (1') обсуждался в разд. 1. Обсудим теперь некоторые подробности этапов (2), (2') и (3).

4. КОНСТРУКЦИИ ВОЛОДИНА

Пусть I — псевдокольцо. Для любого частичного упорядочения σ набора $\{1, 2, \dots, n\}$ обозначим через $T_n^\sigma(I)$ подгруппу σ -треугольных матриц группы $GL_n(I)$:

$$T_n^\sigma(I) = \{(a_{ij}) \in GL_n(I) | a_{ij} = 0, \text{ если } i \not\leq_\sigma j \text{ и } a_{ii} = 1, 1 \leq i \leq n\}.$$

В частности, если σ — обычное упорядочение $1 < 2 < \dots < n$, то $T_n^\sigma(I)$ будет просто подгруппой верхнетреугольных матриц в обычном смысле.

Классифицирующее пространство $BT_n^\sigma(I)$ является подпространством пространства $BGL_n(I)$ и пространство $\underline{T}(I) := \bigcup_{n, \sigma} BT_n^\sigma(I)$ является подпространством пространства $BGL(I)$ (замечание: это не дизъюнктное объединение).

Связь между гомологиями пространства $\underline{T}(I)$ и гомологиями пространства $BGL(I)$ получается в виде спектральной последовательности следующим образом.

Для любой дискретной группы G классифицирующее пространство BG имеет универсальное накрытие EG , которое стягиваемо. Если $\{G_j\}_{j \in J}$ — семейство подгрупп группы G , то существует связанное подпространство $V(G, \{G_j\})$ пространства EG над $\bigcup_j BG_j$, на котором действует G и для которого имеется спектральная последовательность

$$E_{pq}^2 = H_p(BG; H_q(V(G, \{G_j\}))) \Rightarrow H_{p+q} \left(\bigcup_j BG_j \right). \quad (4.1)$$

Это пространство, называемое *пространством Володина*, обладает, кроме того, следующим свойством. Для любого другого набора $(G', \{G'_j\}_{j \in J})$ и любого гомоморфизма $\varphi : G \rightarrow G'$, такого, что $\varphi(G_j) \subset G'_j$ и $\varphi : G/G_{i_1} \cap \dots \cap G_{i_p} \xrightarrow{\sim} G'/G'_{i_1} \cap \dots \cap G'_{i_p}$ есть изоморфизм для любого набора из p индексов,

$$\text{индуцированное отображение } V(G, \{G_j\}) \rightarrow V(G', \{G'_j\}) \quad (4.2)$$

является гомотопической эквивалентностью.

На практике обычно эти конструкции работают в рамках симплициальных объектов.

4.3. Предложение. Для любого псевдокольца I , такого, что $I = I^2$, следующие условия эквивалентны:

$$(a) H_*(GL(I)) \cong H_*(\widetilde{GL}(I)),$$

$$(b) H_*(\underline{T}(I)) \cong H_*(\widetilde{\underline{T}}(I)).$$

Набросок доказательства. Из свойства (4.2) следует, что два пространства Володина $V(I)$ и $\tilde{V}(I)$ гомотопически эквивалентны. Затем можно показать, что $GL(I^2)$ (соотв. $\widetilde{GL}(I^2)$) действует тривиально на $H_*(V(I))$ (соотв. $H_*(\tilde{V}(I))$). Тогда наше предложение является следствием сравнения точных спектральных последовательностей (4.1) для $\underline{T}(I) = \bigcup_{n,\sigma} BT_n^\sigma(I)$ и $\tilde{\underline{T}}(I) = \bigcup_{n,\sigma} B\tilde{T}_n^\sigma(I)$. $\square$

Для доказательства этапа (2) повторяются предыдущие рассуждения, но в рамках алгебр Ли. Разница в технике состоит в том, что классифицирующие пространства заменяются комплексами. Точнее, объект базы — это комплекс Шевалле—Эйленберга

$$C_*(\mathfrak{g}, M) : \dots \rightarrow M \otimes \Lambda^n \mathfrak{g} \xrightarrow{d} M \otimes \Lambda^{n-1} \mathfrak{g} \rightarrow \dots \rightarrow M,$$

где $\mathfrak{g}$ — алгебра Ли над коммутативным кольцом k , $\Lambda^n \mathfrak{g}$ — ее n -я внешняя степень, M — некоторый $\mathfrak{g}$ -модуль и d задается формулой

$$d(g_0, g_1, \dots, g_n) = \sum_{0 \leq i < j \leq n} (-1)^j (g_0, \dots, g_{i-1}, [g_i, g_j], g_{i+1}, \dots, \hat{g}_j, \dots, g_n),$$

где $g_0 \in M$, $g_1, \dots, g_n \in \mathfrak{g}$.

По определению гомологии алгебры Ли $\mathfrak{g}$ с коэффициентами в M , обозначаемые $H_n(\mathfrak{g}, M)$, — это гомологии комплекса $C_*(\mathfrak{g}, M)$.

Для любого семейства $\{g_j\}_{j \in J}$ подалгебр Ли алгебры $\mathfrak{g}$ пространство Володина заменяется некоторым комплексом Володина $v_*(\mathfrak{g}, \{g_j\}) \subset C_*(\mathfrak{g}, U(\mathfrak{g}))$ ($U(\mathfrak{g})$ — универсальная обертывающая алгебры Ли $\mathfrak{g}$).

Аналогом пространства $\underline{T}(I)$ является комплекс $\underline{t}_*(I)$, который строится следующим образом. Пусть $t_n^\sigma(I)$ — алгебра Ли σ -треугольных матриц из $gl_n(I)$. Комплекс $C_*(t_n^\sigma(I))$ — это подкомплекс комплекса $C_*(gl_n(I))$. По определению $\underline{t}_*(I)$ — это следующий подкомплекс комплекса $C_*(gl(I))$:

$$\underline{t}_*(I) := \sum_{n,\sigma} C_*(t_n^\sigma(I)).$$

Свойства комплекса Володина, совершенно аналогичные свойствам пространства Володина, позволяют доказать

4.4. Предложение. Для любой k -алгебры I , такой, что $I^2 = I$, следующие условия эквивалентны:

- (a) $H_*(gl(I)) \cong H_*(\tilde{gl}(I))$,
- (b) $H_*(\underline{t}(I)) \cong H_*(\tilde{\underline{t}}(I))$. $\square$

4.5. Замечание. На самом деле комплексы $\underline{t}(I)$ и $\tilde{\underline{t}}(I)$ ацикличны, поскольку I является H -унитарной (ср. разд. 9 в [S-W]). В новом препринте [Su2] А. Суслин дал прямые указания для доказательства этой ацикличности, сделав ненужными этапы (0), (1) и (2). Тем не менее они очень интересны для полного вычисления препятствия к вырезанию (ср. разд. 6.).

5. СРАВНЕНИЕ ГОМОЛОГИЙ ЛИНЕЙНОЙ ГРУППЫ И ЕЕ АЛГЕБРЫ ЛИ

Теория Мальцева устанавливает взаимно однозначное соответствие между нильпотентными алгебрами Ли $\mathfrak{n}$ над $\mathbb{Q}$ и нильпотентными группами N с однозначным делением. По этому соответствию $T_{\mathfrak{n}}^{\sigma}$ отвечает $t_n^{\sigma}(I)$. Главным моментом здесь является следующий результат.

5.1. Теорема (ср. [P]). Для любой нильпотентной алгебры Ли $\mathfrak{n}$ над $\mathbb{Q}$ существует канонический изоморфизм (градуированных коалгебр)

$$H_*(\mathfrak{n}, \mathbb{Q}) \cong H_*(N, \mathbb{Q}).$$

Слева стоят гомологии алгебры Ли, справа — гомологии дискретной группы.

Набросок доказательства. Универсальная обертывающая $U(\mathfrak{n})$ алгебры $\mathfrak{n}$ имеет пополнение в топологии, определяемой ее идеалом аугментации, которое обозначается $U(\mathfrak{n})^{\wedge}$. Такая же алгебра $\mathbb{Q}[N]$ дискретной группы N допускает аналогичное пополнение, обозначаемое $\mathbb{Q}[N]^{\wedge}$. Тогда в силу условия нильпотентности выполняются следующие условия (ср. [P]):

- (a) вложение $N \subset (U(\mathfrak{n})^{\wedge})^{\times}$ индуцирует изоморфизм алгебр Хопфа $\mathbb{Q}[N]^{\wedge} \cong U(\mathfrak{n})^{\wedge}$,
- (b) $\mathbb{Q}[N]^{\wedge}$ плоская над $\mathbb{Q}[N]$,
- (c) $U(\mathfrak{n})^{\wedge}$ плоская над $U(\mathfrak{n})$.

Два последних условия выполняются в силу свойства Артина—Риса.

Изоморфизм из теоремы 5.1 можно построить, сравнивая абстрактно резольвенты, дающие соответственно комплекс Шевалле–Эйленберга (для n) и комплекс Маклейна (для N). Но доказательство этим методом его каноничности представляется весьма тонким (ср. [S-W, § 5]). Другой метод состоит в сравнении этих двух групп гомологий с $\text{To}^{U(n)\wedge}(\mathbb{Q}, \mathbb{Q}) = \text{To}^{Q[N]\wedge}(\mathbb{Q}, \mathbb{Q})$, как в [L2, XI.3.14]. □

С помощью классических аргументов из точной последовательности Майера–Вьеториса можно вывести следующий результат:

5.2. Теорема. *Для любого псевдокольца I имеется изоморфизм*

$$H_*(\underline{T}(I)) := H_*\left(\bigcup_{n,\sigma} BT_n^\sigma(I)\right) \cong H_*\left(\sum_{n,\sigma} C_*(t_n^\sigma(I))\right) =: H_*(\underline{t}(I)),$$

и аналогичный изоморфизм имеет место для $\tilde{T}(\tilde{I})$ и $\tilde{t}(I)$.

Отсюда мы выведем последний шаг доказательства (3):

5.3. Следствие. *Для любой $\mathbb{Q}$ -алгебры I следующие условия эквивалентны:*

- (a) $H_*(\underline{T}(I)) \cong H_*(\tilde{T}(I))$,
- (b) $H_*(\underline{t}(I)) \cong H_*(\tilde{t}(I))$. □

5.4. Замечание. Теорема 1 (I — псевдокольцо, $I_{\mathbb{Q}}$ H -унитарна) происходит по существу из теоремы 3.1 (справедливой для $I_{\mathbb{Q}}$) и того факта, что группа $H_*(T_n^\sigma(I); \mathbb{Q})$ изоморфна $H_*(T_n^\sigma(I_{\mathbb{Q}}); \mathbb{Q})$ (и аналогичного факта с $\tilde{T}$ вместо T).

5.5. Комментарии к доказательству импликации « I $K_{\mathbb{Q}}$ -вырезается $\Rightarrow I$ H -унитарна».

Этот результат был прежде получен М. Водзицки [W1] с использованием техники того же типа. Главное состоит в том, чтобы выбрать, начиная с I , вырезание, что сводится к вычислению относительной K -группы $K_*(R, J)_{\mathbb{Q}}$, где J — нильпотентный идеал. Ключевым моментом является теорема Гудвилли [G], утверждающая, что тогда имеется изоморфизм

$$K_*(R, J)_{\mathbb{Q}} \cong HC_{*-1}(R, J)_{\mathbb{Q}}. \quad (5.6)$$

Предположение о K -вырезании (обращение в нуль некоторой относительной K -группы) влечет за собой тогда обращение в нуль

некоторой группы относительных циклических гомологий. Пример (R, J) подобран таким образом, что группа относительных циклических гомологий содержит $H_*(B_*(I), b')$ в качестве прямого слагаемого.

Отметим, что представленные выше идеи относительно пространств и комплексов Володина проистекают из доказательства изоморфизма (5.6), предложенного Т. Гудвилли [G], и статьи [Su1] А. Суслина.

6. ПРЕПЯТСТВИЕ К ВЫРЕЗАНИЮ В АЛГЕБРАИЧЕСКОЙ K -ТЕОРИИ

Пусть I — двусторонний идеал кольца A . Пусть $\varphi : A \rightarrow B$ — гомоморфизм колец, такой, что при ограничении на I он становится инъективным и $\varphi(I)$, обозначаемый также I , — двусторонний идеал кольца B . Коммутативная диаграмма

$$\begin{array}{ccc} A & \xrightarrow{\varphi} & B \\ \downarrow & & \downarrow \\ A/I & \xrightarrow{\varphi'} & B/I \end{array} \quad (6.0)$$

декартова. Если I является K -вырезаемым, то существует изоморфизм $K_n(A, I) \cong K_n(B, I)$, из которого выводится точность последовательности Майера—Вьеториса

$$\begin{aligned} \cdots \rightarrow K_n(A) \rightarrow K_n(B) \oplus K_n(A/I) \rightarrow K_n(B/I) \\ \rightarrow K_{n-1}(A) \rightarrow \cdots \rightarrow K_0(B/I). \end{aligned} \quad (6.1)$$

Как уже указывалось во введении, $K_0(A, I) \rightarrow K_0(B, I)$ — всегда изоморфизм. Если к тому же φ сюръективен, то можно показать, что имеется также изоморфизм $K_1(A, I) \xrightarrow{\sim} K_1(B, I)$ (ср. [M, lemme 6.3]). Если положить $J = \ker \varphi$, то это эквивалентно тому, что для кольца A и двух двусторонних идеалов I и J выполняется условие

$$I \cap J = \{0\}.$$

В этом случае естественно построить биотносительные K -группы $K_n(A; I, J)$, составляющие точную последовательность

$$\begin{aligned} \cdots \rightarrow K_n(A; I, J) \rightarrow K_n(A, I) \rightarrow K_n(A/J, I + J/J) \\ \rightarrow K_{n-1}(A; I, J) \rightarrow \cdots \end{aligned} \quad (6.2)$$

Эти группы — *препятствие к точности* последовательности Майера—Вьеториса.

Предыдущие результаты выражаются в следующем виде:

$$K_0(A; I, J) = K_1(A; I, J) = 0.$$

Первый пример, для которого $K_2(A; I, J) \neq 0$, был построен Сваном [Sw]: $A = \mathbf{Z}[x, y]/(xy)$, $I = (x)$, $J = (y)$. В этой ситуации $K_2(A; I, J)$ является прямым слагаемым в $K_2(A)$. Элемент, нетривиальность которого доказал Сван, является образом стандартной порождающей группы $H_2(\mathbf{Z}^2)$ (фундаментального класса тора) при гомоморфизме $\rho_* : H_2(\mathbf{Z}^2) \rightarrow H_2(E(A)) \cong K_2(A)$, индуцируемом гомоморфизмом

$$\rho : \mathbf{Z}^2 \rightarrow E(A), \quad \rho(1, 0) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad \rho(0, 1) = \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix}.$$

На самом деле можно точно вычислить первую группу препятствия.

6.3. Теорема [GW-L, K]. Если $I \cap J = \{0\}$, то

$$K_2(A; I, J) \cong I \otimes_{A^e} J, \quad \text{где } A^e = A \otimes A^{\text{op}}.$$

Заметим, что эта теорема не требует предположения о рациональности.

Если заменить K -теорию циклическими гомологиями (напомним, что имеется сдвиг размерности в обозначениях), то легко доказать, что

$$HC_0(A; I, J) = 0 \quad \text{и} \quad HC_1(A; I, J) = I \otimes_{A^e} J.$$

Эти вычисления, соединенные с изоморфизмом Гудвилли (ср. (5.6)), сравнивающим K -теорию и циклические гомологии для нильпотентного идеала, и с доказательством теоремы Суслина—Водзицки, подсказывают следующую гипотезу:

6.4. Гипотеза. Если $I \cap J = \{0\}$, то

$$K_n(A; I, J)_{\mathbf{Q}} \cong HC_{n-1}(A; I, J)_{\mathbf{Q}}.$$

(На самом деле, если дело обстоит так, достаточно предположения о нильпотентности $I \cap J$.) Хотя эта гипотеза несколько раз анонсировалась, она еще не доказана.

В [G-W] можно найти результаты о препятствии к вырезанию в низких размерностях в случае, когда φ (ср. (6.0)) не сюръективно.

ЛИТЕРАТУРА¹⁾

- [B] Bass H. Algebraic K -theory, Benjamin, 1968. [Имеется перевод: Басс Х. Алгебраическая K -теория. — М.: Мир, 1973.]
- [C] Connes A. Non-commutative differential geometry, Publ. Math. I.H.E.S., **62** (1985), 257–360.
- [G-W] Geller S., Weibel C. $K(A, B, I)$: II, J. K -theory, **2** (1989), 753–760.
- [G] Goodwillie T. G. Relative algebraic K -theory and cyclic homology, Ann. of Math., **124** (1986), 347–402.
- [GW-L] Guin-Waléry D., Loday J.-L. Obstruction à l'excision en K -théorie algébrique, Lect. Notes in Math., vol. 854, Springer-Verlag, 1981, 179–216.
- [H] Hanlon P. On the complete $GL(n, C)$ -decomposition of the stable cohomology of $gl_n(A)$, Trans. Amer. Math. Soc., **308** (1988), 209–225.
- [K] Keune F. The relativization of K_2 , J. Algebra, **54** (1978), 159–177.
- [L1] Loday J.-L. K -théorie algébrique et représentations de groupes, Ann. Sci. Ec. Norm. Sup., **9** (1976), 309–377.
- [L2] Loday J.-L. Cyclic Homology, Grund. Math. Wiss., vol. 301, Springer-Verlag, 1992.
- [L3*] Loday J.-L. Structure multiplicative en K -théorie algébrique, C. R. Acad. Sci., Paris Sér. A, **279** (1974), 321–324.
- [L-Q] Loday J.-L., Quillen D. Cyclic homology and the Lie algebra homology of matrices, Comm. Math. Helvetici, **59** (1984), 565–591.
- [M] Milnor J. Introduction to Algebraic K -theory, Ann. Math. Studies, vol. 72, Princeton Univ. Press, 1974. [Имеется перевод изд. 1971 г.: Милнор Дж. Введение в алгебраическую K -теорию. — М.: Мир, 1974.]
- [P] Pickel P. F. Rational cohomology of nilpotent groups and Lie algebras, Comm. in Alg., **6** (1978), 409–419.
- [Q] Quillen D. Cohomology of groups, Actes du Congrès International des Mathématiciens, t. II, 1970, 47–51.
- [Su1] Suslin A. On the equivalence of K -theories, Comm. in Alg., **9** (1981), 1559–1566.
- [Su2] Suslin A. On the acyclicity of the sum of triangular complexes, Preprint, University of Utrecht, 1991.
- [Su3*] Суслин А. А. Вырезание в целочисленной алгебраической K -теории. — Труды МИАН, т. 208 (1995), с. 290–317.
- [S-W] Suslin A., Wodzicki M. Excision in algebraic K -theory, Ann. of Math., **136** (1992), 51–122.
- [SW] Swan R. G. Excision in algebraic K -theory, J. Pure Appl. Algebra, **1** (1971), 221–252.
- [T] Цыган Б. Л. Гомологии матричных алгебр Ли над кольцами и гомологии Хохшильда. — УМН, т. 38, 1983, вып. 2, с. 217–218.

- [W1] Wodzicki M. Excision in cyclic homology and in rational algebraic K -theory, *Ann. of Math.*, **129** (1989), 591–636.
- [W2] Wodzicki M. Homological properties of rings of functional analytic type, *Proc. Nat. Acad. Sci. USA*, **87** (1990), 4910–4911.
- [W3] Wodzicki M. The long exact sequence in cyclic homology associated with an extension of algebras, *C. R. Acad. Sci. Paris Sér. A-B*, **307** (1988), 249–254.

JEAN-LOUIS LODAY

Institut de Recherche Mathématique Avancée
Université Louis Pasteur et C.N.R.S.
7, rue René-Descartes
F-67084 Strasbourg Cedex
E-mail: loday@math.u-strasbg.fr

ПОЛНОЕ РЕШЕНИЕ ПРОБЛЕМЫ ЗИГЕЛЯ О ЛИНЕАРИЗАЦИИ ГОЛОМОРФНОГО ОТОБРАЖЕНИЯ В ОКРЕСТНОСТИ НЕПОДВИЖНОЙ ТОЧКИ (по Ж.-К. Йоккозу)

Рикардо Перес-Марко

Мы представляем здесь новые геометрические методы, которые оказались эффективными при исследовании проблем малых знаменателей в голоморфной динамике. Они были изобретены Ж.-К. Йоккозом [Y1] для решения проблемы Зигеля. Он также с успехом применил свою революционную технику для того, чтобы получить наилучшие диофантовы условия в проблеме линеаризации аналитических диффеоморфизмов окрестности [Y2].

Для того чтобы предмет не вызывал затруднений у неспециалистов, мы (прося снисхождения у остальных) начинаем с элементарного введения и небольшого исторического обзора. Более исчерпывающее введение и полная библиография имеются в статье М. Эрмана [He1].

Я благодарю П. Сентенаке за большую помощь в подготовке этого текста.

1. ГОЛОМОРФНАЯ ДИНАМИКА В ОКРЕСТНОСТИ НЕПОДВИЖНОЙ ТОЧКИ

Рассмотрим динамическую систему, т. е. отображение $f : X \rightarrow X$ некоторого множества в себя. С наивной точки зрения изучить динамику системы f — это значит исследовать траекторию каждой точки. Более глубокое изучение связано с получением классификации классов сопряженности отображения f относительно морфизмов множества X . В качестве первого шага нужно понять структуру периодических (т. е. имеющих конечную орбиту) точек и динамику в окрестности периодических орбит. Каждая периодическая орбита — это неподвижная точка некоторой итерации. Поэтому естественно изучать локальную динамику в окрестности неподвижной точки.

В голоморфной динамике пространство X наделено комплексной структурой, которую сохраняет отображение f , т.е. f голоморфно. Первичная программа исследований не завершена даже при $\dim_{\mathbb{C}} X = 1$. В дальнейшем мы в основном ограничимся этим случаем. Мы предполагаем, что неподвижная точка находится в начале координат в $\mathbb{C}$ и что f определено в его окрестности. Мы записываем также $f(z) = \lambda z + \sum_{n \geq 2} f_n z^n$.

1.1. Предварительные замечания. Важнейшим понятием динамики в окрестности неподвижной точки является понятие устойчивости.

Определение (устойчивость). Отображение f устойчиво в 0, если для любой окрестности нуля V существует окрестность нуля U , такая, что при $n \geq 0$ отображение f^n определено в U и $f^n(U) \in V$.

Ясно, что это эквивалентно существованию окрестности нуля, в которой положительные итерации отображения f корректно определены и образуют нормальное семейство.

Если линейная часть отображения f в 0 ненулевая ($\lambda \neq 0$), то f хорошо приближается в окрестности нуля отображением $z \mapsto \lambda z$. Естественно возникает вопрос, совпадает ли динамика. Когда это так, говорят, что f линеаризуемо.

Определение (линеаризуемость). Отображение f называется *линеаризуемым* в 0, если $\lambda \neq 0$ и существует голоморфная замена переменных $z \mapsto h(z) = z + \sum_{n \geq 0} h_n z^n$, сопрягающая отображение f с его линейной частью:

$$(h^{-1} \circ f \circ h)(z) = \lambda z.$$

Если λ не является корнем из единицы, отображение h формально однозначно определяется из этого функционального уравнения. В этом случае мы говорим, что существует линеаризующее формальное отображение. Обозначим через $R(f)$ его радиус сходимости. В нелинеаризуемом случае $R(f) = 0$. Образ $h(D_{R(f)})$ называется областью Зигеля или областью линеаризации отображения f (где $D_r = \{z \in \mathbb{C}; |z| < r\}$).

Примечательно, что в голоморфном случае понятия устойчивости для f или f^{-1} и линеаризуемости эквивалентны.

Предложение. Пусть $\lambda \neq 0$ и $|\lambda| \leq 1$. Отображение f линеаризуемо тогда и только тогда, когда оно устойчиво в 0.

Доказательство. Если f линейризуемо, то его положительные итерации образуют нормальное семейство в области линейризуемости отображения f . Обратно, пусть U — окрестность нуля, в которой итерации $(f^n)_{n \geq 0}$ образуют нормальное семейство. Пусть

$$k_n = \frac{1}{n} \sum_{i=0}^{n-1} \lambda^{-i} f^i.$$

Имеем

$$k_n \circ f = \lambda k_n + \frac{\lambda}{n} (\lambda^{-n} f^n - \text{id}).$$

Последовательность $(k_n)_{n \geq 0}$ образует нормальное семейство. Из нее можно извлечь подпоследовательность, сходящуюся к k , и $h \neq k^{-1}$ линейризует f . $\square$

1.2. Сверхпритягивающая неподвижная точка. Мы называем так неподвижную точку, которая является также критической точкой, т. е. такую, что $\lambda = 0$. Динамика проста: все точки в некоторой окрестности нуля притягиваются к нулю. Нетрудно доказать следующее:

Теорема [Bot]. Пусть $f(z) = f_n z^n + O(z^{n+1})$, где $n \geq 0$ и $f_n \neq 0$. Тогда существует голоморфная замена переменных $z \mapsto h(z)$, $h(0) = 0$, сопрягающая f с отображением $z \mapsto z^n$, т. е.

$$(h^{-1} \circ f \circ h)(z) = z^n.$$

1.3. Притягивающая или отталкивающая неподвижная точка. Этот случай соответствует $0 < |\lambda| < 1$ (притягивающая точка) или $1 < |\lambda|$ (отталкивающая точка), и здесь всегда имеет место линейризуемость (применяем критерий устойчивости к f или к f^{-1}).

Теорема [Ko]. Пусть $f(z) = \lambda z + O(z^2)$ и $|\lambda| \neq 0, 1$. Тогда f линейризуемо.

1.4. Безразличная рациональная точка. Этот случай соответствует $|\lambda| = 1$, где λ — корень из единицы, $\lambda = e^{2\pi i \alpha}$, $\alpha = p/q \in \mathbb{Q}$, $p \in \mathbb{Z}$, $q \geq 1$ и $p \wedge q = 1$. Исследование динамики здесь весьма тонкое, и линейризуемость встречается довольно редко, как показывает следующее предложение.

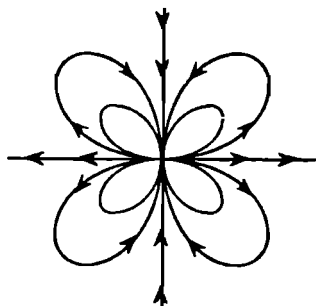


Рис. 0.

Предложение. *Отображение $f(z) = e^{2\pi ip/q} z + O(z^2)$ линеаризуемо тогда и только тогда, когда $f^q = \text{id}$.*

Доказательство. Если $f^q = \text{id}$, то $h = \left(\frac{1}{q} \sum_{i=0}^{q-1} \lambda^{-i} f^i \right)^{-1}$ линеаризует f . Обратно, если $h^{-1} \circ f \circ h(z) = \lambda z$, то $h^{-1} \circ f^q \circ h(z) = \lambda^q z = z$; поэтому $f^q = h \circ h^{-1} = \text{id}$. $\square$

Например, квадратичный многочлен $P_\lambda(z) = \lambda z(1 - z)$ не линеаризуем, так как P_λ^q — многочлен степени 2^q . Тем не менее топологическая динамика для f полностью понята после работ Фату и Лео [Le]. Что касается описания классов топологической сопряженности, мы отсылаем к работе Камачо [Ca]. Точки плоскости после полного оборота (f приближается поворотом на рациональный угол) описывают «цветок», как показано на рисунке. Существует целое число $n \geq 1$ и $2nq$ параболических областей (лепестков), которые переставляются отображением f с числом вращения p/q . При этом возникает $2n$ циклов из q элементов. На рис. 1 изображена динамика отображения f^q .

После работ Экаля и Воронина (см. [Ec, Vo, Ma]) известно полное описание классов аналитической сопряженности, которое, в противоположность уже рассмотренным случаям, много богаче, чем топологическая классификация.

1.5. Безразличная иррациональная неподвижная точка. Проблема Зигеля. Случай, который остается обсудить, это $\lambda = e^{2\pi i \alpha}$ при $\alpha \in \mathbb{R} - \mathbb{Q}$. Формальное линеаризующее отображение существует всегда, но этот ряд может расходиться ввиду наличия

малых знаменателей, которые возникают, если α хорошо приближается рациональными числами. Действительно, можно решить формально уравнение $h^{-1} \circ f \circ h(z) = \lambda z$ и найти $h(z) = \sum_{n=1}^{+\infty} h_n z^n$ с $h_1 = 1$ и

$$h_n = \frac{1}{\lambda^n - \lambda} \left[f_n + \sum_{p=2}^{n-1} f_p \sum_{\substack{j_1 + \dots + j_p = n \\ j_i \geq 1}} h_{j_1} \dots h_{j_p} \right].$$

Есть подозрение, что f может быть неустойчивым для очень «лиувиллевых» чисел α ; мы рассматриваем f как предел отображений $f_n(z) = e^{2\pi i p_n/q_n} z + \mathcal{O}(z^2)$, где $(p_n/q_n)_{n \geq 0}$ хорошо приближают α . Мы видели, что в случае «общего положения» безразличная рациональная неподвижная точка неустойчива. Проблема Зигеля, называемая также проблемой центра или функционального уравнения Шрёдера, состоит в том, чтобы исследовать множество значений α , для которых отображение всегда линеаризуемо.

Как мы убедимся, квадратичный многочлен $P_\lambda(z) = \lambda z(1 - z)$ (единственный с точностью до линейного сопряжения) играет важную роль в проблеме Зигеля: его динамика очень точно отражает динамику общего роста с постоянным мультипликатором λ .

2. КЛАССИЧЕСКИЕ РЕЗУЛЬТАТЫ ПО ПРОБЛЕМЕ ЗИГЕЛЯ

Несмотря на то, что существование нелинеаризуемых отображений, казалось бы, должно было быть известно математикам прошлого столетия, по-видимому, впервые оно было письменно упомянуто Г. А. Пфейфером в 1917 г. [Pf]. На самом деле легко показать, что общий квадратичный многочлен нелинеаризуем, т. е. существует плотное множество типа G_δ значений α , для которых линеаризующий ряд расходится (см. п. 5.1).

Первым, кто обратил внимание на важность арифметики числа α , был Г. Кремер (см. [Cr]). Пусть $(p_n/q_n)_{n \geq 0}$ обозначает последовательность подходящих дробей для α (эта последовательность определена в приложении, посвященном арифметике). Он доказал следующий результат:

Теорема (Кремер [Cr]). Если $\alpha \in \mathbb{R} - \mathbb{Q}$ удовлетворяет условию Кремера

$$\sup_{n \geq 0} \frac{\log q_{n+1}}{q_n} = +\infty, \quad (C)$$

то существует нелинеаризуемое отображение $f(z) = e^{2\pi i \alpha} z + O(z^2)$.

Доказательство элементарно и состоит в последовательном построении коэффициентов отображения f , начиная с первого. Положим $|f_n| = 1$ и положим аргумент отображения f_n равным аргументу второго члена в сумме внутри скобок в рекуррентном соотношении, определяющем последовательность $(h_n)_{n \geq 1}$ (см. п. 1.5). Тогда по меньшей мере

$$|h_n| \geq \frac{1}{|\lambda^n - \lambda|}.$$

Даже в случае, когда условие C не выполнено, получается оценка сверху радиуса сходимости ряда h . С помощью простых и красивых аргументов можно доказать такое утверждение:

Теорема [Cr]. Если $d \geq 2$ и $\alpha \in \mathbf{R} - \mathbf{Q}$ удовлетворяет условию Крмера степени d

$$\sup_{n \geq 0} \frac{\log q_{n+1}}{d^{q_n}} = +\infty, \quad (C_d)$$

то все ростки многочленов $P(z) = e^{2\pi i \alpha} z + \dots + a_d z^d$, $a_d \neq 0$, нелинеаризуемы.

Заметим, что $C_\infty = \bigcap_{d \geq 2} C_d$ — условие, позволяющее применить теорему ко всем многочленам независимо от степени, — записывается в виде

$$\sup_{n \geq 1} \frac{\log \log q_{n+1}}{q_n} = +\infty. \quad (C_\infty)$$

Имеем

$$C_\infty \subset \dots \subset C_3 \subset C_2 \subset C \subset \mathbf{R} - \mathbf{Q}.$$

Проблема существования значений α , для которых все ростки $f(z) = e^{2\pi i \alpha} z + O(z^2)$ линеаризуемы, является тонкой и долгое время оставалась открытой. Г. Жулиа [Ju] даже привел ошибочное доказательство их несуществования. Проблема была решена замечательным образом К.-Л. Зигелем в 1942 г.

Теорема (Зигель [Si], улучшена Брюно [Br]). Если α удовлетворяет арифметическому условию

$$\sum_{n \geq 0} \frac{\log q_{n+1}}{q_n} < +\infty, \quad (B)$$

то все отображения вида $f(z) = e^{2\pi i \alpha} z + O(z^2)$ линеаризуемы.

Множество чисел, удовлетворяющих арифметическому условию теоремы Зигеля 1942 г., имеет полную меру Лебега. Это первый результат, в котором преодолены трудности, связанные с малыми знаменателями. Доказательство основано на оценках коэффициентов линеаризующего отображения непосредственно из рекуррентного соотношения п. 1.5. Малые знаменатели оцениваются снизу с помощью диофантова условия. Позднее, в 50-х и 60-х годах, были открыты новые методы, которые привели к доказательству теорем К.А.М. По этому поводу читатель может обратиться к статье Боста [Bo] в трудах семинара Бурбаки. Эти методы применимы также к проблеме Зигеля, но дают более грубые результаты в том, что касается оценок областей линеаризуемости, и арифметические условия более ограничительны.

Что касается диофантовых условий, Зигель доказал линеаризуемость для всех $\alpha \in \mathbb{R} - \mathbb{Q}$, удовлетворяющих такому условию: существуют $\gamma, \tau > 0$, такие, что для любых рациональных p/q

$$|\alpha - p/q| \geq \frac{\gamma}{q^\tau}.$$

В 1965 г. А.Д. Брюно улучшил это арифметическое условие и усовершенствовал доказательство Зигеля. Одновременно и независимо Т. Черри [Ch] также сформулировал этот результат в качестве предположения и предположил, что условие B является оптимальным. Он не смог при жизни доказать это предположение, но его работа [Ch], хотя и очень неясная, показывает замечательное понимание динамики нелинеаризуемых отображений.

И только в 1987 г. Ж.-К. Йоккоз в своем мемуаре [Y1] доказал это предположение.

3. НОВЫЕ РЕЗУЛЬТАТЫ

3.1. О проблеме Зигеля

3.1.а. Предварительные замечания. Часть I.3 работы [Y1] содержит новое доказательство теоремы Зигеля—Брюно. Это новое геометрическое доказательство дает оптимальные оценки, с точностью до мультипликативной константы, радиуса сходимости линеаризующего отображения. Так как задача инвариантна относительно сопряжения гомотетиями, следует при формулировке результатов нормализовать наши отображения.

Допуская возможность сопряжения f посредством гомотетий, мы предполагаем, что $f \in S_\alpha$, где S_α — множество однолистных (голоморфных и инъективных) отображений единичного круга $\mathbf{D} = \mathbf{D}_1$, таких, что $f(z) = e^{2\pi i \alpha} z + \mathcal{O}(z^2)$. Хорошо известно, что пространство S_α компактно в топологии равномерной сходимости на компактных подмножествах. Мы уже обозначили через $R(f) \geq 0$ радиус сходимости линеаризующего отображения; положим

$$R(\alpha) = \inf_{f \in S_\alpha} R(f).$$

При $\alpha \in B$ положим

$$B(\alpha) = \sum_{n=0}^{+\infty} \frac{\log q_{n+1}}{q_n}.$$

Если $\alpha \notin B$, то $B(\alpha) = +\infty$. Разность между функцией B и функцией Φ , определенной в арифметическом приложении, ограничена универсальной константой там, где она конечна.

3.1.b. Теоремы о ростках. Вот основная теорема из [Y1]:

Теорема. 1. Пусть $\alpha \in \mathbf{R} - \mathbf{Q}$ и

$$\sum_{n=0}^{+\infty} \frac{\log q_{n+1}}{q_n} = +\infty.$$

Тогда существует нелинеаризуемое отображение $f(z) = e^{2\pi i \alpha} z + \mathcal{O}(z^2)$.

2. Существует универсальная константа $C > 0$, такая, что если $B(\alpha) < +\infty$, то $R(\alpha) > 0$ и

$$|B(\alpha) - \log(R(\alpha)^{-1})| \leq C.$$

Пункт 1 показывает, что условие B является оптимальным, а п. 2 улучшает результат Зигеля и Брюно, а именно оценку

$$\log(R(\alpha)^{-1}) \leq C_1 B(\alpha) + C_2,$$

где C_1 и C_2 — универсальные константы и можно положить $C_1 = 2$ (по [Hel]).

Как мы увидим далее, нелинеаризуемые отображения первой части теоремы конструируются «обращением» геометрического доказательства теоремы Зигеля. Это позволяет изучить некоторые черты динамики: нелинеаризуемые ростки имеют последовательность периодических орбит, стремящуюся к началу координат; можно также показать, что точки этих орбит находятся на существенно оптимальном (т. е. наименьшем возможном для орбит с таким периодом) расстоянии от 0. Конструкция обладает большой гибкостью. Она позволяет свободно выбирать динамику возвращения в окрестности периодических орбит. Это дает возможность доказать такую теорему:

Теорема. При $\alpha \notin B$ существует несчетное множество классов сопряженности ростков голоморфных отображений $f(z) = e^{2\pi i \alpha} z + O(z^2)$, не содержащих ни одной целой функции.

Мотивировки в пользу оптимальности условия B содержатся в приложении 2 из [Y1], где показано, что доказательство Брюно неприменимо для квадратичных многочленов при $\alpha \notin B$. В приложении 1 содержится изолированный интересный результат, который показывает, что теорема Зигеля о линеаризации (которая справедлива в высших размерностях при условии, что линейная часть диагонализуема, собственные значения все различны, имеют модуль 1, не резонансны и удовлетворяют диофантовым условиям) не выполняется, если линейная часть не диагонализуема.

Теорема. Пусть $n \geq 2$ и $A \in GL_n(\mathbb{C})$. Предположим, что A имеет собственное значение λ , по модулю равное 1, такое, что характеристическое подпространство, ассоциированное с λ , не совпадает с собственным подпространством. Тогда существуют нелинеаризуемые ростки голоморфных диффеоморфизмов $(\mathbb{C}^n, 0)$ с дифференциалом в нуле, равным A .

3.1.с. Результаты о квадратичных многочленах. До начала работы над теоремой Зигеля Йоккоз интересовался квадратичными многочленами. Уже результаты Кремера показали, что квадратичные многочлены являются в некотором смысле «менее линеаризуемыми», чем другие ростки многочленов. Это было уточнено Йоккозом:

Теорема. Если $\alpha \in \mathbb{R} - \mathbb{Q}$ и квадратичный многочлен $e^{2\pi i \alpha} z(1 - z)$ линеаризуем, то все ростки $f(z) = e^{2\pi i \alpha} z + O(z^2)$ линеаризуемы.

Кроме того, элементарным и независимым способом он доказал теорему Зигеля:

Теорема. Квадратичный многочлен $e^{2\pi i \alpha} z(1-z)$ линейризуем для множества значений α полной меры.

Заметим, что две последние теоремы влекут за собой теорему Зигеля для множества значений α полной меры. Их доказательства можно найти в [Y1] и [Do]. Доказательство первой теоремы связано с результатами о ростках из [Y1, partie II]:

Теорема. 1. При $\alpha \notin B$ квадратичный многочлен P_λ , $\lambda = e^{2\pi i \alpha}$, нелинеаризуем.

2. Существует универсальная константа $C_1 > 0$ и для любого $\varepsilon > 0$ константа $C(\varepsilon)$, такие, что при $\alpha \in B$

$$(1 - \varepsilon)B(\alpha) - C(\varepsilon) \leq \log(R(P_\lambda)^{-1}) \leq B(\alpha) + C_1.$$

Семейство квадратичных многочленов дает примеры нелинеаризуемых ростков при $\alpha \notin B$, и при $\alpha \in B$ радиус сходимости оказывается почти (с точностью до величины порядка ε) наименьшим возможным. Возникает естественный вопрос [Y1, II.1.9]:

Является ли отображение $\alpha \mapsto B(\alpha) - \log(R(P_{e^{2\pi i \alpha}})^{-1})$ ограниченным на B ?

В 1989 г. Йоккоз уточнил первый пункт предыдущей теоремы.

Теорема. Если $\alpha \notin B$, то квадратичный многочлен имеет периодические орбиты, отличные от 0 и содержащиеся в произвольной окрестности 0.

Доказательство можно найти в [PM1]. Те же аргументы применимы также к структурно устойчивым рациональным дробям (при фиксированном мультипликаторе λ в нуле).

Когда $|\lambda| < 1$, нуль является притягивающей неподвижной точкой линейризуемого отображения P_λ . Линейризующее отображение h_λ имеет единственную особую точку $U(\lambda)$ на границе своего круга сходимости и при этом $h_\lambda(U(\lambda)) = 1/2$.

Параграфы II.2 и II.3 в [Y1] посвящены изучению функции $\lambda \mapsto U(\lambda)$, голоморфной и ограниченной на D . Естественным считать, что радиальные пределы функции U при $\lambda \rightarrow \lambda_0 \in S^1 = \partial D$ дают информацию о P_λ . Можно доказать следующий результат:

Теорема. Для любых $\lambda_0 \in S^1$ модуль функции U имеет в λ_0 радиальный предел, равный радиусу сходимости отображения h_{λ_0} .

Это доказывает линейризуемость для п. в. значений α как следствие теоремы Фату, которая утверждает, что функция U , голоморфная и ограниченная в $\mathbf{D}$, имеет ненулевые радиальные пределы для п. в. точек границы. В [Y1] поставлены очень интересные вопросы о функции U , которая, кажется, обладает глубокими арифметическими свойствами.

3.2. Об аналитических диффеоморфизмах окружности. Проблема Зигеля близка к проблеме линейризации аналитических диффеоморфизмов окружности. Эвристическое соотношение сформулировано В. И. Арнольдом во введении к его статье [Ar1]:

«... Проблема центра есть особый случай задачи об отображении окружности, радиус которой, в особом случае, равен нулю...»

(Мы описываем в п. 6.1 конструкцию, делающую аккуратным это эвристическое соотношение между двумя проблемами.)

Йоккоз применил свою геометрическую технику к проблеме линейризации аналитических диффеоморфизмов окружности.

3.2.а. Предварительные замечания. Пусть $f : S^1 \rightarrow S^1$ — гомеоморфизм окружности, сохраняющий ориентацию; его главным динамическим инвариантом является число вращения $\rho(f) \in \mathbf{R}/\mathbf{Z}$, введенное А. Пуанкаре. Если f — диффеоморфизм, например, класса C^2 и $\rho(f)$ иррационально, то f топологически линейризуем по теореме Данжуа [He2]. Возникает вопрос о регулярности линейризующего отображения. Если f — аналитическое отображение, то теорема Арнольда [Ar1] (улучшенная Х. Рюссманом [Ru]) утверждает, что линейризующее отображение является аналитическим для почти всех чисел вращения, если предположить, что f близко к повороту (локальный результат). М. Эрман [He2] доказал глобальную теорему (без предположения близости к повороту) в классе C^∞ и в аналитическом случае, доказав тем самым одну гипотезу Арнольда. Диофантово условие в классе C^∞ было улучшено Ж.-К. Йоккозом [Y3], который дал оптимальное диофантово условие. В аналитическом случае оптимальное диофантово условие неизвестно.

В дальнейшем мы имеем дело с универсальным накрытием окружности S^1 , определенным отображением $E: \mathbf{R} \rightarrow S^1, E(z) = e^{2\pi iz}$. Для $\Delta > 0$ мы обозначаем через $S(\alpha, \Delta)$ пространство аналитических диффеоморфизмов окружности S^1 , поднятых на $\mathbf{R}$, имеющих число вращения $\alpha \in \mathbf{R}$ и продолжаемых до однолистного

отображения полосы $\{z \in \mathbb{C}; |\operatorname{Im} z| < \Delta\}$. Множество поднятых аналитических диффеоморфизмов окружности с числом вращения α обозначаем $S(\alpha, 0) = \bigcup_{\Delta > 0} S(\alpha, \Delta)$.

3.2.b. Теоремы о диффеоморфизмах окружности. Мы отсылаем читателя к арифметическому приложению за определениями арифметических условий $\mathcal{H}, \mathcal{H}(\Delta)$ и $\mathcal{B}$. Заметим просто, что они выражены вполне явно и что $\mathcal{H} \subset \bigcap_{\Delta > 0} \mathcal{H}(\Delta)$ и $\mathcal{B} = \bigcup_{\Delta > 0} \mathcal{H}(\Delta)$. В дальнейшем линейризуемость означает аналитическую линейризуемость. В [Y2] имеются результаты трех типов:

Глобальная теорема. 1. Если $\alpha \in \mathcal{H}$, то все аналитические диффеоморфизмы окружности с числом вращения α линейризуемы.

2. Если $\alpha \notin \mathcal{H}$, то существуют нелинейризуемые аналитические диффеоморфизмы окружности с числом вращения α .

Полулокальная теорема. Существует универсальная константа $C_0 (= 3)$, такая, что

1) если $\alpha \in \mathcal{H}(\Delta - C_0)$, $\Delta > C_0$ и $f \in S(\alpha, \Delta)$, то f линейризуемо;

2) если $\alpha \notin \mathcal{H}(\Delta - C_0)$, то существует нелинейризуемое отображение $f \in S(\alpha, \Delta)$.

Локальная теорема. 1. Если $\alpha \in \mathcal{B}$, $\Delta > \frac{1}{2\pi}\Phi(\alpha) + 3.15$ и $f \in S(\alpha, \Delta)$, то f линейризуемо и линейризирующее отображение принадлежит $S(1, \Delta - \frac{1}{2\pi}\Phi(\alpha) - 3.15)$.

2. Пусть $\alpha \in \mathcal{B}$ и для любых $0 < \delta' < \delta$ существует число $\varepsilon > 0$, такое, что если $|\operatorname{Im} z| < \delta$, то $|f(z) - z - \alpha| < \varepsilon$; тогда f линейризуемо и линейризирующее отображение принадлежит $S(1, \delta')$.

3. Если $\alpha \notin \mathcal{B}$, то дробь Бляшке $f_{a,\mu}(z) = \mu z^2 \frac{z+a}{1+az}$, $a > 3$, $\rho(f_{a,\mu}) = \alpha$, нелинейризуема.

Третья часть последней теоремы показывает, что условие $\mathcal{B}$ является оптимальным. Нелинейризуемость дробей Бляшке выводится из нелинейризуемости квадратичных полиномов с помощью конструкции Э. Гиса (см. [Gh] и [Do]).

4. МЕТОДЫ

4.1. Геометрическое доказательство теоремы Зигеля—Брюно. Доказательство основано на динамическом критерии линейризуемости — устойчивости.

В разных местах появляются универсальные константы. Их точное значение для наших рассуждений неважно. Мы обозначаем их $C, C_0, C_1 \dots$ и иногда применяем одно и то же обозначение для различных констант.

4.1.а. Отображение возвращения в секторе. Отправным пунктом является следующее замечание А. Дуади и Э. Гиса (см. [Do]): множество

$$\{\alpha \in \mathbb{R} - \mathbb{Q}; \text{ все } f(z) = e^{2\pi i \alpha} z + O(z^2) \text{ линеаризуемы}\}$$

инвариантно относительно действия группы $SL(2, \mathbb{Z})$, $\begin{pmatrix} a & b \\ c & d \end{pmatrix} x = \frac{ax+b}{cx+d}$.

Достаточно показать, что множество чисел Зигеля инвариантно при $\alpha \mapsto -\alpha^{-1}$. Доказательство геометрическое. Рассмотрим $f(z) = e^{2\pi i \alpha} z + O(z^2)$ и отрезок $l = [0, \varepsilon]$ с началом в 0. Дуги l и $f(l)$ образуют сектор в окрестности 0, который мы можем замкнуть отрезком l' , соединяющим некоторую точку из l с ее образом (см. рис. 1).

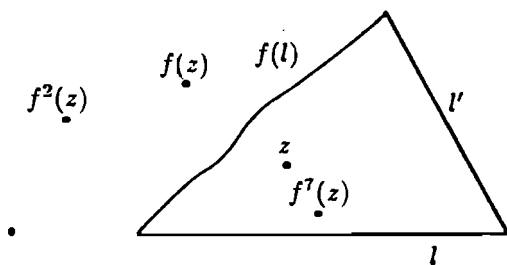


Рис. 1.

Если склеить l и $f(l)$ посредством f , получается топологическое многообразие с границей, гомеоморфное $\overline{\mathbb{D}}$. Легко видеть, что его внутренность биголоморфна $\mathbb{D}$ относительно индуцированной комплексной структуры. Отображение первого возвращения в склеенном секторе корректно определено и голоморфно в окрестности 0. Мы получаем росток $g(z) = e^{2\pi i \beta} z + O(z^2)$. Легко видеть, что $\beta = -\alpha^{-1}$ (чтобы убедиться, рассмотрите случай, когда f — поворот).

Динамика у отображения возвращения g , с точностью до некоторого числа итераций, та же самая, что и у f . В частности, f линеаризуемо тогда же, когда и g . Это доказывает инвариантность относительно действия $SL(2, \mathbf{Z})$. Можно утверждать большее: если f имеет периодическую орбиту, близкую к 0, то g также имеет периодическую орбиту, период и число вращения которой определяются по f . Далее, если f имеет «неустойчивую точку», т. е. близкую к 0 точку, убегающую из окрестности 0, то ей соответствует «неустойчивая точка» g . Это замечание и количественная реализация предшествующей конструкции приводят к теореме об устойчивости, т. е. линеаризуемости.

Начнем с нормализации $f \in S_\alpha$. Удобнее работать на универсальной накрывающей множества $\mathbf{D} - \{0\}$ с накрытием $E: \mathbf{H} = \{z \in \mathbf{C}; \operatorname{Im} z > 0\} \rightarrow \mathbf{D} - \{0\}$, $E(z) = e^{2\pi iz}$. Отображение $f \in S_\alpha$ поднимается до F , такого, что $E \circ F = f \circ E$ и

(1) $F: \mathbf{H} \rightarrow \mathbf{C}$ однолистно;

(2) $F(z) = z + \alpha + \phi(z)$, где ϕ есть $\mathbf{Z}$ -периодическая функция и $\lim_{\operatorname{Im} z \rightarrow +\infty} \phi(z) = 0$, т. е. ее разложение Фурье имеет вид

$$\phi(z) = \sum_{n \geq 1} \phi_n e^{2\pi i n z}.$$

Легко видеть, что ϕ экспоненциально убывает, когда $\operatorname{Im} z \rightarrow +\infty$. Обозначим через $S(\alpha)$ пространство таких отображений F . Оно компактно и является универсальной накрывающей множества S_α . Для построения отображения возвращения нужно взять вертикальную полупрямую и ее образ под действием F , образующие границы вертикальной полуполосы. Положим $\mathbf{H}_r = \mathbf{H} + ir$. Отображение F приближается переносом $z \mapsto z + \alpha$ в $\mathbf{H}_r$ при больших r . Если $\alpha \neq 0$, то в силу компактности пространства $S(\alpha)$ существует высота $t(\alpha)$, такая, что при $z \in \mathbf{H}_{t(\alpha)}$ выполняется фундаментальная оценка

$$|F(z) - z - \alpha| \leq \frac{\alpha}{4}. \quad (*)$$

Легко видеть, что можно всегда положить

$$t(\alpha) = \frac{1}{2\pi} \log \alpha^{-1} + C.$$

Наименьшее значение $t(\alpha)$, при котором удовлетворяется (*), выражает высоту показателей нелинейности отображения F . Выше этого значения F хорошо приближается переносом $z \mapsto z + \alpha$.

Показателями нелинейности F являются, например, неподвижные точки. В силу оценки (*) их мнимые части должны быть меньше $t(\alpha)$.

Теперь можно определить отображение возвращения в полуполосе B , ограниченной кривыми $l = [it(\alpha), +i\infty[, F(l)$ и $[it(\alpha), F(it(\alpha))]$. Если $z \in B$, мы итерируем посредством F , пока не получим $\operatorname{Re} F^n(z) > 1$. Если $\operatorname{Im} z \geq t(\alpha) + C_0$, то положим $z' = F^n(z) - 1 \in B$ и отображение $z \mapsto z'$ будем называть отображением возвращения в полуполосе B . Заметим, что нас интересует значение $t(\alpha)$ только с точностью до универсальной аддитивной константы, т. е. если мы будем рассматривать $t(\alpha) + C$ вместо $t(\alpha)$, то рассуждения не изменятся.

Склеивая границы l и $F(l)$ области B посредством F , получим риманову поверхность $\mathcal{S}$, биголоморфную $\mathbb{D} - \{0\}$. Если ввести подходящую униформизацию поверхности $\mathcal{S}$ на окрестности диска $\mathbb{D}$, то динамика возвращения будет определяться отображением $g \in S_{\alpha^{-1}}$ (для замены $-\alpha^{-1}$ на α^{-1} нужно изменить ориентацию); поднятым до $G \in S(\alpha^{-1})$. Можно показать, что если имеются точки $z \in \mathbb{H}$, убегающие из $\mathbb{H}$ под действием итераций отображения F , то соответствующие им точки $z' \in \mathbb{C}$ убегают из $\mathbb{H}$ под действием итераций отображения G и $\operatorname{Im} z'$ контролируется посредством $\operatorname{Im} z$. Более точно:

Предложение. Пусть $0 < \alpha < 1$, $F \in S(\alpha)$ и $t(\alpha) > 0$ таково, что если $\operatorname{Im} z \geq t(\alpha)$, то $|F(z) - z - \alpha| \leq \alpha/4$. Тогда определенное выше $G \in S(\alpha^{-1})$ обладает следующим свойством: если $z \in \mathbb{H}$, $\operatorname{Im} z \geq t(\alpha)$, $0 \leq i < n$, $F^i(z) \in \mathbb{H}$, но $F^n(z) \notin \mathbb{H}$, то существует $z' \in \mathbb{C}$, такое, что

- 1) $\operatorname{Im} z' \geq \alpha^{-1}(\operatorname{Im} z - t(\alpha) - C_0)$;
- 2) существует m , $0 \leq m < n$, такое, что $G^m(z') \notin \mathbb{H}$.

Оценка п. 1) получается при рассмотрении униформизации поверхности $\mathcal{S}$ методами теории однолистных функций или с помощью оценок модулей кольца, содержащегося в $\mathcal{S}$. В ней нет ничего удивительного: такая оценка получается, если F является трансляцией $z \mapsto z + \alpha$; в этом случае $t(\alpha) = 0$ и униформизация задается явным образом. А ведь на высоте $t(\alpha)$ F аппроксимируется этой трансляцией.

4.1.b. Аргумент бесконечного спуска. Применим теперь по индукции предыдущую конструкцию. Мы отправляемся от $\alpha \in$

$\mathbf{R} - \mathbf{Q}$, $0 < \alpha < 1$ и $F \in S(\alpha)$. Положим (см. приложение) $\alpha_0 = \alpha$ и при $n \geq 0$

$$\alpha_n^{-1} = a_{n+1} + \alpha_{n+1},$$

где $a_{n+1} \in \mathbf{N}^*$ и $0 < \alpha_{n+1} < 1$. Положим $\beta_{-1} = 1$ и $\beta_n = \alpha_0 \dots \alpha_n$.

Построим последовательность $(F_n)_{n \geq 0}$ с $F_n \in S(\alpha_n)$. Положим $F_0 = F$. Предположим, что мы построили F_n . Пусть $G_n \in S(\alpha_n^{-1})$ — отображение возвращения, построенное в п. 4.1.а. Положим $F_{n+1}(z) = G_n(z) - a_{n+1}$. Непосредственной индукцией с использованием предложения из п. 4.1.а получаем

Предложение. *Существует универсальная константа C_0 , такая, что если $z_0 \in \mathbf{H}$,*

$$\operatorname{Im} z_0 \geq \sum_{i=0}^m \beta_{i-1} t(\alpha_i) + C_0$$

и $F^{n_0}(z_0) \notin \mathbf{H}$, то существуют точки $(z_i)_{1 \leq i \leq m}$, $z_i \in \mathbf{H}$, и числа $(n_i)_{1 \leq i \leq m}$, такие, что $F_i^{n_i} \notin \mathbf{H}$ и

$$n_0 > n_1 > \dots > n_m \geq 0.$$

Отсюда выводим с помощью бесконечного спуска:

Теорема. *Если можно найти последовательность $(t(\alpha_n))_{n \geq 0}$, такую, что*

$$\sum_{n=0}^{+\infty} \beta_{n-1} t(\alpha_n) < +\infty,$$

то f устойчиво в 0 и, следовательно, линеаризуемо.

Заметим, что мы получаем круг радиуса

$$C \exp \left(-2\pi \sum_{n=0}^{+\infty} \beta_{n-1} t(\alpha_n) \right),$$

содержащийся в области устойчивости.

Напомним, что всегда можно положить

$$t(\alpha_n) = \frac{1}{2\pi} \log \alpha_n^{-1} + C.$$

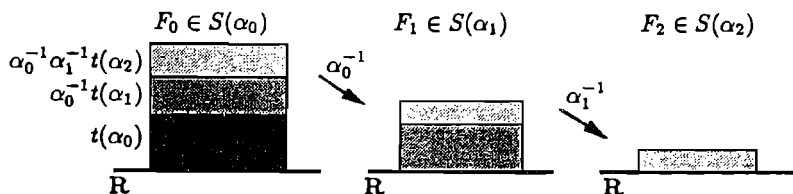


Рис. 2.

Так как $\sum_{n \geq 0} \beta_{n-1} \leq C$ (см. приложение), то универсальная константа, которая прибавляется к $t(\alpha_n)$, не имеет никакого значения и предыдущее арифметическое условие превращается в

$$\sum_{n=0}^{+\infty} \beta_{n-1} \log \alpha_n^{-1} < +\infty.$$

Это условие B (смотри арифметическое приложение) теоремы Зигеля—Брюно. Рисунок 2 схематически резюмирует идею доказательства.

На каждом этапе отображение возвращения F_{n+1} помнит о динамике отображения F_n только выше $t(\alpha_n)$. Отмеченные области, где информация о динамике пропадает, соответствуют на полуплоскости, где действует $F = F_0$, областям, которые «складываются» и толщина которых мажорируется величиной $\beta_{n-1}(t(\alpha_n) + C)$. Далее, если высота этих складывающихся областей не превосходят некоторой фиксированной величины, то точка, находившаяся выше и покидающая H за n_0 итераций отображения F , должна пересечь все эти области. Она соответствует точке, покидающей H за n_1 итераций отображения F_1 и т. д. Так как

$$n_0 > n_1 > \dots \geq 0,$$

получается противоречие.

4.2. Построение нелинеаризуемых ростков

4.2.а. Присоединение нелинейности. Принцип получения оптимальных диофантовых условий состоит в том, чтобы обратить конструкцию отображения возвращения. При переходе от F_{n+1} к

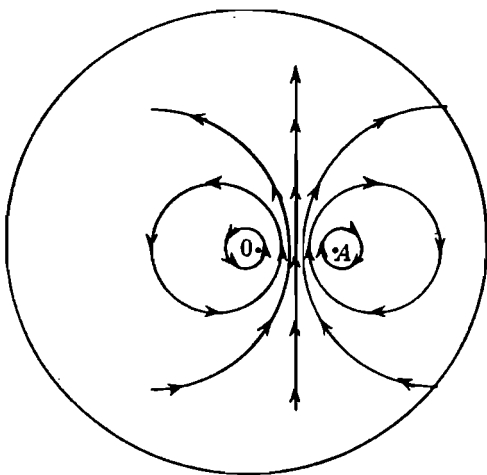


Рис. 3.

F_n нужно реконструировать нелинейности F_n выше $t(\alpha_n)$, препятствующие уменьшению $t(\alpha_n)$. «Типичная» нелинейность создается неподвижными точками с динамикой, подобной изображенной на рис. 3.

Напомним, что оценка (*) из п. 4.1.а показывает, что $t(\alpha_n) > 0$ мажорирует мнимую часть неподвижных точек отображения F_n . Мы покажем, как «приклеить» неподвижные точки к F_{n+1} . Но сначала опишем, как восстановить F_{n+1} без приклеивания неподвижных точек.

Мы отправляемся от $G \in S(\alpha)$ и хотим построить $F \in S(\beta)$ с $\beta = a + \alpha$, $0 < \alpha, \beta < 1$ и $a \in \mathbb{N}^*$ (напомним, что $\alpha_n^{-1} = a_{n+1} + \alpha_{n+1}$). Положим $\tilde{G}(z) = G(z) + a$, $\tilde{G} \in S(a + \alpha) = S(\beta^{-1})$. Имеем $\beta^{-1} > 1$, и выше некоторой высоты $C > 0$ $\tilde{G}$ приближается трансляцией $z \mapsto z + \beta^{-1}$. Рассмотрим полуполосу B' , ограниченную $l_1 = [iC, +i\infty[$, $\tilde{G}(l_1) = l_2$ и $[iC, \tilde{G}(iC)]$. При склеивании l_1 и l_2 посредством $\tilde{G}$ мы получаем риманову поверхность S' , являющуюся полуцилиндром, т. е. биголоморфную $\mathbb{D} - \{0\}$. Трансляция $T(z) = z + 1$, коммутирующая с $\tilde{G}$, индуцирует однолистное отображение на некоторой области в S' , являющейся окрестностью точки 0. Затем выбираем униформизацию $k : S' \rightarrow \mathbb{D} - \{0\}$, сопрягаем динамику, норма-

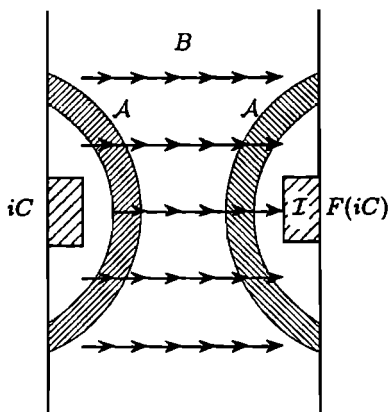


Рис. 4.

лизуем (однолистность на D) и поднимаем на H , чтобы получить динамику $F' \in S(\beta)$.

Построение, которое мы описали, обращает построение отображения возвращения, но не приклеивает неподвижных точек. Для того чтобы это сделать, продолжим полуцилиндр S' до цилиндра S . При дальнейшем построении мы рассматриваем полосу B , ограниченную $l_1, l_2, l_3 =] -i\infty, iC]$ и $l_4 = l_3 + (\tilde{G}(iC) - iC)$, см. рис. 4.

При построении есть определенная свобода: можно взять в качестве l_4 образ l_3 под действием произвольного однолистного отображения, близкого к $z \mapsto z + \beta^{-1}$ и коммутирующего с T . Это позволяет фиксировать класс аналитической сопряженности отображения возвращения в окрестности неподвижной точки, которую мы строим.

Теперь, склеивая l_1 и l_2 , как ранее, а затем l_3 и l_4 посредством отображения $z \mapsto z + (\tilde{G}(iC) - iC)$, мы получаем, как легко видеть, риманову поверхность S , являющуюся цилиндром, т.е. биголоморфную $\mathbb{C} - \{0\}$. Трансляция T определяет на S отображение, однолистное вне некоторой окрестности I точки, соответствующей iC ; пусть это будет $f : S - I \rightarrow S$ (рис. 5).

Как и ранее, униформируем S посредством отображения $k : S \rightarrow \hat{\mathbb{C}} - \{0, A\}$ так, что $k(I)$ остается вне D и точка A находится так близко к 0 , как возможно. Точка 0 соответствует $+i\infty$ на S , а точка A соответствует $-i\infty$. Эти требования необходимы для того, чтобы выполнялись следующие условия:

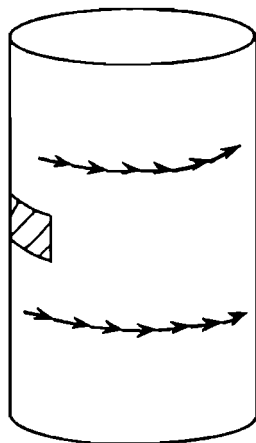


Рис. 5.

1. Сопряженное отображение $k \circ f \circ k^{-1}$ определено и однолистно на $\mathbf{D}$.

2. Нелинейность, соответствующая неподвижной точке A , приклеена настолько близко к началу, насколько возможно.

В этом месте необходимы оценки унифицирующего отображения k . Чтобы их получить, рассмотрим кольцо A , изображенное на рис. 4. Одна связная компонента его дополнения содержит $\mathcal{I}$, а другая содержит 0 и A . Оценивая максимальную длину пути, соединяющего две границы кольца, получим, что его модуль оценивается снизу величиной $\frac{1}{2\pi} \log \beta^{-1} - C$. Если мы выберем k так, что $k(\mathcal{T}) \cap \mathbf{D} \neq \emptyset$, то решение экстремальной задачи Тейхмюллера дает оценку $|A| \leq C\alpha$. Далее, поднимая отображение $k \circ f \circ k^{-1}$ на $\mathbf{H}$, получаем отображение $F \in S(\beta)$, откуда следует, что имеется неподвижная точка с мнимой частью порядка $\frac{1}{2\pi} \log \beta^{-1} - C$. Это, с точностью до универсальной константы, оптимальная высота, на которой можно надеяться приклеить неподвижную точку. На самом деле, как мы видели при доказательстве теоремы Зигеля—Брюно, всегда можно положить $t(\beta) = \frac{1}{2\pi} \log \beta^{-1} + C_0$.

Заметим, что оценки, полученные для униформизации, не завышены: такие же получаются, если мы рассматриваем $G(z) = z + \alpha$; в этом случае имеются явные выражения для k .

Итак, мы выполнили первое построение, G является отображением возвращения для F , и мы получаем также

Предложение. Пусть $0 < \alpha < 1, a \in \mathbf{N}^*, \beta = \frac{1}{a+\alpha}$ и $G \in S(\alpha)$. Тогда существует $F \in S(\beta)$, такое, что

- (1) F имеет неподвижную точку $z_0 \in \mathbf{H}$ с $\operatorname{Im} z_0 \geq \frac{1}{2\pi} \log \beta^{-1} - C$;
- (2) если G имеет периодическую орбиту с числом вращения p/q , т. е. $G^q(z) = z + p$ с $q \geq 1, p \in \mathbf{Z}$ и

$$\min_{1 \leq i \leq q} \operatorname{Im} G^i(z) \geq C_0,$$

то F имеет периодическую орбиту с числом вращения $1/(a+p/q)$ и

$$\min_{1 \leq j \leq aq+p} \operatorname{Im} F^j(z') \geq \beta \min_{1 \leq i \leq q} \operatorname{Im} G^i(z) + \frac{1}{2\pi} \log \beta^{-1} - C_0.$$

4.2.b. Индукция и переход к пределу. Как в нашем доказательстве теоремы Зигеля—Брюно, остается провести индукцию. Начиная с $0 < \alpha < 1$, определим последовательность $(\alpha_n)_{n \geq 0}$. Зафиксируем $n \geq 0$ и положим $F_{n,n+1} = z + \alpha_{n+1} \in S(\alpha_{n+1})$. Для этой трансляции мы присоединяем неподвижную точку, как описано в п. 4.2.a. Получаем отображение $F_{n,n} \in S(\alpha_n)$ и по индукции строим $F_{n,l} \in S(\alpha_l)$, отправляясь от $F_{n,l+1}$. Применяя индукцию к предложению из п. 4.2.a, получаем

Предложение. Пусть $m \geq 0$. Если $0 \leq n \leq m$ и p_n/q_n — хорошая подходящая дробь (см. арифметическое приложение), то $F_{m,0}$ имеет периодическую орбиту с числом вращения p_n/q_n и

$$\min_{1 \leq i \leq q_n} \operatorname{Im} F_{m,0}^i(z_n) \geq \frac{1}{2\pi} \sum_{i=0}^n \beta_{i-1} \log \alpha_i^{-1} - C.$$

Мы должны сделать предположение, что p_n/q_n — хорошая подходящая дробь, чтобы превратить неподвижную точку $F_{m,n} \in S(\alpha_n)$ в орбиту отображения $F_{m,0}$ с числом вращения p_n/q_n . Напомним, что $\alpha \notin \mathcal{B}$, если и только если множество хороших подходящих дробей бесконечно.

Наконец, мы выберем сходящуюся подпоследовательность из $(F_{n,0})_{n \geq 0} \subset S(\alpha)$. Предложение показывает, что при $\alpha \notin \mathcal{B}$ предельное отображение F имеет последовательность периодических орбит, накапливающуюся к 0, и, следовательно, нелинеаризуемо.

4.3. Модификация для диффеоморфизмов окружности. Здесь доказательства весьма техничны, и мы дадим только некоторые указания на изменения по сравнению со случаем ростков. Имея дело с диффеоморфизмами окружности, мы работаем в пространствах $S(\alpha, \Delta)$, и полезно воспринимать вещественную прямую «продолженной до $+\infty$ » по аналогии со случаем ростков. В действительности, когда Δ велико, все проходит, как если бы мы работали в пространствах $S(\alpha)$. Нужно просто все делать симметрично по отношению к вещественной оси.

Для аналитических диффеоморфизмов окружности устойчивость окружности S^1 также эквивалентна линейаризуемости. Доказательства линейаризуемости проводятся в том же стиле, что и в теореме Зигеля—Брюно, по крайней мере в том, что касается локальной теоремы. Мы предоставляем читателю продумать конструкцию отображения возвращения. В полулокальной и глобальной теоремах мы сталкиваемся с новой проблемой: что делать, когда Δ мало?

Действительно, в этом неблагоприятном случае у нас нет больше оценок на мнимые части орбитальных соответствий между F и его отображением возвращения. В случае ростков голоморфизмов мы «срезаем» высоту порядка $\frac{1}{2\pi} \log \alpha^{-1} + C_0$, когда $F \in S(\alpha)$. Для диффеоморфизмов окружности мы не можем действовать подобным образом, если $\Delta \leq \frac{1}{2\pi} \log \alpha^{-1} + C_0$. Напротив, если $\Delta \geq \frac{1}{2\pi} \log \alpha^{-1} + C_0$, мы получаем отображение возвращения $G \in S(\beta, \Delta')$ с $\Delta' = \alpha^{-1}(\Delta - \log \alpha^{-1})$ и $\alpha^{-1} = a + \beta, a \in \mathbb{N}^*, 0 < \beta < 1$. Если можно бесконечно продолжать действовать подобным образом и получить бесконечную последовательность $(F_n)_{n \geq 1}$, то применимо то же доказательство устойчивости, что и в теореме Зигеля, так как мы можем контролировать мнимые части орбитальных соответствий. Диофантовы условия линейаризуемости — это условия, при которых мы сталкиваемся с предыдущей трудностью конечное число раз. После преодоления этого конечного числа препятствий мы можем проводить доказательство, как обычно.

Для того чтобы разобраться с неблагоприятными случаями, Йоккоз доказывает комплексную лемму Данжуа, которую он использует при построении отображения возвращения и для того, чтобы получить большое расширение полосы Δ . Затем в неблагоприятном случае, отправляясь от $F_n \in S(\alpha_n, \Delta_n)$, $\Delta_n < \frac{1}{2\pi} \log \alpha_n^{-1}$, он получает отображение возвращения $F_{n+1} \in S(\alpha_{n+1}, \Delta_{n+1})$ с

$\Delta_{n+1} = Ce^{2\pi\Delta_n}$. Эти формулы объясняют определение условия $\mathcal{H}(\Delta)$.

Все это возможно, только если $F \in S(\alpha, \Delta)$ и Δ больше некоторой универсальной константы. Первый этап состоит в том, чтобы показать, что мы можем привести нашу ситуацию к этому случаю и начинать всегда с $\Delta_0 \geq 1000$. Можно доказать

Предложение. Пусть $f \in S(\alpha, 0)$ и $\Delta > 0$. Тогда существует матрица $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PGL(2, \mathbb{Z})$ и вещественно-аналитическое униформизирующее отображение $h: \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/(f^a \circ T^b)$ (факторпространство пространства $\mathbb{R}$ относительно действия $f^a \circ T^b$), такое, что $h^{-1} \circ (f^c \circ T^d) \circ h \in S\left(\frac{a\alpha+b}{c\alpha+d}, \Delta\right)$.

Полулокальная теорема и это предложение приводят к глобальной теореме. Доказательство предложения проводится в два этапа:

- Сначала показывается, что существует h , такое, что отображение $h^{-1} \circ (f^c \circ T^d) \circ h$ C^2 -близко к трансляции на $\mathbb{R}$.
- Затем используется комплексная лемма Данжуа, которую мы опустили; она позволяет контролировать в комплексной плоскости орбиты диффеоморфизмов. С ее помощью мы получаем существенное расширение полосы $\Delta > 0$.

В части, касающейся построения нелинеаризуемых диффеоморфизмов, мы также должны различать аналогичные случаи. Тем не менее геометрия всегда та же, что и для ростков голоморфизмов: мы приклеиваем неподвижные точки к цилиндрической римановой поверхности S .

5. ДРУГИЕ ПРИЛОЖЕНИЯ И РАСПРОСТРАНЕНИЕ МЕТОДОВ

5.1. Препятствие к линеаризуемости: накопление периодических орбит. Если $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$, $\alpha \in \mathbb{R} - \mathbb{Q}$, — линеаризуемое отображение, то у него нет периодических орбит, отличных от 0, в окрестности начала координат. Это дает в некоторых случаях критерий линеаризуемости. Проиллюстрируем это на примере.

Если $\lambda = e^{2\pi ip/q}$, $p \in \mathbb{Z}$, $q \geq 1$, $p \wedge q = 1$, то квадратичный многочлен P_λ имеет в 0 неподвижную точку порядка q ; это означает, что малым возмущением λ порождается периодическая орбита периода q в окрестности 0. Следовательно, при $n \geq 1$ существует окрестность $U(p/q, n)$ точки p/q на $\mathbb{R}$, такая, что при

$\alpha \in U(p/q, n)$ многочлен P_λ имеет периодическую орбиту периода q , содержащуюся в $D_{1/n}$. Множество $U_n = \bigcup_{p/q \in \mathbb{Q}} U(p/q, n)$ плотно и открыто. По теореме Бэра мы получаем плотное множество $U = \left(\bigcap_{n \geq 1} U_n\right) \cap (\mathbb{R} - \mathbb{Q})$ типа G_δ значений α , таких, что квадратичный многочлен P_λ , $\lambda = e^{2\pi i \alpha}$, имеет последовательность периодических орбит, сходящуюся к 0, и, следовательно, не линеаризуем.

Возникает мысль, что это топологическое препятствие может быть единственным, что мешает линеаризуемости, т.е. все нелинеаризуемые ростки имеют последовательность периодических орбит, накапливающуюся к 0. Аналогичной была гипотеза П. Сада [Sa] и В. И. Арнольда [Ar2] для диффеоморфизмов окружности. Ж.-К. Йоккоз руководил моей диссертацией на эту тему. Имеем:

Теорема [PM1]. 1. Если $\alpha \in \mathbb{R} - \mathbb{Q}$ удовлетворяет условию $\sum_{n \geq 1} \frac{\log \log q_{n+1}}{q_n} < +\infty$ и $f(z) = e^{2\pi i \alpha} z + O(z^2)$ — нелинеаризуемое отображение, то f имеет накапливающуюся к 0 последовательность периодических орбит, периоды которых образуют подпоследовательность $(q_{n_k})_{k \geq 0}$, такую, что

$$\sum_{k \geq 0} \frac{\log q_{n_k+1}}{q_{n_k}} = +\infty.$$

2. Если $\alpha \in \mathbb{R} - \mathbb{Q}$ удовлетворяет условию $\sum_{n \geq 1} \frac{\log \log q_{n+1}}{q_n} = +\infty$, то существует $f \in S_\alpha$, такое, что все положительные орбиты $(f^n(z))_{n \geq 0}$, содержащиеся в D , накапливаются к 0. Следовательно, f нелинеаризуемо и не имеет периодических орбит в $D - \{0\}$.

В [PM1] построены также нелинеаризуемые диффеоморфизмы окружности без периодических орбит. Диофантово условие на число вращения, для которого возможны такие примеры, довольно искусственно. Тем не менее оказывается, что можно получать такие диффеоморфизмы окружности, отправляясь от ростков голоморфизмов.

Теорема ([PM1], недавно завершено). Если $\alpha \in \mathbb{R} - \mathbb{Q}$ удовлетворяет условию

$$\sum_{n \geq 1} \frac{\log \log q_{n+1}}{q_n} = +\infty$$

для всех $\Delta > 0$, то существует аналитический диффеоморфизм окружности f , поднятие которого принадлежит $S(\alpha, \Delta)$, такой, что все положительные орбиты $(f^n(z))_{n \geq 0}$, остающиеся в $U = D_{e^{2\pi\Delta}} - \overline{D_{e^{-2\pi\Delta}}}$, накапливаются к S^1 . Следовательно, f нелинеаризуемо и не имеет периодических орбит в U .

Доказательство п. 1 теоремы о ростках голоморфизмов аналогично доказательству Йоккоза теоремы Зигеля—Брюно. Нужно сделать два добавочных замечания. Во-первых, неподвижная точка отображения $F_n \in S(\alpha_n)$ (см. п. 4.1.b) соответствует периодической орбите отображения F с числом вращения p_n/q_n , которое, как мы знаем, не превышает мнимой части этой точки. Во-вторых, если $F_n \in S(\alpha_n)$ не имеет неподвижных точек, то высота показателей нелинейности весьма невелика. Мы получаем фундаментальную оценку (*) из п. 5.1.a, для которой

$$t(\alpha_n) = \frac{1}{2\pi} \log^+ \log \alpha_n^{-1} + C_0.$$

Пункт 2 теоремы соответствует конструкции нелинеаризуемых ростков, когда $\alpha \notin \mathcal{B}$. Однако нужно полностью изменить геометрию присоединяемой нелинейности. Нам не нужно теперь приклеивать неподвижные точки, желательная нелинейность строится при помощи поверхности S рис. 6.

Мы модифицируем цилиндр Йоккоза (рис. 5) тем, что делаем разрез и приклеиваем бесконечное число плоскостей к одной и другой стороне. Очень приблизительно можно сказать, что мы разрываем неподвижную точку A и получаем точку бесконечного ветвления.

5.2. Симметрии нелинеаризуемых ростков. Рассмотрим группу симметрий ростка голоморфизма f ,

$$\text{Cent}(f) = \{g(z) = \mu z + \mathcal{O}(z^2); \mu \in \mathbb{C}^*, g \circ f = f \circ g\}.$$

Заметим, что централизатор отображения f , $\text{Cent}(f)$, содержит подгруппу, образованную итерациями отображения f и обозначаемую через $f^{\mathbb{Z}}$. Если $df(0) \in \mathbb{C}^*$ не является корнем из единицы, то $\text{Cent}(f)$ вкладывается в $\mathbb{C}^*$ при помощи инъективного отображения

$$\text{Cent}(f) \longrightarrow \mathbb{C}^*,$$

$$g \longmapsto \mu = Dg(0).$$

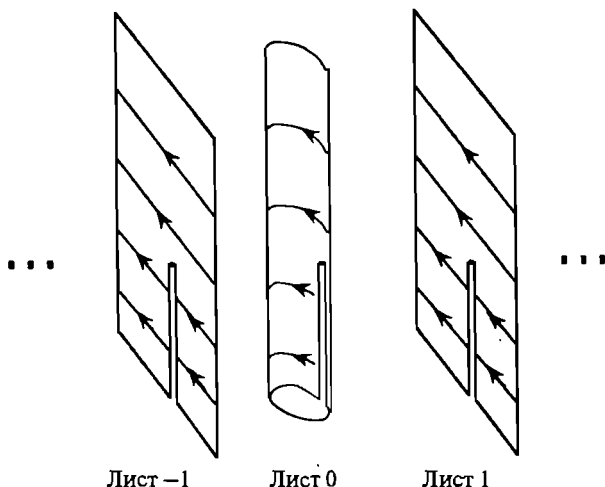


Рис. 6.

Исследование группы симметрий отображения f является обобщением проблемы Зигеля, как показывает следующее предложение.

Предложение. Если $Df(0) \in \mathbb{C}^*$ не является корнем из единицы, то f линеаризуемо тогда и только тогда, когда

$$\text{Cent}(f) = \mathbb{C}^*.$$

В нелинеаризуемом случае легко получаем

Предложение. Пусть f нелинеаризуемо, $f(z) = e^{2\pi i \alpha} z + \mathcal{O}(z^2)$, $\alpha \in \mathbb{R} - \mathbb{Q}$. Тогда $f^{\mathbb{Z}} \subset \text{Cent}(f) \subset S^1$ и множество $\text{Cent}(f)$

- плотно в S^1 ,
- типа F_σ и, следовательно, борелевское в S^1 ,
- имеет нулевую меру Хаара в S^1 .

Если $g \in \text{Cent}(f)$ и $Dg(0)$ не является корнем из единицы, то g нелинеаризуемо.

Если $g \in \text{Cent}(f)$ и $Dg(0)$ — корень из единицы порядка q , то $g^q = \text{id}$.

Естественно поставить задачу о строении группы $\text{Cent}(f)$ для нелинеаризуемого f : какие подгруппы из S^1 имеют форму $\text{Cent}(f)$?

Д. Мозер дает некоторые ограничения:

Теорема [Mo]. Пусть $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$ и $g \in \text{Cent}(f)$ с $Dg(0) = e^{2\pi i\beta}$. Предположим, что существуют $\gamma, \tau > 0$, такие, что при $q \geq 1$ и $p \in \mathbb{Z}$

$$\max(|q\alpha - p|, |q\beta - p|) \geq \frac{\gamma}{q^\tau}.$$

Тогда f линеаризуемо.

М. Эрман, который построил несчетные централизаторы в классе гладких нелинеаризуемых диффеоморфизмов окружности ([He2], см. также [Y4] о других результатах в классе диффеоморфизмов), интересовался, существуют ли такие объекты в классе аналитических отображений. Это показало бы, что диофантовы условия являются необходимыми в теореме Мозера, и дало бы примеры ростков f , таких, что $f^{\mathbb{Z}}$ не является подгруппой конечного индекса в $\text{Cent}(f)$. Оказывается, что геометрический метод присоединения нелинейностей из п. 5.1 позволяет строить такие примеры.

