PROCEEDINGS OF THE
INTERNATIONAL CONGRESS OF
MATHEMATICIANS

AUGUST 3–11, 1986

Berkeley, California, USA

Edited by
Andrew M. Gleason

American Mathematical Society
1987

МЕЖДУНАРОДНЫЙ КОНГРЕСС МАТЕМАТИКОВ В БЕРКЛИ, 1986

Обзорные доклады

Перевод с английского
и немецкого
под редакцией
В. М. Тихомирова

Москва «Мир» 1991

ББК 22.1
М43
УДК 51

Авторы: Атья М., Мазур Б., Милнор Дж., Штрассен Ф., де Бранж Л., Доналдсон С. К., Фальтингз Г., Геринг Ф. У., Громов М., Ленстра-мл. Х. В., Шён Р., Шёнхаге А., Шелах С., Скороход А. В., Смейл С., Стейн И. М., Суслин А. А., Воуган-мл Д. А., Виттен Э.

Международный конгресс математиков в Беркли, 1986.
М43 Обзорные доклады: Сб. статей: Пер. с англ./Сост. В. М. Тихомиров.— М.: Мир, 1991.— 456 с., ил.
ISBN 5-03-002013-6

В сборнике представлены часовые доклады, сделанные на конгрессе, а также сообщения с изложением работ филдсовских медалистов С. Доналдсона (США), Г. Фальтингза (ФРГ), М. Фридмана (США) и лауреата премии Неванлинны Л. Вальянта (США). В числе авторов такие известные ученые, как М. Атья, Дж. Милнор, С. Смейл, Э. Виттен, Л. де Бранж, Х. Ленстра, М. Громов из Великобритании, ФРГ, США, Франции, Нидерландов, Швейцарии.

Для математиков разных специальностей, аспирантов и студентов университетов.

$М\frac{1602010000\text{-}233}{041(01)\text{-}91}$ 5-91 ББК 22.1

Редакция литературы по математическим наукам

ISBN 5-03-002013-6 (русск.)
ISBN 0-8218-0110-4 (англ.)

ПРЕДИСЛОВИЕ

В это издание включены переводы часовых докладов, прочитанных на Международном конгрессе математиков 1986 года, состоявшемся в Беркли (штат Калифорния, США). Подобные книги дважды издавались у нас: в 1961 и 1962 гг. Государственное издательство физико-математической литературы выпустило сборники обзорных докладов Амстердамского (1954 г.) и Эдинбургского (1958 г.) конгрессов (см. стр. 7). Достаточно подробно освещался у нас Московский конгресс 1966 года (Мир, 1968). Были изданы доклады советских математиков на конгрессе 1970 года в Ницце (Наука, 1972).

Международные конгрессы математиков играют исключительную роль в математической жизни всего мира. Каждый из них имеет свои неповторимые отличительные черты, но конгресс в Беркли выделяется из всех как один из самых ярких и содержательных. Это был весьма представительный конгресс — 3711 участников из 82 стран.

Просматривая материалы конгресса, испытываешь гордость от ощущения мощного напора блистательных идей, методов и результатов, свидетельствующих о том, что мы живем в счастливое время бурного и интенсивного развития нашей науки. На конгрессе в Беркли были увенчаны филдсовскими медалями выдающиеся исследования Фальтингза, решившего проблему Морделла, а также ошеломляющие результаты Доналдсона и Фридмана по четырёхмерным многообразиям, там было доложено о решении знаменитой проблемы Бибербаха (де Бранж), были приоткрыты манящие перспективы единой физической картины мира, базирующейся на новейших достижениях математики (Виттен), — не буду продолжать, надеюсь, у читателя уже возникло неудержимое желание поскорее погрузиться в глубины новых и старых теорий, освещённых в данном издании. Воистину часовые доклады Берклеевского конгресса отражают стиль, методы и наиболее горячие точки современной математики.

Материалы данного конгресса невольно ставят вопрос о месте и значении советской математики в рамках мирового творческого процесса. Обсуждение этой темы не может быть проведено в настоящем предисловии, но я надеюсь, что советский читатель получит богатую пищу для сопоставлений и размышлений. И всё же не могу удержаться и не отметить замечательные достижения воспитанников ленинградской математической школы (два часовых доклада: А. Суслина и М. Громова, живущего ныне во Франции) и большую роль ленинградских математиков в решении проблемы Бибербаха (см. доклад де Бранжа). Часовой доклад был сделан в Беркли представителем киевской школы А. Скороходом.

Как москвич я с особенно ревнивым чувством следил за упоминаниями о воспитанниках Московского университета. Мне доставили большую радость высочайшая оценка вклада И. Р. Шафаревича, Ю. И. Манина и их последователей (Паршина, Аракелова, Зархина, Бейлинсона) в проблематику, связанную с гипотезой Морделла (см. доклады Мазура и Фальтингза), а также упоминание имени А. А. Кириллова в связи с замечательной концепцией «метода орбит» (доклад Воугана). Увы — это почти всё. Правда, на следующем конгрессе — в Киото — филдсовской медали был удостоен В. Г. Дринфельд, а премии Неванлинны — А. А. Разборов — оба выпускники мехмата МГУ.

К сожалению, приходится констатировать серьёзное отставание нашего нынешнего университетского математического образования от передовых рубежей современной науки. Сравните обзорные доклады, включённые в этот том, с докладами, прочитанными на Амстердамском и Эдинбургском конгрессах, и вы увидите, насколько далеко обогнало развитие математики в отдельных областях тот уровень подготовки, который дают сейчас наши университеты (сорок лет тому назад такого разрыва почти не было...).

В данное издание, в отличие от аналогичных, выпускавшихся у нас ранее, включены не только часовые доклады, но и выступления, в которых комментировались достижения лауреатов филдсовской и неванлинновской премий; кроме того, подробно освещены организация конгресса и церемонии его открытия и закрытия. Для большей живости сохранены все фотографии, имеющиеся в английском (двухтомном, более 1800 страниц!) оригинале.

Думаю, эта книга будет встречена математической общественностью нашей страны с большим интересом.

В заключение замечу, что издательство «Мир» предполагает издать том, содержащий переводы наиболее ярких часовых докладов, прочитанных на тех конгрессах, которые не были освещены в нашей печати (Ницца 1970 — доклады иностранных участников, Ванкувер 1974, Хельсинки 1978, Варшава 1983).

В. М. Тихомиров

ПРЕДЫДУЩИЕ КОНГРЕССЫ

1897	Цюрих
1900	Париж
1904	Гейдельберг
1908	Рим
1912	Кембридж (Великобритания)
1920	Страссбург
1924	Торонто
1928	Болонья
1932	Цюрих
1936	Осло
1950	Кеймбридж (США)
1954	Амстердам
1958	Эдинбург
1962	Стокгольм
1966	Москва
1970	Ницца
1974	Ванкувер
1978	**Хельсинки**
1982	Варшава (состоялся в 1983 г.)

ИЗДАННЫЕ РАНЕЕ НА РУССКОМ ЯЗЫКЕ СБОРНИКИ ДОКЛАДОВ НА КОНГРЕССАХ МАТЕМАТИКОВ

1. Международный математический конгресс в Амстердаме 1954 г.: Обзорные доклады. — М.: Физматгиз, 1961.
2. Международный математический конгресс в Эдинбурге 1958 г.: Обзорные доклады. — М.: Физматгиз, 1962.
3. Труды Международного конгресса математиков (Москва, 1966). — М.: Мир, 1968.
4. Международный конгресс математиков в Ницце, 1970: Доклады советских математиков. — М.: Наука, 1972.

ФИЛДСОВСКИЕ МЕДАЛИСТЫ И ЛАУРЕАТЫ ПРЕМИИ НЕВАНЛИННЫ ЗА ПРОШЛЫЕ ГОДЫ

ФИЛДСОВСКИЕ МЕДАЛИСТЫ

1936	Ларс Альфорс
	Джесси Дуглас
1950	Атле Сельберг
	Лоран Шварц
1954	Кунихико Кодаира
	Жан-Пьер Серр
1958	Клаус Рот
	Ренэ Том
1962	Ларс Хёрмандер
	Джон Милнор
1966	Майкл Атья
	Пол Дж. Коэн
	Александр Гротендик
	Стивен Смейл
1970	Алан Бейкер
	Хэисукэ Хиронака
	Сергей Новиков
	Джон Томпсон
1974	Энрико Бомбьери
	Дейвид Мамфорд
1978	Пьер Делинь
	Чарлз Фефермен
	Григорий Маргулис
	Дэниел Куиллен
1982	Ален Конн
	Уильям Тёрстон
	Шинтан Яу

ЛАУРЕАТЫ ПРЕМИИ НЕВАНЛИННЫ

1982	Роберт Тарджан

СЕКЦИИ КОНГРЕССА

01	Математическая логика и основания математики
02	Алгебра
03	Теория чисел
04	Геометрия
05	Топология
06	Алгебраическая геометрия
07	Комплексный анализ
08	Группы Ли и теория представлений
09	Вещественный и функциональный анализ
10	Теория вероятностей и математическая статистика
11	Дифференциальные уравнения с частными производными и соответствующие динамические системы
12	Обыкновенные дифференциальные уравнения и соответствующие динамические системы
13	Математическая физика
14	Численные методы и теория вычислений
15	Дискретная математика и комбинаторика
16	Математические аспекты информатики
17	Приложения математики к нефизическим наукам
18	История математики
19	Преподавание математики

ЧИСЛО УЧАСТНИКОВ ПО СТРАНАМ

Австралия	20
Австрия	8
Алжир	3
Англия	94
Аргентина	3
Бангладеш	1
Бахрейн	2
Бельгия	20
Берег Слоновой Кости	9
Болгария	4
Ботсвана	1
Бразилия	20
Венгрия	12
Венесуэла	3
Вест-Индия	1
Виргинские острова (США)	1
Вьетнам	3
Гана	1
Гватемала	2
ГДР	4
Гонконг	16
Греция	6
Дания	14
Египет	2
Израиль	34
Индия	21
Индонезия	1
Иордания	1
Ирак	4
Иран	31
Ирландия	4
Исландия	6
Испания	18
Италия	47
Камерун	3
Канада	167
Кения	2
КНР	30
Колумбия	2
Коста-Рика	1
Куба	3
Кувейт	9
Ливан	1
Люксембург	1
Малайзия	3
Мексика	10
Нигерия	8
Нидерланды	27
Никарагуа	2
Новая Зеландия	10
Норвегия	16
Оман	1
Польша	17
Португалия	7
Румыния	3
Саудовская Аравия	6
Северная Ирландия	2
Сенегал	2
Сингапур	5
СССР	57
США	2324
Таиланд	1
Тайвань	17
Тринидад	1
Турция	4
Уругвай	2
Уэллс	5
Филиппины	2
Финляндия	19
Франция	117
ФРГ	119
Чехословакия	6
Чили	3
Швейцария	33
Швеция	29
Шотландия	12
Эфиопия	1
ЮАР	11
Югославия	10
Южная Корея	6
Ямайка	1
Япония	176

Всего 3711

ОРГАНИЗАЦИЯ КОНГРЕССА. ЦЕРЕМОНИИ ОТКРЫТИЯ И ЗАКРЫТИЯ

ОБ ОРГАНИЗАЦИИ КОНГРЕССА

Национальная академия наук Соединенных Штатов Америки через Национальный математический комитет США пригласила Международный математический союз провести Международный конгресс математиков 1986 года в Соединенных Штатах Америки, в Калифорнийском университете (г. Беркли). Это приглашение было принято Международным математическим союзом на Варшавском конгрессе в августе 1983 г. Ратификация этого решения состоялась несколькими днями позже, когда Варшавский конгресс на своём заключительном заседании принял приглашение в Беркли, представленное конгрессу профессором Джеком К. Хейлом от имени американской математической общественности.

В свою очередь Академия попросила Американское математическое общество взять на себя вопросы, связанные с организацией конгресса. Общество организовало некоммерческую корпорацию ICM-86, исполнительным директором которой была назначена Джилл П. Мезиров. Корпорация ICM-86 использовала службы Отдела по организации конференций при Американском математическом обществе, возглавляемого Х. Хоуп Дейли, которая исполняла функции менеджера конгресса.

Научная программа была организована Программным комитетом, назначенным Международным математическим союзом. Его членами были Энрико Бомбьери, Леннарт Карлесон, Фридрих Э. П. Хирцебрух (председатель), Дейвид Мамфорд, Луис Ниренберг, Михаэль О. Рабин, Ю. А. Розанов, Давид П. Рюэль и И. М. Зингер. Комитет разделил программу конгресса на 19 секций и составил список предполагаемых докладчиков по каждой секции. Комитет также одобрил включение в программу конгресса заседаний, посвящённых кратким (десятиминутным) сообщениям. В октябре 1985 г. Программный комитет отобрал докладчиков, 16 пленарных и 148 секционных. Приглашения были посланы в течение следующих недель. В итоге 164 приглашённых докладчика — 14 пленарных и 132 секционных — были представлены на конгрессе.

Комитет по филдсовским медалям, состоявший из П. Делиня, Дж. Глимма, Л. Хёрмандера, К. Ито, Дж. Милнора, Ю. Мозера (председатель), С. Новикова и К. С. Сешадри, собирался для принятия решений в начале 1986 г. Комитет по присуждению премии Неванлинны, членами которого были С. Кук, Л. Д. Фаддеев (председатель) и Ш. Виноград, завершил свою работу также в начале года.

ЦЕРЕМОНИЯ ОТКРЫТИЯ

Открытие конгресса состоялось в Греческом амфитеатре Калифорнийского университета (г. Беркли) в 9 часов утра 3 августа 1986 г.

Духовой квинтет «Новый Альбион» исполнил «Маленькую увертюру для духового квинтета» Витольда Лютославского.

Профессор Юрген Мозер, президент Международного математического союза, открыл конгресс следующими словами:

«От имени Международного математического союза я рад приветствовать Международный конгресс математиков 1986 года. Одна из важнейших обязанностей ММС — содействовать проведению Международного конгресса, проводимого раз в четыре года.

Может быть, нелишне напомнить, что эти конгрессы имеют долгую историю, берущую начало в прошлом веке. Первый конгресс состоялся в Цюрихе в 1897 г., второй — в Париже в 1900 г. К сожалению, цепочка конгрессов два раза прерывалась мировыми войнами. Но начиная с конгресса 1950 года в Кеймбридже (Массачусетс), на котором был основан Международный математический союз, мы, к нашему общему удовлетворению, имеем длинную непрерывающуюся цепочку конгрессов. Сегодня здесь мы начинаем десятый конгресс этой серии, и я уверен, все мы едины в горячей надежде, что эта серия продожится, не прерываясь, и в будущем столетии.

На первом конгрессе было 216 участников, ныне более 3000 математиков присутствуют на конгрессе. Но несмотря на такой стремительный рост, конгрессы и поныне руководствуются тем же принципом — способствовать установлению личных контактов между математиками разных стран и представлять обзор современного состояния науки.

Во времена возрастающей специализации и стремительно протекающего процесса дробления математики на множество обособленных областей такие конгрессы играют чрезвычайно важную роль, объединяя математиков различных направлений и интересов. Опасность разделения нашей науки на множество обособленных ветвей слишком очевидна. Наша общая надежда состоит в том, что этот конгресс будет противостоять этим сепаратистским тенденциям и даст нам широкую картину единой математики.

Я счастлив приветствовать здесь в Беркли математиков почти из семидесяти стран. Я надеюсь, что эта неделя предоставит массу возможностей для плодотворного обмена идеями, равно как и для установления прочных научных и личных контактов.

Организация этого конгресса осуществлялась под компетентным руководством Организационного комитета, которому оказывало помощь Американское математическое общество. Возглавлял этот комитет профессор Эндру Глисон. Я думаю, что мы единодушно изберём сейчас здесь профессора Глисона президентом Конгресса-86.»

Профессор Глисон, избранный без голосования, на основании единодушного одобрения, произнёс следующую речь:

«Это поистине огромная честь для меня — руководить этим, двадцатым по счёту Международным конгрессом математиков. Для меня также огромное удовольствие приветствовать вас всех в городе Беркли — городе, знаменитом своим Калифорнийским университетом, городе, где погода почти всегда так же хороша, как в это утро. От имени Национальной академии наук и всей математической общественности Америки я особенно горячо приветствую тех, кто приехал сюда из других стран. Надеюсь, ваше пребывание в Соединенных Штатах будет приятным, вы узнаете кое-что новое в математике, приобретёте новых друзей и насладитесь некоторыми из чудесных видов нашей обширной страны.

Как вы знаете, научная программа конгресса разбита на девятнадцать секций. Программный комитет под умелым руководством профессора Фрица Хирцебруха и с помощью многочисленных подкомитетов подготовил замечательную научную программу с 16 пленарными докладчиками и более чем 140 секционными. Кроме того, представлено более 400 статей, и уже открыто несколько специальных семинаров. К сожалению, часть приглашённых докладчиков не смогли приехать, и мы были вынуждены в связи с этим вносить некоторые изменения в расписание в последний момент. Мы будем держать вас в курсе всех изменений через ежедневный информационный бюллетень.

Математика всегда была полезна. Многие из древнейших письменных документов человеческой цивилизации представляют собой «бухгалтерские» документы, да и на сегодня счётное дело всё ещё остаётся самой широкой областью приложений математики. Но мы живём во время, когда открываются всё новые и новые области приложения математики, и этот процесс всё ускоряется. Новые математические вопросы ставят учёные, инженеры, менеджеры — и часто эти вопросы совсем иного сорта, чем те, которые принято было рассматривать раньше. На эти вопросы даются новые математические ответы, часто основанные на идеях, прежде считавшихся совершенно абстрактными и вполне бесполезными. Как математики мы можем испытывать законную гордость от сознания того, что идеи и понятия, над выработкой которых мы упорно трудились, помогают понять реальный мир, точно так же как они помогли нам понять мир наших абстракций. Я полагаю, в этом заключён урок, и он состоит в том, что, когда мы войдём в новую эру, где будут доминировать компьютеры, нам нельзя попасться в ловушку утилитаризма, а следует всегда помнить, что величайшие достижения в математике были связаны с попытками понять фундаментальные структуры, лежащие в основе вещей, а не с решением чисто утилитарных проблем.»

Затем он предоставил слово профессору Мэри Эллен Рудин, председателю Национального математического комитета Соединенных Штатов, которая сказала:

«Я также хотела бы приветствовать всех вас от лица Соединенных Штатов. Ровно пятьдесят лет тому назад на конгрессе, состоявшемся в Норвегии, в Осло, были вручены награды первым филдсовским медалистам. Двух медалей 1936 года были удостоены Джесси Дуглас, которого уже нет в живых, и Ларс Альфорс, который был тогда молодым человеком, не достигшим ещё и тридцати лет. В ознаменование пятидесятой годовщины основания филдсовской премии и учитывая огромный вклад в математику профессора Альфорса на протяжении всех этих пятидесяти лет, я предлагаю избрать профессора Ларса Альфорса Почетным президентом конгресса.»

Это предложение было встречено общим одобрением. Профессор Альфорс, сопровождаемый бурными аплодисментами, взошёл на кафедру и сказал:

«С чистой совестью принимаю я эту великую честь, ибо считаю себя связующим звеном между этим Международным конгрессом и тем, который был в 1936 г., пятьдесят лет назад, когда впервые были присуждены филдсовские медали. И я воспринимаю как свой непременный долг иметь удовольствие вручать на этом конгрессе филдсовские медали и премию имени Неванлинны.

В те далёкие времена обстоятельства были совершенно иными. Идея медалей была одобрена в Цюрихе в 1932 г., но об этом решении не было широко известно. Я не знал, приехав в Осло, что присуждение стало реальностью, а если бы я и знал, то не посчитал бы себя достойным кандидатом. Так или иначе, мне ничего не было официально известно до тех пор, пока я не вошёл в аудиторию, где должна была состояться церемония открытия, и лишь когда мне указали место где-то в первых рядах, я начал волноваться. Впрочем, кое-какие основания для волнения у меня уже были: кто-то по ошибке поздравил меня на день раньше. Но вплоть до самого награждения это был секрет, по крайней мере официально, и я старался об этом не думать. Ещё не было ни традиций, которым нужно было следовать, ни протокола, который требовалось соблюдать. Как уже упоминалось здесь, вручались две медали, одна мне, другая Джесси Дугласу, который тогда работал в Массачусетсском технологическом институте; я же в то время был приглашённым лектором в Гарварде. Иначе говоря, случилось так, что обе медали попали в Кеймбридж, Массачусетс. К сожалению, Дуглас не смог получить свою медаль лично; судя по протокольным записям конгресса, он чувствовал себя слишком усталым. Я не знаю, может он действительно чересчур устал после длительного и напряжённого переезда. Не думаю, чтобы такое могло произойти сейчас. Его медаль принял Норберт Винер как представитель МТИ.

Две традиции восходят к самому началу. Во-первых, комитет по присуждению медалей должен состоять из наиболее сильных современных математиков. В 1936 г. членами комитета были Г. Биркгоф, Каратеодори, Эли Картан, Севери и Такаги. Воистину я могу назвать это собранием олимпийцев. И я считаю, что эта традиция соблюдалась и на последующих конгрессах. Другая традиция заключается в том, что достижения победителей должны быть прокомментированы видными учёными в данной области. В 1936 г. обе медали комментировал Каратеодори.

Как уже упоминалось, в следующий раз конгресс собрался лишь в 1950 г., четырнадцать лет спустя. Во время конгресса 1950 года, проводившегося в Гарварде, медали были вручены Атле Сельбергу и Лорану Шварцу, которых знают и любят все математики. С тех пор филдсовские медали становились всё более и более престижными, и держу пари, что многие мечтают получить их. Ну а правда ли то, что существование медалей внесло свой вклад в необычайный рост математики как в количественном, так и в качественном отношении за последние полвека, пусть об этом судят другие.

Сегодня спокойно можно поздравлять победителей заранее, и я использую эту возможность, чтобы выразить своё искреннее восхищение их достижениями. Я разделяю их чувство гордости и удовлетворения достигнутым и уверен в их дальнейших успехах. Я также разделяю разочарование тех, кто может чувствовать себя обойдённым. Я желаю им большей удачи при следующем присуждении, а если этого не произойдёт, желаю, чтобы будущие поколения признали справедливость их притязаний. Спасибо.»

Затем духовой квинтет сыграл отрывки из «Духового бродячего зверинца» Джона Читхама.

[Далее следуют выступления двух официальных лиц: профессора Калвина К. Мура, помощника вице-президента Калифорнийского университета по академическим делам, и Ричарда Джонсона, исполняющего обязанности советника президента США по науке. В конце своей речи последний зачитал следующее послание президента Рейгана Международному математическому конгрессу (факсимиле этого послания приведено справа).— *Изд. ред.*]

БЕЛЫЙ ДОМ
ВАШИНГТОН

3 июля 1986 г.

Сердечно приветствую тысячи математиков со всего света — участников Международного математического конгресса, собирающегося раз в четыре года. Все мы рады и, более того, горды тем, что Соединенные Штаты были избраны местом проведения этого престижного собрания.

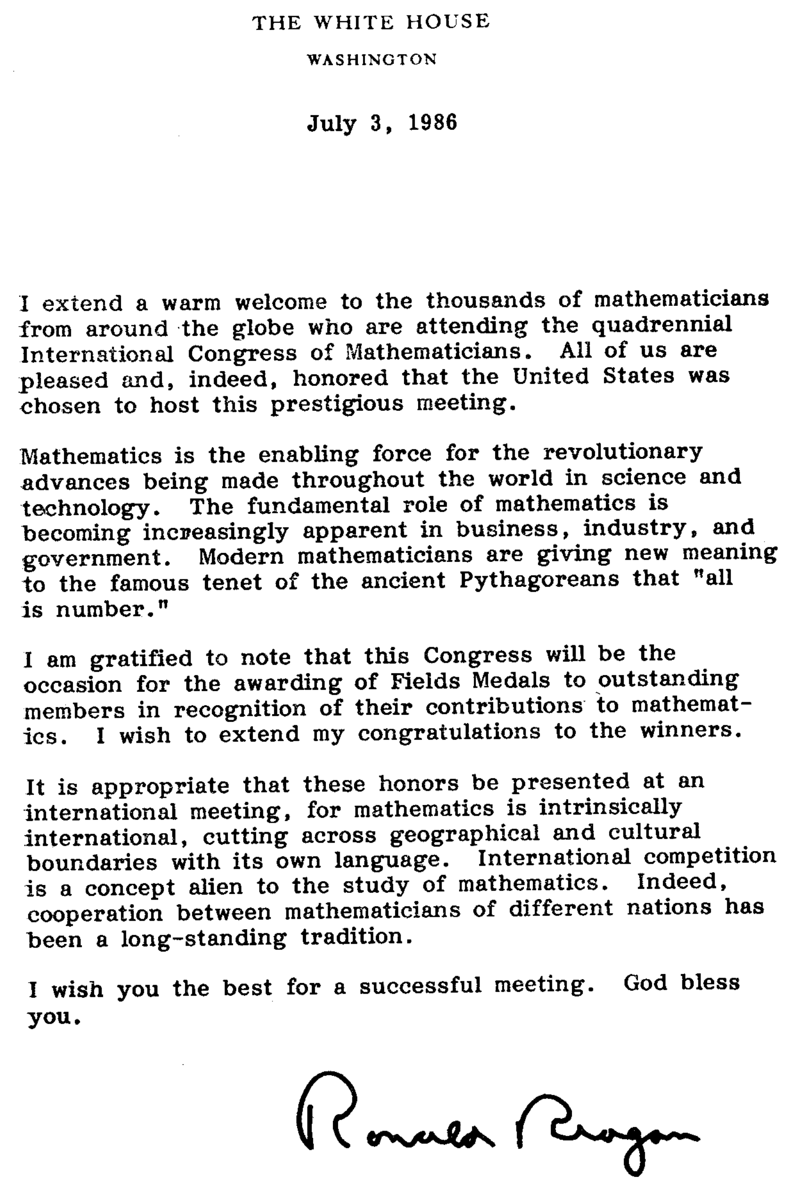
Математика является вдохновляющей силой, приведшей во всём мире к революционным сдвигам в науке и технике. Основополагающая роль математики всё очевиднее возрастает в деловой деятельности, промышленности и управлении. Современные математики придают новый смысл известному принципу древних пифагорейцев, соласно которому «всё есть число».

Я рад, что во время конгресса будут вручены филдсовские медали выдающимся ученым в знак их достижений в математике. Хочу выразить лауреатам свои поздравления.

Представляется естественным, что эти почётные награды присуждаются международным форумом, ибо математика является воистину интернациональной наукой, преодолевающей — благодаря своему особому языку — любые географические и культурные барьеры. Международное противостояние чуждо математическому творчеству. Сотрудничество учёных разных стран действительно имеет давно установившиеся традиции.

Желаю вам всего наилучшего в успешном проведении конгресса. Да поможет вам Бог.

Рональд Рейган



THE WHITE HOUSE
WASHINGTON

July 3, 1986

I extend a warm welcome to the thousands of mathematicians from around the globe who are attending the quadrennial International Congress of Mathematicians. All of us are pleased and, indeed, honored that the United States was chosen to host this prestigious meeting.

Mathematics is the enabling force for the revolutionary advances being made throughout the world in science and technology. The fundamental role of mathematics is becoming increasingly apparent in business, industry, and government. Modern mathematicians are giving new meaning to the famous tenet of the ancient Pythagoreans that "all is number."

I am gratified to note that this Congress will be the occasion for the awarding of Fields Medals to outstanding members in recognition of their contributions to mathematics. I wish to extend my congratulations to the winners.

It is appropriate that these honors be presented at an international meeting, for mathematics is intrinsically international, cutting across geographical and cultural boundaries with its own language. International competition is a concept alien to the study of mathematics. Indeed, cooperation between mathematicians of different nations has been a long-standing tradition.

I wish you the best for a successful meeting. God bless you.

Ronald Reagan

Факсимиле послания президента Рональда Рейгана Международному математическому конгрессу.

Профессор Глисон поблагодарил д-ра Джонсона и попросил его от имени собравшихся выразить признательность президенту за послание. Затем духовой квинтет сыграл пьесу «Фанфара» Дейвида Амрама. Профессор Глисон объявил, что профессор Альфорс будет вручать медали, и предоставил слово академику Людвигу Фаддееву, председателю Комитета по присуждению премии имени Неванлинны, который сказал:

«Неванлинновская премия имеет гораздо более короткую историю, чем филдсовские медали, о первом вручении которых мы только что слышали от профессора Альфорса. Она была учреждена в значительной мере в результате усилий нашего прошлого президента, Леннарта Карлесона, с целью подчеркнуть важность прикладных аспектов математики. Хельсинкский университет щедро выделил необходимые фонды, поэтому вполне уместно, что эта премия носит имя Неванлинны. Более точно, премия вручается молодым математикам за выдающиеся достижения в области математических аспектов информатики. «Молодой» означает то же самое, что и в случае филдсовской медали. Сегодня премия вручается во второй раз. Первое вручение премии состоялось на Генеральной ассамблее Варшавского конгресса. В состав Комитета по неванлинновской премии этого года входили профессор Кук (Торонтский университет), профессор Виноград (IBM) и я. Разрешите мне огласить наше решение. Премия имени Неванлинны за 1966 год присуждается Лесли Вальянту, Гарвардский университет, США. Я хотел бы попросить профессора Альфорса вручить награду.»