Теорема [PM2]. Существует канторово множество $K \subset \mathbb{T} = \mathbb{R}/\mathbb{Z}$ и коммутативное семейство отображений $(f_t)_{t \in K}$, таких, что $f_t \in S(t)$ и f_t нелинеаризуемо при $t \in K \cap (\mathbb{R} - \mathbb{Q})/\mathbb{Z}$.

Множество $K \cap (\mathbb{Q}/\mathbb{Z})$ плотно в K . Канторово множество K порождает несчетную подгруппу $G \subset \mathbb{T}$, такую, что $G \subset \text{Cent}(f)$ при $t \in K$. Конструкция достаточно гибкая и позволяет наложить дополнительные условия: можно потребовать, чтобы все $(f_t)_{t \in K}$ имели одну и ту же последовательность периодических орбит, накапливающуюся к 0, или чтобы f_t не имели ни одной периодической орбиты, отличной от 0, при $t \in K \cap (\mathbb{R} - \mathbb{Q})/\mathbb{Z}$.

При построении мы начинаем с семейства вращений и присоединяем нелинейности того же «типа», что и на рис. 6. При каждом приклеивании исчезают некоторые интервалы и при переходе к пределу остается только компакт K . Средство, используемое для того, чтобы обойти трудности, состоит в том, что мы одновременно приклеиваем нелинейности к нескольким элементам семейства, благодаря чему числа, принадлежащие $K \cap (\mathbb{R} - \mathbb{Q})/\mathbb{Z}$, являются очень «лиувиллевыми».

6. ЭКВИВАЛЕНТНОСТЬ НЕКОТОРЫХ ПРОБЛЕМ

В этом последнем разделе рассматриваются недавние конструкции — полученные после 1991 г. Хотя и не связанные непосредствен-

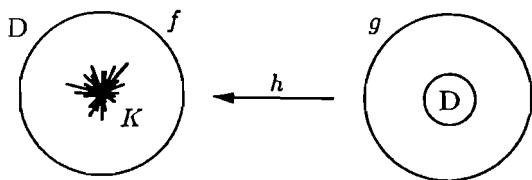


Рис. 7.

но с предшествующей геометрической техникой, они заслуживают упоминания, так как объединяют три задачи, связанные с малыми знаменателями. Результаты, упоминавшиеся ранее, переходят одни в другие.

6.1. Как получать диффеоморфизмы окружности, отправляясь от голоморфных диффеоморфизмов пространства $(\mathbb{C}, 0)$. Если $f(z) = e^{2\pi i \alpha} z + O(z^2)$ линейризуемо или $\alpha \in \mathbb{Q}$, то легко видеть, что в любой окрестности 0 имеются связные компакты K , отличные от $\{0\}$ и такие, что $f(K) = K$. Оказывается, что это общее свойство:

Теорема (П.-М.). Пусть $f \in S(\alpha)$, $\alpha \in \mathbb{R}$ и $0 \leq r \leq 1$. Тогда существует связный компакт K , $0 \in K \subset \overline{D_r}$, такой, что $K \cap \partial D_r \neq \emptyset$ и

$$f(K) = K.$$

Наибольший интерес представляет случай, когда f нелинеаризуемо, и при этом топология компакта K может быть сложной. Г. Биркгоф [Bi] показал, что существует K , удовлетворяющее условию $f(K) \subset K$. Теорема доказывается для линейризуемого f (или $\alpha \in \mathbb{Q}$), а затем мы осуществляем переход к пределу, рассматривая возмущения отображения f .

Когда это установлено, можно отметить, что если $r \neq 0$, то $D - K$ является кольцом, биголоморфным $D_R - \overline{D}$ (рис. 7). Пусть $h: D_R - \overline{D} \rightarrow b f D$ есть K -конформная эквивалентность.

Сопрягая f посредством h , получаем отображение $g = h^{-1} \circ f \circ h$, которое является гомеоморфизмом некоторой окрестности окружности S^1 в $D_R - \overline{D}$. Легко видеть, что g непрерывно продолжается на S^1 , $g(S^1) = S^1$; поэтому по принципу отражения Шварца g определяет аналитический диффеоморфизм окружности S^1 . Можно показать, что $\rho(g) = \alpha = \rho(f)$. Ясно, что если g линейризуемо, то и

f линеаризуемо, и если f не имеет периодических орбит, отличных от 0, то и g не имеет. Далее, если мы выберем $\tau > 0$ достаточно малым, то полученный диффеоморфизм g продолжается однолистно в кольцо, содержащее S^1 , и модуль этого кольца может быть сделан сколь угодно большим.

Мы можем теперь преобразовывать результаты о ростках в результаты о диффеоморфизмах окружности и наоборот. Более точно, прямые результаты о диффеоморфизмах окружности приводят к соответствующим результатам для ростков голоморфизмов. Пример: из локальной теоремы Йоккоза о диффеоморфизмах окружности следует теорема Зигеля—Брюно. Контрпримеры, построенные для ростков голоморфизмов, приводят к аналогичным примерам для диффеоморфизмов окружности. Пример: если f нелинеаризуемо и не имеет периодических орбит, отличных от 0, то и g таково. Мы видим, что локальные проблемы для диффеоморфизмов окружности эквивалентны аналогичным проблемам для ростков голоморфизмов. Можно также рассматривать глобальные теоремы о диффеоморфизмах окружности в рамках теории ростков голоморфизмов

6.2. Как получать особенности голоморфных слоений, отпавляясь от ростков голоморфизмов пространства $(C, 0)$. Обратная процедура является классической. Для $\alpha > 0$ рассмотрим слоение окрестности $(C^2, 0)$ с особенностью в начале, определенное формой

$$\omega = x(1 + \dots)dy + \alpha y(1 + \dots)dx,$$

где многоточия обозначают члены высшего порядка. Проколота ось $\{y = 0\} - \{0\}$ является слоем. Для достаточно малого $\epsilon > 0$ петля $\gamma : [0, 1] \rightarrow \{y = 0\}$, $\gamma(t) = (\epsilon e^{2\pi i t}, 0)$, определяет голономию (отображение последования) в точке $(\epsilon, 0)$ на малом диске, трансверсальном $\{y = 0\}$. Мы получаем класс сопряженности ростка голоморфизма $f(z) = e^{2\pi i \alpha} z + O(z^2)$ и отображение H_α классов сопряженности относительно голоморфизмов пространства $(C^2, 0)$ слоений указанного типа в классы сопряженности ростков голоморфизмов пространства $(C^2, 0)$.

Ж.-Ф. Матье и Р. Муссю показали [М-М], что H_α всегда инъективно. Ж. Мартине и Ж.-П. Рамис доказали сюръективность для $\alpha \in \mathbb{Q}$, а результаты А. Д. Брюно о линеаризации [Br] дают этот результат для $\alpha \in \mathbb{B}$ (имеется один класс сопряженности). В действительности верна

Теорема [Y-P-M]. H_α сюръективно.

Это значит, что всегда можно реализовать росток $f(z) = e^{2\pi i \alpha} z + O(z^2)$ как голономию слоения описанного типа. Для доказательства рассмотрим универсальное накрытие пространства $\mathbb{C}^2 - (\{x = 0\} \cup \{y = 0\})$, задаваемое отображением $E_2 : (z_1, z_2) \mapsto (e^{2\pi i z_1}, e^{2\pi i z_2})$. Рассмотрим слоение пространства $\mathbb{C}^2$ на прямые $z_1 + z_2 = \text{const}$. После факторизации $\mathbb{C}^2$ по действию $T_1 : (z_1, z_2) \mapsto (z_1 + 1, z_2)$ и $T_2 : (z_1, z_2) \mapsto (z_1, z_2 + 1)$ образ при E_2 будет слоением на прямые, задаваемым формой $\omega = xdy + \alpha ydx$. Голономия этого слоения линейризуема.

Пусть $f(z) = e^{2\pi i \alpha} z + O(z^2) \in S_\alpha$ — росток, который мы хотим реализовать, и $F \in S(\alpha)$ — его поднятие на $\mathbb{H}$: $F(z) = z + \alpha + \phi(z)$. Мы реализуем этот росток как голономию слоения, индуцированного линейным слоением, на комплексном многообразии M , полученном склеиванием сторон фундаментальной области для действий $F_1 : (z_1, z_2) \mapsto (z_1 + 1, z_2 + \phi(z_2 + \alpha z_1))$ и $F_2 : (z_1, z_2) \mapsto (z_1, z_2 + 1)$. Мы рассматриваем эти отображения только в области $\{(z_1, z_2) \in \mathbb{C}^2; \text{Im}(z_2 + \alpha z_1) \geq C_0\} \subset \mathbb{C}^2$, в которой F_1 и F_2 близки к T_1 и T_2 . Проблема состоит в том, чтобы показать, что абстрактное многообразие M биголоморфно окрестности начала координат в $\mathbb{C}^2$ с выброшенными осями координат.

Это верно для C^∞ -структуры: существует C^∞ -униформизация нужного нам вида, так как можно сопрячь F_1 и F_2 с T_1 и T_2 посредством C^∞ -отображения. Отличие униформизирующих координат от голоморфных при этом мало (мало $\bar{\delta}$). Если склеивание сторон фундаментальной области произведено хорошо, так что M становится многообразием Штейна, то можно применить метод Л. Хёрмандера [Ho] для решения $\bar{\delta}$ -уравнения с оценками. Можно теперь исправить C^∞ -униформизацию так, что она станет голоморфной. Так как мы хотим сохранить голоморфное слоение, то мы снова решаем второе $\bar{\delta}$ -уравнение для формы, определяющей это слоение.

7. АРИФМЕТИЧЕСКОЕ ПРИЛОЖЕНИЕ

7.1. Разложение в непрерывную дробь. Напомним сначала обычное разложение в непрерывную дробь (см. [La]). Для целого числа x мы обозначаем через $[x]$ его целую часть и через $\{x\} = x - [x]$ — его дробную часть.

Пусть $\alpha \in \mathbb{R} - \mathbb{Q}$. Положим $a_0 = [\alpha] \in \mathbb{Z}$, $\alpha_0 = \{\alpha\}$ и при $n \geq 1$

$$a_n = [\alpha_{n-1}^{-1}], \quad \alpha_n = \{\alpha_{n-1}^{-1}\}.$$

Пусть

$$q_{-2} = p_{-1} = 1,$$

$$q_{-1} = p_{-2} = 0,$$

и при $n \geq 0$

$$q_n = a_n q_{n-1} + q_{n-2},$$

$$p_n = a_n p_{n-1} + p_{n-2}.$$

Последовательность рациональных чисел $(p_n/q_n)_{n \geq 0}$ образует последовательность подходящих дробей для α . Последовательность знаменателей растет тем быстрее, чем лучше α приближается рациональными числами. Далее, определим последовательность $(\beta_n)_{n \geq 1}$: положим $\beta_{-1} = 1$ и при $n \geq 1$

$$\beta_n = \prod_{i=0}^n \alpha_i.$$

Последовательность $(\beta_n)_{n \geq 0}$ всегда убывает не медленнее геометрической прогрессии. Без труда проверяется, что при $n \geq -2$

$$\beta_n = (-1)^n (q_n \alpha - p_n),$$

$$q_{n+1} \beta_n + q_n \beta_{n+1} = 1.$$

Ж.-К. Йоккоз ввел и использовал в [Y1] и [Y2] вариант этого классического разложения. Положим $\|x\| = \inf_{p \in \mathbb{Z}} |x - p|$, $\alpha_0 = \|\alpha\|$ и при $n \geq 1$

$$\alpha_n = \|\alpha_{n-1}^{-1}\|.$$

Величины β_n определим так же, как и ранее, по α_n . Этот вариант имеет то преимущество, что $0 < \alpha_n < 1/2$, а это облегчает геометрические конструкции. Если при обычном определении $0 < \alpha_n < 1/2$, то два разложения в непрерывные дроби совпадают. Для облегчения изложения мы рассматриваем далее только обычное разложение. Разница имеет значение только при формулировке арифметических условий для диффеоморфизмов окружности (где нужно брать модифицированные α_n) и при определении явных значений констант в формулировках теорем о диффеоморфизмах окружности.

Определим функцию

$$\Phi(\alpha) = \sum_{i=0}^{+\infty} \beta_{i-1} \log \alpha_i^{-1}$$

и отображение

$$\tau_{\alpha}(x) = \begin{cases} e^x & \text{при } 0 \leq x \leq \log \alpha^{-1}, \\ \alpha^{-1}(x - \log \alpha^{-1} + 1) & \text{при } x \geq \log \alpha^{-1} \end{cases}$$

7.2. Арифметические условия.

7.2.a. Для ростков. Легко видеть, что условие

$$\sum \frac{\log q_{n+1}}{q_n} < +\infty$$

эквивалентно условию

$$\Phi(\alpha) < +\infty.$$

Обозначим через B множество чисел, удовлетворяющих этим условиям. Это — множество типа F_{σ} на $\mathbf{R}$.

Для $C_0 > 0$ и $0 \leq i \leq n$ положим

$$B_{n,l} = \beta_{l-1}^{-1} \left[\sum_{j=l} n \beta_{j-1} \log \alpha_j^{-1} - C_0 \sum_{j=l}^n \beta_{j-1} \right].$$

Мы говорим, что p_n/q_n — хорошая подходящая дробь, если при $0 \leq l \leq n$

$$B_{n,l} \geq \log \alpha_i^{-1}.$$

Это означает, что дроби p_n/q_n хорошо приближают α . Без труда можно показать, что $\alpha \notin B$, если и только если α обладает бесконечным числом хороших подходящих дробей.

7.2.b. Для диффеоморфизмов окружности. Пусть $\Delta > 0$. Определим последовательность $(\Delta_n(\alpha))_{n \geq 0}$ следующим образом:

$$\Delta_0(\alpha) = 2\pi\Delta,$$

$$\Delta_{n+1}(\alpha) = \tau_{\alpha_n}(\Delta_n(\alpha)).$$

Множество $\mathcal{H}(\Delta)$ состоит из чисел $\alpha \in \mathbf{R} - \mathbf{Q}$, для которых, начиная с некоторого номера, $\Delta_n(\alpha) \geq \log \alpha_n^{-1}$. Это множество типа F_{σ} , и можно проверить, что

$$\bigcup_{\Delta > 0} \mathcal{H}(\Delta) = B.$$

Наконец, множество $\mathcal{H}$ состоит из чисел α , таких, что $\alpha_n \in \mathcal{H}(\Delta)$ при всех $n \geq 0$. Это множество больше не зависит от выбранной нами константы $\Delta > 0$ и инвариантно под действием группы $SL(2, \mathbf{Z})$, т. е. оно определяет диофантово условие. $\mathcal{H}$ является множеством типа $F_{\sigma, \delta}$, т. е. это счетное пересечение множеств типа F_{σ} .

ЛИТЕРАТУРА

- [Ar1] Арнольд В. И. Малые знаменатели. I. Об отображениях окружности на себя. — Изв. АН СССР, серия матем., т. 25, 1961, № 1, с. 21–86.
- [Ar2] Арнольд В. И. Дополнительные главы теории обыкновенных дифференциальных уравнений. — М.: Наука, 1978.
- [Bi] Birkhoff G. D. Surface transformations and their dynamical applications, *Acta Math.*, **43** (1920), 1–119; *Oeuvres compl.*, t. II, 11–129.
- [Bo] Bost J.-B. Tores invariants des systèmes dynamiques hamiltoniens (d'après Kolmogorov, Arnold, Moser, Rüssman, Zehnder, Herman, Pöschel, ...), *Sémin. Bourbaki*, exp. 639, 1984–85; *Astérisque*, **133–134** (1985), 113–157.
- [Bot] Боткер Л. Е. Основные законы сходимости итераций и их применение в анализе. — Изв. Казан. физ.-мат., т. 14, 1904.
- [Br] Брюно А. Д. Аналитическая форма дифференциальных уравнений. — Труды ММО, т. 25, 1971, с. 119–262; т. 26, 1972, с. 199–238.
- [Ca] Camacho C. On the local structure of conformal mappings and holomorphic vector fields, *Asterisque*, **59–60** (1978), 83–84.
- [Cr1] Cremer H. Über die Iteration rationaler Funktionen, *Jahresber. Deutsch Math. Verein*, **33** (1925), 185–210.
- [Cr2] Cremer H. Zum Zentrumproblem, *Math. Ann.*, **98** (1928), 151–153.
- [Cr3] Cremer H. Über das Zentrumproblem (Mit besonderer Berücksichtigung der Lückenreichen), *Ber. Math. Phys. Klasse der Säch. Akad. Wiss. Leipzig*, **82** (1930), 243–250.
- [Cr4] Cremer H. Über das Schrödersche Funktionalgleichung und das Schwarzsche Eckenabbildungsproblem, *ibid.*, **84** (1932), 291–324.
- [Cr5] Cremer H. Über die Häufigkeit der Nichtzentren, *Math. of Ann.*, **115** (1938), 573–580.
- [Ch] Cherry T. M. A singular case of iteration of an analytic function: a contribution to the small divisor problem, in: *Nonlinear Problems of Engineering*, ed. W. F. Ames, Academic Press, New York, 1964, 29–50.
- [Do] Douady A. Disques de Siegel et anneaux de Herman, *Sém. Bourbaki*, exp. 677, 1986–87, *Astérisque*, **152–153** (1987), 151–172.
- [Ec] Écalle J. Les fonctions résurgentes et leurs applications, t. I, II, III, *Publications mathématiques d'Orsay* 81–05, 81–06, 85–05.
- [Gh] Ghys E. Transformations holomorphes au voisinage d'une courbe de Jordan, *C. R. Acad. Sci., série I*, **298**, No. 16 (1984), 385–388.
- [He1] Herman M. R. Recent results and some open questions on Siegel's linearisation theorem of germs of complex analytic diffeomorphisms of $\mathbb{C}^n$ near a fixed point, in: *Proceedings VIIIth Int. Conf. Math. Phys.*, World Scientific Publishers, Singapore, 1987.

- [He2] Herman M. R. Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations, Publ. Math. I.H.E.S., **49** (1979), 5–233.
- [Ju] Julia G. Oeuvres, vol. 1, Gauthiers-Villars, Paris, 1969, 231–232.
- [Ko] Koenigs G. Recherches sur les intégrales de certaines équations fonctionnelles, Ann. Sci. Ecole Norm Sup. (3 série), **1** (1884), suppl., 1–41.
- [Ho] Hörmander L. An Introduction to Complex Analysis in Several Variables, Van Nostrand, 1966. [Имеется перевод: Хёрмандер Л. Введение в теорию функций нескольких комплексных переменных. — М.: Мир, 1968.]
- [La] Lang S. Introduction to Diophantine Approximations, Addison-Wesley, 1966. [Имеется перевод: Ленг С. Введение в теорию диофантовых приближений. — М.: Мир, 1970.]
- [Le] Leau L. Étude sur les équations fonctionnelles à une ou plusieurs variables, Ann. Fac. Sci. Toulouse **11** (1897).
- [Ma] Malgrange B. Travaux d'Écalle et de Martinet-Ramis sur les systèmes dynamiques, Sémin. Bourbaki, exp. 582 (1981), Astérisque **92–93** (1982), 59–73.
- [Mo] Moser J. On commuting circle mappings and simultaneous Diophantine approximations, Math. Z., **205** (1990), 105–121.
- [M-M] Mattei J.-F., Moussu R. Holonomie et intégrales premières, Ann. Sci. Ec. Norm. Sup., **13** (1980), 469–523.
- [M-R] Martinet J., Ramis J.-P. Problème de modules pour les équations différentielles non linéaires du premier ordre, Publ. Math. I.H.E.S., **55** (1982), 63–164.
- [Pf] Pfeiffer G. A. On the conformal mapping of curvilinear angles. The functional equation $\phi[f(x)] = a_1\phi(x)$, Math. Ann. (1917).
- [PM1] Pérez-Marco R. Sur la dynamique des germes de difféomorphismes holomorphes de $(\mathbb{C}, 0)$ et des difféomorphismes analytiques du cercle, Thèse Université de Paris-Sud, décembre 1990.
- [PM2] Pérez-Marco R. Centralizers I: Uncountable centralizers for non linearizable holomorphic germs of $(\mathbb{C}, 0)$, Preprint I.H.E.S., 1991; см. также: C. R. Acad. Sci. Paris, **313** (1991), 461–464.
- [Ru] Rüssmann H. Kleine Nener II: Bemerkungen zur Newtonschen Methode, Nach. Akad. Wiss. Göttingen, Math. Phys. Kl. (1972), 1–10.
- [Sa] Sad P. A note on non linearizable analytic functions, Bol. Soc. Brasil. Mat. (1980), No. 1, 31–36.
- [Si] Siegel C. L. Iterations of analytic functions, Ann. of Math., **43** (1942), 807–812.
- [Bo] Воронин С. М. Аналитическая классификация ростков конформных отображений $(\mathbb{C}, 0) \rightarrow (\mathbb{C}, 0)$ с тождественной линейной частью. — Функц. анализ и его прил., т. 15, 1981, вып. 1, с. 1–17.

- [Y1] Yoccoz J.-C. Théorème de Siegel, polynômes quadratiques et nombres de Brjuno, Preprint (1987); см. также: Asterisque, **231** (1995), 3–88.
- [Y2] Yoccoz J.-C. Conjugaison des difféomorphismes analytiques du cercle, Acta Math.
- [Y3] Yoccoz J.-C. Conjugaison différentiables des difféomorphismes du cercle dont le nombre de rotation vérifie une condition diophantienne, Ann. Sci. Ec. Norm. Sup. (4ème série), **17** (1984), 333–359.
- [Y4] Yoccoz J.-C. Centralisateurs et conjugaison différentiable des difféomorphismes du cercle, Thèse d'Etat Université de Paris-Sud, juin 1985; см. также: Astérisque, **231** (1995), 89–230.
- [Y5] Yoccoz J.-C. Linéarisation des germes de difféomorphismes holomorphes de $(\mathbb{C}, 0)$, C. R. Acad. Sci. Paris, **306**, 55–58.
- [Y-P-M] Yoccoz J.-C., Pérez-Marco R. Germes de feuilletages holomorphes à holonomie prescrite, Preprint Université de Paris-Sud, 92–94; см. также: Astérisque, **222** (1994), 345–371.

RICARDO PÉREZ-MARCO

Université de Paris-Sud

Math. 425-U.R.A. 1169 du C.N.R.S.

F-91405 ORSAY CEDEX

E-mail: Ricardo.Perez-Marco@math.u-psud.fr

0. Эта статья является введением в замечательные недавние работы М. Эрмана по симплектической динамике [Н1–Н7].

В первых двух разделах напоминаются основные понятия. В третьем излагается конструкция контрпримера к C^∞ -лемме о замыкании в категории гамильтоновых полей.

Раздел 4 целиком посвящен вопросу о существовании инвариантных или сдвинутых диофантовых торов коразмерности 1. Как и сдвинутые кривые Рюссмана [R1, Н9], такие торы появляются в неконсервативных ситуациях. В консервативных ситуациях (гамильтоновы потоки, диффеоморфизмы, сохраняющие объем) с помощью таких торов строятся контрпримеры ко многим гипотезам: гипотезе квазиэргодичности, гипотезе Песина, ...

В разд. 5 приведены некоторые из результатов Эрмана о *лагранжевых* инвариантных торах. Геометрия этих торов весьма существенно зависит от сигнатуры кручения; имеется также обширная информация о динамике на торе общего положения.

Чтобы сделать изложение более доступным, излагаемые результаты часто приводятся в наиболее простой, не оптимальной форме. В частности, *если не оговорено противное*, то все утверждения относятся к классу C^∞ .

1. КРИВЫЕ, ИНВАРИАНТНЫЕ ОТНОСИТЕЛЬНО ДИФФЕОМОРФИЗМОВ ЦИЛИНДРА

1.1. Положим $A = T^1 \times \mathbb{R}$ и обозначим канонические координаты на A через $\theta \in T^1$, $r \in \mathbb{R}$. Рассмотрим на A 1-форму $v = r d\theta$ и симплектическую форму $\omega = dv = dr \wedge d\theta$.

Пусть $F: (\theta, r) \mapsto (\Theta(\theta, r), R(\theta, r))$ — диффеоморфизм цилиндра A , изотопный тождественному. Запишем его первую координату в виде

$$\Theta(\theta, r) = \theta + \phi(\theta, r), \quad \phi \in C^\infty(A).$$

Определения. Диффеоморфизм F называется

- *монотонно положительным* (соотв. *отрицательным*), если $\partial_\tau \phi > 0$ (соотв. $\partial_\tau \phi < 0$) на A ;
- *симплектическим*, если $F^* \omega = \omega$;
- *точно симплектическим*, если 1-форма $F^* v - v$ является точной;
- *вполне интегрируемым*, если $R(\theta, \tau) = \tau$ и $\phi(\theta, \tau) = l(\tau)$ не зависят от θ .

1.2. Теория Биркгофа [B1, H19, F, L3]. Пусть F — монотонно положительный диффеоморфизм цилиндра A .

Первая теорема Биркгофа. Если график функции $\psi \in C^0(\mathbb{T}^1)$ инвариантен относительно F , то ψ липшицева и

$$\frac{\partial_\tau R}{\partial_\tau \phi}(F^{-1}(\theta, \psi(\theta))) > D\psi(\theta) > -\frac{\partial_\theta \Theta}{\partial_\tau \phi}(\theta, \psi(\theta))$$

для почти всех $\theta \in \mathbb{T}^1$.

Предположим теперь, что F является также и симплектическим (достаточно, чтобы F сохранял некоторую меру Радона, положительную на открытых множествах).

Вторая теорема Биркгофа. Предположим, что F сохраняет открытое подмножество $U \subset A$, обладающее следующими свойствами:

- существуют такие τ_0, τ_1 , $\tau_0 < \tau_1$, что $\mathbb{T}^1 \times (-\infty, \tau_0) \subset U \subset \mathbb{T}^1 \times (-\infty, \tau_1)$;
- U диффеоморфно A ;
- U совпадает с внутренностью своего замыкания.

Тогда граница множества U является графиком некоторой (липшицевой, согласно предыдущей теореме) функции $\psi \in C^0(\mathbb{T}^1)$.

Следствия. Пусть F — симплектический монотонно положительный диффеоморфизм цилиндра A ; для $\tau_0 > 0$ рассмотрим множество $\mathcal{C}(\tau_0)$ простых кривых, гомотопных сечению $\{\tau = 0\}$ в A , инвариантных относительно F и содержащихся в полосе $\mathbb{T}^1 \times [-\tau_0, \tau_0]$. Всякая кривая $C \in \mathcal{C}(\tau_0)$ представляет собой график некоторого отображения ψ_C , являющегося липшицевым с постоянной, не превышающей $K(\tau_0)$. Объединение всех кривых $C \in \mathcal{C}(\tau_0)$ компактно,

как и множество всех отображений ψ_C (относительно C^0 -топологии). Положим

$$F(\theta, \psi_C(\theta)) = (f_C(\theta), \psi_C(f_C(\theta)))$$

и обозначим через ρ_C число вращения отображения f_C . Множество значений ρ_C также является компактным. Кроме того, если ρ_C иррационально и кривая $C' \in \mathcal{C}(r_0)$ отлична от C , то либо $\rho_{C'} < \rho_C$, $\psi_{C'} < \psi_C$, либо $\rho_{C'} > \rho_C$, $\psi_{C'} > \psi_C$.

Замечание. Для того чтобы множество $\mathcal{C}(r_0)$ было непустым, необходимо, чтобы F был *точным* симплектическим диффеоморфизмом. Однако это требование не является достаточным, как показывает следующий пример:

$$F_0(\theta, r) = (\theta + r, r + \cos 2\pi(\theta + r)), \quad F_0^n(0, 0) = (0, n).$$

1.3. Диофантовы кривые: нормальная форма Биркгофа.
Для $\gamma > 0$, $\tau \geq 0$ положим

$$CD(\gamma, \tau) = \{\alpha \in \mathbb{R}; \forall q \geq 1, \forall p \in \mathbb{Z}, |q\alpha - p| \geq \gamma q^{-1-\tau}\},$$

$$CD(\tau) = \bigcup_{\gamma > 0} CD(\gamma, \tau), \quad CD = \bigcup_{\tau \geq 0} CD(\tau).$$

Число $\alpha \in CD$ называется *диофантовым*. (Простая) кривая C , инвариантная относительно диффеоморфизма F , называется *диофантовой*, если ограничение $F|_C$ сохраняет ориентацию кривой и число вращения этого ограничения диофантово. Если диффеоморфизм F и кривая C гладкие, то по теореме о глобальном сопряжении диффеоморфизмов окружности [HS, Y] на всякой диофантовой кривой C существует координата $\theta \in \mathbb{T}^1$, такая, что ограничение $F|_C$ является сдвигом относительно этой координаты.

Предложение [B2]. Пусть F — симплектический диффеоморфизм цилиндра A , изотопный тождественному, и C — инвариантная диофантова кривая, гомотопная нулевому сечению $\{r = 0\}$. Для всякого целого $n \geq 0$ существует симплектический диффеоморфизм H цилиндра A , изотопный тождественному, такой, что $H(C) = \{r = 0\}$ и

$$H \circ F \circ H^{-1}(\theta, r) = \left(\theta + \sum_1^n b_i r^i + \alpha, r \right) + \mathcal{O}(r^{n+1}),$$

где $\alpha \in CD$ — число вращения ограничения $F|_C$, а числа b_i , $i \geq 1$, однозначно определены (инварианты Биркгофа).

Говорят, что кручение диффеоморфизма F вдоль C положительно, если $b_1 > 0$, отрицательно, если $b_1 < 0$, и вырожденно, если $b_1 = 0$.

Упражнение. Если симплектический диффеоморфизм F монотонно положителен, то его кручение вдоль любой инвариантной диофантовой кривой, гомотопной сечению $\{\tau = 0\}$, положительно. Если кривая не гомотопна сечению $\{\tau = 0\}$, то это не обязательно так.

1.4. Теорема о сдвинутых кривых. Теорема о сдвинутых кривых, принадлежащая Рюссману [R1, H9] — это более гибкая форма теоремы Дж. Мозера [Mo1] об инвариантной кривой, справедливая без предположения о консервативности системы.

Пусть $L(\theta, \tau) = (\theta + l(\tau), \tau)$ — вполне интегрируемый диффеоморфизм цилиндра A . Предположим, что число $l(0) = \alpha$ диофантово и что $l'(0) \neq 0$. Для $\mu \in \mathbb{R}$ обозначим через T_μ сдвиг $(\theta, \tau) \mapsto (\theta, \tau + \mu)$ цилиндра A .

Теорема. Для диффеоморфизма F цилиндра A , достаточно близкого к L в C^∞ -топологии, существуют вещественное число μ_F , близкое к 0, и функция $\psi_F \in C^\infty(T^1)$, малая в C^∞ -топологии, такие, что график функции ψ_F является кривой, инвариантной относительно диффеоморфизма $T_{\mu_F} \circ F$, и число вращения этого диффеоморфизма есть α .

Замечания. 1) Отображение $F \mapsto (\mu_F, \psi_F)$, определенное в окрестности диффеоморфизма L , является хорошим отображением класса C^∞ в смысле Р. Гамильтона [Ha1, H10, Vo].

2) Пусть $\tau \geq 0$, $\gamma > 0$, $\delta > 0$. Предположим, что $l'(\tau) \neq 0$ для $\tau \in [-\delta, \delta]$, и положим $K = l'([-\delta, \delta]) \cap CD(\gamma, \tau)$. Тогда для всех $k > 2\tau + 3$ существует окрестность диффеоморфизма L в C^k -топологии, в которой утверждение теоремы верно для всех чисел вращения из K . Зависимость функции ψ_F и числа μ_F от числа вращения является гладкой в смысле Уитни [La, Po].

3) Пусть F — точный симплектический монотонно положительный диффеоморфизм цилиндра A и α — вещественное диофантово число.

- Если F не сохраняет ни одной кривой, гомотопной $\{r = 0\}$, с числом вращения α , то и ни для какого возмущения диффеоморфизма F таких кривых нет.
- Если F сохраняет *гладкую* кривую, гомотопную $\{r = 0\}$, с числом вращения α , то то же верно для всякого *точного симплектического* возмущения диффеоморфизма F .

1.5. Области неустойчивости. Пусть F — точный симплектический монотонно положительный диффеоморфизм цилиндра A . Объединение кривых, инвариантных относительно F и гомотопных окружности $\{r = 0\}$, является замкнутым подмножеством в A . Связные компоненты дополнения, ограниченные парой таких кривых, называются *областями неустойчивости* диффеоморфизма F ; динамика диффеоморфизма F в этих зонах понята далеко не достаточно, хотя недавно здесь и получены значительные продвижения. Описать их в настоящем докладе невозможно; отошлем читателя к работам [An, B-H, Ch, L1-L4, M1-M5].

2. КРАТКИЕ СВЕДЕНИЯ ИЗ СИМПЛЕКТИЧЕСКОЙ ДИНАМИКИ

2.1. *Симплектическим многообразием* называется пара (M, ω) , где M — многообразие (четной размерности $2n$), а ω — замкнутая всюду невырожденная 2-форма. Пусть N — многообразие половинной размерности n . Иммерсия $j : N \rightarrow M$ называется *лагранжевой*, если $j^*\omega = 0$. Для функции $H \in C^\infty(M)$ векторное поле X_H , определенное равенством

$$\omega(X_H, \cdot) = dH(\cdot),$$

называется *гамильтоновым* с гамильтонианом H .

Если гиперповерхность V в M является поверхностью уровня двух гамильтонианов H_1, H_2 на M , то соответствующие гамильтоновы поля X_{H_1}, X_{H_2} коллинеарны на V и орбиты порождаемых ими потоков совпадают.

2.2. Кокасательное расслоение $p : T^*V \rightarrow V$ многообразия V снабжено канонической структурой симплектического многообразия: положим $\omega = dv$, где 1-форма v в точке $\zeta \in T^*M$ определяется равенством

$$v(Y) = \zeta(Dp(Y)), \quad Y \in T_\zeta(T^*M).$$

Сечение $s : V \rightarrow T^*V$, т.е. 1-форма на V , является лагранжевым вложением в том и только в том случае, когда эта форма замкнута.

Предложение [W]. Если N — вложенное лагранжево подмногообразие симплектического многообразия M , то существует симплектический диффеоморфизм некоторой окрестности подмногообразия N в M на окрестность нулевого сечения расслоения T^*N (переводящий N в нулевое сечение).

2.3. Для $2n \geq k \geq 0$ введем на $T^{2n-k} \times \mathbf{R}^k$ координаты $\theta_1, \theta_2, \dots, \theta_{2n-k} \in T^1, \tau_1, \dots, \tau_k \in \mathbf{R}$. Для $H \in C^\infty(T^{2n-k} \times \mathbf{R}^k)$ положим

$$\nabla H = {}^t(\partial_{\theta_1} H, \dots, \partial_{\tau_k} H).$$

Пусть $A \in GL(2n, \mathbf{R})$ — кососимметричная матрица; симплектическая форма (с постоянными коэффициентами), ассоциированная с A , определяется равенством

$$\omega_A(A\nabla H, \cdot) = dH(\cdot).$$

2.4. отождествим $T^*(T^n)$ с $T^n \times \mathbf{R}^n$ и обозначим это многообразие через A^n . Каноническая симплектическая форма на A^n имеет вид

$$\omega = dv = \sum dr_i \wedge d\theta_i, \quad v = \sum \tau_i d\theta_i.$$

Пусть F — диффеоморфизм многообразия A^n , гомотопный тождественному. Запишем его в виде

$$F(\theta, \tau) = (\Theta(\theta, \tau), R(\theta, \tau)) = (\theta + \phi(\theta, \tau), R(\theta, \tau)).$$

Определения. Диффеоморфизм F называется

- *симплектическим*, если $F^*\omega = \omega$;
- *точным симплектическим*, если форма $F^*v - v$ точна;
- *монотонным*, если отображение $\partial_\tau \phi$ всюду обратимо;
- *строго монотонным*, если для всех $\theta \in T^n$ отображение $\tau \mapsto \phi(\theta, \tau)$ является диффеоморфизмом пространства $\mathbf{R}^n$;
- *вполне интегрируемым*, если $R(\theta, \tau) = \tau$ и $\phi(\theta, \tau) = l(\tau)$ не зависит от θ .

Если отображение $L: (\theta, \tau) \mapsto (\theta + l(\tau), \tau)$ вполне интегрируемое и симплектическое, то для всех $\tau \in \mathbf{R}^n$ дифференциал $Dl(\tau)$ является симметричной матрицей. В этом случае $l = D\hat{l}$ для некоторой функции $\hat{l} \in C^\infty(\mathbf{R}^n)$.

Если график функции $\psi \in C^\infty(\mathbb{T}^n, \mathbb{R}^n)$ представляет собой лагранжево подмногообразие, то $\psi = [\psi] + d\hat{\psi}$ для некоторой функции $\hat{\psi} \in C^\infty(\mathbb{T}^n)$, $[\psi] = \int \psi \, dm$ (где через dm обозначена мера Лебега на $\mathbb{T}^n$).

2.5. Диофантовы условия. Пусть $\gamma > 0$, $\tau \geq 0$, $n \geq 1$. Положим

$$\begin{aligned} CDH_n(\gamma, \tau) &= \{(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n, \forall k \in \mathbb{Z}^n \setminus \{0\}, \left| \sum k_i \alpha_i \right| \geq \gamma (\sup |k_i|)^{1-n-\tau}\}, \\ CDH_n(\tau) &= \bigcup_{\gamma > 0} CDH_n(\gamma, \tau), \quad CDH_n = \bigcup_{\tau \geq 0} CDH_n(\tau), \\ CD_n(\gamma, \tau) &= \{(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n, \forall k \in \mathbb{Z}^n \setminus \{0\}, \forall l \in \mathbb{Z}, \left| l + \sum k_i \alpha_i \right| \geq \gamma (\sup |k_i|)^{-n-\tau}\}, \\ CD_n(\tau) &= \bigcup_{\gamma > 0} CD_n(\gamma, \tau), \quad CD_n = \bigcup_{\tau \geq 0} CD_n(\tau). \end{aligned}$$

Для $\tau > 0$ множества $CD_n(\tau)$ и $CDH_n(\tau)$ имеют полную меру Лебега, однако они тощи в смысле Бэра (для $n \geq 2$ в случае CDH_n). Множества $CD_n(\gamma, \tau)$ инвариантны относительно целых сдвигов; мы сохраняем те же обозначения для образов этих множеств в $\mathbb{T}^n$.

2.6. Теоремы о локальной сопряженности на торах. Для $\alpha \in \mathbb{T}^n$ обозначим через R_α сдвиг $\theta \mapsto \theta + \alpha$ тора $\mathbb{T}^n$. Для $\alpha \in \mathbb{R}^n$ через X_α обозначим векторное поле $\sum \alpha_i \partial_{\theta_i}$. Через $\text{Diff}_0^\infty(\mathbb{T}^n)$ обозначим множество диффеоморфизмов h тора $\mathbb{T}^n$, гомотопных тождественному и удовлетворяющих условию $\int (h(\theta) - \theta) \, dm = 0$.

Теорема. Пусть $\alpha \in CD_n$. Всякий диффеоморфизм f тора $\mathbb{T}^n$, достаточно близкий к R_α , может быть единственным образом записан в виде $f = R_\lambda \circ h \circ R_\alpha \circ h^{-1}$, где точка $\lambda \in \mathbb{T}^n$ достаточно близка к 0 и диффеоморфизм $h \in \text{Diff}_0^\infty(\mathbb{T}^n)$ близок к тождественному.

Теорема. Пусть $\alpha \in CDH_n$. Всякое векторное поле X на $\mathbb{T}^n$, достаточно близкое к полю X_α , единственным образом записывается в виде $X = X_\lambda + h_* X_\alpha$, где точка $\lambda \in \mathbb{R}^n$ достаточно близка к 0 и диффеоморфизм $h \in \text{Diff}_0^\infty(\mathbb{T}^n)$ достаточно близок к тождественному.

Отображение $f \mapsto (\lambda, h)$ (соотв. $X \mapsto (\lambda, h)$), определенное в окрестности сдвига R_α (соотв. в окрестности поля X_α), является хорошим отображением класса C^∞ в смысле Гамильтона [Ha1, H10, Vo].

Если f сохраняет меру Лебега (соотв. если поле X бездивергентно), то диффеоморфизм сопряжения h сохраняет меру Лебега, так как сдвиг R_α (соотв. поле X_α) является однозначно эргодическим (т. е. мера Лебега является единственной инвариантной относительно него вероятностной мерой).

Гораздо более элементарными методами можно получить следующее утверждение.

Предложение. Пусть $\alpha \in CDH_n$, $\psi \in C^\infty(T^n)$, $\psi > 0$. Тогда существуют единственное вещественное число $c > 0$ и единственный диффеоморфизм $h \in \text{Diff}_0^\infty(T^n)$, такие, что $\psi X_\alpha = h_* X_{c\alpha}$.

2.7. Нормальная форма Биркгофа. Пусть F — симплектический диффеоморфизм симплектического многообразия (M, ω) . Тогда k -мерный тор T , инвариантный относительно F , называется *диофантовым*, если существуют точка $\alpha \in CD_k$ и диффеоморфизм $h : T^k \rightarrow T$, такие, что $h \circ R_\alpha \circ h^{-1} = F|_T$.

Замечания. 1) Если заменить диффеоморфизм h диффеоморфизмом $h \circ A$, $A \in GL(k, \mathbb{Z})$, то точку α следует заменить точкой $A^{-1}\alpha$.

2) 2-форма $h^*(\omega|_T)$ инвариантна относительно эргодического сдвига R_α ; поэтому у нее постоянные коэффициенты. Если симплектическая форма ω точна (например, если $M = T^*V$), то должно выполняться условие $h^*(\omega|_T) = 0$; отсюда вытекает, что $k \leq \frac{1}{2} \dim M$.

Пусть теперь T — инвариантный диофантов лагранжес тор.

Предложение. Для всякого $n \geq 0$ существует симплектическое вложение H некоторой окрестности нулевого сечения $T^n \times \{0\}$ в $T^n \times \mathbb{R}^n$ в M , такое, что $H(T^n \times \{0\}) = T$ и

$$H^{-1} \circ F \circ H(\theta, r) = \left(\theta + \sum_1^m l_i(r) + \alpha, r \right) + O(\|r\|^{m+1});$$

здесь $l_i(r)$, $1 \leq i \leq m$, — однородное полиномиальное отображение степени i из $\mathbb{R}^n$ в $\mathbb{R}^n$ вида $D\hat{l}_i$ (где $\hat{l}_i$ — однородный многочлен степени $i+1$ в $\mathbb{R}^n$).

Замечания. Если заменить вложение H вложением $H \circ T^*A$, $A \in GL(n, \mathbb{Z})$, то многочлены $\hat{l}_i$ следует заменить многочленами $\hat{l}_i \circ {}^t A^{-1}$. Сигнатура квадратичной формы $\hat{l}_2$ при этом не меняется.

Определение. Тор T имеет вырожденное (положительное, отрицательное, индефинитное, нулевое) кручение, если квадратичная форма $\hat{l}^2$ вырожденная (положительно определенная, отрицательно определенная, индефинитная, нулевая).

Для гамильтоновых векторных полей справедлив аналог сформулированного выше предложения: в окрестности инвариантного диофантова лагранжева тора всякое гамильтоново поле можно приблизить, с точностью до произвольного порядка, вполне интегрируемым гамильтоновым полем.

2.8. Теоремы КАМ. Полное изложение и весьма детальная библиография по различным вариантам теорем КАМ приведены в [Bo]. Мы процитируем лишь один из результатов.

Теорема. Пусть $(M^{2n}, \omega = dv)$ — точное симплектическое многообразие и F_0 — его точный симплектический диффеоморфизм. Пусть $\alpha \in CD_n$ и $h_0 : \mathbb{T}^n \rightarrow M$ — вложение (автоматически лагранжево), такое, что его образ $T_0 = h_0(\mathbb{T}^n)$ инвариантен относительно F_0 и что

$$h_0 \circ R_\alpha \circ h_0^{-1} = F_0|_{T_0}.$$

Для всякого точного симплектического диффеоморфизма F , достаточно близкого к F_0 , существует (лагранжево) вложение $h_F : \mathbb{T}^n \rightarrow M$, близкое к h_0 , такое, что тор $T_F = h_F(\mathbb{T}^n)$ инвариантен относительно F и что

$$h_F \circ R_\alpha \circ h_F^{-1} = F|_{T_F}.$$

Отображение $F \mapsto h_F$ является хорошим отображением класса C^∞ в смысле Гамильтона.

3. КОНТРПРИМЕР К C^∞ -ЛЕММЕ О ЗАМЫКАНИИ ДЛЯ ГАМИЛЬТОНОВЫХ ВЕКТОРНЫХ ПОЛЕЙ

3.1. Лемма о замыкании. Пусть M — многообразие, на котором задан диффеоморфизм f (или поток $(f^s)_{s \in \mathbb{R}}$, ассоциированный с

векторным полем X). Точка $x \in M$ называется (положительно) *рекуррентной*, если

$$\lim_{s \rightarrow +\infty} d(f^s(x), x) = 0.$$

Проблема «замыкания» состоит в следующем: можно ли так возмутить отображение f (или векторное поле X), чтобы точка x стала *периодической*? Проблема сформулирована для произвольных диффеоморфизмов, однако ее можно ставить и для классов отображений и векторных полей, удовлетворяющих некоторым ограничениям: диффеоморфизмов, сохраняющих объем, симплектических диффеоморфизмов, гамильтоновых векторных полей, биголоморфных диффеоморфизмов, ...

Ответ на эти вопросы существенно зависит от топологии пространства возмущений. В классе всех диффеоморфизмов ответ тривиально положителен для C^0 -малых возмущений; для C^1 -топологии ответ по-прежнему положителен, но это трудная теорема, принадлежащая Пью [Pu]. В C^2 -топологии Гуттиерес [Gu] сконструировал векторное поле на T^2 с рекуррентной точкой $x \neq 0$, которую нельзя сделать периодической C^2 -малым возмущением с компактным носителем в $T^2 \setminus \{0\}$. Для C^2 -топологии на компактном многообразии вопрос остается полностью открытым.

Пью и Робинсон [P-R] распространили результат Пью на класс гамильтоновых векторных полей: если точка x рекуррентна для гамильтонова векторного поля X_{H_0} , то можно так возмутить гамильтониан H_0 в C^2 -топологии, что точка x оказывается периодической для поля X_H при некотором H , C^2 -близком к H_0 .

Приводимый ниже пример был предложен Цендером [Z1] для доказательства того, что компактная поверхность уровня гамильтониана может не содержать периодических орбит. Эрман показал, что тот же самый пример служит контрпримером к «лемме о замыкании» в топологии C^k (при достаточно большом k) в классе гамильтоновых векторных полей.

3.2. Пусть $\alpha \in CDH_{2n-1}$ и $A \in GL(2n, \mathbf{R})$ — кососимметрическая матрица вида

$$A = \begin{pmatrix} B & {}^t\alpha \\ -\alpha & 0 \end{pmatrix}, \quad B \in M(2n-1, \mathbf{R}).$$

Снабдим $T^{2n-1} \times [-\delta, \delta]$ постоянной симплектической формой ω_A , ассоциированной с матрицей A (ср. п. 2.3).

Положим $H_0(\theta, \tau) = \tau$. На каждой поверхности уровня энергии $\{\tau = \tau_0\}$ гамильтоново поле X_{H_0} совпадает с постоянным полем X_α на T^{2n-1} .

Рассмотрим теперь гамильтониан $H \in C^\infty(T^{2n-1} \times [-\delta, \delta])$, C^∞ -близкий к H_0 , и число $\tau_0 \in [-\delta/2, \delta/2]$. Поверхность уровня $\{H = \tau_0\}$ является графиком T некоторой функции $\psi \in C^\infty(T^{2n-1}, [-\delta, \delta])$.

Положим $K(\theta, \tau) = \tau - \psi(\theta)$, так что $T = \{K = 0\}$. Поля X_H и X_K на T коллинеарны.

Поле $Z = -B\nabla_\theta \psi + X_\alpha$ на T^{2n-1} является проекцией ограничения $X_K|_T$. Так как матрица B кососимметрична, поле Z бездивергентно, и так как Z близко к X_α , его можно записать в виде (ср. п. 2.6)

$$Z = X_\lambda + h_* X_\alpha,$$

где $h \in \text{Diff}_0(T^{2n-1})$ сохраняет меру Лебега.

Итак,

$$\int Z dm = \alpha = \int h_* X_\alpha dm,$$

откуда $\lambda = 0$.

Таким образом, поле X_K на T сопряжено полю X_α ; согласно п. 2.6, поле X_H на T сопряжено полю $X_{c\alpha}$ для некоторого c , близкого к 1. В результате мы доказали следующее утверждение.

Теорема [Н1, Н2]. Для всякого гамильтониана H , достаточно близкого к H_0 , и всякой поверхности уровня энергии $\{H = \tau_0\}$, $|\tau_0| \leq \delta/2$, векторное поле X_H на $\{H = \tau_0\}$ сопряжено диофантову полю $X_{c\alpha}$, $c > 0$. В частности, область $\{|H| \leq \delta/2\}$ не содержит периодических орбит поля X_H .

Замечания. 1) Предыдущий пример можно распространить на T^{2n} , а затем разбить точки тора T^{2n} на симплектические слои, получив при этом несчетное семейство компактных симплектических многообразий, на которых воспроизводится описанное выше явление. Этот результат нельзя, однако, распространить на точные симплектические многообразия (ср. п. 2.7), если $n > 1$. «Лемма о замыкании» для гамильтоновых векторных полей на кокасательном расслоении (например) остается открытым вопросом.

2) Для $\alpha \in CDH_{2n-1}(0)$ в условиях предыдущей теоремы достаточно потребовать, чтобы H был C^k -близким к H_0 для всех $k > 2n$.

3) В следующем разделе мы приведем более тонкий пример, иллюстрирующий жесткость числа вращения.

4. ИНВАРИАНТНЫЕ ТОРЫ КОРАЗМЕРНОСТИ 1

4.1. Теорема о сдвинутых торах [Н7]. Для $n \geq 1$, $\delta > 0$ положим $B^n = T^n \times \mathbb{R}$, $B^n(\delta) = T^n \times [-\delta, +\delta]$ и обозначим через $p_1 : B^n \rightarrow T^n$ и $p_2 : B^n \rightarrow \mathbb{R}$ проекции на сомножители.

Для $(\lambda, \mu) \in B^n$ обозначим через $R_{\lambda, \mu}$ сдвиг $(\theta, r) \mapsto (\theta + \lambda, r + \mu)$.

Рассмотрим вложение $F_0 : B^n(\delta) \hookrightarrow B^n$ класса C^∞ , удовлетворяющее условию $F_0(\theta, 0) = (\theta + \alpha, 0)$ при $\alpha \in CD_n$.

Теорема. Для всякого вложения $F : B^n(\delta) \hookrightarrow B^n$, достаточно близкого к F_0 , существуют сдвиг $R_{\lambda, \mu}$ многообразия B^n , функция $\psi \in C^\infty(T^n)$ со средним значением 0 и диффеоморфизм $h \in \text{Diff}_0^\infty(T^n)$, такие, что

- 1) вложение $\tilde{F} = R_{\lambda, \mu} \circ F$ сохраняет график функции ψ ;
- 2) для $\theta \in T^n$

$$\tilde{F}(\theta, \psi(\theta)) = (f(\theta), \psi(f(\theta))), \text{ где } f = h \circ R_\alpha \circ h^{-1}.$$

Дополнения. 1) Для $F = F_0$ имеем $\lambda = \mu = 0$, $\psi = 0$, $h = \text{id}$; если вложение F близко к F_0 и мы потребуем, чтобы λ, μ, ψ были малы, а диффеоморфизм h был близок к тождественному, то эти данные однозначно определяются условиями теоремы.

2) Наилучший подход (по крайней мере, наиболее приятный и гибкий) к формулировке многочисленных теорем о малых знаменателях принадлежит Гамильтону [Ha1, Ha2, H10, Bo]; он вводит понятия «хорошего пространства Фреше» и «хорошего отображения класса C^r » хороших пространств Фреше, для которого остается верным локальная теорема об обращении (или о неявной функции). (При этом, однако, необходимо проверить обратимость дифференциала в целой окрестности рассматриваемой точки.) За более детальным описанием следует обратиться к [Bo]. Доказательство теоремы, набросок которого мы приводим ниже, показывает, что отображение $F \mapsto (\lambda, \mu, \psi, h)$ является хорошим класса C^∞ в смысле Гамильтона.

3) Зафиксируем $\tau \geq 0$, $\gamma > 0$. Для всех $\alpha' \in CD_n(\gamma, \tau)$ можно применить теорему к $R_{\alpha' - \alpha, 0} \circ F_0$; тогда для F из окрестности $U_{\alpha'}$

вложения F_0 теорема обеспечивает существование таких $\lambda_{\alpha'}$, $\mu_{\alpha'}$, $\psi_{\alpha'}$, $h_{\alpha'}$, что

$$F(\theta, \psi_{\alpha'}(\theta)) = (f_{\alpha'}(\theta), \psi_{\alpha'}(f_{\alpha'}(\theta))) - (\lambda_{\alpha'}, \mu_{\alpha'}),$$

$$f_{\alpha'} = h_{\alpha'} \circ R_{\alpha'} \circ h_{\alpha'}^{-1}.$$

В приложениях чрезвычайно существенными оказываются следующие два факта.

- Пересечение всех окрестностей $U_{\alpha'}$ для $\alpha' \in CD_n(\gamma, \tau)$ тоже является окрестностью вложения F_0 .
- Отображение $\alpha' \mapsto (\lambda_{\alpha'}, \mu_{\alpha'}, \psi_{\alpha'}, h_{\alpha'})$ является гладким в смысле Уитни.

Следующее фундаментальное замечание Боста позволяет легко доказать эти утверждения: посмотрим на доказательство теоремы при фиксированном α , основанное на подходе Гамильтона; вводятся хорошие пространства Фреше $F_1, F_2, \dots$ и хорошие отображения $L_1 : F_1 \rightarrow F_2, \dots$. Считая затем α параметром, заменим пространства F_1, F_2 пространствами $\tilde{F}_1 = C^p(CD_n(\gamma, \tau), F_1)$, $\tilde{F}_2 = C^p(CD_n(\gamma, \tau), F_2)$ и отображение L_1 отображением $\tilde{L}_1 : \tilde{F}_1 \rightarrow \tilde{F}_2$. Тогда $\tilde{F}_1, \tilde{F}_2$ вновь являются хорошими пространствами Фреше, а $\tilde{L}_1$ — хорошим отображением, что и позволяет сделать необходимое заключение.

4.2. Некоторые указания к доказательству. Пусть F_1 — диффеоморфизм многообразия $\mathbf{T}^n \times \mathbf{T}^1 = \mathbf{T}^{n+1}$, совпадающий с F_0 в окрестности сечения $\mathbf{T}^n \times \{0\}$. Положим

$$C_1 = \{\psi \in C^\infty(\mathbf{T}^n), \int \psi \, dm = 0\},$$

$$C_2 = C_1^n \subset C^\infty(\mathbf{T}^n, \mathbf{R}^n),$$

$$C_3 = \{g \in C^\infty(\mathbf{T}^{n+1}, \mathbf{R}^{n+1}), g|_{\mathbf{T}^n \times \{0\}} \equiv 0\}.$$

Для $\psi \in C_1$ положим

$$K_\psi(\theta, \tau) = (\theta, \tau + \psi(\theta)).$$

Для $\phi \in C_2$ положим

$$H_\phi(\theta, \tau) = (\theta + \phi(\theta), \tau).$$

Для $g \in C_3$ положим

$$J_g(\theta, r) = (\theta, r) + g(\theta, r).$$

Мы хотим доказать, что отображение

$$\Phi : \mathbf{T}^n \times \mathbf{R} \times C_1 \times C_2 \times C_3 \rightarrow \text{Diff}^\infty(\mathbf{T}^{n+1}),$$

$$(\lambda, \mu, \psi, \phi, g) \mapsto R_{\lambda, \mu} \circ K_\psi \circ H_\phi \circ J_g \circ F_1 \circ H_\phi^{-1} \circ K_\psi^{-1},$$

определенное в окрестности начала координат $(0, 0, 0, 0, 0)$, является хорошим диффеоморфизмом класса C^∞ в смысле Гамильтона окрестности начала координат на окрестность диффеоморфизма F_1 в $\text{Diff}^\infty(\mathbf{T}^{n+1})$. Для этого достаточно проверить выполнение условий теоремы Гамильтона о локальном обращении. Здесь мы проверим только, что дифференциал $D_0\Phi$ отображения Φ в начале координат обратим.

Имеем

$$D_0\Phi(\lambda, \mu, \psi, \phi, g) = (\lambda, \mu) + g \circ F_1 + S_1 + S_2,$$

$$S_1(\theta, r) = (\phi \circ p_1 \circ F_1(\theta, r), 0) - D_{(\theta, r)}F_1(\phi(\theta), 0),$$

$$S_2(\theta, r) = (0, \psi \circ p_1 \circ F_1(\theta, r)) - D_{(\theta, r)}F_1(0, \psi(\theta)).$$

Положив $a(\theta) = \partial_r(p_2 \circ F_1)(\theta, 0)$, получаем

$$S_1(\theta, 0) = (\phi(\theta + \alpha) - \phi(\theta), 0),$$

$$S_2(\theta, 0) = (0, \psi(\theta + \alpha) - a(\theta)\psi(\theta)).$$

Возьмем $\eta_1 \in C^\infty(\mathbf{T}^{n+1}, \mathbf{R}^n)$, $\eta_2 \in C^\infty(\mathbf{T}^{n+1}, \mathbf{R})$ и рассмотрим уравнение

$$D_0\Phi(\lambda, \mu, \psi, \phi, g) = (\eta_1, \eta_2).$$

Получаем

$$\eta_1(\theta, 0) = \lambda + \phi(\theta + \alpha) - \phi(\theta),$$

откуда однозначно находятся $\lambda \in \mathbf{R}^n$ и $\phi \in C_2$. Аналогично,

$$\eta_2(\theta, 0) = \mu + \psi(\theta + \alpha) - a(\theta)\psi(\theta),$$

откуда однозначно определяются $\mu \in \mathbf{R}$ и $\psi \in C_1$: записав $a(\theta) = a^* \frac{b(\theta)}{b(\theta + \alpha)}$, $a^* \neq 0$, $b > 0$, $\log b \in C_1$, решаем затем уравнение (ср. [Н9])

$$b(\theta + \alpha)\eta_2(\theta, 0) - \mu b(\theta + \alpha) = b\psi(\theta + \alpha) - b\psi(\theta)a^*.$$

Наконец, нужно получить

$$g \circ F_1(\theta, r) = (\eta_1(\theta, r) - \eta_1(\theta, 0), \eta_2(\theta, r) - \eta_2(\theta, 0)),$$

откуда g определяется однозначно.

Таким образом, мы показали, что $D_0\Phi$ обратим.

4.3. Жесткость числа вращения. Приводимая ниже лемма служит решающим этапом в применении теоремы 4.1, которое будет дано в п. 4.5.

Возьмем два целых числа k, n , $0 < k < n$, и открытое множество $U \subset \mathbb{R}^k$. Снабдим $T^{2n-k} \times U$ постоянной симплектической формой ω ; обозначим через $p_1 : T^{2n-k} \times U \rightarrow T^{2n-k}$ каноническую проекцию.

Пусть $H \in C^\infty(T^{2n-k} \times U)$ — гамильтониан, X_H — соответствующее гамильтоново векторное поле, $\lambda \in \mathbb{R}^{2n-k}$, $X_\lambda = \sum \lambda_i \partial / \partial \theta_i$, $Y = X_H + X_\lambda$.

Лемма. *Предположим, что*

1) *поле Y касается графика T некоторого гладкого отображения из T^{2n-k} в U ;*

2) *$p_{1*}(Y|_T) = h_*(X_\alpha)$, где h — гомотопный тождественному диффеоморфизм тора T^{2n-k} , а X_α — постоянное эргодическое поле на T^{2n-k} .*

Тогда вектор $\lambda - \alpha$ является ω -ортогональным подпространству $\mathbb{R}^{2n-k} \times \{0\}$.

Доказательство. отождествим тор T^{2n-k} с нулевым сечением $T^{2n-k} \times \{0\}$; положим $j(\theta) = (\theta, \psi(\theta))$, $\omega_1 = \omega|_{T^{2n-k}}$, $\omega_2 = j^*\omega$.

Так как отображение j гомотопно включению $T^{2n-k} \times \{0\} \hookrightarrow T^{2n-k} \times \mathbb{R}^k$, формы ω_1 и ω_2 когомологичны; аналогично, так как h гомотопен тождественному отображению, ω_2 и $h^*\omega_2$ тоже когомологичны. Формы ω_1 и $h^*\omega_2$ когомологичны и имеют постоянные коэффициенты (так как поле X_α эргодично); поэтому они равны.

Положим $Y_1 = p_{1*}(Y|_T)$. Поле Y является симплектическим; следовательно, форма $i_Y\omega$ замкнута; имеем $j^*(i_Y\omega) = i_{Y_1}\omega_2$, и эта форма когомологична форме $h^*(i_{Y_1}\omega_2) = i_{X_\alpha}\omega_1$. С другой стороны, $i_Y\omega = i_{X_H}\omega + i_{X_\lambda}\omega$, где форма $i_{X_H}\omega = dH$ точна и $j^*(i_{X_\lambda}\omega)$ когомологична форме $i_{X_\lambda}\omega_1$.

Поэтому формы $i_{X_\lambda}\omega_1$ и $i_{X_\alpha}\omega_1$ когомологичны; так как их коэффициенты постоянны, то они равны, что мы и хотели показать.

4.4. Гипотеза квазиэргодичности. Гипотеза квазиэргодичности представляет собой попытку (как мы увидим, безуспешную) дать строгое математическое обоснование статистической механики.

Гипотеза эргодичности Больцмана («временные средние равны пространственным средним») на современном языке (Биркгоф, Купман [В-К]) означает, что для (почти) всех собственных гамильтонианов H на симплектическом многообразии (M, ω) и для (почти) всех значений энергии поле X_H является эргодическим на каждой компоненте связности уровня энергии. Теоремы КАМ показали несостоятельность этой гипотезы: для некоторого открытого множества гамильтонианов H существует открытое множество значений энергии, такое, что на каждом из соответствующих поверхностей уровней энергии имеется множество инвариантных лагранжевых диофантовых торов положительной меры.

Требование эргодичности поля X_H на каждой компоненте связности почти всех уровней энергии можно заменить значительно более слабым: для плотного множества значений энергии поле X_H должно иметь всюду плотную орбиту на каждой компоненте связности соответствующего уровня энергии. Отсюда вытекает формулировка гипотезы квазиэргодичности [Еh, Р].

При $\dim M = 2$ справедливость этой гипотезы тривиальна; при $\dim M = 4$ гипотеза неверна, так как в силу теорем КАМ диофантовы инвариантные торы разделяют поверхность уровня энергии. В следующем пункте мы объясним, следуя Эрману, почему на некоторых симплектических (не точных) многообразиях существует открытое множество гамильтонианов и открытое множество уровней энергии, для которых канторово множество инвариантных диофантовых торов коразмерности 1 препятствует появлению плотных орбит на соответствующей поверхности уровня энергии.

4.5. Контрпример к гипотезе квазиэргодичности [Н7]. Введем на многообразии $T^{2n} \times [-\delta, +\delta]^2$ (с координатами $\theta_1, \dots, \theta_{2n}, r_1, r_2$) постоянную симплектическую форму ω_A , ассоциированную с кососимметрической матрицей $A \in GL(2n + 2, \mathbf{R})$ вида

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & B & {}^t\beta & {}^t\beta' \\ 0 & -\beta & 0 & 0 \\ -1 & -\beta' & 0 & 0 \end{pmatrix},$$

где $B \in M(2n - 1, \mathbf{R})$, $\beta' \in CD_{2n-1}$, $\beta \in \mathbf{R}^{2n-1}$, $\beta \neq 0$.

Пусть $H_0 \in C^\infty([-\delta, +\delta]^2)$ — гамильтониан, зависящий только от r_1, r_2 . Предположим, что

$$H_0(0, 0) = \partial_{r_1} H_0(0, 0) = 0,$$

$$\partial_{r_2} H_0(0, 0) \neq 0 \neq \partial_{r_1}^2 H_0(0, 0).$$

Теорема. Пусть $H \in C^\infty(\mathbb{T}^{2n} \times [-\delta, +\delta]^2)$ — гамильтониан, достаточно близкий к H_0 , и b — значение энергии, близкое к нулю. Тогда поле X_H на поверхности энергии $\{H = b\}$ сохраняет канторово множество диофантовых торов размерности $2n$, C^∞ -близких к $\mathbb{T}^{2n} \times (0, 0)$.

Набросок доказательства. 1) Выберем $\gamma > 0$, $\tau > 0$ таким образом, чтобы значение $s = 0$ было точкой уплотнения множества таких $s \in \mathbb{R}$, что $\beta' - \beta s \in CD_{2n-1}(\gamma, \tau)$.

2) Поверхность уровня энергии $\{H = b\}$ в окрестности нулевого сечения $\mathbb{T}^{2n} \times (0, 0)$ задается уравнением

$$K(\theta, r_1, r_2) = r_2 - \eta(\theta, r_1) = 0.$$

Поля X_H , X_K коллинеарны на $\{H = b\}$. Проекция поля X_K на гиперповерхность $\{r_2 = 0\} \cong \mathbb{T}^{2n} \times [-\epsilon, +\epsilon]$ ($0 < \epsilon < \delta$) имеет вид

$$Z = (1, -B\nabla_\theta \eta - \beta \partial_{r_1} \eta + \beta', \langle \beta, \nabla_\theta \eta \rangle).$$

Поле Z является малым C^∞ -возмущением постоянного горизонтального поля $(1, \beta', 0)$.

Поток поля Z за единичное время сохраняет поверхность уровня $\{\theta_1 = 0\}$ и индуцирует на ней диффеоморфизм, который мы обозначаем через F .

Отождествим многообразие $\mathbb{T}^{2n-1} \times \mathbb{R}$ с $\{0\} \times \mathbb{T}^{2n-1} \times \mathbb{R}$.

3) Для $\lambda \in [-1, +1]^{2n-1}$ обозначим через Z_λ векторное поле $Z + (0, \lambda, 0)$ и через F_λ ограничение на $\mathbb{T}^{2n-1} \times \mathbb{R}$ потока этого поля за единичное время.

Зафиксируем r_1 и s , близкие к нулю, такие, что $\alpha = \beta' - \beta s \in CD_{2n-1}(\gamma, \tau)$. По теореме о сдвинутом торе (4.1) существуют такой вектор $\Lambda(\lambda) \in \mathbb{R}^{2n-1}$ и такое число $\mu(\lambda) \in \mathbb{R}$, что диффеоморфизм $\tilde{F}_\lambda = R_{\Lambda(\lambda), \mu(\lambda)} \circ F_\lambda$ сохраняет график некоторой гладкой функции $\mathbb{T}^{2n-1} \rightarrow \mathbb{R}$ со средним значением r_1 и потому C^∞ -сопряжен повороту R_α .

4) Легко проверить, что отображение $\lambda \mapsto \Lambda(\lambda)$ является диффеоморфизмом в окрестности начала координат и что поэтому существует единственный вектор $\lambda_0 = \lambda_0(r_1, s)$, такой, что $\Lambda(\lambda_0) = 0$.

Имеем $|\partial_{r_1}^2 \eta(\theta, 0)| > k > 0$; поэтому можно выбрать $r_1 = r_1(s)$ таким образом, чтобы вектор $\lambda_0(r_1(s), s)$ был ортогонален (в евклидовом смысле) вектору β .

5) Покажем, что $\mu(\lambda_0(r_1(s), s)) = 0$.

Пусть $\omega_{A'}$ — постоянная симплектическая форма на $\mathbf{T}^{2n-1} \times \mathbf{R}$, ассоциированная с матрицей

$$A' = \begin{pmatrix} B & -\beta \\ \beta & 0 \end{pmatrix}.$$

Диффеоморфизм $\hat{F} = F_{\lambda_0(r_1(s), s)}$ является точным симплектическим (θ_1 — параметр гамильтоновой изотопии, связывающей тождественное отображение с $\hat{F}$). Если $d\Omega$ — форма объема, ассоциированная с $\omega_{A'}$, то форма $\hat{F}^* \Omega - \Omega$ точна. Но тогда сдвинутый тор действительно инвариантен.

6) Покажем, наконец, что $\lambda_0(r_1(s), s) = 0$. Применяя лемму о жесткости из п. 4.3 к вектору $Y = Z_{\lambda_0(r_1(s), s)}$, заключаем, что вектор $(1, \alpha) - (0, \lambda_0(r_1(s), s))$ $\omega_{A'}$ -ортогонален подпространству $\mathbf{R}^{2n} \times (0, 0)$ и, значит, является линейной комбинацией векторов $(1, \beta')$ и $(0, \beta)$. Поэтому векторы $\lambda_0(r_1(s), s)$ и β коллинеарны. А так как они еще и ортогональны, то $\lambda_0(r_1(s), s) = 0$.

4.6. Инвариантные торы коразмерности 1 с гамильтоновой точки зрения

4.6.1. Неплоские кривые. Рассмотрим C^∞ -кривую $\phi : [-\delta, +\delta] \rightarrow \mathbf{R}^n$. Пространство $\mathbf{R}^n$ снабжено евклидовой нормой. Пусть $0 < a < b$. Кривая ϕ называется (a, b) -*неплоской*, если для $|x| \leq \delta$, $v \in S^{n-1}$ выполняются неравенства

$$a \leq \sup_{1 \leq j \leq n} |\langle v, D^j \phi(x) \rangle| \leq b.$$

В частности, для всех $x \in [-\delta, +\delta]$ векторы $D\phi(x), \dots, D^n \phi(x)$ образуют базис пространства $\mathbf{R}^n$.

Пример. $\phi(t) = (t, t^2, \dots, t^n)$.

Русские математики, близкие Спринджуку, интересовались диофантовыми аппроксимациями неплоских кривых почти всюду. Здесь нам понадобится лишь элементарный результат Пяртли [Py].

Предложение. Пусть $\tau > n^2 - 1$. Тогда существует постоянная $C = C(a, b, \tau, \delta) > 0$, такая, что если кривая $\phi : [-\delta, +\delta] \rightarrow \mathbf{R}^n$ (a, b) -непланарна, то

$$m\{x \in [-\delta, +\delta], \quad \phi(x) \notin CD_n(\gamma, \tau)\} \leq C\gamma^{1/n}.$$

Замечание. Спринджук высказал гипотезу о том, что (в случае аналитической кривой ϕ) предыдущая оценка выполняется для всех $\tau > 0$. А ведь ее доказательство уже в случае $\tau = o(n^2)$ представляется трудным!

4.6.2. Рассмотрим вложение F_0 многообразия $\mathbf{B}^n(\delta)$ в $\mathbf{B}^n$ (ср. п. 4.1) вида

$$F_0(\theta, r) = (\theta + \ell(r), r).$$

Предположим, что отображение ℓ (a, b) -неплоское для некоторых постоянных $0 < a < b$.

Зафиксируем $\tau > n^2 - 1$ и достаточно маленькое $\gamma > 0$. Пусть $F : \mathbf{B}^n(\delta) \hookrightarrow \mathbf{B}^n$ — вложение, достаточно близкое к F_0 . Применяя теорему о сдвинутом торе, заключаем, что для $\alpha \in CD_n(\gamma, \tau)$ и $r \in [-\delta/2, +\delta/2]$ существуют сдвиг $R_{\lambda(r, \alpha), \mu(r, \alpha)}$ многообразия $\mathbf{B}^n$ и функция ψ на T^n со средним значением r , такие, что отображение $R_{\lambda(r, \alpha), \mu(r, \alpha)} \circ F$ сохраняет график $T_{\alpha, r}$ функции ψ и сопряжено сдвигу R_α .

Для $F = F_0$ имеем $\mu_0(\alpha, r) \equiv 0$, $\lambda_0(\alpha, r) = \alpha - \ell(r)$. Отображение $(\alpha, r) \mapsto \lambda(\alpha, r)$, определенное на $CD_n(\gamma, \tau) \times [-\delta/2, +\delta/2]$, продолжается до C^∞ -отображения $\mathbf{R}^n \times [-\delta/2, \delta/2] \rightarrow \mathbf{R}^n$ (ср. п. 4.1), C^∞ -близкого к λ_0 . Поэтому уравнение $\lambda(\alpha, r) = 0$ неявно задает кривую $\ell_F : [-\delta/2, \delta/2] \rightarrow \mathbf{R}^n$, такую, что $\lambda(\ell_F(r), r) = 0$. Так как кривая ℓ_F C^n -близка к ℓ , то она $(a/2, 2b)$ -неплоская.

Если число $r_0 \in [-\delta/2, +\delta/2]$ таково, что $\ell_F(r_0) \in CD_n(\gamma, \tau)$, то F сдвигает тор $T_{\ell_F(r_0), r_0}$; это имеет место, согласно предложению 4.6.1, для множества K значений r_0 , имеющего относительно большую меру (при маленьких γ) в $[-\delta/2, \delta/2]$.

4.6.3. Диффеоморфизмы, сохраняющие объем. В условиях предыдущего раздела предположим, что вложение F сохраняет форму объема $d\Omega = d(r d\theta_1 \wedge \cdots \wedge d\theta_n)$ и что поток диффеоморфизма F нулевой, т. е. форма $F^*\Omega - \Omega$ точна. В этом случае каждый из сдвинутых торов инвариантен, и мы получаем следующее утверждение.

Теорема [Н7]. Для всякого вложения $F: \mathbb{V}^n \hookrightarrow \mathbb{V}^n$ с нулевым потоком, сохраняющего объем и достаточно близкого к F_0 , существуют канторово множество $K \subset [-\delta/2, \delta/2]$ положительной меры и отображение $r \mapsto \psi_r$ множества K в $C^\infty(\mathbb{T}^n)$, такие, что

- (i) для всех $r \in K$ выполняются условия $\int_{\mathbb{T}^n} \psi_r dt = r$, $\|D\psi_r\| < \varepsilon$;
- (ii) для всех $r \in K$ график отображения ψ_r является диофантовым тором T_r , инвариантным относительно F ;
- (iii) объединение всех торов T_r имеет положительную меру Лебега в $\mathbb{T}^n \times \mathbb{R}$.

Чен и Сунь [17] показали существование таких инвариантных торов при $n = 2$ и аналитическом вложении F . Похоже, что Де ла Ллав независимо (хотя и позднее) получил аналогичный результат. Приведенный здесь более общий и более точный результат принадлежит Эрману [Н7].

4.6.4. Вложим произведение $\mathbb{T}^n \times [-\delta, +\delta]$ в некоторую карту какого-нибудь атласа на произвольном компактном многообразии M^{n+1} . Рассмотрим диффеоморфизм G_0 многообразия M^{n+1} , класс изотопии которого может быть зафиксирован заранее, совпадающий с F_0 на $\mathbb{T}^n \times [-3\delta/4, 3\delta/4]$. Заметим, что нулевое сечение $\mathbb{T}^n \times \{0\}$ разбивает многообразие M .

Для любого диффеоморфизма G из некоторой окрестности U диффеоморфизма G_0 в $\text{Diff}^\infty(M)$ справедлива следующая замечательная альтернатива:

- либо G сохраняет некоторое канторово множество инвариантных торов, объединение которых имеет положительную меру Лебега, и индуцирует на них диофантовы вращения,
- либо G не является *цепно транзитивным*, т. е. в M существуют две точки x, x' , не лежащие одновременно ни в какой орбите никакого C^0 -возмущения диффеоморфизма G .

Если возмущение G сохраняет некоторую вероятностную меру μ , относительно которой все открытые множества имеют положительную меру, то второй член альтернативы невозможен. Заметим, что существенно рассматривать окрестность диффеоморфизма G_0 в C^∞ -топологии: Окстоби [О] показал, что в группе $\text{Homeo}_\mu(M)$ гомеоморфизмов компактного многообразия M , сохраняющих вероятностную меру μ (относительно которой мера каждой точки нуле-

вая, а каждого открытого множества — положительная), свойство иметь плотную орбиту является общим в смысле Бэра.

В случае если мера μ эквивалентна мере Лебега, объединение всех инвариантных торов имеет положительную μ -меру и все показатели Ляпунова ограничения G на это объединение равны нулю; так строится контрпример к гипотезе Песина [Pe].

5. ГЕОМЕТРИЯ И ДИНАМИКА ИНВАРИАНТНЫХ ЛАГРАНЖЕВЫХ ТОРОВ

5.1. Первая теорема Биркгофа. Пусть F — точный симплектический строго монотонный диффеоморфизм многообразия $\mathbf{A}^n$ (ср. п. 2.4). Запишем поднятие $\hat{F}$ диффеоморфизма F на $\mathbf{R}^n \times \mathbf{R}^n$ в виде

$$\hat{F}(\theta, r) = (\Theta, R) = (\theta + \phi_1(\theta, r), \phi_2(\theta, r)),$$

где функции ϕ_1, ϕ_2 $\mathbf{Z}^n$ -периодичны по θ .

Положим $\hat{F}^*v - v = dH$ и будем считать θ, Θ глобальными координатами на $\mathbf{R}^n \times \mathbf{R}^n$; функция H , определенная с точностью до аддитивной постоянной, называется *производящей функцией* диффеоморфизма $\hat{F}$. Она удовлетворяет условиям

$$H(\theta + k, \Theta + k) = H(\theta, \Theta), \quad k \in \mathbf{Z}^n,$$

$$\hat{F}(\theta, r) = (\Theta, R) \iff \begin{cases} \partial_\theta H(\theta, \Theta) = -r, \\ \partial_\Theta H(\theta, \Theta) = R. \end{cases}$$

Определение. Диффеоморфизм $\hat{F}$ (или F) называется *глобально положительным*, если функция H удовлетворяет одному из следующих эквивалентных условий:

- (i) H ограничена снизу;
- (ii) для всех $\theta \in \mathbf{R}^n$ функция $\Theta \mapsto H(\theta, \Theta)$ ограничена снизу;
- (iii) для всех $\Theta \in \mathbf{R}^n$ функция $\theta \mapsto H(\theta, \Theta)$ ограничена снизу;
- (iv) $\lim_{\|\theta - \Theta\| \rightarrow +\infty} \|\Theta - \theta\|^{-1} H(\theta, \Theta) = +\infty$.

Определение. График отображения $\psi \in C^0(\mathbf{T}^n, \mathbf{R}^n)$ называется *C^0 -лагранжевым*, если существуют такие $c \in \mathbf{R}^n, \eta \in C^1(\mathbf{T}^n)$, что $\psi = c + d\eta$ (другими словами, если отображение ψ , рассматриваемое как сечение кокасательного расслоения $T^*(\mathbf{T}^n)$, замкнуто в смысле теории обобщенных функций).

Теорема [Н6]. Пусть F — глобально положительный диффеоморфизм многообразия $\mathbf{A}^n$. Предположим, что F сохраняет C^0 -лагранжев график функции $\psi \in C^0(\mathbf{T}^n, \mathbf{R}^n)$. Тогда отображение ψ липшицево; кроме того, если положить $\widehat{F}(\theta, \psi(\theta)) = (\Theta, \psi(\Theta))$, то для почти всех $\theta \in \mathbf{T}^n$

$$D\psi(\theta) \geq -\partial_{\theta}^2 H(\theta, \Theta),$$

$$D\psi(\Theta) \geq -\partial_{\Theta}^2 H(\theta, \Theta)$$

(неравенства понимаются в смысле порядка на симметрических матрицах).

Набросок доказательства. Положим $\widehat{\psi}(\theta) = \langle c, \theta \rangle + \eta(\theta)$ (ср. определение выше) и обозначим через f гомеоморфизм пространства $\mathbf{R}^n$, такой, что

$$\widehat{F}(\theta, \psi(\theta)) = (f(\theta), \psi(f(\theta))).$$

Положим $K(\theta, \Theta) = H(\theta, \Theta) + \widehat{\psi}(\theta) - \widehat{\psi}(\Theta)$.

Имеем

$$\partial_{\theta} K(\theta, \Theta) = 0 \iff \Theta = f(\theta) \iff \partial_{\Theta} K(\theta, \Theta) = 0.$$

Так как $\lim_{\|\Theta - \theta\| \rightarrow +\infty} K(\Theta, \theta) = +\infty$, критическое многообразие $\{\Theta = f(\theta)\}$ реализует минимум функционала K . Распределения $\theta \mapsto \partial_{\Theta}^2 K|_{\Theta=f(\theta)}$ и $\Theta \mapsto \partial_{\theta}^2 K|_{\theta=f^{-1}(\Theta)}$, принимающие значения в симметрических матрицах, оказываются поэтому положительными. Теорема доказана.

5.2. Вторая теорема Биркгофа (возмущенный вариант). Грассманиан $\Lambda(n)$ лагранжевых плоскостей в $\mathbf{C}^n = \mathbf{R}^n \oplus i\mathbf{R}^n \simeq T^*\mathbf{R}^n$ отождествляется с многообразием $U(n)/O(n)$; выделенная точка соответствует горизонтальному подпространству $h_0 = \mathbf{R}^n \times \{0\}$. Отображение $\delta : U(n)/O(n) \rightarrow S^1$, $u \mapsto (\det_{\mathbf{C}} u)^2$, индуцирует изоморфизмы $\pi_1(\Lambda(n)) \rightarrow \pi_1(S^1)$, $H^1(S^1, \mathbf{Z}) \rightarrow H^1(\Lambda(n), \mathbf{Z})$.

Определение. Пусть $j : N \rightarrow \mathbf{A}^n$ — лагранжева иммерсия. Индексом Маслова пары (N, j) называется класс $J^*(\tau) \in H^1(N, \mathbf{Z})$, где через τ обозначена каноническая образующая группы $H^1(\Lambda(n), \mathbf{Z})$ и $J : x \mapsto T_j(T_x N)$ — это отображение $N \rightarrow \Lambda(n)$, порожденное отображением j .