Профессор Вальянт поднялся на сцену и принял награду из рук профессора Альфорса.

Затем профессор Глисон предоставил слово профессору Мозеру как председателю Комитета по присуждению филдсовских медалей, и тот сообщил следующее:

«В начале тридцатых годов организаторы Торонтского конгресса решили вручать на каждом конгрессе две золотые медали за выдающиеся достижения в математике. Движущей силой, стоявшей за этим решением, был Джон Чарлз Филдс, который выразил желание, чтобы эти награды служили, я цитирую: «знаком признания уже достигнутых достижений, а также поощрением для дальнейших исследований — в отношении получателей и стимулом для приложения бо́льших усилий — в отношении остальных.» Интересно заметить, что Филдс настаивал на том, чтобы при внешнем оформлении медалей был насколько только возможно подчёркнут международный и безличностный характер премии. И действительно, имя самого Филдса даже не фигурирует на медалях. Как мы уже сегодня слышали, первые медали были вручены на конгрессе 1936 года в Осло, четыре года спустя после смерти Филдса. Просматривая впечатляющий список филдсовских медалистов за прошедшие полвека, мы убеждаемся, что замысел Филдса был реализован самым успешным образом.

Начиная со Стокгольмского конгресса 1962 года организация присуждения филдсовских медалей была поручена Международному математическому союзу. Комитет по филдсовским медалистам на этом конгрессе состоял из П. Делиня, Дж. Глимма, Л. Хёрмандера, К. Ито, Дж. Милнора, С. Новикова, К. С. Сешадри и меня в качестве председателя. Перед комитетом стояла трудная задача сделать выбор из впечатляющего списка блестящих математиков. Не надо и говорить о том, что эта задача не имела единственного решения и что каждому из рассматривавшихся нами превосходных математиков смело можно было вручать премию. От имени Международного математического союза я хочу поблагодарить членов комитета за усилия, с которыми они старались принять решение в духе ответственности и сотрудничества.

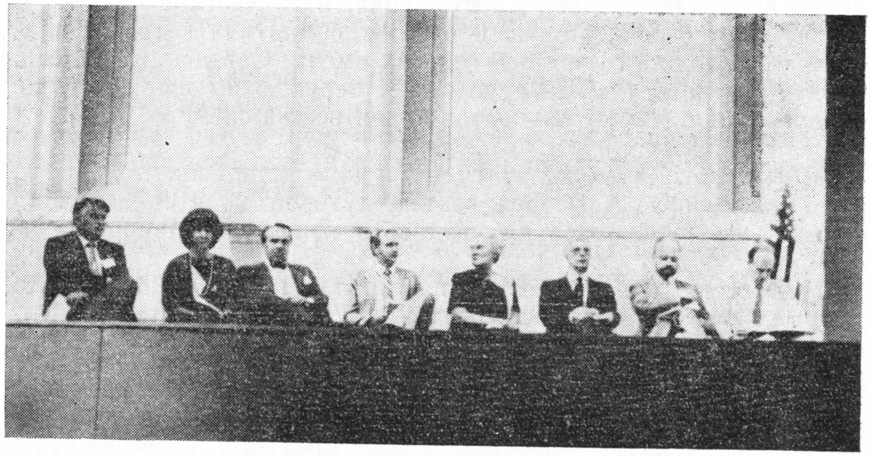
В соответствии с традицией в список кандидатов включались лишь математики не старше сорока лет. После долгого обдумывания и обширных консультаций были отобраны три молодых математика. На всех нас произвели глубокое впечатление их исключительные достижения, и я счастлив сообщить, что комитет был единодушен в своих симпатиях к трём филдсовским медалистам за 1986 год. Их имена: Саймон Доналдсон, Герд Фальтингз, Майкл Фридман. Я передаю им наши горячие поздравления. Прошу лауреатов подняться на сцену.»

Лауреаты поднялись на сцену и получили свои медали из рук профессора Альфорса, затем духовой квинтет исполнил отрывки из «Трёх пьес для духового квинтета» Минору Фудзисиро.

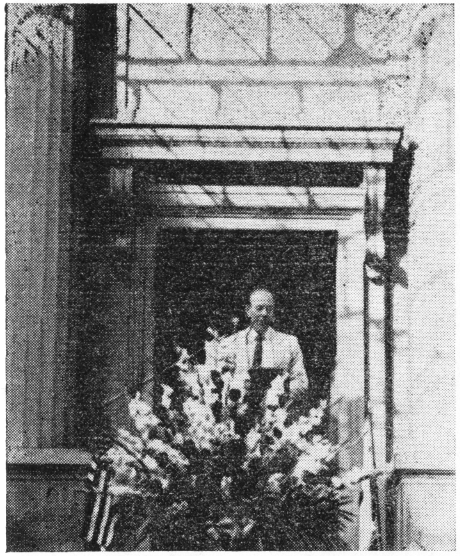
Профессор Глисон поздравил медалистов и сообщил, что доклады о достижениях лауреатов состоятся в аудитории Целлербаха.

Затем был объявлен перерыв до 11.30.

Церемония открытия. Музыканты из духового оркестра.



Церемония открытия. Людвиг Фаддеев, Джилл Мезиров, Кэлвин Мур, Ричард Джонсон, Мэри Эллен Рудин, Ларс Альфорс, Юрген Мозер, Эндру Глисон.

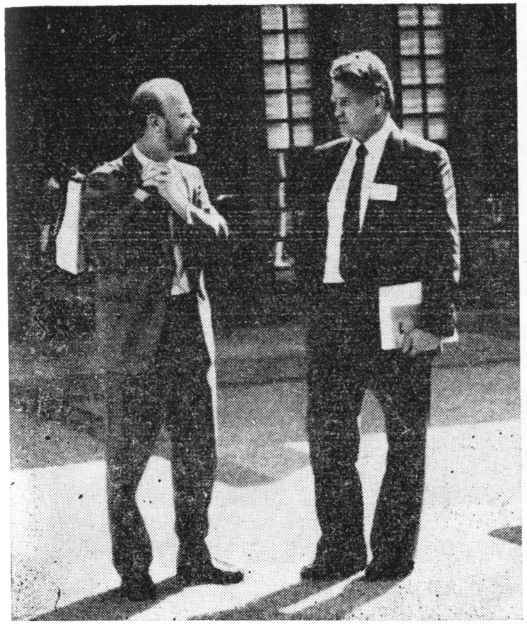


Церемония открытия. Эндру Глисон, президент конгресса.

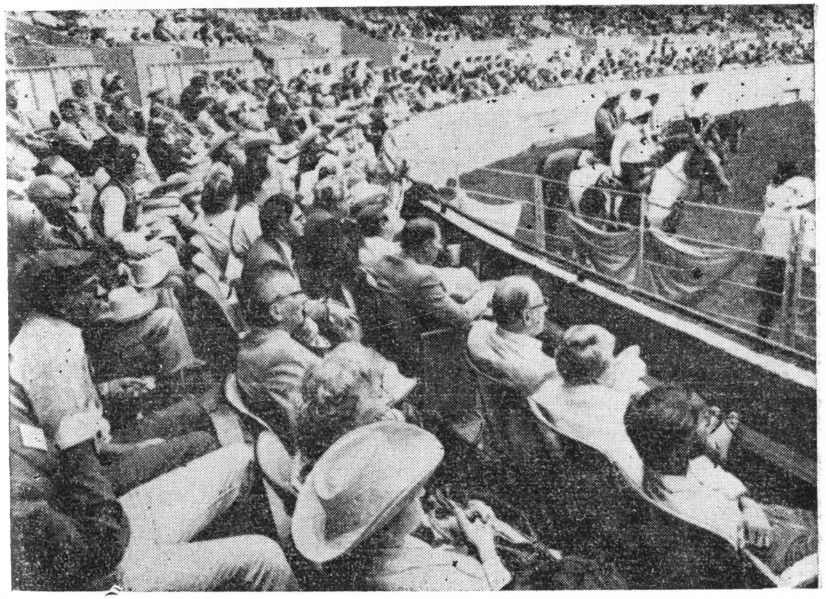
Фриц Хирцебрух, Юзеф Алави и Майкл Атья.

Ларс Альфорс, филдсовский медалист 1936 года. Почётный президент конгресса.

Олли Лехто, Юрген Мозер и Джилл Мезироз.

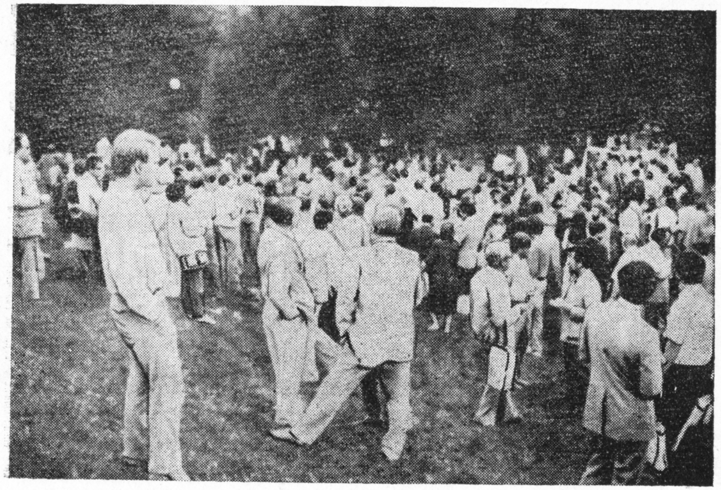


Юрген Мозер и Людвиг Фаддеев — президент ММС и его преемник.

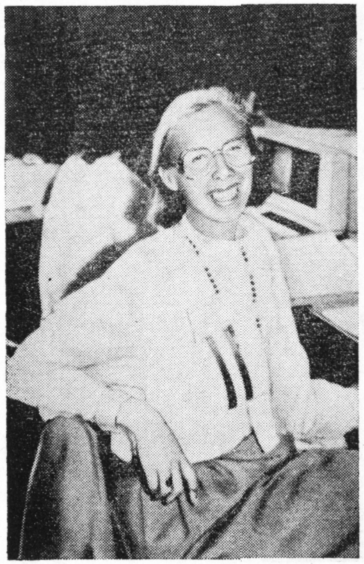


Родео.

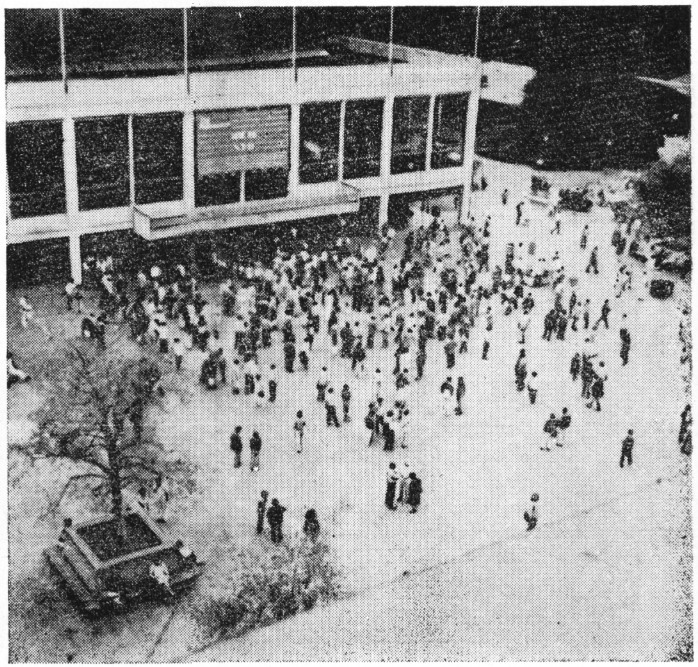
Профессор Нагата приглашает конгресс в Киото.



Приём у ректора университета.

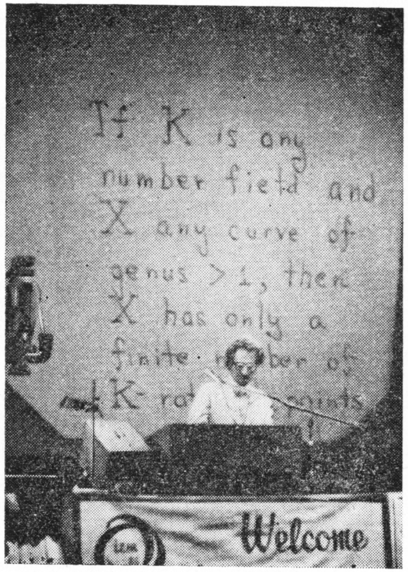


Хоуп Дейли, менеджер конгресса.

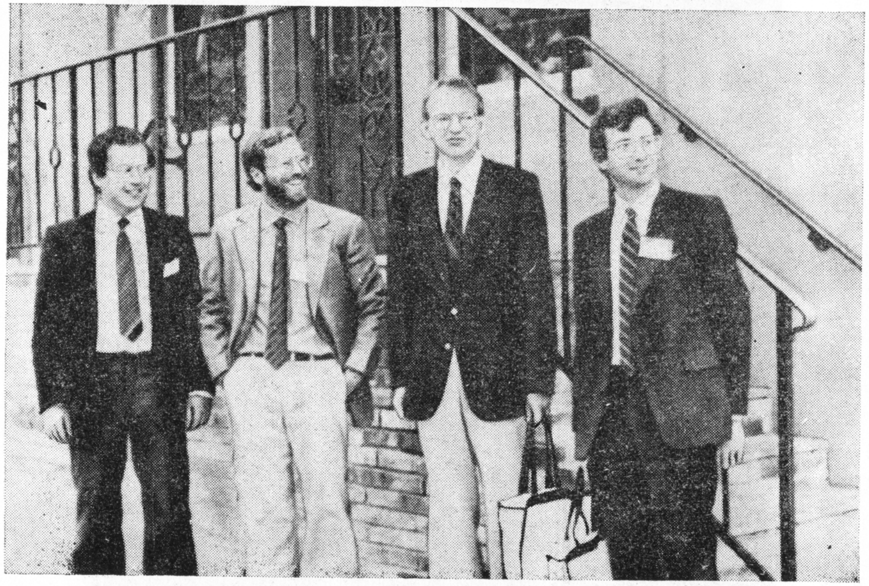


Аудитория Целлербаха.