Теорема [V]. *Индекс Маслова вложенного лагранжева тора, гомотопного нулевому сечению $\{r = 0\} \subset \mathbf{A}^n$, равен нулю.*

Определение. Пусть f — непрерывное отображение метрического пространства (X, d) в себя. Точка $x \in X$ *цепно рекуррентна*, если для всякого $\varepsilon > 0$ существуют такое $n > 0$ и такая последовательность точек $x = x_0, x_1, \dots, x_n = x$, что $d(fx_i, x_{i+1}) < \varepsilon$ для всех i , $0 \leq i < n$. Отображение f *цепно рекуррентно*, если всякая точка цепно рекуррентна.

Для $R > 0$ положим $\mathbf{A}^n(R) = \mathbf{T}^n \times [-R, +R]^n$.

Теорема [H6]. Пусть $L(\theta, r) = (\theta + \ell(r), r)$ — вполне интегрируемый симплектический диффеоморфизм многообразия $\mathbf{A}^n(R)$. Предположим, что $D\ell(r) > 0$ для всех $r \in [-R, +R]^n$.

Пусть $\varepsilon > 0$. Всякое симплектическое вложение $F : \mathbf{A}^n(R) \hookrightarrow \mathbf{A}^n$, достаточно C^1 -близкое к отображению L , обладает следующим свойством: всякое связное лагранжево компактное подмногообразие $N \subset \mathbf{A}^n(R)$ с нулевым индексом Маслова, инвариантное относительно F и такое, что ограничение $F|_N$ цепно рекуррентно, является тором, графиком некоторого отображения $\psi \in C^1(\mathbf{T}^n, \mathbf{R}^n)$, удовлетворяющего условию $\|D\psi\| < \varepsilon$.

Следующее замечание является решающим в доказательстве теоремы: обозначим через $\tilde{L} : \mathbf{A}^n(R) \times \tilde{\Lambda}(n) \rightarrow \mathbf{A}^n(R) \times \tilde{\Lambda}(n)$ (где $\tilde{\Lambda}(n)$ — универсальная накрывающая многообразия $\Lambda(n)$) отображение, отвечающее TL ; если точка $(z, \tilde{h}) \in \mathbf{A}^n(R) \times \tilde{\Lambda}(n)$ цепно рекуррентна относительно некоторого малого C^0 -возмущения отображения $\tilde{L}$, то проекция плоскости $\tilde{h}$ в $\Lambda(n)$ близка к горизонтальной плоскости h_0 .

Относительно невозмущенных обобщений второй теоремы Биркгофа см. [Vi, B-P2, Pol].

5.3. Динамика общего положения [H5]. Рассмотрим, как и выше, симплектический вполне интегрируемый диффеоморфизм

$$L(\theta, r) = (\theta + \ell(r), r)$$

многообразия $\mathbf{A}^n(R)$ и предположим дополнительно, что $D\ell(r) > 0$ для всех $r \in [-R, +R]^n$.

Для симплектического вложения $F : \mathbf{A}^n(R) \hookrightarrow \mathbf{A}^n$, C^1 -близкого к L , обозначим через $IT^\infty(F)$ множество гладких инвариантных

диофантовых торов, гомотопных нулевому сечению $\{r = 0\}$. Такой тор автоматически лагранжев (ср. п. 2.7), и по теореме Витербо его индекс Маслова равен нулю. По второй теореме из п. 5.2 это означает, что существует C^1 -окрестность $\mathcal{U}$ отображения L , такая, что для $F \in \mathcal{U}$ всякий тор $T \in IT^\infty(F)$ является графиком некоторой функции $\psi_T \in C^\infty(T^n, [-R, +R]^n)$, удовлетворяющей условию $\|D\psi_T\| \leq 1$.

Введем на пространстве S липшицевых отображений из T^n в $[-R, +R]^n$ с постоянной Липшица ≤ 1 C^0 -топологию. Пространство S компактно. Отображение $T \mapsto \psi_T$ превращает множество $IT^\infty(F)$ в подмножество пространства S ; обозначим через $IT(F)$ его замыкание в S . Для всякого отображения $\psi \in IT(F)$ его график T_ψ является C^0 -лагранжевым F -инвариантным тором.

Для любого $F \in \mathcal{U}$ и всякого отображения $\psi \in IT(F)$ ограничение $F|_{T_\psi}$ *цепно рекуррентно*: график T_ψ является пределом диофантовых торов, а множество цепно рекуррентных отображений замкнуто.

Для всех $F \in \mathcal{U}$ и некоторого остаточного множества отображений $\psi \in IT(F)$

- отображение ψ принадлежит классу C^1 ;
- ограничение отображения F на T_ψ минимально, однозначно эргодично, и его число вращения есть число Лиувилля.

Наконец, для общего отображения $F \in \mathcal{U}$ и остаточного множества отображений $\psi \in IT(F)$ ограничение отображения F на T_ψ имеет нулевую топологическую энтропию, не является перемешивающим, является слабо перемешивающим и единственная мера, инвариантная относительно этого отображения, сингулярна и ее носитель является борелевским множеством хаусдорфовой размерности 0.

Эти последние свойства, прежде всего свойство слабого перемешивания, показывают, насколько динамика на общем торе отличается от динамики поворота!

Другие общие свойства и доказательства приведены в [Н5], где демонстрируется вся мощь методов, основанных на категории Бэра.

Замечание. Может ли лагранжев тор, инвариантный относительно общего симплектического диффеоморфизма симплектического многообразия M^{2n} , содержать периодическую орбиту? При $n = 1$ легко показать, что это не так; Арно [А] показал, что для $n = 2$ это также невозможно. При $n > 2$ вопрос остается открытым.

5.4. Индефинитное кручение [Н4].

5.4.1. Предположения на F из предыдущих пп. 5.2, 5.3 выполняются, в частности, для всех симплектических диффеоморфизмов некоторой окрестности гладкого инвариантного лагранжева диофантова тора, если кручение этого тора *положительно* (ср. 2.7); это вытекает из нормальной формы Биркгофа. Аналогичное утверждение имеет место и для отрицательного кручения.

При возмущении вполне интегрируемого диффеоморфизма с вырожденным или индефинитным кручением ситуация изменяется драматически.

5.4.2. Пример. Положим

$$L(\theta_1, \theta_2, r_1, r_2) = (\theta_1 + r_2, \theta_2 + r_1, r_1, r_2),$$

$$F(\theta_1, \theta_2, r_1, r_2) = (\theta_1 + r_2, \theta_2 + r_1, r_1 + \varepsilon \varphi(\theta_1 + r_2), r_2),$$

где

$$0 < \varepsilon \ll 1, \quad \varphi \in C^\infty(\mathbb{T}^1), \quad \int_{\mathbb{T}^1} \varphi dm = 0.$$

Для всех $\alpha \in \mathbb{R}$ подмногообразие $M_\alpha = \{r_2 = \alpha\}$ инвариантно относительно F . Для $n \geq 0$ имеем $F^n(\theta_1, \theta_2, r_1, \alpha) = (\Theta_1^{(n)}, \Theta_2^{(n)}, R_1^{(n)}, \alpha)$, где

$$\Theta_1^{(n)} = \theta_1 + n\alpha,$$

$$\Theta_2^{(n)} = \theta_2 + nr_1 + \varepsilon \sum_1^n (n-i) \varphi(\theta_1 + i\alpha),$$

$$R_1^{(n)} = r_1 + \varepsilon \sum_1^n \varphi(\theta_1 + i\alpha).$$

1-й случай: уравнение

$$\psi(\theta_1 + \alpha) - \psi(\theta_1) = \varphi(\theta_1 + \alpha)$$

имеет решение $\psi \in C^0(\mathbb{T}^1)$, для которого $\int_{\mathbb{T}^1} \psi dm = 0$. Так происходит (при $\psi \in C^\infty(\mathbb{T}^1)$), если α диофантово. Торы T_c , заданные уравнением $\{r_1 = \varepsilon \psi(\theta_1) + c, r_2 = \alpha\}$, инвариантны относительно F , а их объединение есть M_α . На торе T_c отображение F C^0 -сопряжено повороту $R_{\alpha, c}$ в том и только в том случае, если у уравнения

$$\eta(\theta_1 + \alpha) - \eta(\theta_1) = \psi(\theta_1)$$

есть решение $\eta \in C^0(\mathbb{T}^1)$.

2-й случай: у уравнения

$$\psi(\theta_1 + \alpha) - \psi(\theta_1) = \varphi(\theta_1 + \alpha)$$

нет решения $\psi \in C^0(\mathbf{T}^1)$. Согласно Готтшалку и Хедлунду, для всех $(\theta_1, \theta_2, r_1, \alpha) \in M_\alpha$ выполняется равенство

$$\sup_{n \geq 1} |R_1^{(n)}| = +\infty.$$

Предположим, что среди коэффициентов Фурье функции φ бесконечно много ненулевых. Тогда

- для почти всех $\alpha \in \mathbf{R}$ M_α является объединением гладких инвариантных лагранжевых торов;
- для некоторого G_δ -плотного множества значений α орбита каждой точки из M_α неограниченна.

Легко проверить, что инвариантные лагранжевы торы не могут удовлетворять никакому из априорных неравенств. Сопрягая диффеоморфизм F точным симплектическим диффеоморфизмом, близким к тождественному, можно получить гладкие инвариантные диофантовы лагранжевы торы, гомотопные нулевому сечению $\{r_1 = r_2 = 0\}$ и не являющиеся графиками.

5.4.3. Неустойчивость кручения. Для возмущения F вполне интегрируемого симплектического диффеоморфизма L с *положительным* кручением, как в пп. 5.2, 5.3, нетрудно показать, что всякий инвариантный тор $T \in IT^\infty(F)$ тоже имеет положительное кручение. В частности, это имеет место, когда F описывает (посредством нормальной формы Биркгофа) динамику симплектического диффеоморфизма в окрестности инвариантного лагранжева диофантова тора с положительным кручением.

Ситуация меняется радикально в случае indefinitного кручения. Пусть F — симплектический диффеоморфизм симплектического многообразия (M, ω) и T — (гладкий) диофантов лагранжев тор, инвариантный относительно F . Предположим, что кручение диффеоморфизма F вдоль T вырожденно или indefinitно.

Теорема. *Существует C^∞ -возмущение $\tilde{F}$ диффеоморфизма F , которое сохраняет некоторый диофантов лагранжев тор $\tilde{T}$, C^∞ -близкий к T , с произвольной заранее заданной, отличной от тождественного нуля сигнатурой кручения вдоль $\tilde{T}$.*

Неизвестно, можно ли полностью аннулировать кручение малым C^∞ -возмущением.

5.4.4. Неустойчивость геометрии торов. Патология примера 5.4.2 при условии сохранения indefinitного кручения частично распространяется и на более общую ситуацию.

Рассмотрим вполне интегрируемый симплектический диффеоморфизм $L(\theta, r) = (\theta + \ell(r), r)$ многообразия $\mathbf{A}^n$. Предположим, что матрица $D\ell(0)$ indefinitна или вырожденна; пусть $v \in \mathbf{R}^n$, $v \neq 0$, — такой вектор, что $\langle v, D\ell(0)v \rangle = 0$.

Возмущая при необходимости L (в классе вполне интегрируемых симплектических диффеоморфизмов), можно предполагать, что $\ell(0) \in \mathbf{Q}^n$, $v \in \mathbf{Q}^n$, $\langle v, D_0^2 \ell(v, v) \rangle \neq 0$. С помощью линейной замены координат и взятия конечного накрытия добьемся того, чтобы $\ell(0) = 0$, $v = (1, 0, \dots, 0)$.

Тогда функция $\widehat{\ell}(r)$, такая, что $D\widehat{\ell} = \ell$, имеет вид

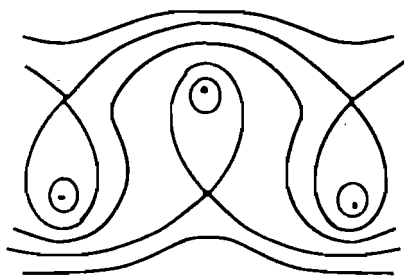
$$\widehat{\ell}(r_1, 0) = ar_1^3 + O(r_1^4), \quad a \neq 0.$$

Рассмотрим гамильтониан

$$H(\theta, r) = \widehat{\ell}(r) - \varepsilon \cos 2\pi\theta_1 + \sigma_1^n \beta_i r_i,$$

где $0 < |\beta_1|, \dots, |\beta_n|, |\varepsilon| \ll 1$.

Гамильтоново поле X_H имеет n интегралов в инволюции $r_2, \dots, r_n, H$. Поверхности уровня при фиксированных значениях $r_2 = \dots = r_n = 0$ первых $n - 1$ интегралов и переменном значении последнего интеграла H представляют собой произведение $(n - 1)$ -мерного тора (с координатами $\theta_1, \dots, \theta_n$) на кривые уровня функции $(\theta_1, r_1) \mapsto \widehat{\ell}(R_1, 0) - \varepsilon \cos 2\pi\theta_1 + \beta_1 r_1$. Эти кривые уровня изображены на рисунке ниже (для $a > 0$, $\beta_1 < 0$).



Поток поля X_H за единичное время является малым возмущением F диффеоморфизма L (отвечающего случаю $\varepsilon = \beta_1 = \dots = \beta_n = 0$) в C^∞ -топологии. У диффеоморфизма F есть инвариантные гладкие лагранжевы диофантовы торы (при подходящем выборе β_i), не являющиеся графиками, которые накапливаются в топологии Хаусдорфа на инвариантном компакте, не являющемся многообразием.

Кроме того, можно добиться, чтобы такой тор снова имел indefinite кручение, а затем повторить предыдущую конструкцию

ЛИТЕРАТУРА

- [A.M.] Abraham R., Marsden J. Foundations of Mechanics, 2nd edition, Benjamin Cummings, Reading, Mass., 1970.
- [An] Angenent S. B. Monotone recurrence relations, their Birkhoff orbits and topological entropy. Erg. Th. Dyn. Sys., 1 (1990), 15–41.
- [A] Arnaud M.-C. Points périodiques et tores lagrangiens invariants des difféomorphismes symplectiques, Thèse Univ. Paris VII, 1990.
- [Ar1] Арнольд В. И. Математические методы классической механики. — М.: Наука, 1974 (переиздания 1984, 1989).
- [Ar2] Арнольд В. И. Малые знаменатели. I. Об отображениях окружности на себя. — Изв. АН СССР, сер. матем., т. 25, 1961, № 1, с. 21–86.
- [Ar3] Арнольд В. И. Доказательство теоремы А. Н. Колмогорова о сохранении условно-периодических движений при малом изменении функции Гамильтона. — УМН, т. 18, 1963, вып. 5, с. 13–40.
- [Ar4] Арнольд В. И. (ред.), Итоги науки и техники. Современные проблемы математики. Фундаментальные направления. Т. 3: Динамические системы-3. — М.: ВИНТИ, 1985.
- [Bi] Bialyi M. L. Aubry-Mather sets and Birkhoff's theorem for geodesic flows on the two-dimensional torus, Comm. Math. Phys., 126 (1989), 13–24.
- [B-P1] Бялый М. Л., Полтерович Л. В. Геодезические потоки на двумерном торе и фазовые переходы «соизмеримость-несоизмеримость». — Функц. анализ и прил., т. 20, 1986, вып. 4, с. 9–16.
- [B-P2] Bialyi M. L., Polterovich L. Hamiltonian diffeomorphisms and Lagrangian distributions, Geom. Funct. Anal., 2 (1992), № 2, 173–210.
- [B1] Birkhoff G. D. Surface transformations and their dynamical applications, Acta Math., 43 (1920), 1–119.
- [B2] Birkhoff G. D. Note sur la stabilité, J. Math., 15 (1936), 339–344; Coll. Math. Papers, vol. 2, 1968, 662–667.

- [B-K] Birkhoff G.D, Koopman B.O. Recent contributions to the ergodic theory, Proc. Nat. Ac. Sc., **18** (1932), 279–282; Coll. Math. Papers, vol. 2, 1968, 462–466.
- [B-H] Boyland P., Hall G.R. Invariant circles and the order structure of periodic orbits in monotone twist maps, Topology, **26** (1987), 21–35.
- [Bo] Bost J.B. Tores invariantes des systèmes dynamiques hamiltoniens, Sémin. Bourbaki, exp. 639, Astérisque, **133–134** (1986), 113–157.
- [Ch] Chenciner A. La dynamique au voisinage d'un point fixe elliptique conservatif: de Poincaré et Birkhoff à Aubry et Mather, Sémin. Bourbaki, exp. 662, Astérisque, **121–122** (1985), 147–170.
- [C-S] Cheng C.Q., Sun Y.S. Existence of invariant tori in three-dimensional measure preserving mappings, Cel. Mech., **47** (1990), 275–293.
- [D] Douady R. Stabilité et instabilité des points fixes elliptiques, Ann. Sci. Ecole Norm. Sup., 4^e série, **21** (1988), 1–46.
- [Eh] Ehrenfest P., Ehrenfest T. The conceptual foundations of statistical mechanics (перевод на англ. язык работы 1912 г.), Dover publ., 1959, 1990.
- [E] Eliasson H. Invariant tori for Hamiltonian systems, Ann. Sc. Norm. Pisa, **15** (1988), 115–147.
- [F] Fathi A. Une interprétation plus topologique de la démonstration du théorème de Birkhoff, Приложение к гл. I в [H9, vol. 1].
- [G] Gutierrez C. A counterexample to a C^2 closing lemma, Erg. Th. Dyn. Sys., **7** (1987), 509–530.
- [Ha1] Hamilton R.S. The inverse function theorem of Nash and Moser, Preprint, 1974.
- [Ha2] Hamilton R.S. The inverse function theorem of Nash and Moser, Bull. Amer. Math. Soc., **7** (1982), 65–222.
- [H1] Herman M.R. Exemple de flots hamiltoniens dont aucune perturbation en topologie C^∞ n'a d'orbites périodiques sur un ouvert de surfaces d'énergie, C. R. Acad. Sci. Paris, serie I, **312** (1991), 989–994.
- [H2] Herman M.R. Différentiabilité optimale et contre-exemples à la fermeture en topologie C^∞ des orbits récurrentes des flots hamiltoniens, C. R. Acad. Sci. Paris, serie I, **313** (1991), 49–51.
- [H3] Herman M.R. Stabilité topologique des systèmes dynamiques conservatifs, Atas do 18^o Col. Bras. Mat. (1992).
- [H4] Herman M.R. Dynamics connected to indefinite torsion, Proc. Conf. «Twist maps and their applications», I.M.A., Univ. Minnesota, Springer-Verlag.
- [H5] Herman M.R. On the dynamics on Lagrangian tori invariant under symplectic diffeomorphisms, Preprint École Polytechnique, 1990.
- [H6] Herman M.R. Inégalités a priori pour des tores lagrangiens invariants par des difféomorphismes symplectiques, Publ. Math. I.H.E.S., **70** (1989), 47–101.

- [H7] Herman M. R. Théorème des tores translatés et quelques applications à la stabilité topologique des systèmes dynamiques conservatifs, *готовится к печати*.
- [H8] Herman M. R. Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations, *Publ. Math. I.H.E.S.*, **49** (1979), 5–233.
- [H9] Herman M. R. Sur les courbes invariantes par les difféomorphismes de l'anneau, vol. 1, *Astérisque*, **103–104** (1983), vol. 2, *Astérisque*, **144** (1986).
- [H10] Herman M. R. Démonstration du théorème des translatées par les difféomorphismes de l'anneau; démonstration du théorème des tores invariants, *рукопись*, 1980.
- [Ho] Hörmander L. The boundary problem of physical geodesy, *Arch. Rat. Mech. Anal.*, **62** (1976), 1–52.
- [K] Колмогоров А. Н. Общая теория динамических систем и классическая механика. — *Proc. Intern. Cong.*, 1954, v. 1, North Holland, Amsterdam, 315–333; см. также в кн.: Труды Международного математического конгресса, Амстердам, 1954, Обзорные докл. — М.: Изд-во АН СССР, 1961, с. 187–208.
- [La] Лазуткин В. Ф. Существование каустик для бильярдной задачи в выпуклой области. — *Изв. АН СССР*, т. 7, 1973, № 1, с. 186–216.
- [L1] Le Calvez P. Propriétés dynamiques des zones d'instabilité, *Ann. Sci. École Norm. Sup.*, 4^e série, **20** (1987), 443–464.
- [L2] Le Calvez P. Propriétés générales des applications déviant la verticale, *Bull. Soc. Math. France*, **117** (1989), 69–102.
- [L3] Le Calvez P. Étude topologique des applications déviant la verticale, *Ensaio Mat.*, Vol. 2 (1990).
- [L4] Le Calvez P. Propriétés dynamiques des difféomorphismes de l'anneau et du tore, *Astérisque*, **204** (1991).
- [M1] Mather J. The existence of quasiperiodic orbits for twist homeomorphisms of the annulus, *Topology*, **21** (1982), 457–467.
- [M2] Mather J. Minimal measures, *Comm. Math. Helv.*, **64** (1989), 375–394.
- [M3] Mather J. Variational construction of orbits of twist diffeomorphisms, *Preprint*, 1990.
- [M4] Mather J. A criterion for the non existence of invariant circles, *Publ. Math. I.H.E.S.*, **63** (1986), 153–204.
- [M5] Mather J. Destruction of invariant circles, *Erg. Th. Dyn. Sys.*, **8** (1988), 199–214.
- [Mo1] Moser J. On invariant curves of area preserving mappings of an annulus, *Nachr. Ak. Wiss. Gott., Math. Phys. Kl.* (1962), 1–20.
- [Mo2] Moser J. A rapidly convergent iteration method and nonlinear partial differential equations 2, *Ann. Sc. Norm. Pisa*, **20** (1966), 499–535.
- [O] Oxtoby J. C. Note on transitive transformations, *Proc. Nat. Ac. Sc.* (1937), 443–446.
- [Pe] Песин Я. Б. Характеристические показатели Ляпунова и гладкая эргодическая теория. — *УМН*, т. 92, 1977, № 4, с. 60.

- [P] Poincaré H. Les méthodes nouvelles de la mécanique céleste, 3 vol., Gauthier-Villars, Paris, 1892.
- [Pol] Polterovich L. V. The second Birkhoff theorem for optical Hamiltonian systems, Proc. Amer. Math. Soc., 113 (1991), No. 2, 513–516.
- [Po] Poschel J. Integrability of Hamiltonian systems on Cantor sets, Comm. Pure Appl. Math., 35 (1982), 653–695.
- [Pu] Pugh C. The closing Lemma, Amer. J. Math., 89 (1967), 956–1009.
- [P-R] Pugh C., Robinson C. The C^1 closing lemma including Hamiltonians, Erg. Th. Dyn. Sys., 3 (1983), 261–314.
- [Py] Пяртли А. С. Диофантовы приближения на подмногообразиях евклидова пространства. — Функц. анализ и прил., т. 3, 1969, вып. 4, 59–62.
- [R1] Rüssmann H. Kleine Nenner I: Über invarianten Kurven differenzierbarer Abbildungen eines Kreisinges, Nachr. Ak. Wiss. Gött., Math. Phys. Kl. (1970), 67–105.
- [R2] Russmann H. On the existence of invariant curves of twist mappings of an annulus, in: Lect. Notes in Math., vol. 1007, 1983, 677–712.
- [R3] Russmann H. Non-degeneracy in the perturbation theory of integrable dynamical systems, in: Number Theory and Dynamical Systems, Dodson-Vickers (eds.), Lect. Notes, vol. 134, Lond. Math. Soc., 1989, 1–18.
- [S-Z] Salamon D., Zehnder E. KAM theory in configuration space, Comm. Math. Helv., 64 (1989), 84–142.
- [V] Viterbo C. A new obstruction to embedding Lagrangian tori, Invent. Math., 100 (1990), 301–320.
- [Y] Yoccoz J.-C. Conjugaison différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition diophantienne, Ann. Sci. Éc. Norm. Sup., 4^e série, 17 (1984), 333–359.
- [Z1] Zehnder E. Remarks on periodic solutions on hypersurfaces, in: Periodic solutions of hamiltonian systems and related topics, P. H. Rabinowitz et coll., D. Reidel Co., Dordrecht, 1987, 91–140.
- [Z2] Zehnder E. Generalized implicit function theorems with application to some small divisors problems I, Comm. Pure Appl. Math., 28 (1975), 91–140.

JEAN-CHRISTOPHE YOCOZ

Université de Paris-Sud
 Département de Mathématiques
 Bâtiment 425
 F-91405 ORSAY CEDEX

ГАРМОНИЧЕСКИЙ АНАЛИЗ НА ФРАКТАЛЬНЫХ ПРОСТРАНСТВАХ

Мартин Барлоу

1. ВВЕДЕНИЕ

Интерес к этой области возник в начале 1980-х годов у специалистов по математической физике, изучавших *свойства проводимости* неупорядоченных сред. Имелись основательные причины верить, что фракталы могут оказаться хорошей моделью для таких сред — это и привело к изучению вопросов, связанных с волновым уравнением и уравнением теплопроводности на фрактальных пространствах; см., например, [RT].

До сих пор математическая работа концентрировалась главным образом на более простых уравнениях — уравнении Лапласа и уравнении теплопроводности. Она началась с вероятностных рассмотрений в работах Кусуоки [K1], Голдстейна [G] и Барлоу—Перкинса [BP], а аналитические рассмотрения, существенно использующие формы Дирихле, были проведены преимущественно в Японии; см. [Kig1, Kig2, F2, K2].

Наибольшие проблемы и трудности возникли, естественно, на микроскопических масштабах из-за отсутствия какой-либо евклидовой структуры. Однако для регулярного фрактала, например салфетки Серпинского, можно определить *предфрактальное* многообразие или граф, крупномасштабная структура которого изображает структуру настоящего фрактала. Эти предфракталы являются совершенно классическими объектами, но классическая техника не способна указать, например, правильные границы ядра теплопроводности на таком объекте.

Теперь более конкретно опишем проблемы, которые будут рассмотрены.

Пусть $F \subset \mathbb{R}^d$ — связный самоподобный фрактал, $d_f(F)$ — его размерность Хаусдорфа, а μ_F — хаусдорфова $d_f(F)$ -мера на F . Уравнение теплопроводности на F имеет вид

$$\Delta_F u = \frac{\partial u}{\partial t}, \quad u(x, 0) = u_0(x), \quad x \in F,$$

где $u : F \times \mathbf{R}_+ \rightarrow \mathbf{R}$, $u_0 \in C(F)$, а Δ_F — оператор Лапласа, действующий на подпространстве $\mathcal{D}(\Delta_F) \subset C(F)$. Немедленно возникают следующие проблемы:

- (i) *Существование*. Конструкция подходящего оператора Δ_F , являющегося F -изотропным, т. е. локально инвариантным относительно локальных изометрий фрактала F .
- (ii) *Единственность*. Определяется ли оператор Δ_F свойством F -изотропности?
- (iii) *Свойства*. Свойства решений уравнения Лапласа и уравнения теплопроводности, ассоциированных с Δ_F , и форма спектра оператора Δ_F .

К этому мы добавим и вопросы о марковском процессе X_t с генератором Δ_F и полугруппой $P_t = \exp(t\Delta_F)$.

2. ОПРЕДЕЛЕНИЕ ФОРМЫ ДИРИХЛЕ

В первых работах рассматривалось множество, которое является простейшим нетривиальным связным фракталом, — салфетка Серпинского. Дальнейшие работы распространили большинство полученных результатов на более широкий класс множеств: *гнездовые фракталы* [L], *p.s.f.-самоподобные множества*, а также на более широкий класс, включающий ковры Серпинского [KZ]. Однако в этом обзоре я, по большей части, ограничусь подмножеством гнездовых фракталов, которое достаточно широко для того, чтобы прочувствовать основные особенности предмета. На самом деле ясно, что более сложные множества демонстрируют тот же тип поведения, что и простые, и лишь доказательства аналогичных результатов становятся сложнее.

Пусть $F_0 = \{a_1, \dots, a_k\}$ — множество вершин правильного k -угольника в $\mathbf{R}^2$, а H_0 — замкнутая выпуклая оболочка множества F_0 . Пусть $\lambda \geq 1$, $M \geq k$, $a_{k+1}, \dots, a_M \in \mathbf{R}^2$ и преобразования ψ_i , $1 \leq i \leq M$, определены формулой

$$\psi_i(x) = a_i + \lambda^{-1}(x - a_i).$$

Мы будем предполагать, что

- (A1) (Симметричность) Множество $\{a_i, 1 \leq i \leq M\}$ обладает такими же симметриями, как и F_0 .
- (A2) (Вложенность) $\psi_i(H_0) \subset H_0$.
- (A3) (Связность) Множество $H_1 = \bigcup_{i=1}^M \psi_i(H_0)$ является связным.

- (A4) (Условие на открытые множества) Множества $\text{Int}(\psi_i(H_0))$, $1 \leq i \leq M$, не пересекаются.
- (A5) (Конечная разветвленность) Если $x, y \in H_0$ и $\psi_i(x) = \psi_j(y)$ при $i \neq j$, то $x, y \in F_0$.

Замечание. Отметим, что предположение (A5) отмечает фракталы типа ковра Серпинского. Это предположение на самом деле весьма сильно и, как мы увидим, позволяет рассматривать разные «уровни» фрактала по-отдельности. В большей части работ, сделанных до сих пор, это предположение используется существенным образом. Тем не менее, ценой нелегкой работы возможно разобраться с ковром Серпинского, по крайней мере, в размерности два (см. [BB1, BB2, BB3, KZ]).

Пусть $\Psi(\cdot) = \bigcup_{i=1}^M \psi_i(\cdot)$. Положим

$$F = \bigcap_{n=0}^{\infty} \Psi^n(H_0), \quad F_n = \Psi^n(H_0).$$

Множество F является *гнездовым фракталом* размерности $d_f(F) = \log M / \log \lambda$ и аппроксимируется конечными множествами F_n .

Полезно ввести координаты с помощью абстрактных пространств последовательностей. Пусть $I = \{1, \dots, M\}$; для $w = (i_1, \dots, i_n) \in I^n$ положим

$$\psi_w = \psi_{i_1, i_2, \dots, i_n} = \psi_{i_1} \circ \dots \circ \psi_{i_n}.$$

Мы будем называть множества $\psi_w(F_0)$, $w \in I^n$, *n-клетками*, а множества $\psi_w(H_0)$, $w \in I^n$, *n-комплексами*. Для $w \in I^N$ пишем $w|n = (i_1, \dots, i_n) \in I^n$ и заметим, что пересечение

$$\psi_w(H_0) = \bigcap_{n=0}^{\infty} \psi_{w|n}(H_0)$$

состоит из одной точки, обозначаемой $\phi(w)$. Легко проверяется, что отображение $\phi : I^N \rightarrow F$ непрерывно и сюръективно, но не инъективно. В случае салфетки Серпинского, например, имеем $\phi(12) = \phi(21)$. Фактически можно определить фрактал чисто абстрактным способом, рассматривая множество I^N с точностью до эквивалентности, отождествляющей подходящие последовательности указанного вида — это сделано в [Kig2].

Приступим теперь к определению форм Дирихле на конечных множествах F_n .

Пусть $\Lambda = \{\alpha \in \mathbf{R}^{k-1} : \alpha_i = \alpha_{k-i}, 1 \leq i \leq k-1\}$, и положим для $x, y \in F_n, \alpha \in \Lambda$

$$a_n(\alpha, x, y) = \begin{cases} \alpha_i, & \text{если } x, y \text{ принадлежат одной } n\text{-клетке} \\ & \text{и отстоят на } i \text{ шагов по окружности,} \\ 0 & \text{в противном случае,} \end{cases}$$

$$\mu_n(\alpha, x) = \sum_y a_n(\alpha, x, y).$$

Заметим, что $\mu_n(\alpha, F_n) = M^n \sum \alpha_i$ и что $\mu_n(\alpha, x)$ зависит от α только через сумму $\sum \alpha_i$. Для $f \in B(F_n) = \mathbf{R}^{F_n}$ положим

$$\mathcal{E}_\alpha^n(f, f) = \sum_x \sum_y a_n(\alpha, x, y)(f(x) - f(y))^2. \quad (2.1)$$

Мы можем интерпретировать форму Дирихле $\mathcal{E}$ несколькими способами.

1. Если мы рассмотрим $(F_n, a_n(\alpha))$ как электрическую цепь, в которой перемычка между x и y имеет проводимость $a_n(\alpha, x, y)$, то $\mathcal{E}_\alpha^n(f, f)$ есть диссипация энергии, если к цепи приложен потенциал f .

2. Дискретный лапласиан Δ_α^n на F_n определяется формулой

$$\Delta_\alpha^n f(x) = \mu_n(\alpha, x)^{-1} \sum_y a_n(\alpha, x, y)(f(x) - f(y)),$$

и, обозначая через (f, g) скалярное произведение с весом $\mu(\alpha, \cdot)$, получаем

$$\mathcal{E}_\alpha^n(f, g) = -(\Delta_\alpha^n f, g), \quad f, g \in C(F_n).$$

Случайное блуждание X_r^n , $r \geq 0$, с генератором Δ_α^n и формой Дирихле $\mathcal{E}_\alpha^n$ определяется условием

$$P(X_r^n = y | X_{r-1}^n = x) = p_\alpha^n(x, y) = a_n(\alpha, x, y) / \mu_n(\alpha, x).$$

Мы обозначим через $P_\alpha^n(x, y)$ распределение для этого случайного блуждания.

3. Иногда полезно записать определение (2.1) в матричной форме. Пусть $A (= A^n(\alpha))$ есть матрица, определенная формулой

$$A_{xy} = \delta_{xy} \mu_n(\alpha, x, y) - a_n(\alpha, x, y);$$

тогда

$$\mathcal{E}_\alpha^n(f, f) = f^T A f. \quad (2.2)$$

Фундаментальное свойство конечно разветвленных фракталов состоит в том, что операция прореживания¹⁾ действует в этом случае самым непосредственным способом. Сейчас мы определим ее. Для данной формы $\mathcal{E}_\alpha^n$ и функции $g \in B(F_{n-1})$ определим прореженную форму Дирихле

$$\tilde{\mathcal{E}}_\alpha^n(g, g) = \inf \{ \mathcal{E}_\alpha^n(f, f) : f|_{F_{n-1}} = g \}. \quad (2.3)$$

Теорема 2.1. *Отображение $\tilde{\mathcal{E}}_\alpha^n$ является формой Дирихле и существует такой элемент $\tilde{\alpha}(\alpha) \in \Lambda$, что*

$$\tilde{\mathcal{E}}_\alpha^n = \mathcal{E}_{\tilde{\alpha}}^n.$$

Это утверждение проверяется непосредственно. Отметим, что минимизирующая функция f в (2.3) удовлетворяет соотношению

$$\Delta_n^\alpha f(x) = 0, \quad x \in F_n - F_{n-1} \quad (2.4)$$

и что f внутри $(n-1)$ -клетки $\Psi_w(H_0)$ зависит только от значений функции g на $I_w(F_0)$. Таким образом, доказательство немедленно сводится к случаю $n = 1$.

Теперь мы прокомментируем разные интерпретации этой операции.

1. В терминах электрических цепей мы можем рассматривать $(F_n, a_n(\alpha))$ как «черный ящик», в котором доступны лишь узлы из F_{n-1} . Тогда цепь $(F_n, a_n(\alpha))$ дает тот же отклик (в терминах диссипации энергии, токов и т. п.) на возбуждение в F_{n-1} , что и цепь $(F_{n-1}, a_{n-1}(\tilde{\alpha}))$.

2. Мы можем «проредить» случайное блуждание $(X_r^n, r \leq 0)$, положив

$$T_0 = \min\{r \geq 0 : X_r^n \in F_{n-1}\},$$

$$T_m = \min\{r \geq T_{m-1} : X_r^n \in F_{n-1} - \{X_{T_{m-1}}\}\},$$

$$\tilde{X}_r^n = X_{T_r}^n, \quad r \geq 0.$$

Тогда случайное блуждание $(\tilde{X}^n, P_\alpha^n)$ эквивалентно по распределению блужданию $(X^{n-1}, P_{\tilde{\alpha}}^{n-1})$.

¹⁾ В оригинале — decimation.

3. В матричных терминах мы запишем векторы $f \in \mathbf{R}^{F_n}$ так, чтобы в начале стояли координаты в F_{n-1} . Тогда

$$A_n(\alpha) = \begin{pmatrix} B_n(\alpha) & C_n(\alpha) \\ C_n(\alpha)^T & D_n(\alpha) \end{pmatrix},$$

а матричные коэффициенты, отвечающие форме Дирихле $\tilde{\mathcal{E}}_\alpha^n$, суть

$$\tilde{A}_n(\alpha) = B_n(\alpha) - C_n(\alpha)D_n(\alpha)^{-1}C_n(\alpha)^T.$$

Операция прореживания позволяет нам свести многие вопросы, касающиеся форм Дирихле $\mathcal{E}_\alpha^n$, к поведению итераций отображения $\tilde{\alpha}$. В частности, представляют особый интерес (аффинные) неподвижные точки функции $\tilde{\alpha}$, т. е. такие элементы $\beta \in \Lambda$, что

$$\tilde{\alpha}(\beta) = \beta/\rho \quad \text{для некоторого } \rho > 0. \quad (2.5)$$

Эти неподвижные точки отвечают процессам и операторам на F с правильными скейлинговыми свойствами относительно преобразований ψ_i . Поскольку отображение $\tilde{\alpha}$ является непрерывным, теорема о неподвижной точке, примененная к симплексу $\Lambda' = \{\alpha \in \Lambda : \sum \alpha_i = 1\}$ показывает, что хотя бы одна неподвижная точка существует. Применение отражений [L, Ch. V] показывает, что для гнездовых фракталов существует хотя бы одна неподвижная точка, которая невырождена в том смысле, что $\alpha_i > 0$ для всех i , а потому электрическая цепь $(F_1, a_1(\alpha))$ связна.

Общая проблема существования и единственности (невырожденных) неподвижных точек представляется достаточно сложной. Естественный подход состоит в том, чтобы показать, что отображение $\tilde{\alpha}$ есть стягивание на Λ' в смысле некоторой метрики d , но найти такую метрику нелегко. В [B1] имеются некоторые частные результаты, приводящие к подходу, который может оказаться эффективным для общих гнездовых фракталов, но который не работает для общих р.с.ф.-самоподобных множеств. Кажется также правдоподобным, что любая невырожденная неподвижная точка β будет также устойчивой; и опять это было бы установлено, если бы можно было найти подходящее стягивание.

Проблема. Может ли отображение $\tilde{\alpha}$ иметь более одной невырожденной неподвижной точки? Всегда ли невырожденные неподвижные точки являются устойчивыми?

Замечание. Легко видеть, что отображение $\tilde{\alpha}$ может иметь одну или более вырожденных неподвижных точек — примеры имеются в [L] и [B1]. Возможно также, что для р.с.ф.-самоподобных множеств не существует невырожденных неподвижных точек — см. пример в [HNW].

Скаляр ρ в формуле (2.5) будет играть основную роль при изучении аналитических свойств фрактала F . В терминах электрических цепей замена цепи $(F_0, a_0(\beta))$ на цепь $(F_1, a_1(\beta))$ приводит (с точки зрения наблюдателя, которому доступны лишь узлы в F_0) к цепи, эквивалентной $(F_0, a_0(\rho^{-1}\beta))$. Поскольку

$$\text{Сопротивление} = 1/\text{Проводимость},$$

естественно назвать ρ *масштабным фактором сопротивления*.

Тесная связь между электрическими цепями и симметричными случайными блужданиями делает возможным вероятностное истолкование величины ρ . Пусть G — конечное множество, а $a = (a_{xy})$, $x, y \in G$, — матрица проводимости, удовлетворяющая условиям

$$a_{xx} = 0, \quad a_{xy} = a_{yx} \geq 0,$$

и пусть Y_n , $n \geq 0$, — марковская цепь с вероятностями перехода $p_{xy} = a_{xy}/\mu_x$, где $\mu_x = \sum_y a_{xy}$. Пусть $S_x = \min\{n \geq 0 : Y_n = x\}$ и $R(x, y)$ — эффективное сопротивление цепи (G, a) между точками x и y . Тогда [C]

$$E^x S_y + E^y S_x = R(x, y) \sum_x \mu_x. \quad (2.6)$$

Поскольку последний член может быть истолкован как «масса» цепи, мы получаем неформальное соотношение

$$\text{Время} = \text{Сопротивление} \times \text{Масса}.$$

Основываясь на этом, легко доказать, что для случайного блуждания X^n на $(F_n, a_n(\beta))$

$$ET_m = M\rho.$$

Мы назовем M (количество 1-клеток в F_1) *масштабным массовым фактором* фрактала F , $\tau = M\rho$ — его *масштабным временным фактором* и λ (коэффициент сжатия в преобразованиях ψ_i) —

масштабным фактором длины. Отсюда мы извлекаем еще два показателя, связанных с F :

$$d_w(F) = \frac{\log \tau}{\log \lambda},$$

$$d_s(F) = -\frac{2 \log M}{\log \tau}.$$

Следуя физической литературе, мы назовем их *размерностью блуждания* и *спектральной размерностью* фрактала F . (Причины таких наименований станут ясны позднее.)

Большинство аналитических свойств фрактала F могут быть сформулированы в терминах этих размерностей и размерности Хаусдорфа $d_f(F)$. Разумеется, среди этих величин независимы лишь две, поскольку

$$d_w(F) = \frac{2d_f(F)}{d_s(F)}.$$

Замечания. 1. Для рассматриваемых здесь фракталов, как легко убедиться, $\rho \geq 1$ или, эквивалентно, $\tau \geq M$; поэтому $d_s(F) \leq 2$. Кроме того, $d_s(F) \leq d_f(F)$, откуда $d_w(F) \geq 2$.

2. Хотя единичный куб $C \subset \mathbf{R}^d$ не попадает в множество рассматриваемых здесь фракталов, для него можно определить эти размерности. Имеем

$$d_f(C) = d, d_w(C) = 2, d_s(C) = d.$$

3. ПРЕДЕЛЬНЫЕ ПРОЦЕССЫ И РЕГУЛЯРНОСТЬ

Зафиксируем теперь какую-нибудь неподвижную точку α с масштабным фактором сопротивления ρ и будем опускать α в обозначениях типа Δ_α^n . Мы масштабируем α так, чтобы $\mu_n(\alpha, F) = M^n$.

Мы хотим получить предел (подходящим образом масштабированного) оператора Δ^n . Один из методов — вероятностный. Если X^n есть случайное блуждание на F_n с генератором Δ^n , то, используя прореживание, получаем случайные блуждания F_{n-1} , F_{n-2} и т. д. с генераторами Δ^{n-1} , Δ^{n-2} и т. д. Переходя к проективному пределу, мы получаем последовательность блужданий X^n , $n \geq 0$, связанных этим прореживанием. Если масштабировать время, рассматривая процессы

$$Y_t^n = X_{[t^n t]}^n, \quad t \geq 0,$$

то каждый из этих процессов пересечет F_0 за среднее время 1, и нетрудно видеть, что в действительности $Y_t^n \rightarrow Y_t$ почти наверное и что Y есть непрерывный F -значный процесс. К сожалению, установление марковости процесса Y весьма технично и утомительно. Имеются, стало быть, известные преимущества в использовании более аналитического подхода, намеченного в [F2].

Для $f \in B(F_m)$ и $n \geq m$ обозначим через $L_{n,m}f \in B(F_n)$ такую единственную функцию, что

$$L_{n,m}f(x) = f(x), \quad x \in F_m, \quad (3.1)$$

$$\Delta^n L_{n,m}f(x) = 0, \quad x \in F_n - F_m.$$

Назовем функцию f , для которой $\Delta^n f(x) = 0$ при $x \in F_n - F_m$, m -гармонической. Легко проверяется, что $L_{m,m-1}f$ есть функция g , доставляющая минимум в формуле (2.3); следовательно,

$$\mathcal{E}^m(L_{m,m-1}f, L_{m,m-1}f) = \rho^{-1} \mathcal{E}^{m-1}(f, f). \quad (3.2)$$

Поэтому, поскольку $\mathcal{E}^m(f, f - L_{m,m-1}(f|_{F_{m-1}})) = 0$, мы имеем

$$\rho^m \mathcal{E}^m(f, f) \geq \rho^{m-1} \mathcal{E}^{m-1}(f|_{F_{m-1}}, f|_{F_{m-1}}). \quad (3.3)$$

Далее мы будем избегать обозначений типа $f|_{F_{m-1}}$, распространяя $\mathcal{E}^{m-1}$ и т. п. на функции на F . Неравенство (3.3) приводит к простому определению предельной формы Дирихле на F : для $f \in C(F)$ положим

$$\mathcal{E}(f, f) = \lim_n \rho^n \mathcal{E}^n(f, f), \quad \mathcal{D} = \{f \in C(F) : \mathcal{E}(f, f) < \infty\}.$$

Теперь мы рассмотрим свойства регулярности формы Дирихле $(\mathcal{E}, \mathcal{D})$.

Если $f \in B(F_0)$, то, поскольку $\alpha_i > 0$ для всех i , имеем

$$\sup_{x, y \in F_0} |f(x) - f(y)|^2 = \text{Osc}(f, F_0)^2 \leq c_0 \mathcal{E}^0(f, f).$$

Так как любая пара точек из F_1 может быть соединена цепочкой, состоящей из не более чем M 1-клеток, то для $f \in B(F_1)$

$$\text{Osc}(f, F_1)^2 \leq c_0 M \mathcal{E}^1(f, f). \quad (3.4)$$

Стандартное рассуждение, использующее связывание x, y подходящей башней из r -клеток с $0 \leq r \leq n$, приводит к следующему неравенству Соболева:

$$\text{Osc}(f, F_n)^2 \leq c_1 \rho^n \mathcal{E}^n(f, f), \quad f \in B(F_n). \quad (3.5)$$

Если x, y лежат в одном m -комплексе, то $|x - y| < c\lambda^{-m}$ и цепочка состоит лишь из r -клеток с $m \leq r \leq n$. Отсюда следует, что

$$|f(x) - f(y)|^2 \leq c_2 |x - y|^{d_w - d_f} \rho^n \mathcal{E}^n(f, f). \quad (3.6)$$

Пусть $\mathcal{H}_{n,m} = \{f \in B(F_n) : \Delta^n f(x) = 0, x \in F_n - F_m\}$ есть множество m -гармонических функций на F_n . Из свойства прореженности следует, что если $f \in \mathcal{H}_{n,m}$, то $f|_{F_{n-1}} \in \mathcal{H}_{n-1,m}$ при $n - 1 \geq m$. (Возможно, проще всего увидеть это с вероятностной точки зрения — $f(X^n)$ является мартингалом до первого момента, когда X^n попадет в F_m , а значит, таков же и процесс $f(\tilde{X}^n)$.) Итак, если $n \geq j \geq m$, то $L_{n,j} L_{j,m} f = L_{n,m} f$. Если $g \in B(F_m)$, то для $x \in F_\infty = \bigcup_{n=0}^\infty F_n$ положим

$$L_m g(x) = L_{n,m} g(x) \text{ при } x \in F_n.$$

Поскольку $\rho^n \mathcal{E}^n(L_{n,m} g, L_{n,m} g) = \rho^m \mathcal{E}^m(g, g)$, то из формулы (3.6) следует, что функция $L_m g$ равномерно непрерывна на F_∞ и может быть продолжена до функции (также обозначаемой $L_m g$) из $C(F)$, удовлетворяющей равенству $\mathcal{E}(L_m g, L_m g) = \rho^m \mathcal{E}^m(g, g)$.

Отсюда получается, что

$$\bigcup_{m=0}^\infty \{L_m g, g \in B(F_m)\} \subseteq \mathcal{D}$$

и множество $\mathcal{D}$ плотно в $C(F)$. А это, вместе с тем фактом, что $\mathcal{D} \subset C(F)$, доказывает регулярность формы Дирихле $(\mathcal{E}, \mathcal{D})$.

Пусть μ есть слабый предел мер $M^{-n} \mu_n$, так что $\mu(F) = 1$ и в смысле меры μ каждый n -комплекс имеет массу M^{-n} . (Мера μ кратна хаусдорфовой x^{d_f} -мере на F , а также является образом при отображении ϕ равномерной меры произведения на абстрактном пространстве последовательностей I^N .) Поскольку

$$\rho^n \mathcal{E}^n(f, g) = - \sum_{x \in F_n} M^{-n} \mu_n(x) \tau^n \Delta^n f(x) g(x),$$

определяя Δ_F как самосопряженный оператор на $L^2(F, \mu)$, ассоциированный с $(\mathcal{E}, \mathcal{D})$, формулой $\mathcal{E}(f, g) = -(\Delta_F f, g)$, получаем, что Δ_F аппроксимируется последовательностью $\tau^n \Delta^n$.

Подытожим свойства $\mathcal{E}$, Δ_F и ассоциированного диффузионного процесса X в следующей теореме.

Теорема 3.1. (а) Пара $(\mathcal{E}, \mathcal{D})$ является регулярной локальной формой Дирихле на $L^2(F, \mu)$.

(b) Для $f \in \mathcal{D}$, $x, y \in F$

$$|f(x) - f(y)|^2 \leq c|x - y|^{d_w - d_f} \mathcal{E}(f, f). \quad (3.7)$$

(c) Для $f \in \mathcal{D}$

$$\rho^n \mathcal{E}^n(f|_{F_n}, f|_{F_n}) \uparrow \mathcal{E}(f, f).$$

(d) Оператор Δ_F на $L^2(F, \mu)$ самосопряжен и $\mathcal{D}(\Delta_F) \subset \mathcal{D}$, причем

$$\mathcal{E}(f, g) = -(\Delta_F f, g),$$

$$\Delta_F f(x) = \lim_{n \rightarrow \infty} \tau^n \Delta^n f(x), \quad x \in F_\infty.$$

(e) Если $(X_t, t \geq 0, P_x, x \in F)$ есть диффузия (непрерывный строго марковский процесс) с полугруппой $P_t = e^{t\Delta_F}$, то X является слабым пределом процессов $Y_t^n = X_{[\tau^n t]}$. Процесс X симметричен относительно меры μ .

Стандартный лапласиан на $\mathbf{R}^d$ может быть охарактеризован (с точностью до умножения на константу) как единственный оператор второго порядка, инвариантный относительно всех изометрий пространства $\mathbf{R}^d$. Естественно поинтересоваться аналогичной характеристикой для Δ_F . В общем случае ответ неизвестен — это связано с упоминавшейся ранее проблемой единственности неподвижных точек. Однако для салфетки Серпинского (когда неподвижная точка очевидным образом одна) в [BP] доказано, что, с точностью до детерминистской замены времени, X есть единственный процесс, который локально инвариантен относительно локальных изометрий фрактала F . Отсюда вытекает соответствующая единственность для оператора Δ_F .

4. АНАЛИТИЧЕСКИЕ СВОЙСТВА ОПЕРАТОРА Δ_F

Мы начнем с определения плотностей потенциалов u_α , $\alpha > 0$, $x, y \in F$. Формально они являются решениями уравнения

$$(\alpha - \Delta_F)u_\alpha(\cdot, y) = \delta_y(\cdot).$$

Они могут быть определены с помощью процесса X как плотности относительно меры μ α -резольвент

$$U_\alpha f(x) = E^x \int_0^\infty e^{-\alpha t} f(X_t) dt. \quad (4.1)$$

Естественно, требуется некоторая работа, чтобы показать, что $U_\alpha(\cdot, y) \ll \mu$. Наиболее быстрый способ — использовать форму Дирихле $\mathcal{E}_\alpha = \alpha(f, g) + \mathcal{E}(f, g)$. Из неравенства (3.7) следует, что

$$\|f\|_\infty^2 \leq c_\alpha \mathcal{E}_\alpha, \quad f \in \mathcal{D};$$

таким образом, отображение $T_y : \mathcal{D} \rightarrow \mathbf{R}$, определенное формулой $T_y(g) = g(y)$, является ограниченным. Отсюда следует (как в [F1, 3.3.3]), что $\mathcal{E}_\alpha$ допускает такое воспроизводящее ядро $u_\alpha(x, y)$, что

$$\mathcal{E}_\alpha(u_\alpha(\cdot, y), g) = g(y), \quad g \in \mathcal{D}. \quad (4.2)$$

Теорема 4.1. (a) $u_\alpha(x, y) = u_\alpha(y, x)$, $x, y \in F$.

(b) u_α есть непрерывная функция класса Гёльдера $d_w - d_f$ на $F \times F$.

(c) $U_\alpha f(x) = \int_F u_\alpha(x, y) f(y) \mu(dy)$, $f \in C(F)$.

Из этого результата вытекают разнообразные вероятностные свойства процесса X . В силу ограниченности функции $u_\alpha(\cdot, y)$, записав $T_y = \inf\{t > 0 : X_t = y\}$, получаем

$$u_\alpha(x, y) = E^x e^{-\alpha T_y} u_\alpha(y, y),$$

так что $P^y(T_y = 0) = 1$ и y регулярна для $\{y\}$. Из гёльдеровской непрерывности функции u_α вытекает, что X имеет локальное время, непрерывное по совокупности пространственной и временной переменных (плотность времени пребывания) $(L_t^x, x \in F, t \geq 0)$, откуда следует, что процесс X «заполняет пространство»:

$$P^x(\{X_t, 0 \leq t \leq N\} = F \text{ для некоторого } N < \infty) = 1.$$

Некоторые свойства процесса X могут быть сформулированы проще, если вместо него рассмотреть процесс на неограниченной версии фрактала F . Возьмем $a_1 = 0$, чтобы подобие ψ_1 задавалось формулой $\psi_1(x) = \lambda^{-1}x$. Затем положим

$$\tilde{F} = \bigcup_{n=0}^{\infty} \lambda^n F.$$

Построение формы Дирихле $\mathcal{E}$, лапласиана Δ и диффузии X на $\tilde{F}$ происходит непосредственно путем использования покрытий. При этом процесс $\tilde{X}$ удовлетворяет скейлинговому соотношению

$$(X_t, t \geq 0, P^0) \stackrel{(d)}{=} (\lambda^{-1} X_{\tau t}, t \geq 0, P^0), \quad (4.3)$$

которое полезно сравнить с соотношением $\theta^{-1}B_{\theta^2 t} \stackrel{(d)}{=} B_t$ для стандартного броуновского движения в $\mathbf{R}^d$. Поскольку $|X_1 - X_0| = O(1)$, из (4.3) следует, что $|X_t - X_0| = O(t^{1/d_w})$ при $t > 0$. (Напомним, что $1/d_w = \log \lambda / \log \tau$.) Точнее, мы получаем

Предложение 4.2. (а) *Процесс X непрерывен по Гёльдеру класса $d_w^{-1} - \epsilon$ для любого $\epsilon > 0$, но не является непрерывным класса d_w^{-1} .*
 (б) *Существуют такие константы c_1, c_2 , что при $t \geq 0$*

$$c_1 t^{2/d_w} \leq E^x |X_t - x|^2 \leq c_2 t^{2/d_w}.$$

Более тонкие результаты, дающие точный модуль непрерывности для траекторий процесса X , содержатся в [BP, BV3, Kum]. Этот результат дает интуитивное объяснение «размерности» d_w : она управляет пространственно-временным скейлингом диффузии X таким же образом, как число 2 управляет пространственно-временным скейлингом обычного броуновского движения.

Следствие 4.3. (а) *Процесс X не является семимартингалом.*
 (б) *Если $f \in C^1(\mathbf{R}^d)$ и $f|_F \in \mathcal{D}(\Delta_F)$, то $f|_F$ — константа.*

Пункт (а) следует из предложения 4.2 и того, что траектории семимартингалов имеют конечную квадратичную вариацию. Что касается п. (б), то для $f \in \mathcal{D}(\Delta)$ процесс

$$M_t^f = f(X_t) - f(X_0) - \int_0^t \Delta f(X_s) ds$$

является мартингалом, а потому $f(X_t)$ — семимартингал, что невозможно, если $f|_F$ не константа.

Отсюда видно, что «гладкие функции» в $\mathcal{D}(\Delta_F)$ не имеют ничего общего с обычными гладкими функциями на $\mathbf{R}^d$. На настоящий момент не слишком много известно о структуре и форме функций из $\mathcal{D}(\Delta_F)$ или $\mathcal{D}$. Из теоремы 3.1 мы видим, что они непрерывны класса $d_w - d_f$ по Гёльдеру всюду; они также непрерывны класса $\frac{1}{2}d_w$ почти всюду в смысле меры μ на F . (Отметим, что поскольку $\frac{1}{2}d_w > 1$, последнее утверждение не может быть верно всюду.) Несколько более точный результат получен в [K2]: если ν_f — локальная мера, ассоциированная с $\mathcal{E}(f, f)$, так что

$$\mathcal{E}(f, f) = \int_F \nu_f(dx), \quad f \in \mathcal{D},$$

то существует такая мера ν на F , что $\mu \perp \nu$ и $\nu_f \ll \nu$ для всех $f \in \mathcal{D}$. Неформально мы можем сказать, что для $f \in \mathcal{D}$ получаем $|\nabla f| = 0$ μ -почти всюду, где $|||\nabla f|||_\infty = \infty$.

Так как функции $u_\alpha(x, y)$ непрерывны, а операторы U_α компакты, мы можем записать

$$u_\alpha(x, y) = \sum_{i=1}^{\infty} (\alpha + \lambda_i)^{-1} \phi_i(x) \phi_i(y),$$

где $0 \leq \lambda_1 \leq \lambda_2 \leq \dots$ суть собственные значения оператора $-\Delta_F$, а ϕ_i — нормированные собственные функции. Отметим, что λ_i и ϕ_i удовлетворяют соотношению

$$\mathcal{E}(\phi_i, f) = \lambda_i(f, g) \text{ для всех } g \in \mathcal{D}.$$

Теперь мы изучим асимптотическое поведение собственных значений. Положим

$$\mathcal{D}_i = \{f \in \mathcal{D} : f = 0 \text{ на } F_i\}, \quad i = 0, 1,$$

и обозначим через Δ_i соответствующий лапласиан (который отвечает граничному условию Дирихле на F_i). Заметим, что если $(\mathcal{D}_i, \mathcal{F}_i)$, $i = 1, 2$, — формы Дирихле с $\mathcal{D}_1 \subset \mathcal{D}_2$ и $\mathcal{E}_1 = \mathcal{E}_2$ на $\mathcal{D}_1$, то принцип минимакса (см. [Ch]) показывает, что для λ_n^i , $n \geq 1$, являющихся собственными значениями оператора $-\Delta_i$, мы имеем $\lambda_n^1 \geq \lambda_n^2$. Следовательно, если $N_i(x) = \#\{\lambda_n^i : \lambda_n^i \leq x\}$, то $N_1(x) \leq N_2(x)$. Назовем функции N_i *интегрированными плотностями состояний* (для $(\mathcal{D}_i, \mathcal{F}_i)$).

Пусть N , N_i — интегрированные плотности состояний для $(\mathcal{D}, \mathcal{F})$, $(\mathcal{D}_i, \mathcal{F}_i)$ соответственно. Если $f, g \in B(F_n)$, то

$$\mathcal{E}^n(f, g) = \sum_{i=1}^M \mathcal{E}^{n-1}(f \circ \psi_i, g \circ \psi_i);$$

значит, для $f, g \in \mathcal{D}$ имеем

$$\mathcal{E}(f, g) = \sum_{i=1}^M \rho \mathcal{E}(f \circ \psi_i, g \circ \psi_i). \quad (4.4)$$

Так как $\mathcal{D}_1 \subset \mathcal{D}_2$, то $N_1(x) \leq N_0(x)$. Однако если g есть собственная функция оператора $-\Delta_0$ и $\Delta_0 g + \lambda g = 0$, то для $1 \leq i \leq M$ определим $h_i \in C(F)$ формулой

$$h_i(x) = \begin{cases} g \circ \psi_i^{-1} & \text{для } x \in \psi_i(F), \\ 0 & \text{для } x \notin \psi_i(F). \end{cases}$$

Тогда для $v \in D_1$ из формулы (4.4) и из того, что $h_i \circ \psi_i = g$, получаем

$$\mathcal{E}(h_i, v) = \rho \mathcal{E}(g, v \circ \psi_i) = \rho \lambda(g, v \circ \psi_i) = M \rho \lambda(h_i, v).$$

Поэтому $\tau \lambda$ есть собственное значение оператора Δ_1 с кратностью M и $N_1(x) = M N_0(x/\tau)$.

Сходную операцию мы сделаем и для неймановских собственных значений на F , только на этот раз она будет включать сечения в точках из F_1 , заменяя F на дизъюнктное объединение пространств $\psi_i(F)$, $1 \leq i \leq M$. Обозначая через D' ассоциированное пространство Дирихле, а через N' интегрированную плотность состояний, мы получаем (осуществляя подходящее вложение) $\mathcal{D} \subset \mathcal{D}'$. То же рассуждение, что и в случае Дирихле, приводит к соотношению

$$\begin{aligned} N_0(y) = M^{-1} N_1(\tau y) &\leq M^{-1} N_0(\tau y) \leq M^{-1} N(\tau y) \\ &\leq M^{-1} N'(Ty) = N(y) \end{aligned}$$

(здесь $y = x/\tau$). Поэтому существуют такие константы c_1 и c_2 , что

$$c_1 x^{\frac{1}{2}d_s} \leq N_0(x) \leq N(x) \leq c_2 x^{\frac{1}{2}d_s} \text{ для } x \geq x_0. \quad (4.5)$$

Теперь естественно спросить о более тонких деталях асимптотик $N(\cdot)$. Однако Фукусима и Сима [FS] доказали, что для салфетки Серпинского

$$\liminf_{x \rightarrow \infty} x^{-\frac{1}{2}d_s} N(x) < \limsup_{x \rightarrow \infty} x^{-\frac{1}{2}d_s} N(x);$$

таким образом, невозможно надеяться на асимптотическое разложение в стиле того, которое имеется для областей в $\mathbb{R}^n$.

Замечания. 1. Работы [S] и [FS] дают весьма детальное описание спектра оператора Δ для салфетки Серпинского.

2. Уравнение (4.5) объясняет происхождение термина *спектральная размерность* для числа d_s .

5. ГРАНИЦЫ ЯДРА ТЕПЛОПРОВОДНОСТИ

До сих пор мы уделяли мало внимания геометрии фрактала F . Однако когда мы рассматривали форму ядра теплопроводности,

метрика $|x - y|$ оказалась неадекватной: лучше рассматривать внутреннюю метрику $d(x, y)$, определяемую кратчайшим путем от x до y по F . Пусть

$$H_n = \Psi^n(H_0)$$

— замкнутое связное множество, получаемое в результате n итераций отображения Ψ , определенного в (1.2), и пусть $d_\tau(x, y)$ — длина кратчайшего пути в H_n , соединяющего x и y .

Для фракталов типа салфетки Серпинского имеем $|x - y| \leq d_n(x, y) \leq c|x - y|$ для всех n , но для более общих фракталов, предельное множество которых не содержит отрезков прямых, это неверно. В общем случае существует такое число $b \geq \lambda$, что точки из F_0 могут быть соединены цепочкой из $O(b^n)$ n -клеток; таким образом, положив

$$d_c = \log b / \log \lambda,$$

получаем

$$d_n(x, y) \approx |x - y|^{d_c} (b/\lambda)^n.$$

Переходя к пределу по подпоследовательностям, получаем метрику на F , для которой

$$d(x, y) \approx |x - y|^{d_c}.$$

Эта метрика называется в физической литературе *химическим расстоянием*, а показатель d_c известен как *химическая экспонента*. (В противоположность тому, что могли бы ожидать математики, физики обычно не измеряют длину полимеров с помощью этой метрики.)

Полезно переопределить фрактальную размерность и размерность блуждания в терминах новой метрики, полагая

$$d_f^l = \log M / \log b = d_f / d_c,$$

$$d_w^l = \log \tau / \log b = d_w / d_c.$$

Отметим, что спектральная размерность удовлетворяет соотношениям $\frac{1}{2}d_s = d_f / d_w = d_f^l / d_w^l$ и не изменяется при замене метрики.

Рассмотрим теперь случайный процесс X на неограниченном фрактале $\tilde{F}$, и пусть $p(t, x, y)$ — соответствующие плотности вероятностей перехода. (Существование таких функций следует из μ -симметричности процесса X и из существования плотности ре-

зольвент U_α .) Плотность p является решением уравнения теплопроводности на $\tilde{F}$:

$$\begin{aligned}\frac{\partial}{\partial t} p(t, x, y) &= \Delta_x p(t, x, y) \\ p(0, x, y) &= \delta_y(x)\end{aligned}$$

Верхние границы на $p(t, x, y)$ могут быть получены с помощью масштабирования формы $\mathcal{E}$ и общей теории из [CKS]. Приводимое здесь рассуждение принадлежит Фицсиммонсу и Хэмбли. Поскольку [F1, p. 21]

$$\mathcal{E}(f, f) = \sup_{\alpha} (\alpha f, f - U_{\alpha} f) \geq (f, f - U_1 f),$$

имеем

$$\|f\|_2^2 \leq \|u_1\|_{\infty} \|f\|_1^2 + \mathcal{E}(f, f).$$

Заменяя f на $\tilde{f}(x) = f(\lambda^n x)$, используя масштабирование и оптимизируя по n , получаем *неравенство Нэша*

$$\|f\|_2^{2+4/d_s} \leq c \mathcal{E}(f, f) \|f\|_1^{4/d_s}. \quad (5.1)$$

Отсюда следует [CKS], что

$$p(t, x, x) \leq c_1 t^{\frac{1}{2}d_s}. \quad (5.2)$$

Нижние границы на диагонали также весьма просто получить — они вытекают из оценки

$$\limsup_{\theta \uparrow \infty} \sup_{x, t} \left(P^x |X_t - x| > \theta t^{1/d_w} \right) = 0.$$

Отсюда следует, что для подходящего $r > 0$, обозначив $z = r t^{1/d_w}$, имеем

$$\frac{1}{2} < \int_{B(x, z)} p(t, x, y) \mu(dy) \leq \mu(B(x, z)) p(t, x, x)^{\frac{1}{2}} \sup_y p(t, y, y)^{\frac{1}{2}}.$$

Используя оценку $\mu(B(x, z)) \geq c z^{d_f}$, получаем нижнюю границу

$$p(t, x, x) \geq c t^{\frac{1}{2}d_s}, \quad 0 < t < \infty. \quad (5.3)$$

Замечания. 1. Связь между асимптотиками ядра теплопроводности и количеством собственных значений означает, что можно ожидать границ такого вида для малых t из (4.5). Из скейлингового соотношения (4.3) будет следовать тогда, что эти оценки верны при всех t .

2. Отметим, что показатели в (5.2) и (5.3) не связаны с изопериметрической размерностью, которая равна нулю у конечно разветвленных фракталов. Таким образом, ситуация здесь существенно отличается от изученной в работе [V], в которой из изопериметрического неравенства вытекало неравенство Нэша с правильным показателем. Следует также сравнить границы для ядра теплопроводности, полученные в [O] и [BB3].

Оценки вне диагонали оказываются более тонкими. Имеется общая «машина» для получения внедиагональных оценок сверху — это «метод Дэвиса» (см. [CKS]). Он существенно опирается на рассмотрение полугруппы $e^{-\psi} P_t e^{\psi}$, использует глобальную верхнюю границу, полученную из верхней границы на диагонали, а также подходящим образом варьирует ψ . К сожалению, оказывается, что этот метод не работает в случае фракталов. Основная причина состоит в том, что, с одной стороны, требуется, чтобы $\psi \in \mathcal{D}$, но в то же время оценки сверху в более или менее явной форме используют величину $\|\nabla \psi\|_{\infty}$, которая, как показано в [K2], равна $+\infty$ для гнездовых фракталов.

Приемлемая техника для получения верхних и нижних границ называется *техникой связывания*. Для нижних границ она совпадает с классической. Предположим, что

$$p(t, x, y) \geq ct^{-d_s/2} \text{ при } t > 0, |x - y| < c_2 t^{1/d_w}. \quad (5.4)$$

Положив $D = |x - y|$, мы получаем, что x и y могут быть соединены цепочкой из $n = c(D/\epsilon)^{d_c}$ шаров евклидова радиуса ϵ ; обозначим их $B_i = B(x_i, \epsilon)$, $1 \leq i \leq n$ с $x_0 = x$, $x_n = y$. Возьмем $\epsilon = c_2(t/n)^{1/d_w}$; тогда

$$\begin{aligned} p(t, x, y) &\geq \int_{B_1} \cdots \int_{B_{n-1}} p(t/n, x_0, x_1) \cdots p(t/n, x_{n-1}, y) \mu(dx_1) \cdots \mu(dx_{n-1}) \\ &\geq \prod_{i=1}^{n-1} \mu(B_i) c_1^n (t/n)^{-nd_s/2} \geq (c\epsilon^{d_f})^{n-1} c_1^n (t/n)^{-d_s/2} (\epsilon/c_2)^{-d_f(n-1)} \\ &\geq t^{-d_s/2} c_3^n = t^{-d_s/2} \exp(-c_4^n). \end{aligned}$$

Подставляя выражение для n , получаем

$$p(t, x, y) \geq t^{-d_s/2} \exp(-c(|x - y|^{d_w}/t)^{d_c/(d_w - d_c)}); \quad (5.5)$$

экспоненту в (5.5) можно записать в виде

$$c(d(x, y)^{d_w}/t)^{1/(d_w - 1)}.$$

Итак, связывание превращает «полуфабрикатную» оценку (5.4) в глобальную нижнюю границу, которая с точностью до константы является наилучшей.

Теперь остается доказать оценку (5.4). Но она выполняется при $x = y$, и гёльдерова непрерывность функций из $\mathcal{D}$ позволяет распространить ее на некоторый шар.

Похожая техника годится и для получения верхних оценок, но рассуждения здесь являются более вероятностными. Для $x \in \tilde{F}$, $\epsilon > 0$ положим

$$S(x, \epsilon) = \inf\{t \geq 0 : |X_t - x| < \epsilon\}.$$

Пространственно-временной скейлинг процесса X означает, что величина $\epsilon^{-1/d_w} S(x, \epsilon)$ имеет порядок $O(1)$; используя это соображение и элементарное неравенство

$$P(\xi \leq t) \leq (\text{Var}(\xi) + 2tE\xi)/E\xi^2,$$

которое выполняется для любой неотрицательной случайной величины ξ , мы получаем

$$P(S(x, \epsilon) < t) \leq p + a\epsilon^{-1/d_w} t, \quad t \geq 0. \quad (5.6)$$

Следующая лемма [BB1, Lemma 1.1] является ключевой для применения техники связывания.

Лемма 5.1. Пусть $T, S_1, \dots, S_n$ — неотрицательные случайные величины, удовлетворяющие следующим условиям при некоторых $p < 1$, $a > 0$:

$$(a) \quad T \geq \sum_1^n S_i;$$

$$(b) \quad P(S_i < t | \sigma(S_1, \dots, S_{i-1})) \leq p + at.$$

$$\text{Тогда } P(T \leq t) \leq \exp(2(ant/p)^{1/2} - n \log p^{-1}).$$

Зафиксируем теперь x, y, t и положим $A = \{z \in F : |z - y| < |z - x|\}$. Если обозначить $S_0 = 0$, $S_{i+1} = S(X_{S_i}, \epsilon)$, то, так как

кратчайший путь от x до A пересекает $n = c(|x - y|/\epsilon)^{d_c}$ шаров радиуса ϵ , мы получаем, обозначив $T_A = \inf\{t \geq 0 : X_t \in A\}$,

$$\begin{aligned} \log P^x(T_A \leq t) &\leq (2a\epsilon^{-1/d_w} n t/p)^{1/2} - n \log p^{-1} \\ &\leq c(|x - y|^{d_w}/t)^{d_c/(d_w - d_c)}. \end{aligned}$$

Комбинируя это с глобальной верхней границей (5.1), получаем такой результат:

Теорема 5.2. Пусть $\phi(z, t) = (z^{d_w}/t)^{1/(d_w - 1)}$. Тогда плотность ядра теплопроводности $p(t, x, y)$ на $\tilde{F}$ при $0 < t < \infty$, $x, y \in \tilde{F}$, удовлетворяет неравенствам

$$\begin{aligned} c_1 t^{-d_s/2} \exp(-c_2 \phi(d(x, y), t)) &\leq p(t, x, y) \\ &\leq c_3 t^{-d_s/2} \exp(-c_4 \phi(d(x, y), t)). \end{aligned}$$

Замечание. Оценки такого вида непривычны для классической ситуации, но тем не менее они возникают в менее необычных предположениях, чем сделанные здесь. Например, пусть $\tilde{G}$ есть неограниченная двумерная салфетка Серпинского (у которой $d_c = 1$, $d_w = \log 5 / \log 2$ и $d_f = \log 3 / \log 2$). Вложим $\tilde{G}$ в $\mathbf{R}^3$, полагая

$$M' = \bigcup_{x \in \tilde{G}} B(x, 1)$$

и $M = \partial M'$. Тогда M является многообразием с богатой скейлинговой структурой, которая имитирует структуру салфетки Серпинского. Если $p(t, x, y)$ — ядро теплопроводности на M , то при малых t и $|x - y| = O(1)$ его поведение подобно поведению ядра теплопроводности на $\mathbf{R}^2$. Однако следует ожидать, что оценки теоремы 5.2 имеют место в области $1 \leq t \leq |x - y|$. (При $|x - y| > t$ мы оказываемся в классической ситуации большого уклонения.)

ЛИТЕРАТУРА

- [B1] Barlow M. T. Random walks, electrical resistance, and nested fractals, in: Asympt. Probl. in Prob. Th., K. D. Elworthy, N. Ikeda (eds.), Pitman Res. Notes Math. Ser. vol. 283, Longman Sci. Tech., Harlow, 1993.
- [BB1] Barlow M. T., Bass R. F. The construction of Brownian motion on the Sierpinsky carpet. Ann. Inst. H. Poincaré. **25** (1989), no. 3, 225–257.

- [BB2] Barlow M. T., Bass R. F. On the resistance of the Sierpinsky carpet. *Proc. Roy. Soc. Lond. A.*, **431** (1990), 345–360.
- [BB3] Barlow M. T., Bass R. F. Transition densities for Brownian motion on the Sierpinsky carpet. *Probab. Th. Rel. Fields*, **91** (1992), 307–330.
- [BP] Barlow M. T., Perkins E. A. Brownian motion on the Sierpinsky gasket, *Probab. Th. Rel. Fields*, **79** (1988), 543–623.
- [CKS] Carlen E. A., Kusuoka S., Stroock D. W. Upper bounds for symmetric Markov transition functions. *Ann. Inst. H. Poincaré*, **23** (1987), no. 2, suppl., 245–287.
- [C] Chandra A. K., Raghaven P., Ruzzo W. L., Smolensky R., Tiwary P. The electrical resistance of a graph captures its commute and cover times, in: *Proceedings of the 21st ACM Symposium on theory of computing*, 1989.
- [Ch] Chavel I. *Eigenvalues in Riemannian Geometry*. Academic Press, 1984.
- [F1] Fukushima M. *Dirichlet Forms and Markov Processes*, North Holland, 1980.
- [F2] Fukushima M. Dirichlet forms, diffusion processes, and spectral dimensions for nested fractals, in: *Ideas and Methods in Stochastic Analysis, Stochastics and Applications*. S. Albeverio et al. (eds.), Cambridge Univ. Press., Cambridge 1992, 151–161.
- [FS] Fukushima M., Shima T. On a spectral analysis for the Sierpinsky gasket. *Potential Anal.*, **1** (1992), no. 1, 1–35.
- [G] Goldstein S. Random walks and diffusion on fractals, in: H. Kesten (ed.). *Percolation Theory and Ergodic Theory of Infinite Particle Systems* (IMA Math. Appl., vol. 8) Springer-Verlag, New York, 1987, 121–129.
- [HHW] Hattori K., Hattori T., Watanabe H. Gaussian field theories and the spectral dimensions. *Prog. Th. Phys. Supp.*, **92** (1987), 108–143.
- [Kig1] Kigami J. A harmonic calculus on the Sierpinsky space. *Japan J. Appl. Math.*, **6** (1989), 259–290.
- [Kig2] Kigami J. A harmonic calculus on p.c.f. self-similar sets. *Trans. Amer. Math. Soc.*, **335** (1993), No. 2, 721–755.
- [Kum] Kumagai T. Estimates of the transition densities for Brownian motion on nested fractals, Preprint, 1991.
- [K1] Kusuoka S. A diffusion process on a fractal, in: *Symposium on Probabilistic Methods in Mathematical Physics*, K. Ito, N. Ikeda (eds.), Taniguchi, Katata. Academic Press, Amsterdam, 1987, 251–274.
- [K2] Kusuoka S. Dirichlet forms on fractals and products of random matrices. *Publ. RIMS Kyoto Univ.*, **25** (1989), 659–680.
- [KZ] Kusuoka S., Zhou X. Y. Dirichlet forms on fractals: Poincaré constant and resistance. *Probab. Th. Rel. Fields*, **93** (1992), 169–186.
- [L] Lindstrøm T., Brownian motion on nested fractals. *Mem. Amer. Math. Soc.*, **83** (1990), No. 420.

- [O] Osada H. Isoperimetric dimension and estimates of heat kernels of pre-Sierpinsky carpets. *Probab. Th. Rel. Fields*, **86** (1990), 469–490.
- [RT] Rammal R., Toulouse G. Random walks on fractal structures and percolation clusters. *J. Phys. Lettres*, **44** (1983), L13–L22.
- [S] Shima T. On eigenvalue problems for the random walk on the Sierpinsky pre-gasket. *Japan J. Appl. Ind. Math.*, **8** (1991), 127–142.
- [V] Varopoulos N. Th. Isoperimetric inequalities and Markov chains. *J. Funct. Anal.*, **63** (1985), 215–239.

M. T. BARLOW

Department of Mathematics
 University of British Columbia
 Vancouver
 British Columbia
 Canada V6T 1Z2
 E-mail: barlow@math.ubc.ca

КОМПАКТНЫЕ МНОГООБРАЗИЯ ДРИНФЕЛЬДА (по работам Ломона, Рапопорта и Штулера)

Анри Карайоль

0. ВВЕДЕНИЕ

0.1. Пусть $f = \sum a_n q^n$ — нормализованная (т. е. с $a_1 = 1$) параболическая модулярная форма веса 2 относительно группы $\Gamma_0(N)$, являющаяся собственным вектором операторов Гекке. Классическая теория Эйхлера—Шимуры (обобщенная Делинем на случай форм веса > 2) позволяет построить по такой форме f двумерное представление σ_ℓ группы Галуа $\text{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ над полем $\bar{\mathbb{Q}}_\ell$. Представление σ_ℓ не разветвлено над точками $p \nmid N\ell$ и удовлетворяет соотношениям

$$\text{tr } \sigma_\ell(\text{Frob}_p) = a_p, \quad \det \sigma_\ell(\text{Frob}_p) = p,$$

где Frob_p обозначает автоморфизм Фробениуса точки p (мы неявно выбираем вложение в $\bar{\mathbb{Q}}_\ell$ поля, порожденного элементами a_p). Известно, что представления σ_ℓ возникают при разложении ℓ -адических когомологий модулярных кривых относительно действия операторов Гекке.

Более общая теория, глубоко разработанная за последние годы (см. [Ko]), пытается сделать то же самое для гомологий многообразий Шимуры, т. е. по некоторым автоморфным представлениям редуктивных групп над числовыми полями построить ℓ -адические представления Галуа (точнее говоря, некоторые мотивы). Эта программа, являющаяся составной частью «философии Ленглендса» (см. [La, Cl]), в настоящий момент полностью завершена только для некоторых типов групп, по существу в тех случаях, когда соответствующие многообразия Шимуры компактны.

0.2. С 1973 г. Дринфельд начал развивать аналогичную теорию для поля функций от одной переменной над конечным полем, т. е. для поля $F = \mathbb{F}_q(X)$ рациональных функций на алгебраической кривой $X/\mathbb{F}_q$. Аналогия между функциональными и числовыми полями хорошо известна в теории чисел, в которой как первый, так и второй тип полей называются глобальными полями. «Философия»

Лейнлендса, так же как и теория полей классов, которую она обобщает, должна применяться ко всем глобальным полям.

0.3. Первое понятие, введенное Дринфельдом, — это «эллиптический модуль ранга d ». Это понятие зависит от выбора точки поля F , обозначаемой ∞ , и использует подкольцо A поля F , состоящее из рациональных функций, регулярных вне ∞ . Эллиптические модули обладают формальными свойствами, аналогичными свойствам эллиптических кривых. Например, снабженные структурой уровня J (где J — идеал в A), они параметризуются многообразием модулей M_J^d , которое является гладким аффинным многообразием относительной размерности $d - 1$ над $A[J^{-1}]$. В частности, при $d = 2$ мы получаем семейство кривых над F , которое является аналогом модулярных кривых над полем $\mathbb{Q}$. Изучая когомологии этих кривых (подходящим образом компактифицированных), Дринфельд развил теорию, параллельную теории Эйхлера—Шимуры, и доказал, что любому *параболическому автоморфному представлению* Π адельной группы $\mathrm{GL}_2(A_F)$, такому, что компонента Π_∞ является специальным представлением, можно сопоставить двумерное ℓ -адическое представление группы $\mathrm{Gal}(\bar{F}/F)$, где $\bar{F}$ обозначает сепарабельное замыкание поля F (эти представления удовлетворяют соотношениям, аналогичным тем, которые определяют представление σ_ℓ (см. выше)). В следующей своей работе [Dr2] Дринфельд построил систему этальных накрытий многообразия M_J^d , с помощью которой (при $d = 2$) он сумел сопоставить представления Галуа автоморфным представлениям Π с каспидальной компонентой Π_∞ .

0.4. Как будет видно из этого обзора, определение эллиптического модуля алгебраически элементарно. В 1976 г. Дринфельд ввел новые, более геометрические объекты, называемые «эллиптическими пучками», которые представляют собой специального вида диаграммы векторных расслоений на кривой X . Он доказал, что понятия эллиптического пучка и эллиптического модуля эквивалентны, а возникающие многообразия модулей изоморфны. В следующем году он модернизировал понятие эллиптического пучка так, чтобы точка ∞ не играла особой роли. Используя полученные объекты, называемые «штуками» или F -пучками, он сумел сопоставить ℓ -адическое представление Галуа любому автоморфному параболическому представлению группы $\mathrm{GL}_2(A_F)$.