Джон Милнор говорит о достижениях Фридмана.



Барри Мазур рассказывает о достижениях Фальтингза.



Лесли Вальянт, Майкл Фридман, Герд Фальтингз и Саймон Доналдсон.

ЦЕРЕМОНИЯ ЗАКРЫТИЯ

Заключительное заседание конгресса проходило в Аудитории Целлербаха в 11 часов утра 9 августа.

Профессор Глисон открыл заседание и предоставил слово профессору Мозеру для сообщения о Международном математическом союзе. Профессор Мозер сказал:

«Вот и подошла к концу богатая и интересная программа конгресса. И по сложившейся традиции настало время президенту Международного математического союза доложить конгрессу о работе союза и о решениях, принятых на его Генеральной ассамблее.

Но перед этим разрешите мне приветствовать присутствующего здесь профессора Маршалла Стоуна, одного из предыдущих президентов Международного математического союза. Это большая честь для меня. Именно профессор Стоун сыграл решающую роль в восстановлении ММС в 1950 г. после прекращения его деятельности в 1932 г. Профессор Стоун, мы рады видеть Вас среди нас на этом конгрессе.

С 1962 г. ММС отвечает за подготовку научной программы конгресса. Этим занимается Программный комитет, назначаемый частично ММС и частично страной-хозяйкой. В состав Программного комитета этого конгресса вошли: Фриц Хирцебрух (председатель), Энрико Бомбьери, Леннарт Карлесон, Дейвид Мамфорд, Луис Ниренберг, Михаэль Рабин, Юрий Розанов, Давид Рюэль и Изидор Зингер. Я хотел бы выразить нашу благодарность профессору Хирцебруху и его Комитету за то, что они представили нам отличную разнообразную программу.

Однако, к нашему великому разочарованию, в Беркли не приехали многие приглашённые докладчики из Советского Союза; фактически не присутствовала почти половина советских докладчиков. Это — серьёзная потеря для всех нас и серьёзная помеха для осуществления целей конгресса. Для любого конгресса наиболее важно, чтобы все приглашенные докладчики присутствовали на нём, могли лично прочесть свои лекции и принять участие в обмене идеями.

Нам известно, что наши советские коллеги приложили очень много сил для разрешения этой проблемы, и мы ценим их усилия. Кроме того, от большинства отсутствующих докладчиков поступили тексты их сообщений, так что эти тексты могли быть зачитаны на конгрессе другими математиками.

Независимо от всяких обстоятельств, невозможность для докладчиков из любой страны присутствовать на конгрессе всегда разочаровывает, так что позвольте мне выразить нашу общую надежду, что на конгрессе 1990 года будут присутствовать все приглашённые докладчики из всех стран.

Как вы, может быть, знаете, ММС — член Международного совета научных союзов (МСНС) и как таковой действует в соответствии с провозглашенным МСНС принципом свободного перемещения учёных. Я рад сообщить вам, что, насколько мне известно, страна-хозяйка выдала все визы, о которых мы просили. В некоторых трудных случаях помощь МСНС была абсолютно необходима. Это вновь демонстрирует важность указанного принципа МСНС для нашего Союза. Добавлю в этой связи, что две недели назад на Генеральной ассамблее ММС была принята резолюция, подтверждающая статью устава МСНС о недопустимости расовой дискриминации.

Позвольте мне теперь дать краткий отчёт о Генеральной ассамблее ММС. Генеральная ассамблея собиралась 31 июля и 1 августа в Окленде, Калифорния, и я хотел бы проинформировать вас о главных её решениях.

Прежде всего, в ММС были приняты два новых члена — Берег Слоновой Кости и Китайская Народная Республика. Стоит напомнить, что вопрос о членстве Китая имеет давнюю и сложную историю. В связи с этим я хотел бы особенно поблагодарить секретаря ММС, профессора Олли Лехто за его непрестанные усилия, направленные на разрешение этой проблемы. Мы хотим также выразить нашу признательность администрации Тайваня за сотрудничество в этом вопросе.

Общее число членов ММС составляет теперь пятьдесят три. Далее, на Генеральной ассамблее был избран новый состав комитетов, и я хотел бы сообщить вам результаты выборов. Начну с Исполнительного комитета ММС на 1987—1990 гг.: президент — Людвиг Фаддеев; вице-президенты — Уолтер Фейт и Ларс Хёрмандер; секретарь — Олли Лехто; члены — Джон Коутс, Хикосабуро Комацу, Ласло Ловас, Жакоб Палис-мл. и К. С. Сешари. В качестве почётного члена в состав комитета включён также бывший президент.

Перехожу к подкомиссиям. В ММС существуют две подкомиссии — Международная комиссия по математическому обучению (МКМО) и Комиссия по развитию и обмену (КРО). В новый состав Исполнительного комитета МКМО входят: президент — Жан-Пьер Кахан; два вице-президента — Пен-Е Ли и Эмилио Луис Рьера; секретарь — А. Дж. Хоусон; три члена — Хироси Фудзита, Джемери Килпатрик и Могенс Нисс. Новый состав КРО: председатель — М. С. Нарасимхан; члены — Жан-Пьер Бургиньон, Филлип Гриффитс, М. Иманалиев, Адереми О. Куку, Ле Донг Транг, Синго Мураками и Джованни Видоссич. В качестве почётных членов в Исполнительный комитет МКМО и в КРО включены президент и секретарь ММС.

Наконец, Комитет по выбору места проведения представил Генеральной ассамблее свои предложения о месте и времени проведения следующего конгресса.

Прежде чем предоставить слово следующему оратору, позвольте поделиться личными замечаниями. Работая эти четыре года в качестве президента ММС, я имел счастье получать советы и поддержку от многих коллег из всех частей света. Я хотел бы поблагодарить их всех, особенно членов Исполнительного комитета и Комитета по филдсовским медалям. Наибольшую поддержку, однако, ощущал от Олли Лехто, который вместе со своей секретаршей, г-жой Тууликки Мякеляйнен, был готов помочь мне в любое время и в любой ситуации. Без его ободрения и полезных советов я не смог бы выполнить свою задачу. Им обоим — моя горячая благодарность.»

Затем профессор Мозер передал слово профессору Нагате, пригласившему слушателей на очередной Международный конгресс. Профессор Нагата сказал:

«От имени Японского математического комитета я имею честь пригласить вас в Киото на следующий Международный математический конгресс. Город Киото был столицей Японии в течение примерно тысячи лет и может показать вам некоторые стороны древней японской культуры.

Мы полностью отдаём себе отчёт в том, как сложно организовать такое грандиозное мероприятие. Однако с помощью Международного математического союза и опираясь на поддержку мирового математического сообщества, мы, я уверен, сможем преодолеть все трудности.

В надежде на то, что вы примете наше приглашение, жду вас всех на следующем Международном математическом конгрессе в Киото в августе 1990 года.»

Затем профессор Глисон предоставил слово профессору Адамсу, сказавшему:

«Дамы и господа! Настало время гостям поблагодарить своих хозяев.

Мы, математики всего мира, собрались в Беркли из многих стран. Мы провели в высшей степени успешную работу и хотим теперь выразить свою признательность. Мы имели возможность встречаться и общаться друг с другом; мы слышали множество замечательных докладов и благодарили докладчиков своими аплодисментами. Чтобы организовать всё это, был затрачен большой труд множества людей.

Как объяснил профессор Мозер, выбор докладчиков был задачей входящего в ММС комитета под председательством профессора Хирцебруха. Мы знаем, что в этом им охотно помогали математики всех стран; мы снова благодарим их всех за сделанный от нашего имени квалифицированный выбор.

Почти вся работа, однако, лежала на хозяевах. Для этой цели Американское математическое общество создало специальный комитет под председательством профессора Глисона, и мы выражаем признательность всем принимавшим участие в его работе. Мы также благодарим штат сотрудников Американского математического общества, которых вы постоянно могли видеть в Бюро конгресса. Эти люди заказывали номера в гостиницах и билеты на самолеты, улаживали всё, что нужно было уладить, брали на себя ответственность за продажу сувениров (кстати, осталось ещё несколько непроданных ковбойских шляп...). Какой похвалы и благодарности они заслуживают, лучше всего оценят те, кто сам раньше пробовал себя в таком роде деятельности, например професора Чеслав Олех, Олли Лехто и Морис Сион.

Я прошу всех математиков из других стран присоединиться ко мне в выражении признательности нашим американским хозяевам, и все вместе мы благодарим тех, кто трудился для конгресса.»

Поблагодарив профессора Адамса, профессор Глисон сказал:

«Успешная работа научного конгресса определяется двумя главными компонентами — научной программой и материальным оформлением. Программный комитет подготовил превосходный список докладчиков, действительно замечательный список, и я хочу поблагодарить за это всех челнов комитета. Я выражаю признательность докладчикам, многие из которых приехали издалека, чтобы поделиться с нами своими идеями. Доклады были прекрасны. Научная программа конгресса была также обогащена несколькими специальными семинарами и предоставленными в распоряжение участников документами. Организаторы и докладчики заслуживают нашей благодарности. Успех научного форума зависит и от участников. На нашем конгрессе присутствовало свыше 3500 математиков. Всего было зарегистрировано 3970 человек. Спасибо вам всем за участие. От имени американской математической общественности я хочу особенно поблагодарить всех приехавших из других стран. Вы оказали нам честь своим присутствием, и мы желаем вам дальнейших успехов в решении проблем, поставленных в докладах.

Многие представители американской математической общественности напряжённо трудились последние два года для этого конгресса, и я бы не смог даже перечислить их всех. Я хотел бы выразить им лично свою признательность и присоединиться к словам благодарности, высказанным профессором Адамсом, который, по моему мнению, очень верно выразил чувства участников конгресса. Заслуживают быть отмеченными также Руководящий комитет и, в частности, Кеннет Росс и Хьюго Росси, занимавшиеся составлением расписания; Совет директоров под председательством Роналда Грейема; Комитет специальных фондов под председательством Ричарда Андерсона; Комитет общественной информации под председательством Юзефа Алави и Комитет по связи с местными организациями под председательством Джона Эддисона. Все эти комитеты сделали больше, чем могли, как мы отлично видели.

Из математиков, работавших для конгресса, один заслуживает особого упоминания. Это — Джилл Мезиров, Исполнительный директор корпорации ICM-86. С самого начала конгресса она ежедневно занималась текущими вопросами подготовки заседаний, и мы выражаем ей глубокую признательность.

Я хотел бы упомянуть здесь выдающийся вклад в подготовку этого конгресса организаторов Варшавского конгресса. В 1983 г. они сердечно приняли менеджера нашего конгресса и щедро поделились с ней своим опытом. Их советы были бесценны.

Наконец, я хочу поблагодарть только что упомянутого менеджера нашего конгресса, Хоуп Дейли. Последние четыре года она была нашим фельдмаршалом, руководила всеми ответственными операциями и никогда не уклонялась от менее важных дел, если не хватало рук.»

Затем профессор Глисон вручил г-же Дейли серебряную подвеску и букет. После бурных аплодисментов он продолжал:

«Этот конгресс — часть давней традиции интернационализма. Со времён Архимеда математики переписывались друг с другом и лично преодолевали огромные расстояния, чтобы учиться, учить и обсуждать. По мере того как издавать книги и путешествовать становилось всё проще, эта традиция всё укреплялась. Теперь ежегодно публикуются сотни математических книг и журналов. Они свободно проходят через границы и распространяют новые математические идеи по всему миру. Математики всё чаще переезжают из одного университета или института в другой. Давайте же поддерживать и расширять наши великие традиции свободы исследования, перемещения и обсуждения; приложим все усилия, чтобы конгресс математиков Киото был в ещё большей степени международным.»

На этом профессор Глисон объявил конгресс закрытым.

ВЫСТУПЛЕНИЯ О ДОСТИЖЕНИЯХ ЛАУРЕАТОВ КОНГРЕССА

О ДОСТИЖЕНИЯХ САЙМОНА ДОНАЛДСОНА

Майкл Атья[1]

В 1982 г., на втором году аспирантуры, Саймон Доналдсон получил результат [1], который потряс весь математический мир. С учётом важных достижений Майкла Фридмана (о них расскажет Джон Милнор), из результата Доналдсона вытекало, что существуют весьма «экзотические» четырёхмерные пространства, а именно четырёхмерные дифференцируемые многообразия, которые топологически, но не дифференциально эквивалентны стандартному четырёхмерному эвклидову пространству $\mathbb{R}^4$. И самым удивительным во всём этом было то, что $n=4$ — единственное натуральное число, для которого подобные экзотические пространства вообще существуют. Эти экзотические четырёхмерные пространства (в отличие от $\mathbb{R}^4$) обладают таким, к примеру, поразительным свойством: они содержат компактные подмножества, которые не могут быть заключены внутрь никакого диффеоморфа трёхмерной сферы!

Для того чтобы заключить картину в исторические рамки, разрешите мне напомнить, что Милнором в 1958 г. были открыты экзотические семимерные сферы и что в шестидесятых годах структура дифференцируемых многообразий в размерностях $n \geqslant 5$ активно исследовалась Милнором, Смейлом (оба — филдсовские медалисты) и другими, и это привело к весьма содержательной теории таких многообразий. Размерность два (римановы поверхности) — это классический случай, изученный ранее, и, таким образом, оставалось разобраться лишь с размерностями 3 и 4. На предыдущем конгрессе в Варшаве Тёрстон получил филдсовскую медаль за свои замечательные результаты по трёхмерным многообразиям. А ныне на этом конгрессе пришёл черёд и четырёхмерным.

---

[1] Michael Atiyah, On the work of Simon Donaldson, ICM86, pp. 3—6. [Здесь и далее мы ссылаемся на оригинальное издание Трудов конгресса (см. контртитул) как на ICM86. — *Изд. ред.*]

Я хотел бы подчеркнуть, что случаи $n=3$, $n=4$ и $n \geqslant 5$ — совершенно разные, причём малые размерности доставляют гораздо больше хлопот.

Хотя я и выделил экзотические четырёхмерные пространства как самое эффектное следствие результатов Фридмана — Доналдсона, но необходимо подчеркнуть, что это лишь побочный продукт их деятельности. На самом деле их исследования посвящены *замкнутым* четырёхмерным многообразиям. Замкнутые четырёхмерные многообразия обладают стандартными топологическими инвариантами. В частности, для *ориентируемого* многообразия можно задать некоторую симметричную целочисленную матрицу с детерминантом ± 1, определяемую свойствами пересечений 2-циклов (и зависящую от выбора базисов). Фридман показал, что для топологических четырёхмерных многообразий при этом получаются все такие матрицы. Из результата Доналдсона следует, что среди положительно-определённых матриц лишь матрицы, эквивалентные единичной, могут встретиться для четырёхмерных дифференцируемых многообразий[1]. Это — весьма существенное ограничение. Оно показывает, что дифференцируемая и топологическая структуры принципиально отличаются друг от друга.