0.5. При $d \geq 3$ многообразия эллиптических модулей (или пучков) ранга d имеют относительную размерность $d - 1$ и играют для

группы $GL_d(F)$ роль многообразий Шимур (которые, однако, для группы GL_d над числовым полем при $d \geq 3$ не существуют). Изучение когомологий этих многообразий должно было бы позволить сопоставить d -мерные ℓ -адические представления Галуа автоморфным параболическим представлениям Π группы $GL_d(A_F)$, таким, что представление Π_∞ специально. Однако такое изучение весьма затруднено тем, что рассматриваемые многообразия не являются собственными. В связи с этим возникает множество сложных проблем, как геометрических (вопросы компактификации, формула следа Лефшеца), так и автоморфных (формула следа Артура—Сельберга не доказана для функционального поля). Отметим, однако, что, воспользовавшись некоторыми гипотезами (а именно гипотезой Делиня, дающей простую формулу Лефшеца для соответствия, скрученного большой степенью оператора Фробениуса, а также формулой следа Артура—Сельберга), Ломон [Lau2] вычислил виртуальный модуль $\sum (-1)^i H_c^i(M_J^d \otimes_F \overline{F}, \overline{\mathbb{Q}}_\ell)$ относительно действия группы $\text{Gal}(\overline{F}/F)$ и операторов Гекке. Проблемы, связанные с некомпактностью, для штук оказываются еще сложнее. Даже в случае $d = 2$ они весьма серьезны, хотя и успешно решены.

0.6. Чтобы обойти эти трудности, Делинь сразу после появления первой статьи Дринфельда предложил определить и изучить собственные аналоги модулярных многообразий Дринфельда. Эти новые многообразия должны были быть сопоставлены уже не группе $GL_d(F)$, а мультипликативной группе G центрального тела D над F размерности d^2 , расщепляющегося над точкой ∞ . Именно эта проблема, решенная в статье [L-R-S], является объектом настоящего обзора.

Существенным моментом было нахождение хорошего определения аналога эллиптического модуля. Более жизненным оказалось понятие эллиптического пучка, а не более элементарное понятие эллиптического модуля. Его непосредственным обобщением стало понятие *эллиптического D -пучка*. При этом, если база является спектром совершенного поля, эти объекты могут быть интерпретированы как «векторные расслоения над некоммутативной проективной прямой» — понятие, которое более непосредственно обобщает понятие эллиптического модуля в духе работы Андерсона о « t -мотивах» [An].

0.7. Ломон, Рапопорт и Штулер доказывают, что эллиптические D -пучки, снабженные структурой уровня I (где I — замкнутая конечная подсхема в $X - \{\infty\}$), параметризуются многообразием модулей

$\mathcal{E}ll_{X,D,I}$, которое является собственной гладкой схемой относительной размерности $d-1$ над $X-I-\mathcal{R}-\{\infty\}$ ($\mathcal{R}$ обозначает множество точек ветвления тела D). Затем они разлагают когомологии $H^{d-1}(\mathcal{E}ll_{X,D,I} \times_X \operatorname{Spec} \overline{F}, \overline{Q}_\ell)$ относительно действия операторов Гекке и доказывают, что таким образом автоморфным представлениям Π группы $G(\mathbf{A}_F)$ со специальной компонентой Π_∞ можно сопоставить представления Галуа. При этом возникают именно те представления группы $\operatorname{Gal}(\overline{F}/F)$, которые и должны были возникнуть (в частности d -мерные), при условии, что кратность, с которой представление Π входит в пространство автоморфных форм, равна 1. Гипотетически это верно всегда (это следует из существования соответствия Жак—Ленглендса между автоморфными представлениями групп $G(\mathbf{A}_F)$ и $\operatorname{GL}_d(\mathbf{A}_F)$). К несчастью, полного доказательства этого факта для функционального поля не существует (см. [De-Ka-Vi] и [Ro]). Поэтому с кратностью $m(\Pi)$, входящей в условие основной теоремы работы [L-R-S], связаны многочисленные технические сложности на протяжении всей статьи.

0.8. Изучение когомологий $H^*(\mathcal{E}ll_{X,D,I} \times_X \operatorname{Spec} \overline{F}, \overline{Q}_\ell)$ и их разложения по действию операторов Гекке использует технику, хорошо знакомую специалистам по многообразиям Шимур: описание множества точек многообразия над конечным полем, вычисление орбитальных интегралов, сравнение формул следа Лефшеца и Сельберга и т. д. Кроме того, при доказательстве того, что существенная часть когомологий расположена в размерности $d-1$, используются сильная теорема Лефшеца и теорема Делиня о чистоте монодромической фильтрации. Эта часть доказательства существенно упрощается, если воспользоваться обобщенным соответствием Жак—Ленглендса.

0.9. Наконец, в завершение, в статье [L-R-S] доказана *локальная гипотеза Ленглендса для группы GL_d над полем равной характеристики*: здесь используется глобальная конструкция представлений Галуа (в применении к автоморфному представлению Π кратности 1), а также многочисленные результаты Эннъяра. В семидесятых годах Делинь получил аналогичные результаты для $d=2$, опираясь на теорию Дринфельда.

0.10. В этом обзоре мы ограничимся описанием теории эллиптических модулей, затем эквивалентной ей теории эллиптических пучков, после чего изложим содержание работы [L-R-S]. Мы не будем затрагивать вопросы, связанные с накрытиями пространств M_J^d ,

построенными в [Dr2], и с аналогичными накрытиями пространств $\mathcal{E}ll_{X,D,I}$, построенными в [L-R-S], а также со штуками.

0.11. Обозначения. Начиная с этого момента, X обозначает алгебраическую гладкую проективную и геометрически связную кривую над конечным полем F_q (где $q = p^e$ — степень простого числа p). Обозначим через $F = F_q(X)$ поле рациональных функций на X ; $\bar{F}$ обозначает сепарабельное замыкание поля F .

Для всякой замкнутой точки $x \in |X|$ (соответствующей точке поля F) обозначим через $\mathcal{O}_x$ и F_x пополнения локального кольца $\mathcal{O}_{X,x}$ и поля F в точке x . Поле вычетов кольца $\mathcal{O}_x$ обозначается через $\kappa(x)$, а его мощность — через $q_x = q^{\deg(x)}$.

Пусть $v_x : F_x \rightarrow \mathbf{Z}$ — нормирование, такое, что $v_x(\varpi_x) = 1$, где ϖ_x — униформизирующая кольца $\mathcal{O}_x$; обозначим через $|\cdot|_x = q_x^{-v_x(\cdot)}$ соответствующую абсолютную величину.

Мы фиксируем замкнутую точку ∞ кривой X и обозначаем через $A = H^0(X - \{\infty\}, \mathcal{O}_X)$ подкольцо поля F , состоящее из элементов, являющихся целыми во всех точках $x \neq \infty$. Для такой точки x пополнение $\hat{A}_x$ отождествляется с кольцом $\mathcal{O}_x$.

Вся теория становится несколько проще при условии $\deg(\infty) = 1$. Для упрощения изложения мы будем считать это условие выполненным.

0.12. Полезно иметь в виду аналогию между функциональными и числовыми полями. При этом поле F соответствует полю $\mathbf{Q}$, точка ∞ соответствует вещественному нормированию поля $\mathbf{Q}$, а $|X| - \{\infty\}$ соответствует множеству всех простых чисел. Аналогом кольца A является $\mathbf{Z}$; а различным пополнениям $\mathcal{O}_x$ (соотв. F_x) при $x \neq \infty$ соответствуют $\mathbf{Z}_p$ (соотв. $\mathbf{Q}_p$), в то время как аналогом F_∞ является поле $\mathbf{R}$.

0.13. Наконец, через A_F , или просто A , обозначим кольцо аделей поля F (ограниченное произведение полей F_x). Для конечного множества точек S через A^S обозначим ограниченное произведение полей F_x при $x \notin S$. Часто мы будем рассматривать $A^\infty := A^{\{\infty\}}$.

1. ЭЛЛИПТИЧЕСКИЕ МОДУЛИ

1.1. Определение эллиптического A -модуля над A -схемой S .

1.1.1. Начнем с изучения частного случая, когда S является спектром поля K , и, таким образом, мы имеем гомоморфизм $i : A \rightarrow K$.

Напомним, что если K — поле характеристики $p > 0$, то аддитивная группа $G_{a,K} = \text{Spec } K[T]$ имеет, в частности, следующие эндоморфизмы: гомотетии $\lambda : T \mapsto \lambda T$ ($\lambda \in K$), гомоморфизм $\tau : T \mapsto T^p$ и его степени. Легко проверить, что произвольный эндоморфизм f группы $G_{a,K}$ можно единственным образом представить в виде суммы

$$f = \lambda_0 + \lambda_1 \tau + \cdots + \lambda_n \tau^n, \quad \lambda_i \in K.$$

Однако, вообще говоря, это кольцо эндоморфизмов не коммутативно, так как имеется очевидное соотношение

$$\tau \lambda = \lambda^p \tau.$$

Дринфельд обозначает через $K\{\tau\}$ множество полиномиальных выражений от τ с коэффициентами в K с операцией покомпонентного сложения и операцией умножения, удовлетворяющей приведенному выше соотношению. Кольцо $K\{\tau\}$ таким образом отождествляется с $\text{End}(G_{a,K})$. Мы будем говорить, что элемент $f = \lambda_0 + \cdots + \lambda_n \tau^n \in K\{\tau\}$ имеет степень p^n , если $\lambda_n \neq 0$. Производная соответствующего эндоморфизма вычисляется формулой $f' = \lambda_0$.

1.1.2. Определение. Эллиптический A -модуль (с жесткостью) над полем K — это гомоморфизм колец $\phi : A \rightarrow K\{\tau\} = \text{End}(G_{a,K})$, такой, что для любого $a \in A$ выполнено $\phi(a)' = i(a)$.

Можно доказать, что существует целое число $d \geq 0$, называемое рангом эллиптического A -модуля, такое, что для любого $a \in A - \{0\}$ выполнено $\deg \phi(a) = |a|_\infty^d$. Тривиальный случай $d = 0$, соответствующий гомоморфизму $\phi(a) = i(a)$, мы исключаем из рассмотрения.

Легко видеть, что ϕ принимает значения в подкольце $K\{\tau^e\} \subset K\{\tau\}$, порожденном элементом τ^e .

Пример. Если $X = \mathbf{P}_{F_q}^1$, а ∞ — «бесконечно удаленная точка», то A отождествляется с кольцом полиномов $F_q[Y]$. Задание эллиптического A -модуля ранга d над полем K равносильно заданию элемента $\phi(Y)$, который должен иметь вид

$$\phi(Y) = i(Y) + \mu_1 \tau^e + \mu_2 \tau^{2e} + \cdots + \mu_d \tau^{de} \quad (\mu_i \in K, \mu_d \neq 0).$$

1.1.3. Предыдущее определение естественным образом обобщается на случай произвольной A -схемы S . Если L — линейное расслоение

над S (т.е. локально свободный $\mathcal{O}_S$ -модуль ранга 1) с послойным действием кольца A , то слои расслоения L изоморфны аддитивным группам поля (изоморфизм определен с точностью до гомотетии) и легко проверить, являются ли слои эллиптическими модулями ранга d .

Определение. *Эллиптический A -модуль ранга d над S* — это линейное расслоение L над S с гомоморфизмом $\phi : A \rightarrow \text{End}_S(L)$, таким, что для любого $a \in A$ дифференциал эндоморфизма $\phi(a)$ задается структурным гомоморфизмом $A \rightarrow \mathcal{O}_S$ и для произвольной точки $s : \text{Spec } K \rightarrow S$ слой L_s является эллиптическим A -модулем ранга d в указанном выше смысле.

Гомоморфизм (соотв. изоморфизм) эллиптических модулей определяется естественным образом.

1.2. Пример: аналитическое описание эллиптических модулей над F_∞ . Обозначим через $\widehat{F}_\infty$ пополнение сепарабельного замыкания поля F_∞ , и пусть Γ есть d -мерная решетка в $\widehat{F}_\infty$, т.е. дискретный (проективный) A -подмодуль ранга d . Рассмотрим бесконечное произведение

$$\eta_\Gamma(z) = z \prod_{\gamma \in \Gamma - \{0\}} (1 - z/\gamma).$$

Легко проверить его сходимости в любой точке $z \in \widehat{F}_\infty$. Мы получаем, таким образом, отображение η_Γ поля $\widehat{F}_\infty$ в себя. Легко видеть, что это аддитивный сюръективный эндоморфизм, ядро которого равно Γ , т.е. биекция между $\widehat{F}_\infty/\Gamma$ и $\widehat{F}_\infty$. Дринфельд доказал, что, перенося посредством этой биекции естественное действие кольца A с $\widehat{F}_\infty/\Gamma$ на $\widehat{F}_\infty$, мы получаем структуру эллиптического A -модуля ϕ_Γ ранга d на $\widehat{F}_\infty$: по определению при $a \in A$ имеем

$$\phi_\Gamma(a)(\eta_\Gamma(z)) = \eta_\Gamma(az).$$

Более того, любой эллиптический модуль на $\widehat{F}_\infty$ получается таким способом из некоторой решетки Γ , которая определяется по нему однозначно с точностью до гомотетии. При $d = 2$ (соотв. $d = 1$) это описание аналогично обычному аналитическому описанию эллиптических кривых (соотв. мультипликативной группы) над $\mathbb{C}$. Существенная разница состоит в том, что если дискретная

$\mathbb{Z}$ -решетка в $\mathbb{C}$ не может иметь ранг, больший чем 2, то в нашем случае ранг d может быть произвольным, так как поле $\widehat{F}_\infty$ имеет бесконечную степень над F_∞ .

1.3. Многообразие модулей.

1.3.1. Как и в классическом случае эллиптических кривых, мы начинаем с введения понятия *структуры уровня* J , где $\{0\} \neq J \subsetneq A$ — идеал. Предположим для простоты, что (L, ϕ) — эллиптический модуль ранга d над $A[J^{-1}]$ -схемой S . Обозначим через L_J подгруппу в L , определяемую уравнениями $\phi(a)(x) = 0$ при $a \in J$. Можно доказать, что эта подгруппа (с естественным действием кольца A) *этална* над S , а ее слои изоморфны $(J^{-1}/A)^d$. Структура уровня J на (L, ϕ) по определению состоит из A -линейного изоморфизма $\psi : (J^{-1}/A)_S^d \xrightarrow{\sim} L_J$.

1.3.2. Теорема (Дринфельд). Пусть $\{0\} \neq J \subsetneq A$ — идеал. Тогда функтор, сопоставляющий каждой $A[J^{-1}]$ -схеме S множество классов изоморфизма эллиптических модулей ранга d над S , снабженных структурой уровня J , представим гладкой аффинной $A[J^{-1}]$ -схемой M_J^d чистой относительной размерности $d - 1$.

1.3.3. Дринфельд показал, что при $d = 1$ пространство M_J^d является спектром нормализации кольца $A[J^{-1}]$ в таком абелевом расширении поля F , кондуктором которого является J , которое вполне расщепляется в точке ∞ и является максимальным среди расширений, обладающих такими свойствами (этот результат похож на теорему Кронекера—Вебера).

При $d = 2$ пространства M_J^d являются аналогами модулярных кривых $Y(N)$. Аналогия становится более четкой, если воспользоваться результатами *p -адической униформизации*, дающей описание, похожее на обычное описание кривой $Y(N)_\mathbb{C}$ как фактора полуплоскости Пуанкаре. При $d \geq 2$ обозначим через Ω^d подмножество в $\mathbb{P}^{d-1}(\widehat{F}_\infty)$, состоящее из точек, не принадлежащих ни одной из гиперплоскостей, определенных над F_∞ (например, $\Omega^2 = \mathbb{P}^1(\widehat{F}_\infty) - \mathbb{P}^1(F_\infty)$ — аналог «двойной» полуплоскости Пуанкаре $\mathbb{P}^1(\mathbb{C}) - \mathbb{P}^1(\mathbb{R})$). Аналитическая теория эллиптических модулей (п. 1.2) дает следующее описание пространства $M_J^d(\widehat{F}_\infty)$:

$$M_J^d(\widehat{F}_\infty) \cong \mathrm{GL}_d(F) \backslash [\Omega^d \times (\mathrm{GL}_d(\mathbf{A}^\infty)/U_J)],$$

где U_J обозначает подгруппу в $GL_d(\hat{A}) = \prod_{x \neq \infty} GL_d(A_x)$, состоящую из элементов, равных 1 по модулю J . Можно уточнить это описание и наделить пространство Ω^d структурой *жесткого аналитического пространства*. Тогда приведенная выше биекция становится биекцией жестких аналитических пространств.

1.4. Когомологии многообразия модулей ($d = 2$)

1.4.1. При изменении идеала J аффинные кривые $M_{J,F}^2 = M_J^2 \otimes_A F$ над F образуют *проективную систему*: если $J \subset J'$, то $((J')^{-1}/A)^2$ вложено в $(J^{-1}/A)^2$; поэтому структура уровня J ограничивается до структуры уровня J' , определяя таким образом (этальный) морфизм из $M_{J,F}^2$ в $M_{J',F}^2$. Кроме того, эти кривые обладают каноническими компактификациями, которые являются гладкими проективными кривыми над F и обозначаются $\overline{M}_{J,F}^2$. Кривые $\overline{M}_{J,F}^2$ тоже образуют проективную систему.

Зафиксируем простое число $\ell \neq p$. Мы будем изучать индуктивный предел групп ℓ -адических когомологий

$$H^1 = \varinjlim_J H^1(\overline{M}_{J,F}^2 \otimes_F \overline{F}, \overline{\mathbb{Q}}_\ell).$$

На каждой из этих групп, а следовательно, и на их пределе действует группа Галуа $\text{Gal}(\overline{F}/F)$. С другой стороны, группа $GL_2(\hat{A}) = \varprojlim GL_2(A/J)$ действует справа на $M_{J,F}^2$ изменением структуры уровня, причем действие продолжается и на компактификацию. Аналогично тому, как на модулярных кривых вводится действие операторов Гекке, можно продолжить действие группы $GL_2(\hat{A})$ до действия группы $GL_2(A^\infty)$ на проективной системе пространств $\overline{M}_{J,F}^2$ (но не на отдельных пространствах $\overline{M}_{J,F}^2$). Точная конструкция содержится в [Dr1]. Кроме того, в п. 4.2 этого обзора содержится аналогичная конструкция в случае эллиптических $\mathcal{D}$ -пучков.

Таким образом, в пространстве H^1 имеется представление произведения групп $\text{Gal}(\overline{F}/F) \times GL_2(A^\infty)$.

1.4.2. Точное определение *автоморфного параболического представления* группы $GL_2(A)$ содержится, например, в [Bo-Ja]. Грубо говоря, это допустимое неприводимое представление, являющееся подфактором пространства функций на однородном пространстве $GL_2(F) \backslash GL_2(A)$. Обычно рассматриваются представления над полем комплексных чисел, однако в случае функционального поля

определение имеет смысл над любым алгебраически замкнутым полем характеристики 0, например над полем $\overline{\mathbb{Q}}_\ell$ (с другой стороны, можно было бы воспользоваться изоморфизмом $\overline{\mathbb{Q}}_\ell \cong \mathbb{C}$).

Всякое представление Π разлагается в ограниченное тензорное произведение по всем точкам x поля F своих локальных компонент Π_x (которые являются допустимыми неприводимыми представлениями групп $\mathrm{GL}_2(F_x)$). В частности, можно записать $\Pi = \Pi^\infty \otimes \Pi_\infty$, где Π^∞ обозначает ограниченное тензорное произведение компонент Π_x по $x \neq \infty$.

Напомним также, что *специальное представление* (или *представление Штейнберга*) St_∞ группы $\mathrm{GL}_2(F_\infty)$ реализуется в факторпространстве пространства локально постоянных функций на пространстве $\mathbb{P}^1(F_\infty)$, на котором эта группа действует правыми сдвигами, по подпространству постоянных функций.

Почти все компоненты Π_x автоморфного параболического представления группы $\mathrm{GL}_2(A)$ являются *неразветвленными представлениями из основной серии*, т. е. получаются унитарной индукцией из представления подгруппы верхнетреугольных матриц, заданного парой неразветвленных квазихарактеров μ_x и μ'_x мультипликативной группы F_x^* .

1.4.3. Основная теорема [Dr1] полностью описывает H^1 как представление группы $\mathrm{Gal}(\overline{F}/F) \times \mathrm{GL}_2(A^\infty)$.

Теорема (Дринфельд). *Определенное выше представление H^1 изоморфно прямой сумме $\bigoplus \Pi^\infty \otimes \sigma(\Pi^\infty)$. Суммирование ведется по множеству (классов эквивалентности) допустимых неприводимых $\overline{\mathbb{Q}}_\ell$ -представлений Π^∞ группы $\mathrm{GL}_2(A^\infty)$, таких, что представление $\Pi = \Pi^\infty \times \mathrm{St}_\infty$ является автоморфным параболическим представлением. Здесь $\sigma(\Pi^\infty)$ — двумерное $\overline{\mathbb{Q}}_\ell$ -представление группы Галуа $\mathrm{Gal}(\overline{F}/F)$, характеризующееся следующим свойством: для всякой точки x поля F , такой, что компонента Π_x является неразветвленным представлением из основной серии, ассоциированным с парой квазихарактеров μ_x и μ'_x группы F_x^* , ограничение представления $\sigma(\Pi^\infty)$ на группу разложения в точке x есть сумма пары неразветвленных квазихарактеров, переводящих геометрический элемент Фробениуса в $\mu(\varpi_x)^{-1}q^{1/2}$ и $\mu'(\varpi_x)^{-1}q^{1/2}$ соответственно.*

Эта теорема является полным аналогом результатов Эйхлера—Шимуры. Доказательство, данное Дринфельдом, весьма близко к классической теории: аналитическое (жесткое) вычисление

когомологий в духе «изоморфизма Шимуры» и использование «конгруэнц-формулы».

2. ЭЛЛИПТИЧЕСКИЕ ПУЧКИ

Напомним, что мы предположили, что $\deg(\infty) = 1$. Для сокращения обозначений в тех случаях, когда это не может привести к недоразумению, расслоенное произведение F_q -схем S и T над F_q будем обозначать просто через $S \times T$.

2.1. Определение. Пусть S есть F_q -схема. *Эллиптический пучок ранга d* (с полюсом в ∞) на S — это коммутативная диаграмма

$$\begin{array}{ccccccc} \cdots & \xrightarrow{j} & \mathcal{F}_{i-1} & \xrightarrow{j} & \mathcal{F}_i & \xrightarrow{j} & \mathcal{F}_{i+1} & \xrightarrow{j} & \cdots \\ & \nearrow t & & \nearrow t & & \nearrow t & & \nearrow t & \\ \cdots & \xrightarrow{\tau j} & {}^{\tau}\mathcal{F}_{i-1} & \xrightarrow{\tau j} & {}^{\tau}\mathcal{F}_i & \xrightarrow{\tau j} & {}^{\tau}\mathcal{F}_{i+1} & \xrightarrow{\tau j} & \cdots \end{array}$$

в которой $\mathcal{F}_i$ — локально свободные $\mathcal{O}_{X \times S}$ -модули (или, что то же самое, векторные расслоения) ранга d ($i \in \mathbb{Z}$),

$${}^{\tau}\mathcal{F}_i = (\text{id}_X \times \text{Frob}_{S/F_q})^* \mathcal{F}_i$$

(т. е. обратный образ относительно автоморфизма Фробениуса схемы S), а j и t суть $\mathcal{O}_{X \times S}$ -линейные вложения. Кроме того, должны быть выполнены следующие условия:

(i) Периодичность. $\mathcal{F}_{i+d} = \mathcal{F}_i(\{\infty\} \times S)$; d -кратная композиция морфизма j является естественным вложением (индуцированным вложением $\mathcal{O}_X \hookrightarrow \mathcal{O}_X(\infty)$).

(ii) Полюс. $\mathcal{F}_i/j(\mathcal{F}_{i-1})$ равен прямому образу $(\Gamma_{\infty})_* \mathcal{A}_i$ обратного $\mathcal{O}_S$ -модуля $\mathcal{A}_i$ относительно сечения ∞ :

$$\Gamma_{\infty} : S \rightarrow X \times S, \quad s \mapsto (\infty, s).$$

(iii) Нуль. $\mathcal{F}_i/t({}^{\tau}\mathcal{F}_{i-1})$ равен прямому образу $(\Gamma_z)_* \mathcal{B}_i$ обратимого $\mathcal{O}_S$ -модуля $\mathcal{B}_i$ относительно сечения $\Gamma_z : S \rightarrow X \times S$, являющегося графиком морфизма F_q -схем $z : S \rightarrow X - \{\infty\}$.

(iv) Нормализация. Для всякой геометрической точки s схемы S характеристика Эйлера—Пуанкаре $\chi(\mathcal{F}_{0|X \times s})$ равна нулю.

2.2. Замечания

2.2.1. В силу условий (i) или (ii) переход от $\mathcal{F}_i$ к $\mathcal{F}_{i+1}$ увеличивает характеристику Эйлера—Пуанкаре на 1; поэтому для любого $i \in \mathbb{Z}$ имеем $\chi(\mathcal{F}_{i|X \times s}) = i$.

2.2.2. Морфизм z из п. (iii) определен однозначно. Таким образом, каждому эллиптическому пучку на S соответствует S -точка кривой $X - \{\infty\}$, которая называется его *нулем*.

2.3. Структуры уровня. Пусть $I \subset X - \{\infty\}$ — конечная непустая подсхема и $(\mathcal{F}_i, j, t)$ — эллиптический пучок на S , такой, что I не пересекается с образом соответствующего морфизма z (нуля). При этих условиях пучок $\mathcal{F}_i|_{I \times S}$ не зависит (отождествляется посредством морфизма j) от i . Обозначим этот пучок через $\mathcal{F}_I$. Тогда морфизм t индуцирует изоморфизм ${}^t\mathcal{F}_I \cong \mathcal{F}_I$. Определим *структуру уровня* I как задание $\mathcal{O}_{I \times S}$ -линейного изоморфизма $\iota: \mathcal{O}_{I \times S}^d \cong \mathcal{F}_I$, такого, что следующая диаграмма коммутативна:

$$\begin{array}{ccc} & & {}^t\mathcal{F}_I \\ & \nearrow \tau_\iota & \downarrow t \\ \mathcal{O}_{I \times S}^d & & \mathcal{F}_I \\ & \searrow \iota & \end{array}$$

2.4. Эквивалентность. В своей короткой, но глубокой статье [Dr3] Дринфельд доказал, что понятия эллиптического пучка и эллиптического модуля эквивалентны. Вдохновившись идеей, происходящей из работы Кричевера [Kr] по дифференциальным уравнениям, он подчеркнул волнующую связь между этими двумя, далекими на первый взгляд, объектами. Об этом см. также статью [Mu].

Пусть $\{0\} \neq J \subset A$ — идеал в кольце A , соответствующий замкнутой подсхеме I в $X - \{\infty\} = \text{Spec}(A)$. Результат Дринфельда можно сформулировать так:

Теорема. Пусть $S \xrightarrow{z} \text{Spec}(A) - I$ является $A[J^{-1}]$ -схемой. Существует функториальная по S биекция между множеством классов изоморфизма эллиптических A -модулей ранга d на S , снабженных структурой уровня J , и множеством классов изоморфизма эллиптических пучков ранга d на S с нулем z , снабженных структурой уровня I .

Функтор, сопоставляющий схеме S множество классов изоморфизма эллиптических пучков ранга d на S со структурой уровня I , следовательно, представим $A[J^{-1}]$ -схемой, изоморфной M_J^d .

2.5. Теперь мы изложим идею конструкции Дринфельда, предположив для простоты, что S является спектром поля K .

2.5.1. Выберем эллиптический пучок $(\mathcal{F}_i, j, t)$ на $\text{Spec } K$ с нулем $z : A \rightarrow K$. Рассмотрим K -векторные пространства $M_i \stackrel{\text{def}}{=} H^0(X \otimes K, \mathcal{F}_i)$, которые морфизмом j объединяются в возрастающую последовательность. Их объединение M отождествляется для всех i с $H^0((X - \{\infty\}) \otimes K, \mathcal{F}_i)$ (заметим, что j является изоморфизмом над $X - \{\infty\}$). Пространство M , очевидно, является $(A \otimes_{F_q} K)$ -модулем.

Воспользуемся теперь отображением $t : M \rightarrow M$, индуцированным морфизмом $t : {}^t\mathcal{F}_i \rightarrow \mathcal{F}_{i+1}$. Это отображение вкладывает M_i в M_{i+1} и является A -линейным и K -полулинейным, т. е.

$$(am) = at(m) \quad \text{при } a \in A,$$

$$(\lambda m) = \lambda^q t(m) \quad \text{при } \lambda \in K.$$

2.5.2. Если пользоваться обозначениями п. 1.1, то полулинейность отображения t означает, что пространство M можно рассматривать как левый $K\{\tau^e\}$ -модуль, в котором τ^e действует оператором t (мы временно забываем о действии кольца A). Докажем теперь, что M — свободный одномерный модуль.

(а) Нуль и полюс эллиптического пучка не пересекаются; поэтому для всякого i морфизм t индуцирует вложение (а если поле K совершенно, то изоморфизм) $t : {}^t(\mathcal{F}_i/j(\mathcal{F}_{i-1})) \rightarrow \mathcal{F}_{i+1}/j(\mathcal{F}_i)$; следовательно, отображение $t : M_i/M_{i-1} \rightarrow M_{i+1}/M_i$ инъективно.

(b) $M_0 = \{0\}$. Действительно, в противном случае существовали бы $i \leq 0$ и элемент $x \in M_i - M_{i-1}$. Воспользовавшись п. (а), мы получили бы, что $t^m x \in M_{i+m} - M_{i+m-1}$ для всякого $m \geq 0$; следовательно, $\dim_K M_{i+m} \geq m + 1$ (так как элементы $x, tx, \dots, t^m x$ линейно независимы). Однако для больших m это противоречит замечанию 2.2.1, поскольку из него следует, что

$$\dim_K M_{i+m} = \chi(\mathcal{F}_{i+m}) = i + m.$$

(с) Те же самые аргументы, что и в п. (b), показывают, что при $i \geq 0$ пространство M_{i+1}/M_i одномерно над K . Отсюда по индукции следует, что если μ — базисный вектор пространства M_1 , то $\{\mu, t\mu, \dots, t^{n-1}\mu\}$ является базисом пространства M_n над K . Следовательно, пространство M является свободным одномерным модулем над $K\{\tau^e\} = K\{t\}$ с базисным вектором μ .

2.5.3. Зафиксировав базис μ , мы получим изоморфизм $M \cong K\{t\}$ (мы отождествляем t и τ^e). Вспомним теперь о действии кольца A

на пространстве M . Это действие коммутирует с действием кольца $K\{t\}$; поэтому мы имеем гомоморфизм

$$\phi : A \rightarrow \text{End}_{K\{t\}}(K\{t\}) = K\{t\}^{\text{opp}},$$

который в силу коммутативности кольца A можно рассматривать и как гомоморфизм $A \rightarrow K\{t\} = K\{\tau^e\}$. Для вычисления производной $(\phi(a))'$ надо рассмотреть действие A на $M/K \cdot t(M) \cong M_1$. Это пространство естественным образом вкладывается в слой пучка $\mathcal{F}_1$ над нулем; поэтому $(\phi(a))' = z(a)$.

Наконец, для каждого $a \in A - \{0\}$ имеем

$$aM_i \subset H^0(X \otimes K, \mathcal{F}_i(-v_\infty(a))) = M_{i-dv_\infty(a)},$$

причем $-dv_\infty(a)$ — это самое маленькое целое число, при котором включение всегда верно. Отсюда следует, что степень элемента $\phi(a) \in K\{\tau\}$ равна $p^{e(-dv_\infty(a))} = (q^{-v_\infty(a)})^d = |a|_\infty^d$.

Таким образом, мы получили эллиптический модуль ранга d .

2.5.4. Продолжим отождествление M и $K\{t\}$. Задание структуры уровня I на эллиптическом пучке — это задание $(A \otimes K)$ -линейного изоморфизма

$$\iota : (A/J)^d \otimes_{F_q} K \xrightarrow{\sim} M/MJ,$$

такого, что действие автоморфизма Фробениуса на левой части соответствует действию оператора t на правой части.

Пусть $\lambda \in [(A/J)^d]^*$ — элемент из F_q -двойственного пространства к $(A/J)^d$. Его композиция с отображением ι^{-1} дает K -линейный гомоморфизм h_λ из $M/MJ = K\{\tau\}/K\{\tau\} \cdot \phi(J)$ в K , удовлетворяющий условию $h_\lambda(tm) = h_\lambda(m)^q$.

Любой такой гомоморфизм однозначно определяется своим значением $\psi_\lambda = h_\lambda(1)$ (так как $h_\lambda(t^m) = \psi_\lambda^{q^m}$). Пусть $a \in J$ и $\phi(a) = u_0 + u_1 t + \dots + u_n t^n$. Тогда значение ψ_λ должно удовлетворять уравнению

$$\phi(a)(\psi_\lambda) = u_0 \psi_\lambda + u_1 \psi_\lambda^q + \dots + u_n \psi_\lambda^{q^n} = h_\lambda(\phi(a)) = 0.$$

Иначе говоря, элемент ψ_λ должен быть точкой J -кручения эллиптического модуля ϕ . Таким образом, отображение $\lambda \rightarrow \psi_\lambda$ является A -линейной инъекцией (а следовательно, биекцией)

$$\psi : [(A/J)^d]^* \xrightarrow{\sim} (\phi)_{J\text{-tors}}.$$

Значит, на модуле ϕ возникает структура уровня J (зависящая от выбора изоморфизма $(A/J)^* \cong J^{-1}/A$).

Легко проверить, что все предыдущие конструкции с точностью до изоморфизма не зависят от выбора элемента μ .

3. ЭЛЛИПТИЧЕСКИЕ $\mathcal{D}$ -ПУЧКИ

3.1. Зафиксируем простую центральную алгебру D размерности d^2 над своим центром F , *расщепляющуюся в бесконечности*, т.е. такую, что $D \otimes_F F_\infty \cong M_d(F_\infty)$. Кроме того, выберем когерентный локально свободный пучок $\mathcal{O}_X$ -алгебр $\mathcal{D}$ с общим слоем D , такой, что для любой замкнутой точки x кривой X его слой $\mathcal{D}_x = D \otimes_{\mathcal{O}_X} \mathcal{O}_x$ является *максимальным порядком* в $D_x = D \otimes_F F_x$. Выбор пучка $\mathcal{D}$ равносильно такому выбору всех $\mathcal{D}_x$, чтобы они были когерентны в том смысле, что для почти всех x слой $\mathcal{D}_x$ должен порождаться над кольцом $\mathcal{O}_x$ одним и тем же базисом алгебры D над полем F .

Обозначим через $\mathcal{R}$ множество тех точек, в которых алгебра $\mathcal{D}_x$ не расщепляется. Тогда для всех прочих точек $x \notin \mathcal{R}$ пара $(D_x, \mathcal{D}_x)$ изоморфна паре $(M_d(F_x), M_d(\mathcal{O}_x))$.

Напомним, что мы для простоты ввели предположение, что $\deg(\infty) = 1$. Эллиптический $\mathcal{D}$ -пучок определен Ломоном, Рапопортом и Штулером следующим образом (напомним, что для F_q -схем S и T мы обозначаем через $S \times T$ произведение над F_q).

3.1.1. Определение. Пусть S есть F_q -схема. *Эллиптический $\mathcal{D}$ -пучок ранга d (с полюсом в ∞) на S* — это коммутативная диаграмма

$$\begin{array}{ccccccc} \cdots & \xrightarrow{j} & \mathcal{E}_{i-1} & \xrightarrow{j} & \mathcal{E}_i & \xrightarrow{j} & \mathcal{E}_{i+1} & \xrightarrow{j} & \cdots \\ & \nearrow t & & \nearrow t & & \nearrow t & & \nearrow t & \\ \cdots & \xrightarrow{\tau j} & \tau \mathcal{E}_{i-1} & \xrightarrow{\tau j} & \tau \mathcal{E}_i & \xrightarrow{\tau j} & \tau \mathcal{E}_{i+1} & \xrightarrow{\tau j} & \cdots \end{array}$$

в которой $\mathcal{E}_i$ — локально свободный $\mathcal{O}_{X \times S}$ -модуль ранга d^2 ($i \in \mathbb{Z}$) с правым действием $\mathcal{D}$, согласованным с действием $\mathcal{O}_X$. Далее, $\tau \mathcal{E}_i$ обозначает то же, что и в предыдущем разделе, а j и t — это $\mathcal{O}_{X \times S}$ -линейные вложения, согласованные с действием $\mathcal{D}$. Кроме того, должны быть выполнены следующие условия:

(i) Периодичность. $\mathcal{E}_{i+d} = \mathcal{E}_i(\{\infty\} \times S)$, d -кратная композиция морфизма j является естественным вложением.

(ii) Полюс. $\mathcal{E}_i/j(\mathcal{E}_{i-1})$ как $\mathcal{O}_{X \times S}$ -модуль изоморфен прямому образу $(\Gamma_\infty)_* A_i$ локально свободного $\mathcal{O}_S$ -модуля A_i ранга d относительно сечения ∞ .

(iii) Нуль. $\mathcal{E}_i/t(\Gamma\mathcal{E}_{i-1})$ как $\mathcal{O}_{X \times S}$ -модуль изоморфен прямому образу $(\Gamma_z)_* \mathcal{B}_i$ локально свободного $\mathcal{O}_S$ -модуля $\mathcal{B}_i$ ранга d относительно сечения $\Gamma_z : S \rightarrow X \times S$, являющегося графиком морфизма Γ_q -схем $z : S \rightarrow X - \{\infty\} - \mathcal{R}$.

3.1.2. Нормализация. Условия (ii) и (iii) показывают, что характеристика Эйлера—Пуанкаре пучка $\mathcal{E}_i$ в любой геометрической точке s схемы S при переходе к $\mathcal{E}_{i+1}$ увеличивается на d . Поэтому имеет смысл наложить условие нормализации, аналогичное условию (iv) п. 2.1, например $\chi(\mathcal{E}_0|_{X \times s}) \in [0, d[$. В нашем обзоре мы будем считать, что эллиптические $\mathcal{D}$ -пучки нормализованы таким способом.

В статье [L-R-S], напротив, никаких условий нормализации не налагается. Вместо этого рассматривается естественное действие (сдвигами) группы $\mathbb{Z}$ на стэке модулей эллиптических $\mathcal{D}$ -пучков, и после перехода к фактору получается то же самое.

3.1.3. Точно так же, как и в случае эллиптических пучков, морфизм z корректно определен; таким образом, каждому эллиптическому $\mathcal{D}$ -пучку на схеме S соответствует S -точка кривой $X - \{\infty\} - \mathcal{R}$.

3.1.4. Пример: случай матричной алгебры D . Предположим, что $D = M_d(F)$ и $\mathcal{D} = M_d(\mathcal{O}_X)$. Каждому эллиптическому пучку $(\mathcal{F}_i, j, t)$ можно сопоставить тройку $(\mathcal{F}_i \boxtimes \mathcal{O}_X^d, j \boxtimes \text{id}, t \boxtimes \text{id})$. Эта тройка является эллиптическим $\mathcal{D}$ -пучком $(\mathcal{E}_i, j', t')$, в котором $\mathcal{E}_i$ есть прямая сумма d экземпляров пучка $\mathcal{F}_i$, а алгебра $\mathcal{D}$ действует справа матричным умножением. Обратно, разлагая эллиптический $\mathcal{D}$ -пучок относительно действия диагональных идемпотентов алгебры $\mathcal{D}$, несложно проверить, что любой нормализованный эллиптический $\mathcal{D}$ -пучок получается таким образом (эквивалентность Мориты). Поэтому в случае матричной алгебры ничего нового мы не получаем.

3.2. Структуры уровня. Пусть $I \subset X - \{\infty\}$ — конечная непустая подсхема и $(\mathcal{E}_i, j, t)$ — эллиптический $\mathcal{D}$ -пучок на S , такой, что I не пересекается с образом соответствующего морфизма z . При этих условиях пучок $\mathcal{E}_I = \mathcal{E}_i|_{I \times S}$ не зависит от i (пучки $\mathcal{E}_{i-1}|_{I \times S}$ и $\mathcal{E}_i|_{I \times S}$ отождествляются морфизмом j) и морфизм t индуцирует изоморфизм ${}^t\mathcal{E}_I \cong \mathcal{E}_I$.

Определим теперь *структуру уровня* I как задание $\mathcal{O}_{I \times S}$ -линейного изоморфизма $\iota : \mathcal{D}_I \boxtimes \mathcal{O}_S \xrightarrow{\sim} \mathcal{E}_I$, согласованного с дей-

ствием $\mathcal{D}_I$ на обеих частях, такого, что следующая диаграмма коммутативна:

$$\begin{array}{ccc} & & {}^{\tau}\mathcal{E}_I \\ & \nearrow \tau_t & \downarrow t \\ \mathcal{D}_I \boxtimes \mathcal{O}_S & & \mathcal{E}_I \\ & \searrow t & \end{array}$$

3.3. Многообразие модулей. Зафиксируем непустую конечную подсхему $I \subset X - \{\infty\}$. Пусть $z : S \rightarrow X - \{\infty\} - I - \mathcal{R}$ есть S -точка схемы $X - \{\infty\} - I - \mathcal{R}$. Обозначим через $\mathcal{E}ll_{X, \mathcal{D}, I}(S)$ множество классов изоморфизма эллиптических $\mathcal{D}$ -пучков на S с нулем z , снабженных структурой уровня I и нормализованных, как в п. 3.1.2.

Первый важный результат, доказанный в [L-R-S], — это

3.3.1. Теорема. Функтор $\mathcal{E}ll_{X, \mathcal{D}, I}$ представим схемой, квазипроективной над $X - \{\infty\} - I - \mathcal{R}$:

$$\mathcal{E}ll_{X, \mathcal{D}, I} \xrightarrow{\text{zero}} X - \{\infty\} - I - \mathcal{R}.$$

Схема $\mathcal{E}ll_{X, \mathcal{D}, I}$ гладкая, чистой относительной размерности $d-1$ над $X - \{\infty\} - I - \mathcal{R}$. Если $\mathcal{D}$ — алгебра с делением, то схема $\mathcal{E}ll_{X, \mathcal{D}, I}$ проективна над $X - \{\infty\} - I - \mathcal{R}$.

Например, если $\mathcal{D}$ — матричная алгебра, то рассуждения п. 3.1.4 и теорема п. 2.4 показывают, что в этом случае схема $\mathcal{E}ll_{X, \mathcal{D}, I}$ изоморфна схеме M_J^d , где J — идеал в кольце A , соответствующий подсхеме I .

3.4. Наметим вкратце доказательство теоремы о представимости.

3.4.1. Одна из важных составных частей доказательства — это утверждение о (почти) стабильности векторных расслоений $\mathcal{E}_i$. А именно, существует константа C , такая, что для всякого эллиптического $\mathcal{D}$ -пучка $(\mathcal{E}_i, j, t)$ над полем и всякого собственного векторного подрасслоения $\mathcal{F}$ в $\mathcal{E}_0$ выполнено неравенство

$$\frac{\deg \mathcal{F} - C}{\text{rk } \mathcal{F}} < \frac{\deg \mathcal{E}_0 - C}{\text{rk } \mathcal{E}_0},$$

где $\deg$ обозначает степень, а rk — ранг векторного расслоения. Для доказательства этого утверждения используются свойства канонической фильтрации Хардера—Нарасимхана—Квиллена [Ha-Na].

3.4.2. Если степень I больше чем C (мы всегда можем это предполагать, переходя впоследствии к фактору по конечной группе) и $(\mathcal{E}_i, j, t)$ снабжен структурой уровня I , то приведенное выше неравенство означает, что расслоение $\mathcal{E}_0$ стабильно как расслоение со структурой уровня. Поэтому существует схема над $\mathbf{F}_q$ (квазипроективная и гладкая), классифицирующая такие расслоения. Отсюда можно вывести, что существует $\mathbf{F}_q$ -схема $\text{Vect}_{X,D,I}^\bullet$, классифицирующая последовательности

$$\dots \xrightarrow{j} \mathcal{E}_{-1} \xrightarrow{j} \mathcal{E}_0 \xrightarrow{j} \mathcal{E}_1 \xrightarrow{j} \mathcal{E}_2 \xrightarrow{j} \dots$$

векторных расслоений ранга d^2 над $X \times S$ с (согласованным) действием пучка алгебр $\mathcal{D}$ и с (согласованными) структурами уровня I $\mathcal{D}_I \boxtimes \mathcal{O}_S \cong \mathcal{E}_I$, такие, что условия (i) и (ii) п. 3.1.1 выполнены, а расслоение $\mathcal{E}_0$ стабильно (как расслоение со структурой уровня) и нормализовано, как в п. 3.1.2.

Далее, существует также $\mathbf{F}_q$ -схема $\text{Hecke}_{X,D,I}$, классифицирующая коммутативные диаграммы

$$\begin{array}{ccccccc} \dots & \xrightarrow{j} & \mathcal{E}_{i-1} & \xrightarrow{j} & \mathcal{E}_i & \xrightarrow{j} & \mathcal{E}_{i+1} & \xrightarrow{j} & \dots \\ & \nearrow t & & \nearrow t & & \nearrow t & & \nearrow t & \\ \dots & \xrightarrow{j'} & \mathcal{E}'_{i-1} & \xrightarrow{j'} & \mathcal{E}'_i & \xrightarrow{j'} & \mathcal{E}'_{i+1} & \xrightarrow{j'} & \dots \end{array}$$

в которых строки — это последовательности из $\text{Vect}_{X,D,I}^\bullet$, а отображения t являются $\mathcal{O}_{X \times S}$ -линейными вложениями, согласованными с действием $\mathcal{D}$ и со структурами уровня, и удовлетворяют условию (iii) п. 3.1.1.

3.4.3. Можно доказать, что схема $\text{Vect}_{X,D,I}^\bullet$ гладкая и проективная над $\mathbf{F}_q$ и что морфизм

$$\text{Hecke}_{X,D,I} \xrightarrow{(z,r_1)} (X - \{\infty\} - I - \mathcal{R}) \times \text{Vect}_{X,D,I}^\bullet,$$

переводящий диаграмму в пару (нуль и первая строка), — гладкий морфизм относительной размерности $d - 1$.

Остается заметить, что следующее расслоенное произведение является гладкой квазипроективной схемой относительной размер-

ности $d-1$ над $X - \{\infty\} - I - \mathcal{R}$ и, очевидно, представляет функтор $\mathcal{E}ll_{X,D,I}$:

$$\begin{array}{ccc}
 \mathcal{E}ll_{X,D,I} & \xrightarrow{r_1} & \text{Vect}_{X,D,I}^\bullet \\
 \downarrow & & \downarrow (\text{id}, \text{Frob}) \\
 \text{Hecke}_{X,D,I} & \xrightarrow{(r_1, r_2)} & \text{Vect}_{X,D,I}^\bullet \times \text{Vect}_{X,D,I}^\bullet \\
 \downarrow z & & \\
 X - \{\infty\} - I - \mathcal{R} & &
 \end{array}$$

3.4.4. Проективность схемы в том случае, когда D является алгеброй с делением, доказывается с помощью валюативного критерия: для этого применяется *теорема о полустабильной редукции*, доказанная Дринфельдом в работе о компактификации многообразий модулей штурк [Dr6].

4. КОГОМОЛОГИИ МНОГООБРАЗИЯ МОДУЛЕЙ: ОСНОВНАЯ ТЕОРЕМА

Всюду в дальнейшем мы предполагаем, что D — алгебра с делением. Для упрощения обозначений мы будем писать $\mathcal{E}ll_I$ вместо $\mathcal{E}ll_{X,D,I}$ и $\mathcal{E}ll_{I,F}$ вместо $\mathcal{E}ll_I \times_X \text{Spec } F$.

4.1. Многообразия $\mathcal{E}ll_{I,F}$ — гладкие проективные $(d-1)$ -мерные многообразия над F . Для $I' \supset I$ ограничение структуры уровня определяет (конечную этальную) проекцию многообразия $\mathcal{E}ll_{I',F}$ на $\mathcal{E}ll_{I,F}$. Мы интересуемся пространствами ℓ -адических когомологий (где $\ell \neq p$ — фиксированное простое число и $0 \leq n \leq 2d-2$)

$$H_I^n = H^n(\mathcal{E}ll_{I,F} \otimes_F \bar{F}, \bar{\mathcal{Q}}_\ell),$$

а также их индуктивным пределом $H^n = \varinjlim_I H_I^n$. На каждом из этих пространств действует группа $\text{Gal}(\bar{F}/F)$.

4.2.1. Обозначим через G алгебраическую группу над F , определяемую равенством $G(B) = (D \otimes_F B)^*$, где B есть F -алгебра. В частности, $G(F) = D^*$. Обозначим через $G(A)$ (соотв. $G(A^{\{\infty\}})$) группу ее аделльных точек (соотв. точек со значениями в «конечных» аделях). Определим *действие справа* группы $G(A^{\{\infty\}})$ на проективной системе многообразий $\mathcal{E}ll_{I,F}$, в результате чего мы получим левое действие (коммутирующее с действием группы Галуа)

на пространстве H^n . Для того чтобы задать это действие, достаточно определить действие «целых» элементов $g \in G(A^{\{\infty\}}) \cap \mathcal{D}_{\hat{A}}$, где $\hat{A} = \prod_{x \neq \infty} \mathcal{O}_x$ и $\mathcal{D}_{\hat{A}} = \mathcal{D} \otimes_A \hat{A} = \varprojlim \mathcal{D}_I = \prod_{x \neq \infty} \mathcal{D}_x$.

4.2.2. Пусть $I = \text{Spec}(A/J)$ — замкнутое множество. Обозначим теперь через U_I конгруэнц-подгруппу группы $\mathcal{D}_{\hat{A}}^*$, состоящую из элементов, равных 1 по модулю J . Пусть g — целый элемент, а множества I и I' таковы, что $gU_I g^{-1} \subset U_{I'}$.

Задание структуры уровня I на эллиптическом $\mathcal{D}$ -пучке $(\mathcal{E}_i, j, t)$, определенном на сепарабельно замкнутом расширении K поля F , равносильно заданию морфизма $\iota: \mathcal{D}_{\hat{A}} \otimes K \xrightarrow{\sim} \mathcal{E}_i \otimes_{A \otimes K} (\hat{A} \otimes K)$, определенного по модулю действия группы U_I .

Элемент g задает эндоморфизм пучка $\mathcal{D}_{\hat{A}}$, а следовательно, посредством изоморфизма ι , и пучка $\mathcal{E}_i \otimes_{A \otimes K} (\hat{A} \otimes K)$.

Теперь мы можем построить эллиптический $\mathcal{D}$ -пучок $(\mathcal{E}', j', t')$, входящий в декартов квадрат

$$\begin{array}{ccc} \mathcal{E}'_i & \longrightarrow & \mathcal{E}_i \otimes_{A \otimes K} (\hat{A} \otimes K) \\ \downarrow & & \downarrow g \\ \mathcal{E}_i & \longrightarrow & \mathcal{E}_i \otimes_{A \otimes K} (\hat{A} \otimes K) \end{array}$$

Точнее говоря, локальные сечения пучка $\mathcal{E}'_i$ — это локальные сечения пучка $\mathcal{E}_i$, удовлетворяющие некоторым условиям (лежать в образе g) в тех точках x , в которых компонента g_x не биективна.

Мы снабжаем пучок $(\mathcal{E}'_i, j', t')$ структурой уровня ι' (корректно определенной по модулю действия группы $U_{I'}$) так, чтобы следующая диаграмма стала коммутативной:

$$\begin{array}{ccc} \mathcal{D}_{\hat{A}} \otimes K & \xrightarrow{\iota'} & \mathcal{E}'_i \otimes_{A \otimes K} (\hat{A} \otimes K) \\ \downarrow g & & \downarrow \\ \mathcal{D}_{\hat{A}} \otimes K & \xrightarrow{\iota} & \mathcal{E}_i \otimes_{A \otimes K} (\hat{A} \otimes K) \end{array}$$

Можно также проверить, что отображение $(\mathcal{E}_i, j, t, \iota) \rightarrow (\mathcal{E}'_i, j', t', \iota') \bmod \mathbf{Z}$ (т.е. с точностью до сдвига, необходимого для выполнения условия нормализации п. 3.1.2) задает морфизм $g: \mathcal{E}\ell\ell_{I,F} \rightarrow \mathcal{E}\ell\ell_{I',F}$.

4.2.3. Предыдущая конструкция задает действие группы $G(A^{\infty})$ на проективной системе $\mathcal{E}\ell\ell_{I,F}$, а также на ее проективном пределе $\mathcal{E}\ell\ell_F$. Можно, кроме того, ввести *соответствия Гекке*. Элемент $g \in G(A^{\infty})$ (или, точнее говоря, его двойной класс смежности

по модулю U_I) индуцирует на проективной системе $\mathcal{E}ll_{I,F}$ соответствие

$$\begin{array}{ccc} & \mathcal{E}ll_F / (U_I \cap g^{-1}U_I g) & \\ \alpha \swarrow & & \searrow \beta \\ \mathcal{E}ll_F / U_I & \cdots \cdots \cdots & \mathcal{E}ll_F / U_I \end{array}$$

где морфизм α индуцирован вложением $U_I \cap g^{-1}U_I g$ в U_I , а морфизм β индуцирован действием элемента g^{-1} на $\mathcal{E}ll_F$ и вложением $u \mapsto gug^{-1}$ из $U_I \cap g^{-1}U_I g$ в U_I .

Можно проверить, что если компонента g_x в точке $x \neq \infty$ тривиальна, то соответствие продолжается до соответствия на $\mathcal{E}ll_I \times_X \text{Spec}(\mathcal{O}_{X,x})$, причем α и β — этальные конечные морфизмы.

4.3.1. Рассмотрим пространство $L^2 = L^2(G(F) \backslash G(\mathbb{A}) / \varpi_\infty^{\mathbb{Z}})$ квадратично интегрируемых функций на (компактном) пространстве $G(F) \backslash G(\mathbb{A}) / \varpi_\infty^{\mathbb{Z}}$. Пространство L^2 разлагается относительно естественного действия группы $G(\mathbb{A})$ в гильбертову прямую сумму:

$$L^2 = \widehat{\bigoplus_{\Pi} m(\Pi) \Pi},$$

с конечными кратностями $m(\Pi)$. Неприводимые представления Π , входящие в это разложение, являются по определению *автоморфными представлениями* (с тривиальным на $\varpi_\infty \in \mathbb{A}^*$ центральным характером) группы $G(\mathbb{A})$. Можно, однако, дать и более алгебраическое определение. Для этого вместо пространства L^2 возьмем пространство локально постоянных функций и рассмотрим все его неприводимые подфакторы (см. [Bo-Ja]). Это определение сохраняет смысл, если поле $\mathbb{C}$ заменить произвольным алгебраически замкнутым полем характеристики 0, например полем $\overline{\mathbb{Q}}_\ell$. Именно такой точки зрения мы и будем придерживаться в дальнейшем.

Каждое автоморфное представление Π разлагается в ограниченное тензорное произведение $\bigotimes \Pi_x$ (допустимых неприводимых) представлений групп $G(F_x)$. Мы будем пользоваться обозначением $\Pi = \Pi^\infty \otimes \Pi_\infty$, введенным в п. 1.4.2.

4.3.2. Среди автоморфных представлений группы $G(\mathbb{A})$ выделим *характеры* $\chi \circ \text{Nr}$, разлагающиеся в композицию приведенной нормы $\text{Nr} : G(\mathbb{A}) \rightarrow \mathbb{A}^*$ и характера Гекке χ группы $\mathbb{A}^*$. Кратность таких

представлений равна 1. Все остальные автоморфные представления бесконечномерны (точнее говоря, их компоненты Π_x в точках $x \notin \mathcal{R}$ бесконечномерны).

4.3.3. Рассмотрим точку $x \notin \mathcal{R}$. Представление Π_x группы $G(F_x) \cong \mathrm{GL}_d(F_x)$ (допустимое и неприводимое) называется *неразветвленным*, если оно обладает вектором, инвариантным относительно максимальной компактной подгруппы $\mathrm{GL}_d(\mathcal{O}_x)$ (см. [Ca1]). Такие представления реализуются как компоненты в индуцированном представлении $\mathrm{Ind}(\mu_1, \dots, \mu_d)$ (унитарное индуцирование) из d неразветвленных квазихарактеров группы F_x^* . Числа $z_i = \mu_i(\varpi_x)$, корректно определенные с точностью до порядка, называются *собственными значениями Гекке* представления Π_x .

4.3.4. Наконец, представлением Стейнберга St_∞ группы $G(F_\infty) \cong \mathrm{GL}_d(F_\infty)$ называется единственное неприводимое факторпредставление индуцированного представления

$$\mathrm{Ind} \left(\left| \cdot \right|_\infty^{-\frac{d-1}{2}}, \left| \cdot \right|_\infty^{-\frac{d-3}{2}}, \dots, \left| \cdot \right|_\infty^{-\frac{d-3}{2}}, \left| \cdot \right|_\infty^{-\frac{d-1}{2}} \right)$$

(унитарное индуцирование с группы верхнетреугольных матриц). Это представление дискретной серии.

4.4.1. Рассмотрим полупростую оболочку $(H^n)^{ss}$ представления H^n . Это представление произведения $G(\mathbb{A}^\infty) \times \mathrm{Gal}(\overline{F}/F)$ разлагается в прямую сумму изотипических компонент по классам эквивалентности допустимых неприводимых представлений Π^∞ группы $G(\mathbb{A}^\infty)$:

$$(H^n)^{ss} = \bigoplus_{\Pi^\infty} \Pi^\infty \otimes V_{\Pi^\infty}^n,$$

где $V_{\Pi^\infty}^n$ — конечномерные представления группы $\mathrm{Gal}(\overline{F}/F)$. Главная теорема [L-R-S] описывает это разложение. Говоря коротко, она утверждает, что в указанную выше сумму нетривиально входят лишь те представления Π^∞ , которые являются «конечной» компонентой автоморфных представлений Π , причем представление Π либо должно быть характером (см. п. 4.3.2), либо таким, что $\Pi_\infty \cong \mathrm{St}_\infty$. С представлениями первого типа связаны характеры Галуа (они входят в четномерные когомологии), в то время как представления второго типа задают более интересные представления Галуа, которые связаны с глобальными гипотезами Ленглендса. Эти представления входят в средние $(d-1)$ -мерные когомологии.

4.4.2. Теорема. (i) В изотипическое разложение представления $(H^n)^{ss}$ нетривиально входят только такие представления Π^∞ , что либо $\Pi^\infty \otimes 1$, либо $\Pi^\infty \otimes \text{St}_\infty$ — автоморфное представление. В первом случае Π^∞ разлагается в композицию вида $\chi^\infty \circ \text{Nr}$, где χ^∞ — характер группы $(A^\infty)^*$, такой, что $\chi = \chi^\infty \otimes 1$ является характером Гекке группы A^* .

(ii) В первом случае ($\Pi^\infty = \chi^\infty \circ \text{Nr}$) представление $V_{\Pi^\infty}^0$ является характером Галуа, соответствующим характеру χ по теории полей классов (используется геометрическая нормализация изоморфизма теории полей классов). При $0 \leq i \leq d-1$ имеем $V_{\Pi^\infty}^{2i} = V_{\Pi^\infty}^0(-i)$ (подкрутка Тэйта), в то время как $V_{\Pi^\infty}^{2i+1}$ равно нулю.

(iii) Во втором случае (когда $\Pi = \Pi^\infty \otimes \text{St}_\infty$ автоморфно) представления $V_{\Pi^\infty}^n$ равны нулю при $n \neq d-1$, в то время как $V_{\Pi^\infty}^{d-1}$ является полупростым $m(\Pi)d$ -мерным представлением группы $\text{Gal}(\overline{F}/F)$, характеризующимся следующим свойством: если компонента Π_x в точке $x \notin \mathcal{R} \cap \{\infty\}$ неразветвлена, то представление $V_{\Pi^\infty}^{d-1}$ неразветвлено в точке x и

$$\text{tr}(\text{Frob}_x; V_{\Pi^\infty}^{d-1}) = m(\Pi)q_x^{r(d-1)/2}(z_1(\Pi_x)^r + \dots + z_d(\Pi_x)^r),$$

где Frob_x обозначает (геометрический) элемент Фробениуса, а $z_i(\Pi_x)$ — собственные значения Гекке представления Π_x .

(iv) Помимо этого, во втором случае полупростая оболочка относительно Frob_∞ (см. [De]) представления $V_{\Pi^\infty}^{d-1} |_{\text{Gal}(\overline{F}_\infty/F_\infty)}$ изоморфна прямой сумме $m(\Pi)$ экземпляров представления $\text{Sp}_\infty(-d+1)$, т. е. подкрученного по Тэйту специального представления.

4.4.3. Из п. (iii) предыдущей теоремы и из гипотез Вейля легко получить доказательство гипотезы Рамануджана—Петерсона для автоморфных представлений Π группы $G(A)$, таких, что $\Pi_\infty \cong \text{St}_\infty$, утверждающей, что собственные значения Гекке представления Π_x (в тех точках, где Π_x неразветвлено) имеют модуль 1 при любом погружении в поле комплексных чисел.

4.4.4. Комментарий. Гипотетически кратность $m(\Pi)$, входящая в утверждение теоремы, всегда равна 1. Это должно следовать из обобщенного соответствия Жакс—Ленглендса между автоморфными представлениями группы $G(A)$ и некоторыми автоморфными представлениями группы $\text{GL}_d(A)$ (см. [De-Ka-Vi] и [Ro]). Однако

это соответствие еще не построено для функциональных полей (за исключением случаев $d = 2$ и $d = 3$). В тех случаях когда $m(\Pi) = 1$, представление $V_{\Pi\infty}^{d-1}$ является (с точностью до подкрутки на квази-характер) тем представлением, которое гипотезы Ленглендса сопоставляют представлению Π .

5. ПЛАН ДОКАЗАТЕЛЬСТВА ОСНОВНОЙ ТЕОРЕМЫ

Доказательство основной теоремы занимает большую часть статьи [L-R-S]. Мы приведем здесь лишь план доказательства. В общих чертах он сводится к тому, чтобы сначала с помощью формул следа вычислить виртуальные представления $H_I^* = \sum (-1)^n H_I^n$, а затем из геометрических соображений получить и сами H_I^n .

5.1. Зафиксируем замкнутую точку $x \in |X| - \{\infty\} - \mathcal{R}$. Пусть I — замкнутая подсхема в $X - \{\infty, x\}$. Рассмотрим алгебру Гекке $\mathcal{H}_I^{x,\infty}$ локально постоянных ($\overline{\mathbb{Q}_\ell}$ -значных) функций с компактным носителем на $G(A^{x,\infty})$, бинвариантных относительно U_I (см. п. 4.2.2). Эта алгебра действует на H_I^n . Если $g^{x,\infty} \in G(A^{x,\infty})$, то характеристическая функция двойного класса смежности $U_I g^{x,\infty} U_I$ действует посредством соответствия Гекке (п. 4.2.3), отвечающего элементу $g^{x,\infty}$ (к которому добавлена компонента 1 в точке x).

Когомологии H_I^n пространства $\mathcal{E}ll_I \times_X \text{Spec } \overline{F}$ отождествляются с когомологиями специального слоя $\mathcal{E}ll_I \times_X \text{Spec } \overline{\kappa(x)}$, так как рассматриваемая схема собственная и гладкая над $\mathcal{O}_{X,x}$. Кроме того, как мы видели выше (п. 4.2.3), соответствия Гекке обладают «хорошей редукцией» в точке x . Поэтому мы можем воспользоваться формулой следа Лефшеца, применив ее к специальному слою над точкой x , чтобы вычислить для функции $f^{x,\infty} \in \mathcal{H}_I^{x,\infty}$ и $r > 0$ след

$$\text{Lef}_r(f^{x,\infty}) \stackrel{\text{def}}{=} \text{tr}(\text{Frob}_x^r \times f^{x,\infty}; H_I^*).$$

Для вычислений необходимо иметь точное описание точек специального слоя, а также действия Фробениуса и соответствий Гекке.

5.2. Описание точек специального слоя происходит в два этапа, подобно тому как описывается множество абелевых многообразий над конечным полем. Сначала мы разбиваем изучаемое множество на классы изогении (следуя Хонде—Тэйту), затем параметризуем каждый класс изогении.

Если $(\mathcal{E}_i, j, t)$ — эллиптический $\mathcal{D}$ -пучок над полем $k = \overline{\kappa(x)}$ (с нулем $z = x : \text{Spec } k \rightarrow X$), то его общий слой V — это слой

расслоения $\mathcal{E}_0$ в общей точке $\text{Spec}(F \otimes_{F_q} k)$ схемы $X \otimes_{F_q} k$, т. е. векторное пространство над полем $F \otimes k$ с действием алгебры D . С другой стороны, морфизм j отождествляет V с общим слоем каждого из расслоений $\mathcal{E}_i$; поэтому морфизм t индуцирует полулинейное отображение $\varphi : V \rightarrow V$. Следуя терминологии Дринфельда, мы говорим, что V есть φ -пространство с действием алгебры D . Два эллиптических D -пучка *изогенны*, если их общие слои изоморфны.

С помощью классификации φ -пространств, полученной Дринфельдом, в статье [L-R-S] доказывается, что классы изогении эллиптических D -пучков находятся во взаимно однозначном соответствии с классами изоморфизма « (D, x, ∞) -типов», т. е. пар $(\tilde{F}, \tilde{\pi})$, где $\tilde{F}$ — конечное сепарабельное расширение поля F и $\tilde{\pi} \in \tilde{F}^* \otimes \mathbb{Q}$, удовлетворяющих следующим условиям:

- Если $\tilde{F}' \subsetneq \tilde{F}$ — подполе, то $\tilde{\pi} \notin (\tilde{F}')^* \otimes \mathbb{Q}$.
- Степень $[\tilde{F} : F]$ делит d .
- Поле $\tilde{F}$ имеет только одну точку $\widetilde{\infty}$, лежащую над ∞ , причем

$$\deg(\widetilde{\infty}) \cdot v_{\infty}(\tilde{\pi}) = -[\tilde{F} : F]/d.$$

- Существует только одна точка $\tilde{x} \neq \widetilde{\infty}$ поля $\tilde{F}$, такая, что $v_{\tilde{x}}(\tilde{\pi}) \neq 0$. Эта точка лежит над точкой x .
- Если точка $\tilde{y}$ поля $\tilde{F}$ лежит над точкой y поля F , то

$$(d[\tilde{F}_{\tilde{y}} : F_y]/[\tilde{F} : F])\text{inv}_y(D) \in \mathbb{Z}$$

(v_{∞} (соотв. $v_{\tilde{x}}$) обозначает нормирование поля $\tilde{F}$, продолжающее нормирование v_{∞} (соотв. v_x)).

Затем дается комбинаторное описание класса изогении (в том числе и действие Фробениуса и соответствий Гекке), соответствующего каждому (D, x, ∞) -типу.

5.3. Следующий шаг заключается в применении формулы следа Лефшеца, т. е., по существу, в перечислении неподвижных точек. Комбинаторные вычисления, основанные на приведенном выше описании точек, приводят к следующему выражению для $\text{Lef}_r(f^{x, \infty})$ ($r > 0$):

$$\begin{aligned} \text{Lef}_r(f^{x, \infty}) &= \sum_{\gamma} \text{vol}(G_{\gamma}(F) \backslash G_{\gamma}(\mathbf{A}) / \varpi_{\infty}^{\mathbb{Z}}) \\ &\times \frac{\varepsilon_{\infty}(\gamma)}{\text{vol}(\bar{D}_{\infty, \gamma}^* / \varpi_{\infty}^{\mathbb{Z}})} \cdot O_{\gamma}(f_x^{(r)} f^{x, \infty}). \end{aligned} \quad (5.3.1)$$

Суммирование производится по классам сопряженности γ группы $G(F) = D^*$, являющимся эллиптическими в точке ∞ . Символ G_γ обозначает централизатор класса γ . По определению $\varepsilon_\infty(\gamma) = (-1)^{d'-1}$, где $d' = d/[F_\infty(\gamma) : F_\infty]$. Далее, $\overline{D}_{\infty, \gamma}^*$ — это тело с центром $F_\infty(\gamma)$ и инвариантом $1/d'$ (так что $\overline{D}_{\infty, \gamma}^*$ является внутренней формой группы $G_\gamma(F_\infty) \cong \text{GL}_{d'}(F_\infty(\gamma))$). Как обычно, если f^∞ — локально постоянная функция на $G(A^\infty)$ с компактным носителем, то $O_\gamma(f^\infty)$ обозначает *орбитальный интеграл*

$$O_\gamma(f^\infty) = \int_{G_\gamma(A^\infty) \backslash G(A^\infty)} f^\infty((h^\infty)^{-1} \gamma h^\infty) \frac{dh^\infty}{dh_\gamma^\infty}$$

(для придания смысла формуле (5.3.1) необходимо согласованным образом выбрать встречающиеся в ней меры Хаара, однако на этом мы не будем останавливаться).

Наконец, $f_x^{(r)} \in C_c^\infty(D_x^* // \mathcal{D}_x^*) \cong C_c^\infty(\text{GL}_d(F_x) // \text{GL}_d(\mathcal{O}_x))$ обозначает локально постоянную функцию с компактным носителем на D_x^* , биинвариантную относительно $\mathcal{D}_x^*$, с *преобразованием Сатаке* (см. [Ca1]), равным

$$(f_x^{(r)})^\vee(z_1, \dots, z_d) = q_x^{r(d-1)/2} (z_1^r + \dots + z_d^r).$$

Эта функция обладает следующим свойством. Если Π_x — допустимое неприводимое представление группы $G(F_x)$, то

$$\text{tr } \Pi_x(f_x^{(r)}) = \begin{cases} 0, & \text{если } \Pi_x \text{ разветвленное,} \\ q_x^{r(d-1)/2} (z_1^r + \dots + z_d^r), & \text{если } \Pi_x \text{ неразветвленное} \\ & \text{с собственными значениями Гекке } z_1, \dots, z_d \end{cases} \quad (5.3.2)$$

(по отношению к мере, относительно которой объем $\mathcal{D}_x^*$ равен 1).

В более ранних работах Дринфельд получил явное выражение для функций $f_x^{(r)}$ и вычислил их орбитальные интегралы. Эти результаты играют ключевую роль в доказательстве формулы (5.3.1).

5.4. Пусть $f = \bigotimes f_v$ — локально постоянная функция с компактным носителем на $G(A)/\varpi_\infty^Z$, являющаяся произведением функций f_v на $G(F_v)$, почти всюду равных характеристической функции $\mathcal{D}_v^*$. *Формула следа Сельберга* выражает след оператора, индуцированного функцией f на пространстве $L^2(G(F) \backslash G(A)/\varpi_\infty^Z)$, т. е. (см.

п. 4.3.1) сумму $\sum m(\Pi) \operatorname{tr} \Pi(f)$ по всем автоморфным представлениям с тривиальным на ϖ_∞ центральным характером. В нашем случае формула имеет простой вид, поскольку фактор компактен:

$$\sum_{\Pi} m(\Pi) \operatorname{tr} \Pi(f) = \sum_{\gamma} \operatorname{vol}(G_{\gamma}(F) \backslash G_{\gamma}(\mathbf{A}) / \varpi_{\infty}^{\mathbf{Z}}) O_{\gamma}(f). \quad (5.4.1)$$

В правой части формулы суммирование ведется по всем классам сопряженности γ группы $G(F)$, а O_{γ} обозначает орбитальный интеграл

$$O_{\gamma}(f) = \int_{G_{\gamma}(\mathbf{A}) \backslash G(\mathbf{A})} f(h^{-1} \gamma h) \frac{dh}{dh_{\gamma}}.$$

5.5. Существует «функция Эйлера–Пуанкаре» f_{∞} — локально постоянная функция на группе $G(F_{\infty})$, имеющая компактный носитель по модулю центра, линейно выражающаяся через характеристические функции паракорических подгрупп [Lau2] и удовлетворяющая следующим условиям:

5.5.1. Орбитальные интегралы

$$O_{\gamma}(f_{\infty}) = \int_{G_{\gamma}(F_{\infty}) \backslash G(F_{\infty})} f_{\infty}(h_{\infty}^{-1} \gamma h_{\infty}) \frac{dh_{\infty}}{dh_{\gamma, \infty}}$$

определяются формулой

$$O_{\gamma}(f_{\infty}) = \begin{cases} 0, & \text{если } \gamma \text{ не эллиптический в } G(F_{\infty}), \\ \frac{\varepsilon_{\infty}(\gamma)}{\operatorname{vol}(\overline{D}_{\infty, \gamma}^* / \varpi_{\infty}^{\mathbf{Z}})}, & \text{если } \gamma \text{ эллиптический в } G(F_{\infty}). \end{cases}$$

5.5.2. Если Π_{∞} — допустимое неприводимое унитаризуемое представление группы $G(F_{\infty}) / \varpi_{\infty}^{\mathbf{Z}}$, то

$$\operatorname{tr} \Pi_{\infty}(f_{\infty}) = \begin{cases} 0, & \text{если } \Pi_{\infty} \not\cong 1, \operatorname{St}_{\infty}, \\ 1, & \text{если } \Pi_{\infty} \cong 1, \\ (-1)^{d-1}, & \text{если } \Pi_{\infty} \cong \operatorname{St}_{\infty} \end{cases}$$

(другими словами, f_{∞} выделяет 1 и $\operatorname{St}_{\infty}$ в унитарном спектре группы $G(F_{\infty}) / \varpi_{\infty}^{\mathbf{Z}}$).

5.6. Подставляя формулы п. 5.5.1 в выражение для $\operatorname{Lef}_{\tau}$ из п. 5.3.1, мы получаем

$$\operatorname{Lef}_{\tau}(f^{x, \infty}) = \sum_{\gamma} \operatorname{vol}(G_{\gamma}(F) \backslash G_{\gamma}(\mathbf{A}) / \varpi_{\infty}^{\mathbf{Z}}) O_{\gamma}(f_x^{(\tau)} f_{\infty} f^{x, \infty}),$$

причем суммирование ведется по всем классам сопряженности γ группы $G(F)$. Теперь эта сумма интерпретируется с помощью формулы Сельберга. С учетом формул п. 5.5.2 мы получаем

$$\begin{aligned} \text{Lef}_r(f^{x,\infty}) \\ = \sum_{\Pi_\infty \cong 1} m(\Pi) \text{tr } \Pi^\infty(f^\infty) + (-1)^{d-1} \sum_{\Pi_\infty \simeq \text{St}_\infty} m(\Pi) \text{tr } \Pi^\infty(f^\infty), \end{aligned}$$

где $f^\infty = f_x^{(r)} f^{x,\infty}$. Теперь воспользуемся формулой (5.3.2), вычисляющей $\text{tr } \Pi_x(f_x^{(r)})$, тем фактом, что если Π — характер, то $m(\Pi) = 1$, а также тем, что если Π_x — неразветвленное представление группы $G(F_x)$, то его собственные значения Гекке равны $\Pi_x(\varpi_x) q_x^{(1-d)/2+i}$, $0 \leq i \leq d-1$. Отсюда получаем окончательно

$$\begin{aligned} \text{Lef}_r(f^{x,\infty}) = \sum_{\substack{\Pi - \text{характер,} \\ \Pi_\infty \cong 1, \Pi_x \text{ неразветвлено}}} \Pi^{x,\infty}(f^{x,\infty}) \Pi_x(\varpi_x)^r (1 + q_x^r + \dots + q_x^{r(d-1)}) \\ + (-1)^{d-1} \sum_{\substack{\Pi_\infty \simeq \text{St}_\infty, \\ \Pi_x \text{ неразветвлено}}} \text{tr } \Pi^{x,\infty}(f^{x,\infty}) m(\Pi) q_x^{r(d-1)/2} (z_1(\Pi_x)^r + \dots + z_d(\Pi_x)^r). \end{aligned} \quad (5.6.1)$$

5.7. Полученная в предыдущем пункте формула показывает, что виртуальное представление $H^* = \sum (-1)^n H^n$ имеет в точности такой след, как предсказывала теорема 4.4.2. Отсюда легко вывести пп. (i) и (ii) этой теоремы. При доказательстве п. (ii) надо воспользоваться «гипотезой» Вейля, из которой следует, что собственные значения $\Pi_x(\varpi_x) q_x^i$ происходят из пространства когомологий степени $2i$. Аналогичные простые соображения позволили бы получить и п. (iii), если бы мы уже располагали гипотезой Рамануджана—Петерсона. Достаточно было бы располагать более слабым ограничением, состоящим в том, что представление Π_x *общее* (т. е. обладает моделью Уиттекера). Последнее, гипотетически, верно всегда (это следует из обобщенного соответствия Жакс—Ленглендса). На самом деле известно [Ja-Sh], что собственные значения Гекке общего неразветвленного унитарного представления удовлетворяют неравенству $|z_i(\Pi_x)| < q_x^{1/2}$.

Однако на данный момент неизвестно, как проверять общность представления Π_x ; поэтому Ломон, Рапопорт и Штулер для доказательства пп. (iii) и (iv) теоремы пользуются более сложными геометрическими соображениями. Эти соображения, применя-

емые для изучения представления Галуа, ограниченного на группу $\text{Gal}(\overline{F}_\infty/F_\infty)$, опираются на сильную теорему Лефшеца, теорему Делиня о чистоте монодромической фильтрации и используют свойства L -функции.

6. ЛОКАЛЬНАЯ ГИПОТЕЗА ЛЕНГЛЕНДСА ДЛЯ ПОЛЕЙ РАВНОЙ ХАРАКТЕРИСТИКИ

6.1. Пусть K — локальное неархимедово поле и $d \geq 1$ — целое число. Обозначим через $A_K(d)$ множество (классов эквивалентности) комплексных допустимых неприводимых представлений группы $\text{GL}_d(K)$ и через $\mathcal{G}_K(d)$ множество (классов эквивалентности) непрерывных комплексных ϕ -полупростых d -мерных представлений группы Вейля–Делиня W'_K . *Локальная гипотеза Ленглендса* (см. [He1]) предсказывает для каждого d существование биекций между $A_K(d)$ и $\mathcal{G}_K(d)$, обладающих многочисленными свойствами (сохранение L - и ε -множителей, определенных для пары представлений в обоих случаях, согласованность с подкруткой и переходом к контраградиентному представлению, ...). Кроме того, эти биекции должны ограничиваться до биекций между $A_K^0(d)$ — множеством неприводимых *каспидальных* представлений с центральным характером конечного порядка — и $\mathcal{G}_K^0(d)$ — множеством *неприводимых* d -мерных представлений группы $\text{Gal}(\overline{K}/K)$.

Обратно, если для всех d имеются согласованные биекции между $A_K^0(d)$ и $\mathcal{G}_K^0(d)$, то их можно каноническим образом продолжить до биекций между $A_K(d)$ и $\mathcal{G}_K(d)$ (см. [He1, He4]).

Локальные гипотезы были доказаны Куцко в случае $d = 2$ и Энньяром в случае $d = 3$. Последний важный результат работы [L-R-S] — это доказательство локальной гипотезы Ленглендса (для всех d), если K — *локальное поле равной характеристики p* (другими словами, $K \cong \mathbb{F}_q((T))$).

6.2. Теорема. Пусть K — *локальное поле равной характеристики p* . Существует семейство биекций

$$A_K^0(d) \rightarrow \mathcal{G}_K^0(d), \quad \pi \mapsto \sigma_\pi,$$

такит, что выполняются следующие условия:

(i) для любых $\pi \in A_K^0(d)$, $\pi' \in A_K^0(d')$

$$L(\sigma_\pi \otimes \sigma'_{\pi'}, s) = L(\pi \times \pi', s), \quad \varepsilon(\sigma_\pi \otimes \sigma'_{\pi'}, s) = \varepsilon(\pi \times \pi', s);$$

(ii) для любого $\pi \in A_K^0(d)$ выполнено $\sigma_\pi = \tilde{\sigma}_\pi$ (контраградиентное представление);

(iii) *детерминант представления σ_π соответствует, согласно теории полей классов (в геометрической нормализации), центральному характеру представления π ;*

(iv) *если $\pi \in A_K^0(d)$, а χ — характер конечного порядка поля K^* , то $(\chi$ обозначает также и соответствующий характер группы $\text{Gal}(\bar{K}/K)$)*

$$\sigma(\pi \otimes \chi) = \sigma_\pi \otimes \chi \quad (\text{согласованность с подкруткой}).$$

В приложении [He3] к статье [L-R-S] Эннъяр доказал, что указанные выше свойства однозначно *характеризуют* полученное семейство биекций. Смотри также [He4], где содержится описание продолжения таких биекций до биекций $A_K(d) \rightarrow \mathcal{G}_K(d)$.

6.3. На самом деле предыдущая теорема доказывается в [L-R-S] не для комплексных, а для ℓ -адических представлений, т.е. представлений над полем $\bar{\mathbb{Q}}_\ell$. Однако эти утверждения равносильны: как в автоморфном случае, так и в случае группы Галуа все рассматриваемые объекты могут быть определены чисто алгебраически; поэтому биекция может быть перенесена на комплексный случай с помощью изоморфизма $\bar{\mathbb{Q}}_\ell \cong \mathbb{C}$, причем L -множители и ε -множители сохраняются (см. [He4]). Изменим соответственно наши обозначения. Зафиксируем простое число $\ell \neq p$. Начиная с этого момента, $A_K^0(d)$ обозначает множество классов эквивалентности неприводимых каспидальных $\bar{\mathbb{Q}}_\ell$ -представлений группы $\text{GL}_d(K)$ с центральным характером конечного порядка, а $\mathcal{G}_K^0(d)$ обозначает множество классов эквивалентности неприводимых d -мерных $\bar{\mathbb{Q}}_\ell$ -представлений группы $\text{Gal}(\bar{K}/K)$ с детерминантом конечного порядка. Теперь мы опишем основные моменты в доказательстве теоремы 6.2.