Достижения Доналдсона поражают ещё и тем, что его методы были совершенно новыми, навеянными теоретической физикой, а именно уравнениями Янга — Миллса. Уравнения Янга — Миллса представляют собой существенно нелинейное обобщение максвелловых уравнений электродинамики; это вариационные уравнения, связанные с некоторым естественным геометрическим функционалом. Дифференциальные геометры изучают связности и кривизну в расслоенных пространствах, а функционал Янга — Миллса — это не что иное, как L_2-норма кривизны.

Если структурной группой расслоения является окружность, мы возвращаемся к линейной максвелловой теории. Неабелевым же группам Ли отвечает нелинейная теория. Доналдсон использует лишь простейшую неабелеву группу, SU(2), хотя в принципе могут и должны быть изучены и другие группы.

Физики интересуются подобными уравнениями в пространстве-времени Минковского (где эти уравнения гиперболические), а также в эвклидовом четырёхмерном пространстве, где уравнения эллиптические. В эвклидовом случае представляют особый интерес решения, дающие абсолютный минимум (при заданных граничных условиях на бесконечности). Их называют *инстантонами.*

---

[1] В действительности в [1] Доналдсон доказал это лишь для односвязных многообразий, но недавно ему удалось снять условие односвязности.

Ряд математиков (в том числе и я) занимались изучением инстантонов и испытывали удовлетворение от сознания того, что тем самым они содействуют развитию физики. Доналдсон же возымел дерзкий замысел — пойти в обратном направлении и применить инстантоны как новый геометрический инструмент к теории общих четырёхмерных многообразий. И добился на этом пути блестящего успеха — ему удалось добраться до совершенно новых явлений и тем продемонстрировать, что уравнения Янга — Миллса прекрасно приспособлены для исследования всей этой новой области.

Разумеется, применение дифференциальных уравнений в геометрии — не новость; классические примеры — изучение геодезических, теория минимальных поверхностей. Однако в этих случаях решение дифференциальных уравнений (скажем, для минимальных поверхностей) представляет собой геометрический объект. Доналдсон же использует инстантоны совершенно по-другому. Здесь следует сказать, что инстантоны как решения задачи минимизации не обладают свойством единственности; они, вообще говоря, зависят от конечного числа непрерывно изменяющихся параметров; эти инстантонные параметры образуют нелинейное пространство. Его-то Доналдсон и использует как геометрическое средство исследования.

Из более ранних примеров подобного подхода ближе всего сюда (линейная) теория гармонических форм Ходжа. Действительно, исследования Ходжа были прямо навеяны уравнениями Максвелла, и инстантоны — естественное нелинейное обобщение гармонических форм. В линейном случае пространство параметров будет, разумеется, линейным и полностью определяется своей размерностью, в нелинейном же случае пространство параметров несёт гораздо больше информации и представляет собой топологически нетривиальное многообразие.

Для своего успеха доналдсоновский замысел требовал глубокого проникновения в теорию уравнений Янга — Миллса. Нужны были соответствующие теоремы о существовании и регулярности решений, а также теоремы сходимости, а большинство этих результатов являются весьма тонкими, ибо включают в себя и локальные, и глобальные аспекты. К счастью, стараниями Таубса [6, 7] и Уленбекк [8, 9] необходимые аналитические заготовки уже имелись в распоряжении, и можно было непосредственно приступить к использованию инстантонов как эффективного геометрического средства. Однако инстантоны — это не товар, готовый к употреблению; для того чтобы применять их, необходимо понимать вплоть до тонкостей все детали и методы анализа. И Доналдсону пришлось освоиться со всем этим, прежде чем приступить к геометрическим приложениям инстантонов.

Уравнения Янга — Миллса зависят от фиксированной метрики на четырёхмерном многообразии, и, как и в случае теории Ходжа, Доналдсону, чтобы получить результаты, которые зависят лишь от самого многообразия, пришлось изучить эффект варьирования метрики. Из-за нелинейности здесь всё значительно более сложно, чем в теории Ходжа, и преодоление возникающих трудностей требует больших усилий.

На самом деле уравнения Янга — Миллса зависят лишь от конформного класса метрики, и эта конформная инвариантность играет фундаментальную роль в физике, ибо из-за неё отсутствует основной масштаб длины. С точки зрения анализа в этом заключён источник многих трудностей, ибо указанная конформная инвариантность приводит к тому, что соответствующие уравнения принадлежат тем пограничным случаям, когда обычные соображения компактности не проходят, так что последовательность инстантонных решений может в пределе сходиться к дельта-функции Дирака. Однако именно этот тонкий «дефект» Доналдсон использует в своих геометрических целях: дельта-функции, являющиеся нежелательными сингулярностями, вместе с тем оказываются нитью, связывающей четырёхмерные многообразия с пространством инстантонных параметров. Можно сказать, что суть доналдсоновской теории с физической точки зрения — это двойственность частиц и полей.

Когда Доналдсон доказал свой первый результат, было далеко не ясно, особый ли это случай или инстантоны могут найти и более широкие применения. Но Доналдсон продолжал с большим искусством и силой развивать и применять теорию инстантонов и добился в этом замечательных успехов. Он распространил свои результаты на случай незнакоопределённых матриц пересечения, наложив новые ограничения на топологию дифференцируемых четырёхмерных многообразий. Идя в другом направлении, он построил новые инварианты четырёхмерных многообразий, приспособленные для различения гладких гомеоморфных многообразий. В частности, он показал, что здесь призваны сыграть ключевую роль комплексные алгебраические поверхности (комплексной размерности два, т. е. вещественной размерности четыре). В своей весьма изящной работе [2] он доказал теорему существования, которая показывает, что на алгебраических поверхностях инстантоны (точнее, их пространства параметров) допускают чисто алгебраическое описание; соответствующая конструкция совпадает с тем, что алгебраические геометры называют стабильными векторными расслоениями. Поэтому новые доналдсоновские инварианты могут быть сосчитаны алгебраически, и он использовал это [3] для различения двух алгебраических поверхностей, гомеоморфных, но не диффеоморфных между собой. Одна из этих поверхностей рациональна, и результаты Доналдсона настойчиво подводят к мысли, что рациональность алгебраической поверхности может быть дифференциальным свойством (топологическим свойством она *не* является).

Я уже упоминал, что математики занимались и собственно физической проблемой точного описания всех инстантонов в четырёхмерном эвклидовом пространстве. В короткой, но содержательной работе [4] Доналдсон связал эту проблему с теорией алгебраических векторных расслоений на комплексной проективной плоскости (реализованной как компактификация пространства $R^4 = C^2$). Далее, применив ряд простых соображений [5], он решил связанную с предыдущей, но более трудную физическую проблему — проблему магнитных монополей. Он доказал следующий замечательный простотой своей формулировки результат: параметрическое пространство монополей заряда k можно отождествить с пространством рациональных функций комплексной переменной степени k.

В те времена когда Доналдсон получал свои первые результаты о четырёхмерных многообразиях, его идеи были столь новы и необычны для топологов и геометров, что они просто немели от удивления и восхищения. Но постепенно ко всему этому стали привыкать, и ныне идеи Доналдсона начинают новую жизнь в исследованиях других математиков, применяющих их на множество ладов.

Из сказанного мною с очевидностью вытекает, что Доналдсону суждено было открыть целую новую область: взору открылись неожиданные и таинственные явления, касающиеся геометрии четырёхмерных многообразий. Далее, его методы, основанные на использовании трудных нелинейных дифференциальных уравнений, новы и замечательно тонки. И в то же время эта теория заведомо лежит в главном русле математического потока, имея глубокие корни в прошлом, черпая идеи из теоретической физики и обнаруживая красивые связи с алгебраической геометрией.

Внушает надежды тот поистине замечательный факт, что столь молодой математик за такое короткое время сумел овладеть столь широким кругом идей и методов и так блистательно применить их. Это знак того, что математика не утратила ни своего единства, ни жизнеспособности.

ЛИТЕРАТУРА

1. S. K. Donaldson, Self-dual connections and the topology of smooth 4-manifolds, Bull. Amer. Math. Soc. **8** (1983), 81—83.

2. —, Anti-self-dual connections over complex algebraic surfaces and stable vector bundles, Proc. London Math. Soc. **50** (1985), 1—26.

3. —, La topologie differentielle des surfaces complexes, C. R. Acad. Sci. Paris **301** (1985), 317—320.

4. —, Instantons and geometric invariant theory, Comm. Math. Phys. **93** (1984), 453—460.

5. —, Nahm's equations and the classification of monopoles, Comm. Math. Phys. **96** (1984), 387—407.

6. C. H. Taubes, Self-dual connections on non-self-dual 4-manifold. J. Differential Geom. **17** (1982), 139—170.

7. —, Self-dual connections on manifolds with indefinite intersection matrix, J. Differential Geom. **19** (1984), 517—560.

8. K. K. Uhlenbeck, Connections with L^p bounds on curvature, Comm. Math. Phys. **83** (1982), 11—30.

9. —, Removable singularities in Yang-Mills fields, Comm. Math. Phys. **83** (1982), 31—42.

ОКСФОРДСКИЙ УНИВЕРСИТЕТ, МАТЕМАТИЧЕСКИЙ ИНСТИТУТ,
ОКСФОРД OX1 3LB, АНГЛИЯ

О МАТЕМАТИЧЕСКИХ ОТКРЫТИЯХ ГЕРДА ФАЛЬТИНГЗА

Барри Мазур[1]

Одним из крупнейших событий в математике последнего времени явилась разработка Гердом Фальтингзом цикла идей, приведших его к доказательству гипотезы Морделла (см. [1], а также [2; 3]).

Эта гипотеза, замечательная простотой своей формулировки, более полувека была для всех соблазнительной и неуловимой целью. Она даже старше филдсовской медали! На современном языке она звучит так:

Если K — произвольное числовое поле и X — любая кривая рода >1, то X имеет лишь конечное число K-рациональных точек.

Чтобы дать прочувствовать степень нашего незнания в этом круге вопросов, заметим, что до результатов Фальтингза не было известно ни одной кривой X (рода, большего 1), для которой было бы доказано, что данное утверждение верно для всех числовых полей K, над которыми эта кривая определена!

Ещё в двадцатые годы в работах А. Вейля и Зигеля была предпринята серьёзная попытка взять штурмом эту проблему. Под влиянием диссертации Вейля Зигель применил здесь методы алгебраической геометрии и доказал, что множество *целочисленных* решений полиномиального уравнения $f(X, Y) = 0$ (т. е. решений в кольце целых чисел числового поля K) конечно, в предположении что f определяет кривую над K рода >0 или же кривую рода нуль, имеющую на бесконечности не менее трёх рациональных точек.

В своей диссертации А. Вейль обобщил теорему Морделла о конечной порождённости группы рациональных точек эллиптической кривой на случай абелевых многообразий любой

---

[1] Barry C. Mazur, On some of the mathematical contributions of Gerd Faltings, ICM86, pp. 7—12.

размерности. Вейль надеялся, применив этот результат (о конечной порождённости) к якобиану кривой, показать затем, что при вложении кривой рода >1 в её якобиан лишь конечное число рациональных точек якобиана может лежать на кривой. Не найдя способа строго обосновать это утверждение, он решил представить в качестве диссертации доказательство результата о конечной порождённости (т. е. «теоремы Морделла — Вейля»), хотя Адамар советовал ему не останавливаться на этом, как он считал, половинчатом результате!

После работ Вейля и Зигеля в течение тридцати лет не было почти никакого прогресса.

Существенные сдвиги в алгебраической геометрии и алгебраической теории чисел (оказавшие потом влияние на работы Фальтингза) произошли лишь в шестидесятые годы и в начале семидесятых (работы Гротендика, Серра, Мамфорда, Нерона, Тэйта, Манина, Шафаревича, Паршина, Аракелова, Зархина, Рэйно и других).

Эти достижения, существенно использованные в работах Фальтингза, группируются в три крупные математические темы, и Фальтингз доказал гипотезу Морделла, установив сперва справедливость ряда других выдающихся гипотез, касавшихся фундаментальных явлений в арифметике и арифметической геометрии. В ближайшие несколько минут я попытаюсь дать обзор достижений Фальтингза, затронув эти темы и эти гипотезы.

1. Арифметизация геометрии и геометризация арифметики. Аналогия между числовыми полями и полями рациональных функций на алгебраической кривой (над конечным полем) в настоящее время настолько прочно вошла в наши представления, связанные как с теорией чисел, так и с теорией алгебраических кривых, что уже невозможно и вообразить, как можно работать с одной из этих теорий, не привлекая другой. Понимание плодотворности этой аналогии было осознано уже в конце прошлого столетия в работах Кронекера, а также Дедекинда и Вебера. Это понимание ещё более углубилось с дальнейшим развитием алгебраической теории чисел — в трудах Артина и Шевалле, — а также алгебраической геометрии, начала которой были заложены Зарисским, А. Вейлем и Серром, а ближе к нашему времени, — арифметической алгебраической геометрии, основания которой излагаются ныне на языке схем Гротендика.

С точки зрения теории схем гладкие кривые над конечным полем и кольца целых чисел числовых полей не просто аналогичны друг другу, но являются двумя воплощениями одного и того же понятия (*регулярные одномерные схемы конечного типа над* $\mathbf{Z}$). Подобным же образом соотносятся друг с другом семейство кривых над некоторой «базовой» кривой над конечным полем и кривая над кольцом целых чисел некоторого числового поля.

Однако нельзя сказать, что эта аналогия полностью понята! Лишь относительно недавно мы начали, благодаря основополагающим работам Аракелова, осознавать место *архимедовых точек* числового поля в целостной геометрической картине.

К тому же этот «синтетический взгляд», исключительно эффективный при переносе проблем и гипотез из царства функциональных полей в царство числовых полей, работает куда менее эффективно, когда приходится переносить *доказательства* этих гипотез из одного царства в другое.

К примеру, аналог гипотезы Морделла для функциональных полей был установлен Маниным ещё в 1963 г. Другое доказательство было дано Грауэртом в 1965 г. Используя красивую идею Паршина, Аракелов нашёл в 1971 г. новое доказательство. И ещё одно доказательство, также основанное на идее Паршина, было дано Зархиным в 1974 г.

Но даже будучи вооружены всеми этими подходами в функциональном случае, математики в течение ещё почти десятилетия ничего не могли поделать с числовым случаем, пока, наконец, Фальтингз не обнаружил метод, аналогичный методу Паршина — Зархина, который и привёл к доказательству классической гипотезы Морделла над числовыми полями. Однако до сего дня неизвестно аналогов для числового случая других подходов к этой проблеме — скажем, аналога подхода Манина или же Аракелова (существуют ли они?) Если судить по работе Фальтингза [4], относящейся к теории Аракелова, то создаётся впечатление, что он параллельно развивал два подхода к гипотезе Морделла в числовом случае: один, базирующийся на работах Аракелова, другой — в духе Зархина. К доказательству гипотезы привёл второй путь.

Метод Фальтингза в числовом случае и метод Зархина в функциональном случае состоят в том, что гипотеза Морделла выводится из ответа на один более «геометрический» вопрос.

А именно, Кодаира поставил проблему изучения или, возможно, «классификации» всех (непостоянных) семейств гладких кривых данного рода над фиксированной (не обязательно полной) базовой кривой. На Международном конгрессе 1962 года Шафаревич впервые привлёк внимание исследователей к арифметическому аналогу проблемы Кодаиры и объяснил его большое значение. Одна из версий подобной аналогии, ныне известная как *гипотеза Шафаревича для кривых*, утверждает следующее:

Существует лишь конечное число неизоморфных кривых данного рода > 1, определённых над фиксированным числовым полем и имеющих хорошую редукцию вне фиксированного конечного множества простых идеалов кольца целых чисел этого поля.

Один из способов перефразировать первоначальную проблему Кодаиры состоит в формулировании «функционального аналога» гипотезы Шафаревича, когда кольцо целых числового поля заменяется на некоторую базовую кривую над конечным полем. В 1968 г. Паршин нашёл удивительный приём, одинаково хорошо работающий и в числовом, и в функциональном случае, который позволил ему доказать, что гипотеза Морделла следует из гипотезы Шафаревича.

Грубо говоря, идея Паршина состоит в следующем. Зафиксируем $g > 1$. Для кривой X рода g и рациональной точки P на X над числовым полем K Паршин построил накрытие Y кривой X, разветвлённое только над P, у которого поле определения и множество точек плохой редукции «равномерно ограничены» в терминах исходных данных (g, поле определения кривой X и P, а также множество точек плохой редукции для X). Поскольку Y определяет пару (X, P) с точностью до конечной неопределённости и поскольку из гипотезы Шафаревича следует, что таких Y лишь конечное число, то мы получаем, что и таких P тоже лишь конечное число (гипотеза Морделла).

В соответствующем контексте и Зархин, и Фальтингз доказывали гипотезу, устанавливая справедливость гипотезы Шафаревича (над функциональным и над числовым полем соответственно) и привлекая затем идею Паршина.

2. Кривые и абелевы многообразия. Впервые переход от кривых к абелевым многообразиям был предпринят А. Вейлем при доказательстве «гипотезы Римана для кривых над конечными полями»; этот переход позволил получить важные следствия для арифметики кривых.

«Геометрическая» интуиция подсказывала и в более давние времена сводить рассмотрение вопросов, связанных с кривыми, к их якобианам. Действительно, факт существования тесных связей между кривыми и их якобианами относится к богатому наследию итальянской школы алгебраической геометрии.

Тенденция арифметизации геометрии приводит к естественному вопросу о построении моделей для якобианов кривых и, более общо, для абелевых многообразий над кольцами целых числовых полей. В 1964 г. Нерон обнаружил факт, замечательный и сам по себе, и для приложений: любое абелево многообразие над числовым полем имеет «наилучшую» модель — «наилучшую» с точки зрения хороших свойств редукции этой модели по модулю простых идеалов кольца. Хотя Нерон, когда занимался этим вопросом, и не знал об исследованиях Кодаиры для комплексно-аналитического случая, теорию Нерона можно рассматривать как далеко идущее дополнение и арифметизацию программы, начатой Кодаирой.

Теперь модели Нерона играют важную роль при любом серьёзном изучении арифметических свойств абелевых многообразий, и в частности при детальном анализе компактификаций пространств модулей абелевых многообразий. Систематическое *арифметическое* исследование пространств модулей и их компактификаций — исследование, начатое впечатляющей работой Мамфорда, — играет в свою очередь ключевую роль в подходе Фальтингза. К компактификации пространств модулей абелевых многообразий над **Z** Фальтингз возвращался и позже. Уточнив в определённых аспектах результаты Цзин-Ли Чжая, Фальтингз прояснил некоторые вопросы арифметической компактификации и тем самым существенно упростил логическую структуру своего доказательства гипотезы Морделла, в результате чего оно стало гораздо более естественным. В первоначальном доказательстве [1] Фальтингз шёл куда более окольным путём, используя многообразия модулей кривых, а не абелевых многообразий (и данное в [1] изложение технических моментов доказательства чересчур сжато и рассчитано лишь на весьма искушённого читателя).

Ввиду тесной связи кривых с их якобианами (классическая теорема Торелли плюс результат о конечности при учёте поляризаций), гипотезу Шафаревича для кривых (рода, большего 1) можно свести к аналогичной гипотезе (также называемой *гипотезой Шафаревича*) для абелевых многообразий:

Имеется лишь конечное число абелевых многообразий данной размерности над фиксированным числовым полем, у которых модель Нерона имеет хорошую редукцию вне некоторого фиксированного конечного множества простых данного числового поля.

Эта гипотеза также была доказана Фальтингзом. Исследование было проведено в тандеме с другим фундаментальным арифметическим вопросом:

3. Абелевы многообразия и представления Галуа. Рассматривая теоретико-числовой аналог классической гипотезы Ходжа (об алгебраических циклах) и геометрические аналоги гипотезы

Бёрча — Суиннертон-Дайера, Тэйт в 1963 г. сформулировал следующую гипотезу, связывающую проблему классификации абелевых многообразий (с точностью до изогении) с проблемой классификации ассоциированных с ними представлений Галуа:

Пусть l — простое число. Пусть K — числовое поле и $\bar{K}$ — его алгебраическое замыкание. Всякое абелево многообразие A над K определяется однозначно с точностью до изогении естественным представлением группы Галуа $\mathrm{Gal}(\bar{K}/K)$ *на* $\mathbf{Q}$*-векторном пространстве*

$$V_l = \mathrm{Hom}\,(\mathbf{Q}_l,\ A_l(\bar{K})),$$

где $A_l(\bar{K})$ *обозначает группу* $\bar{K}$*-значных точек многообразия* A*, порядок которых есть степень числа* l.

Это фундаментальное утверждение о связи между абелевыми многообразиями (обитателями алгебраической геометрии) и представлениями Галуа (на первый взгляд, более «элементарными» существами) было доказано самим же Тэйтом в 1967 г. для случая конечного поля K, но над числовыми полями это не было доказано даже для эллиптических кривых, вплоть до работы Фальтингза.

С помощью красивого рассуждения (существенно использующего две теории, о которых вскоре будет сказано, и как челнок снующего туда-сюда между гипотезой Шафаревича для абелевых многообразий и гипотезой Тэйта) Фальтингз показал, что обе эти гипотезы справедливы. Выражение «снующего как челнок» лишь очень неточно характеризует идею рассуждения Фальтингза, в ткань которого вплетаются все те математические темы, которых я выше коснулся.

Указанные «две теории» — это *теория высот* и *теория p-делимых групп* (или, более общо, *групповых схем экспоненты p*).

Начала *теории высот* были заложены А. Вейлем в 1928 г. как техника «подсчёта» рациональных точек на абелевых многообразиях; эта техника существенно использовалась им при доказательстве теоремы «Морделла — Вейля». Эта теория была далее развита Нероном, Тэйтом, а в более недавнее время повернулась к нам новой стороной в работах Аракелова.

Теория p-делимых групп развита была Серром и Тэйтом и, независимо, Барзотти в середине шестидесятых годов для того, чтобы проанализировать характер «вырождения» точек, порядок которых есть степень p, при специализации в характеристику p. В 1966 г. Тэйт доказал аналог своей гипотезы об абелевых многообразиях для p-делимых групп над локальными полями характеристики 0. Фальтингз использовал и саму эту теорему, и её важное уточнение (покрывающее случай групповых схем экспоненты p), принадлежащее Рэйно.

Данное выше перечисление не исчерпывает полностью список давно стоявших гипотез, которые были решены Фальтингзом в ходе доказательства гипотезы Морделла. К примеру, важным дополнением к гипотезе Тэйта о представлениях группы $\mathrm{Gal}(\bar{K}/K)$ на точках l-примарного кручения абелева многообразия A, определённого над числовым полем K, является утверждение о *полупростоте* представления $\mathrm{Gal}(\bar{K}/K)$ на $V_l(A)$. Эта гипотеза о полупростоте также была доказана Фальтингзом (её «функциональный аналог» был ранее доказан в работах Зархина). При этом доказательство гипотезы о полупростоте играет ключевую роль в доказательствах других гипотез.

Указанная *гипотеза о полупростоте* была сформулирована Гротендиком как «одномерный случай» более общего вопроса (о полупростоте представления Галуа, действующего на d-мерных l-адических когомологиях гладких неприводимых проективных многообразий над числовыми полями, для любых d). В то время когда была предложена эта гипотеза, свидетельствующие в её пользу факты могли быть найдены в исследованиях Серра о богатстве действия Галуа на точках кручения эллиптических кривых, определенных над числовыми полями. Благодаря Фальтингзу мы знаем теперь, что гипотеза Гротендика о полупростоте справедлива при $d = 1$. Так случилось, что этот результат Фальтингза, вместе с техникой, использованной при его доказательстве, нашёл совсем недавно применение в глубоком исследовании Серра действия Галуа на точках кручения абелевых многообразий произвольной размерности g, определённых над числовыми полями.

Общий случай гипотезы Гротендика (т. е. случай $d > 1$) всё ещё является открытой проблемой.

Мы обсудили лишь подход Герда Фальтингза к доказательству гипотезы Морделла, однако и все другие его математические работы — рассматривает ли он многообразия модулей абелевых многообразий, теорему Римана — Роха для арифметических поверхностей или же p-адическую теорию Ходжа — сразу же впечатляют как работы поразительно оригинального ума, от которого в будущем мы можем ждать столь же удивительных вещей.

ЛИТЕРАТУРА

1. G. Faltings, Endlichkeitssätze für abelsche Varietäten über Zahlkörpern, Invent. Math. **73** (1983), 349—366.

2. G. Faltings, G. Wüstholtz, et al., Rational points (Seminar Bonn/Wuppertal, 1983/84), Aspects of Math., vol. E6, Vieweg, Braunschweig-Wiesbaden, 1984.

3. L. Szpiro, Séminaire sur les pinceaux arithmétiques: La conjecture de Mordell, Astérisque, no. **127**, Soc. Math. France, Paris, 1985. [Имеется перевод: Шпиро Л. Гипотеза Морделла. — В сб.: Алгебра и теория чисел с приложениями. — М.: Мир, 1987, с. 125—150.]

4. G. Faltings, Calculus on arithmetic surfaces, Ann. of Math. (2) **119** (1984), 387—424.

ГАРВАРДСКИЙ УНИВЕРСИТЕТ, КЕЙМБРИДЖ, МАССАЧУСЕТС 02138, США

О ДОСТИЖЕНИЯХ МАЙКЛА ФРИДМАНА

Джон Милнор[1]

Майкл Фридман не только доказал гипотезу Пуанкаре для четырёхмерных топологических многообразий, охарактеризовав тем самым сферу S^4, но и дал нам теоремы классификации для значительно более общих четырёхмерных многообразий, теоремы, которые трудно доказать, но легко сформулировать и использовать. Простота формулировки его результатов, относящихся к топологическому случаю, находится в разительном контрасте с теми чрезвычайными сложностями, которые, как теперь нам известно, возникают при изучении дифференцируемых и кусочно-линейных четырёхмерных многообразий.

«n-мерная гипотеза Пуанкаре» — это гипотеза о том, что всякое топологическое n-мерное многообразие, имеющее те же группы гомологий и ту же фундаментальную группу, что и n-мерная сфера, в действительности гомеоморфно n-мерной сфере. Для случаев $n = 1$ и $n = 2$ это было известно ещё в девятнадцатом веке, а для случая $n \geqslant 5$ гипотеза была доказана Смейлом и независимо Столлингсом, Зиманом и Уоллесом в 1960—1961 гг. (Первоначальные доказательства нуждались в дополнительном предположении о дифференцируемости или кусочной линейности, которое несколько лет спустя было снято Ньюманом.) Трёхмерный и четырёхмерный случаи намного труднее.

Фридмановское доказательство (1982 г.) четырёхмерной гипотезы Пуанкаре было проявлением поистине необычайной изобретательности. Его методы были настолько тонки и сильны, что в действительности привели к полной классификации компактных односвязных топологических четырёхмерных многообразий, породив множество неизвестных ранее примеров таких многообразий, а также множество неизвестных ранее гомеоморфизмов

---

[1] John Milnor, The work of M. H. Freedman, ICM86, pp. 13—15.

между известными многообразиями. Фридман показал, что компактное односвязное четырёхмерное многообразие M характеризуется, с точностью до гомеоморфизма, двумя простыми инвариантами. Первый — это двумерная группа гомологий

$$H_2 = H_2(M;\ Z) \cong Z \oplus \ldots \oplus Z$$

вместе с симметричной билинейной формой пересечений

$$\omega\colon\ H_2 \otimes H_2 \to Z.$$

Эта форма, которая определена, коль скоро на M выбрана ориентация, должна иметь определитель ± 1 в силу двойственности Пуанкаре. Второй инвариант — это препятствующий класс Кёрби — Зибенманна, т. е. элемент

$$\sigma \in H^4(M;\ Z/2) \cong Z/2,$$

обращающийся в нуль в том и только том случае, когда M стабильно сглаживаемо. Иными словами, σ равно нулю, если и только если на произведении $M \times R$ можно задать дифференцируемую структуру, или, что равносильно, кусочно-линейную структуру. Эти инварианты ω и σ могут быть заданы произвольно, за тем лишь исключением, что в одном специальном случае должно выполняться некоторое соотношение. А именно, если форма ω чётна, т. е. $\omega(x, x) = 0 \pmod 2$ для каждого $x \in H_2$, то препятствие Кёрби — Зибенманна должно совпадать с инвариантом Рохлина:

$$\sigma \equiv \text{signature}(\omega)/8 \pmod 2.$$

(Первоначальное фридмановское доказательство того, что эти два инварианта характеризуют M с точностью до гомеоморфизма, нуждалось в дополнительном предположении о «почти-дифференцируемости», которое позднее было снято Куинном.)