6.4. Основная идея в построении отображения $\pi \mapsto \sigma_\pi$ — *глобализация* ситуации и применение теоремы 4.4.2. Выберем для этого глобальное поле F характеристики p , а также точку x_0 поля F так, чтобы пополнение F_{x_0} было изоморфно полю K (зафиксируем этот изоморфизм). Выберем затем еще две различные точки ∞ и x_1 , отличные от точки x_0 . Для каждого $d \geq 1$ обозначим через D_d алгебру с делением, d^2 -мерную над центром F , с инвариантом

$$\text{inv}_x(D_d) = \begin{cases} 1/d, & \text{если } x = x_0, \\ -1/d, & \text{если } x = x_1, \\ 0 & \text{в противном случае.} \end{cases}$$

Следующая лемма, доказанная в [L-R-S] с помощью простой формулы следа Сельберга, позволяет рассматривать любое представление $\pi \in \mathcal{A}_K^0(d)$ как компоненту в точке x_0 некоторого автоморфного представления группы $\mathrm{GL}_d(\mathbf{A}_F)$.

6.4.1. Лемма. Пусть $\pi \in \mathcal{A}_K^0(d)$ и $x_2 \notin \{x_0, x_1, \infty\}$ — точка поля F . Тогда существует автоморфное параболическое представление $\Pi = \bigotimes \Pi_v$ группы $\mathrm{GL}_d(\mathbf{A}_F)$, такое, что $\Pi_\infty \cong \mathrm{St}_\infty$, $\Pi_{x_0} \cong \pi$, а компоненты Π_{x_1} и Π_{x_2} каспидальны.

Затем лемма Эннъяра позволяет превратить представление Π в представление группы $G_d(\mathbf{A}_F) = (D_d \otimes \mathbf{A}_F)^*$. Это частный случай обобщенного соответствия Жаке—Ленглендса, которое, конечно же, хотелось бы иметь и в общем случае. Заметим, что в нашем случае $G_d(F_y) \cong \mathrm{GL}_d(F_y)$ для всех $y \notin \{x_0, x_1\}$.

6.4.2. Лемма. Пусть представление Π такое же, как и в предыдущей лемме. Тогда существует (единственное) автоморфное представление $\tilde{\Pi}$ группы $G_d(\mathbf{A}_F)$, такое, что для всех $y \notin \{x_0, x_1\}$ имеем $\tilde{\Pi}_y \cong \Pi_y$. Кроме того, кратность $m(\Pi)$ представления Π равна 1.

6.5. Теперь применяем теорему 4.4.2. Представлению $\tilde{\Pi}$ соответствует d -мерное представление $V_{\tilde{\Pi}_\infty}^{d-1}$ группы $\mathrm{Gal}(\bar{F}/F)$. Обозначим через $\Sigma(\tilde{\Pi})_{x_0}$ ограничение этого представления на группу разложения в точке x_0 , подкрученное на неразветвленный квазихарактер, переводящий Frob_{x_0} в $q_{x_0}^{(1-d)/2}$ (чтобы его детерминант имел конечный порядок). Таким образом, мы получаем искомое представление группы $\mathrm{Gal}(\bar{K}/K)$.

Наша конструкция зависит как от выбора точки x_2 , так и от выбора представления Π . Априори неясно, почему представление $\Sigma(\tilde{\Pi})_{x_0}$ от этих выборов не зависит. Обозначим временно через $\Sigma(\pi)$ множество представлений $\Sigma(\tilde{\Pi})_{x_0}$ (с точностью до эквивалентности), которые могут получиться при всевозможных выборах. Рассуждения, использующие теорию глобальных L -функций (которые, как известно, как в автоморфном случае, так и в случае представлений Галуа удовлетворяют функциональным уравнениям, с помощью которых можно выразить глобальные константы как произведение локальных), позволяют вывести из теоремы 4.4.2 следующую лемму.

6.5.1. Лемма. (i) Пусть $\pi \in \mathcal{A}_K^0(d)$ и $\sigma \in \Sigma(\pi)$. Тогда $\det \sigma$ соответствует (по теории полей классов) центральному характеру представления π .

(ii) Если $\pi \in \mathcal{A}_K^0(d)$ и $\sigma \in \Sigma(\pi)$, то $\check{\sigma} \in \Sigma(\check{\pi})$.

(iii) Если $\pi \in \mathcal{A}_K^0(d)$, $\sigma \in \Sigma(\pi)$ и χ — характер конечного порядка поля K^* , то $\sigma \otimes \chi \in \Sigma(\pi \otimes \chi)$.

(iv) Если $\pi \in \mathcal{A}_K^0(d)$, $\pi' \in \mathcal{A}_K^0(d')$, $\sigma \in \Sigma(\pi)$ и $\sigma' \in \Sigma(\pi')$, то

$$\begin{cases} L(\sigma \otimes \sigma', s) &= L(\pi \times \pi', s), \\ \varepsilon(\sigma \otimes \sigma', s) &= \varepsilon(\pi \times \pi', s). \end{cases}$$

6.6. Следующая лемма доказывает, что множество $\Sigma(\pi)$ состоит из одного элемента σ_π .

Лемма. Пусть $\pi \rightarrow \Sigma(\pi)$ — отображение, которое каждому $\pi \in \mathcal{A}_K^0(d)$ сопоставляет непустое множество $\Sigma(\pi)$ классов эквивалентности d -мерных представлений группы Галуа $\text{Gal}(\bar{K}/K)$ с детерминантом конечного порядка. Предположим, что выполняются следующие условия:

(i) $\forall \pi, \pi' \in \mathcal{A}_K^0(d), \forall \sigma \in \Sigma(\pi), \forall \sigma' \in \Sigma(\pi'): L(\sigma \otimes \sigma', s) = L(\pi \times \pi', s)$;

(ii) $\sigma \in \Sigma(\pi) \implies \check{\sigma} \in \Sigma(\check{\pi})$.

Тогда для любого $\pi \in \mathcal{A}_K^0(d)$ множество $\Sigma(\pi)$ состоит из одного элемента $\sigma_\pi \in \mathcal{G}_K^0(d)$. Более того, отображение $\pi \mapsto \sigma_\pi$ инъективно.

6.7. Таким образом, мы уже получили вложения $\pi \mapsto \sigma_\pi$, удовлетворяющие всем условиям, изложенным в теореме 6.2. Теперь нам осталось лишь воспользоваться локальной численной гипотезой Ленглендса, доказанной Эннъяром.

Теорема [He2]. Если инъективное отображение $\mathcal{A}_K^0(d) \rightarrow \mathcal{G}_K^0(d)$ сохраняет кондукторы и согласовано с подкруткой на неразветвленные характеры, то оно является биекцией.

ЛИТЕРАТУРА

- [An] Anderson G. t -motives. Duke Math. J., **53** (1986), 457–502.
 [Bo-Ja] Borel A., Jacquet H. Automorphic forms and automorphic representations, in: Automorphic forms, representations and L -functions. Proc. Symp. Pure Math., Vol. 33, Part 1, Amer.

Math. Soc., Providence, 1979, 189–202. [Имеется перевод: Борель А., Жакс Э. Автоморфные формы и автоморфные представления. — В кн.: Автоморфные формы, представления и L -функции. — М.: Мир, 1984.]

- [Ca1] Cartier P. Representations of p -adic groups: a survey. In: Automorphic Forms, Representations and L -functions. Proc. Symp. Pure Math., Vol. 33, Part 1, Amer. Math. Soc., Providence, 1979, 111–156.
- [Ca2] Cartier P. La conjecture locale de Langlands pour $GL(2)$ et la démonstration de Ph. Kutzko, Sémin. Bourbaki, Fév. 1980, exp. 550, Lecture Notes in Math., vol. 842, Springer-Verlag, 1981, 112–138.
- [Cl] Clozel L. Motifs et formes automorphes: applications du principe de fonctorialité, in: Automorphic Forms, Shimura Varieties and L -functions. Proceedings of the Ann Arbor Conference, T. 1, Academic Press, 1990, 77–159.
- [De] Deligne P. Les constantes des équations fonctionnelles des fonctions L , Modular functions of one variable II, Lecture Notes in Math., vol. 349, Springer-Verlag, 1973, 501–597.
- [De-Hu] Deligne P., Husemöller D. Survey of Drinfeld modules, in: Current Trends in Arithmetical Algebraic Geometry, Contemp. Math., vol. 67, Amer. Math. Soc., 1987, 25–91.
- [De-Ka-Vi] Deligne P., Kazhdan D., Vignéras M.-F. Représentations des algèbres centrales simples p -adiques, in: Représentations des groupes réductifs sur un corps local (Bernstein, Deligne, Kazhdan, Vignéras), Hermann, 1984, 33–118.
- [Dr1] Дринфельд В. Г. Эллиптические модули. — Матем. сб., т. 94, 1974, вып. 4, с. 594–627.
- [Dr2] Дринфельд В. Г. Эллиптические модули II. — Матем. сб., т. 102, 1977, вып. 2, с. 182–194.
- [Dr3] Дринфельд В. Г. О коммутативных подкольцах некоторых некоммутативных колец. — Функц. анализ и прил., т. 11, 1977, вып. 1, с. 11–14.
- [Dr4] Дринфельд В. Г. Многообразия модулей F -пучков. — Функц. анализ и прил., т. 21, 1987, вып. 2, с. 23–41.
- [Dr5] Дринфельд В. Г. Доказательство гипотезы Петерсона для $GL(2)$ над глобальным полем характеристики p . — Функц. анализ и прил., т. 22, 1988, вып. 1, с. 34–54.
- [Dr6] Drinfeld V. G. Cohomology of compactified manifolds of modules of F -sheaves of rank 2, J. of Soviet Math., 46, (1989), No. 1, 1789–1821.
- [Ha-Na] Harder G., Narasimhan M. S. On the cohomology of moduli spaces of vector bundles on curves, Math. Ann., 212 (1975), 215–248.
- [He1] Henniart G. Le point sur la conjecture de Langlands pour $GL(n)$ sur un corps local, Séminaire de théorie des nombres, Paris 1983–1984. Prog. in Math., Birkhäuser, 1985, 115–131.

- [He2] Henniart G. La conjecture de Langlands locale numérique pour $GL(n)$, Ann. Sci. Éc. Norm. Sup. (IV), **21** (1988), 145–203.
- [He3] Henniart G. Caractérisation de la correspondance de Langlands locale par les facteurs ε de paires. Invent. Math., **113** (1993), No. 2, 339–350.
- [He4] Henniart G. La correspondance de Langlands locale: caractérisation et propriétés fonctorielles, готовится к печати.
- [Ja-Sh] Jacquet H., Shalika J. On Euler products and the classification of automorphic representations, I, Amer. J. Math., **103** (1981), 499–558, II, Amer. J. Math., **103** (1981), 777–815,
- [Ko] Kottwitz R. On the λ -adic representations associated to some simple Shimura varieties, Preprint, 1989.
- [Kr] Кричевер И. М. Алгебро-геометрическое построение уравнений Захарова—Шабата и их периодических решений. — ДАН СССР, т. 227, 1976, № 2, с. 291–294.
- [La] Langlands R. P. Automorphic representations, Shimura varieties, and motives. Ein Märchen, in: Automorphic forms, Representations and L -functions. Proc. Symp. Pure Math., vol. 33, Tome 2, Amer. Math. Soc., Providence, 1979, 205–246.
- [Lau1] Laumon G. Transformation de Fourier, constantes d'équations fonctionnelles et conjecture de Weil, Publ. Math. I.H.E.S., **65** (1987), 131–210.
- [Lau2] Laumon G. Cohomology with compact support of Drinfeld modular varieties, part I, II, prepublications de l'Université de Paris-Sud, 91–01, 92–12.
- [L-R-S] Laumon G., Rapoport M., Stuhler U. $\mathcal{D}$ -elliptic sheaves and the Langlands correspondence, Invent. Math., **113** (1993), No. 2, 217–338.
- [Mu] Mumford D. An algebro-geometric construction of commuting operators and of solutions to the Toda lattice equations, Korteweg–de Vries equation and related non-linear equations, Internat. Symp. on Algebraic geometry (Kyoto 1977), Kinokuniya, Tokyo, 1977, 115–153.
- [Ro] Rogawski J. Representations of $GL(n)$ and division algebras over a p -adic field, Duke Math. J., **50** (1983), 161–196.

HENRI CARAYOL

Institut de Recherche Mathématique Awaucée
 Université Louis Pasteur et C.N.R.S.
 7, rue René-Descartes
 67084 Strasbourg Cedex
 E-mail: carayol@math.u-strasbg.fr

НЕДАВНИЕ РЕЗУЛЬТАТЫ, ОТНОСЯЩИЕСЯ К ДВУМЕРНОМУ ПОТОКУ ИДЕАЛЬНОЙ НЕСЖИМАЕМОЙ ЖИДКОСТИ (по Ж.-И. Шемену и Ж.-М. Делору)

Патрик Жерар

Уравнение Эйлера для потока идеальной несжимаемой жидкости является одним из важнейших уравнений в истории развития теории нелинейных уравнений математической физики; изучению его решений посвящено множество исследований, особенно в последние тридцать лет, параллельно с развитием методов нелинейного анализа. Недавно Шемен и Делор доказали два результата, относящиеся к сингулярным решениям этого уравнения в двумерном случае. В первой части мы напомним математический аппарат и попытаемся изложить эти два результата в русле общего подхода к задаче Коши для уравнения Эйлера в двумерном случае. Во второй части мы дадим доказательства двух важных классических результатов, принадлежащих соответственно Волибнеру и Юдовичу. Затем мы более подробно изложим работы Шемена и Делора.

Я благодарен Ж.-Б. Босту, Ж.-И. Шемену, Ж.-М. Делору и К. Сен-Раймону за помощь в подготовке этой статьи.

0.1. Предварительные результаты. Пусть (M, g) — компактное связное ориентированное риманово многообразие размерности d^1 . В дальнейшем будут использоваться следующие обозначения дифференциальной геометрии:

Если X — векторное поле на многообразии M , то через L_X обозначается производная Ли вдоль X , а ∇_X — ковариантная производная вдоль X . Обозначим через τ положительную форму объема на M . Дивергенция поля X — это функция $\operatorname{div} X$, определяемая формулой

$$L_X \tau = (\operatorname{div} X) \tau.$$

¹⁾ Большинство обсуждаемых здесь результатов было получено в пространстве $\mathbb{R}^d$, а не на компактном многообразии. Мы предпочитаем последний случай, так как он не требует тонкого анализа условий на бесконечности и позволяет лучше продемонстрировать инвариантность используемой техники.

Норма поля X обозначается через $|X| = g(X, X)^{1/2}$. Ковариантная производная индуцирует дифференциальный оператор ∇ порядка 1 на векторных полях со значениями в сечениях расслоения $\text{End}(TM)$ или расслоений тензорных полей. Следовательно, $\text{div } X = \text{tr}(\nabla X)$.

Обозначим через $u \mapsto \tilde{u}$ изоморфизм пространства TM на T^*M , индуцированный метрикой g . Градиент ∇f функции f является векторным полем, соответствующим дифференциалу df при этом изоморфизме, а лапласиан функции f определяется равенством $\Delta f = \text{div } \nabla f$. Дивергенция тензорного поля T является векторным полем $\text{div } T$, соответствующим полю $-{}^t\nabla T$ при этом изоморфизме. Здесь символом ${}^t\nabla$ обозначен транспонированный оператор к дифференциальному оператору ∇ по отношению к плотности τ . Если u — бездивергентное векторное поле, то $\text{div}(u \otimes \tilde{u}) = \nabla_u u$.

Наконец, обозначим через $*$ оператор Ходжа на формах.

Если M имеет размерность 2, то оператор J определяется на векторных полях формулой

$$(JX)^{\sim} = * \tilde{X}.$$

Тогда $J^2 = -1$. Ротором (или вихрем) векторного поля X называется функция $\text{rot } X$, определяемая формулой $d\tilde{X} = (\text{rot } X)\tau$. Кроме того, $\text{rot } u = -\text{tr}(J\nabla X)$. Наконец, если u — бездивергентное векторное поле, то $\text{rot } \nabla_u u = L_u \text{rot } u$.

0.2. Уравнение Эйлера. В 1755 г. для описания движения идеальной несжимаемой жидкости на многообразии M Эйлер вывел следующую систему уравнений. Если обозначить через $u(t, x)$ скорость в момент времени $t \in \mathbf{R}$ частицы жидкости, находящейся в положении $x \in M$, то фундаментальный принцип динамики и условие несжимаемости запишутся соответственно в виде

$$\partial_t u + \nabla_u u = -\nabla p, \quad (E_1)$$

$$\text{div } u = 0, \quad (E_2)$$

где p — скалярное поле давления. В дальнейшем мы будем обозначать через (E) указанную выше систему уравнений. Заметим, что градиент давления в постановке задачи не задан, а является дополнительным неизвестным и может быть исключен, если взять дивергенцию от уравнения (E_1) :

$$-\Delta p = \text{div } \nabla_u u;$$

это уравнение определяет ∇p как функцию векторного поля u .

Можно также рассматривать систему (E) как систему уравнений геодезических на бесконечномерной группе Ли диффеоморфизмов многообразия M , сохраняющих объем, с метрикой, определенной кинетической энергией

$$E_c = \int_M |u(t, x)|^2 \tau(x).$$

В этом случае p интерпретируется как множитель Лагранжа. Мы не будем развивать здесь этот подход, отсылая читателя к работам Арнольда [Ar], Эбена—Марсдена [EB], Шнирельмана [Sh] и Бреньера [Br] для детального изучения. Однако позаимствуем из этого подхода тот факт, что если u — решение системы (E) , то кинетическая энергия не зависит от времени, что, впрочем, легко получить, взяв скалярное произведение в L^2 двух обеих частей уравнения (E_1) на u и воспользовавшись (E_2) .

Фундаментальной проблемой гидромеханики, разумеется, является задача Коши для уравнений (E) , т. е. нахождение функции u , удовлетворяющей уравнениям (E) и условию $u(0, x) = u^0(x)$, где u^0 — заданное бездивергентное векторное поле на многообразии M . Единственность и существование на малом интервале времени, по-видимому, впервые были доказаны Лихтенштейном [Li] в 1925 г. Существование такого глобального решения в размерности n остается открытой проблемой, но оно было доказано в размерности 2 Волибнером [W] в 1933 г. (см. также статью Като [Ka]). Тем не менее результат Волибнера мало освещает качественные свойства решения; полученные оценки позволяют предположить, что поведение линий тока становится при больших значениях t совершенно хаотическим (см. разд. 1).

В настоящее время для изучения качественных свойств решений этой задачи делаются попытки решить эту проблему с сингулярными данными и описать эволюцию особенностей. Эти сингулярные решения (например разрывное поле скоростей) являются, вообще говоря, «предельными» моделями физических явлений и указывают на наличие очень разномасштабных величин (например, эволюцию границы раздела двух сред). Два результата, которые являются предметом данного сообщения, прямо связаны с этими проблемами для двумерного потока. Прежде чем их представлять, нам необходимо ввести фундаментальную величину гидромеханики — вихрь.

0.3. Вихрь. В последующем изложении многообразии M предполагается двумерным. Вихрь ω поля скоростей u определяется формулой

$$\omega = \text{rot } u.$$

Заметим, что $\int_M \omega(x) \tau(x) = 0$. Помимо его очевидной физической интерпретации, это понятие интересно тем, что допускает формулировку, эквивалентную системе (E) . А именно, применив оператор rot к (E_1) , получаем уравнение переноса

$$\partial_t \omega + L_u \omega = 0, \quad (T)$$

показывающее, что ω постоянен вдоль линий тока. Кроме того, возможно (почти) выразить u через ω . Для этого используется тот факт, что лапласиан является автоморфизмом векторного пространства функций класса C^∞ со средним значением нуль (см., например, [R]); обозначим через Δ^{-1} обратный ему. Пусть, кроме того, через $\mathcal{H}$ обозначается конечномерное векторное пространство гармонических векторных полей на M и через H ортогональный проектор на $\mathcal{H}$ в пространстве векторных полей на M относительно скалярного произведения в L^2 , ассоциированного с τ . Учитывая (E_2) , получаем, что разложение Ходжа—де Рама [R] поля u дает «соотношение Био—Савара»

$$u = B\omega + h, \quad (B)$$

где $B = J\nabla\Delta^{-1}$ — ограничение на функции с нулевым средним соответствующего псевдодифференциального оператора порядка -1 и $h = H(u)$. Наконец, эволюция h в пространстве $\mathcal{H}$ описывается уравнением, которое получается применением H к уравнению (E_1) ,

$$\partial_t h + Q(u, u) = 0, \quad (H)$$

где

$$Q(v, w) = - \sum_j \left(\int_M g(\nabla_v X_j, w) \tau \right) X_j$$

и (X_j) — ортонормированный базис в $\mathcal{H}$. Поскольку h принадлежит конечномерному пространству и его L^2 -норма мажорируется нормой u и, следовательно, ограничена по времени, можно предсказать, что эволюция h не влияет на качественные свойства решения u . Заметим, что в двух простых случаях эта эволюция тривиальна: если M — сфера, то $\mathcal{H} = \{0\}$, и если M — плоский тор, то h не зависит от времени.

Легко проверить, что система (T, B, H) эквивалентна системе (E) .

0.4. Вихревые карманы. Перейдем к сингулярным решениям системы (E) . Здесь основным является результат Юдовича [Y], полученный им в 1963 г. и относящийся к течению с разрывным вихрем. Более точно, Юдович доказал, что если начальные условия имеют существенно ограниченный вихрь, то существует единственное решение задачи Коши с существенно ограниченным вихрем в любой момент времени — уравнения при этом должны рассматриваться в слабом смысле. Чтобы пояснить этот результат, заметим, что соотношения Био—Савара (B) не позволяют заключить, что поле скоростей липшицево; действительно, псевдодифференциальный оператор порядка -1 переводит функции из L^∞ не в функции, удовлетворяющие условию Липшица, а только в несколько менее регулярные функции. Юдович, однако, доказал, что поле такой регулярности порождает поток и что вихрь в момент времени t получается из начального вихря под действием этого потока в соответствии с уравнением (T) . К сожалению, регулярность этого потока экспоненциально убывает со временем, что затрудняет анализ эволюции особенностей (см. разд. 2).

В прекрасной обзорной статье, посвященной завихрению в гидромеханике [M1], Майда в 1985 г. поставил следующую проблему, названную им проблемой «вихревых карманов». Предположим, что вихрь в начальный момент принимает только два значения и множество разрывов является жордановой кривой γ^0 класса C^∞ . Согласно результату Юдовича, вихрь в момент t тоже принимает только два значения и множество разрывов γ^t получается из γ^0 под действием указанного выше слабо регулярного потока. Что можно сказать о регулярности γ^t ?

В той же статье, после описания интегро-дифференциальных уравнений, которым должна удовлетворять регулярная параметризация кривой γ^t , Майда анонсировал утверждение, что для малых промежутков времени контур остается регулярным; более того, опираясь на численное моделирование [Z], он предположил появление каспов и нерегулярность кривой γ^t на конечном промежутке времени.

После этого появилось множество работ, в которых делается попытка обосновать это предположение; помимо численных методов [D, D1], отметим подход, предложенный Константином и Тити

[СТ], которые изучали возмущения стационарного вихревого кармана (например, плоского диска). В этом контексте Алиак [А1] доказал, что квадратичное приближение уравнения вихревого кармана достаточно для выявления особенностей.

Все эти вклады подчеркивают особое значение следующего результата, доказанного Шеменом в 1991 г.:

Теорема А (Шемен [С4]). *Контур вихревого кармана остается в классе C^∞ для любого t .*

Доказательство Шемена основано на двух основных идеях.

Первая состоит в замечании, что если обозначить через X^0 ненулевое векторное поле класса C^∞ на M , касательное к кривой γ^0 , и если X^t — сингулярное поле, получающееся из X^0 под действием потока жидкости в момент времени t , то регулярность γ^t вытекает из того факта, что, хотя X^t не принадлежит классу C^∞ , оно и все его ковариантные производные $\nabla_{X^t}^k X^t$ принадлежат одному и тому же гёльдерову пространству C^α . Этот тип регулярности, называемый «конормальным» или «вторично микролокальным», является в последние десятилетия центральным в работах Бони и его школы по распространению особенностей в гиперболических нелинейных уравнениях. Для того чтобы познакомить читателя с этими результатами, мы отсылаем его к обзорной статье Бони [В2] и к докладу Лебо [L] на этом семинаре. Применяя методы, используемые в этих работах, и собственные достижения [С5], Шемен доказал в 1989 г. [С1, 3] локальную регулярность во времени контура вихревого кармана в том виде, как она была анонсирована Майдой. Заметим, что теми же методами Шемен доказал следующий замечательный результат, который относится к уравнению Эйлера в произвольной размерности: «*Всякое решение класса C^r ($r > 1$) уравнения Эйлера на $[0, T] \times M$ обладает линиями тока класса C^∞* » [С2].

Вторая идея является решающим шагом в направлении глобальной регулярности во времени и является предметом статьи [С4]. Она состоит в оценке нормы Липшица поля скоростей через функцию от нормы вихря ω в L^∞ и «конормальных» величин, измеряющих регулярность геометрии, причем главное здесь то, что эти величины появляются как аргумент логарифма (см. разд. 3, предложение 8).

Заметим, что эти оценки являются более общими и что результат Шемена распространяется на гораздо более сложную геометрическую ситуацию, чем вихревой карман (см. разд. 3).

Заметим, что работы Шемена, упомянутые выше, послужили предметом для многочисленных улучшений и переформулировок: так, локальная регулярность вихревых карманов была независимо доказана Серфати [S], который работал непосредственно с интегро-дифференциальными уравнениями контура и доказал, кроме того, аналитическую регулярность этого контура по времени. Прямой подход к локальной регулярности был также развит недавно Бертоцци [Be] в его диссертации. Наконец, Константин и Бертоцци [BC] дали несколько менее длинное доказательство теоремы А, тоже основанное на логарифмической оценке, но существенно зависящее от конкретной геометрической структуры вихревого кармана.

В заключение введения в эту проблему заметим, что оценки геометрии контура, полученные Шеменом, являются дважды экспоненциальными; было бы интересно узнать, являются ли они оптимальными; если это так, то это могло бы объяснить предсказание появления особенностей, вытекавшее из численного моделирования. Другая открытая проблема — эволюция изначально сингулярного контура (с углами, каспами и т. д.).

0.5. Вихревые поверхности. Перейдем теперь к еще более сингулярным решениям. Предположим, что задано начальное векторное поле скоростей u^0 , разрывное вдоль регулярной кривой на M ; тогда вихрь является мерой Радона, сингулярная часть которой является плотностью на этой кривой. Возможно ли решить задачу Коши для системы (E) в этом случае (заметим, что теорема Юдовича уже неприменима)?

Еще в большей степени, чем в предыдущих случаях, эта задача моделирует большое число физических ситуаций, например течение воздуха вблизи крыла самолета, и явилась источником обширной физической литературы и литературы по численным методам; для более детального ознакомления мы отсылаем к статье Сафмана и Бейкера [SE]. Отметим, что речь идет об очень неустойчивой задаче, что было замечено Гельмгольцем [H], Кельвином [K] и Тейлором [T]. В отличие от случая вихревых карманов, упомянутое интегро-дифференциальное уравнение, описывающее эволюцию границы раздела сред, некорректно определено на малых промежутках времени в том смысле, что ассоциированная *линеаризованная* задача не имеет решения в классе C^∞ (см. работу Биркгофа [Bi]). Однако для этой задачи имеется результат о локальном существовании в случае, когда граница раздела и ограничение u^0 на ее дополнение аналитичны, причем методы базируются на абстракт-

ных формах теоремы Коши—Ковалевской (см. [BF, DR, CO]). Но численное моделирование (см., в частности, [Kr]) позволяет предположить гораздо более сложное поведение на конечном отрезке времени.

Более общо, можно поставить следующий вопрос: существует ли решение задачи Коши при заданном начальном поле скоростей в L^2 , вихрь которого является ограниченной мерой? Методы отыскания слабых решений нелинейных уравнений в частных производных подсказывают следующую стратегию: регуляризуя данные Коши и используя теорему Волибнера, получаем последовательность регулярных решений (u_n) с ограниченной кинетической энергией, а также L^1 -нормами соответствующих вихрей. Некоторая подпоследовательность $(u_{n'})$ будет слабо сходиться к полю скоростей u , интегрируемому с квадратом, вихрь которого является мерой. Проблема состоит в том, чтобы выяснить, является ли u решением уравнения Эйлера

$$\partial_t u + \operatorname{div}(u \otimes \bar{u}) = -\nabla p, \quad \operatorname{div} u = 0, \quad (E')$$

где уравнение (E_1) записано в консервативной форме для того, чтобы оно имело смысл, когда u — только функция, интегрируемая с квадратом. Трудность заключается в том, что неизвестно, является ли сходимостью $u_{n'}$ к u сильной и можно ли, следовательно, перейти к пределу в члене $(u'_{n'} \otimes \bar{u}'_{n'})$. Диперна и Майда в трех статьях [DM1, 2, 3] первыми приступили к этой проблеме. Они доказали, что последовательность сильно сходится вне некоторого множества хаусдорфовой размерности, меньшей или равной 1; несколько позже Грингард и Томанн [GT] доказали, что если эта размерность < 1 , то сходимостью сильная. Диперна и Майда также предполагали возможность явления «рассеянной концентрации», когда, хотя сходимостью и не сильная, u все-таки является решением уравнения Эйлера; Алинанк [A2] показал, что это явление имеет место, если точки накопления последовательности $(|u_{n'} - u|^2)$ в слабой топологии концентрируются на «достаточно малых» множествах.

Только в 1990 г. был дан положительный ответ на этот вопрос в важном частном случае:

Теорема В (Делор [B1, 2]). Пусть (u_n) — последовательность регулярных решений системы (E) , последовательность (u_n^0) начальных данных которой сходится в среднеквадратичном к (u_n) ; предположим, что $\omega^0 = \operatorname{rot} u^0$ — мера Радона, сингулярная часть которой положительна (или отрицательна), и что последователь-

ность ($\text{rot } u_n^0$) сходится к ω_0 в слабой топологии. Тогда каждая точка накопления последовательности (u_n) является решением системы (E') . В частности, задача Коши для (E') с начальными данными (u_0) имеет решение.

Заметим, что из теоремы Делора в действительности следует и устойчивость относительно перехода к слабому пределу для решений системы (E') , что позволяет найти решения с конечной энергией, вихрь которых является в каждый момент времени мерой Радона с положительной (или отрицательной) сингулярной частью. Однако единственность подобных решений не выяснена; вопрос о сохранении со временем кинетической энергии остается также открытым — имеется только неравенство, равенство было бы эквивалентно сильной сходимости последовательностей решений, введенных выше.

Доказательство теоремы В базируется на двух соображениях. Прежде всего, исключив функцию давления, Делор вывел эквивалентную форму для (E) , которая позволила ему заключить, что достаточно перейти к пределу в *некоторых* квадратичных выражениях от u_n . Этот последний факт ранее отмечался цитируемыми выше авторами в связи с «рассеянной концентрацией». Отсюда получается, что предположение о положительности для сингулярной части меры ω^0 дает замечательное явление компенсации в псевдодифференциальных квадратичных выражениях от $\omega_n = \text{rot } u_n$, что делает возможным эффективный переход к пределу (см. разд. 4). Отметим, наконец, что это явление компенсации недавно было интерпретировано Эвансом и Мюллером [ЕМ] в терминах пространства Харди в духе результатов Куафмана—Лионса—Мейера—Семеса [CL].

0.6. Об используемых функциональных пространствах. Если E — векторное расслоение над M и $k \in \mathbb{N} \cup \{\infty\}$, то обозначим через $C^k(M, E)$ пространство сечений класса C^k расслоения E . Для каждого расслоения E , используемого в дальнейшем, и для всех $k \neq \infty$ выбирается норма на $C^k(M, E)$, которая определяет топологию C^k -сходимости; обозначим эту норму через $[\]_k$.

Если $\alpha \in]0, 1[$, то обозначим через $C^\alpha(\mathbb{R}^n)$ пространство ограниченных функций f на $\mathbb{R}^n$, таких, что существует $C > 0$, при котором выполняется неравенство

$$|f(x) - f(y)| \leq C|x - y|^\alpha.$$

Если α — положительное действительное нецелое число, то $C^\alpha(\mathbf{R}^n)$ означает пространство ограниченных функций класса $C^{[\alpha]}$, производные которых порядка $[\alpha]$ принадлежат классу $C^{\alpha-[\alpha]}(\mathbf{R}^n)$. Каждое из этих пространств обладает естественной структурой банахова пространства.

Если $\alpha - m$ — положительное нецелое число, то псевдодифференциальный оператор порядка m из $C^\alpha(\mathbf{R}^n)$ в $C^{\alpha-m}(\mathbf{R}^n)$ ограничен. Далее определяется гёльдерово пространство для других значений α так, чтобы это свойство оставалось верным для всех действительных значений α и m . Эти пространства обозначают также $C^\alpha(\mathbf{R}^n)$, кроме случая, когда $\alpha = k$ — натуральное число, в котором предпочитают обозначение $C_*^k(\mathbf{R}^n)$, чтобы отличать это пространство от пространства ограниченных функций класса C^k с ограниченными производными, которое оно строго содержит. Эти пространства наделены естественной структурой банахова пространства и через $\|\cdot\|_\alpha$ обозначается норма, определяющая эту структуру. Например, $C_*^1(\mathbf{R}^n)$ — «класс Зигмунда», т. е. пространство ограниченных функций, удовлетворяющих условию

$$|f(x+h) + f(x-h) - 2f(x)| \leq C|h|.$$

Гёльдеровы пространства, определенные таким образом, легко описать с использованием диадического разложения в пространстве Фурье следующим образом (см., например, [СМ]). Пусть $\varphi = \varphi(\xi)$ — функция класса C^∞ на $\mathbf{R}^n$ с компактным носителем в $\mathbf{R}^n \setminus \{0\}$, такая, что носитель функции

$$\chi(\xi) = 1 - \sum_{p=0}^{\infty} \varphi(2^{-p}\xi),$$

лежит в компакте пространства $\mathbf{R}^n$. Если f — обобщенная функция умеренного роста на $\mathbf{R}^n$, то положим для $p \geq 0$

$$\Delta_p f(x) = \varphi(2^{-p}D)f(x), \quad S_p f(x) = \chi(2^{-p}D)f(x)$$

и $\Delta_{-1}f(x) = \chi(D)f(x)$, так что

$$S_p f = \sum_{-1 \leq q \leq p-1} \Delta_q f, \quad f = \sum_{-1 \leq q} \Delta_q f.$$

Заметим, что $S_p f$ является функцией класса C^∞ и что, обозначив через $[\cdot]_k$ C^k -норму на $\mathbf{R}^n$, получаем

$$[S_p f]_k \leq C_k 2^{kp} [S_p f]_0,$$

где C_k зависит только от k . Таким образом, обобщенная функция f умеренного роста принадлежит $C^\alpha(\mathbb{R}^n)$ для $\alpha \in \mathbb{R} \setminus \mathbb{N}$ (соответственно $C_*^\alpha(\mathbb{R}^n)$ для $\alpha \in \mathbb{N}$) тогда и только тогда, когда существует константа C , такая, что для всех $p \geq -1$

$$[\Delta_p f]_p \leq C 2^{-p\alpha},$$

причем наилучшая из констант C определяет норму на гёльдеровом пространстве порядка α и индуцирует на нем структуру банахова пространства.

Так определенные гёльдеровы пространства выдерживают умножение на функцию из класса C^∞ с компактным носителем и локально инвариантны относительно диффеоморфизмов. Если E — векторное расслоение над компактным многообразием M , то можно определить пространство гёльдеровых сечений порядка α расслоения E над M , которое обозначается $C^\alpha(M, E)$, если α не является натуральным числом, и $C_*^\alpha(M, E)$ в противном случае; обозначим через $\|\cdot\|_\alpha$ норму, определяющую естественную топологию этого пространства.

Если Ω открыто в M и f — гёльдерова функция порядка α на Ω , то положим

$$\|f\|_{\alpha, \Omega} = \inf\{\|g\|_\alpha, \quad g \in C^\alpha(M), \quad g|_\Omega = f\}.$$

Обозначим, наконец, L^2 -пространство сечений расслоения E над M через $L^2(M, E)$.

1. РЕГУЛЯРНЫЕ РЕШЕНИЯ

Теорема 1 (Волибнер [W], 1933; Като [Ka], 1967). Пусть u^0 — бездивергентное векторное поле класса C^∞ на M . Существует единственное решение $u \in C^\infty(\mathbb{R} \times M, TM)$ уравнения (E), такое, что $u(0, x) = u^0(x)$.

Доказательство. Оно проводится в три этапа: сначала доказывается существование и единственность решения на малом интервале времени; затем доказывается, что если решение невозможно продолжить за момент времени $T > 0$, то

$$\lim_{t \rightarrow T^-} \sup [u(t)]_1 = +\infty.$$

В завершение доказательства мы находим оценку для $[u(t)]_1$ на любом ограниченном интервале времени. Если два первых этапа обобщаются на любую гиперболическую нелинейную систему порядка 1,

то третий использует частный вид уравнения (E) и двумерность M ; напомним, что аналогичная проблема в размерности 3 остается открытой.

1.1. Существование и единственность на малом интервале времени. Будем рассматривать случай $t \geq 0$; случай $t \leq 0$ разбирается аналогично. Если u и v — регулярные решения системы (E) на интервале времени $[0, T]$ с одинаковыми начальными условиями и давлениями соответственно p и q , то

$$\partial_t(u - v) + \nabla_u(u - v) = -\nabla(p - q) + \nabla_{v-u}v.$$

Приравнявая скалярные произведения в $L^2(M, TM)$ обеих частей этого уравнения на $u - v$ и используя равенство $\operatorname{div} u = \operatorname{div} v = 0$, получаем

$$\frac{1}{2} \frac{d}{dt} \|u(t) - v(t)\|_{L^2}^2 = (\nabla_{v(t)-u(t)}v(t), u(t) - v(t)).$$

Поскольку $[v(t)]_1 \leq C$,

$$\frac{d}{dt} \|u(t) - v(t)\|_{L^2}^2 \leq C \|u(t) - v(t)\|_{L^2}^2,$$

и лемма Гронуолла позволяет заключить, что $u = v$.

Перейдем к локальному существованию. Будем использовать формулировку уравнения (E) в виде системы (T), (H), (B); определим рекуррентно три последовательности (u_n) , (ω_n) , (h_n) следующим образом: $u_0(t, x) = u^0(x)$, $\omega_0 = \operatorname{rot} u_0$, $h_0 = H(u_0)$ и для всех $n \geq 0$

$$\begin{aligned} \partial_t \omega_{n+1} + L_{u_n} \omega_{n+1} &= 0, & \omega_{n+1}(0) &= \operatorname{rot} u^0, \\ \partial_t h_{n+1} + Q(u_n, B\omega_{n+1} + h_{n+1}) &= 0, & h_{n+1}(0) &= H(u^0), \\ u_{n+1} &= B\omega_{n+1} + h_{n+1}. \end{aligned}$$

Заметим, что первое уравнение является уравнением линейного переноса, второе — система линейных дифференциальных уравнений в конечномерном пространстве $\mathcal{H}$ гармонических полей; кроме того, $Q(h_n, h_{n+1}) = 0$, поскольку h_n и h_{n+1} гармонические. Из уравнения переноса выводим равенство

$$[\omega_n(t)]_0 = [\operatorname{rot} u^0]_0$$

для всех t и n и, используя его и второе уравнение, получаем, что $h_n(t)$ ограничена в конечномерном пространстве $\mathcal{H}$. Легко также доказать, что достаточно равномерной ограниченности $[\omega_n(t)]_1$

на $[0, T]$, чтобы последовательность (ω_n) (соотв. (h_n)) была последовательностью Коши в пространстве $C^0([0, T] \times M)$ (соотв. $C^0([0, T], \mathcal{H})$). Псевдодифференциальный оператор B , имея порядок -1 , является ограниченным оператором из C^ϵ в C^1 при всех $\epsilon > 0$ и, разумеется, из C^1 в C^1 . Поскольку h_n ограничена, существует константа C_1 , такая, что для любых n и t имеем

$$[u(t)]_1 \leq C_1([\omega_n(t)]_1 + 1).$$

Тогда $[\omega_n(t)]_1$ оценивается интегрированием по методу характеристик уравнения переноса, определяющего ω_n . Отсюда следует, что

$$\frac{d}{dt}[\omega_n(t)]_1 \leq C_2[\omega_n(t)]_1[u_{n-1}(t)]_1.$$

Из этого последнего уравнения выводится, что если $[u_{n-1}(t)]_1 \leq M$ для всех $t \in [0, T]$, то

$$[\omega_n(t)]_1 \leq [\text{rot } u^0]_1 e^{C_2 M T},$$

откуда

$$[u_n(t)]_1 \leq C_1(1 + [\text{rot } u^0]_1) e^{C_2 M T},$$

и достаточно выбрать T и M , такие, что

$$C_1(1 + [\text{rot } u^0]_1) e^{C_2 M T} \leq M,$$

чтобы заключить по индукции, что $[u_n(t)]_1$, а значит, и $[\omega_n(t)]_1$ ограничены на $[0, T]$. Наконец, оцениваются с помощью уравнения переноса производные порядка k от $\omega_n(t)$:

$$\frac{d}{dt}[\omega_n(t)]_k \leq C_k([u_{n-1}(t)]_1)[\omega_n(t)]_k + [u_{n-1}(t)]_k[\omega_n(t)]_1.$$

Учитывая оценки

$$\begin{aligned} [u_{n-1}(t)]_k &\leq C_k(\|\omega_{n-1}(t)\|_{k-1/2} + 1) \\ &\leq C_k([\omega_{n-1}(t)]_k^{1/2} [\omega_{n-1}(t)]_{k-1}^{1/2} + 1), \end{aligned}$$

легко заключаем индукцией по n и k , что $[\omega_n(t)]_k \leq M_k$ для всех $t \in [0, T]$, а это обеспечивает сходимость (u_n) в C^∞ к решению уравнения (E). Заметим, кроме того, что так полученное время существования T зависит только от мажоранты для $[u_0]_2$.

1.2. Глобальное существование. Предположим, что u — решение системы (E) на интервале $[0, T]$. Исходя из оценок, полученных из уравнения переноса (T),

$$\frac{d}{dt}[\omega_n(t)]_k \leq C_k([u(t)]_1[\omega(t)]_k + [u(t)]_k[\omega(t)]_1), \quad (1)$$

а также оценок, полученных с помощью соотношения Био—Савара,

$$[u(t)]_k \leq C_k([\omega(t)]_k + |h(t)|),$$

заключаем, что если $[u(t)]_1$ ограничена на $[0, T]$, то это верно и для $[u(t)]_k$ при любом k , и предыдущий локальный результат позволяет продолжить решение u за момент времени T .

Нам остается доказать, что $[u(t)]_1$ ограничена на $[0, T]$. Здесь главным является то, что в силу (T)

$$[\omega(t)]_0 = [\operatorname{rot} u^0]_0$$

для всех t . К сожалению, оператор B отображает C^0 не в C^1 , а только в C_*^1 . Чтобы избавиться от этого недостатка, используем следующую нелинейную оценку:

Лемма 2. Пусть B — псевдодифференциальный оператор порядка -1 на компактном многообразии M . Существует число $C > 0$, такое, что для любой регулярной функции f

$$[Bf]_1 \leq C[f]_0 \log \left(1 + \frac{[f]_1}{[f]_0} \right).$$

Этот тип оценки является также одной из главных составляющих результата Шемена, и мы вернемся к нему в разд. 3. А сейчас посмотрим, как это неравенство позволяет завершить доказательство теоремы 1. Из уравнения (E) получаем сохранение энергии,

$$\|u(t)\|_{L^2} = \|u^0\|_{L^2},$$

а из этого, в частности, вытекает, что $h(t)$ ограничена в $\mathcal{H}$. Из соотношения Био—Савара выводим

$$[u(t)]_1 \leq C \log(2 + [\omega(t)]_1),$$

в то время как неравенство (1) для $k = 1$ и лемма Гронуолла дают

$$[\omega(t)]_1 \leq [\omega(0)]_1 (1 + e^{\int_0^t [u(s)]_1 ds}).$$

Сравнивая эти две оценки и снова используя лемму Гронуолла, получаем, что существует такая константа A , для которой

$$[u(t)]_1 \leq Ae^{At}, \quad (2)$$

чем завершается доказательство теоремы 1.

Замечания. а) Удивительно различие роста со временем для двух норм, C_*^1 и C^1 , поля $u(t)$, на вид очень близких; в то время как первая ограничена, единственные известные оценки для второй экспоненциальны (см. неравенство (2)). Неизвестно, является ли этот рост действительно экспоненциальным; заметим, что он индуцировал бы двойной экспоненциальный рост для первой производной вихря ω . Впрочем, в настоящее время неизвестен какой-либо хороший пример регулярного нестационарного решения системы (E). Несмотря на теорему 1, можно утверждать, что изучение ее регулярных решений остается абсолютно открытой проблемой.

б) Когда многообразие M имеет размерность 3, вихрь ω будет уже не функцией, а дифференциальной 2-формой на M . Уравнение переноса (B) не позволяет оценить $[\omega(t)]_0$ без появления величины $[u(t)]_1$, что делает приведенное выше доказательство неприменимым.

2. КВАЗИЛИПШИЦЕВЫ РЕШЕНИЯ

2.1. Функция f на $\mathbf{R}^n$ называется *квазилипшицевой*, если существует константа $C > 0$, такая, что для любых x и y

$$|f(x) - f(y)| \leq C|x - y| \left(1 + \log_+ \frac{1}{|x - y|} \right),$$

где через $\log_+ t$ обозначается наибольшее из чисел 0 и $\log t$. Такая функция принадлежит всем гёльдеровым пространствам порядка $\alpha < 1$. Нетривиальный пример квазилипшицевой функции дается следующей леммой:

Лемма 3. Все функции класса C_*^1 квазилипшицевы.

Доказательство. Выпишем для любого целого числа N равенство

$$f(x) - f(y) = \sum_{p \leq N} (\Delta_p f(x) - \Delta_p f(y)) + \sum_{p > N} (\Delta_p f(x) - \Delta_p f(y)).$$

Используя оценку $[\Delta_p f]_k \leq C2^{(k-1)p}$ при $k = 1$ в первой сумме и при $k = 0$ во второй, получаем

$$|f(x) - f(y)| \leq C(N|x - y| + 2^{-N}),$$

и выбор целой части числа $\log|x - y|/\log 2$ в качестве значения N дает искомую оценку.

Поскольку множество локально квазилипшицевых функций на открытом подмножестве пространства $\mathbf{R}^n$ устойчиво относительно умножения на функции с компактным носителем и замены переменных, можно говорить о локально квазилипшицевых векторных полях на дифференцируемых многообразиях. Имеется следующая теорема Коши—(квази)Липшица:

Предложение 4. Пусть v — квазилипшицево векторное поле на компакте M . Для любой точки $x_0 \in M$ существует единственная кривая $\gamma \in C^1(\mathbf{R}, M)$, являющаяся решением дифференциального уравнения

$$\frac{d}{dt}\gamma = v(\gamma), \quad \gamma(0) = x_0.$$

Кроме того, существует число $k > 0$, такое, что таким образом определенный поток (ϕ^t) удовлетворяет условию

$$\phi^t \in C^{e^{-k|t|}}(M, M) \quad \text{для любого } t \in \mathbf{R}.$$

Доказательство. Можно ограничиться локальными рассуждениями в $\mathbf{R}^n$. Если γ_j для $j = 1, 2, \dots$ — интегральная кривая поля v на интервале $[0, T]$ с $\gamma_j(0) = x_j$, то

$$\frac{d}{dt}|\gamma_1 - \gamma_2|^2 = 2(\gamma_1 - \gamma_2, v(\gamma_1) - v(\gamma_2)).$$

Положив $f(t) = |\gamma_1(t) - \gamma_2(t)|^2$, из квазилипшицевости поля v получаем

$$\frac{d}{dt}f \leq Cf \left(1 + \log_+ \frac{1}{f}\right).$$

Отсюда выводится, что если $f(0) \leq 1$, то для всех $t \in [0, T]$

$$f(t) \leq f(0)e^{-Ct},$$

что доказывает единственность и регулярность потока. Существование вытекает из аппроксимации поля v регулярными полями и теоремы Асколи—Арцела.

Заметим, что приведенная выше теорема переносится на поле $v = v(t, x)$, непрерывно зависящее от параметра t .

2.2. Теперь перейдем к изучению сингулярных решений уравнения (E) , начальные данные которого удовлетворяют условию $\operatorname{rot} u^0 \in L^\infty(M)$. Учитывая уравнение переноса (T) , мы ожидаем, что для любого момента времени t функция $\omega(t)$ находится в классе $L^\infty(M)$ и, следовательно, поле скоростей квазилипшицево. Уравнения (E) и (T) нужно тогда интерпретировать в смысле обобщенных функций следующим образом:

$$\partial_t u = \operatorname{div}(u \otimes \tilde{u}) = -\nabla p, \quad \operatorname{div} u = 0, \quad (E')$$

$$\partial_t \omega + \operatorname{div}(u\omega) = 0, \quad (T')$$

причем условие $\operatorname{div} u = 0$ обеспечивает эквивалентность двух формулировок в случае регулярного решения u .

Теорема 5 (Юдович [Y], 1963). Пусть u^0 — такое векторное поле, что $\operatorname{rot} u^0 \in L^\infty$, $\operatorname{div} u^0 = 0$. Тогда существует единственное решение u уравнения (E') , удовлетворяющее условиям $u \in C^0(\mathbf{R} \times M, TM)$, $\operatorname{rot} u \in L^\infty(\mathbf{R} \times M)$. Кроме того, поле u квазилипшицево по x , поток ϕ^t , начиная с момента $t = 0$, сохраняет объем и для всех действительных t и почти всех $x \in M$ выполняется равенство

$$\omega(t, \phi^t(x)) = \operatorname{rot} u^0(x).$$

Доказательство. Центральная идея этого результата появляется в следующем доказательстве единственности. Воспроизводя доказательство, данное в п. 1.1, замечаем, что, поскольку ∇v теперь уже не принадлежит L^∞ , невозможно просто воспользоваться леммой Гронуолла. Однако так как псевдодифференциальный оператор порядка 0 ограничен на L^p для всех $p \in]1, \infty[$, то $\nabla u(t) \in L^p(M)$, и неравенство Гёльдера дает

$$\begin{aligned} \frac{d}{dt} \|u(t) - v(t)\|_{L^2}^2 &\leq 2 \|\nabla v(t)\|_{L^p} \|u(t) - v(t)\|_{L^{2p}}^2 \\ &\leq 2M^{2/p} \|\nabla v(t)\|_{L^p} \|u(t) - v(t)\|_{L^2}^{2/p'}, \end{aligned}$$

где p' — показатель, сопряженный к p , и $M = \sup_{t \in [0, T]} [u(t) - v(t)]_0$. Красивая идея Юдовича состоит в том, чтобы изучать это дифференциальное неравенство при p , стремящемся к бесконечности, опираясь на следующий результат, являющийся следствием интерполяционной теоремы Марцинкевича (см., например, Стейн [St]) и теории Кальдерона—Зигмунда (см., например, [CM]):

Лемма 6. Пусть M — компактное многообразие и A — псевдодифференциальный оператор порядка 0 на M . Тогда существует такая константа C , что для всех $p < \infty$ верна оценка

$$\|Af\|_{L^p} \leq Cp\|f\|_{L^\infty}.$$

Используя лемму 6 и полагая $g(t) = \|u(t) - v(t)\|_{L^2}^2$, получаем

$$\frac{d}{dt}g \leq CpM^{2/p}g^{1-1/p}.$$

Интегрируя это неравенство и учитывая, что $g(0) = 0$, получаем, что для всех $t \geq 0$

$$g(t) \leq M^2(Ct)^p.$$

Поскольку это неравенство верно для всех $p < \infty$, то $g(t) = 0$ для всех $t < 1/C$ и, наконец, $g = 0$, что доказывает единственность. Для доказательства существования регуляризуем начальные данные u^0 и, используя теорему 1, получаем последовательность (u_n) регулярных решений уравнения (E); из уравнения (T) получаем, что $\text{rot } u_n$ ограничен в L^∞ , а следовательно, $u_n(t)$ равномерно ограничена в C_*^1 по t . Кроме того, рассуждения, использованные выше при доказательстве единственности, показывают, что последовательность (u_n) является последовательностью Коши в $C^0([0, T], L^2)$ для всех $T > 0$. Отсюда выводится с помощью интерполяции, что u_n равномерно сходится к решению u уравнения (E'); остальные анонсированные свойства тогда легко получаются переходом к пределу.

3. ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ А.

3.1. Постановка задачи. Пусть u^0 — поле скоростей на M , вихрь которого имеет вид

$$\omega^0 = a1_{D^0} + b(1 - 1_{D^0}),$$

где D^0 — регулярное открытое подмножество в M , а a и b — числа, удовлетворяющие условию

$$a \operatorname{vol}(D^0) + b (\operatorname{vol}(M) - \operatorname{vol}(D^0)) = 0.$$

Для простоты предположим, что граница области D^0 связна. Обозначим через u решение Юдовича для этих начальных данных, а через ϕ^t поток поля u , начиная с $t = 0$. Тогда вихрь поля $u(t)$ задается формулой

$$\omega(t) = a 1_{D^t} + b(1 - 1_{D^t}),$$

где $D^t = \phi^t(D^0)$. Выберем регулярное ненулевое поле X^0 на компактной окрестности K^0 границы ∂D^0 , касающееся этой кривой, и обозначим через $\gamma^0 : S_1 \rightarrow \partial D^0$ такой диффеоморфизм, что

$$\frac{d}{ds} \gamma^0 = X^0(\gamma^0).$$

Чтобы изучить регулярность кривой $\gamma^t = \phi^t(\gamma^0)$, заметим, что если поле $X^t = \phi_*^t(X^0)$ достаточно регулярно, то

$$\frac{d}{ds} \gamma^t = X^t(\gamma^t), \quad \nabla_{d/ds}^k \frac{d}{ds} \gamma^t = \nabla_{X^t}^k X^t(\gamma^t).$$

Следовательно, всякая информация относительно регулярности поля X^t тотчас превращается в аналогичную информацию относительно γ^t . Мы будем пользоваться этим рассуждением, регулярizando данные и рассматривая соответствующую регулярную последовательность решений (u_n) . Заметив, что $\gamma_n^t = \phi_n^t(\gamma^0)$ равномерно сходится к γ^t , мы сводим все к оценке для поля $X_n^t = \phi_{n*}^t(X^0)$ и его ковариантных производных относительно его самого, например, по норме в C^α для фиксированного $\alpha \in]0, 1[$. Мы ограничимся случаем $[X_n^t]_\alpha$, из чего следует принадлежность кривой к классу $C^{1+\alpha}$. С производными высшего порядка работать после этого проще. Главная информация относительно последовательности регуляризованных начальных данных (ω_n^0) состоит в том, что $L_{X^0} \omega_n^0$ ограничены в пространстве L^∞ и что ω_n^0 ограничены в C^α на дополнении $(K^0)^c$ к K^0 . Помимо оценки для $[X_n^t]_\alpha$ отсюда выводится оценка нормы Липшица для $u_n(t)$, которая с ней тесно связана, и доказывается, что поле $u(t)$ в действительности липшицево для любого t .

Если K — компакт в M и X — векторное поле на M , то положим

$$I(K, X) = \min_{x \in K} |X(x)|.$$

Ожидаемый результат получается применением к последовательности регуляризованных решений следующей теоремы:

Теорема 7 (Шемен [С4]). Для любого $\alpha \in]0, 1[$ существует константа C_α , такая, что для любого регулярного решения u уравнения Эйлера, для любого компакта K^0 в M и для любого регулярного векторного поля X^0 , не обращающегося в нуль на K^0 , верны для всех $t > 0$ следующие оценки, где u^0 — начальные данные для u и ω — его вихрь:

$$\begin{aligned} [u(t)]_1 &\leq C_\alpha M(u^0) \log N_\alpha(K^0, X^0, \omega^0) e^{C_\alpha M(u^0)t}, \\ \|X^t\|_\alpha &\leq C_\alpha \left(\|X^0\|_\alpha + [\operatorname{div} X^0]_0 + \frac{[LX^0\omega^0]_0}{[\omega^0]_0} \right) \\ &\quad \times N_\alpha(K^0, X^0, \omega^0)^2 \exp C_\alpha M(u^0)t, \end{aligned}$$

где $M(u^0) = [\omega^0]_0 + \|u^0\|_{L^2}$,

$$\begin{aligned} N_\alpha(K^0, X^0, \omega^0) \\ = 2 + \frac{\|X^0\|_\alpha + [\operatorname{div} X^0]_0}{I(K^0, X^0)} + \frac{[LX^0\omega^0]_0}{I(K^0, X^0)[\omega^0]_0} + \frac{\|\omega^0\|_{\alpha, (K^0)^c}}{[\omega^0]_0}. \end{aligned}$$

Доказательству теоремы 7 посвящены два следующих пункта.

3.2. Основная оценка. Основным моментом доказательства теоремы 7 является логарифмическое неравенство для лапласиана в размерности 2.

Предложение 8. Для любого $\alpha \in]0, 1[$ существует константа $C_\alpha > 0$, такая, что для любых векторных полей u и X класса C^∞ на M , удовлетворяющих условию $\operatorname{div} u = 0$, и для любого компакта K в M , на котором X не обращается в нуль, имеет место неравенство

$$\begin{aligned} [\nabla u]_0 &\leq C([\omega]_0 + \|u\|_{L^2}) \log \left(2 + \frac{\|X\|_\alpha + [\operatorname{div} X]_0}{I(K, X)} \right. \\ &\quad \left. + \frac{[LX\omega]_0}{I(K, X)[\omega]_0} + \frac{\|\omega\|_{\alpha, (K)^c}}{[\omega]_0} \right), \end{aligned}$$

где $\omega = \operatorname{rot} u$.

Доказательство. Оно опирается на три следующие леммы.

Лемма 9. Пусть $\alpha \in]0, 1[$. Верны следующие неравенства между гёльдеровыми нормами и L^∞ -нормой:

$$(i) [f]_0 \leq C_\alpha \|f\|_0 \log \left(2 + \frac{\|f\|_\alpha}{\|f\|_0} \right),$$

$$(ii) [a.b]_0 \leq C_{\alpha,p} [a]_0 \|b\|_0 \log \left(2 + \frac{\|a\|_\alpha}{[a]_0} + \frac{\|a.b\|_\alpha}{[a]_0 \|b\|_0} \right),$$

где $a = (a_1, \dots, a_p)$, $b = (b_1, \dots, b_p)$, $a.b = a_1 b_1 + \dots + a_p b_p$,

$$(iii) \|L_Z B f\|_\alpha \leq C_{\alpha, \alpha', B} ([L_Z f]_0 + (\|Z\|_{\alpha'} + [\operatorname{div} Z]_0) [f]_0)$$

для всех $\alpha' > \alpha$, всякого псевдодифференциального оператора B порядка -1 и любого векторного поля Z .

Разумеется, достаточно доказать приведенные выше неравенства в $\mathbf{R}^n$. Они вытекают из характеристики гёльдеровых пространств через диадическое разложение и парадифференциальное исчисление Бони [B1]; неравенство (iii) является адаптацией коммутаторной леммы из [B1]. Чтобы доказать (i), запишем для всех N

$$\begin{aligned} [f]_0 &\leq \sum_{p \leq N} [\Delta_p f]_0 + \sum_{p > N} [\Delta_p f]_0 \\ &\leq CN \|f\|_0 + C_\alpha 2^{-N\alpha} \|f\|_\alpha \end{aligned}$$

и выведем нужное неравенство, выбирая оптимальное N , как в доказательстве леммы 3. Наконец, доказательство (ii) является комбинацией логарифмической интерполяции с парапроизведением Бони. Заметим, между прочим, что, применяя (i) к $f = Ag$, где A — псевдодифференциальный оператор порядка 0, получаем.

$$\begin{aligned} [Af]_0 &\leq C_\alpha \|f\|_0 \log \left(2 + \frac{\|f\|_\alpha}{\|f\|_0} \right) \\ &\leq C_\alpha [f]_0 \log \left(2 + \frac{\|f\|_\alpha}{[f]_0} \right), \end{aligned}$$

поскольку A ограничен в C_*^0 , что является уточненной формой леммы 2 (с $A = \nabla B$.) Вот один из вариантов оценки такого типа, где гёльдерова регулярность на этот раз заменена на предположение локализации:

Лемма 10. Пусть A — псевдодифференциальный оператор порядка 0. Существует $C > 0$, такое, что если F и G — два непересе-

секающихся замкнутых подмножества в M и f — существенно ограниченная функция с носителем в F , то

$$\sup_{x \in G} |Af(x)| \leq C[f]_0 \left(1 + \log_+ \frac{1}{d(F, G)} \right).$$

Здесь достаточно провести рассуждения в локальных координатах, и лемма сразу вытекает из формулы

$$Af(z) = \int K(x, y) f(y) dy,$$

где K — ядро с компактным носителем, удовлетворяющее неравенству $K(x, y) \leq C|x - y|^{-n}$. Этот подход дает новое доказательство леммы 2.

Наконец, тот факт, что размерность равна 2, используется в следующем элементарном замечании:

Лемма 11. Пусть E — действительная векторная плоскость, Y — ненулевой вектор из E и J — такой эндоморфизм плоскости E , что $J^2 = -1$. Тогда отображение

$$\begin{aligned} \text{End}(E) &\rightarrow \mathbb{R}^2 \times E, \\ T &\mapsto (\text{tr}(T), \text{tr}(JT), T(Y)), \end{aligned}$$

является изоморфизмом векторных пространств.

Теперь кратко изложим, как завершить доказательство предложения 8. Согласно лемме 10, оценка для ∇u легко получается в области L , достаточно удаленной от K ; действительно, запишем $\omega = \omega_1 + \omega_2$, где носитель ω_1 лежит в окрестности K' области K , не пересекающейся с L , а носитель ω_2 лежит в K^c . Согласно лемме 10, вклад ω_1 в ∇u на L оценивается величиной $[\omega]_0 \log(2 + \theta^{-1})$, где $\theta = d(L, K')$, в то время как вклад ω_2 оценивается в силу леммы 9(i) величиной $[\omega]_0 \log(2 + \theta^{-1} + \|\omega\|_{\alpha, K^c} / [\omega]_0)$, если дополнительно предполагать, что $d(K, K'^c)$ также имеет порядок θ . Параметр θ будет уточнен позже.

Остается оценить ∇u в окрестности $K'' = L^c$ области K . Для этого используем такое следствие из леммы 11: существует число $C > 0$, такое, что для любого сечения T расслоения $\text{End } TM$ и для любого векторного поля Y на M , для которого $|Y(x)| = 1$ на K'' , для всех $x \in K''$ имеет место неравенство

$$|T(x)| \leq C(|\text{tr}(T(x))| + |\text{tr}(JT(x))| + |T(x)Y(x)|).$$

Применяя это неравенство к $T = \nabla u$, получаем

$$|\nabla u(x)| \leq C(|\omega(x)| + |\nabla_Y u(x)|). \quad (3)$$

Выбираем $Y = \chi_\theta X/|X|$, где χ_θ — функция класса C^∞ , принимающая значение 1 в некоторой окрестности области K'' , с носителем в окрестности K''' порядка θ области K'' , такая, что для любого $\beta \in]0, 1[$

$$\|\chi_\theta\|_\beta \leq C\theta^{-\beta}.$$

Выбор θ применяется здесь для того, чтобы гарантировать возможность уменьшать $|X|$ на K''' . Выбирая $\theta = (I(K, K)/A\|X\|_\alpha)^{1/\alpha}$, где A — достаточно большая константа, легко получить, что

$$I(K''', X) \geq \frac{1}{2}I(K, X),$$

$$\|Y\|_\alpha \leq C \frac{\|X\|_\alpha}{I(K, X)}, \quad \|\nabla_Y u\|_\alpha \leq C \frac{\|X\|_\alpha \|\nabla_X u\|_\alpha}{I(K, X)^2}.$$

Тогда в силу леммы 9(ii)

$$[\nabla_Y u] \leq C([\omega]_0 + \|u\|_{L^2}) \log \left(2 + \frac{\|X\|_\alpha}{I(K, X)} + \frac{\|\nabla_X u\|_\alpha}{I(K, X)([\omega]_0 + \|u\|_{L^2})} \right),$$

где, кроме того, использовалось, что $\|u\|_0 \leq C([\omega]_0 + \|u\|_{L^2})$. Возвращаясь к неравенству (3) и применяя лемму 9(iii) для оценки $\|\nabla_X u\|_\alpha$, получаем предложение 8.

3.3. Распространение регулярности. Конец доказательства теоремы 7 опирается на следующую лемму о распространении гёльдеровской регулярности для уравнения переноса.

Лемма 12. Пусть $\alpha \in]0, 1[$. Существует константа C , такая, что для любого $u \in C^\infty(\mathbf{R}_+ \times M, TM)$ и для любых функций f, g на $\mathbf{R}_+ \times M$, удовлетворяющих условию $\partial_t f + L_u f = g$, имеет место неравенство

$$\|f(t)\|_\alpha \leq \|f(0)\|_\alpha e^{\int_0^t [u(s)]_1 ds} + \int_0^t \|g(s)\|_\alpha e^{\int_s^t [u(\sigma)]_1 d\sigma} ds.$$

Приведенная выше оценка легко получается с помощью метода характеристик.

Вернемся к обозначениям п. 3.1; из уравнения переноса

$$\partial_t L_X \omega + L_u L_X \omega = 0$$

сначала выводим, что $[L_X \omega]_0 = [L_X^0 \omega^0]_0$. Кроме того, поскольку ϕ^t сохраняет объем,

$$[\operatorname{div} X]_0 = [\operatorname{div} X^0]_0.$$

Применяем лемму 12 к уравнению переноса

$$\partial_t X + \nabla_u X = \nabla_X u$$

и учитывая оценку

$$\|\nabla_X u\|_\alpha \leq C_\alpha ([L_X \omega]_0 + [u]_1 (\|X\|_\alpha + [\operatorname{div} X]_0)),$$

которая является вариантом леммы 9(iii), из леммы Гронуолла получаем

$$\|X^t\| \leq C_\alpha \left(\|X^0\|_\alpha + [\operatorname{div} X^0]_0 + \frac{[L_X^0 \omega^0]_0}{[\omega^0]_0} \right) e^{2 \int_0^t [u(s)]_1 ds}.$$

Кроме того, можно непосредственно показать, что

$$\begin{aligned} I(K^0, X^0) &\leq I(K^t, X^t) e^{\int_0^t [u(s)]_1 ds}, \\ \|\omega(t)\|_{\alpha, (K^t)^c} &\leq \|\omega^0\|_{\alpha, (K^0)^c} e^{\int_0^t [u(s)]_1 ds}. \end{aligned}$$

Подставляя эти три оценки в неравенство, выведенное в предложении 8, получаем, наконец,

$$[u(t)]_1 \leq C_\alpha ([\omega^0]_0 + \|u^0\|_{L^2}) \left(\log N_\alpha(K^0, X^0, \omega^0) + \int_0^t [u(s)]_1 ds \right),$$

где

$$\begin{aligned} N_\alpha(K^0, X^0, \omega^0) \\ = 2 + \frac{\|X^0\|_\alpha + [\operatorname{div} X^0]_0}{I(K^0, X^0)} + \frac{[L_X^0 \omega^0]_0}{I(K^0, X^0)[\omega^0]_0} + \frac{\|\omega^0\|_{\alpha, (K^0)^c}}{[\omega^0]_0}, \end{aligned}$$

откуда вытекает теорема 7.

4. ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ В

4.1. Исключение давления.

Лемма 13. Пусть $T > 0$ и $u \in L^2([0, T] \times M, TM)$. Следующие условия эквивалентны:

(i) u — решение системы Эйлера

$$\partial_t u + \operatorname{div}(u \otimes \tilde{u}) = -\nabla p, \quad \operatorname{div} u = 0; \quad (E')$$

(ii) u — решение уравнения

$$\partial_t u + A \left(u \otimes \tilde{u} - \frac{1}{2} g(u, u) I \right) = 0, \quad \operatorname{div} u = 0, \quad (E'')$$

где $A = J \nabla \Delta^{-1} \operatorname{rot} \operatorname{div} + H \operatorname{div}$.

Рассмотрим псевдодифференциальный оператор

$$P = J \nabla \Delta^{-1} \operatorname{rot} + H,$$

такой, что $A = P \operatorname{div}$. Согласно разложению Ходжа обобщенных сечений расслоения TM (или обобщенных полей), P является проектором на пространство бездивергентных полей параллельно пространству полей градиентов. Следовательно, система Эйлера (E') эквивалентна системе

$$\partial_t u + P \operatorname{div}(u \otimes \tilde{u}) = 0, \quad \operatorname{div} u = 0.$$

Для любой скалярной обобщенной функции f имеем $\operatorname{div}(fI) = \nabla f$; следовательно, можно добавить к аргументу оператора $P \operatorname{div}$ любой тензор типа fI ; взяв $f = -\frac{1}{2} g(u, u)$, получаем требуемый результат.

Замечания. Совершенно произвольный выбор f , который мы собираемся сделать, будет обоснован ниже в п. 4.3.

Уравнение (E') , в частности, показывает, что любое решение u системы Эйлера, интегрируемое с квадратом на $]0, T[\times M$, непрерывно на $]0, T[$ со значениями в обобщенных полях на M . Поэтому для любого $t \in [0, T]$ можно говорить об обобщенном поле $u(t)$ и обобщенной функции $\omega(t)$.

4.2. Теорема устойчивости. Напомним, что семейство $\mathcal{F}$ функций, τ -интегрируемых на M , называется равномерно τ -интегрируемыми, если для любого $\epsilon > 0$ существует $\delta > 0$, такое, что для

любой борелевской части E множества M , удовлетворяющей условию $\int_E \tau \leq \delta$, имеет место неравенство

$$\int_E |f| \tau \leq \epsilon \quad \text{для любой функции } f \in \mathcal{F}.$$

Критерий Данфорда—Петтиса (см., например, [Gr]) гарантирует, что семейство $\mathcal{F}$ относительно компактно в $L^1(M, \tau)$ в слабой топологии $\sigma(L^1, L^\infty)$. Теорема В является в действительности следствием такого общего результата:

Теорема 14 (Делор [D1, 2]). Пусть $T > 0$ и (u_n) — последовательность интегрируемых с квадратом решений на $]0, T[\times M$ консервативной системы Эйлера (E'). Предположим, что $u_n(t)$ для любого $t \in [0, T]$ — функция, интегрируемая с квадратом, что ее вихрь $\omega_n(t)$ является мерой Радона на M и что существует такая константа C , что

$$\int_M |u_n(t)|^2 \tau + \int_M |\omega_n(t)| \leq C \quad \text{для любого } t \in [0, T]. \quad (4)$$

Кроме того, предположим, что для любого $t \in [0, T]$ существует последовательность $\sigma_n(t)$ положительной (соотв. отрицательной) меры Радона на M , такая, что $(\omega_n(t) - \sigma_n(t))$ равномерно τ -интегрируема на M для $n \in \mathbb{N}$, $t \in [0, T]$. Тогда всякая точка накопления последовательности (u_n) в слабой топологии пространства $L^2([0, T] \times M, TM)$ является решением системы Эйлера (E').

Доказательство теоремы 14 является предметом следующего пункта. Здесь мы покажем, как из нее выводится теорема В.

Пусть u^0 — бездивергентное поле класса L^2 , вихрь ω^0 которого является мерой Радона с положительной (соотв. отрицательной) сингулярной частью σ^0 . Положим $\omega_n^0 = e^{\Delta/n} \omega^0$ и $u_n^0 = B\omega_n^0 + H(u^0)$. Ясно, что u_n^0 сходится к u^0 в среднем квадратичном. Пусть (u_n) — последовательность решений класса C^∞ уравнения (E) на $[0, \infty[\times M$ с начальными данными (u_n^0) . Тогда из условия сохранения кинетической энергии и уравнения переноса (T) получается оценка (4). Кроме того, пусть σ_n — решение уравнения переноса (T), причем $u = u_n$ и $\sigma_n(0) = e^{\Delta/n} \sigma^0$. Тогда $\sigma \geq 0$ (соотв. ≤ 0). Заметим, что $\text{rot } u_n^0 - \sigma_n^0$ равен $e^{\Delta/n} f^0$, где f^0 — производная Радона—Никодима от ω^0 по τ . Отсюда вытекает, что последовательность $(\omega_n^0 - \sigma_n^0)$ равномерно τ -интегрируема, и, используя снова уравнение переноса (T) и тот факт, что поток поля u_n сохраняет объем, заключаем,

что семейство $(\omega_n(t) - \sigma_n(t))_{n,t}$ равномерно τ -интегрируемо. Поскольку, согласно (4), последовательность (u_n) ограничена в классе $L^2([0, T] \times M)$, она допускает точку слабого накопления u , которая, следовательно, является решением для (E') . Более того, из уравнения (E'') вытекает, что последовательность (u_n) равномерно непрерывна на $[0, T]$ со значениями в обобщенных полях на M , а значит, $u(0)$ является пределом u_n^0 , т.е. равно u^0 . Подобным же образом, $u(t)$ для любого $t \in [0, T]$ является точкой накопления для $u_n(t)$, а следовательно, она является функцией, интегрируемой с квадратом, и ее вихрь есть мера Радона с положительной (соотв. отрицательной) сингулярной частью, что завершает доказательство теоремы В.

4.3. Доказательство теоремы 14. Используя равностепенную непрерывность по времени, вытекающую из уравнения (E'') , заключаем, что достаточно доказать, что если $u_n(t)$ для всех $t \in [0, T]$ слабо сходится к $u(t)$ в $L^2(M, TM)$, то u — решение уравнения (E'') . Принимая во внимание уравнение (E'') и теорему о мажорируемой сходимости, получаем теорему 14 как следствие такого результата, не зависящего от t :

Лемма 15. Пусть (u_n) — ограниченная последовательность бездивергентных векторных полей, интегрируемых с квадратом, последовательность вихрей (ω_n) которой ограничена в пространстве мер Радона. Предположим, что существует последовательность положительных (соотв. отрицательных) мер Радона на M , таких, что последовательность $(\omega_n - \sigma_n)$ равномерно τ -интегрируема. Тогда если u_n слабо сходится к u , то $u_n \otimes \bar{u} - g(u_n, u_n)I/2$ сходится к $u \otimes \bar{u} - g(u, u)I/2$ в смысле обобщенных тензор-функций.

Разумеется, мы будем доказывать лемму 15, используя соотношение Био—Савара

$$u_n = B\omega_n + H(u_n).$$

Поскольку последовательность $H(u_n)$ ограничена в конечномерном пространстве $\mathcal{H}$, она относительно компактна, и все сводится к случаю, когда $u_n = B\omega_n$. Таким образом, лемма 15 является следствием общего результата о «микролокальной компенсации» положительностью, который мы анонсировали в произвольной размерности.

Для дальнейшего полезно ввести некоторые понятия. Пусть E — действительное векторное пространство размерности d , и пусть f —

непрерывная однородная функция на E степени $-d$. Обозначим через p радиальное векторное поле на $E \setminus \{0\}$, задаваемое в линейных координатах $(\xi_1, \dots, \xi_d)$ формулой $p = \sum \xi_j \partial_{\xi_j}$. Пусть λ — мера Лебега на E , и пусть φ — функция класса C^1 на E с компактным носителем, принимающая значение 1 в окрестности начала координат. Тогда величина $\int_E f L_p \varphi d\lambda$ не зависит от φ . Действительно, если ψ обладает теми же свойствами, что и φ , то носитель функции $\varphi - \psi$ класса C^1 лежит в компакте $E \setminus \{0\}$, и интегрирование по частям дает

$$\int_E f L_p (\varphi - \psi) d\lambda = \int_E (\varphi - \psi) (-L_p - d) f d\lambda,$$

что равно нулю согласно соотношению Эйлера. Следуя Водзицки (см., например, сообщение Кассела [К] на этом семинаре), обозначим эту величину через $\text{res}(f, \lambda)$. Заметим, что она допускает очень простую метрическую интерпретацию: если Q — скалярное произведение на E , элементом объема которого является λ , то, приближая характеристическую функцию единичного шара относительно Q такими функциями φ , получаем, что $\text{res}(f, \lambda)$ является интегралом от f по единичной сфере относительно Q по мере Лебега.

Теперь предположим, что M есть d -мерное многообразие и что f — непрерывная однородная функция на $T^*M \setminus M$ степени $-d$. Согласно предыдущему, выражение в локальных координатах

$$\text{res}(f) = |dx^1 \wedge \dots \wedge dx^d| \text{res}(f(x, \cdot), |d\xi_1 \wedge \dots \wedge d\xi_d|)$$

определяет на M плотность (т.е. скрученную d -форму в смысле [VAR]), не зависящую от выбора координат; условимся называть эту плотность вычетом функции f . Результат Делюра может быть сформулирован следующим образом:

Лемма 16. Пусть M — компактное d -мерное многообразие, снабженное положительной плотностью, отождествляющей обобщенные функции и распределения. Пусть (ω_n) — ограниченная последовательность мер Радона на M , ограниченная также на пространстве Соболева $H^{-d/2}(M)$ и слабо сходящаяся к ω . Предположим, что существует последовательность (σ_n) положительных мер Радона и последовательность (f_n) , равномерно интегрируемая на M , такие, что

$$\omega_n = \sigma_n + f_n.$$

Тогда для любого скалярного псевдодифференциального оператора на M порядка $-d$, главный символ которого имеет нулевой вычет, имеем

$$(A\omega_n, \omega_n)_{H^{d/2}, H^{-d/2}} \rightarrow (A\omega, \omega)_{H^{d/2}, H^{-d/2}}.$$

Доказательство. Предположим, например, что $\sigma_n \geq 0$. Используя разбиение единицы, сведем доказательство к доказательству аналогичного результата на $\mathbf{R}^d$ с последовательностью ω_n , носители которой принадлежат некоторому фиксированному компактному.

Первый этап. Если ω — мера Радона с компактным носителем в $\mathbf{R}^d$, принадлежащая, кроме того, $H^{-d/2}$, то она является рассеянной, т. е. $\omega(\{x_0\}) = 0$ для любой точки x_0 .

Действительно, если φ — функция класса C^∞ с компактным носителем вблизи начала координат, то теорема о мажорируемой сходимости дает

$$\omega(\{x_0\}) = \lim_{r \rightarrow 0} \int \varphi\left(\frac{x - x_0}{r}\right) \omega(dx).$$

Поскольку семейство $(\varphi((x - x_0)/r))$ ограничено в $H^{d/2}$ и слабо сходится к 0, отсюда вытекает доказываемое утверждение.

Второй этап. Если, кроме того, $\omega = \sigma + f$, где $\sigma \geq 0$ и $f \in L^1$, то для любой функции $g \in L^1$ функция $\sigma + g$ является рассеянной.

Действительно, $\sigma + g = \omega + g - f$.

Третий этап. Ядро оператора A имеет вид $k(x, x - y)$, где $k = k(x, z)$ — функция класса C^∞ на $\mathbf{R}^d \times \mathbf{R}^d \setminus \{0\}$, и предположение о нулевом вычете в точности означает, что $k(x, z)$ ограничена вблизи особого значения $z = 0$ (см., например, [СМ]). Кроме того, если φ является функцией класса C^∞ с компактным носителем, принимающей значение 1 вблизи начала координат, то для f и g , принадлежащих $H^{-d/2}$, имеет место равенство

$$(Af, g)_{H^{d/2}, H^{-d/2}} = \lim_{r \rightarrow 0} \int k(x, x - y) \left(1 - \varphi\left(\frac{x - y}{r}\right)\right) f(y) \bar{g}(x) dx dy.$$

С этого момента, если ω такая же, как на первом этапе, то

$$(A\omega, \omega)_{H^{d/2}, H^{-d/2}} = \int k(x, x - y) \omega(dy) \bar{\omega}(dx)$$

в силу теоремы о мажорируемой сходимости, поскольку по теореме Фубини мера $\omega \otimes \bar{\omega}$ нулевая на диагонали.

Четвертый этап. Теперь используем следующую элементарную лемму теории меры:

Лемма 17. Пусть X — компактное метрическое пространство, (μ_n) — последовательность мер Радона на X , сильно сходящаяся к μ , и (ν_n) — последовательность положительных мер Радона, сильно сходящаяся к ν , такая, что $|\mu_n| \leq \nu_n$. Если f — борелевская функция, ограниченная на X и непрерывная вне замкнутого множества ν -меры нуль, то

$$\int_X f d\mu_n \rightarrow \int_X f d\mu.$$

Достаточно применить эту лемму к $X = K \times K$, где K — компакт в $\mathbf{R}^d$, и $\mu_n = \omega_n \otimes \bar{\omega}_n$, $\mu = \omega \otimes \bar{\omega}$, $\nu_n = \sigma_n + |f_n|$. Поскольку последовательность (f_n) равномерно интегрируема, можно предположить, что f_n слабо сходится к $f \in L^1$ и что $|f_n|$ слабо сходится к $g \in L^1$. Тогда σ_n сильно сходится к положительной мере $\sigma = \omega - f$, и можно взять $\nu = (\sigma + g) \otimes (\sigma + g)$. Остается убедиться, что ν равна нулю на диагонали, что является прямым следствием второго этапа и теоремы Фубини.

Теперь легко доказать лемму 16. Пусть α есть 1-форма и X — регулярное векторное поле на M . Тогда

$$(u_n \otimes \bar{u}_n - \frac{1}{2}g(u_n, u_n), \alpha \otimes X)_{\mathcal{D}', \mathcal{D}} = (A\omega_n, \omega_n)_{H^1, H^{-1}},$$

где $A = {}^t\langle \alpha, B \rangle \langle \bar{B}, X \rangle - \frac{1}{2}{}^t\bar{B}(\alpha, X)B$. Следовательно, его главный символ на единичной сфере относительно g в T^*M дается формулой

$$a(x, \xi) = g(*\xi, \alpha)g(*\xi, \bar{X}) - \frac{1}{2}g(\alpha, \bar{X}).$$

Значит, как показывают элементарные выкладки, его вычет равен

$$\tau \int_S \left(g(\xi, \alpha)g(\xi, \bar{X}) - \frac{1}{2}g(\alpha, \bar{X}) \right) d\sigma(\xi) = 0,$$

что завершает доказательство.

Замечание. Приведенное выше доказательство распространяется на последовательность (u_n) решений уравнения Навье—Стокса

$$\partial_t u_n + \operatorname{div}(u_n \otimes \tilde{u}_n) - \nu_n \Delta u_n = -\nabla p_n, \quad \operatorname{div} u_n = 0,$$

где ν_n — последовательность чисел > 0 , сходящаяся к нулю, которая описывает вязкость жидкости. Решения, найденные Делором, можно рассматривать как ограниченную модель вихревых поверхностей в слабо вязкой жидкости.

ЛИТЕРАТУРА

- [A1] Alinhac S. Remarques sur l'instabilité du problème des poches de tourbillon, *J. Funct. Anal.*, **98** (1991), 361–379.
- [A2] Alinhac S. Un phénomène de concentration évanescence pour des flots non-stationnaires incompressibles en dimension 2, *Comm. Math. Phys.*, **127** (1990), 585–596.
- [Ar] Arnold V. Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits, *Ann. Inst. Fourier*, **16** (1966), 319–361.
- [BF] Bardos C., Frisch U., Sulem C., Sulem P.-L. Finite time analyticity for the two and three-dimensional Kelvin—Helmholtz instability, *Comm. Math. Phys.*, **80** (1981), 485–516.
- [Ba] Batchelor G. K. *An Introduction in Fluid Mechanics*, Cambridge University Press, 1970. [Имеется перевод: Бэтчелор Дж. Введение в динамику жидкости. — М.: Мир, 1973.]
- [Be] Bertozzi A. L. Existence, uniqueness, and a necessary and sufficient condition for blow up of smooth solutions to the contour dynamics equation, *Thèse, University of Princeton*, May 1992.
- [BC] Bertozzi A. L., Constantin P. Global regularity for vortex patches, *Comm. Math. Phys.*, **152** (1993), No. 1, 19–28.
- [Bi] Birkhoff G. Helmholtz and Taylor instability, in: *Hydrodynamic instability*, *Proc. Symp. Appl. Math.*, vol. 13, Amer. Math. Soc., 1962, 55–76.
- [B1] Bony J.-M. Calcul symbolique et propagation des singularités pour les équations aux dérivées partielles non linéaires, *Ann. Sci. Éc. Norm. Sup.*, **14** (1981), 209–246.
- [B2] Bony J.-M. Analyse microlocale et singularités non linéaires, in: *Nonlinear Hyperbolic Problems*, *Lect. Notes in Math.*, vol. 1402, Springer-Verlag, 1989.
- [Br] Brenier Y. The least action principle and the related concept of generalized flows for incompressible perfect fluids, *J. Amer. Math. Soc.*, **2** (1989), 225–255.
- [CO] Caffish R., Orellana O. Longtime existence for a slightly perturbed vortex sheet, *Comm. Pure Appl. Math.*, **39** (1986), 807–838.

- [C1] Chemin J.-Y. Autour du problème des vortex patches, in: Séminaire sur les équations aux dérivées partielles, 1989–1990, exp. no. XI, École Polytechnique, 1990.
- [C2] Chemin J.-Y. Régularité de la trajectoire des particules d'un fluide parfait incompressible remplissant l'espace, *J. Math. Pures Appl.*, **71** (1992), No. 5, 407–417.
- [C3] Chemin J.-Y. Sur le mouvement des particules d'un fluide parfait incompressible bidimensionnel, *Invent. Math.*, **103** (1991), 599–629.
- [C4] Chemin J.-Y. Persistance de structures géométriques dans les fluides incompressibles bidimensionnels, *Ann. École Norm. Sup.*, **26** (1993), No. 4, 517–542.
- [C5] Chemin J.-Y. Évolution d'une singularité ponctuelle dans des équations strictement hyperboliques non linéaires, *Amer. J. Math.*, **112** (1990), No. 5, 805–860.
- [CL] Coifman R., Lions P.-L., Meyer Y., Semmes S. Compensated compactness and Hardy space, prépublication, Université Paris-Dauphine, n° 9123, 1991.
- [CM] Coifman R., Meyer Y. Au-delà des opérateurs pseudo-différentiels, *Astérisque*, **57** (1978).
- [CT] Constantin P., Titi E. On the evolution of nearly circular vortex patches, *Comm. Math. Phys.*, **119** (1988), 177–198.
- [D1] Delort J.-M. Existence de nappes de tourbillon pour l'équation d'Euler sur le plan, in: Séminaire sur les équations aux dérivées partielles, 1990–1991, exp. no. 2, École Polytechnique, 1991.
- [D2] Delort J.-M. Existence de nappes de tourbillon en dimension deux, *J. Amer. Math. Soc.*, **4** (1991), No. 3, 553–586.
- [DM1] Diperna R. J., Majda A. Oscillations and concentrations in weak solutions of the incompressible fluid equations, *Comm. Math. Phys.*, **108** (1987), 667–689.
- [DM2] Diperna R. J., Majda A. Concentrations in regularizations for 2-D incompressible flow, *Comm. Pure Appl. Math.*, **40** (1987), 301–345.
- [DM3] Diperna R. J., Majda A. Reduced Hausdorff dimension and concentration-cancellation for 2-D incompressible flow, *J. Amer. Math. Soc.*, **1** (1988), 59–95.
- [D] Dritschel D. G. Contour dynamics and contour surgery: numerical algorithms for extended, high-resolution modeling of vortex dynamics in two-dimensional, inviscid, incompressible flow, *Computer Physics Reports*, **10** (1989), 77–146.
- [DI] Dritschel D. G., McIntire M. E. Does contour dynamics go singular? *Physics of Fluids*, **A 2** (1990), 748–753.
- [DR] Duchon J., Robert R. Global vortex sheet solutions of Euler equations in the plane, *J. Differential Equations*, **73** (1988), 215–224.
- [EB] Ebin D., Marsden J. E. Groups of diffeomorphisms and the motion of an incompressible fluid, *Ann. of Math.*, **40** (1970), 102–163.

- [E] Euler L. Principes généraux du mouvement des fluides, Mémoires de l'Académie des Sciences de Berlin, **11** (1755), 274–315, 316–361, Opera Omnia, series Secunda, livre 12.
- [EM] Evans L. C. Müller S. Hardy space and two-dimensional Euler equations with nonnegative vorticity, J. Amer. Math. Soc., **7** (1994), No. 1, 199–219.
- [GT] Greengard C., Thomann E. On DiPerna-Majda concentration sets for two-dimensional incompressible flow, Comm. Pure Appl. Math., **41** (1988), 295–303.
- [Gr] Grothendieck A. Espaces Vectoriels Topologiques, S. Paulo, 1958.
- [H] von Helmholtz H. L. F. Über discontinuierlich Flüssigkeitsbewegungen, Monatsberichte der konigl. Akad. Wissenschaften zu Berlin (1868), 215–228.
- [K] Kassel C. Le résidu non commutatif (d'après M. Wodzicki), Séminaire Bourbaki, 1988–89, exposé n° 708, Astérisque, **177–178** (1989), 199–229. [Имеется перевод: В кн.: Труды семинара Бурбаки за 1989 г. — М.: Мир, 1991.]
- [Ka] Kato T. On classical solutions of the two-dimensional non-stationary Euler equation, Arch. Rat. Mech. Anal., **27** (1968), 188–200.
- [Ke] Lord Kelvin (Sir W. Thompson), Mathematical and physical papers, vol. 4, Cambridge University Press, 1910.
- [Kr] Krasny R. Desingularization of periodic vortex sheet roll-up, J. Comput. Phys., **65** (1986), 292–313.
- [L] Lebeau G. Interaction des singularités pour les équations aux dérivées partielles non linéaires (d'après J.-M. Bony et al.), Séminaire Bourbaki, 1984–85, exp. 642, Astérisque, **133–134** (1986), 209–222.
- [Li] Lichtenstein L. Über einige Existenzprobleme der Hydrodynamik homogener unzusammen drückbarer, reibungsloser Flüssigkeiten und die Helmholtzschen Wirbelsätze, Math. Z., **23** (1925), 89–154, 310–316, **26** (1927), 196–323, 387–415, **725**, **32** (1930) 608.
- [Lo] Lions P.-L. Remarks on incompressible models of Fluid Mechanics, prépublication Université de Paris-Dauphine, n° 9206, 1992; см. также: Developments in partial differential equations and applications to mathematical physics (Ferrara, 1991), Plenum Press, New York, 1992, 59–64.
- [M1] Majda A. Vorticity and the mathematical theory of incompressible fluid flow, Comm. Pure Appl. Math., **39** (1986), supplément, 187–220.
- [M2] Majda A. Vortex dynamics: Numerical Analysis, scientific computing, and mathematical theory, in: Proceeding of the First International Congress for Industrial and Applied Mathematics, S.I.A.M. Publ., 1988, 153–182.
- [M3] Majda A. The interaction of nonlinear analysis and modern applied mathematics, in: Proc. I.C.M., Kyoto, 1990.

- [R] de Rham G. Variétés différentiables, Hermann, Actualités Sci. Indust., Paris, 1960. [Имеется перевод I изд.: Де Рам Ж. Дифференцируемые многообразия. — М.: ИЛ, 1956.]
- [SB] Saffman P., Baker G. Vortex interactions, Ann. Rev. Fluid Mech., 11 (1979), 95–122.
- [S] Serfati P. Vortex patches bidimensionnels, рукопись, Février 1990; thèse Université de Paris VI, septembre 1992.
- [Sh] Шнирельман А. И. О геометрии групп диффеоморфизмов и динамике идеальной несжимаемой жидкости. — Матем. сб., нов. сер., 128 (170), (1985), № 1, 82–109.
- [St] Stein E. Singular Integrals and Differentiability Properties of Functions, Princeton University Press, 1970.
- [T] Sir Taylor G. The instability of liquid surfaces when accelerated in a direction perpendicular to their planes, Proc. Roy. Soc. London, Ser. A, 201 (1950), 192–196.
- [VAR] Bourbaki N. Variétés différentielles et analytiques, fascicule de résultats, paragraphes 8 à 15, Hermann, Paris, 1971. [Имеется перевод: Бурбаки Н. Дифференцируемые и аналитические многообразия. Сводка результатов. — М.: Мир, 1975, §§ 8–15.]
- [W] Wolibner W. Un théorème sur l'existence du mouvement plan d'un fluide parfait, homogène, incompressible, pendant un temps infiniment long, Math. Z., 37 (1933), 698–726.
- [Y] Юдович В. И. Нестационарные течения идеальной несжимаемой жидкости. — Журнал вычисл. матем. и матем. физ., т. 3, 1963, № 6, с. 1032–1066.
- [Z] Zabusky N., Hughes M., Roberts K. Contour dynamics for the Euler equations in two dimensions, J. Comput. Phys., 30 (1979), 96–106.

PATRICK GÉRARD

Université de Paris-Sud
 Département de Mathématiques
 Bâtiment 425
 F-91405 ORSAY Cedex

E-mail: Patrick.Gerard@math.u-psud.fr

РАСПРЕДЕЛЕНИЕ СОБСТВЕННЫХ ЗНАЧЕНИЙ СЛУЧАЙНЫХ ОПЕРАТОРОВ И МАТРИЦ

Леонид Па́стур

1. ВВЕДЕНИЕ

Изучение распределения собственных значений для различных классов самосопряженных операторов составляет один из старейших и наиболее популярных разделов спектральной теории. Вспомним, например, знаменитую проблему Г. Вейля об асимптотическом поведении при высоких энергиях функции

$$\mathcal{N}_\Lambda(\lambda) = \#\{\lambda_i \leq \lambda\}, \quad (1)$$

считающей собственные значения λ_i в самосопряженной граничной задаче для эллиптического оператора A_Λ в компактной области $\Lambda \subset \mathbb{R}^d$. Согласно Г. Вейлю, в простейшем, но важном случае, когда $A_\Lambda = -\Delta$,

$$\mathcal{N}_\Lambda(\lambda) = c_d |\Lambda| \lambda^{d/2} + o(\lambda^{d/2}), \quad \lambda \rightarrow \infty, \quad (2)$$

где $|\Lambda|$ — объем области Λ , а c_d зависит только от размерности пространства.

Изучение следующих членов этой и подобных асимптотических формул для разных случаев породило множество красивых и важных результатов и вскрыло чрезвычайно глубокие связи спектральной теории с геометрией, топологией, эргодической теорией, различными аналитическими и асимптотическими методами и т. д.

Напомню, что одной из причин, побудивших Вейля заняться этой задачей, послужила необходимость обоснования вывода Рэлея и Джинсом формулы для спектрального распределения излучения черного тела и вывода Дебаем формулы для теплоемкости кристалла. Эти задачи, сыгравшие важную роль в развитии квантовой механики, можно коротко сформулировать как задачи построения термодинамики идеального Бозе-газа. Одним из принципиальных понятий современной математической физики является понятие термодинамического предела, введенное для того, чтобы

изучать объемные свойства макроскопических систем. С этой точки зрения довольно естественно рассмотреть последовательность $\{\Lambda_k\}$ компактных областей, заполняющих все $\mathbf{R}^n$ при $k \rightarrow \infty$, доказать существование для каждого $\lambda \in \mathbf{R}$ предела

$$\lim_{k \rightarrow \infty} N_{\Lambda_k} \equiv N(\lambda), \quad (3)$$

где

$$N_{\Lambda}(\lambda) = |\Lambda|^{-1} \mathcal{N}_{\Lambda}(\lambda), \quad (4)$$

проверить его независимость от выбора последовательности $\{\Lambda_k\}$ расширяющихся множеств при достаточно слабых предположениях об этих последовательностях, а затем изучить неубывающие функции (3) для различных областей значений λ (в частности, при $\lambda \rightarrow \infty$) и других параметров.

В случае лапласиана и достаточно регулярных последовательностей областей эту задачу можно свести к задаче о высокочастотной асимптотике функции $N_{\Lambda}(\lambda)$ при фиксированном Λ . Поэтому Г. Вейль был так заинтересован в доказательстве независимости старшего члена в (2) от вида области и считал этот результат одним из важных своих достижений.

Функцию $N(\lambda)$, известную под названием интегрированной плотности состояний (ИПС), можно ввести и в дискретном случае, т. е. для последовательности $n \times n$ -матриц при $n \rightarrow \infty$. Аналог функции (4) в этом случае имеет вид

$$N_n(\lambda) = \#\{\Lambda_i \leq \lambda\} n^{-1}. \quad (5)$$

Случай, когда матрица представляет собой ограничение на конечное множество $\Lambda \subset \mathbf{Z}^d$ теплового оператора, изучался Гренандером и Серге [1].

В этой статье я собираюсь обсудить три класса случайных дифференциальных и матричных операторов, для которых сформулированная выше задача может быть изучена достаточно полно. К первому классу относятся дифференциальные и конечно-разностные операторы со случайными коэффициентами. Во второй класс попадают случайные матрицы с независимыми и одинаково распределенными элементами. Таким образом, главное различие этих двух классов состоит в том, что матрицы первого класса содержат ненулевые элементы только на конечном числе диагоналей вблизи главной диагонали, причем это число не зависит от Λ , а в

матрицах второго класса все члены имеют, грубо говоря, одинаковый порядок. Эта разница оказывается весьма существенной и приводит к различной постановке задач, различным формам ответов и к различным подходам к решению. Удивительно поэтому, что существует класс операторов, объединяющий, в некотором смысле, оба предыдущих класса. Это и будет наш третий класс.

2. ДИФФЕРЕНЦИАЛЬНЫЕ И КОНЕЧНО-РАЗНОСТНЫЕ ОПЕРАТОРЫ СО СЛУЧАЙНЫМИ КОЭФФИЦИЕНТАМИ

Ясно, что если мы собираемся доказать существование предела (3), то мы должны наложить некоторые условия на коэффициенты соответствующих операторов. В самом деле, рассмотрим, например, оператор Шрёдингера H_Λ , имеющий на Λ вид

$$-\Delta + q(x) \quad (6)$$

и действующий на функциях, удовлетворяющих каким-нибудь граничным условиям, скажем условиям Дирихле, на $\partial\Lambda$. Тогда легко видеть, что если $q(x) \rightarrow \infty$ при $|x| \rightarrow \infty$, то $N(\lambda) \equiv 0$, а если $q(x) \rightarrow 0$ при $|x| \rightarrow \infty$, то $N(\lambda) = N_0(\lambda)$, где функция

$$N_0(\lambda) = c_d \lambda^{d/2} \quad (7)$$

представляет собой, согласно (2), интегрированную плотность состояний оператора $-\Delta$ во всем пространстве. Таким образом, чтобы получить нетривиальный результат (отличный от 0 и $N_0(\lambda)$), необходимо рассматривать ненулевые потенциалы $q(x)$, не растущие и не стремящиеся к нулю на бесконечности. Кроме того, ясно, что поведение $q(x)$ на бесконечности должно быть довольно «регулярным», чтобы функция $N_\Lambda(\lambda)$ не слишком осциллировала при больших Λ .

Простейший нетривиальный случай — это случай периодического потенциала. Соответствующий оператор Шрёдингера описывает движение электрона в идеальном твердом теле. Довольно общий класс потенциалов, для которых предел (3) существует, составляют реализации (пробные функции) метрически транзитивных (эргодических) случайных полей в $\mathbb{R}^d$, которые моделируют неупорядоченные твердые тела. Это означает, что мы рассматриваем не отдельный оператор H_Λ при фиксированном $q(x)$, а семейство операторов $H_\Lambda(\omega)$, отвечающих потенциалам вида

$$q(x, \omega) = Q(T_x \omega), \quad (8)$$

где $Q(\omega)$, $\omega \in \Omega$, — измеримая функция на вероятностном пространстве Ω , в котором действует группа T_x , $x \in \mathbb{R}^d$, сохраняющих меру и метрически транзитивных автоморфизмов. В дискретном случае, т. е. для конечно-разностного аналога оператора (6)

$$-\Delta_{\text{disc}} + q(x), \quad x \in \mathbb{Z}^d, \quad (9)$$

простой пример семейства (8) дают независимые одинаково распределенные (н.о.р.) случайные величины. Этот оператор представляет собой дискретную модель движения электрона в полностью неупорядоченных твердых телах: аморфных средах, сплавах замещения и т. д.

В начале семидесятых годов нами было доказано существование предела (3) с вероятностью 1 и его неслучайность для оператора Шрёдингера с потенциалами вида (8), ограниченными снизу. В дальнейшем эта задача изучалась рядом авторов, и полученные результаты близки к оптимальным (см., например, книги [2, 3] и библиографию в них). Например, в случае оператора (9) ИПС существует для всякого метрически транзитивного поля (8) на $\mathbb{Z}^d$, которое конечно с вероятностью 1.

Обсудим теперь некоторые свойства ИПС, ограничиваясь случаем операторов Шрёдингера (6) и (9) (соответствующие доказательства и ссылки см., например, в книгах [2, 3]).

(i) **Гладкость.** В дискретном случае (т. е. для операторов (9)) и в одномерном непрерывном случае (т. е. для операторов (6) при $d = 1$) ИПС всегда непрерывна. Это можно доказать на основе простых эргодических рассуждений. Более тонкая техника, основанная на понятии показателей Ляпунова, доказывает $\log$ -гёльдеровость ИПС:

$$|N(\lambda_1) - N(\lambda_2)| \leq \text{const} |\log |\Lambda_1 - \Lambda_2||^{-1}.$$

Имеются контрпримеры, показывающие, что эта оценка оптимальна на всем классе метрически транзитивных потенциалов. Если, однако, мы рассмотрим дискретный оператор (9) с н.о.р. случайным потенциалом, плотность которого имеет вид $F(dq) = f(q)dq$, $\sup_{q \in \mathbb{R}} f(q) \leq f_0 < \infty$, то, согласно Вегнеру,

$$N(d\lambda) = \rho(\lambda)d\lambda, \quad \sup_{\lambda \in \mathbb{R}} \rho(\lambda) \leq f_0.$$

(Здесь и ниже одним символом $N(\cdot)$ обозначаются неубывающая функция и очевидным образом связанная с ней мера.)

Предполагая достаточную степень гладкости функции $f(q)$, можно вывести более сильные утверждения о гладкости функции $\rho(\lambda)$, вплоть до бесконечной гладкости или вещественной аналитичности. С другой стороны, известно, что если при $d = 1$ н.о.р. случайный потенциал принимает всего два значения, скажем 0 и q_0 , то для достаточно большого q_0 у ИПС имеется сингулярно непрерывная компонента [2, 3].

Гладкость ИПС, сама по себе чрезвычайно интересная, играет важную роль в доказательствах андерсоновской локализации, т.е. присутствия точечной компоненты в спектре операторов (6) и (9).

Необходимо подчеркнуть, что большинство известных результатов о гладкости ИПС и других спектральных свойствах случайного оператора относится к дискретному случаю (9). Непрерывный случай намного более сложен с технической точки зрения и намного хуже изучен.

(ii) **Явные формулы.** В одномерном непрерывном случае с марковским случайным потенциалом, принимающим два значения, и в многомерном дискретном случае с независимым одинаково распределенным по Коши потенциалом или с некоторыми специальными квазипериодическими потенциалами ИПС найдена явно. В последних случаях ИПС представляет собой функцию, аналитическую в некоторой полосе $|\operatorname{Im} \lambda| \leq \text{const}$. В этих формулах содержится много количественной информации о поведении ИПС на различных частях спектра.

(iii) **Асимптотическое поведение.**

(а) *Высокие энергии.* Предположим, что потенциал $q(x)$ в (6) — это метрически транзитивное поле на $\mathbb{R}^d$, такое, что

$$E\{|q(0)|^{p+1}\} < \infty,$$

где p — наименьшее четное число, превосходящее $d/2$. Тогда

$$N(\lambda) = N_0(\lambda)(1 + o(1)), \quad \lambda \rightarrow \infty, \quad (10)$$

где $N_0(\lambda)$ — это ИПС для $-\Delta$, имеющая вид (7), т.е. аналог асимптотики Вейля (2), содержащей только старший член. Эвристическая теория возмущений в одномерном случае дает следующий член в виде

$$N(\lambda) = N_0(\lambda) - (4\pi\lambda)^{-1} \int_0^\infty B(x) \cos 2\sqrt{\lambda}x dx. \quad (11)$$

Здесь мы предполагаем, что $E\{q(x)\} = 0$, и полагаем $E\{q(x)q(0)\} = B(x)$. Эту формулу можно обосновать во многих интересных случаях. Но доказательства ее многомерных аналогов до сих пор отсутствуют.

Однако задача обоснования этой формулы представляется довольно важной, так как это — простейший нетривиальный случай квазиклассической асимптотики для случайного оператора Шрёдингера. В течение последнего десятилетия физики обнаружили большое число красивых результатов с помощью различных версий теории возмущений и квазиклассического приближения. Эти результаты составляют так называемую теорию слабой локализации. Их математический смысл остается совершенно неисследованным.

(b) *Низкие энергии.* Эта область значений спектрального параметра не имеет аналога в традиционной спектральной теории; однако для случайных операторов она оказывается довольно богатой и интересной. Упомянем несколько типичных результатов.

В соответствии с идеологией квантовой механики низкоэнергетическая часть спектра существенно зависит от выбранного потенциала. Несмотря на это, можно выделить несколько хорошо определенных типов асимптотического поведения ИПС при низких энергиях.

Самый простой из них соответствует неограниченному потенциалу. Рассмотрим, например, потенциал, для которого функция распределения вероятности $F(q)$ удовлетворяет условию

$$\lim_{q \rightarrow -\infty} \frac{\log F(q)}{\log F(q+a)} = 1 \quad \forall a \in \mathbb{R}. \quad (12)$$

Тогда для операторов (6) и (9)

$$\log N(\lambda) = \log F(\lambda)(1 + o(1)), \quad \lambda \rightarrow -\infty. \quad (13)$$

В дискретном случае, т. е. для операторов (9), имеется совсем простое доказательство асимптотики (13), основанное на элементарных вариационных рассуждениях. В непрерывном случае, т. е. для операторов (6), доказательство оказывается более сложным, хотя и его можно изложить в вариационных терминах. Нами это доказательство дано для случайного гауссовского потенциала и для так называемого пуассоновского потенциала

$$q(x) = \sum_j u(x - x_j), \quad (14)$$

где $u(x)$ — неположительная функция с компактным носителем, а $\{x_j\}$ — случайно распределенная по Пуассону последовательность точек в $\mathbf{R}^d$.

Асимптотические соотношения (13), по-видимому, имеют место в непрерывном случае для довольно широкого класса случайных потенциалов, удовлетворяющих достаточно слабым предположениям о непрерывности.

Другой тип асимптотического поведения демонстрируют случайные потенциалы, ограниченные снизу, например потенциалы вида (14) с неотрицательной функцией $u(x)$. В этом случае спектр совпадает с $\mathbf{R}^+$ и

$$\log N(\lambda) = -c_d n \lambda^{-d/2} (1 + o(1)), \quad \lambda \rightarrow 0_+, \quad (15)$$

где c_d зависит только от d , а n — это концентрация пуассоновых точек.

Эта асимптотика была предложена И. Лифшицем и доказана рядом авторов и, в частности, нами. Доказательство основано на технике винеровских интегралов, и существенную его часть составляют глубокие результаты Донскера и Варадхана о больших отклонениях для винеровского процесса или, другими словами, о бесконечномерном методе Лапласа.

Естественно заняться изучением следующих членов асимптотик (13) и (15), в частности, найти коэффициент при экспоненте. Точное значение этого коэффициента найдено лишь в некоторых одномерных случаях.

Упомяну еще один тип асимптотического поведения ИПС. Рассмотрим эллиптический оператор

$$-\sum_{i,j=1}^d \frac{\partial}{\partial x_i} \left(a_{ij}(x) \frac{\partial}{\partial x_j} \right), \quad (16)$$

где $a_{ij}(x)$ — случайные поля вида (8). Спектр такого оператора также совпадает с $\mathbf{R}^+$, и нас интересует поведение функции $N(\lambda)$ при $\lambda \rightarrow +0$. Замена переменных $\lambda = \varepsilon^2$, $x_i = \xi_i \varepsilon^{-1}$ и $\varepsilon^d N(\varepsilon^2) = \nu(\varepsilon)$ сводит нашу задачу к изучению операторов (16), в которых функции $a_{ij}(x)$ заменены быстро сходящимися функциями $a_{ij}(\xi/\varepsilon)$. Решению этой задачи посвящена теория гомогенизации, созданная в последние десятилетия.

С. Козлову удалось доказать довольно общий результат, применимый к этой спектральной задаче. С помощью этого результата можно доказать, что для случайного оператора (16) (ср. (7) и (10))

$$N(\lambda) = c(a)\lambda^{d/2}(1 + o(1)), \quad \lambda \rightarrow +0,$$

где $c(a)$ — довольно сложный функционал на a_{ij} . Явный вид этого функционала известен в одномерном случае и в некоторых двумерных примерах.

3. СЛУЧАЙНЫЕ МАТРИЦЫ

Рассмотрим симметрическую случайную $n \times n$ -матрицу $M^{(n)}$, элементы которой имеют вид

$$M_{ij}^{(n)} = n^{-1/2}W_{ij}, \quad i, j = 1, \dots, n, \quad (17)$$

где W_{ij} — симметричные ($W_{ij} = W_{ji}$) независимые при $i \leq j$ случайные величины, такие, что

$$E\{W_{ij}\} = 0, \quad E\{W_{ij}W_{kl}\} = v^2(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \quad (18)$$

Простейший и лучший всего изученный случай таких матриц отвечает гауссову распределению случайных величин W_{ij} , и соответствующую плотность можно представить в виде

$$Z_n^{-1} \exp(\text{Tr } W^2 / 4a^2) dW, \quad dW = \prod_{i=1}^n dW_{ii} \prod_{i < j} dW_{ij}, \quad (19)$$

где Z_n — нормирующая константа. Вигнер ввел этот матричный ансамбль для описания некоторых свойств спектров энергетических уровней тяжелых ядер.

Для таких матричных ансамблей получено много красивых результатов (см. книги [4, 5] и библиографию в них). В частности, для последовательности $N_n(\lambda)$, заданной формулой (5), где $\lambda_1, \dots, \lambda_n$ — собственные числа соответствующей случайной матрицы, с вероятностью 1 имеем

$$\lim_{n \rightarrow \infty} N_n(\lambda) \equiv N(\lambda) = \int_{-\infty}^{\lambda} \rho(\mu) d\mu, \quad (20)$$

где

$$\rho(\lambda) = \frac{1}{2\pi v^2} \begin{cases} \sqrt{4v^2 - \lambda^2}, & |\lambda| \leq 2v, \\ 0 & \text{в противном случае.} \end{cases} \quad (21)$$

Это хорошо известный закон Вигнера, или полукруговой закон. Вигнер и другие авторы распространили этот результат на ряд более общих случаев. В [6] доказан оптимальный результат, состоящий в том, что если заменить в (20) сходимость с вероятностью 1 на сходимость по вероятности, то (20) и (21) остаются верными при условии, что случайные величины W_{ij} при $i < j$ одинаково распределены и это распределение удовлетворяет условию (18).

Это условие, представляющее собой аналог хорошо известного в теории вероятностей условия Линдберга, в несколько более общем виде оказывается необходимым и достаточным [5, 6]. Таким образом, полукруговой закон (20) и (21) является столь же универсальной формой предельного распределения собственных чисел случайных матриц с независимыми элементами, сколь нормальное распределение оказывается универсальным для нормированных сумм независимых случайных величин.

Последние результаты были получены путем изучения асимптотики рекуррентных соотношений по n для $n^{-1}\text{Tr}(M^{(n)} - z)^{-1}$, $\text{Im } z \neq 0$, при $n \rightarrow \infty$. Эта техника позволяет изучать и матричные ансамбли более общего вида

$$H_{ij}^{(n)} = h_i \delta_{ij} + n^{-1/2} W_{ij}, \quad i, j = 1, \dots, n, \quad (22)$$

с теми же функциями W_{ij} и такими числами h_i , что для каждого λ , являющегося точкой непрерывности некоторой функции распределения $N_0(\lambda)$, существует предел

$$\lim_{n \rightarrow \infty} \#\{h_i \leq \lambda\} n^{-1} \equiv N_0(\lambda). \quad (23)$$

Этот предел можно назвать невозмущенной интегрированной плотностью состояний. Рассмотрим ее преобразование Стильтеса

$$r_0(z) = \int (\mu - z)^{-1} N_0(d\mu), \quad \text{Im } z \neq 0.$$

Согласно [6], если

$$r_n(z) = \int (\mu - z)^{-1} N_n(d\mu),$$

то предел $r(z) = \lim_{n \rightarrow \infty} r_n(z)$ по вероятности существует для всех $\text{Im } z \neq 0$ и его можно найти как единственное решение функционального уравнения

$$r(z) = r_0(z + v^2 r(z)) \quad (24)$$

в классе функций, аналитических при всех $\text{Im } z \neq 0$ и таких, что $\text{Im } r(z) \cdot \text{Im } z > 0, \text{Im } z \neq 0$.

ИПС, отвечающая решению функционального уравнения (24), называется деформированным полукруговым законом. Из уравнения (24) можно вывести существование плотности состояний, ее ограниченность, описать носитель плотности, его асимптотическое поведение вблизи граничных точек носителя и т.д.

Заслуживает внимания то, что приведенные выше результаты для ансамбля (22) можно рассматривать как частный случай более общего результата (см. [6]). Рассмотрим следующий ансамбль $n \times n$ -матриц:

$$H^{(n)} = h^{(n)} + \sum_{i=1}^m \tau_i(\cdot, q_i) q_i, \quad (25)$$

где $h^{(n)}$ — произвольная матрица с предельной ИПС $N_0(d\lambda)$, τ_i и q_i — соответственно независимые одинаково распределенные случайные величины и единичные векторы в $\mathbb{R}^n$, причем распределение последних «достаточно близко» к равномерному. Тогда при $n \rightarrow \infty, m \rightarrow \infty, m/n \rightarrow c < \infty$ существует предельная ИПС для ансамбля (25), а преобразования Стильтьеса $r(z)$ и $r_0(z)$ плотностей $N(d\lambda)$ и $N_0(d\lambda)$ связаны соотношением

$$r(z) = r_0 \left(z + c \int_{-\infty}^{\infty} \tau(1 + \tau r(z))^{-1} \sigma(d\tau) \right), \quad (26)$$

где $\sigma(d\tau)$ — это плотность распределения величин τ_i . Легко видеть, что, заменив в (26) $z + cE\{\tau\}$ на z и сделав предельный переход $c \rightarrow \infty, E\{\tau^2\} \rightarrow 0, cE\{\tau^2\} \rightarrow v^2$, мы придем к уравнению (24).

Уравнения (24) и (26) можно считать полным решением задачи о собственных значениях для ансамблей (22), (25) и (27).

Как я уже упоминал выше, мотивировкой постановки задач изначально послужила ядерная физика. В последние десятилетия ансамбли случайных матриц использовались и во многих других областях квантовой физики. В частности, задача изучения квантового ротора, подверженного периодическим по времени толчкам, архетипической модели квантового хаоса, связана с изучением распределения собственных значений случайных матриц

$$M_{ij}^{(n,b)} = (2b_n + 1)^{-1/2} \phi((i - j)/b_n) W_{ij}. \quad (27)$$

Здесь $b_n \rightarrow \infty$ при $n \rightarrow \infty$ и $\phi(t) = \phi(-t) \in \mathbb{R}$ — кусочно непрерывная ограниченная функция с компактным носителем, такая, что

$$\int \phi^2(t) dt = 1. \quad (28)$$

В частности, если $\phi(t) \equiv 1$ и $2b_n + 1 = n$, то мы получаем (17), а если $\phi(t)$ — характеристическая функция интервала $(-1, 1)$, то условия (27), (28) задают так называемые «ленточные» матрицы, т.е. матрицы, в которых ненулевые н.о.р. элементы расположены в полосе шириной $2b_n + 1$ вокруг главной диагонали матрицы $M^{(n,b)}$.

ИПС для таких матричных ансамблей была найдена недавно несколькими коллективами авторов [8–10]. Наиболее общий результат доказан в [10]. Согласно ему, если $\lim_{n \rightarrow \infty} b_n n^{-1}$ равен 0 или $1/2$, то ИПС для матриц (27) дается полукруговым законом (20), (21). Если, однако, $0 < \lim_{n \rightarrow \infty} b_n n^{-1} < 1/2$, то предельное распределение собственных значений существует, но не совпадает с полукруговым законом. Его преобразование Стильтьеса можно найти как единственное решение некоторого нелинейного интегрального уравнения. Ситуацию можно «подправить» в том смысле, что предельное распределение можно все-таки сделать полукруговым законом, подобрав подходящую периодическую функцию $\phi(t)$ в (27).

Доказательство этих результатов основывается на одном новом подходе. Он состоит в построении бесконечной системы линейных уравнений для моментов диагональных элементов матрицы резольвенты $(V^{(n,b)} - z)^{-1}$ и асимптотического решения этой системы относительно b_n при $n \rightarrow \infty$. Этот метод оказывается довольно эффективным и общим и позволяет изучить широкий класс смежных задач (см. обзор [11]). Можно, в частности, доказать, что деформированный полукруговой закон (24) выполняется для произвольной, не обязательно диагональной, «невозмущенной» матрицы $h^{(n)} = \{h_{ij}\}_{i,j=1}^n$ в (22), для которой существует ИПС, т.е. если $\mu_1, \dots, \mu_n$ — собственные числа матрицы $h^{(n)}$, то нам нужно предполагать лишь существование предела (5).

В следующем разделе мы рассмотрим другие задачи, к решению которых применим тот же подход.

Выше обсуждались матричные ансамбли, которые, наряду с гауссовым ансамблем (19), обладают свойством статистической независимости всех функционально независимых элементов. У всех

этих ансамблей одинаковая ИПС, описываемая полукруговым законом (21). Имеется еще одно обобщение плотности (19), задаваемое вероятностным распределением вида

$$Z_n^{-1} \exp(-n \text{Tr} V(M)) dM, \quad (29)$$

где $V(t)$ — вещественнозначная функция, растущая на бесконечности быстрее, чем $a \log |t|$ для всех $a \geq 0$. Гауссов ансамбль (19), очевидно, соответствует случаю $V(t) = t^2/4a^2$. Многочлены четной степени $p > 2$ появляются в квантовой теории поля, статистической механике случайных поверхностей, комбинаторике и т. д. (см. [7] и приведенную там библиографию).

ИПС для таких ансамблей отличается от полукругового закона. Ее носитель состоит, вообще говоря, из $p/2$ интервалов, причем в концах этих интервалов плотность состояний ведет себя как корень квадратный из малого параметра. Однако, меняя функцию $V(t)$ (т. е. коэффициенты многочлена $V(t)$), можно добиться того, чтобы число интервалов стало меньшим, а поведение плотности $\rho(\lambda)$ на концах интервалов стало другим. Например, для $V(t) = |t|^\alpha$, $\alpha > 1$,

$$\rho(\lambda) = v_\alpha (\lambda B_\alpha^{-1}) B_\alpha^{-1}, \quad (30)$$

где

$$v_\alpha(t) = v_\alpha(-t) = \frac{2\alpha}{\pi} \begin{cases} \int_0^1 \tau^{\alpha-1} (\tau^2 - t^2)^{-1/2} d\tau, & |t| \leq 1, \\ 0, & |t| > 1, \end{cases}$$

$$B_\alpha = \left(2\alpha\pi^{-1} \int_0^1 t^\alpha (1 - t^2)^{-1/2} dt \right)^{1/\alpha}. \quad (31)$$

Существует красивый подход, который позволяет изучить ансамбли (29). Он основан на выражении математического ожидания унитарно-инвариантной функции на случайных матрицах из ансамбля (29) через вещественные многочлены, ортогональные с весом $\exp(-nV(\lambda))$ [4, 7]. С помощью этого подхода физики обнаружили много интересных свойств распределения собственных значений ансамбля (29). Строгих доказательств, однако, немного. Одна из причин отсутствия строгих результатов состоит в том, что они должны основываться на точных асимптотических формулах для соответствующих ортогональных многочленов. Лучший из недавних результатов [13] справедлив лишь для случая $V(t) = |t|^\alpha$.

Соответствующие асимптотические формулы позволяют строго вывести вид (30), (31) плотности состояний для ансамбля (29) с той же функцией $V(t)$ [14]. Эти формулы, однако, не настолько точны, чтобы с их помощью можно было вычислить другие важные характеристики случайных матриц. Кроме того, нет почти никаких строгих результатов для немонотонных $V(t)$, особенно в невыпуклом случае, например для $V(t) = at^2 + bt^4$ при $a < 0$, $b > 0$. Соответствующая достаточно точная асимптотическая формула для ортогональных многочленов, которая была бы аналогична с многих точек зрения квазиклассическим формулам квантовой механики, принесла бы существенную пользу в спектральной теории.

Прогресс в изучении моделей двумерной гравитации и теории струн [12] вызвал новую волну активности в этой области в последние три года. Переведенные на язык случайных матриц, полученные там результаты описывают поведение средней допредельной плотности состояний $\rho_n(\lambda) = E\{n^{-1} \sum_1^n \delta(\lambda - \lambda_i)\}$ ансамбля (29) со специальными многочленами $V(\lambda)$ при $n \rightarrow \infty$ и λ , одновременно стремящемся к граничной точке носителя предельной плотности состояний $\rho(\lambda) = \lim_{n \rightarrow \infty} \rho_n(\lambda)$. Эти достижения выявили много красивых связей теории случайных матриц с теорией интегрируемых систем, спектральной теорией матриц Якоби, теорией ортогональных многочленов, а также поставили много задач, представляющих большой интерес как для математиков, так и для специалистов по математической физике.

4. «ИНТЕРПОЛИРУЮЩИЕ» СЕМЕЙСТВА СЛУЧАЙНЫХ ОПЕРАТОРОВ

Происхождение двух классов случайных операторов, рассмотренных в предыдущих разделах, совершенно различно, и их изучение велось независимо и различными методами. Естественно поэтому попробовать найти связи между изучаемыми классами. В настоящем разделе мы рассмотрим три семейства случайных операторов, относящиеся к первому классу. Эти семейства зависят, однако, от некоторых параметров (мы называем их радиусом взаимодействия, размерностью пространства и числом компонент соответственно) таким образом, что пределы ИПС соответствующих случайных операторов для бесконечных значений этих параметров

совпадают с возмущенным полукруговым законом, который, в свою очередь, типичен для операторов второго класса.

Первое семейство состоит из операторов $H^{(R)} = h + V^{(R)}$, действующих на $l^2(\mathbb{Z}^d)$ и определенных с помощью матриц

$$H^{(R)}(x, y) = h(x - y) + R^{-d/2} \phi((x - y)R^{-1})W(x, y), \quad x, y \in \mathbb{Z}^d, \quad (32)$$

где $h(-x) = \overline{h(x)}$, $h(x) \in l^1(\mathbb{Z}^d)$, $R < \infty$, а $\phi(t)$, $t \in \mathbb{R}^d$, — кусочно непрерывная функция с компактным носителем и $W(x, y) = W(y, x)$ — независимые (не считая условия симметричности $W(x, y) = W(y, x)$) и одинаково распределенные случайные величины, такие, что (ср. (18))

$$\mathbb{E}\{W(x, y)\} = 0, \quad \mathbb{E}\{W(x, y)W(s, t)\} = v^2(\delta_{xs}\delta_{yt} + \delta_{xt}\delta_{ys}). \quad (33)$$

Матрица (32) определяет d -мерный конечно-разностный оператор. В частности, если $h(x) = 0$, $|x| \neq 1$, $R = 1$ и $\phi(t) = 0$ для $|t| > 1$, то это оператор второго порядка. Он отличается, однако, от дискретного оператора Шрёдингера (6) тем, что содержит случайные внедиагональные элементы.

В терминологии книги [3] случайный оператор $H^{(R)}$, задаваемый равенством (32), метрически транзитивен. В частности, он допускает неслучайную ИПС $N^{(R)}(\lambda)$, заданную равенствами (3) и (4), где $H_{\Lambda}^{(R)}$ представляет собой ограничение оператора $H^{(R)}$ на конечный параллелепипед $\Lambda \subset \mathbb{Z}^d$. Недавно нам удалось доказать [15], что существует предел $\lim_{R \rightarrow \infty} N^{(R)}(\lambda)$ и что его преобразование Стильтеса удовлетворяет уравнению (24), в котором функция $r_0(z)$ является тем же преобразованием ИПС для оператора Тёплица $h(x - y)$ и ее можно выразить через символ этого оператора [1, 3]. Таким образом, $\lim_{R \rightarrow \infty} N^{(R)}(\lambda)$ совпадает с возмущенным полукруговым законом. В частности, если неслучайная часть в (32) отсутствует (т.е. $h(x) \equiv 0$), то этот предел совпадает с полукруговым законом (20), (21).

Второе семейство случайных операторов $H^{(d)}$ также действует на $l^2(\mathbb{Z}^d)$ и задается случайной матрицей

$$H^{(d)}(x, y) = h_d(x - y) + d^{-1/2} \delta(|x - y| - 1)W(x, y), \quad (34)$$

где

$$h_d(x) = d^{-1/2} \sum_{i=1}^d h_1(x_i - y_i) \prod_{k \neq i} \delta(x_k - y_k), \quad (35)$$

$h_1(x) \in l_1(\mathbb{Z})$ и случайные величины $W(x, y)$ такие же, как в (33). Этот оператор также принадлежит классу метрически транзитивных операторов, в частности, он допускает ИПС $N^{(d)}(\lambda)$. Устремляя d к бесконечности, мы вновь получаем предельное распределение собственных значений, совпадающее с деформированным полукруговым законом. Невозмущенная часть (35) семейства (34) имеет более специальный вид, чем в семействе (32). Этот специальный вид обеспечивает существование предела при $d = \infty$ для невозмущенной ИПС, которая в данном случае имеет гауссову форму.

Третье семейство операторов $H^{(n)}$ действует на $l^2(\mathbb{Z}^d) \otimes C_n$ и задается случайными «блочными» матрицами

$$H^{(n)}(\alpha, x; \beta, y) = h(x - y)\delta_{\alpha\beta} + N^{-1/2}\delta(x - y)W_{\alpha\beta}(x), \quad (36)$$

где $\alpha, \beta = 1, \dots, n$, $x, y \in \mathbb{Z}^d$, функция $h(x)$ такая же, как в (32), а случайные функции $W_{\alpha\beta}(x)$ независимы (не считая условия симметричности $W_{\alpha\beta}(x) = W_{\beta\alpha}(x)$) и одинаково распределены, причем (ср. (18) и (32))

$$E\{W_{\alpha\beta}(x)\} = 0, \quad E\{W_{\alpha\beta}(x)W_{\gamma\delta}(y)\} = v^2\delta(x - y)(\delta_{\alpha\gamma}\delta_{\beta\delta} + \delta_{\alpha\delta}\delta_{\beta\gamma}).$$

ИПС $N^{(n)}(\lambda)$ операторов $H^{(n)}$ определяется как предел последовательности $(n|\Lambda|)^{-1}\mathcal{N}_{\Lambda, n}(\lambda)$, где $\mathcal{N}_{\Lambda, n}(\lambda)$ — считающая функция для собственных значений ограничения $H_{\Lambda}^{(n)}$ семейства (36) на конечную область $\Lambda \subset \mathbb{Z}^d$.

Размер блока n при этом стремится к бесконечности. В результате ИПС опять оказывается возмущенным полукруговым законом.

Операторы $H^{(n)}$ были введены Вегнером [16], который вывел последний результат на уровне строгости теоретической физики с помощью возмущения случайной части семейства (36).

Мы вновь видим, что возмущенный полукруговой закон — довольно общий и универсальный вид предельного распределения собственных значений.

Результаты этого раздела доказаны в [15] с помощью подхода, набросок которого приведен в предыдущем разделе. Основной технический результат, используемый при этом, состоит в том, что при $R, d, n \rightarrow \infty$ моменты $E\{\prod_{i=1}^l G(x_i, y_i; z_i)\}$ принимают вид $\prod_{i=1}^l E\{G(x_i, y_i; z_i)\}$ для любых фиксированных $l \geq 1$, $x_i, y_i \in \mathbb{Z}^d$ и $|\operatorname{Im} z_i| \geq \eta$, $i = 1, \dots, l$, при достаточно большом η . Здесь $G(x, y; z) = (H^{(a)} - z)^{-1}(x, y)$ при $a = R, d$ и $G(x, y; z) = n^{-1} \sum_{\alpha=1}^n (H^{(n)} - z)^{-1}(\alpha, x; \alpha, y)$ для $a = n$.

Для доказательства этого результата мы строим некоторую бесконечную систему уравнений на указанное семейство моментов и показываем, что у построенной системы имеется единственное решение, которое при $R, d, n = \infty$ можно записать в виде произведения. Функцию $G(x - y; z) = \lim_{R, d, n \rightarrow \infty} E\{G(x, y; z)\}$, входящую в это произведение, можно найти как единственное решение некоторого нелинейного интегрального уравнения. Условие его разрешимости приводит к равенству (24).

Читатель, знакомый со статистической механикой, обнаружит близость описанных выше результатов с предельным переходом к бесконечному радиусу взаимодействия или к бесконечной размерности пространства, широко используемым в статистической механике для обоснования приближения среднего поля. Таким образом, наши предельные результаты можно считать аналогами предельных результатов в статистической механике.

ЛИТЕРАТУРА

- [1] Grenander U., Szegő G. Toeplitz Forms and Their Applications, Univ. California Press, Berkeley, 1958.
- [2] Carmona R., Lacroix J. Spectral Theory of the Random Schrödinger Operator, Birkhäuser, Basel, 1990.
- [3] Pastur L. A., Figotin A. L. Spectra of Random and Almost Periodic Operators, Grundlehren der Math. Wiss., Springer-Verlag, Heidelberg, 1992.
- [4] Mehta M. L. Random Matrices, Academic Press, N. Y., 1967.
- [5] Гирко В. Л. Спектральная теория случайных матриц. М.: Наука, 1988.
- [6] Пастур Л. А. Спектры случайных самосопряженных операторов. — УМН, т. 28, 1973, вып. 1, с. 3–64.
- [7] Bessis D., Itzykson C., Zuber J. Quantum fields theory techniques, Adv. App. Math., 1 (1980), No. 2, 109–157.
- [8] Kus M., Lewenstein M., Haake F. Density of eigenvalues of random band matrices, Phys. Review, A44 (1991), 2800–2808.
- [9] Борачев Л. В., Молчанов С. А., Пастур Л. А. О плотности состояний случайных ленточных матриц, Матем. заметки, 50 (1991), № 6, 31–42.
- [10] Хорунжий А. М., Молчанов С. А., Пастур Л. А. Распределение собственных значений случайных ленточных матриц при стремлении их порядка к бесконечности. — Теор. матем. физика, 90 (1992), № 2, 163–178.

- [11] Khorunzhy A., Khoruzhenko B., Pastur L., Shcherbina M. The large- n limit in statistical mechanics and spectral theory of disordered systems, in: Phase Transitions and Critical Phenomena, C. Domb and J. L. Lebowitz (eds.), vol. 15, Academic Press, London, 1992, 74–239.
- [12] Brezin E., Kazakov V. Exactly solvable field theories of closed strings, Phys. Lett., B236, 2 (1990), 144–150; Douglas M., Shenker S. Nucl. Phys., B335 (1990), 635; Gross D., Migdal A. Nonperturbative solution of the Ising model on a random surface, Phys. Rev. Lett., 64, 7 (1990), 717–720.
- [13] Lubinsky D. S. Strong asymptotics for extremal errors and polynomials associated with Erdős-type weights, Pitman Research Notes in Math., vol. 202, Longmans, Harlow, 1989.
- [14] Pastur L. On the universality of the level spacing distribution for some ensembles of random matrices, Lett. Math. Phys., 25 (1992), 259–265.
- [15] Khorunzhy A., Pastur L. Limits of infinite interaction radius, dimensionality and the number of components for random operators with off-diagonal randomness, Comm. Math. Phys., 153 (1993), № 3, 605–646.
- [16] Wegner F. Disordered system with n orbitals per site: $n = \infty$ limit, Phys. Rev., B19 (1979), 783–792.

Примечания

Отметим

некоторые результаты, полученные в ходе изучения статистики случайных матриц с момента публикации настоящей работы (1992 г).

1) В работе [a] (см. также [b]) дано доказательство существования ИПС ансамбля (29), в котором функция V является локально липшицевской и растет на бесконечности не медленнее чем $(2 + \varepsilon) \log |t|$, $\varepsilon > 0$. Эта ИПС абсолютно непрерывна и ее плотность (плотность состояний) может быть найдена либо как единственное решение нелинейного интегрального уравнения, либо как единственный минимум (распределение зарядов) электростатической энергии линейных зарядов во внешнем поле V .

2) В работе [c] с использованием результатов из [a] и [b] получены асимптотические формулы для ортогональных многочленов на всей оси с весом e^{-nV} , где V есть вещественно-аналитическая функция (см. также [d], где такие формулы получены для случая, когда V есть многочлен четвертого порядка). С помощью этих формул в [c, d] установлено одно из наиболее фундаментальных свойств распределения собственных значений случайных матриц: универсальность (независимость от вида V) их локальной статистики (см. [4] и [e], где дана постановка задачи). В работе [f] свойство универсальности установлено для более широкого класса функций V , принадлежащих C^3_{loc} . Это доказательство не использует указанных

асимптотических формул, но существенно опирается на ряд общих соотношений теории случайных матриц, записанных с использованием ортогональных полиномов. Некоторые мотивированные этим доказательством результаты об асимптотических свойствах ортогональных полиномов с весом e^{-nV} , $V \in C^3_{loc}$, представлены в [g].

Литература к примечаниям

- [a] Boutet de Monvel A., Pastur L., Shcherbina M. J. Stat. Phys., 79 (1995), 585–611.
- [b] Deift P., Kriecherbauer T., McLaughlin K. J. Approx. Theory, 95 (1998), 388–475.
- [c] Deift P., Kriecherbauer T., McLaughlin K., Venakides S., Zhou X. Comm. Pure Appl. Math., 52 (1999), 1491–1552.
- [d] Bleher P., Its A. Ann of Math., 150 (1999), 185–266.
- [e] Pastur L. In: Algebraic and Geometric Methods in Mathematical Physics, eds. A. Boutet de Monvel and V. Marchenko, Kluwer, Dordrecht, 1996, 207–247.
- [f] Pastur L., Shcherbina M. J. Stat. Phys., 86 (1997), 109–147.
- [g] Albeverio S., Pastur L., Shcherbina M. Математическая физика, анализ, геометрия, 4 (1997), 263–277.

ЛЕОНИД ПАСТУР

Отдел математики
Института физики
низких температур
Академии наук Украины
310164, Харьков, Украина

О НЕКОТОРЫХ НЕДАВНИХ ДОСТИЖЕНИЯХ ТЕОРИИ СОБСТВЕННЫХ ВЛОЖЕНИЙ МИНИМАЛЬНЫХ ПОВЕРХНОСТЕЙ В $\mathbb{R}^3$

Гарольд Розенберг

В последние десять лет наблюдался значительный прогресс в понимании минимальных поверхностей в трехмерных многообразиях. Я хотел бы обсудить небольшую часть работ, в которых рассматриваются собственно вложенные минимальные поверхности (далее мы будем называть их также m -поверхностями) в $\mathbb{R}^3$.

До 1982 г. единственными примерами таких поверхностей были периодические минимальные поверхности, а также катеноид и плоскость, причем получили мы их от прошлого века: геликоид, поверхность Шерка, римановы поверхности, поверхности Шварца и т.д. Мы называем m -поверхность периодической, если она инвариантна относительно нетривиальной дискретной группы изометрий, свободно действующей в пространстве $\mathbb{R}^3$. Рассматриваемые поверхности всегда будут предполагаться связными, если явно не оговорено противное. Через $C(M)$ обозначим полную кривизну поверхности M , т.е. $C(M) = \int_M K$, где K — гауссова кривизна поверхности M .

В 1982 г. С. Коста выписал формулу для полной минимальной поверхности с $C(M) = -12\pi$, моделируемой тором с тремя выколотыми точками, которая по его предположению являлась вложенной [Cost.-1,2]. Д. Хоффман и У. Микс смогли *посмотреть* на эту поверхность с помощью компьютера и, используя соображения симметрии, доказали, что поверхность Косты вложенная (Джеймс Хоффман построил графики). Впоследствии были построены семейства m -поверхностей конечной полной кривизны [Н.-М.-2] (см. рис. 1, 2).

Все известные сегодня примеры m -поверхностей в $\mathbb{R}^3$ — это либо периодические поверхности, либо поверхности конечной полной кривизны. Одна из важных открытых проблем — выяснить, существуют ли другие примеры.

Мы обсудим некоторые главные результаты о m -поверхностях конечной полной кривизны — теоремы Р. Шёна [Sch.-1] и Лопе-

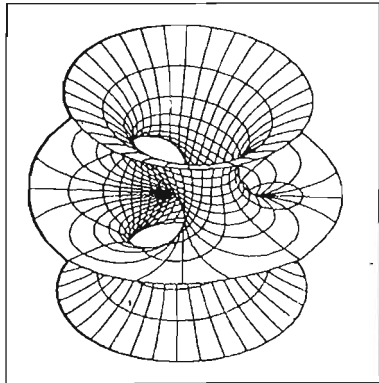
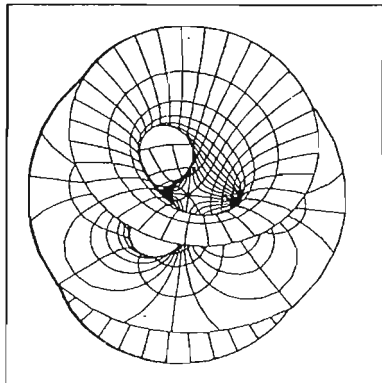


Рис. 1.

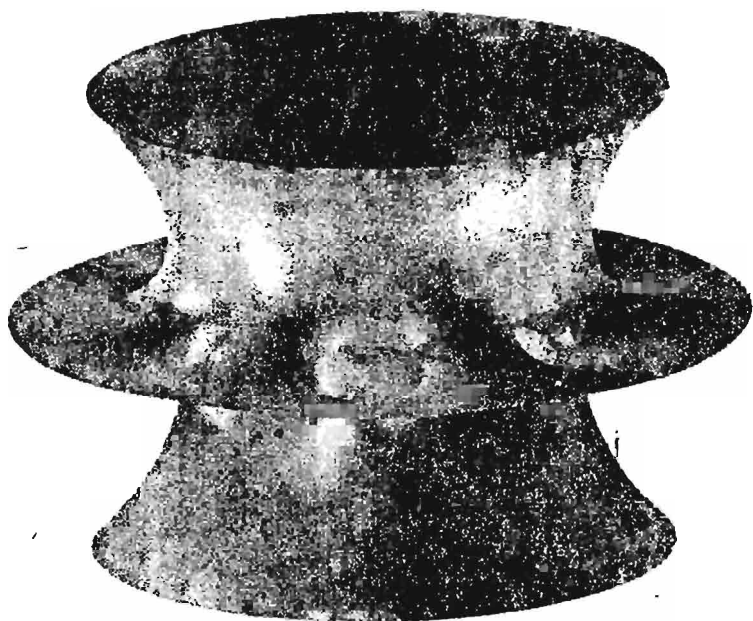


Рис. 2.

са—Роса [Lo.-Ros]. Каждая из этих теорем характеризует катеноид в классе m -поверхностей конечной полной кривизны. В теореме Шёна предполагается, что поверхность имеет в точности два конца, а в теореме Лопеса—Роса — что род поверхности равен нулю.

Мы обсуждаем оценку кривизны для устойчивых минимальных поверхностей, выполненную Гейнцем для графиков, а в общем случае полученную Р. Шёном. Мы показываем, каким образом оценка кривизны используется для построения устойчивых пределов поверхностей наименьшей площади, и рассматриваем приложения.

Мы обсуждаем теорему о кольцевых концах и сильную теорему о полупространстве Хоффмана—Микса. Последний результат означает, что две собственно погруженные непересекающиеся минимальные поверхности в $\mathbf{R}^3$ являются плоскостями; это очень полезный результат.

Мы обсуждаем работы Микса и автора, посвященные гипотезе о конечной полной кривизне: m -поверхности в $\mathbf{R}^3$ конечной топологии, обладающие по крайней мере двумя концами, имеют конечную полную кривизну. Следствием из наших работ является тот факт, что такие поверхности имеют конечный конформный тип.

Мы также обсуждаем исследования Микса и автора по периодическим минимальным поверхностям. Здесь главным результатом является тот факт, что конечность топологии факторповерхности влечет за собой конечность полной кривизны этой факторповерхности. В этом случае (фактор)поверхность параметризуется мероморфными данными на компактной римановой поверхности (представление типа Вейерштрасса).

Эти теоремы доставляют нам топологические и геометрические препятствия к существованию таких поверхностей. Например, число концов таких поверхностей всегда не меньше двух (исключая плоскость). Если поверхность является двоякопериодической и ориентируемой (при факторизации), то число ее концов не меньше четырех. Это топологические препятствия; геометрические мы обсудим в разд. VII. Например, если все концы поверхности не параллельны (как у двоякопериодической поверхности Шерка), то группа G соизмерима. Это означает, что существуют два независимых элемента G одинаковой длины.

Мы докажем, что плоскость и геликоид — единственные односвязные m -поверхности в $\mathbf{R}^3$, обладающие бесконечной группой симметрии.

Кроме того, мы обсудим суммы минимальных поверхностей и некоторые приложения.

Топологические препятствия к реализации полной ориентируемой некомпактной поверхности в виде m -поверхности в $\mathbf{R}^3$ неизвестны.

Наконец, мы обсудим некоторые проблемы, гипотезы и близкие результаты.

Я решил не обсуждать конструкции замечательных примеров Косты, Кархера, Хоффмана и Микса. Их влияние на рассматриваемую теорию было огромным, и, как сказал Г. Кархер: «Что за волшебная картина конформного отображения». Я хотел бы поблагодарить Дэвида Хоффмана и Германа Кархера за их работы и их вдохновляющее воздействие. Некоторые из компьютерных графиков были сделаны в Лаборатории графики, геометрии и анализа университета Массачусетса Джимом Хоффманом, Эдом Тайером и Фушенгом Веем. Другие компьютерные графики построили Герман Кархер и Конрад Полтьер, работающие с SFB256 в Бонне. Я благодарен им всем. Большую помощь при подготовке этой рукописи оказали Герман Кархер и Катрин Вендланд. Я благодарю их обоих.

Статья организована следующим образом.

1. Как построены классические примеры.
2. Представление Вейерштрасса и геометрия концов минимальных поверхностей в $\mathbf{R}^3$ конечной полной кривизны.
 - 2.1. Параметризация Оссермана поверхностей конечной полной кривизны.
 - 2.2. Геометрия концов конечной полной кривизны.
3. Характеризации катеноида в работах Р. Шёна и Лопеса—Роса.
 - 3.1. Теорема Шёна.
 - 3.2. Теорема Лопеса—Роса.
 - 3.3 и 3.4. Принцип максимума на бесконечности.
 - 3.5. Формула монотонности.
4. Оценки кривизны для устойчивых минимальных поверхностей.
 - 4.1. Критерий устойчивости Барбосы—До Кармо.
 - 4.2. Идея доказательства теоремы Гейнца.
5. Компактность семейств наименьшей площади и построение дополнительных поверхностей конечной полной кривизны.
6. Теорема о кольцевых концах и сильная теорема о полупространстве Хоффмана—Микса.
 - 6.1. Теорема о кольцевых концах.

- 6.2, 6.3, 6.4 и 6.5. Теорема о конечности конформного типа и ее следствия.
- 6.6. Сильная теорема о полупространстве.
7. Двойкопериодические минимальные поверхности.
- 7.1, 7.2 и 7.3. Теорема о конечной полной кривизне для двойкопериодических минимальных поверхностей.
- 7.4. Формула полной кривизны.
- 7.5, 7.6 и 7.7. Глобальные топологические и геометрические свойства.
- 7.8, 7.9, 7.10 и 7.11. Сумма минимальных поверхностей и приложения.
8. Однопериодические минимальные поверхности.
- 8.1. Теорема о конечной полной кривизне для однопериодических минимальных поверхностей.
- 8.2, 8.3, 8.4. Обобщенное представление Вейерштрасса.
- 8.5. Геометрия концов конечной полной кривизны.
- 8.6. Число вращения конца.
- 8.7 -- 8.12. Приложения теоремы о конечной полной кривизне.
9. Некоторые задачи, гипотезы и смежные вопросы.

1. КАК ПОСТРОЕНЫ КЛАССИЧЕСКИЕ ПРИМЕРЫ

Пусть Γ — ломаная жорданова кривая в $\mathbf{R}^3$ и M_0 — компактная минимальная поверхность с границей $\partial M_0 = \Gamma$. Принцип отражения Шварца [Oss.-1] позволяет нам продолжить M_0 через каждое ребро ломаной Γ , поворачивая поверхность M_0 вокруг каждого ребра на π . Выполнив отражение относительно каждого ребра, ограничивающего поверхность, мы получим минимальную поверхность M , которая, вообще говоря, будет иметь особенности и самопересечения, т. е. полученная поверхность M будет погруженной.

Если угол, под которым сходятся два ребра ломанной Γ , иррационален, то поверхность M будет бесконечное число раз поворачиваться вокруг этой вершины ломанной. Поэтому если мы хотим получить пример вложенной поверхности, то контур Γ следует выбирать так, чтобы углы при вершинах имели вид $2\pi/n$. Поверхность M_0 при этом должна быть вложенной. Например, с помощью многоугольников, изображенных на рис. 3 и 4, можно построить примеры собственно вложенных поверхностей, которые инвариантны относительно трех независимых сдвигов (тройкопериодические поверхности).

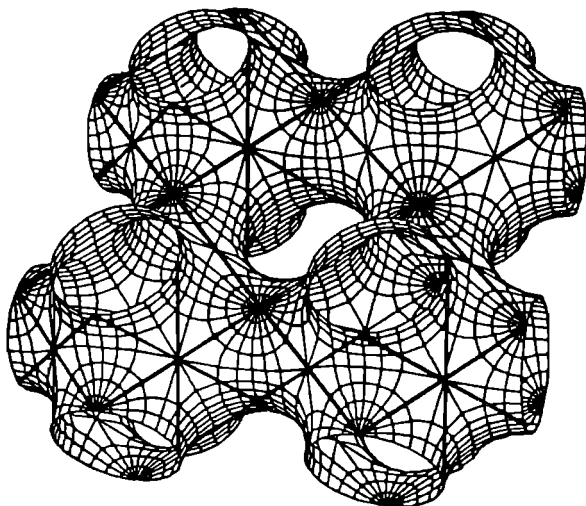


Рис. 3.

Риман, Шварц и Вейерштрасс нашли минимальные поверхности M_0 с условием $\partial M_0 = \Gamma$, явно задавая конформные отображения (в соответствии с теорией Римана) и представления Вейерштрасса (это хорошо изложено в книге Дарбу [Darb.]). Сегодня мы находим M_0 с помощью другой техники. Дуглас и Радо доказали, что любая спрямляемая жорданова кривая Γ в $\mathbb{R}^n$ ограничивает некоторый диск M_0 наименьшей площади [Doug., Rado-1, 2]. Впоследствии Р. Оссерман доказал, что M_0 не имеет геометрических точек ветвления (т. е. диск наименьшей площади, ограниченный кривой Γ , является погруженным [Oss.-1]). Наконец, Райфенберг при помощи геометрической теории меры доказал, что всегда существует такая вложенная минимальная поверхность M_0 , что $\partial M_0 = \Gamma$ [Reif.].

Итак, если контур Γ и поверхность M_0 выбраны подходящим образом, то полная поверхность M , полученная отражением M_0 относительно каждого ограничивающего M_0 ребра, будет вложенной троякопериодической поверхностью. Ее фактор по группе G , порожденной тремя независимыми сдвигами, будет компактной минимальной поверхностью конечного рода, вложенной в плоский тор $\mathbb{R}^3/G$. Геометрия и топология этой поверхности были изучены Миксом [М.-3] и Кархером [К.-3, К.-4].

Большинство других известных примеров поверхностей бесконечной полной кривизны (однопериодические и двоякопериоди-

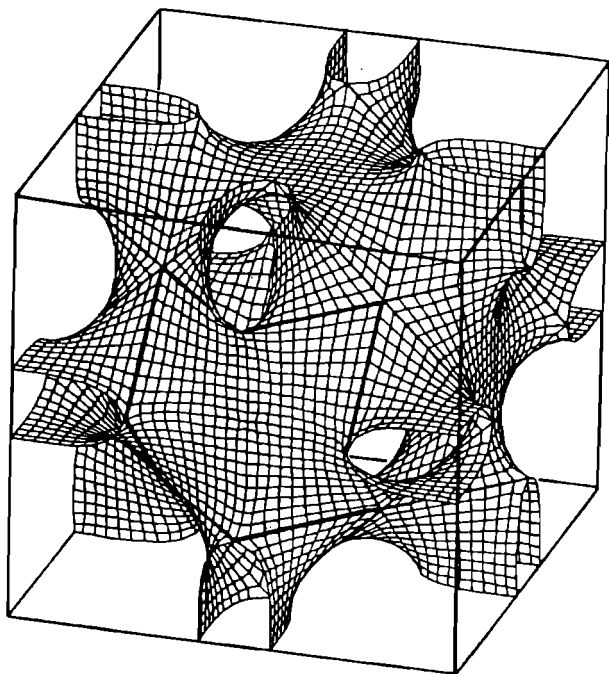


Рис. 4.

ческие поверхности) строятся при помощи некомпактного многоугольника Γ и полной вложенной минимальной поверхности M_0 с границей Γ . Снова произведя все возможные отражения поверхности M_0 относительно ребер Γ (и их образов), построим полную минимальную поверхность M . Далее возникает вопрос: как найти поверхность M_0 , если контур Γ бесконечен? Существует общая теория (теорема Дженкина—Серрина [J.-S.] и сопряженная конструкция Плато [К.-1]), но я предпочитаю вместо обсуждения этой теории описать, как можно непосредственно получить несколько примеров.

Сначала построим первую поверхность Шерка, решая задачу Плато и переходя к пределу. Рассмотрим многоугольник $\Gamma(n)$, изображенный на рис. 5,а. Для каждого целого числа n выберем $\Gamma(n)$ (рис. 5,а) таким образом, что $\Gamma(n)$ проектируется на квадрат в горизонтальной плоскости, причем верхние ребра находятся на высоте n , а нижние — на высоте $-n$.

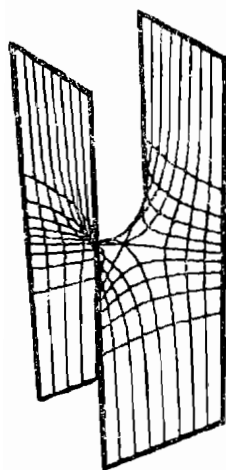


Рис. 5, а.

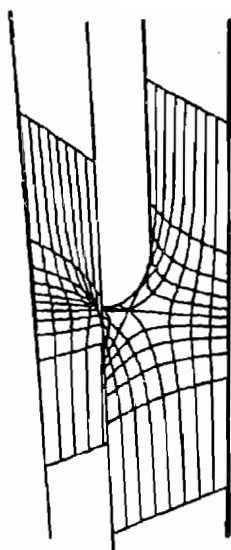


Рис. 5, б.

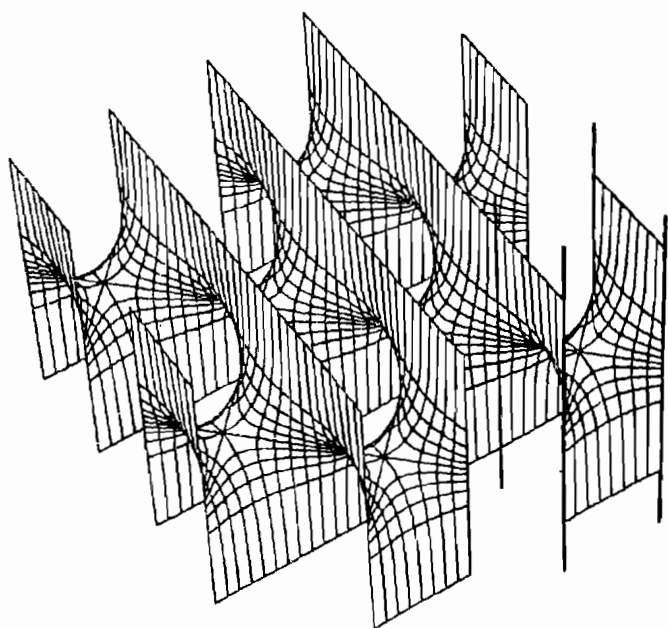


Рис. 5, в.

Обозначим через $\Sigma(n)$ диск наименьшей площади с границей $\Gamma(n)$. Несложно доказать, что $\Sigma(n)$ является графиком над квадратом в горизонтальной плоскости, в который проектируется $\Gamma(n)$. (Радо доказал более общее утверждение [М.-2]: если жорданова кривая Γ проектируется на выпуклую плоскую кривую C , то любая минимальная поверхность, ограниченная кривой Γ , является графиком над плоской областью, ограниченной кривой C .)

Поверхность $\Sigma(n)$ обладает теми же симметриями, что и $\Gamma(n)$; поэтому существует точка $p_n \in \Sigma(n)$ на высоте 0, в которой касательная плоскость к $\Sigma(n)$ горизонтальна.

Мы видим, что при $n \rightarrow \infty$ все поверхности $\Sigma(n)$ проходят через ту же точку $p_n = p$. Поэтому функции, задающие графики $\Sigma(n)$, сходятся к функции f , определенной на внутренности квадрата и принимающей значения $+\infty$ на двух противоположных сторонах квадрата и $-\infty$ на двух других. График функции f — минимальная поверхность, границей которой являются четыре вертикальные прямые, проходящие через вершины квадрата (рис. 5, б).

Применим принцип отражения Шварца, отражая график функции f относительно этих четырех вертикальных прямых и всех новых прямых, которые будут получаться. В результате мы получим минимальную поверхность M , называемую поверхностью Шерка. Поверхность M проектируется на бесконечное множество квадратов в горизонтальной плоскости, которые расположены как черные (или белые) квадраты бесконечной шахматной доски (см. рис. 5, в).

Образуя факторы этой поверхности по независимым параллельным переносам, мы получаем собственно вложенные поверхности конечной топологии (и конечной полной кривизны) в плоском многообразии $T^2 \times \mathbb{R}$, где T^2 — плоский 2-тор.

Простейший вариант этой конструкции приводит к проективной плоскости с двумя выколотыми точками. Опишем некоторые примеры.

Пусть P — квадрат, на который проектируется $\Gamma(n)$, а v_1 и v_2 — векторы, задающие стороны этого квадрата. Через $G(v_1 + v_2, v_1 - v_2)$ обозначим группу, порожденную параллельными переносами на $v_1 + v_2$, $v_1 - v_2$. Тогда поверхность M инвариантна относительно группы $G(v_1 + v_2, v_1 - v_2)$, а фактор по этой группе топологически является проективной плоскостью без двух точек; полная кривизна факторповерхности равна -2π (рис. 6, а).

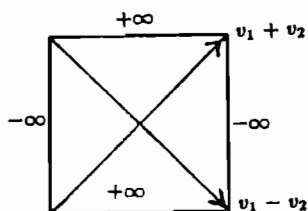


Рис. 6,а.

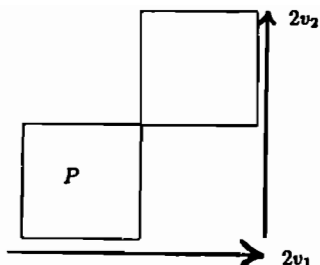


Рис. 6,б.

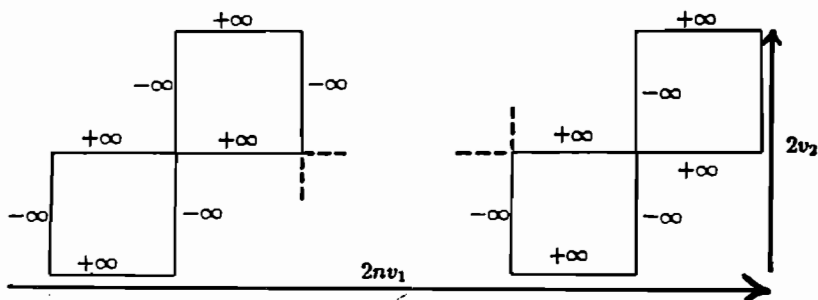


Рис. 6,в.

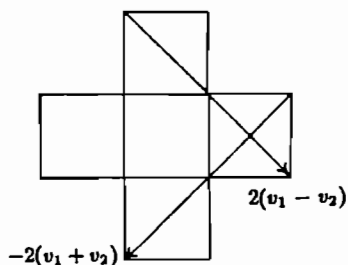


Рис. 6,д.

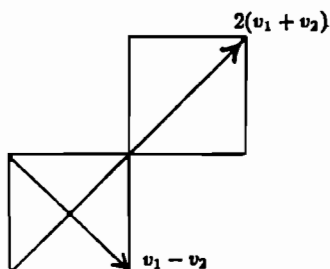


Рис. 6,е.

Фундаментальной областью для группы $G(2v_1, 2v_2)$ являются два смежных экземпляра P (рис. 6,б). Фактор поверхности M по этой группе конформно диффеоморфен сфере без четырех точек и имеет полную кривизну -4π в $T^2 \times \mathbb{R}$.

Этим способом можно реализовать сферу S^2 без любого четного числа точек. Пусть $G = G(2nv_1, v_2)$. Фундаментальная область этой группы состоит из $2n$ экземпляров квадрата P (рис. 6, *c*).

Для того чтобы получить тор без четырех точек, надо рассмотреть группу $G(2(v_1 + v_2), 2(v_1 - v_2))$. Фундаментальная область в этом случае состоит из четырех экземпляров квадрата P (рис. 6, *d*). Факторповерхность имеет четыре конца, и ее полная кривизна равна -8π .

При помощи группы $G(v_1 - v_2, 2(v_1 + v_2))$ можно получить бутылку Клейна. Фундаментальная область этой группы изображена на рис. 6, *e*. Факторповерхность имеет два конца, а ее полная кривизна равна -4π . Располагая n экземпляров квадрата P по диагонали и рассматривая группу $G = G(v_1 - v_2, n(v_1 + v_2))$, мы получим связную сумму n проективных плоскостей без двух точек. Выбирая подходящим образом ориентированное двулистное накрытие над любым из описанных неориентированных примеров, можно получить любую возможную ориентируемую поверхность без четырех точек.

Заметим, что во всех рассмотренных примерах концы минимальных поверхностей асимптотичны плоским цилиндрам, которые оказывались вертикальными. Также во всех случаях верхние концы оказывались не параллельными нижним. Существуют примеры, в которых все концы параллельны и не вертикальны [М.-Р.-1]. Г. Кархер построил наглядный пример тора без четырех точек в $T \times \mathbb{R}$, все концы которого параллельны. Эту поверхность мы будем называть седлом Кархера [К.-2].

Построение этого примера начинается с рассмотрения прямоугольника P и минимального графика над частью этого прямоугольника, ограниченной L_1, L_2, C_1 и C_2 . См. рис. 7, *a*. Будем считать, что функция равна 0 на C_1 и C_2 и $+\infty$ на L_1 и L_2 . Кроме того, график вертикален над $C_1 \cup C_2$. Тогда C_1 и C_2 — проекции на плоскость линий уровня кривизны и график может быть продолжен отражением в плоскости прямоугольника P . Новая поверхность имеет четыре граничных вертикальных линии (рис. 7, *b*).

Для того чтобы получить тор без четырех точек, расположим два экземпляра P по диагонали и рассмотрим фактор по группе $G(2v_1, 2v_2)$ (рис. 7, *c*).

Во всех рассмотренных примерах геометрия и топология факторповерхностей связаны формулой

$$C(M) = 2\pi\chi(M).$$

Это частный случай следующего общего результата.

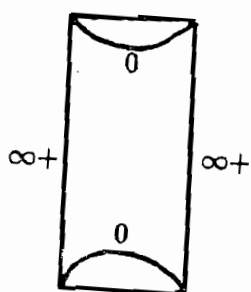


Рис. 7,а.

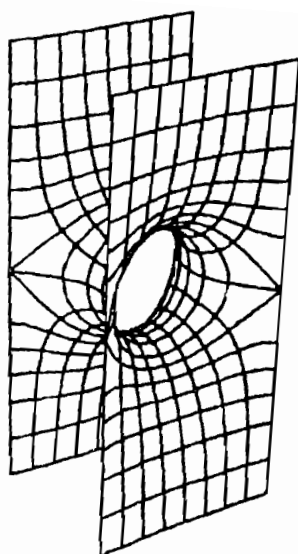


Рис. 7,б.

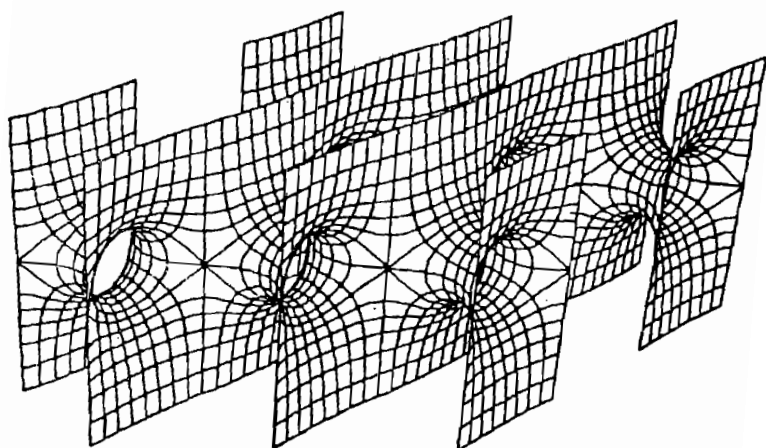


Рис. 7,с.

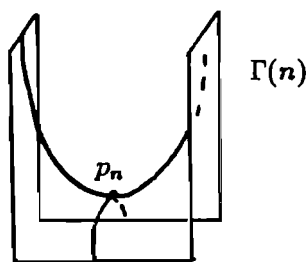


Рис. 8.

Теорема 1.1 [М.-Р.-1]. Пусть $M \subset T \times R$ — собственно вложенная минимальная поверхность конечной топологии. Тогда M имеет конечную полную кривизну, которая вычисляется по формуле

$$C(M) = 2\pi\chi(M).$$

При построении поверхности Шерка мы начинали с квадрата P , над которым рассматривался предел минимальных графиков. Если в качестве P мы выберем ромб и повторим ту же конструкцию, графики с границей $\Gamma(n)$ по-прежнему будут иметь общую точку p_n с нулевой высотой. Поэтому поверхность Шерка существует над «шахматной доской» из ромбов. Однако если в качестве P выбрать параллелограмм с неравными сторонами, то точки p_n будут всегда уходить в бесконечность и предельная поверхность будет иметь две непересекающиеся вертикальные полосы (рис. 8).

Это частный случай нашего следующего результата.

Теорема 1.2 [М.-Р.-1]. Пусть M — собственно вложенная минимальная поверхность конечной топологии в $T \times R$. Если ее концы не параллельны, то $T \times R$ имеет соизмеримую решетку и концы поверхности M вертикальны.

Под соизмеримостью решетки мы подразумеваем, что $T \times R = R^3/G$, где G имеет два линейно независимых вектора равной длины.

Существует теорема Дженкина и Серрина, которая гарантирует существование поверхности Шерка над ромбом (и, следовательно, полной поверхности, полученной отражениями относительно вертикальных прямых, проходящих через вершины). Сформулируем частный случай этого результата.