Если форма пересечений $\omega \neq 0$ является незнакоопределённой или же её ранг не превосходит одиннадцати, то из известных результатов о квадратичных формах следует, что многообразие M может быть представлено (не единственным способом) как связная сумма простых строительных блоков четырёх сортов, каждый из которых наделяется либо стандартной ориентацией, либо противоположной. А именно, нужны произведение $S^2 \times S^2$, комплексная проективная плоскость CP^2 и ещё два экзотических многообразия, впервые построенные Фридманом. Одно из них — некоторый недифференцируемый аналог комплексной проективной плоскости, а другое — единственное многообразие, для которого форма пересечений ω положительно-определённа, чётна и имеет ранг восемь. (Эта форма ω может быть отождествлена с решёткой, порождённой корневыми векторами группы Ли E_8. Как отметил в 1952 г. Рохлин, четырёхмерное многообразие с такой формой пересечений не может быть дифференцируемым.)

Если же форма пересечений является положительно-определённой, то с увеличением среднего числа Бетти число различных односвязных многообразий растёт быстрее чем по экспоненте.

Методы Фридмана распространяются также на некомпактные четырехмёрные многообразия. Например, он показал, что можно ввести такую экзотическую дифференцируемую структуру на произведении $S^3 \times R$, чтобы оно содержало гладко вложенную трёхмерную гомологическую сферу Пуанкаре и тем самым не допускало гладкого вложения в четырёхмерное эвклидово пространство [11, 16]. Его методы применимы также ко многим неодносвязным многообразиям [22]. Например, «плоская» двумерная сфера в четырёхмерном пространстве незаузлена тогда и только тогда, когда её дополнение имеет свободную циклическую фундаментальную группу; плоская одномерная сфера в S^3 имеет тривиальный полином Александера, если и только если она ограничивает плоский двумерный диск в единичном четырёхмерном диске, дополнение которого имеет свободную циклическую фундаментальную группу.

Доказательства этих результатов чрезвычайно трудны. Основная идея, которую для малых размерностей использовали Мёбиус и Пуанкаре, а для высоких размерностей Смейл и Уоллес, заключается в том, чтобы строить заданное четырёхмерное многообразие по индукции, начиная с четырёхмерного диска и последовательно добавляя ручки. Существенная трудность, не возникающая в высших размерностях, появляется при попытке контролировать фундаментальную группу, вставляя двумерные ручки, поскольку двумерный диск, погружённый в четырёхмерное многообразие, обычно будет иметь самопересечения. Первую атаку на эту проблему предпринял Кассон, который показал, как в четырёхмерном многообразии построить двумерную ручку некоторого обобщённого вида с заданной границей. Главный технический инструмент Фридмана — теорема, утверждающая, что каждая ручка Кассона в действительности гомеоморфна стандартной открытой ручке, т. е. произведению (замкнутый двумерный диск)$\times$(открытый двумерный диск). Доказательство основано на тонко регулируемом бесконечном итерационном процессе в духе топологической школы Бинга и в некотором смысле является существенно недифференцируемым.

ЛИТЕРАТУРА

Работы Майкла Фридмана

1. Automorphisms of circle bundles over surfaces, Geometric Topology, Lecture Notes in Math., vol. 438, Springer-Verlag, 1974, pp. 212—214.

2. On the classification of taut submanifolds, Bull. Amer. Math. Soc. **81** (1975), 1067—1068.

3. Uniqueness theorems for taut submanifolds, Pacific J. Math. **62** (1976), 379—387.

4. Surgery on codimension 2 submanifolds, Mem. Amer. Math. Soc. No. 119 (1977).

5. Une obstruction élémentaire à l'existence d'une action continue de groupe dans une variété (with W. Meeks), C. R. Acad. Sci. Paris Ser. **A286** (1978), 195—198.

6. Λ-splitting 4-manifolds (with L. Taylor), Topology **16** (1977), 181—184.

7. A geometric proof of Rochlin's theorem (with R. Kirby), Proc. Sympos. Pure Math., vol. 32, part 2, Amer. Math. Soc., Providence, R. I., 1978, pp. 85—98.

8. Remarks on the solution of first degree equations in groups, Algebraic and Geometric Topology, Lecture Notes in Math., vol. 664, Springer-Verlag, 1978, pp. 87—93.

9. Quadruple points of 3-manifolds in S^4, Comment. Math. Helv. **53** (1978), 385—394.

10. A converse to (Milner-Kervaire treorem)$\times R$ etc. ..., Pacific J. Math. **82** (1979), 357—369.

11. A fake $S^3 \times R$, Ann. of Math. **110** (1979), 177—201.

12. Cancelling 1-handles and some topological imbeddings, Pacific J. Math. **80** (1979), 127—130.

13. A quick proof of stable surgery (with F. Quinn), Comment. Math. Helv. **55** (1980), 668—671.

14. Planes triply tangent to curves with nonvanishing torsion, Topology **19** (1980), 1—8.

15. Slightly singular 4-manifolds (with F. Quinn), Topology (1981), 161—173.

16. The topology of 4-manifolds, J. Differential Geom. **17** (1982), 357—454 (see also F. Quinn, ibid., p. 503).

17. A surgery sequence in dimension four; the relations with knot concordance, Invent. Math. **68** (1982), 195—226.

18. Closed geodesics on surfaces (with J. Hass, P. Scott), Bull. London Math. Soc. **14** (1982), 385—391.

19. A conservative Dehn's Lemma, Low Dimensional Topology, Contemp. Math., vol. 20, Amer. Math. Soc., Providence, R. I., 1983, pp. 121—130.

20. Imbedding least area incompressible surfaces (with J. Hass, P. Scott), Invent. Math. **71** (1983), 609—642.

21. Homotopically trivial symmetries of Haken manifolds are toral (with S.-T. Yau), Topology **22** (1983), 179—189.

22. The disk theorem for 4-dimensional manifolds, Proc. Internat. Congr. Math. Warsaw 1983, vol. 1, Pol. Sci. Publ., 1984, pp. 647—663.

23. There is no room to spare in four-dimensional space, Notices Amer. Math. Soc. 31 (1984), 3—6.

24. Atomic surgery problems (with A. Casson), Four-Manifold Theory, Contemp. Math., vol. 35, Amer. Math. Soc., Providence, R. I., 1984, pp. 181—199.

Работы других авторов

25. H. Poincaré, Cinquième complément á l'Analysis situs (1904), Oeuvres 6, Paris, 1953, pp. 435—498.

26. L. Siebenmann, Amorces de la chirurgie en dimension 4, un $S^3 \times R$ exotique (d'apres A. Casson et M. H. Freedman), Séminaire Bourbaki (1978—79), Exp. No. 536, Lecture Notes in Math., vol. 770, Springer-Verlag, 1980, pp. 183—207.

27. —, La conjecture de Poincaré topologique en dimension 4 (d'apres M. H. Freedman), Séminaire Bourbaki (1981—82), Exp. No. 588, Astérisque, no. 92—93, Soc. Math. France, Paris, 1982, pp. 219—248.

ИНСТИТУТ ВЫСШИХ ИССЛЕДОВАНИЙ, ПРИНСТОН, НЬЮ-ДЖЕРСИ 08540, США

О ДОСТИЖЕНИЯХ ЛЕСЛИ ВАЛЬЯНТА

Фолькер Штрассен[1]

Теоретическая информатика (computer science) очень молода по сравнению, например, с теорией чисел, геометрией или топологией. В то время как эти классические области подобны величественным старым дубам, вознёсшимся до таких головокружительных высот, что нелегко охватить их взглядом, теоретическая информатика напоминает быстро растущее молодое дерево, свежая зелень которого доставит радость любому, кто подойдёт к нему.

Лесли Вальянт решающим образом повлиял на рост практически всех ветвей этого молодого дерева. Чтобы дать некоторое представление о размахе его работы и о стремительной поступи его творчества, я сначала расскажу о трёх ранних работах, опубликованных в один год и содержащих яркие результаты в трёх совершенно различных областях. Затем я перейду к наиболее, вероятно, важным и зрелым исследованиям Вальянта, концентрирующимся вокруг его теории проблем подсчёта.

ЯЗЫКИ

Контекстно-свободные грамматики были введены Н. Хомским в 1956 г. как средство анализа естественных языков. Теперь их широко используют для описания структуры языков программирования. Проблема распознавания предложений языка, определённого такой грамматикой, составляет центральную алгоритмическую проблему. В течение ряда лет были найдены алгоритмы, распознающие предложения длины n за время порядка n^3, но, несмотря на все приложенные усилия, никак не удавалось, кроме как в специальных случаях, получить значительных улучшений этого результата.

---

[1] Volker Strassen, The work of Leslie G. Valiant, ICM86, pp. 16—22.

Вальянт, предприняв смелую вылазку, показал в 1975 г., что проблема распознавания может быть решена за менее чем кубическое время, сведя её к умножению целочисленных матриц. Он никогда не возвращался к этому предмету, но остроумные алгоритмические редукции комбинаторных проблем к алгебраическим стали одной из главных тем его творчества.

ГРАФЫ

Пусть m — натуральное число. Под *m-суперконцентратором* понимается ориентированный граф с m входными и m выходными вершинами, такой что для каждого $r \leqslant m$ любые r входных вершин можно соединить с любыми r выходными вершинами в некотором порядке с помощью r попарно непересекающихся ориентированных путей. Размером m-суперконцентратора называется число его рёбер.

Суперконцентраторы впервые появились в алгебраической теории сложности: любой прямолинейный (straightline) алгоритм вычисления *дискретного преобразования Фурье* порядка m приводит к некоторому m-суперконцентратору, размер которого пропорционален длине алгоритма. В частности, *быстрое преобразование Фурье* даёт примеры m-суперконцентраторов размера $\mathrm{const} \cdot m \log m$, и любое улучшение быстрого преобразования Фурье привело бы к m-суперконцентраторам ещё меньшего размера. Ахо, Хопкрофт и Ульман среди прочих поставили проблему: доказать или опровергнуть, что минимальный размер m-суперконцентратора растёт как $m \log m$. Согласно предыдущим замечаниям, положительное решение привело бы к доказательству оптимальности с точностью до порядка для быстрого преобразования Фурье.

Вальянт разрушил подобные надежды, показав, что существуют m-суперконцентраторы линейного по m размера. Если учесть, насколько гибки соответствующие графы, в этот результат почти невозможно поверить. В доказательстве Вальянта, основанном на предшествующих результатах Пинскера, некий подсчёт сочетается с элегантной рекурсивной конструкцией. (Хочу отметить, что Маргулис и Габбер-Галил сумели и этот подсчёт тоже заменить некоторой явной конструкцией.) Суперконцентраторы линейного размера в течение десятилетия, последовавшего за их открытием, стали полезным инструментом в теории информации и связи, выйдя далеко за рамки своего первоначального предназначения.

Этот результат Вальянта — лишь один из примеров, характеризующих проведённое им систематическое и глубокое исследование свойств графов, связанных с эффективным поиском путей и вложениями; эти исследования привели в последние годы к теории параллельных компьютеров общего назначения (так называемые суперкомпьютеры).

МАШИНЫ ТЬЮРИНГА

С тех пор как А. Тюринг изобрёл их в 1936 г., машины Тьюринга служат основной теоретической моделью, на которой базируются понятия вычислимости и сложности вычислений. Машины Тьюринга часто используют в качестве разрешающей процедуры для тех или иных свойств чисел, графов, логических формул и т. д., которые после подходящей кодировки представляются машине как двоичные слова. В то время как теория рекурсии занимается разрешимостью проблем разрешения, теория сложности учитывает также, сколько времени и пространства требуется для того, чтобы получить решение.

Для заданной функции t: $\mathbb{N} \to \mathbb{N}$ пусть $\mathrm{TIME}(t)$ — класс тех проблем разрешения (т. е. множеств двоичных слов), которые могут быть решены на многоленточной машине Тьюринга за $O(t(n))$ шагов на входах длины n. Аналогично определим $\mathrm{SPACE}(t)$ через число используемых ячеек ленты. Очевидно, $\mathrm{TIME}(t) \subset \mathrm{SPACE}(t)$, так как за один шаг машина Тьюринга может достичь не более некоторого постоянного числа новых ячеек.

Пространственно-временной дуализм, в значительной степени свойственный естественным наукам, присутствует также в дискретном мире идеализированных компьютеров. Прежде всего можно задать фундаментальный вопрос, является ли приведённое выше включение строгим:

$$\mathrm{TIME}(t) \subsetneqq \mathrm{SPACE}(t)? \qquad (A)$$

Опыт вычислений со всей убедительностью указывает на то, что это действительно так, но простого доказательства, по-видимому, не существует. Настоящей научной сенсацией явился результат Хопкрофта, Пола и Вальянта, которые показали, что фактически имеет место включение

$$\mathrm{TIME}(t) \subset \mathrm{SPACE}(t/\log t), \qquad (B)$$

откуда вытекает (A), поскольку $\mathrm{SPACE}(t/\log t)$ строго содержится в $\mathrm{SPACE}(t)$, как устанавливается стандартным диагональным рассуждением. (Перефразируя классическое изречение Г. Вейля, этот результат можно сформулировать так: пространство и время — это разные вещи.)

Хопкрофт, Пол и Вальянт свели доказательство утверждения (B) к нахождению хорошей стратегии для «игры в камешки» — некоторой игры на графах, которую придумали ранее для других целей Патерсон и Хьюитт. Пол, Тарджан и Челони показали позднее, что эта стратегия оптимальна, так что (B) не может быть улучшено тем же методом. (Приятным совпадением явилось то, что в этом доказательстве оптимальности в качестве существенного ингредиента используются суперконцентраторы линейного размера.)

У результата Хопкрофта, Пола и Вальянта есть одно слабое место: сложностные классы $\mathrm{TIME}(t)$ и $\mathrm{SPACE}(t)$ связаны с выбором машин Тьюринга в качестве вычислительной модели, и свойства (A) и (B) могут утратиться при изменении модели. Более «устойчивой» была бы формулировка проблемы времени и пространства, использующая сложностные классы

$$\mathrm{PTIME} = \bigcup_k \mathrm{TIME}(n^k) \quad \text{и} \quad \mathrm{PSPACE} = \bigcup_k \mathrm{SPACE}(n^k)$$

всех проблем, разрешимых за полиномиальное время или с полиномиальным объёмом. Нет сомнений, что включение

$$\mathrm{PTIME} \subset \mathrm{PSPACE}$$

также строгое, однако это ещё не доказано.