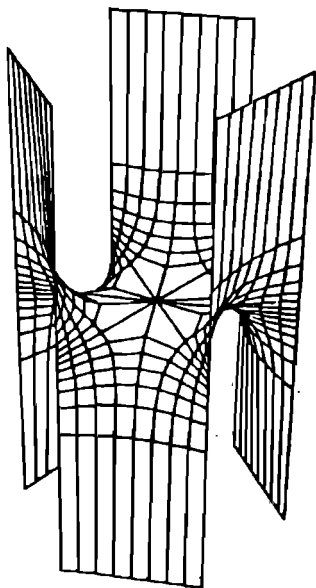


Рис. 9.

Теорема 1.3 [J.-S.]. Пусть C — ломаная жорданова кривая в плоскости, имеющая четное число сторон. Далее, пусть P — компактная плоская область, ограниченная кривой C , и φ принимает значение $+\infty$ и $-\infty$ на смежных сторонах этой кривой. Необходимое и достаточное условие для продолжения φ до минимального графика над P (принимającego лишь конечные значения) состоит в том, что сумма длин ребер кривой C , на которых φ равна $+\infty$, равна сумме длин ребер этой кривой, на которых φ равна $-\infty$.

В случае когда φ продолжается до минимального графика, этот график ограничен вертикальными прямыми, проходящими через вершины кривой C . Например, график Дженкина—Серрина над правильным шестиугольником изображен на рис. 9.

Очевидным следствием теоремы Дженкина—Серрина является то, что поверхность Шерка существует над параллелограммом только тогда, когда параллелограмм является ромбом.

Существуют обобщения теоремы Дженкина—Серрина для некомпактных областей, которые успешно применяются для построения полных поверхностей [R.-S.E., K.-1].

2. ПРЕДСТАВЛЕНИЕ ВЕЙЕРШТРАССА И ГЕОМЕТРИЯ КОНЦОВ МИНИМАЛЬНЫХ ПОВЕРХНОСТЕЙ В $\mathbf{R}^3$ КОНЕЧНОЙ ПОЛНОЙ КРИВИЗНЫ

Рассмотрим риманову поверхность M и конформное отображение $\phi : M \rightarrow \mathbf{C}^3$, удовлетворяющее условию $\sum_{i=1}^3 \phi_i^2 = 0$, $\phi = (\phi_1, \phi_2, \phi_3)$. Тогда $X(z) = \operatorname{Re} \int_{z_0}^z \phi$ является минимальной поверхностью в $\mathbf{R}^3$. Несложно показать (см., например, [Oss.-1]), что любая поверхность в $\mathbf{R}^3$, средняя кривизна которой равна нулю, локально имеет такой вид. Для того чтобы поверхность в $\mathbf{R}^3$ моделировалась поверхностью M , необходимо, чтобы интеграл от ϕ не зависел от пути на M между точками z_0 и z . Это условие называется условием периодов: для любого цикла γ на M выполнено равенство

$$\operatorname{Re} \int_{\gamma} \phi(z) dz = 0.$$

Для того чтобы M была погруженной в $\mathbf{R}^3$, требуется выполнение условия $\sum_{i=1}^3 |\phi_i(z)| \neq 0$, $z \in M$.

Соберем все сказанное в определении: минимальной поверхностью в $\mathbf{R}^3$, моделируемой римановой поверхностью M , называется конформное отображение $\phi : M \rightarrow \mathbf{C}^3$, удовлетворяющее следующим условиям:

- $\phi_1^2 + \phi_2^2 + \phi_3^2 = 0$ на M и $|\phi_1|^2 + |\phi_2|^2 + |\phi_3|^2$ не обращается в нуль;
- $\operatorname{Re} \int_{\gamma} \phi(z) dz = 0$ для любого цикла γ на M (условие периодов).

В случае, если условие периодов не выполнено, рассматривают минимальную поверхность, моделируемую универсальным конформным накрытием поверхности M (т.е. $\mathbf{C}$ или открытым единичным диском).

Образ $X(M)$ называют минимальной поверхностью, моделируемой поверхностью M , даже в случае, когда X не является вложением.

Задание трех координатных функций отображения ϕ и одного уравнения $\phi_1^2 + \phi_2^2 + \phi_3^2 = 0$ может быть сведено к двум условиям. Один из классических способов сделать это называется представлением Вейерштрасса. Предположим, что $\phi_1 - i\phi_2$ не равно тождественно нулю (это соответствует тому, что M — не плоскость), и положим

$$g = \frac{\phi_3}{\phi_1 - i\phi_2}, \quad \omega(z) = (\phi_1 - i\phi_2) dz.$$

Тогда g — мероморфная функция на M , ω — голоморфная 1-форма, а поверхность M восстанавливается по (g, ω) с помощью формулы

$$X(z) = \operatorname{Re} \int_{z_0}^z \left(\frac{\omega}{2}(1 - g^2), \frac{i\omega}{2}(1 + g^2), g\omega \right). \quad (W)$$

Это и есть представление Вейерштрасса поверхности M . Заметим, что полюсы функции g являются нулями формы ω , причем полюсу функции g порядка k соответствует нуль формы ω порядка $2k$.

Нетрудно видеть, что пара (g, ω) с указанными свойствами (функция g мероморфна на M , а форма ω голоморфна на M), удовлетворяющая условию нулей—полюсов, задает минимальную поверхность при помощи формулы (W). Конечно же, для поверхности, моделируемой поверхностью M , должно быть выполнено условие периодов.

Мероморфное отображение g имеет важный геометрический смысл: оно является гауссовым отображением поверхности M . Точнее говоря, оно является композицией обычного гауссова отображения поверхности $X(M)$ и стереографической проекции сферы единичного радиуса (с центром в начале координат) на экваториальную плоскость из северного полюса.

Геометрические инварианты поверхности M выражаются через функцию g и форму ω . Так, индуцированная метрика на M задается формулой

$$ds = \frac{|\omega|}{2}(1 + |g|^2),$$

а кривизна поверхности M — формулой

$$K = - \left[\frac{4|g'|}{|\omega|(1 + |g|^2)^2} \right]^2.$$

2.1. Параметризация Оссермана поверхностей конечной полной кривизны. Риманова поверхность M называется поверхностью конечного конформного типа, если существует такая компактная риманова поверхность $\overline{M}$, что M конформно эквивалентна поверхности $\overline{M}$ без конечного числа точек.

В теории полных минимальных поверхностей в $\mathbb{R}^3$ поверхности конечной полной кривизны занимают важное место, что хорошо поясняет следующая теорема Оссермана.

Теорема 2.2 [Oss.-1]. Пусть M — полная минимальная поверхность, погруженная в $\mathbf{R}^3$, и $|C(M)| = \int_M |K| dA < \infty$. Тогда M является поверхностью конечного конформного типа и может быть параметризована при помощи мероморфных данных на компактной римановой поверхности. Точнее говоря, если через $\bar{M}$ обозначить конформную компактификацию поверхности M (т. е. M конформно эквивалентна $\bar{M} \setminus \{p_1, \dots, p_n\}$), то представление Вейерштрасса (g, ω) поверхности M мероморфно продолжается на $\bar{M}$.

Таким образом, в некотором смысле теория минимальных поверхностей конечной полной кривизны в $\mathbf{R}^3$ является частью теории римановых поверхностей. Но мы очень далеки от понимания сути этой теории. Как увидеть поверхность M в терминах пары (g, ω) ? Когда M является вложенной? Какие поверхности M существуют?

Интересно было бы понять, какую роль играет минимальность в теореме Оссермана, так как в действительности важная часть этой теоремы от минимальности не зависит. Полные римановы двумерные многообразия конечной полной кривизны имеют конечный конформный тип. Это теорема Губера [Hub.], современное доказательство которой можно найти в [Wh.]. Трудной частью теоремы Губера является доказательство конформности, поскольку из неравенства Кон-Фоссена (напомним, что оно имеет вид $C(M) \leq 2\pi\chi(M)$ для полных двумерных многообразий M неположительной кривизны) следует, что топологический тип конечен, если конечна $C(M)$. Приняв это условие, нетрудно продолжить (g, ω) на выколотые точки. Конец A многообразия M конформно является пунктированным диском: $A = D^* = \{0 < z \leq 1\}$. Гауссово отображение g мероморфно продолжается в точку $z = 0$, поскольку полная кривизна конца A есть площадь образа A при гауссовом отображении, подсчитанная с учетом кратности. Если бы рассматриваемая точка была существенной особенностью, то почти каждое значение отображения принималось бы бесконечное число раз и сферическая площадь была бы бесконечной. Теперь повернем M так, чтобы значение функции g в точке $z = 0$ было конечным. Так как метрика на M задается формулой $\frac{|\omega|^2}{2}(1 + |g|^2)$ и эта метрика полная, $\int_\gamma |\omega| = \infty$ для любого расходящегося пути γ на A , т. е. γ сходится к центру диска D^* . Из теории функций (см., например, [Oss.-1]) следует, что ω имеет полюс в рассматриваемой точке.

2.2. Геометрия концов конечной полной кривизны. Теперь для нас не составит труда изучить геометрию конца A конечной полной кривизны. Пусть A конформно параметризуется при помощи диска D^* , и пусть (g, ω) — представление Вейерштрасса конца A . После поворота A в $\mathbb{R}^3$ можно считать, что $g(0) = 0$, так что после конформной перепараметризации некоторого подконца конца A получим, что $g(z) = z^k$. Из полноты метрики в нуле следует, что форма ω должна иметь полюс в нуле: в окрестности нуля должно выполняться равенство $\omega(z) = \left(\frac{c}{z} + O(|z|^{-l-1})\right) dz$.

Непосредственное вычисление с использованием представления Вейерштрасса (W) дает

$$2[x_1(z) - ix_2(z)] = \int_{z_0}^z \omega - \overline{\int_{z_0}^z g^2 \omega}. \quad (E)$$

Мы знаем, что форма $g^2 \omega$ имеет в точке 0 полюс меньшего порядка, чем ω ; следовательно, отображение $x_1 - ix_2$ имеет в этой точке полюс порядка l . Рассмотрим образ окружности $re^{i\theta}$, $0 \leq \theta \leq 2\pi$, $r > 0$, где r — малое положительное число, при действии отображения $x_1 - ix_2$. Этот образ $l - 1$ раз оборачивается вокруг оси x_3 . Поскольку эта кривая должна быть замкнутой (т. е. A — кольцо), мы получаем, что $l > 1$. В действительности коэффициент при $1/z$ в форме ω должен быть равен нулю, поскольку отображения x_1 и x_2 однозначны на A .

Если конец A является вложенным, то кривая $(x_1 - ix_2)(re^{i\theta})$, $0 \leq \theta \leq 2\pi$, один раз оборачивается вокруг оси x_3 ; следовательно, $l = 2$ и в окрестности точки 0 имеем

$$\omega = \left(\frac{c}{z^2} + h(z)\right) dz, \quad \text{где } h \text{ — голоморфная функция.}$$

Далее, $x_3 = \operatorname{Re} \int g \omega$ и $g(z) = z^k$ в окрестности точки 0; следовательно, если отображение $|x_3|$ ограничено, то $k \geq 2$; это случай плоскостных концов. Заметим, что если $k = 1$, то коэффициент c действительный, поскольку координата x_3 корректно определена на конце.

Интегрируя ϕ_3 , мы получим следующее разложение x_3 (входящие в формулу константы — это постоянные интегрирования):

$$x_3(z) = \begin{cases} c \ln |z| + c_0 + O(|z|^2), & \text{если } k = 1, \\ d_0 + O(|z|), & \text{если } k \neq 1. \end{cases}$$

Из уравнения (Е) мы получаем

$$|z| = \frac{|c|}{2|x|} + O(|x|^{-2}).$$

Подставляя это выражение в формулу для нахождения $x_3(z)$, получим

$$x_3 = a \ln |x| + b + \frac{c_1 x_1 + c_2 x_2}{|x|^2} + O(|x|^{-2}).$$

Коэффициенты этого уравнения — действительные числа ($a = -c$).

Таким образом, вложенный конец, имеющий конечную полную кривизну, асимптотичен плоскостному или катеноидному концу.

3. ХАРАКТЕРИЗАЦИИ КАТЕНОИДА В РАБОТАХ Р. ШЁНА И ЛОПЕСА—РОСА

Для класса минимальных поверхностей конечной полной кривизны в $\mathbf{R}^3$ существуют две фундаментальные теоремы, каждая из которых характеризует катеноид.

Теорема 3.1 [Sch.-1]. *Пусть M — полная минимальная поверхность конечной полной кривизны, погруженная в $\mathbf{R}^3$, с двумя концами, каждый из которых является вложенным. Тогда M есть катеноид.*

Замечание. Мы убедимся, что поверхность M является вложенной; это следует непосредственно из формулы монотонности.

Теорема 3.2 [Lo.-Ros, P.-Ros]. *Пусть M есть m -поверхность конечной полной кривизны и нулевого рода в $\mathbf{R}^3$. Тогда M является плоскостью или катеноидом.*

Сделаем краткое отступление. Утверждение теоремы Шёна весьма неожиданно. Почему нельзя присоединить к катеноиду ручку (рис. 10)?

С этим вопросом связана следующая недоказанная гипотеза У. Микса. Пусть C_1 и C_2 — выпуклые кривые, лежащие в параллельных плоскостях, и пусть M — такая компактная связная минимальная поверхность, что $\partial M = C_1 \cup C_2$. Тогда M имеет нулевой род.

Как показывает следующий пример, выпуклость необходима. Пусть M_1, M_2 — две части катеноидов, размещенные как на рис. 11, а.

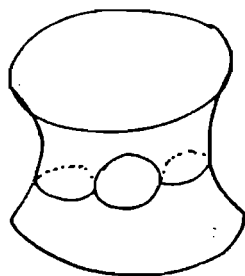


Рис. 10.

Границы поверхностей M_1 и M_2 лежат в параллельных плоскостях. Соединим нижние и верхние граничные окружности узкими перемычками, «мостами» (см. рис. 11, б). По известному принципу моста (см. [Cour., Smale]) существует минимальная поверхность M , заполняющая мосты и близкая к M_1 и M_2 (вблизи M_1 и M_2).

Теперь мы можем приступить к доказательству теорем 3.1 и 3.2. В качестве полезного средства можно использовать принцип максимума на бесконечности. Из обычного принципа максимума следует, что расстояние между двумя непересекающимися минимальными поверхностями, собственно погруженными в $\mathbb{R}^3$, не может реализоваться как расстояние между точками $p_1 \in \text{int}(M_1)$ и $p_2 \in \text{int}(M_2)$,

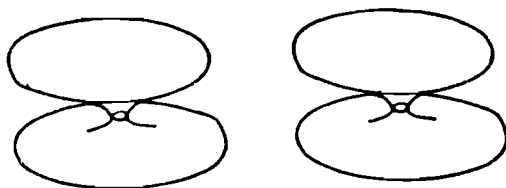


Рис. 11, а.

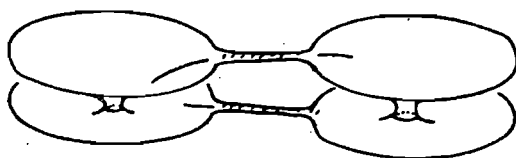


Рис. 11, б.

если только M_1 и M_2 не являются параллельными плоскостями. А что же происходит, если M_1 и M_2 асимптотичны на бесконечности?

Теорема 3.3 (принцип максимума на бесконечности [L.-R, M.-R.-2]). Пусть M_1 и M_2 — непересекающиеся минимальные поверхности с компактной границей, собственно погруженные в полное плоское трехмерное многообразие. Если $\partial M_1 = \partial M_2 = \emptyset$, то поверхности M_1 и M_2 плоские. В противном случае имеет место формула

$$\text{dist}(M_1, M_2) = \min\{\text{dist}(M_1, \partial M_2), \text{dist}(M_2, \partial M_1)\}.$$

Случай $\partial M_1 = \partial M_2 = \emptyset$ составляет содержание сильной теоремы о полупространстве Хоффмана—Микса. Он не рассматривается в доказательствах теорем Шёна и Лопеса—Роса. Сильная теорема о полупространстве будет рассмотрена в разд. 6. В настоящий же момент нам понадобится частный случай принципа максимума на бесконечности, впервые доказанный Лангевинем и Розенбергом (и мы приведем его доказательство).

Теорема 3.4 [L.-R.]. Пусть M_1 и M_2 — непересекающиеся минимальные поверхности конечной полной кривизны, вложенные в $\mathbb{R}^3$, с компактными границами. Тогда $\text{dist}(M_1, M_2) > 0$.

При доказательстве этой теоремы используется понятие потока на минимальной поверхности M .

Пусть α — ориентированный цикл на M , и пусть J — оператор комплексной структуры поверхности M (J — поворот в каждой касательной плоскости к M на угол $\pi/2$). Тогда поток цикла α определяется по формуле

$$\text{Flux}(\alpha) = \int_{\alpha} J(\alpha'),$$

где α' — единичный касательный вектор к α ; тогда $J(\alpha')$ — конормальное поле к M вдоль α , т. е. единичная нормаль к α , касательная к M .

Поскольку координатные функции гармоничны на M , непосредственное применение теоремы о дивергенции показывает, что $\text{Flux}(\alpha)$ зависит только от класса гомологий цикла α . Таким образом, поток можно рассматривать как $\mathbb{R}^3$ -значную функцию на гомологиях поверхности M .

Во втором разделе настоящей статьи установлено, что конец конечной полной кривизны минимальной поверхности, вложенный в $\mathbf{R}^3$, имеет предельный нормальный вектор (который мы здесь предполагаем вертикальным), и этот конец можно представить как график функции $u(x)$ для больших $|x|$, где

$$u(x) = a \ln |x| + b + \frac{c_1 x_1 + c_2 x_2}{|x|^2} + O(|x|^{-2}).$$

Мы говорим, что конец является концом катеноидного типа, если $a \neq 0$ (a — скорость логарифмического роста конца), и что конец плоскостной, если $a = 0$. В первом случае конец геометрически асимптотичен катеноиду, а во втором — горизонтальной плоскости.

Разложение функции u можно продифференцировать почленно, откуда легко получить, что направленный наружу вектор, конормальный к концу вдоль кривой $C_R = \{(x, y, u(x, y)) \mid x^2 + y^2 = R^2\}$, определяется по формуле

$$\nu = \frac{1}{R}(x, y, a) + O(|R|^{-2}).$$

Следовательно, $\text{Flux}(C_R) = (0, 0, 2\pi a) + O(|R|^{-1})$.

Поскольку поток зависит только от класса гомологий кривой C_R , мы получаем $\text{Flux}(C_R) = (0, 0, 2\pi a)$. В частности, вектор Flux зависит только от скорости логарифмического роста конца и параллелен предельному нормальному вектору.

Теперь мы можем доказать теорему 3.4. Допустим, что M_1 и M_2 — вложенные непересекающиеся минимальные поверхности конечной полной кривизны, имеющие компактные границы и такие, что $\text{dist}(M_1, M_2) = 0$, и покажем, что это предположение приведет к противоречию.

Поскольку конечность полной кривизны поверхностей M_1 и M_2 влечет за собой конечность топологического типа, каждая из этих поверхностей имеет конечное число концов и каждый конец топологически представляет собой кольцо. Так как $\text{dist}(M_1, M_2) = 0$ и поверхности M_1, M_2 имеют компактные границы, должны существовать такой конец E_1 поверхности M_1 и такой конец E_2 поверхности M_2 , что $\text{dist}(E_1, E_2) = 0$. После поворота в $\mathbf{R}^3$ мы можем допустить, что E_1 и E_2 асимптотичны одному и тому же горизонтальному катеноидному концу со скоростью роста a (а если $a = 0$, то горизонтальной плоскости).

Поскольку $E_1 \cap E_2 = \emptyset$, мы можем допустить, что E_1 лежит выше, чем E_2 (можно считать E_1 и E_2 графиками). Незначитель-

но сдвинув E_1 вертикально вниз, можно получить такое множество E'_1 , что $\partial E'_1$ по-прежнему лежит над E_2 , но вне большого шара E'_1 лежит ниже, чем E_2 . Следовательно, $E'_1 \cap E_2$ представляет собой компактное непустое одномерное аналитическое подмножество как множества E'_1 , так и множества E_2 .

Теперь покажем, что $E'_1 \cap E_2$ есть простая замкнутая кривая γ , гомотопически нетривиальная на E'_1 и E_2 , и что E'_1 трансверсально E_2 вдоль γ . Поскольку E'_1 представляет собой график, образ множества $E'_1 \cap E_2$ при проекции $\pi: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ есть компактное непустое одномерное аналитическое многообразие в $\mathbf{R}^2$. Пусть D — такой диск в $\mathbf{R}^2$, что E'_1 является графиком над $\mathbf{R}^2 \setminus D$. Если предположить, что $\pi(E'_1 \cap E_2)$ не является связной гомотопически нетривиальной простой замкнутой кривой в $\mathbf{R}^2 \setminus D$, то множество $\mathbf{R}^2 \setminus \pi(E'_1 \cap E_2)$ содержит компактную компоненту, не пересекающуюся с D . Но это невозможно, поскольку поднятия этой компоненты на E_2 и E'_1 соответствуют различным решениям уравнения минимальной поверхности для одних и тех же граничных условий (что противоречит обычному принципу максимума). Следовательно, E'_1 пересекает E_2 трансверсально по единственной кривой γ , которая гомотопически нетривиальна на E'_1 и E_2 . Пусть $\tilde{E}_1$ и $\tilde{E}_2$ — концы поверхностей E'_1 и E_2 с границей γ .

Поверхности $\tilde{E}_1$ и $\tilde{E}_2$ представляют собой различные решения уравнения минимальной поверхности (они являются графиками) в неограниченной области Δ пространства $\mathbf{R}^2$ с границей $\pi(\gamma)$ для одних и тех же граничных значений на $\pi(\gamma)$. Поскольку $\tilde{E}_1$ и $\tilde{E}_2$ асимптотичны сдвигам фиксированного вертикального катеноида, они обладают одинаковой скоростью логарифмического роста.

Пусть X_1 и X_2 — градиенты третьей координатной функции поверхностей $\tilde{E}_1$ и $\tilde{E}_2$ соответственно. Если обозначить через ν_1, ν_2 конормальные (направленные вверх) векторы к $\tilde{E}_1$ и $\tilde{E}_2$ вдоль γ , то

$$\text{Flux}(\tilde{E}_1) = \left(0, 0, \int_{\gamma} X_1 \cdot \nu_1\right) = \left(0, 0, \int_{\gamma} X_2 \cdot \nu_2\right).$$

Однако вдоль кривой γ поверхность $\tilde{E}_1$ лежит ниже, чем $\tilde{E}_2$; следовательно, $X_1 \cdot \nu_1 < X_2 \cdot \nu_2$ в каждой точке этой кривой, что противоречит условию $\int_{\gamma} X_1 \cdot \nu_1 = \int_{\gamma} X_2 \cdot \nu_2$.

Основной результат теории минимальных поверхностей — формула монотонности. Ее доказательство можно найти в [G.-T.].

Теорема 3.5 (формула монотонности). Пусть M — минимальная поверхность, собственно погруженная в $\mathbf{R}^3$, $x \in M$, $D_R(x)$ — евклидов шар в $\mathbf{R}^3$ радиуса R с центром в точке x , k — число листов поверхности M , проходящих через x , и $\Sigma(R) = M \cap D_R(x)$. Тогда $\frac{|\Sigma(R)|}{\pi R^2 k}$ — монотонно возрастающая функция аргумента R , стремящаяся к единице при $R \rightarrow 0$. (Здесь $|\Sigma(R)|$ обозначает площадь поверхности $\Sigma(R)$). Площадь каждого листа поверхности $\Sigma(R)$, проходящего через x , возрастает не медленнее, чем πR^2 — площадь плоского диска радиуса R с центром в x .

Следствие. Пусть M — полная поверхность конечной полной кривизны в $\mathbf{R}^3$, имеющая в точности два вложенных конца. Тогда поверхность M вложенная.

Доказательство следствия. Каждый конец E поверхности M можно представить в виде графика функции u над $\mathbf{R}^2 \setminus$ (компактный диск). Несложные вычисления показывают, что площадь конца E имеет евклидову скорость роста, т.е. $\frac{|E \cap D_R|}{\pi R^2} \rightarrow 1$ при $R \rightarrow \infty$. Поскольку M имеет в точности два конца, каждый из которых вложенный, мы заключаем, что $f(R) = \frac{|M \cap D_R|}{2\pi R^2} \rightarrow 1$ при $R \rightarrow \infty$. Если M имеет точку самопересечения, из формулы монотонности следует, что $f(R) = 1$ для всех R ; поэтому M представляет собой объединение двух плоскостей.

Доказательство теоремы 3.1. Пусть E_1 и E_2 — концы поверхности M , и пусть γ_1 и γ_2 — гладкие жордановы кривые на E_1 и E_2 соответственно, каждая из которых гомотопически нетривиальна на своем конце, а Σ — компактное подмногообразие многообразия M , ограниченное кривой $\gamma_1 \cup \gamma_2$. Поскольку кривые γ_1 и γ_2 гомологичны в Σ , их потоки равны. Следовательно, если поток кривой γ_1 отличен от нуля, то предельные нормальные векторы к E_1 и E_2 параллельны, и если обозначить через a_1 и a_2 скорости логарифмического роста концов E_1 и E_2 соответственно, то $a_1 = -a_2$. Из формулы потоков также следует, что если $a_1 = 0$, то и $a_2 = 0$. Таким образом, оба конца одновременно являются либо концами катеноидного типа, либо концами плоскостного типа. Кроме того, так как поверхность M вложенная, концы всегда параллельны, т.е. их предельные нормальные векторы параллельны.

Осуществив поворот поверхности M , мы можем считать, что концы горизонтальны.

Сначала заметим, что ни один из концов поверхности M не является плоскостным. Действительно, если предположить, что конец E_1 плоскостной, то мы можем найти горизонтальную плоскость P , не пересекающуюся с M . Осуществляя параллельный перенос P по направлению к M , мы получим первую точку пересечения с этой горизонтальной плоскостью (это будет конечная точка поверхности M или бесконечность), что противоречит принципу максимума на бесконечности или обычному принципу максимума.

Итак, мы можем предположить, что E_1 — конец катеноидного типа (расположенный над E_2) со скоростью роста $a_1 \neq 0$, $a_1 = -a_2$.

Теперь докажем, что существует горизонтальная плоскость P , которая является плоскостью симметрии поверхности M . Тогда можно будет сделать вывод, что катеноиды, которым асимптотичны E_1 и E_2 , имеют общую вертикальную ось. Плоскость P можно найти, применяя метод отражения Александра и принцип максимума на бесконечности.

Пусть $P(t)$ — горизонтальная плоскость $x_3 = t$. Для каждого t обозначим через M_t^+ множество точек поверхности M , лежащих на плоскости $P(t)$ и над ней, а через M_t^- — множество точек поверхности M , лежащих на плоскости $P(t)$ и под ней. Пусть поверхность M_t^* симметрична M_t^+ относительно $P(t)$:

$$M_t^* = \{(x, 2t - x_3) \mid (x_1, x_3) \in M_t^+\}.$$

Поверхность S имеет локально ограниченный тангенс угла наклона, если касательные плоскости во всех внутренних точках этой поверхности не содержат вертикальных прямых. Пусть $p: \mathbf{R}^3 \rightarrow \mathbf{R}^2$ — проекция на горизонталь. Мы будем говорить, что подмножество A лежит выше подмножества B ($A \geq B$), если для любого такого $x \in \mathbf{R}^2$, что $p^{-1}(x) \cap A \neq \emptyset$ и $p^{-1}(x) \cap B \neq \emptyset$, каждая точка множества $p^{-1}(x) \cap A$ лежит над всеми точками множества $p^{-1}(x) \cap B$.

Для больших t множество M_t^+ является графиком с локально ограниченным тангенсом угла наклона над плоскостью $P(t)$ (точнее, над ее частью). Если t достаточно велико, то M_t^+ представляет собой часть E_1 . Следовательно, M_t^* — конец катеноидного типа со скоростью роста $-a_1$ (скорость роста такая же, как и у E_2). Таким образом, для достаточно больших t поверхность M_t^* лежит над M_t^- .

Теперь перейдем к случаю, когда t убывает и стремится, скажем, к s , и рассмотрим поверхность M_s^* . Мы утверждаем, что если для любого τ , $s \leq \tau \leq t$, поверхность M не является вертикальной

вдоль $P(\tau)$, то M_s^* лежит выше, чем M_s^- . Действительно, в противном случае при некотором τ существовала бы первая внутренняя точка пересечения M_τ^* и M_τ^- и в соответствии с обычным принципом максимума мы получили бы, что поверхность $M_\tau^* = M_\tau^- \circ M$ вертикальна вдоль $P(\tau) \cap M$. Здесь мы использовали принцип максимума на бесконечности, чтобы убедиться, что конец поверхности M_τ^* находится на строго положительном расстоянии от E_2 .

Поскольку невозможно, чтобы для всех s поверхность M_s^* лежала выше, чем M_s^- , существует наибольшее такое τ , что поверхность M вертикальна в некоторой точке $p \in M \cap P(\tau)$. Поверхность M_τ^* лежит выше, чем M_τ^- ; следовательно, M_τ^* и M_τ^- имеют общую границу. Они касаются в точке p и локально одна находится по одну сторону от другой. Таким образом, в соответствии с принципом максимума на границе $M_\tau^* = M_\tau^-$, и поверхность M инвариантна при отражении относительно $P(\tau)$.

Теперь рассмотрим представление конца E_2 в виде графика:

$$u_2(x) = a \ln x + b + \frac{c_1 x_1}{|x|^2} + \frac{c_2 x_2}{|x|^2} + \mathcal{O}(|x|^{-2}).$$

Нам нужно сдвинуть поверхность M таким образом, чтобы ось катеноида проходила через начало координат. Положим $x_1 = y_1 + \alpha_1$, $x_2 = y_2 + \alpha_2$. После такой замены u_2 можно переписать в виде

$$u_2(y) = a \ln |y| + b + \frac{\tilde{c}_1 y_1}{|y|^2} + \frac{\tilde{c}_2 y_2}{|y|^2} + \mathcal{O}(|y|^{-2}),$$

где $\tilde{c}_i = c_i + a\alpha_i$, $i = 1, 2$. Таким образом, положив $\alpha_i = -c_i/a$, $i = 1, 2$, и вернувшись от переменных y к переменным x , мы получим

$$u_2(x) = a \ln |x| + b + \mathcal{O}(|x|^{-2}).$$

Считая, что горизонтальная плоскость $x_3 = 0$ является плоскостью симметрии поверхности M (мы можем принять это предположение, осуществив вертикальный сдвиг поверхности M), мы получим $u_2(x) = -u_1(x)$, и, следовательно, разложение функции $u_1(x)$ имеет вид

$$u_1(x) = -a \ln |x| - b + \mathcal{O}(|x|^{-2}).$$

Теперь докажем, что M является поверхностью вращения вокруг оси x_3 и, следовательно, катеноидом. Поскольку выражение для u_2 инвариантно относительно поворотов вокруг начала координат в плоскости (x_1, x_2) , достаточно доказать, что плоскость $x_1 = 0$ является плоскостью симметрии поверхности M .

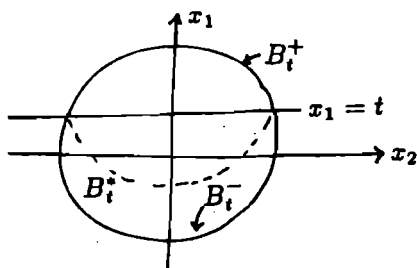


Рис. 12.

Обозначим через $P(t)$ плоскость $x_1 = t$ (теперь мы рассматриваем ось x_1 как вертикальную) и для (достаточно большой) константы K положим $M \cap \{|x_3| = K\} = B_1 \cup B_2 = B$, где $B_2 = M \cap \{x_3 = K\}$, $B_2 = M \cap \{x_3 = -K\}$.

Приведем идею дальнейших рассуждений. Зафиксируем $t > 0$. Для больших K множества B_1 и B_2 близки (C^1 -близки) к окружностям в плоскости $|x_3| = K$ с центрами на оси x_3 . Итак, B_t^+ является графиком с ограниченным тангенсом угла наклона над $P(0)$ и $B_t^+ \geq B_t^-$ (рис. 12).

Пусть $\Sigma = \Sigma(K)$ — компактная часть поверхности M , имеющая границу B . Применив принцип отражения Александрова к горизонтальным плоскостям $P(s)$ и опустившись с верхней части поверхности $\Sigma(K)$ на $P(t)$, можно доказать, что Σ_t^+ — график с ограниченным тангенсом угла наклона над $P(0)$ и $\Sigma_t^+ \geq \Sigma_t^-$. По построению это свойство будет выполняться и для больших значений K . Таким образом, $M_t^+ \geq M_t^-$. Поскольку это верно для любого t , мы получаем $M_0^+ \geq M_0^-$. Далее, проделав те же рассуждения снизу, т.е. начиная со значения $-t$ и поднимаясь вверх по горизонтальным плоскостям, получим, что $M_0^- \leq M_0^+$. Следовательно, поверхность M инвариантна при отражении относительно $P(0)$.

Теперь проделаем описанные выше действия более точно. Прежде всего докажем, что B_t^+ является графиком с ограниченным тангенсом угла наклона и $B_t^+ \geq B_t^-$. Очевидно, что достаточно провести доказательство для B_2 (для B_1 все рассуждения аналогичны). Поскольку

$$u_2(x) = a \ln |x| + b + O(|x|^{-2}),$$

мы имеем

$$\frac{\partial \dot{u}_2}{\partial x_1} = \frac{ax_1}{|x|^2} + \mathcal{O}(|x|^{-3}).$$

Отсюда следует, что при $x_1 \geq t$ и достаточно больших $|x|$ (в зависимости от t) должно выполняться неравенство $\frac{\partial u_2}{\partial x_1} > 0$. Нормальный вектор η к B_2 в плоскости $x_3 = K$ имеет вид

$$\eta = \left(\frac{ax_1}{|x|^2} + \mathcal{O}(|x|^{-3}), \frac{ax_2}{|x|^2} + \mathcal{O}(|x|^{-3}), 0 \right).$$

Первая координата этого вектора отлична от нуля для больших $|x|$ и $x_1 \geq t$; следовательно, B_t^+ — график с ограниченным тангенсом угла наклона.

На B_2 выполняется равенство $u_2 = K$; следовательно,

$$\ln |x| + \mathcal{O}(|x|^{-2}) = \frac{1}{a}(K - b)$$

и потому $|x|e^{\mathcal{O}(|x|^{-2})} = R$, где R — большое число. Поскольку $e^{\mathcal{O}(|x|^{-2})} = 1 + \mathcal{O}(|x|^{-2})$, отсюда следует, что $|x| = R + \mathcal{O}(|x|^{-1})$, и, таким образом, для больших $|x|$ множество B_2 близко к окружности. Обозначим через C окружность радиуса R в плоскости $x_3 = K$ с центром в начале координат. Очевидно, что

$$\text{dist}(C_t^*, C_{t/2}^-) \geq \epsilon(t),$$

где $\epsilon(t) > 0$, причем $\epsilon(t)$ зависит только от t . Следовательно, если K достаточно велико, то $B_{2,t}^* \geq B_{2,t/2}^-$. Поскольку $B_{2,t/2}^+$ является графиком над $P(0)$, отсюда следует, что $B_{2,t}^* \geq B_{2,t}^-$.

Чтобы завершить доказательство теоремы Р. Шёна, осталось убедиться в том, что $\Sigma_t^* \geq \Sigma_t^-$, где $\Sigma = \Sigma(K)$ — компактное подмногообразие поверхности M , ограниченное B .

Пусть T — максимальное значение x_1 на Σ ; оно достигается на B , поскольку функция x_1 гармоническая.

Пусть $J = \{s \in [t, T] \mid \Sigma_s^+ \text{ — график с локально ограниченным тангенсом угла наклона над } P(s) \text{ и } \Sigma_s^* \geq \Sigma_s^-\}$. Мы докажем теорему, показав, что J является непустым открытым и замкнутым подмножеством отрезка $[t, T]$, откуда следует, что $t \in J$.

Пусть $p \in B$ — точка, в которой $x_3 = T$. Тогда $T \in J$, и, следовательно, множество J непусто. Поскольку компактная минимальная поверхность лежит в выпуклой оболочке своей границы (это

следует из принципа максимума, если применить его к координатным функциям), Σ представляет собой «плиту» между плоскостями $x_3 = K$ и $x_3 = -K$. В соответствии с принципом максимума на границе поверхность Σ не может касаться плоскостей $|x_3| = K$; следовательно, она трансверсальна этим плоскостям вдоль B . Таким образом, для значений s , близких к T , $s < T$, Σ_s^+ является графиком над $P(s)$ с локально ограниченным тангенсом угла наклона. Заметим также, что если $s \in J$ и $s < s_1 \leq T$, то $s_1 \in J$.

Сначала докажем, что множество J замкнуто. Предположим, что $(s, T) \subset J$. Если Σ_s^+ не является графиком, то существуют такое число s_1 , $s < s_1 \leq T$, и такая точка $x \in P(0)$, что точки (x, s) и (x, s_1) лежат на одной вертикали. Выберем s_1 таким образом, чтобы на вертикали, соединяющей точки (x, s) и (x, s_1) , не было других точек поверхности Σ . Мы знаем, что (x, s) является внутренней точкой поверхности Σ , поскольку B_t^+ является графиком и Σ касается плоскостей $|x_3| = K$ только вдоль B .

В окрестности точки (x, s_1) поверхность Σ является графиком и не вертикальна. Таким образом, вертикальные прямые, опущенные из этой окрестности, заполняют некоторую окрестность точки (x, s) . Отсюда следует, что Σ лежит ниже $P(s)$ в окрестности точки (x, s) , так как в противном случае существовала бы такая точка $(\bar{x}, \bar{s})$, расположенная на Σ вблизи от (x, s) , что $\bar{s} > s$. Но тогда на вертикали над точкой $(\bar{x}, \bar{s})$ находилась бы еще одна точка поверхности Σ и, следовательно, Σ_s^+ не была бы графиком. Очевидно, что если бы в точке (x, s) поверхность Σ была расположена ниже, чем $P(s)$, то по принципу максимума $\Sigma = P(s)$. Полученное противоречие доказывает замкнутость множества J .

Теперь докажем, что множество J открыто. Возьмем $(x, s) \in \Sigma$ и будем считать, что $[z, T] \subset J$.

Пусть $D \subset \Sigma$ — диск, содержащий точку (x, s) . Заметим, что поверхность Σ не вертикальна в точке (x, s) . Действительно, предположим, что она вертикальна, и рассмотрим диски D_s^* и D_s^- . Диск D_s^* локально в окрестности точки (x, s) лежит с одной стороны от D_s^- , а следовательно, в соответствии с принципом максимума на границе $D_s^* = D_s^-$ и $P(s)$ является плоскостью симметрии для Σ . Но это невозможно, поскольку множество B не симметрично относительно $P(s)$.

Таким образом, поверхность Σ является графиком в окрестности U плоскости $P(s)$ и не вертикальна в этой окрестности. Осталось показать, что $\Sigma_\tau^* \geq \Sigma_\tau^-$, если τ достаточно близко к s . По-

сколькx $\Sigma \cap U$ является графиком, мы имеем $\Sigma_\tau^* \setminus V \geq \Sigma_\tau^-$, где V — окрестность плоскости $P(s)$, $V \subset U$, и τ близко к s . Также отметим, что поверхность $\Sigma_\tau^+ \setminus V$ является компактной и ее образ при отражении относительно $P(s)$ не пересекается с Σ_τ^- ; следовательно, в силу непрерывности для значений τ , близких к s , мы имеем $\Sigma_\tau^* \setminus V \geq \Sigma_\tau^-$. Это означает, что $\Sigma_\tau^* \geq \Sigma_\tau^-$ для значений τ , близких к s .

Последнее рассуждение с использованием принципа отражения Александpова позволяет доказать даже более сильное утверждение: при определенных обстоятельствах минимальная поверхность «наследует» симметрии своей границы. Более точно, Р. Шён доказал следующую теорему.

Теорема 3.6. Пусть $\Omega \subset \mathbf{R}^n$ — компактная область, граница которой выпукла в среднем, и пусть $B^{n-1} \subset \mathbf{R}^{n+1}$ — компактное вложенное многообразие (не обязательно связное), удовлетворяющее следующим условиям: $B \subset (\partial\Omega) \times \mathbf{R}$, B_0^+ — график с локально ограниченным тангенсом угла наклона над $P(0) = \mathbf{R}^n \times (0)$ и $B_0^+ \geq B_0^-$. Если M — такая вложенная минимальная гиперповерхность, что $\partial M = B$ и все внутренние точки этой гиперповерхности лежат в $\Omega \times \mathbf{R}$, то M_0^+ является графиком над $P(0)$ с локально ограниченным тангенсом угла наклона и $M_0^+ \geq M_0^-$.

Р. Шён доказал эту теорему, используя лишь тот факт, что поверхность M погруженная, но, чтобы получить это обобщение, необходимо провести более сложные рассуждения, чем проделали мы. Предположение о том, что внутренние точки поверхности M лежат в $\Omega \times \mathbf{R}$, несущественно, поскольку если внутренняя точка p этой гиперповерхности содержится в $(\partial\Omega) \times \mathbf{R}$, то по принципу максимума связная компонента точки p в M полностью содержится в $(\partial\Omega) \times \mathbf{R}$, так что такие компоненты можно не принимать во внимание.

Нетрудно построить примеры границ B , симметрии которых не продолжаются до симметрий минимальной поверхности M с $\partial M = B$. Например, можно рассмотреть два экземпляра «штангообразных» кривых в параллельных плоскостях (рис. 13).

(Асимметричную) минимальную поверхность M можно получить, применив принцип моста к катеноиду, ограниченному двумя окружностями, справа и к двум дискам, ограниченными двумя окружностями, слева.

Приступим теперь к доказательству теоремы Лопеса—Роса: плоскость и катеноид — единственные собственно вложенные минимальные поверхности конечной полной кривизны и рода нуль.

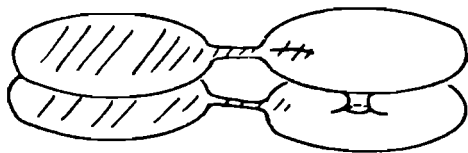


Рис. 13.

Приведем идею доказательства. Пусть M — поверхность конечной полной кривизны, вложенная в $\mathbf{R}^3$. Мы знаем, что M имеет конечное число концов, которые после поворота поверхности M в $\mathbf{R}^3$ можно считать горизонтальными (т. е. предельные нормали к этим концам вертикальны). Каждый конец асимптотичен горизонтальной плоскости или катеноиду.

Лопес и Рос деформировали поверхность M в классе минимальных поверхностей, меняя данные Вейерштрасса. Данные Вейерштрасса (g, ω) поверхности M заменялись данными $(\lambda g, \frac{\omega}{\lambda})$, где λ — действительное положительное число. Можно проверить, что эти данные определяют погружение X_λ поверхности M . Действительно, это так для всех случаев, когда поток поверхности M вертикален. В нашем случае это верно, поскольку M имеет нулевой род и весь поток сосредоточен на концах. Плоскостные концы имеют нулевой поток, а концы катеноидного типа имеют вертикальный поток, поскольку предельные нормали к этим концам вертикальны.

Далее заметим, что каждый катеноидный конец поверхности M является (вертикальным) катеноидным концом поверхности X_λ и имеет ту же самую скорость логарифмического роста; при деформации он может двигаться вверх или вниз. Плоскостные концы остаются плоскостными на той же самой высоте.

В соответствии с принципом максимума на бесконечности расстояние между концами поверхности X_λ строго положительно, когда λ меняется в пределах от 1 до бесконечности. Следовательно, каждая поверхность X_λ является вложенной.

Если на поверхности M существует точка p , в которой нормальный вектор вертикален, то найдутся такая окрестность D точки p и такое число $\lambda > 1$, что поверхность $X_\lambda(D)$ не является вложенной (мы предполагаем, что M не является плоскостью). В этом можно убедиться, доказав, что поверхности X_λ вблизи точки p при $\lambda \rightarrow \infty$ сходятся к поверхности Эннепера (которая не является погружен-

ной). Таким образом, на поверхности M не может быть точек, в которых нормаль вертикальна.

Если поверхность M имеет плоскостной конец A , то с помощью подобных рассуждений можно доказать, что если λ достаточно велико и D является подконцом конца A , то поверхность $X_\lambda(D)$ не вложенная (мы предполагаем, что M не плоская).

Таким образом, все концы поверхности M катеноидного типа и гауссово отображение не имеет нулей и полюсов на M . Конформная компактификация поверхности M — это сфера S ; слое-ние на кривые уровня функции x_3 невырожденно на M и имеет особенность положительного индекса в каждой выколотой точке. Следовательно, поверхность M имеет в точности два конца.

Теперь с помощью теоремы Р. Шёна можно заключить, что M является катеноидом, однако это несложно доказать и непосредственно. Поскольку все нули и полюсы гауссова отображения расположены на концах, отображение g имеет первую степень. Проведя конформную перепараметризацию поверхности M , можно считать, что $g(z) = z$. Из теории вычетов получаем, что $\omega(z) = c dz/z^2$, $c \in \mathbb{R}$. Следовательно, M — катеноид [Oss.-1].

Обратимся теперь к деталям доказательства.

Для любого цикла γ на M можно вычислить его поток по следующей формуле:

$$\int_{\gamma} (\phi_1, \phi_2, \phi_3) = i \text{Flux}(\gamma).$$

Из этой формулы нетрудно вывести, что следующие три условия эквивалентны тому, что M имеет вертикальный поток:

- 1) формы ϕ_1 и ϕ_2 точны;
- 2) формы ω и $g^2\omega$ точны;
- 3) для любого $\lambda > 0$ погружение X_λ корректно определено на M .

Формула для метрики и кривизны показывает, что каждая поверхность $X_\lambda(M) = M_\lambda$ является полной минимальной поверхностью той же полной кривизны, что и M .

Теперь предположим, что $p \in M$ — точка, в которой нормаль к M вертикальна, например $(0, 0, -1)$. Конформно параметризуем окрестность точки p диском $\{0 \leq |z| < \epsilon\}$ таким образом, что

$$g(z) = z^k, \quad \omega = (a + zh(z))dz,$$

где $a \in \mathbb{C}^*$ и функция h голоморфна в $D(t) = \{|z| < \epsilon\}$. Введем на $D(\lambda^{1/k}\epsilon)$ конформную координату $\xi = \lambda^{1/k}z$. Тогда X_λ параме-

тризуется данными

$$g_\lambda(\xi) = \xi^k, \quad \omega_\lambda = \frac{1}{\lambda^{1+1/k}} \left(a + \frac{\xi}{\lambda^{1/k}} h \left(\frac{\xi}{\lambda^{1/k}} \right) \right) d\xi.$$

Растягивая X_λ в $\lambda^{1+1/k}$ раз, получим $\tilde{X}_\lambda$. При $\lambda \rightarrow \infty$ поверхности $\tilde{X}_\lambda$ на компактных подмножествах пространства $\mathbf{C}$ сходятся к минимальной поверхности $X_\infty: \mathbf{C} \rightarrow \mathbf{R}^3$ с данными Вейерштрасса

$$g_\infty(\xi) = \xi^k, \quad \omega_\infty = a d\xi.$$

Это полная поверхность, конец которой имеет трансверсальные самопересечения; следовательно, для больших λ поверхности X_λ имеют самопересечения.

Если нормаль в точке p равна $(0, 0, 1)$, то следует перевернуть поверхность M .

Теперь предположим, что A — плоскостной конец поверхности M (причем M не является плоскостью), и пусть предельный нормальный вектор к A — это $(0, 0, -1)$. Параметризуем подконец конца A следующими данными Вейерштрасса в $D(\epsilon)$:

$$g(z) = z^k, \quad \omega = \left(\frac{a}{z^2} + h(z) \right) dz,$$

где $a \in \mathbf{C}^*$ и функция h голоморфна в $D(\epsilon)$.

Мы получаем следующую параметризацию конца поверхности M_λ в $D(\lambda^{1/k}\epsilon)$:

$$g_\lambda(\xi) = \xi^k, \quad \omega_\lambda = \frac{1}{\lambda^{1-1/k}} \left(\frac{a}{\xi^2} + \frac{1}{\lambda^{2/k}} h \left(\frac{\xi}{\lambda^{1/k}} \right) \right) d\xi, \quad k > 1,$$

где $\xi = \lambda^{1/k}z$. Осуществив гомотегию с коэффициентом $\lambda^{1-1/k}$, мы получим новую минимальную поверхность $\widehat{M}_\lambda$. При $\lambda \rightarrow \infty$ поверхности $\widehat{M}_\lambda$ на компактных подмножествах пространства $\mathbf{C}^*$ равномерно сходятся к поверхности $X_\infty: \mathbf{C}^* \rightarrow \mathbf{R}^3$, определяемой формулами

$$g_\infty(\xi) = \xi^k, \quad \omega_\infty = \frac{a}{\xi^2} d\xi, \quad \xi \in \mathbf{C}^*.$$

Если $k = 1$, то эта поверхность является катеноидом, а если $k > 1$, то она имеет невложенный конец на бесконечности. Следовательно, для достаточно больших λ поверхности M_λ не являются вложенными, поскольку A — плоскостной конец.

Осталось доказать, что все X_λ являются вложениями. Пусть $J = \{\lambda \mid X_\lambda \text{ инъективно}\}$. Если $\lambda_0 \in J$, то, согласно принципу максимума на бесконечности, расстояние между двумя фиксированными концами поверхности $X_{\lambda_0}(M)$ строго положительно. Очевидно, что это расстояние является непрерывной функцией аргумента λ (возможно, принимающей бесконечные значения). Следовательно, если λ близко к λ_0 , то X_λ также является вложением, и множество J открыто.

Пусть $\lambda_n \in J$, $\lambda_n \rightarrow \lambda$ при $n \rightarrow \infty$. Если X_λ не инъективно, то существуют такие точки $x, y \in M$, $x \neq y$, что $X_\lambda(x) = X_\lambda(y)$. Пересечения поверхности M_λ с $X_\lambda(x)$ и $X_\lambda(y)$ не могут быть одномерными; следовательно, согласно обычному принципу максимума, окрестности точек x и y имеют один и тот же образ при действии X_λ (мы использовали условие $\lambda_n \in J$). Поэтому $X_\lambda: M \mapsto M_\lambda$ — конечное накрытие (вложенной) минимальной поверхности M_λ . Вновь применяя принцип максимума на бесконечности, мы доказываем, что существует вложенная ϵ -трубчатая окрестность U поверхности M_λ . Концы поверхностей M_{λ_n} меняются непрерывно, так что при больших n мы имеем $M_{\lambda_n} \subset U$. Но тогда ортогональная проекция поверхности M_{λ_n} на M_λ является диффеоморфизмом, и, таким образом, X_λ тоже представляет собой диффеоморфизм. Это противоречие показывает, что множество J замкнуто, и это завершает доказательство теоремы Лопеса—Роса.

4. ОЦЕНКИ КРИВИЗНЫ ДЛЯ УСТОЙЧИВЫХ МИНИМАЛЬНЫХ ПОВЕРХНОСТЕЙ

В 1952 г. Е. Гейнц доказал, что если M — минимальный график над диском D_R радиуса R ($D_R = \{x^2 + y^2 \leq R^2\}$) и K_0 — гауссова кривизна поверхности M в начале координат, то [Heinz]

$$|K_0| \leq \frac{4\pi^3}{3R^2}.$$

Е. Хопф, Финн и Оссерман [E. Hopf-1, F.-Oss.] обобщили этот результат на случай параметрических минимальных поверхностей, для которых гауссово отображение не принимает значений из некоторого открытого множества.

Наиболее общая теорема была получена Р. Шёном [Sch.-2]: существует такая универсальная константа $C > 0$, что если M — устой-

чивая минимальная (погруженная и полная) поверхность в плоском трехмерном многообразии, то

$$|K(p)| \leq \frac{C}{d(p)^2},$$

где $p \in M$, а $d(p)$ — внутреннее расстояние от p до ∂M . Устойчивость означает, что на любой компактной области D поверхности M достигается минимальное с точностью до второго порядка малости значение площади в классе всех нормальных вариаций области D , оставляющих границу неподвижной. Теперь приведем вкратце более строгие рассуждения.

Теорема Шёна представляет собой очень важный инструмент для изучения минимальных поверхностей в трехмерных многообразиях. Заметим, что из нее вытекает следующий факт: в плоских трехмерных многообразиях полные погруженные устойчивые минимальные поверхности, не имеющие границы, и только они являются вполне геодезическими подмногообразиями. Так, например, в $\mathbb{R}^3$ это плоскости. Этот результат также доказан До Кармо и Пенгом [Do C.-P.].

Почему результат Шёна является обобщением теоремы Гейнца, т.е. почему минимальный график устойчив? Вообще говоря, когда мы говорим о слоении на минимальные гиперповерхности, мы подразумеваем устойчивость каждого листа. Единичное векторное поле n , нормальное к слоению, является бездивергентным в объемлющем пространстве. Пусть D — компактная область в одном из листов и $\tilde{D}$ — такая цепь, что $\partial \tilde{D} = \partial D$. Тогда $D \cup \tilde{D}$ — цикл, гомологичный нулю. Согласно теореме о дивергенции, поток поля n через D равен потоку поля n через $\tilde{D}$, т.е.

$$\text{площадь}(D) = \int_D \langle n, n \rangle = \int_{\tilde{D}} \langle n, n_{\tilde{D}} \rangle \leq \text{площадь}(\tilde{D}),$$

где $n_{\tilde{D}}$ — единичное векторное поле, нормальное к $\tilde{D}$. Следовательно, D минимизирует площадь в своем классе гомологий.

Вертикальные сдвиги графика заполняют трехмерный цилиндр, и приведенные выше рассуждения показывают, что для любого $R' < R$ на части поверхности M , лежащей над $D_{R'}$, достигается минимум площади с точностью до второго порядка малости. Перейдя к пределу при $R' \rightarrow R$, мы убедимся, что этим свойством обладает вся поверхность M .

4.1. Критерий устойчивости Барбосы—До Кармо. Барбоса и До Кармо получили важный критерий устойчивости области на минимальной поверхности в $\mathbb{R}^3$, из которого следует устойчивость графиков. Согласно доказанной ими теореме, минимальная поверхность, погруженная в $\mathbb{R}^3$, устойчива, если площадь ее сферического образа (при гауссовом отображении) меньше чем 2π .

Мне хотелось бы сделать ряд замечаний в связи с теоремой Барбосы—До Кармо. Пусть D — компактная область на минимальной поверхности M , n — единичное векторное поле, нормальное к M , и f — кусочно-гладкая функция на D , обращающаяся в нуль на ∂D . Векторное поле $Y = fn$ на D индуцирует нормальную вариацию области D , и вторая производная площади этой вариации имеет вид

$$- \int_D f(\Delta f - 2Kf),$$

где Δ — внутренний лапласиан поверхности M . Оператор $L = \Delta - 2K$ является оператором устойчивости (оператором Якоби) поверхности M . Устойчивость поверхности M означает строгую положительность приведенного выше интеграла для всех компактных областей D и непостоянной функции f на D , обращающейся в нуль на ∂D . Следовательно, если можно найти непостоянную функцию f , $f = 0$ на ∂D , в ядре оператора L (такие функции f называются полями Якоби), то область D неустойчива.

Теперь предположим, что D — область, в которой гауссово отображение g представляет собой разветвленное накрытие над $g(D)$. Для этого случая Шварц доказал, что если первое собственное значение λ_1 сферического лапласиана Δ_s на $g(D)$ меньше чем два, то область D неустойчива. Приведем доказательство этого утверждения. Пусть u — функция на $g(D)$, положительная во внутренней части этой области, равная нулю на $\partial g(D)$ и такая, что $\Delta_s u + \lambda_1 u = 0$. Положим $f = u \circ g$. Поскольку $g(\partial D) = \partial(gD)$, функция f обращается в нуль на ∂D и положительна во внутренней части D . Вторая вариация, определяемая функцией f , имеет вид

$$- \int_D f \Delta_s f + 2f^2 = (\lambda_1 - 2) \int_D f^2 < 0.$$

Следовательно, область D неустойчива (приведенные выше интегралы берутся по дополнению к множеству точек ветвления отображения g).

Теперь приведем идею доказательства критерия устойчивости Барбосы—До Кармо. Если область D неустойчива, то можно найти

такую область $\tilde{D} \subset D$ и такую функцию u на $\tilde{D}$, что $u > 0$ в $\text{int } \tilde{D}$, $u = 0$ на $\partial \tilde{D}$ и $\Delta u - 2Ku = 0$.

Затем, усреднив u с помощью гауссова отображения, мы получим функцию f на $g(\tilde{G})$, удовлетворяющую неравенству

$$\int_{g(\tilde{G})} |\text{grad } f|^2 \leq \int_{g(\tilde{G})} f^2.$$

Из этого неравенства следует, что $\lambda_1(g(\tilde{D})) \leq 2$.

Тем не менее среди всех сферических областей фиксированной площади минимум первого собственного значения лапласиана достигается на сферической «шарике». Но для сферической «шарике» в открытой полусфере $\lambda_1 > 2$ (координатные функции пространства $\mathbb{R}^3$ удовлетворяют условию $\Delta_s + 2 = 0$, положительны на открытой полусфере и равны нулю на ее границе), что приводит к противоречию.

4.2. Идея доказательства теоремы Гейнца. Пусть M — график функции f , градиент которой обращается в нуль в начале координат. Это предположение относительно градиента упрощает доказательство. Основная идея состоит в сравнении поверхности M с графиком поверхности Шерка.

Мы можем считать, что f определяет минимальный график на D_R , $f(0, 0) = 0$ и $|\nabla f(0, 0)| = 0$. Повернем график функции f , искривляя его вверх, так, чтобы ось x имела главное направление.

Пусть N — график Шерка, определенный над квадратом со стороной длины 2 и с центром в начале координат, и пусть его граничные значения на вертикальных сторонах квадрата равны $+\infty$, а на горизонтальных $-\infty$. Предположим также, что этот график проходит через начало координат; очевидно, что N горизонтален в начале координат.

Пусть $\hat{K}_0$ — гауссова кривизна поверхности N в начале координат. При гомотетии с коэффициентом $C > 0$ относительно начала координат N переходит в минимальный график N_C , определенный над квадратом $\diamond(C)$, содержащим D_C . Поскольку при такой гомотетии кривизна умножается на $\frac{1}{C^2}$, кривизна K_C поверхности N_C в начале координат должна удовлетворять неравенству

$$|K_C| \leq \frac{|\hat{K}_0|}{C^2}.$$

Заметим, что поверхность N_C горизонтальна в начале координат и одна из главных кривизн этой поверхности направлена вверх вдоль оси x .

Выберем $C > 0$ таким образом, чтобы главная кривизна поверхности N_C , направленная вдоль оси x в начале координат, была равна соответствующей главной кривизне поверхности M в начале координат. Тогда в этой точке выполнено равенство $K_0 = K_C$.

Если $R \leq C$, то $D_R \subset D_C$ и

$$|K_0| = |K_C| \leq \frac{|\tilde{K}_0|}{C^2} \leq \frac{|\tilde{K}_0|}{R^2}.$$

Если $\diamond(C) \subset D_R$, то рассмотрим $M \cap N_C$. Обе поверхности касаются в начале координат; поэтому либо они совпадают, либо $M \cap N_C$ является одномерной аналитической кривой, имеющей особенность в начале координат, через которую проходят и по крайней мере шесть ветвей.

Так как N_C асимптотична бесконечности на границе квадрата $\diamond(C)$, за исключением четырех вершин, то должна существовать компактная компонента кривой $M \cap N_C$, строго содержащаяся в вертикальной области над внутренностью $\diamond(C)$ (по крайней мере одна ветвь кривой $M \cap N_C$ должна проходить через фиксированную вершину квадрата $\diamond(C)$, поскольку M является графиком). Следовательно, существует жорданова кривая α в $\diamond(C)$, вдоль которой M и N_C совпадают. Так как обе поверхности являются графиками над внутренностью α и, согласно принципу максимума, существует единственный такой минимальный график, то $M = N_C$. Мы получили противоречие. Таким образом, $\diamond(C)$ не содержится в D_R , а значит, $C > R/\sqrt{2}$ и

$$|K_0| = |K_C| \leq \frac{|\tilde{K}_0|}{C^2} < \frac{2|\tilde{K}_0|}{R^2}.$$

Тем самым справедливость оценки Гейнца установлена.

5. КОМПАКТНОСТЬ СЕМЕЙСТВ НАИМЕНЬШЕЙ ПЛОЩАДИ И ПОСТРОЕНИЕ ДОПОЛНИТЕЛЬНЫХ ПОВЕРХНОСТЕЙ КОНЕЧНОЙ ПОЛНОЙ КРИВИЗНЫ

Для изучения полной минимальной поверхности M в плоском трехмерном многообразии N нередко используется следующий метод: строятся такие минимальные поверхности Σ конечной полной

кривизны, что множество $\partial\Sigma'$ компактно и непусто, поверхность Σ не является компактной, $\partial\Sigma \subset M$ и $\text{int}(\Sigma) \cap M = \emptyset$. Такие поверхности Σ «улавливают» M в малых областях многообразия N , что делает геометрию поверхности M доступной для понимания. Мы приведем несколько примеров применения этого метода.

Сначала объясним, как можно получить поверхность Σ . Пусть Ω — полная область в N , граница которой не допускает решения задачи Плато о наименьшей площади (эта теория была разработана Миксом и Яу [M.-Y.]). Это означает, что $\partial\Omega = C$ — двумерное многообразие, гладкость которого нарушается лишь вдоль аналитического одномерного многообразия, такое, что

- C выпукло в среднем в точках гладкости, т. е. вектор средней кривизны в этих точках направлен внутрь Ω (считаем, что нулевой вектор направлен внутрь Ω), и

- в тех точках, где гладкость многообразия C нарушается, угол между гладкими гранями этого многообразия не превосходит π (этот угол измеряется в Ω).

Микс и Яу доказали, что любой гладкий вложенный 1-цикл Γ в $\partial\Omega$, т. е. цикл, гомологичный нулю в Ω , является границей компактной поверхности наименьшей площади Σ_Γ в Ω , причем поверхность Σ_Γ является гладкой и вложенной. Основная идея состоит в том, чтобы решить задачу Плато в N , перейдя к пределу вложенных поверхностей с границей Γ , площади которых сходятся к точной нижней грани всех возможных площадей. Нетрудно убедиться, что можно построить эту минимизирующую последовательность таким образом, чтобы ее элементы оставались в пределах области Ω . Из выпуклости в среднем (и углового условия) следует, что площадь поверхностей, выходящих за пределы Ω , возрастает при пересечении $\partial\Omega$. Теперь остается (с помощью весьма трудоемкой процедуры) выделить подпоследовательность, сходящуюся к гладкой вложенной поверхности.

Можно получить Σ_Γ также и с помощью геометрической теории меры [Simon]. Вновь предположим, что граница $C = \partial\Omega$ не допускает решения задачи Плато и $\Gamma \subset \partial\Omega$ — гладкий 1-цикл (т. е. совокупность непересекающихся гладких жордановых кривых). Если Γ является границей ориентированной 2-цепи в Ω , то Γ является границей гладкой вложенной ориентируемой поверхности Σ_Γ в Ω , на которой достигается минимум площади на множестве всех ориентируемых 2-цепей в Ω с границей Γ . Если Γ есть Z_2 -граница в Ω , то Γ является границей гладкой вложенной поверхности наименьшей площади в том же классе относительных Z_2 -гомологий. Если

Γ является границей ориентируемой (погруженной) поверхности рода n в Ω , то можно выбрать поверхность Σ_Γ рода не более чем n , имеющую наименьшую площадь в своем гомотопическом классе.

Теперь перейдем к рассмотрению того, как компактные минимальные поверхности наименьшей площади Σ_Γ могут сходиться к некомпактным минимальным поверхностям Σ конечной полной кривизны.

Допустим, что поверхность M ориентируема, A является концом поверхности M , $A \subset \partial\Omega$ и Γ — гладкая жорданова кривая на A , не гомологичная нулю в Ω . Пусть $A_1 \subset A_2 \subset \dots$ — возрастающая последовательность компактных подмногообразий многообразия A , заполняющая все это многообразие, и пусть $\partial A_i = \Gamma \cup \Gamma_i$. Из нашего предыдущего обсуждения решения задачи Плато в A с использованием геометрической теории меры мы знаем, что в Ω существует такая гладкая вложенная поверхность наименьшей площади Σ_i , что $\partial\Sigma_i = \Gamma \cup \Gamma_i$ и Σ_i является Z_2 -гомологичной A_i относительно ∂A_i . Поскольку многообразие A_i ориентируемо, а поверхность $\Sigma_i \cup A_i$ является Z_2 -гомологичной нулю, Σ_i также ориентируема. Поскольку кривая Γ не гомологична нулю в Ω , поверхность Σ_i можно выбрать связной.

Теперь покажем, что некоторая подпоследовательность последовательности Σ_i сходится к устойчивой вложенной минимальной поверхности Σ , для которой $\partial\Sigma = \Gamma$.

Заметим, что для семейства Σ_i существуют равномерные локальные границы площади. Действительно, если $B \subset \Omega$ — шар радиуса r и граница ∂B трансверсальна к Σ_i , то $\partial B \cap \Sigma_i$ есть 1-цикл на ∂B , ограничивающий 2-цепь (mod 2) на ∂B площади не более чем $2\pi r^2$. Поскольку на Σ_i достигается минимум площади, ограниченной ∂A_i (в классе Z_2 -гомологий), мы заключаем, что площадь поверхности $B \cap \Sigma_i$ не превосходит $2\pi r^2$. Аналогично, если B — шар с центром в точке, лежащей на $\partial\Sigma_i$, то площадь поверхности $\Sigma_i \cap B$ не превосходит площади ∂B .

Теперь пусть $B(r) \subset \Omega$. Согласно оценкам кривизны, приведенным Р. Шёном, выбрав достаточно малое r , мы можем представить каждую компоненту множества $\Sigma_i \cap B(r)$, пересекающую $B(r/2)$, в виде графика с малым градиентом над плоскостью P_i в $B(r)$, проходящей через центр шара, причем P_i не зависит от компоненты. В силу равномерных оценок площади число компонент множества $\Sigma_i \cap B(r/2)$ ограничено и не зависит от i ; следовательно,

существует ограниченное число соответствующих графиков. Сначала предположим, что для любого i множество $\Sigma_i \cap B(r/2)$ содержит одну компоненту. Выберем подпоследовательность последовательности P_i , которая сходится к плоскости P , проходящей через центр шара. Тогда из стандартной теоремы о компактности для минимальных графиков следует сходимости некоторой подпоследовательности графиков $\Sigma_i \cap B(r/2)$ к минимальному графику над $P \cap B(r/2)$. Если $\Sigma_i \cap B(r/2)$ имеет более одной компоненты, мы проводим приведенные выше рассуждения для каждой компоненты и получаем (равномерно ограниченное) конечное число графиков над $P \cap B(r/2)$, к которым сходится подпоследовательность последовательности $\Sigma_i \cap B(r/2)$.

Область Ω обладает счетным базисом шаров B_n , где для любого n и для любой подпоследовательности Σ_{i_λ} последовательности Σ_i из $\Sigma_{i_\lambda} \cap B_n$ можно выделить подпоследовательность, сходящуюся в B_n . Предположим, что подпоследовательность $\Sigma_{i_\lambda} \cap B_1$ сходится в B_1 . Тогда из соответствующей последовательности графиков в $B_2 \cap \Sigma_{i_\lambda}$ можно выделить подпоследовательность, сходящуюся в $B_2 \cup B_1$. Продолжим эти рассуждения, переходя от B_i к B_{i+1} , и возьмем диагональную подпоследовательность. Тем самым мы получим подпоследовательность последовательности Σ_i , сходящуюся к некоторой гладкой минимальной поверхности Σ , для которой $\partial \Sigma = \Gamma$. Нетрудно видеть, что поверхность Σ является вложенной и устойчивой (поскольку она является пределом вложенных поверхностей наименьшей площади). Из теоремы Хардта и Симонса о регулярности границ также следует, что поверхность Σ является гладкой вдоль Γ [H.-S.]. И наконец по теореме Дорис Фишер-Колбри мы получаем, что Σ имеет конечную полную кривизну [F.C.].

Эта техника позволяет, в частности, получить следующую лемму.

Лемма 5.1. Пусть M — собственно вложенная минимальная поверхность в $\mathbb{R}^3$ с более чем одним концом. Тогда в дополнении к M существует конец катеноида или плоскости.

Доказательство. Пусть Γ — гладкая жорданова кривая на M , разделяющая M на две некомпактные компоненты. Обозначим одну из этих компонент через A . Поверхность M разделяет $\mathbb{R}^3$ на две связные компоненты, и кривая Γ не может быть гомологична нулю в каждой из этих компонент. Пусть Ω — та компонента, в которой Γ не гомологична нулю.

В соответствии с нашими предыдущими рассуждениями в Ω существует такая вложенная минимальная поверхность Σ_Γ конечной полной кривизны, что $\partial\Sigma_\Gamma = \Gamma$. Точнее, $C = \partial\Omega = M$ — минимальная поверхность; следовательно, она не допускает решения проблемы Плато. Пусть $A_i \subset A_{i+1}$ — такое исчерпывание кривой A , что $\partial A_i = \Gamma \cup \Gamma_i$, и пусть Σ_i — вложенная минимальная поверхность в Ω , $\mathbb{Z}$ -гомологичная A_i и такая, что $\partial\Sigma_i = \partial A_i$. Как мы видели ранее, подпоследовательность последовательности Σ_i сходится к поверхности Σ_Γ .

Может случиться так, что $\Sigma_\Gamma \subset M$ (если эта поверхность касается M в одной внутренней точке, то, поскольку она находится с одной стороны от M в этой точке, она содержится в M). В этом случае хотя бы один конец поверхности M имеет конечную полную кривизну и, таким образом, асимптотичен некоторому плоскостному или катеноидному концу B . В соответствии с принципом максимума на бесконечности расстояние между концами поверхности M строго положительно. Следовательно, конец B можно перенести в Ω таким образом, чтобы он не пересекался с M . Аналогично, если $\text{int } \Sigma_\Gamma \subset \text{int } \Omega$, то концы поверхности Σ_Γ находятся на строго положительном расстоянии от M , откуда следует утверждение леммы.

Рассмотрим полезную модификацию этой леммы.

Лемма 5.2. Пусть B — шар в $\mathbb{R}^3$, и пусть A_1, A_2 — такие некомпактные собственно вложенные минимальные поверхности, что $\partial A_1, \partial A_2$ — гладкие жордановы кривые, $B \cap (A_1 \cup A_2) = \partial A_1 \cup \partial A_2$ и $A_1 \cap A_2 = \emptyset$. Обозначим через Δ кольцо на ∂B с границей $\partial A_1 \cup \partial A_2$, а через Ω — связанную компоненту множества $\mathbb{R}^3 \setminus (A_1 \cup A_2 \cup \Delta)$, не пересекающуюся с B . Тогда внутри Ω найдется конец плоскости или катеноида. Кроме того, ∂A_1 является границей гладкой вложенной поверхности Σ , лежащей в Ω , причем вне большего шара $\tilde{B}$, содержащего B , Σ является минимальной поверхностью конечной полной кривизны, разделяющей концы поверхностей A_1 и A_2 , т. е. любой путь в $\mathbb{R}^3 \setminus \tilde{B}$ с концами в A_1 и A_2 пересекает Σ .

Доказательство. Пусть $\Gamma = \partial A_1$. Рассмотрим такую область Ω , что $\partial\Omega = A_1 \cup \Delta \cup A_2$. Если бы граница $\partial\Omega$ не допускала решения задачи Плато, то построение поверхности $\Sigma = \Sigma_\Gamma$ осуществлялось бы в точности так же, как в доказательстве предыдущей леммы. Однако граница ∂B не является выпуклой в среднем относительно Ω . Но мы можем изменить риманову метрику пространства $\mathbb{R}^3$

в окрестности кольца Δ в Ω таким образом, что $\partial\Omega$ в новой метрике не будет допускать решения задачи Плато (подробности см. в. [М.-У.]). После этого можно применять изложенные выше рассуждения. (Эта лемма остается верной даже в случае, когда поверхности A_1 и A_2 являются собственно погруженными; см. [М.-Р.-2].)

6. ТЕОРЕМА О КОЛЬЦЕВЫХ КОНЦАХ И СИЛЬНАЯ ТЕОРЕМА О ПОЛУПРОСТРАНСТВЕ ХОФФМАНА—МИКСА

Здесь представлена идея доказательства следующего важного результата.

Теорема 6.1 [Н.-М.-3]. *Пусть M — собственно вложенная минимальная поверхность в $\mathbb{R}^3$. Тогда M имеет не более двух кольцевых концов бесконечной полной кривизны.*

Набросок доказательства. Пусть A_1 , A_2 и A_3 — различные кольцевые концы поверхности M . Нетрудно найти такой шар B , что $B \cap (A_1 \cup A_2 \cup A_3) = \partial A_1 \cup \partial A_2 \cup \partial A_3$. Используя предыдущую лемму, можно поместить один из концов, скажем A_1 , между стандартными концами E_1 и E_2 (каждый из которых является катеноидным или плоскостным концом).

Можно доказать, что A_1 имеет конечную полную кривизну. Это наиболее сложная часть доказательства. Докажем, что касательная плоскость к A_1 не может быть вертикальной вне некоторого компактного множества (тогда гауссов образ этого подконца лежит в полусфере; следовательно, его площадь меньше чем 2π , и, таким образом, по теореме Барбосы—До Кармо об устойчивости конец A_1 устойчив и потому имеет конечную полную кривизну). Чтобы доказать, что касательная плоскость к A_1 в итоге никогда не будет вертикальной, построим слоения области между E_1 и E_2 на минимальные кольца, границы которых лежат на $E_1 \cup E_2$, а затем изучим касание конца A_1 с таким слоением. Точки касания могут быть лишь седлового типа (это следует из обычного принципа максимума), и простота топологии конца A_1 позволяет показать, что в итоге он будет трансверсален слоению, откуда следует отсутствие вертикальных касательных плоскостей. Таким образом, не более чем два кольцевых конца поверхности M имеют бесконечную полную кривизну.

А могут ли два оставшихся кольцевых конца иметь бесконечную полную кривизну? Это неизвестно, и на сегодняшний день это одна из самых важных задач в данной области. Приведем следующий результат, доказанный автором и Миксом.

Теорема 6.2. [М.-Р.-3]. Пусть M — собственно вложенная минимальная поверхность в $\mathbb{R}^3$ с более чем одним концом. Если A — кольцевой конец поверхности M , то (после поворота поверхности M в $\mathbb{R}^3$) либо A асимптотичен горизонтальной плоскости (а следовательно, имеет конечную полную кривизну), либо $\chi|_A$ — собственная гармоническая функция. В частности, каждый такой конец A конформно представляет собой проколотый диск $D^* = \{z \in \mathbb{C} \mid 0 < |z| \leq 1\}$.

Следствие 6.3. Если M — собственно вложенная минимальная поверхность конечной топологии и с более чем одним концом, то M — поверхность конечного конформного типа.

Следствие 6.4. Если M — собственно вложенное минимальное кольцо, то, поворачивая его, можно добиться того, чтобы M пересекало каждую горизонтальную плоскость по простой замкнутой кривой.

Следствие 6.5. Любая m -поверхность в $\mathbb{R}^3$ с концом геликоидного типа имеет в точности один конец.

Сильная теорема о полупространстве. В полупространстве пространства $\mathbb{R}^3$ существуют полные погруженные неплоские минимальные поверхности. Йорг и Ксавье построили такие примеры для «плиты» [J.-Хав.]. Неизвестно, существуют ли примеры такого рода для шара.

Однако Хоффман и Микс доказали, что для собственного погружения это невозможно. Более того, они получили следующий результат.

Теорема 6.6 (сильная теорема о полупространстве [Н.-М.-4]). Если M_1 и M_2 — непересекающиеся собственно погруженные минимальные поверхности в $\mathbb{R}^3$, то они являются параллельными плоскостями.

Доказательство. Сначала предположим, что M_2 — плоскость (скажем, с координатами (x, y)), а M_1 лежит в верхнем полупространстве. После вертикального сдвига мы можем считать, что $\text{dist}(M_1, M_2) = 0$.

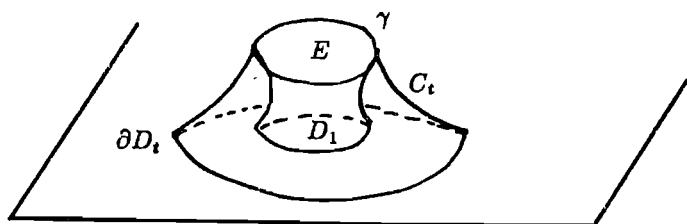


Рис. 14.

Пусть D_t — диск радиуса t в M_2 с центром в начале координат. Поскольку поверхность M_1 собственно погруженная, существует такое $t > 0$, что $\text{dist}(D_1, M_1) > t$. Выберем $t < 1/4$. Пусть кривая γ есть результат сдвига ∂D_1 вертикально вверх на расстояние t . В силу выбора t и D_1 мы получаем, что $\gamma \cup \partial D_1$ является границей устойчивого катеноида C_1 . Для любого $t > 1$ множество $\gamma \cup \partial D_1$ является границей устойчивого катеноида C_t , и катеноид C_{t_1} расположен выше C_{t_2} , если $1 \leq t_2 \leq t_1$. При $t \rightarrow \infty$ катеноиды C_t сходятся к горизонтальной плоскости, расположенной на высоте t , из которой удален диск E с границей γ (рис. 14).

Очевидно, что $E \cup C_t \cup D_t$ является границей компактного топологического шара и пределом таких шаров при $t \rightarrow \infty$ будет «плита», заключенная между M_2 и горизонтальной плоскостью, расположенной на высоте t .

Поверхность M собственно погружена в $\mathbb{R}^3$ и $\text{dist}(M_2, M_1) = 0$; следовательно, существует наименьшее t , такое, что $C_t \cap M_2 = \emptyset$. Но тогда C_t лежит с одной стороны от M_2 в этой точке первого касания, а следовательно, по принципу максимума $C_t = M_2$. Тем самым сильная теорема о полуплоскости доказана для случая, когда M_2 является плоскостью.

Теперь предположим, что M_1 и M_2 — непересекающиеся собственно погруженные поверхности. Найдем плоскость между M_1 и M_2 . Тогда из только что доказанного будет следовать, что M_1 и M_2 тоже являются плоскостями.

Пусть Ω — связная область в $\mathbb{R}^3$, граница которой содержится в $M_1 \cup M_2$ и включает в себя как точки поверхности M_1 , так и точки поверхности M_2 . Заметим, что $\partial\Omega$ не допускает решения задачи Плато.

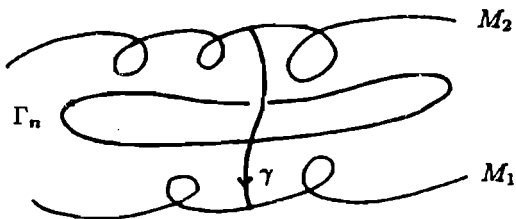


Рис. 15.

Пусть γ — дуга в Ω , соединяющая точку поверхности M_1 с точкой поверхности M_2 , и пусть Γ_n — такие жордановы кривые в Ω , что коэффициент зацепления кривых Γ_n и γ равен единице и Γ_n лежит в дополнении к шару радиуса n с центром в фиксированной точке кривой γ (рис. 15).

Пусть Σ_n — такая гладкая погруженная минимальная поверхность наименьшей площади в Ω , что $\partial \Sigma_n = \Gamma_n$. Как показано в разд. 5, подпоследовательность последовательности Σ_n сходится к полной устойчивой минимальной поверхности $\Sigma \subset \Omega$. Поверхность Σ непуста, поскольку каждая поверхность Σ_n пересекает кривую γ в силу нашего условия на коэффициент зацепления.

По теореме Шёна Σ является плоскостью. Очевидно, что если Σ касается поверхности M_1 или M_2 в какой-либо точке, они тоже будут плоскостями. Это завершает доказательство сильной теоремы о полупространстве.

7. ДВОЯКОПЕРИОДИЧЕСКИЕ МИНИМАЛЬНЫЕ ПОВЕРХНОСТИ

Назовем минимальную поверхность в $\mathbf{R}^3$ периодической, если она связна и инвариантна относительно нетривиальной дискретной группы изометрий, свободно действующих на $\mathbf{R}^3$. Обозначим эту группу через G . Мы изучаем минимальную поверхность в $\mathbf{R}^3/G$, полученную в результате факторизации указанной поверхности в $\mathbf{R}^3$. На самом деле все связные собственно вложенные минимальные поверхности M в $\mathbf{R}^3/G$ могут быть получены таким образом, поскольку по сильной теореме о полупространстве поднятие поверхности M на $\mathbf{R}^3$ есть связная минимальная поверхность, инвариантная относительно G (мы предполагаем, что M не является плоскостью). Заметим, что из этого при наших предположениях относительно M следует сюръективность отображения $\pi_1(M) \rightarrow \pi_1(\mathbf{R}^3/G)$.

Наш главный результат связывает топологию поверхности M с ее полной кривизной $C(M)$.

Теорема 7.1 (о конечной полной кривизне [М.-Р.-4]). Пусть M — собственно вложенная минимальная поверхность в неодносвязном полном плоском трехмерном многообразии N . Тогда поверхность M имеет конечную топологию в том и только в том случае, когда $C(M)$ конечна. Если $C(M)$ конечна, то имеет место равенство

$$C(M) = 2\pi(\chi(M) - W(M)),$$

где $W(M)$ — полное число вращения концов поверхности M (мы определим его позже). Если $N = \mathbb{T}^2 \times \mathbb{R}$, то $W(M) = 0$.

Заметим, что неодносвязность многообразия N предполагать необходимо; геликоид в $\mathbb{R}^3$ имеет бесконечную полную кривизну и конечную топологию.

Обсудим доказательство этой теоремы (по крайней мере для двоякопериодических поверхностей) и приведем некоторые приложения. Для полного плоского трехмерного многообразия существует конечное накрытие поверхностями $\mathbb{T}^3$, $\mathbb{T}^2 \times \mathbb{R}$ или $\mathbb{R}^3/S_\theta$, где S_θ — винтовое движение вокруг оси x_3 с поворотом на угол θ относительно этой оси. Таким образом, в данной теореме мы рассматриваем $\mathbb{T}^2 \times \mathbb{R}$ и $\mathbb{R}^3/S_\theta$ (двоякопериодические и однопериодические поверхности).

Пусть теперь группа G порождена двумя независимыми сдвигами, причем $\mathbb{R}^3/G = \mathbb{T} \times \mathbb{R}$, где $\mathbb{T}$ — плоский двумерный тор. Пусть $x_3: \mathbb{T} \times \mathbb{R} \rightarrow \mathbb{R}$ — третья координатная функция, $T_t = \mathbb{T} \times \{t\}$ — множество уровня функции x_3 на высоте h и $D^* = \{0 < |z| \leq 1\}$ — проколотый диск в $\mathbb{C}$.

Лемма 7.2. Пусть A — кольцо, диффеоморфное диску D^* , и $X: A \rightarrow \mathbb{T} \times \mathbb{R}$ — собственное минимальное погружение кольца A . Тогда A содержит собственное подкольцо A' , которое можно канонично параметризовать диском D^* . В этой параметризации $x_3|_{A'}(z) = c \ln |z|$, где c — постоянная величина.

Доказательство. Пусть $X_3 = x_3 \circ X: A \rightarrow \mathbb{R}$; тогда X_3 — собственное отображение. Поскольку кольцо A имеет один конец, X_3 ограничено снизу или сверху, но не может быть ограничено одновременно снизу и сверху. Поэтому предположим, что X_3 ограничено снизу. Сдвигая $X(A)$ вертикально вниз, мы можем добиться, чтобы граница кольца имела отрицательную координату x_3 и

нуль был регулярным значением отображения X_3 . Следовательно, $\Delta = X_3^{-1}(-\infty; 0]$ — компактное гладкое подмногообразие многообразия A . Оно имеет в точности одну компоненту, содержащую ∂A , а на всех остальных компонентах координата x_3 равна нулю. Из принципа максимума для гармонической функции X_3 следует, что множество Δ связно, а из элементарных топологических соображений следует, что Δ представляет собой кольцо и $A' = X_3^{-1}[0; \infty)$ — собственное подкольцо кольца A .

Функция $X_3|_{A'}$ — собственная неотрицательная гармоническая функция, обращающаяся в нуль на границе. Оставим в качестве несложного упражнения по элементарному комплексному анализу доказательство того факта, что кольцо A' может быть конформно параметризовано диском D^* и $X_3 = c \ln |z|$, где c — некоторая константа.

Пусть теперь M — собственно погруженная минимальная поверхность в $T \times R$ с конечной топологией. Согласно предыдущей лемме, каждый кольцевой конец поверхности M конформно представляет собой диск D^* ; следовательно, M имеет конечный конформный тип. Мы хотим убедиться в том, что если M вложена, то она обладает конечной полной кривизной, так что (переходя к двулистному накрытию) можем считать, что M ориентируема. Тогда отображение Гаусса $g: M \rightarrow S^2$ определено и конформно; два поднятия точки поверхности M в R^3 отличаются друг от друга на сдвиг, который оставляет инвариантным поле ориентированных единичных нормальных векторов к накрытию.

Чтобы доказать, что поверхность M имеет конечную полную кривизну, покажем, что выколотые точки кольцевых концов этой поверхности представляют собой устранимые особенности отображения Гаусса g . Поскольку полная кривизна равна площади сферического образа поверхности M при действии g , выполнения этого условия достаточно. Вообще, можно показать, что выколотая точка не является существенной особенностью, заключив конец A в область пространства, в котором можно проверять значения отображения g . Если вблизи выколотой точки g не принимает многих значений, то особенность устранимая. Теперь мы можем доказать следующую теорему.

Теорема 7.3. Пусть $X: A \rightarrow T \times R$ — собственное минимальное вложение диска D^* . Тогда A имеет конечную полную кривизну.

Доказательство. По лемме 7.2 мы можем положить $A = D^*$ и $X_3(z) = c \ln |z|$; отождествим A с $X(A)$. Выберем такое $c < 0$, что

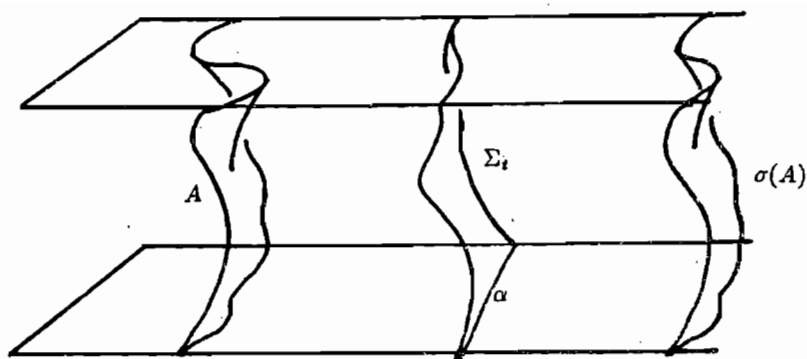


Рис. 16.

$X_3 \geq 0$ на A . Пусть $C_t = A \cap T_t$; тогда C_t — простая замкнутая кривая¹ для каждого t . Доказательство проводится отдельно для двух случаев: C_0 порождает $\pi_1(T_0)$ или не порождает. Мы рассмотрим здесь первый случай; второй случай рассмотрен в работе [М.-Р.-1].

Кривая C_0 порождает циклическую подгруппу G в $\pi_1(T \times \mathbb{R})$. Пусть $p: \widetilde{T \times \mathbb{R}} \rightarrow T \times \mathbb{R}$ — такое риманово накрывающее пространство, что $p_*\pi_1(\widetilde{T \times \mathbb{R}}) = G$. Тогда $\widetilde{T \times \mathbb{R}}$ изометрично $(S^1 \times \mathbb{R}) \times \mathbb{R}$ и порождающий элемент группы $\pi_1(T \times \mathbb{R})/G$ естественным образом действует на множестве $H = p^{-1}(x_3^{-1}[0; \infty))$ как сдвиг. Для удобства поднятие конца A в H также будем обозначать через A . Поскольку граница ∂A компактна, а множество $S^1 \times \mathbb{R} = \partial H$ не является компактным, мы можем выбрать такую замкнутую геодезическую α в ∂H , что $\alpha \cap \partial A = \emptyset$. Выберем такой сдвиг σ в накрывании, что кривая α содержится внутри компактного кольца Δ с границей, состоящей из ∂A и $\sigma(\partial A)$ (рис. 16).

Пусть $\Omega \subset \widetilde{T \times \mathbb{R}}$ — компонента множества $H \setminus (A \cup \sigma A)$, граница которой содержит $A \cup \sigma A$, и пусть Ω_t — множество точек поверхности Ω на высоте, не превышающей t . Заметим, что поверхность Ω_t не допускает решения задачи Плато; ее граница состоит из четырех минимальных поверхностей, которые пересекаются под углами не более чем π .

Пусть α_t — жорданова кривая, лежащая внутри гладкого кольца $\partial \Omega_t$ на высоте t и гомотопная кривой α , и пусть Σ_t — такая вложенная поверхность наименьшей площади, что $\partial \Sigma_t = \alpha \cup \alpha_t$, $\Sigma_t \subset \Omega_t$ (рис. 16). Прежде всего заметим, что поверхность Σ_t ори-

ентируема. Чтобы убедиться в этом, покажем, что Σ_t разделяет Ω_t . Если это не так, то должна существовать простая замкнутая кривая δ в Ω_t , пересекающая Σ_t трансверсально в одной точке. Но группа $\pi_1(\Omega_t)$ порождена $\pi_1(\partial A)$, а следовательно, кривая δ гомотопна кратному границы ∂A и ∂A имеет нулевой индекс пересечения с Σ_t . Поскольку индексы $\mathbb{Z}^2$ -пересечения корректно определены в гомотопических классах, это невозможно, и, значит, поверхность Σ_t ориентируема.

В соответствии с рассуждениями из разд. 5 подпоследовательность поверхностей Σ_t сходится к некоторой гладкой вложенной устойчивой поверхности Σ , $\partial\Sigma = \alpha$, $\Sigma \subset W$. Согласно обычному принципу максимума, $\text{int } \Sigma \subset \text{int } W$.

Теперь докажем, что Σ является частью плоскости. Поскольку α представляет собой фактор прямой линии в $\mathbb{R}^3$, мы можем продолжить Σ с помощью отражения Шварца R_α до собственно вложенной минимальной поверхности $\tilde{\Sigma} \subset \widetilde{T \times \mathbb{R}}$. Заметим, что, поскольку $\sigma(\Omega) \cap \Omega = \sigma(A)$ и $\Sigma \cap \partial\Omega = \alpha$, должно выполняться условие $\sigma\Sigma \cap \Sigma = \emptyset$. Пусть $R_{\sigma\alpha}$ — отражение Шварца относительно $\sigma\alpha$ (т.е. поворот на угол π вокруг $\sigma\alpha$). Мы утверждаем, что $\tilde{\Sigma}$ и $\Sigma' = \sigma\Sigma \cup R_{\sigma\alpha}(\sigma\Sigma)$ — две непересекающиеся собственно вложенные минимальные поверхности. Заметим, что $R_{\sigma\alpha} \circ \sigma = R_{\tilde{\alpha}}$ и $R_{\tilde{\alpha}} \cdot R_{\sigma\alpha} \cdot \sigma = \text{id}$, где $\tilde{\alpha}$ — геодезическая на ∂H , лежащая посередине между α и $\sigma\alpha$. Следовательно, если $\tilde{\Sigma} \cap \Sigma' \neq \emptyset$, то $R_\alpha \Sigma \cap R_{\sigma\alpha}(\sigma\Sigma) \neq \emptyset$. Применив к последнему соотношению симметрию $R_{\tilde{\alpha}}$, получим $R_{\tilde{\alpha}} R_\alpha(\Sigma) \cap \Sigma \neq \emptyset$. Но $R_{\tilde{\alpha}} R_\alpha = \sigma$; следовательно, $\sigma\Sigma \cap \Sigma \neq \emptyset$. Мы получили противоречие.

Поскольку поверхности $\tilde{\Sigma}$ и Σ' не пересекаются, по сильной теореме о полупространстве их поднятия на $\mathbb{R}^3$ являются плоскостями, а следовательно, Σ и $\sigma\Sigma$ представляют собой параллельные плоские кольца в H .

Пусть $P(\theta)$ — плоское кольцо, содержащее α и образующее угол θ с горизонтальной плоскостью ∂H . Выберем угол θ таким малым, чтобы $P(\theta)$ пересекало область, ограниченную поверхностями Σ и $\sigma\Sigma$, по компактному множеству, а также пересекало A трансверсально по гладкой кривой. Это возможно, поскольку A пересекает ∂H трансверсально по одной кривой.

Рассмотрим слоение пространства $T \times \mathbb{R}$ на плоскости, параллельные $P(\theta)$ (на самом деле плоские кольца). Заметим, что каждый лист пересекает $P(\theta)$ по компактному множеству. Это слоение определяется множествами уровня линейной функции, ограниче-

ние которой на A есть собственная гармоническая функция. Следовательно, эта гармоническая функция не имеет критических точек на A над $P(\theta)$. В частности, нормаль к $P(\theta)$ нигде не совпадает с нормальными векторами к части поверхности A , лежащей над $P(\theta)$.

Поскольку θ может принимать различные значения из некоторого интервала, отображение Гаусса не принимает значений, лежащих на некоторой кривой, а значит, выколота точка не является существенной особенностью и A имеет конечную полную кривизну.

Теперь мы можем изучить геометрию концов m -поверхности в пространстве $\mathbf{T} \times \mathbf{R}$ конечного топологического типа. Мы убедимся, что они геометрически сходятся к плоскому кольцу. Прежде чем приступить к доказательству этого утверждения, изучим погруженные поверхности конечной полной кривизны в $\mathbf{T} \times \mathbf{R}$.

Теорема 7.4. Пусть M — собственно погруженная минимальная поверхность конечной полной кривизны в $\mathbf{T} \times \mathbf{R}$, $A_1, \dots, A_l$ — концы поверхности M с вертикальными предельными нормальными векторами и n_i — порядок ветвления отображения Гаусса в конце A_i . Тогда $C(M) = 2\pi(\chi(M) - \sum_{\lambda=1}^l n_\lambda)$. В частности, если M не имеет горизонтальных концов, то $C(M) = 2\pi\chi(M)$.

Доказательство. Если поверхность M неориентируема и мы переходим к ориентированному двулистному накрытию этой поверхности, то все члены приведенной выше формулы умножаются на 2. Это очевидно для $C(M)$ и $\chi(M)$; каждый конец поверхности M поднимается до двух концов в накрытии. Следовательно, мы можем допустить, что M — ориентируемая поверхность.

Пусть $M_t = M \cap (\mathbf{T} \times [-t; t])$. По теореме Гаусса—Бонне

$$C(M) = \lim_{t \rightarrow \infty} \int_{M_t} K = \lim_{t \rightarrow \infty} \left(2\pi\chi(M_t) - \int_{\partial M_t} \kappa_g \right).$$

Для больших значений t выполнено равенство $\chi(M_t) = \chi(M)$, так что мы должны вычислить $\int_{\partial M_t} \kappa_g$.

Прежде всего рассмотрим компоненту C_t множества ∂M_t , обладающую на одном конце E неvertикальным предельным нормальным вектором v . Докажем, что $\int_{C_t} \kappa_g \rightarrow 0$ при $t \rightarrow \infty$. Пусть $\vec{a}$ — горизонтальный единичный вектор, ортогональный вектору v . Поскольку v не является вертикальным, существует в точности два таких вектора. Выберем ориентацию вектора $\vec{a}$ таким образом, чтобы

C'_t сходились к $\vec{a}$ при $t \rightarrow \infty$ (ориентация C'_t задается ориентацией M_t , а штрихом обозначена производная по длине дуги).

Пусть $d\vec{a}$ — замкнутая 1-форма, определяемая ортогональной проекцией на $\vec{a}$ (на прямую, параллельную $\vec{a}$). Так как множество $C_{t_1} \setminus C_{t_2}$ ограничено на E , то $\int_{C_{t_1}} d\vec{a} = \int_{C_{t_2}} d\vec{a}$. Мы имеем $C'_t \rightarrow \vec{a}$ при $t \rightarrow \infty$; следовательно, $\int_{C_t} ds$ сходится к $\int_{C_t} d\vec{a}$. В частности, длины кривых C_t равномерно ограничены.

Пусть X — кономальное векторное поле вдоль C_t , т.е. поле X касается M , $\langle X, C'_t \rangle = 0$, $|X| = 1$ и X направлено внутрь M_t . Пусть $\vec{a}^\perp$ — единичный нормальный вектор к $\vec{a}$, касательный к T_t , направление которого определяется производной C''_t при $C''_t \neq 0$.

Мы имеем

$$\kappa_g(C_t) = \langle C''_t, X \rangle = |C''_t| \cos(\angle(C''_t, X)) = \kappa \cos(\angle(\vec{a}^\perp, X) + \epsilon),$$

где κ — кривизна компоненты C_t , рассматриваемой как плоская горизонтальная кривая, и $\epsilon \rightarrow 0$ при $t \rightarrow \infty$.

Теперь вычислим κ , рассматривая C_t как плоское сечение поверхности M . Пусть P — плоскость в $C_t(s)$, порожденная нормалью n к M в $C_t(s)$ и $C'_t(s)$, и пусть $\kappa_n(s)$ — нормальная кривизна, т.е. кривизна кривой $P \cap M$ в $C_t(s)$.

Мы получаем $\kappa_n(s) = \kappa \cos \psi$, где ψ — угол между $C''_t(s)$ и n . Поскольку предельная нормаль не вертикальна, функция $\cos \psi$ отделена от нуля. Следовательно, если $\kappa \rightarrow 0$ при $t \rightarrow \infty$, то то же верно и для κ и κ_g .

Поскольку поверхность M минимальна, главные кривизны κ_1 , κ_2 этой поверхности равны по модулю. Нормальные кривизны лежат между κ_1 и κ_2 , так что достаточно доказать, что

$$K = \kappa_1 \kappa_2 \rightarrow 0 \quad \text{при} \quad t \rightarrow \infty.$$

В конформной параметризации конца диском D^* индуцированная метрика имеет вид $ds = \lambda |dz|$, где $\lambda = \frac{|f(z)|}{2} (1 + |g(z)|^2)$, f — голоморфная функция, не обращающаяся в нуль на D^* , а g — отображение Гаусса. Кривизна K вычисляется по формуле $K = \frac{-\Delta \ln \lambda}{2\lambda}$. Поскольку функция $\ln |f|$ является гармонической в D^* и g голоморфно продолжается в нуль, мы получаем, что функция $\Delta \ln \lambda = \Delta \ln(1 + |g(z)|^2)$ ограничена в окрестности нуля. Метрика ds является полной в нуле; поэтому $\lambda \rightarrow \infty$ при $|z| \rightarrow 0$. Отсюда мы получаем, что $K \rightarrow 0$, а следовательно, $\int_{C_t} \kappa_g ds \rightarrow 0$ при $t \rightarrow \infty$.

Теперь рассмотрим конец A поверхности M с вертикальным предельным нормальным вектором v . Согласно лемме 7.2, мы можем конформно параметризовать A диском D^* таким образом, чтобы выполнялось равенство $x_3 = C \ln |z|$. Отображение Гаусса имеет в нуле нуль или полюс порядка n , где n — это порядок ветвления конца A .

Мы можем считать, что $g(z)$ представляет собой сумму z^n и членов более высокого порядка малости в окрестности нуля. Если z один раз обходит окружность $|z| = r$ против часовой стрелки, то, в предположении, что r достаточно мало, нормальный вектор к A вдоль C_t обходит вертикальный вектор v n раз, не принимая вертикального направления. Следовательно, нормальный вектор к кривой C_t в T_t монотонно поворачивается вскруг начала координат против часовой стрелки n раз.

Пусть κ — плоская кривизна кривой C_t (в T_t). Согласно результатам последнего абзаца, $\kappa > 0$ и $\int_{C_t} \kappa ds = 2\pi n$. Итак, $\kappa_g = \kappa \cos \phi$, где ϕ — угол между конормалью к C_t в M и горизонталью. Мы видим, что $\phi \rightarrow 0$ при $t \rightarrow \infty$; следовательно, $\int_{C_t} \kappa_g ds \rightarrow 2\pi n$. Это завершает доказательство теоремы 7.4.

Замечание. Из приведенного выше рассуждения можно заключить, что если конец A вложенный, то он не может обладать вертикальным предельным нормальным вектором v . Действительно, если вектор v вертикальный, кривые C_t имеют положительную кривизну κ и $\int_{C_t} \kappa ds = 2\pi n$. Это, очевидно, означает, что C_t — выпуклая кривая, гомотопная нулю в T_t . Следовательно, A поднимается до вложенного конца конечной полной кривизны в $\mathbb{R}^3$, который должен представлять собой катеноид (асимптотически). Очевидно, что он не может быть вложен в $T \times \mathbb{R}$.

Теорема 7.5. Пусть A — собственно вложенный минимальный кольцевой конец в $T \times \mathbb{R}$. Тогда A асимптотичен плоскому цилиндру. Кроме того, два различных кольцевых конца m -поверхности в $T \times \mathbb{R}$ сходятся к различным плоским цилиндрам.

Доказательство. Как говорилось выше, можно считать, что конец A параметризован диском D^* и $X_3 = c \ln |z|$, $c < 0$. В соответствии с рассуждениями из доказательства теоремы 7.4 мы можем допустить, что для любого малого $\epsilon > 0$ существует такое $T > 0$, что при $t > T$ каждая кривая $C_t = A \cap T_t$ содержится внутри ϵ -трубчатой окрестности B_t геодезической, $B_t \subset T_t$. Зафиксируем $\epsilon > 0$ и будем считать, сдвигая, если нужно, A вниз, что кривые C_t

обладают этим свойством для $t \geq 0$. Пусть α_t и β_t — кривые, ограничивающие B_t . Очевидно, что для малых t существует единственное плоское кольцо $F_t \subset T \times R \setminus A$ с границей $\alpha_0 \cup \alpha_t$. Проверим, что такие кольца существуют для всех t и непрерывно изменяются с изменением t . Поскольку множества F_t и $A \cap (T \times [0; t])$ компактны, легко видеть, что множество всех t , для которых F_t существует, открыто. А так как $\partial F_t \cap \partial A = \emptyset$, из принципа максимума следует, что предел таких колец F_t обладает тем же свойством.

Подобным образом мы определим плоские кольца E_t с границей $\beta_0 \cup \beta_t$, не пересекающиеся с A . Подпоследовательность колец E_t сходится к такому плоскому кольцу E , что $\partial E = \alpha_0$, $E \cap A = \emptyset$ (по принципу максимума). Аналогично можно выделить подпоследовательность колец F_t , сходящуюся к плоскому кольцу F , $\partial F = \beta_0$, $F \cap A = \emptyset$ и $E \cap F = \emptyset$. Следовательно, кольца E и F параллельны и находятся на расстоянии ϵ друг от друга, а A лежит между ними. Далее проведем те же самые рассуждения на такой высоте, что кривая C_t попадает в $\epsilon/2$ -окрестность геодезической на T_t . Перейдя к пределу при $\epsilon \rightarrow 0$, получим искомое предельное плоское кольцо.

Из принципа максимума на бесконечности следует, что различные концы сходятся к различным плоским кольцам.

7.6. Глобальные топологические и геометрические свойства. Напомним, что множество $T \times R = R^3/G$ обладает соизмеримой решеткой, если G содержит два линейно независимых вектора равной длины.

Теорема 7.6. Пусть M есть m -поверхность пространства $T \times R$ конечного топологического типа (не плоская). Тогда

1. Если поверхность M ориентируема, то она разделяет $T \times R$. В этом случае как число верхних концов поверхности M , так и число ее нижних концов четно. В частности, поверхность M имеет не менее четырех концов.

2. Если поверхность M неориентируема, то как число ее верхних концов, так и число ее нижних концов нечетно. В частности, независимо от того, является ли поверхность M ориентируемой, общее число концов всегда четно.

3. Верхние концы поверхности M параллельны нижним концам тогда и только тогда, когда подгруппа группы $H_1(T \times R)$, порожденная петлями на концах поверхности M , является циклической. Если концы параллельны, то число верхних концов равно числу нижних концов. В частности, в соответствии с первым

пунктом в случае, когда поверхность M ориентируема, число концов кратно четырем.

4. Если концы поверхности M не параллельны, то они вертикальны и $T \times R$ имеет соизмеримую решетку.

Доказательство. Пусть поверхность M ориентируема, $\widehat{M}$ — связное поднятие M на R^3 , а G — группа сдвигов, определяющая M . Если $\sigma \in G$, то по сильной теореме о полупространстве $\sigma\widehat{M} = \widehat{M}$. Поверхность $\widehat{M}$ разделяет R^3 на две связные компоненты A и B , а σ сохраняет ориентацию; следовательно, $\sigma A = A$. Поэтому поверхность M ограничивает A/G в $T \times R$ и разделяет $T \times R$.

Мы знаем, что для больших t множество $M \cap T_t$ состоит из конечного числа попарно непересекающихся простых замкнутых кривых $C_1, \dots, C_n$ и каждая кривая C_i приближенно является геодезической. Здесь n — число верхних концов. Аналогично, если $t < 0$ и $|t|$ — большое число, то $M \cap T_t = D_1 \cup \dots \cup D_m$, где каждая кривая D_j является почти геодезической, а m — число нижних концов. Поскольку M разделяет $T \times R$ на две компоненты, скажем A и B , каждая кривая C_i имеет две стороны в T_t : одну в A , другую в B . Следовательно, n и m — четные числа. Это доказывает первый пункт. Доказательство второго пункта мы оставляем читателю; см. также [М.-Р.-1].

Пусть P — плоское кольцо, параллельное предельным верхним концам, а Q — плоское кольцо, представляющее собой предел нижних концов. Пусть $\vec{a}$ и $\vec{b}$ — предельные направления кривых C_i и D_j соответственно, $|\vec{a}| = |\vec{b}| = 1$. Обозначим через X конормальное векторное поле к ∂M_t ; оно касается M , $|X| = 1$, $X \perp \partial M_t$ и X направлено вверх. Пусть $\vec{v}$ — поле направленных вверх единичных векторов, касательных к P и нормальных к $\vec{a}$. Аналогично определим $\vec{w}$ как единичное поле, касательное к Q , нормальное к $\vec{b}$ и направленное вверх.

Поток поля $\vec{v}$ через кривую C_j равен $\int_{C_j} \langle \vec{v}, X \rangle ds$. При $t \rightarrow \infty$ поле X сходится к $\vec{v}$, а C_j сходится к геодезической A_j . Следовательно, поток поля $\vec{v}$ через ∂M_t для больших t определяется по формуле

$$\sum_{j=1}^n \int_{|A_j|} \langle \vec{v}, \vec{v} \rangle ds = n|A_1|.$$

Аналогичным образом, поток поля $\vec{v}$ через D_j равен $\int_{D_j} \langle \vec{v}, \vec{w} \rangle |B_j|$, где $|B_j|$ — предельная геодезическая кривой D_j .

Поскольку $\vec{v}$ является градиентом координатной функции, гармонической в M , поток поля $\vec{v}$ через $\partial(M \cap T \times [-t; t])$ равен нулю. Следовательно, $n|A_1| = \langle \vec{v}, \vec{w} \rangle m|B_1|$. В частности, $n|A_1| \leq m|B_1|$, причем равенство достигается в том и только том случае, когда $\vec{v} = \vec{w}$. А перевернув поверхность M , мы можем заключить, что $m|B_1| \leq \langle \vec{v}, \vec{w} \rangle n|A_1|$. Следовательно, $\vec{v} = \vec{w}$ и $n|A_1| = m|B_1|$.

Если поле $\vec{v}$ не вертикально, то существует единственное горизонтальное направление $\vec{a}$, нормальное к $\vec{v}$. Следовательно, $\vec{a} = \vec{b}$ в том случае, когда направление $\vec{v}$ не вертикально и верхние концы параллельны нижним.

Если все концы параллельны, то подгруппа группы $H_1(T \times R)$, порожденная концами, представляет собой циклическую подгруппу, порожденную A_1 . Если концы порождают циклическую подгруппу с порождающим элементом A , то $\vec{a} = \vec{b}$ и $\vec{v} = \vec{w}$, а следовательно, концы параллельны.

Если концы не параллельны, то они вертикальны и $n|A_1| = m|B_1|$. Поскольку направления $\vec{a}$ и $\vec{b}$ независимы, векторы $n|A_1|\vec{a}$ и $m|B_1|\vec{b}$ независимы и имеют равную длину. Следовательно, решетка соизмерима. Это завершает доказательство теоремы 7.6.

Теперь не составит труда привести необходимые условия для того, чтобы концы заданной двоякопериодической минимальной поверхности были непараллельны, что влечет за собой существование соизмеримой решетки в объемлющем пространстве. Доказательство предоставляется читателю; см. также [М.-Р.-1].

Теорема 7.7. Пусть $M \subset T \times R$ — неплоская m -поверхность конечной топологии. Тогда ее концы параллельны, если выполняется одно из следующих трех условий:

- 1) поверхность M ориентируема и число ее концов не кратно четырем;
- 2) поверхность M — область в плоскости;
- 3) $\chi(M)$ нечетно.

7.8. Сумма минимальных поверхностей конечной полной кривизны. Пусть M_1, M_2 — полные неплоские минимальные поверхности конечной полной кривизны в R^3 с отображениями Гаусса g_1, g_2 . Пусть $p_1, \dots, p_n, q_1, \dots, q_m$ — выколотые точки поверхно-

стей M_1 и M_2 соответственно, а $\overline{M_1}, \overline{M_2}$ — соответствующие компактифицированные римановы поверхности. Если зафиксировать единичный вектор $z \in S^2$, можно сложить (в $\mathbf{R}^3$) все точки пространства $\mathbf{R}^3$, для которых вектор z является нормальным. Варьируя z в S^2 , мы получаем полную (разветвленную) минимальную поверхность или точку. Точнее говоря, имеет место следующий результат Розенберга и Тубианы.

Теорема 7.8 [R.-T.-2]. *Множество*

$$M_1 + M_2 = \left\{ \sum_{x \in g_1^{-1}(z)} x + \sum_{y \in g_2^{-1}(z)} y \mid z \in S^2 \setminus W \right\}$$

является полной минимальной поверхностью в $\mathbf{R}^3$ (или точкой), полная кривизна которой равна -4π . Здесь W — некоторое подмножество множества $g_1\{p_1, \dots, p_n\} \cup g_2\{q_1, \dots, q_m\}$.

Нормальный вектор к $M_1 + M_2$ в точке $\sum_{x \in g_1^{-1}(z)} x + \sum_{y \in g_2^{-1}(z)} y$ равен z .

Таким образом, $M_1 + M_2$ естественным образом параметризуется множеством $S^2 \setminus W$. Обозначим соответствующую параметризацию через $\hat{g}$. Если $M_1 + M_2$ не является точкой, то $\hat{g}$ представляет собой конформное инъективное отображение. Это объясняет тот факт, что полная кривизна поверхности $M_1 + M_2$ равна -4π .

(Возможные) точки ветвления поверхности $M_1 + M_2$ являются геометрическими точками ветвления, но, тем не менее, данные Вейерштрасса этой поверхности мероморфны в этих точках ветвления; для этой поверхности $\hat{\omega}$ обращается в нуль в точках ветвления. Заметим, что эти точки отличны от точек ветвления отображения Гаусса. Вообще $\hat{g}$ инъективно в тех точках, в которых $\hat{\omega}$ обращается в нуль.

Операция суммы очень полезна для обнаружения симметрий поверхности M . Так, например, Розенберг и Тубиана доказали следующую теорему.

Теорема 7.9 [R.-T.-2]. *Пусть M — полная минимальная поверхность конечной полной кривизны в $\mathbf{R}^3$. Тогда если все концы поверхности M асимптотичны плоскостям (плоскостные концы), то $M + M$ является точкой.*

Идея доказательства проста. В плоскостном конце поверхности M точки, имеющие фиксированное нормальное направление

(вблизи предельной нормали), распределены в пространстве таким образом, что они имеют общий центр масс (подобно корням из единицы). Следовательно, выколота точка плоскостного конца становится регулярной точкой поверхности $M + M$.

Поскольку $T \times R$ есть абелева группа относительно сложения, а отображение Гаусса инвариантно относительно сдвига, сумма $M_1 + M_2$ также определена в $T \times R$ и ее полная кривизна равна -4π или 0.

Микс и Розенберг доказали следующую теорему.

Теорема 7.10 [М.-Р.-1]. Пусть M — полная погруженная минимальная поверхность конечной полной кривизны в $T \times R$. Тогда

- 1) если концы поверхности M сходятся к параллельным плоским кольцам, то $M + M$ является точкой;
- 2) если M — вложенная поверхность и ее концы не параллельны, то $M + M$ является поверхностью Шерка.

Применяя эту теорему, можно получить следующий результат.

Теорема 7.11. Пусть M есть t -поверхность конечной топологии в $T \times R$ и $T \times R$ обладает несоизмеримой решеткой. Тогда

- 1) поверхность $M + M$ является точкой;
- 2) если M — поверхность первого рода с четырьмя параллельными концами (т. е. седло Картера), то после сдвига этой поверхности (такого, что гауссова кривизна примет нулевое значение в начале координат) точки порядка два группы $(R^2/G) \times R$ станут нулями гауссовой кривизны поверхности M . В этом случае M разделяет $T \times R$ на две изометричные компоненты.

8. ОДНОПЕРИОДИЧЕСКИЕ МИНИМАЛЬНЫЕ ПОВЕРХНОСТИ

Существует множество замечательных примеров однопериодических t -поверхностей. Простейший для понимания пример — геликоид. Его можно построить следующим образом. Возьмем горизонтальную прямую l , пересекающую ось x_3 , и будем поворачивать ее с постоянной скоростью, двигаясь при этом вертикально с постоянной скоростью. Полученная поверхность $\widetilde{M}$ инвариантна относительно винтовых движений S_θ , и для фиксированного θ поверхность $M = \widetilde{M}/S_\theta$ конформно является сферой с двумя выколотыми точками конечной полной кривизны -2θ . В R^3/S_θ поверх-

ность $\widetilde{M}$ имеет два кольцевых конца, каждый из которых является геликоидным. Заметим, что для M уже не существует корректно определенного отображения Гаусса. Отображение Гаусса g поверхности $\widetilde{M}$ индуцирует многозначное мероморфное отображение поверхности M , при котором различные значения отличаются друг от друга на λ^m , где $\lambda = e^{i\theta}$: если p и q — такие точки пространства $\mathbf{R}^3$, что $S_\theta^m(p) = q$, $p, q \in \widetilde{M}$, то нормальные векторы к $\widetilde{M}$ в точках p и q могут быть получены друг из друга поворотом на угол $m\theta$ вокруг оси x_3 ; следовательно, их стереографические проекции на горизонтальную комплексную плоскость отличаются друг от друга множителем λ^m .

С точки зрения представления Вейерштрасса поверхности $\widetilde{M}$ в $\mathbf{R}^3$ она является сопряженной к катеноиду. В $\mathbf{C}^* = \mathbf{C} \setminus (0)$ данные $g(z) = z$, $\omega_\tau(z) = e^{i\tau} dz/z^2$ определяют полную минимальную поверхность $\widetilde{M}_\tau$ для каждого действительного τ ; при $\tau = 0$ это будет катеноид, а при $\tau = \pi/2$ — геликоид. При $0 < \tau < \pi/2$ поверхности $\widetilde{M}_\tau$ не являются вложенными, однако каждая из них имеет два кольцевых конца, которые уже будут вложенными. Каждый конец является концом геликоидно-катеноидного типа. Пересечение поверхности M_τ с большим вертикальным цилиндром радиуса R , ось которого совпадает с осью x_3 , состоит из двух спиральных линий. При $R \rightarrow \infty$ спирали поднимаются по цилиндру подобно функции $\ln R$; таким образом, они похожи на геликоиды и катеноиды (на самом деле одна спираль поднимается, а другая опускается). Далее мы покажем, что в случае, когда $\widetilde{M}$ является вложенной однопериодической поверхностью, кольцевых концов такого типа не существует. Число концов тогда будет четно, и половина из них будет подниматься при $R \rightarrow \infty$, а другая половина — опускаться. Таким образом, поверхность $\widetilde{M}$ не может быть вложенной. Хоффман и Вей показали [H.-Wei], что можно периодическим образом добавить ручку к геликоиду; см. рис. 17, а, б).

Еще один пример однопериодической поверхности — поверхность, сопряженная к двоякопериодической m -поверхности Шерка. В терминах данных Вейерштрасса в $\mathbf{T} \times \mathbf{R}$ $g(z) = z$ и $\omega(z) = dz/(z^4 - 1)$ параметризуют первую поверхность Шерка сферой, из которой выколоты корни четвертой степени из единицы. Читатель легко может убедиться, что функции x_1 и x_2 многозначны, а функция x_3 однозначна. Функции $g(z) = z$, $\omega = idz/(z^4 - 1)$ параметризуют сопряженную поверхность (вторую поверхность Шерка)



Рис. 17,а.

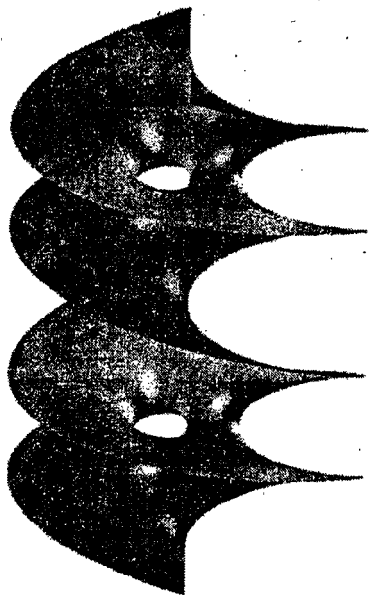


Рис. 17,б.

также сферой S^2 , из которой выколоты корни четвертой степени из единицы. Итак, функция x_3 имеет период, а функции x_2, x_3 однозначны; поверхность инвариантна относительно вертикального сдвига T . Существуют четыре кольцевых конца, которые являются факторами по T вертикальных концов, асимптотичных плоскостям. Такие концы называются концами типа Шерка; см. рис. 18,а).

Кархер показал, что можно построить однопериодические m -поверхности такого вида, имеющие $2n$ концов типа Шерка, для любого $n \geq 2$. Кроме того, он смог деформировать эти поверхности в однопериодические m -поверхности, инвариантные относительно винтовых движений S_θ , таким образом, чтобы концы типа Шерка перешли в концы геликоидного типа [К.-2]; см. рис. 18,б). Это было сделано с использованием обобщенного представления Вейерштрасса, которое мы рассматриваем в этом разделе.

Приведем еще один пример — пример Римана. Рассмотрим поверхность, инвариантную относительно сдвига T (не вертикального) и такую, что горизонтальные сечения $x_3 = \text{const}$ представляют собой окружности или прямые. Концы в $\mathbf{R}^3$ асимптотичны горизонтальным плоскостям (расположенным на высоте, соответствующей

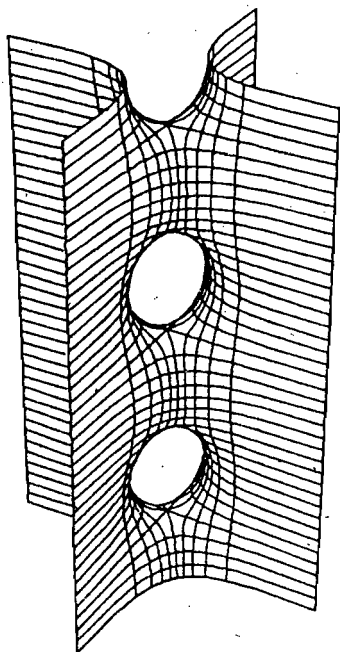


Рис. 18,a.

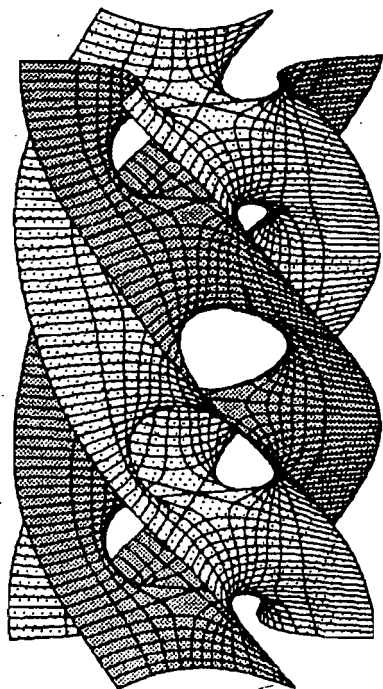


Рис. 18,b.

щей прямым линиям в сечении); см. рис. 19. Простейший ориентируемый фактор в этом случае — тор с двумя выколотыми точками, полная кривизна которого равна -8π . Поверхности из примера Римана образуют однопараметрическое семейство, и сопряженные к ним поверхности также принадлежат этому семейству (это хорошее упражнение для читателя). Превосходное исследование таких поверхностей проведено в [С.-Н.-М.]. Каллахан, Хоффман и Микс [С.-Н.-М.] обобщили пример Римана; см. рис. 20,a). Хоффман и Вей также доказали [Н.-Wei], что можно присоединить к поверхности Римана ручку (одну ручку между каждой парой плоскостных концов) и получить однопериодическую m -поверхность первого рода с тремя выколотыми точками; см. рис. 20,b).

8.1. Теорема о конечной полной кривизне. В теореме 7.1 утверждается, что поверхность M в $\mathbb{R}^3/S_\theta$ есть поверхность конечной топологии тогда и только тогда, когда она имеет конечную

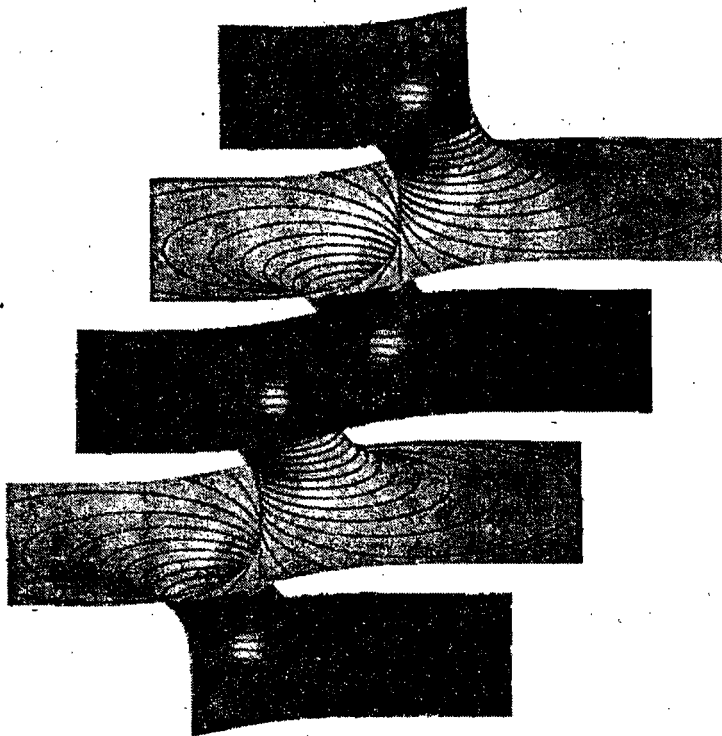


Рис. 19.

полную кривизну. Приведем краткий набросок доказательства. За подробностями решительного читателя мы отсылаем к [М.-Р.-4].

Пусть $M \subset \mathbb{R}^3/S_\theta$ есть m -поверхность конечной топологии. Задача состоит в том, чтобы показать, что (топологически кольцевые) концы имеют конечную полную кривизну. Для этого следует заключить конец поверхности M между *стандартными концами*, т. е. двумя концами конечной полной кривизны, геометрия которых нам понятна. Затем, используя слоение области между стандартными концами (между которыми заключен конец поверхности M) на устойчивые минимальные кольца, можно доказать, что этот конец поверхности M устойчив, а следовательно, имеет конечную полную кривизну.

Для проведения первой части доказательства необходимо понимание геометрии кольцевых концов конечной полной кривизны. В разд. 2 настоящей статьи геометрия вложенных концов конечной

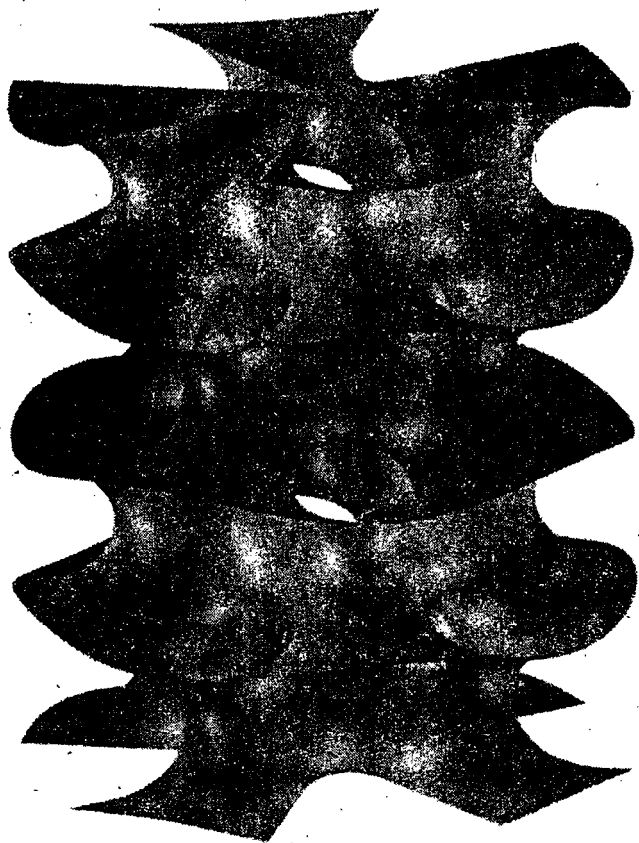


Рис. 20, а.

полной кривизны в $\mathbf{R}^3$ объясняется с использованием представления Вейерштрасса и того факта, что данные Вейерштрасса (g, ω) мероморфно продолжаются в выколотую точку. Для кольцевого конца A в $\mathbf{R}^3/S_\theta = N$ отображение Гаусса многозначно, так что прежде всего нам необходимо установить существование предельной касательной плоскости к A в бесконечности. Предполагая, что конец A имеет конечную полную кривизну, по теореме Губера мы можем заключить, что A может быть конформно параметризован диском D^* . Мы получаем теорему пикаровского типа (при доказательстве которой используется элементарный комплексный анализ).

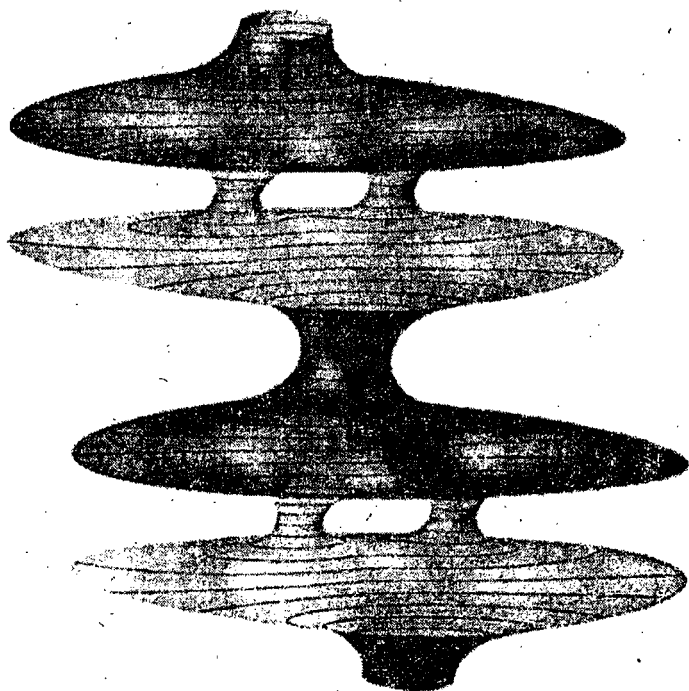


Рис. 20, б.

Теорема 8.2 [М.-Р.-4]. Пусть g — многозначное мероморфное отображение на D^* , $g = \hat{g}(\exp^{-1})$, где $\hat{g}(z + 2\pi i) = \lambda \hat{g}(z)$ при $z \in \{x + iy/x \leq 0\}$ и $|\lambda| = 1$. Если площадь образа отображения g (т. е. ограничения отображения g на диск D^* , разрезанный по радиальной линии), подсчитанная с учетом кратности, конечна, то g непрерывно продолжается в начало координат.

8.3. Обобщенное представление Вейерштрасса. Теперь, используя приведенный выше результат, мы получим представление Вейерштрасса для конца A (мероморфное на D). Поступим следующим образом. Мы можем допустить, что $\lambda \neq 1$, поскольку это обычное представление Вейерштрасса. Тогда предельное значение отображения g равно нулю или бесконечности, поскольку оно не меняется при умножении на λ . Итак, предположим, что $g(0) = 0$. Положим $\theta = 2\pi a$, где $0 < a < 1$. Очевидно, что отображение $z^{1-a}g(z)$ ограничено в окрестности нуля, а следовательно, $g(z) = z^{1-a}h(z)$, где отображение h голоморфно в окрестности нуля. Поэтому $dg/g -$

корректно определенная мероморфная 1-форма на D^* , которая мероморфно продолжается в нуль. С помощью этой формы можно получить многозначное отображение g по формуле $g = \exp(\int dg/g)$.

Заметим, что третья координатная функция x_3 определена на N с точностью до константы; следовательно, форма dx_3 корректно определена на N . Положим $\eta = dx_3 + i(*dx_3)$. Легко видеть, что форма η мероморфна на D^* и мероморфно продолжается в нуль.

Теперь можно взять в качестве данных Вейерштрасса на A пару $(dg/g, \eta)$, которая мероморфно продолжается в выколотую точку. В общем случае справедлив следующий результат.

Теорема 8.4 [М.-Р.-4]. Пусть M — полная минимальная поверхность конечной полной кривизны в $\mathbf{R}^3/S_\theta$. Тогда существуют такая конформная компактификация $\overline{M}$ поверхности M и такие мероморфные 1-формы $(dg/g, \eta)$ на $\overline{M}$, что M параметризуется функцией

$$X(z) = \operatorname{Re} \int \left(d + \frac{1}{g}, ig - \frac{i}{g}, 2 \right) \eta,$$

где $g = \exp(\int dg/g)$.

8.5. Геометрия концов конечной полной кривизны. Теперь, используя полученную параметризацию, опишем асимптотическую геометрию вложенных кольцевых концов.

Теорема 8.5 [М.-Р.-4]. Собственно вложенное минимальное кольцо конечной полной кривизны в $\mathbf{R}^3/S_\theta$ асимптотично плоскости, плоскому вертикальному кольцу (концу типа Шерка) или концу геликоидно-катеноидного типа (с горизонтальной предельной касательной плоскостью). Если конец A является частью t -поверхности конечной полной кривизны, то он не может быть концом геликоидно-катеноидного типа. Если $\theta \neq 0$ и конец A асимптотичен плоскости, то эта плоскость горизонтальна. Если θ иррационально, то A не является концом типа Шерка.

8.6. Число вращения конца. Используя приведенную выше теорему, мы можем вычислить поток и полную кривизну t -поверхности M конечной полной кривизны. Для вычисления последней поступим следующим образом. Пусть $\gamma \subset \mathbf{R}^3/S_\theta$ — фактор оси x_3 , и пусть T_R — трубчатая окрестность радиуса R кривой γ . Для больших R поверхность $M_R = M \cap T_R$ ограничена k жордановыми кривыми $C_1, \dots, C_k$, лежащими на ∂T_R , попарно непересекающимися

и такими, что каждая кривая C_i сходится к вертикальной прямой (конец типа Шерка), к горизонтальной окружности (плоскостной конец) или к спирали на ∂T_R . Теперь можно вычислить полную кривизну поверхности M_R , используя теорему Гаусса—Бонне и переходя к пределу при $R \rightarrow \infty$. Граничный член мы называем числом вращения конца.

Рассмотрим более общую постановку задачи. Пусть A — собственно погруженный кольцевой конец в $\mathbf{R}^3/S_\theta$. Мы знаем, что некоторый подконец конца A не пересекается с γ , и, таким образом, можем считать, что $A \cap \gamma = \emptyset$. Тогда граница ∂A гомотопна циклу на ∂T_R , имеющему вид $n\alpha + m\beta$, где α — горизонтальная окружность на ∂T_R , а β — фактор дуги правостороннего геликоида, которая соединяет точку p с $S_\theta(p)$ и проектируется во вложенный цикл на ∂T_R . Число вращения конца A определяется формулой $\frac{1}{2\pi}|2\pi n + m\theta|$. Нетрудно видеть, что при больших R это число не зависит от R и в случае стандартных концов является пределом полной геодезической кривизны кривых $C_1, \dots, C_k$.

Если M — полная минимальная поверхность конечной полной кривизны в $\mathbf{R}^3/S_\theta$, то число вращения этой поверхности определяется как сумма чисел вращения ее концов. Если поверхность M вложенная, то ее число вращения равно числу вращения одного конца, умноженному на k , где k — число концов.

Теперь формула

$$C(M) = 2\pi(\chi(M) - W(M))$$

из п. 7.1 станет более понятной для читателя. Если концы являются концами типа Шерка, эта формула принимает вид $C(M) = 2\pi\chi(M)$. Если концы плоскостные и их k штук, мы получаем $C(M) = 2\pi(\chi(M) - k)$.

Приложения теоремы о конечной полной кривизне. Мы видели в разд. 7 данной статьи, что неплоская ориентируемая m -поверхность в плоском трехмерном многообразии разделяет пространство (это легко следует из сильной теоремы о полупространстве). Из этого факта, а также из п. 8.1 и утверждений относительно геометрии концов конечной полной кривизны, полученных в п. 8.5, мы получаем следующее топологическое условие, не допускающее существования некоторых m -поверхностей.

Теорема 8.7 [М.-Р.-4]. Пусть M — ориентируемая неплоская m -поверхность конечной топологии в неодносвязном плоском

трехмерном многообразии. Тогда число концов поверхности M четно.

Эрик Тоубиана доказал [Т], что m -поверхность в $\mathbb{R}^3/T$, где T — сдвиг, имеющая конечную ненулевую полную кривизну и обладающая топологией сферы с двумя выколотыми точками (т. е. кольца), представляет собой геликоид. Используя п. 8.1, мы обобщим этот результат на случай $\mathbb{R}^3/S_\theta$.

Теорема 8.8 [М.-Р.-4]. Пусть M — неплоская m -поверхность в $\mathbb{R}^3/S_\theta$, топологически представляющая собой кольцо. Тогда M является геликоидом.

Следующая теорема Переса и Роса представляет собой обобщение теоремы Тоубианы на случай нулевого рода.

Теорема 8.9 [Р.-Ros.]. Геликоид — единственная m -поверхность нулевого рода в $\mathbb{R}^3$ с конечным числом концов геликоидного типа.

Техника доказательства, применяемая Пересом и Росом, основана на использовании деформации Лопеса—Роса, описанной в разд. 3, и теоремы 7.1. Они также доказали (применяя ту же технику), что в $\mathbb{R}^3/S_\theta$, $\theta \neq 0$, не существует m -поверхности первого рода с конечным числом плоскостных концов, т. е., другими словами, невозможно «закрутить» поверхность из примера Римана. Заметим, что деформация Кархера однопериодической поверхности Шерка показывает [К.-2], что поверхность Шерка можно «закрутить»; см. рис. 18, б).

Из теоремы 8.8 можно получить следующую теорему единственности для геликоида в $\mathbb{R}^3$.

Теорема 8.10 [М.-Р.-4]. Плоскость и геликоид — единственные односвязные m -поверхности в $\mathbb{R}^3$ с бесконечной группой симметрий.

Еще одно приложение позволяет классифицировать суммы поверхностей.

Теорема 8.11 [М.-Р.-4]. Пусть $M \subset \mathbb{R}^3/T$ есть m -поверхность конечной топологии и T — нетривиальный сдвиг. Если поверхность M имеет геликоидный конец, то $M + M$ является геликоидом. Если M имеет плоскостной конец, то $M + M$ является

точкой. Если M имеет четыре конца типа Шерка, то $M + M$ — поверхность Шерка.

Каллахан, Хоффман и Микс доказали следующую полезную структурную теорему для однопериодических m -поверхностей с более чем одним концом.

Теорема 8.12 [С.-Н.-М.]. Пусть $M \subset \mathbf{R}^3$ есть m -поверхность с бесконечной группой симметрий и более чем одним концом. Тогда либо поверхность M является катеноидом, либо для нее выполняются следующие условия:

1) группа $\text{Sym}(M)$ содержит бесконечную циклическую подгруппу S , имеющую конечный индекс, порожденную винтовым движением S_θ ;

2) поверхность M/S обладает конечной топологией в точности в тех случаях, когда она имеет конечную полную кривизну;

3) существует плоскость, пересечение которой с M состоит из конечного числа простых замкнутых кривых.

Как следствие из этой теоремы Каллахан, Хоффман и Микс доказали, что двоякопериодическая m -поверхность в $\mathbf{R}^3$ имеет один конец.

9. НЕКОТОРЫЕ ЗАДАЧИ, ГИПОТЕЗЫ И СМЕЖНЫЕ ВОПРОСЫ

Возможно, в $\mathbf{R}^3$ существует много m -поверхностей бесконечной полной кривизны и конечной топологии. Однако на данный момент нам известен только один пример — геликоид. Есть ли другие примеры? Марк Соре доказал, что вблизи геликоида (т. е. среди графов над геликоидом в ϵ -трубчатой окрестности геликоида) других примеров нет [М.-С.].

Вопрос менее общего характера состоит в том, чтобы выяснить, являются ли геликоид и плоскость единственными односвязными m -поверхностями в $\mathbf{R}^3$. Мы знаем, что это верно в случае, если поверхность имеет бесконечную группу симметрий (см. теорему 8.10).

Возможно, каждая m -поверхность бесконечной полной кривизны в $\mathbf{R}^3$ обладает бесконечной группой симметрий (хотя я в этом сомневаюсь), и в этом случае ответ будет утвердительным. Все известные на сегодняшний день примеры m -поверхностей бесконечной полной кривизны были построены с использованием симметрий, но нет (насколько я могу судить) никакой убедительной

причины полагать, что не существует других примеров. Может быть, можно добавить к геликоиду всего одну ручку (а может быть, и больше), и мы получим непериодическую m -поверхность бесконечной полной кривизны.

Случай, когда m -поверхность в $\mathbb{R}^3$ имеет более чем один конец, принципиально отличается от рассмотренного выше. В разд. 5 показано, что в этом случае можно найти плоскостные или катеноидные концы в дополнении к M . Хоффман и Микс доказали, что для таких поверхностей M не более двух кольцевых концов имеют бесконечную полную кривизну (см. теорему 5.1). Это побудило их выдвинуть следующую гипотезу.

Гипотеза о конечной полной кривизне [Н.-М.-3]. *Кольцевой конец m -поверхности в $\mathbb{R}^3$, имеющей не менее двух концов, имеет конечную полную кривизну.*

Это утверждение сходно с гипотезой Ницше, заключающейся в том, что минимальная поверхность, которая пересекает каждую горизонтальную плоскость по жордановой кривой, есть катеноид. Ницше доказал эту гипотезу в предположении, что жордановы кривые звездообразны [N].

Далее мы с Миксом доказали (см. теорему 6.2), что кольцевой конец m -поверхности в $\mathbb{R}^3$, имеющей не менее двух концов, либо имеет конечную полную кривизну, либо содержит подконец, который пересекает каждую горизонтальную плоскость в верхнем полупространстве пространства $\mathbb{R}^3$ по жордановой кривой (после евклидова сдвига поверхности).

Таким образом, гипотеза о конечной полной кривизне является следствием утвердительного ответа на следующую гипотезу, принадлежащую Миксу и автору.

Обобщенная гипотеза Ницше [М.-Р.-3]. *Пусть A — такой минимальный кольцевой конец, что $A \cap \{x_3 = c \geq 0\}$ — жорданова кривая для любого $c \geq 0$. Тогда A имеет конечную полную кривизну.*

Заметим, что данный вопрос связан с голоморфной функцией g в проколоте диске, и проблема состоит в выяснении того, является ли начало координат существенной особенностью. Поскольку конец A (или его подконец) может быть конформно параметризован таким диском D^* , что $x_3 = K \ln |z|$, данные Вейерштрасса для этого конца имеют вид $(g, 1/zg)$. Однако представляется затруднительным связать особенность функции g в начале координат со свойством вложенности конца A . Мы с Эриком Тубиа-

ной построили примеры погруженных колец бесконечной полной кривизны, трансверсально пересекающих каждую горизонтальную плоскость (g имеет существенную особенность [R.-T.-1]). Мы даже построили такие погружения в «плиту» в $\mathbf{R}^3$. Время от времени я обращаюсь к гипотезе Ницше, используя технику комплексного анализа (той его части, которая касается существенных особенностей), но не используя того факта, что конец A является вложенным. К счастью, такое случается со мной не так уж часто.

Утвердительный ответ на обобщенную гипотезу Ницше означал бы, что m -поверхности конечной топологии, имеющие более одного конца, могут быть параметризованы мероморфными данными на компактной римановой поверхности. Все известные нам примеры собственно вложенных m -поверхностей в $\mathbf{R}^3$ должны обладать этим свойством: все поверхности с бесконечной полной кривизной, которые мы знаем, периодические и имеют факторы конечной топологии. В разд. 8 показано, что обобщенное представление Вейерштрасса мероморфно на компактной римановой поверхности.

Какие m -поверхности в $\mathbf{R}^3$ имеют в точности один конец (топологически кольцевой)? К настоящему времени из таких поверхностей мы знаем только плоскость и геликоид, но, как я говорил выше, возможно, что добавление ручки к геликоиду дает такой же пример. Может быть, можно реализовать все компактные поверхности произвольного рода с одним проколом. Назовем кольцевой конец алгебраическим, если он конформно является проколотым диском и dg/g и η мероморфно продолжаются в выколотую точку. Является ли каждая m -поверхность конечной топологии в $\mathbf{R}^3$ алгебраической? Является ли собственно вложенный минимальный кольцевой конец алгебраическим? Можно ли по крайней мере определить, является ли он конформно проколотым диском? Я могу доказать, что минимально погруженное кольцо, полная кривизна которого растет полиномиально (т. е. $\int_{D_r} |K| \leq cr^n$, где D_r — геодезический диск радиуса r), конформно представляет собой проколотый диск. Условия на рост могут подразумевать алгебраичность. Интересная смежная проблема состоит в изучении минимальных поверхностей, пересечение которых с каждой из плоскостей $x_3 = \text{const}$ есть одна и та же собственно вложенная действительная прямая. Совпадает ли эта поверхность конформно с S ? Будет ли это плоскость или геликоид?

Какие m -поверхности в $\mathbf{R}^3$ имеют нулевой род? Единственные известные нам примеры — плоскость, геликоид и однопери-

одическая поверхность из примера Римана. Микс предположил [М.-1], что если поверхность также является и периодической, то она относится к одному из этих трех типов.

Какие m -поверхности имеют нулевой род в плоских трехмерных многообразиях, отличных от $\mathbf{R}^3$? Мы думаем, что в $\mathbf{T}^2 \times \mathbf{R}^2$ все такие поверхности поднимаются до поверхности Шерка в $\mathbf{R}^3$ (заметим, что поверхность Шерка имеет бесконечный род в $\mathbf{R}^3$). Мы с Миксом доказали этот факт для случая поверхности с четырьмя концами [М.-R.-1], а Вей обобщил результат на случай шести концов [Wei].

По теореме Персса—Роса 8.9 в пространстве $\mathbf{R}^3/S_\theta$, $\theta = 0$, единственной поверхностью нулевого рода конечной топологии, имеющей концы геликоидного типа, является геликоид. Какие примеры поверхностей нулевого рода существуют в $\mathbf{R}^3/S_\theta$? Заметим, что поверхность из примера Римана имеет нулевой род в $\mathbf{R}^3$ и род два в $\mathbf{R}^3/\mathbf{T}$.

(Чрезвычайно) общий вопрос состоит в классификации m -поверхностей рода g конечной топологии в $\mathbf{T}^2 \times \mathbf{R}$ или в $\mathbf{R}^3/S_\theta$. Я полагаю, что при $g = 0$ и $g = 1$ задача уже сейчас доступна для нашего понимания. Но конечно, решение подобной задачи для $\mathbf{R}^3$ нам пока не по силам. До недавнего времени единственными известными примерами двоякопериодических поверхностей были накрытия поверхности Шерка или седла Кархера вместе со своими семействами. Эти примеры были построены мною и Миксом [М.-R.-1]. Затем Ф. Вей нашел очень красивый пример построения двоякопериодической поверхности второго рода с двумя верхними и двумя нижними концами, которые все параллельны (он использовал сопряженную технику Плато или представление Вейерштрасса). Этот пример отличался от всех известных ранее. Поверхность Вея не содержит прямых линий, как поверхность Шерка и седло Кархера [Wei]; см. рис. 21,а). Используя идею Вея, Кархер, как он сообщил мне в личной беседе, смог добавить ручку к поверхности Шерка (так, чтобы концы новой поверхности вели себя так же, как у поверхности Шерка, и чтобы новая поверхность была поверхностью первого рода); см. рис. 21,б). Рабах Соуам доказал, что ни поверхность Вея, ни поверхность Кархера не могут существовать, если попытаться разместить на них четыре вертикальные прямые [Souam].

Дж. Хоффман предполагает, что если M есть m -поверхность конечной полной кривизны в $\mathbf{R}^3$, то число концов этой поверхности не превосходит суммы рода поверхности M и числа 2. Он счита-



Рис. 21, а.

ет, что, добавив конец к вложенной минимальной поверхности конечной полной кривизны в $\mathbb{R}^3$, мы увеличим род поверхности (это противоречит примеру Римана).

Еще один интересный объект для изучения — соотношение между внутренними изометриями m -поверхности M (т. е. группой симметрий) и внешними изометриями, оставляющими поверхность M инвариантной (группой изометрий). Микс предположил, что, когда M является m -поверхностью в $\mathbb{R}^3$, любая симметрия поверхности M продолжается до симметрии пространства $\mathbb{R}^3$. Микс и автор этой статьи доказали этот факт для двоякопериодических m -поверхностей; в действительности доказано большее (жесткость): пусть $f_1 : M \rightarrow \mathbb{R}^3$ — двоякопериодическая m -поверхность, и предположим, что $f_2 : M \rightarrow \mathbb{R}^3$ — другое минимальное погружение поверхности M ; тогда существует такая изометрия ϕ пространства $\mathbb{R}^3$, что $\phi f_2 = f_1$ [М.-Р.-1]. Чоу, Микс и Вей доказали,

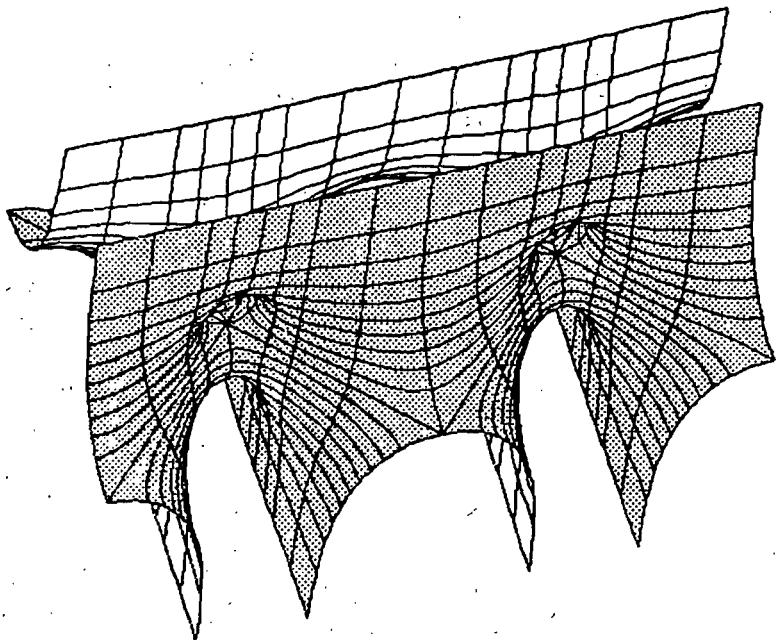


Рис. 21, b.

что m -поверхность в $\mathbb{R}^3$, обладающая более чем одним концом, является жесткой [С.-М.-В.].

Однопериодические m -поверхности в $\mathbb{R}^3$ не жесткие (например, геликоид), однако верно, что их группа симметрий совпадает с группой изометрий, когда M/S_θ имеет конечную топологию [М.-1].

Микс также предположил, что неодносвязные m -поверхности в $\mathbb{R}^3$ жесткие (может быть, геликоид — единственная нежесткая m -поверхность в $\mathbb{R}^3$?): любое другое изометрическое собственное погружение поверхности M конгруэнтно M [М.-1].

Возможно, понятие жесткости следует ограничить изометрическими минимальными вложениями поверхности M (а не всеми погружениями). Тогда, вероятно, геликоид будет жесткой поверхностью.

Микс распространил теорему 7.1 о конечной полной кривизне на двоякопериодические поверхности конечного рода. Он доказал, что M -поверхность в $T^2 \times \mathbb{R}$ конечного рода имеет конечную полную кривизну [М.-1]. Остается ли это верным в $\mathbb{R}^3/S_\theta$?

При какой степени обобщения принцип максимума на бесконечности остается справедливым? Можно ли изменить предположение о компактности ∂M_1 и ∂M_2 ? Минимум расстояния между M_1 и M_2 (в предположении, что они не пересекаются) не обязательно достигается на *внутренних точках на бесконечности*.

Антонио Рос задал мне следующий вопрос в том же направлении: если M есть m -поверхность в $\mathbf{R}^3$, может ли ее конец быть точкой накопления других концов этой поверхности? Более точно, могут ли существовать сходящаяся последовательность x_n точек конца A поверхности M и последовательность $y_n \in M \setminus A$, такие, что $\text{dist}(x_n, y_n) \rightarrow 0$ при $n \rightarrow \infty$?

При изучении m -поверхностей конечной полной кривизны в $\mathbf{R}^3$ возникает следующий важный вопрос: имеют ли концы катеноидного типа одну и ту же ось? Ответ неизвестен, даже когда поверхность обладает тремя такими концами.

Важная работа в направлении классификации m -поверхностей конечной полной кривизны в $\mathbf{R}^3$ была выполнена Селсу Костой. Он классифицировал такие поверхности полной кривизны -12π [Cost.-3]. Его доказательство использует сложные вычисления из теории эллиптических функций. Было бы интересно понять этот результат с другой точки зрения.

Недавно была сделана красивая и важная работа по минимальным поверхностям, которую я здесь не обсуждаю. Это мое большое упущение. Теорема Фромана и Микса утверждает, что две m -поверхности в $\mathbf{R}^3$ с одним концом, обладающие одинаковым родом, объемлемо изотропны (ambiently isotropic) [F.-M.].

Есть также очень красивая работа Фудзимото о значениях отображения Гаусса [Fuj.-1, 2], но мне осталось неясным, можно приписать ее к вложенным поверхностям или нет.

Микс и Вей изучали пространство минимальных подмногообразий $\mathbf{R}^3$, ограниченных двумя выпуклыми жордановыми кривыми C_1 и C_2 . Они доказали, что когда C_1 и C_2 находятся в параллельных плоскостях, существует 0, 1 или 2 минимальных кольца с границей $C_1 \cup C_2$ [M.-Wh.].

Наконец, я упомяну проблему десингуляризации (разрешения особенностей) минимального многообразия: когда для данных двух m -поверхностей M_1 и M_2 в $\mathbf{R}^3$ существует m -поверхность M , которая близка к $M_1 \cup M_2$ вне некоторой окрестности $M_1 \cup M_2$? Во многих примерах десингуляризация поверхности M выглядит как серия ручек вдоль $M_1 \cup M_2$. Вот некоторые примеры. Однопери-

одическая поверхность Шерка является десингуляризацией двух плоскостей. Однопериодическое обобщение этой поверхности Шерка, построенное Кархером, есть десингуляризация n плоскостей, пересекающихся вдоль некоторой оси; см. рис. 18, а).

Геликоид и его образ при повороте вокруг его оси пересекаются по этой оси. Пример Кархера десингуляризации (в общем случае n геликоидов, пересекающихся вдоль оси) осуществляется набором ручек вдоль этой оси (см. рис. 18, б).

m -поверхность Косты конечной полной кривизны с тремя концами можно рассматривать как десингуляризацию вертикального катеноида и горизонтальной плоскости, проходящей через «горловину» катеноида; см. рис. 19. Примеры поверхностей большего рода с тремя концами, построенные Хоффманом и Миксом, лучше иллюстрируют этот эффект (рис. 20): ручки располагаются вдоль окружности пересечения катеноида и горизонтальной плоскости.

В случае когда $M_1 \cap M_2$ — жорданова кривая C , необходимое условие существования десингуляризации выглядит следующим образом: $\int_C \nu \cdot n = 0$, где ν — кономаль к C в M_1 и n — нормаль к M_2 вдоль C .

Не совсем понятен смысл этого условия. Какой должна быть, например, *минимальная перестройка* для того, чтобы $M_1 \cup M_2$ обладала десингуляризацией?

ЛИТЕРАТУРА

- [B.Do C.] Barbosa J.L., Do Carmo M. On the size of a stable minimal surface in $\mathbf{R}^3$, Amer. J. Math., **98** (1976), No. 2, 515–528.
- [C.-H.-M] Callahan M., Hoffman D., Meeks W.H., III. The structure of singly-periodic minimal surfaces, Invent. Math., **99** (1990), No. 3, 455–481.
- [Cost.-1] Costa C. Imersões mínimas em $\mathbf{R}^3$ de gênero um e curvatura total finita. PhD thesis, IMPA, Rio de Janeiro, Brazil, 1982.
- [Cost.-2] Costa C. Example of a complete minimal immersion in $\mathbf{R}^3$ of genus one and three embedded ends, Bull. Soc. Bras. Mat., **15** (1984), No. 3, 47–54.
- [Cost.-3] Costa C. Uniqueness of minimal surfaces embedded in $\mathbf{R}^3$ with total curvature -12π , J. Differential Geom., **30** (1989), No. 3, 597–618.
- [Cour.] Courant R. Dirichlet's Principle, Conformal Mapping and Minimal Surfaces, Interscience Publishers, New York, 1950.
- [Darb.] Darboux G. Leçons sur la théorie générale des surfaces et les applications géométriques du calcul infinitésimal, Gauthier-Villars, Paris, 1st part, 2nd edition, 1914.

- [Do C.-P.] Do Carmo M., Peng C.K. Stable minimal surfaces in $\mathbb{R}^3$ are planes, *Bull. Amer. Math. Soc.*, **1** (1979), 903-906.
- [Doug.] Douglas J. Solution of the problem of Plateau, *Trans. Amer. Math. Soc.*, **33** (1931), 263-321.
- [F.-Oss.] Finn R., Osserman R. On the Gauss curvature of non-parametric minimal surfaces, *J. Anal. Math.*, **12** (1964), 351-364.
- [F.C.] Fischer-Colbrie D. On complete minimal surfaces with finite Morse index in 3-manifolds, *Invent. Math.*, **82** (1985), 121-132.
- [Fr.-M.] Frohman C., Meeks W.H., III. The topological uniqueness of complete one-ended minimal surfaces and Heegard surfaces in $\mathbb{R}^3$, Preprint.
- [Fuj.-1] Fujimoto H. On the number of exceptional values of the Gauss maps of minimal surfaces, *J. Math. Soc. Japan*, **40** (1988), No. 2, 235-247.
- [Fuj.-2] Fujimoto H. Modified defect relations for the Gauss map of minimal surfaces, *J. Differential Geom.*, **29** (1989), 245-262.
- [G.-T.] Gilbarg D., Trudinger N.S. *Elliptic Partial Differential Equations of Second Order*, Springer-Verlag, New York, 2nd edition, 1983.
- [H.-S.] Hardt R., Simon L. Boundary regularity and embedded minimal solutions for the oriented Plateau problem, *Ann. of Math.*, **110** (1979), 439-486.
- [Heinz] Heinz E. Über die Lösungen der Minimallächengleichung, *Nachr. Akad. Wiss. Göttingen Math. Phys.*, **K1**, II (1952), 51-56.
- [H.-M.-1] Hoffman D., Meeks W.H., III. A complete embedded minimal surface in $\mathbb{R}^3$ with genus one and three ends, *J. Differential Geom.*, **21** (1985), 109-127.
- [H.-M.-2] Hoffman D., Meeks W.H., III. Properties of properly embedded minimal surfaces of finite total curvature, *Bull. Amer. Math. Soc.*, **17** (1987), No. 2, 296-300.
- [H.-M.-3] Hoffman D., Meeks W.H., III. The asymptotic behavior of properly embedded minimal surfaces of finite topology, *J. Amer. Math. Soc.*, **2** (1989), No. 4, 667-681.
- [H.-M.-4] Hoffman D., Meeks W.H., III. The strong halfspace theorem for minimal surfaces, *Invent. Math.*, **101** (1990), 373-377.
- [H.-M.-5] Hoffman D., Meeks W.H., III. Minimal surfaces based on the catenoid, *Amer. Math. Monthly*, Special Geometry Issue, **97** (1990), No. 8, 702-730.
- [H.-Wei] Hoffman D., Wei F. Adding handles to the helicoid, Preprint.
- [E.H.] Hopf E. On an inequality for minimal surfaces $z = f(x, y)$, *J. Rat. Mech. Anal.*, **2** (1953), 519-522.
- [Hub.] Huber A. On subharmonic functions and differential geometry in the large, *Comment. Math. Helv.*, **32** (1957), 181-206.
- [J.-S.] Jenkins H., Serrin J. Variational problems of minimal surface type II, *Arch. Rat. Mech. Analysis*, **21** (1966), 321-342.
- [J.-Xav.] Jorge L., Xavier F. A complete minimal surface in a slab of $\mathbb{R}^3$, *Ann. of Math.*, **112** (1980), No. 1, 203-206.

- [K.-1] Karcher H. Construction of minimal surfaces, in: Surveys in Geometry, 1-96, University of Tokyo, 1989, and Lecture Notes No. 12, SFB256, Bonn, 1989.
- [K.-2] Karcher H. Embedded minimal surfaces derived from Scherk's examples, *Manuscripta Math.*, **62** (1988), 83-114.
- [K.-3] Karcher H. The triply periodic minimal surfaces of Alan Schoen and their constant mean curvature companions, *Manuscripta Math.*, **64** (1989), 291-357.
- [K.-4] Karcher H. Construction of higher genus embedded minimal surfaces, in: Geometry and Topology of Submanifolds, III (Leeds, 1990), World Sci. Publ., 1991, 174-191.
- [L.-R.] Langevin R., Rosenberg H. A maximum principle at infinity for minimal surfaces and applications, *Duke Math. J.*, **57** (1988), 819-828.
- [Lo.-Ros] Lopez F. J., Ros A. On embedded complete minimal surfaces of genus zero, *J. of Differential Geom.*, **33** (1991), No. 1, 293-300.
- [M.-1] Meeks W. H., III. The geometry, topology and existence of periodic minimal surfaces, Preprint.
- [M.-2] Meeks W. H., III. Lectures on Plateau's Problem, Instituto de Matematica Pura e Aplicada (IMPA), Rio de Janeiro, Brazil, 1978.
- [M.-3] Meeks W. H., III. The theory of triply-periodic minimal surfaces, *Indiana Univ. Math. J.*, **39** (1990), No. 3, 877-936.
- [M.-R.-1] Meeks W. H., III, Rosenberg H. The global theory of doubly periodic minimal surfaces, *Invent. Math.*, **97** (1989), 351-379.
- [M.-R.-2] Meeks W. H., III, Rosenberg H. The maximum principle at infinity for minimal surfaces in flat three-manifolds, *Comment. Math. Helv.*, **65** (1990), 255-270.
- [M.-R.-3] Meeks W. H., III, Rosenberg H. The geometry and conformal structure of properly embedded minimal surfaces of finite topology in $\mathbb{R}^3$, *Invent. Math.*, **114** (1993), No. 3, 625-639.
- [M.-R.-4] Meeks W. H., III, Rosenberg H. The geometry of periodic minimal surfaces, *Comment. Math. Helv.*, **68** (1993), No. 4, 538-578.
- [M.-Wh.] Meeks W. H., III, White B. Minimal surfaces bounded by convex curves in parallel planes, *Comment. Math. Helv.*, **66** (1991), 263-278.
- [M.-Y.] Meeks W. H., III, Yau S. T. The existence of embedded minimal surfaces and the problem of uniqueness, *Math. Z.*, **179** (1982), 151-168.
- [N.] Nitsche J. C. C. A characterization of the catenoid, *J. Math. Mech.*, **11** (1962), 293-302.
- [Oss.-1] Osserman R. Global properties of minimal surfaces in E^3 and E^n , *Ann. of Math.*, **80** (1964), No. 2, 340-364.
- [Oss.-2] Osserman R. On the Gauss curvature of minimal surfaces, *Trans. Amer. Math. Soc.*, **96** (1960), 115-128.
- [P.-Ros] Perez J., Ros A. Some uniqueness and nonexistence theorems for embedded minimal surfaces, Preprint.

- [Rado-1] Rado T. The problem of the least area and the problem of Plateau, *Math. Z.*, **32** (1930), 763-796.
- [Rado-2] Rado T. On the problem of Plateau, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Springer-Verlag, Berlin, 1933.
- [Reif.] Reifenberg R. Solution for the Plateau problem for m -dimensional surfaces of varying topological type, *Acta Math.*, **104** (1960), 1-94.
- [R.-T.-1] Rosenberg H., Toubiana E. A cylindrical type complete minimal surface in a slab of $\mathbb{R}^3$, *Bull. Sci. Math.*, **111** (1987), 241-245.
- [R.-T.-2] Rosenberg H., Toubiana E. Complete minimal surfaces and minimal herissons, *J. Differential Geom.*, **28** (1988), 115-132.
- [R.-S.E.] Sa Earp, Rosenberg H. The Dirichlet problem for the minimal surface equation on unbounded planar domains, *J. Math. Pures Appl.*, **68** (1989), 163-183.
- [Sch.-1] Schoen R. Uniqueness, symmetry, and embeddedness of minimal surfaces, *J. Differential Geom.*, **18** (1983), 791-809.
- [Sch.-2] Schoen R. Estimates for Stable Minimal Surfaces in Three Dimensional Manifolds, *Annals of Math. Studies*, vol. 103, Princeton University Press, 1983.
- [Simon] Simon L. Lectures on geometric measure theory. Proceedings of the Center for Mathematical Analysis, vol. 3, Australian National University, Canberra, 1983.
- [Smale] Smale N. A bridge principle for minimal and constant mean curvature submanifolds of $\mathbb{R}^n$, *Invent. Math.*, **90** (1987), 505-549.
- [M.S.] Soret M. Deformations de surfaces minimales, Thèse Univ. Paris VII, 1992.
- [Souam] Souam R. Stabilité et unicité des surfaces minimales, Thèse Univ. Paris VII, 1992.
- [T.] Toubiana E. On the uniqueness of the helicoid, *Ann. Inst. Four.*, **38** (1988), 121-132.
- [Wei] Wei F. Some existence and uniqueness theorems for doubly periodic minimal surfaces, *Invent. Math.*, **109** (1992), No. 1, 113-115.
- [Wh.] White B. Complete surfaces of finite total curvature, *J. Differential Geom.*, **26** (1987), 315-326.

HAROLD ROSENBERG

Université de Paris VII
 U.F.R. de Mathématiques
 URA 212 du C.N.R.S.
 Tour 45-55 — 5eme etage
 2, place Jussieu
 F-75251 PARIS CEDEX 05