Класс PTIME, обозначаемый ниже просто P, занял центральное положение в теории сложности, так как он наилучшим образом подходит для того, чтобы различать, что может, а что не может быть вычислено на практике. Для краткости я буду называть проблемы из P «лёгкими», а остальные — «трудными».

ПОЛНЫЕ ПРОБЛЕМЫ

Рассмотрим отображение f: $\mathbb{N} \to 2\mathbb{N}$, такое что

$$\text{проблема } \{(x, y) : y \in f(x)\} \text{ — лёгкая,} \qquad (1)$$

$$y \in f(x) \Rightarrow |y| \leqslant |x|^k \qquad (2)$$

для подходящей константы $k(|x|$ обозначает длину соответствующего двоичного слова). Множества всех чисел x, для которых $f(x)$ непусто, называется *проблемой перебора*. Примером служит дополнение к множеству простых чисел: за $f(x)$ можно принять множество всех собственных делителей числа x. Название «проблема перебора» указывает на возможность перебирать все y, удовлетворяющие неравенству (2), в поисках элемента из $f(x)$. Хотя каждая проверка на принадлежность легка в силу (1), такой исчерпывающий перебор может потребовать экспоненциального времени ввиду количества необходимых проверок.

Класс NP всех проблем перебора лежит между P и PSPACE. Оказывается, что большое число проблем разрешения, возникающих в математике и её приложениях, после подходящей кодировки попадает в класс NP. Вот несколько примеров: решить,

выполнима ли пропозициональная формула;
изоморфен ли один граф подграфу другого;
есть ли в графе гамильтонов цикл;
допускает ли граф 3-раскраску;
имеет ли диофантово уравнение $F(X_1, \ldots, X_n) = c$, где c и коэффициенты многочлена F — натуральные числа, решение в натуральных числах.

В 1971 г. С. Кук показал в своей основополагающей работе «Сложность процедур доказательства теорем», что проблема выполнимости, как и проблема подграфов, *полна* в классе NP относительно полиномиальной редукции, или, короче, NP-*полна*. Грубо говоря, NP-полные проблемы имеют максимальную трудность среди всех проблем перебора. Отсюда следует, что если проблема выполнимости или проблема подграфов легка, то любая проблема перебора легка.

Из более точной формулировки теоремы Кука вытекает даже более сильное утверждение, а именно такое: быстрый алгоритм для решения проблемы выполнимости или проблемы подграфов вполне механическим способом приводил бы к быстрой разрешающей процедуре для любой эффективно заданной проблемы перебора. Такой алгоритм служил бы отмычкой к проблемам перебора из всех областей математики. Например, не понадобится никаких дополнительных знаний из теории чисел, чтобы придумать способ быстрой проверки чисел на простоту или на представимость заданным положительным многочленом. Это кажется настолько невероятным, что мало кто сомневается в справедливости так называемой *гипотезы Кука*: проблема выполнимости и проблема подграфов на самом деле трудны, или, что равносильно,

$$P \neq NP$$

Р. Карп в 1972 г. значительно расширил список NP-полных проблем, включив в него среди многого прочего упоминавшиеся выше проблемы гамильтонова цикла и раскраски графа. К настоящему моменту большинство естественно возникающих проблем перебора классифицированы либо как лёгкие, либо как NP-полные. (Как показали Мандерз и Адлеман, проблема разрешимости положительных диофантовых уравнений NP-полна, даже если ограничиться бинарными квадратичными многочленами.)

Обратимся теперь к достижениям Вальянта в этой области. Для заданного отображения $f\colon \mathbb{N} \to 2^{\mathbb{N}}$, такого как в определении проблемы перебора, можно не только ставить вопрос о непустоте множества $f(x)$, но и интересоваться его размером. Вальянт называет функцию $x \longmapsto \# f(x)$ *проблемой подсчёта* (counting problem) и показывает, что среди всех проблем подсчёта также имеются полные; таковы, например, проблемы подсчёта, соответствующие NP-полным проблемам перебора из нашего списка. Поскольку пересчитать решения по меньшей мере так же трудно, как и решить вопрос об их существовании, то по гипотезе Кука полные проблемы подсчёта трудны. Весьма волнующим событием стало открытие Вальянтом в 1979 г. разнообразных полных проблем подсчёта, соответствующих лёгким проблемам перебора, что существенно раздвинуло рамки теории NP-полноты. Приведу три примера:

(I) Пересчёт поддеревьев ориентированного графа. Заметим, что если заменить «поддерево» на «остовное поддерево» (spanning subtree), то проблема подсчёта становится лёгкой ввиду одного варианта классической теоремы Кирхгофа (1847 г.).

(II) Оценка вероятности отказа в ненадёжной сети соединения. Согласно лапласовскому определению вероятности как отношения, это равносильно проблеме подсчёта.

(III) Пересчёт совершенных паросочетаний в двудольном графе. Здесь соответствующая проблема перебора нетривиальна, в отличие от случаев I и II, но легка ввиду хорошо известного алгоритма М. Холла.

Пример III — ключевой в подходе Вальянта. В доказательстве полноты этой проблемы переплетаются идеи из математической логики, теории графов и алгебры.

Число совершенных паросочетаний в двудольном графе равно значению перманента на представляющей граф матрице из нулей и единиц над кольцом целых чисел. Что произойдёт, если $\mathbb{Z}$ заменить на $\mathbb{Z}/m\mathbb{Z}$? При $m = 2$ перманент совпадает с определителем, так что вычислить его легко. Вальянт показывает, что в случае, когда m — степень двойки, проблема остаётся лёгкой. С другой стороны, он получает удивительный результат, согласно которому из существования быстрого алгоритма для перманента по модулю m для некоторого m, не являющегося степенью двойки, вытекает, что всякая полиномиально ограниченная теоретико-числовая функция, график которой легко разрешим, сама легко вычислима. («Полиномиально ограниченная» следует понимать в терминах длин.) Отсюда можно вывести, что если перманент по модулю (скажем) 3 лёгок, то таково же разложение натуральных чисел на простые множители.

Перманент — это полиномиальная функция от элементов матрицы. Тем самым исследования Вальянта естественным образом привели его к *алгебраической теории сложности*, которой мы уже касались при обсуждении суперконцентраторов. Основная модель здесь — это прямолинейный алгоритм, т. е. конечная последовательность арифметических инструкций, которые должны быть выполнены над подходящей алгебраической структурой. Долгое время казалось, что понятие NP-полноты не имеет отношения к этому предмету. Однако Вальянт, отталкиваясь от своих результатов о перманенте, развил убедительный аналог теории проблем перебора и подсчёта в чисто алгебраических рамках. Новая теория значительно отличается от своего прототипа, и не представляется возможным описать здесь её даже в основных чертах. В то же время аналог гипотезы Кука — назовём его *гипотезой Вальянта* — может быть зажат между двумя краткими утверждениями в духе классической алгебры. А именно, фиксируем поле F характеристики $\neq 2$, и пусть $t(n)$ — наименьшее из чисел r, таких что перманент размера n может быть получен из определителя порядка r простой подстановкой (т. е. такой, при которой переменные могут замещаться лишь переменными или элементами из F). Тогда из гипотезы Вальянта следует, что $t(n)$ растёт быстрее любой степени n. В свою очередь, если $t(n)$ растёт даже быстрее, чем $e^{(\log n)^q}$ при некотором q (например, если оно растёт экспоненциально), то гипотеза Вальянта справедлива над полем F.

Некоторым из вас может показаться, что обсуждаемые здесь теории покоятся на шатком фундаменте. Это не так. Свидетельств в пользу гипотез Кука и Вальянта имеется несметное число, а следствия их невыполнения настолько нелепы, что по статусу их можно уподобить скорее физическим законам, чем обычным математическим гипотезам. Тем не менее традиционное доказательство представляло бы большой интерес, и мне кажется, что из них двоих легче установить гипотезу Вальянта (например, используя мощные методы алгебраической геометрии).

В представленной выше выборке из работ Вальянта не отдано должное многим другим его достижениям, сравнимым по важности, — результатам, относящимся к булевой сложности, вероятностным алгоритмам, монотонным и параллельным вычислениям, искусственному интеллекту. Все они убедительно свидетельствуют о его проницательности, оригинальности и вкусе.

Теоретическая информатика находится в стадии формулирования центральных проблем и создания методов для их решения. Вальянт сыграл в этом процессе выдающуюся роль, не только дав ответ на ряд открытых вопросов, упорно не поддававшихся решению, но и, главное, развив важные новые идеи, которые привели его к открытию глубоких и красивых связей между задачами, казавшимися совершенно не связанными друг с другом.

Найти новые плодотворные идеи — вот что больше всего требуется в любой науке. Особенно же справедливо это в отношении новой области, где существует столь много концептуальных возможностей. Но вознаградятся усилия — молодые побеги станут главными ветвями дерева, когда оно поднимется в полный рост.

У меня нет сомнений, что сказанное в полной мере приложимо к исследованиям Лесли Вальянта. Позвольте пожелать ему, как и трём филдсовским медалистам, такого же блестящего будущего, каково их научное прошлое.

ИЗБРАННЫЕ РАБОТЫ ЛЕСЛИ ВАЛЬЯНТА

1. The equivalence problem for deterministic finite-turn pushdown automata, Inform. and Control **25** (1974), 123—133.

2. Parallelism in comparison problems, SIAM J. Comput. **4** (1975), 348—355.

3. General context-free recognition in less than cubic time, J. Comput. System Sci. **10** (1975), 308—315.

4. On non-linear lower bounds in computational complexity, Seventh Annual ACM Symposium on Theory of Computing (Albuquerque, N. M., 1975), Assoc. Comput. Mach., New York, 1975, pp. 45—53. (See also J. Comput. System Sci. **13** (1976), 278—285.)

5. On time versus space and related problems (with J. E. Hopcroft and W. J. Paul), 16th Annual Symposium on Foundations of Computer Science (Berkeley, Calif., 1975), IEEE Computer Society, Long Beach, Calif., 1975, pp. 57—64. (See also J. Assoc. Comput. Mach. **24** (1977), 332—337.)

6. Graph-theoretic arguments in low-level complexity, Lecture Notes in Comput. Sci., vol. 53, Springer-Verlag, 1977, pp. 162—176.

7. Fast probabilistic algorithms for Hamiltonian circuits and matchings (with D. Angluin), J. Comput. System Sci. **18** (1979), 155—193.

8. The complexity of computing the permanent, Theoret. Comput. Sci. **8** (1979), 189—201.

9. The complexity of enumeration and reliability problems, SIAM J. Comput. **8** (1979), 410—421.

10. Completeness classes in algebra, Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing (Atlanta, GA., 1979), Assoc. Comput. Mach., New York, 1979, pp. 249—261.

11. Negation can be exponentially powerful, Theoret. Comput. Sci. **12** (1980), 303—314.

12. Reducibility by algebraic projections, Logic and Algorithms (Zurich, 1980), Monograph. Enseign. Math., No. 30, Univ. Genève, Geneva, 1982, pp. 365—380.

13. A scheme for fast parallel communication, SIAM J. Comput. **11** (1982), 350—361.

14. Universal schemes for parallel communication (with G. J. Brebner), Proceedings of the 13th ACM Symposium on Theory of Computing (Milwaukee, Wisconsin, 1981), Assoc. Comput. Mach., New York, pp. 263—277.

15. Fast parallel computation of polynomials using few processors (with S. Skyum, S. Berkowitz, and C. Rackoff), SIAM J. Comput. **12** (1983), 641—644.

16. Short monotone formula fer the majority function, J. Algorithms **5** (1984), 363—366.

17. A theory of the learnable, Comm. ACM **27** (1984), 1134—1142.

18. Learning disjunctions of conjunctions, Proceedings of Ninth International Joint Conference on Artificial Intelligence (Los Angeles, Calif., 1985), pp. 560—566.

19. A complexity theory based on Boolean algebra (with S. Skyum), J. Assoc. Comput. Mach. **32** (1985), 484—502.

20. A logarithmic time sort for linear size networks (with J. H. Reif), J. Assoc. Comput. Mach. **34** (1987), 60—76.

21. NP is as easy as detecting unique solutions (with V. V. Vazirani), Theoret. Comput. Sci. (to appear).

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ ПРИ ЦЮРИХСКОМ УНИВЕРСИТЕТЕ, 8032 ЦЮРИХ, ШВЕЙЦАРИЯ

ЧАСОВЫЕ ДОКЛАДЫ

ИДЕИ, ЛЕЖАЩИЕ В ОСНОВЕ ДОКАЗАТЕЛЬСТВА ГИПОТЕЗЫ БИБЕРБАХА

Луи де Бранж[1]

Большинство математических событий находит отклик лишь у небольшого числа знатоков, обладающих необходимыми знаниями, чтобы понять и оценить случившееся. Доказательство гипотезы Бибербаха было приятным исключением из этого правила. Этот результат привлёк внимание более широкой, но также и менее однородной, аудитории. А потому не будет лишним ещё раз рассказать об этом открытии.

ХРОНИКА СОБЫТИЙ

Наше обсуждение разумно начать с теоремы Римана о конформных отображениях[2]. Каждый её знает: любая собственная односвязная подобласть комплексной плоскости является образом единичного круга при конформном отображении, такое отображение задаётся степенным рядом, который сходится в единичном круге и представляет функцию, принимающую различные значения в различных точках единичного круга.

Риман сформулировал, но не дал приемлемого доказательства этой теоремы. Она доказывается с использованием аппроксимационных соображений, которые требуют оценок конформных отображений. Любая однолистная в единичном круге функция может быть оценена через значение функции и её производной в нуле. Гипотеза Бибербаха появилась в результате поиска наилучших возможных оценок.

Нормируем однолистную функцию так, чтобы свободный член равнялся нулю, а коэффициент при z был положитель-

---

[1] Lous de Branges, Underlying concepts in the proof of the Bieberbach conjecture, ICM86, pp. 25—42.

[2] Автор всюду использует термин «Riemann mapping function». В переводе употребляются общепринятые термины «конформное отображение» или — чаще — «однолистная функция». — *Прим. перев.*

ным. Таким образом, наше конформное отображение имеет неподвижную точку в нуле. Положительность производной в нуле позволяет однозначно определить однолистную функцию по той области, на которую она отображает единичный круг.

Примером однолистной функции служит функция Кёбе

$$\frac{z}{(1-\omega z)^2}=z+2\omega z^2+3\omega^2z^3+\ \ldots,$$

где ω — константа, по модулю равная единице. Функция Кёбе отображает единичный круг на область, которая получается из комплексной плоскости в результате проведения радиального разреза, начинающегося на расстоянии 1/4 от начала координат.

Гипотеза Бибербаха [4] утверждает, что для любой однолистной функции $f(z)=\sum a_nz^n$ неравенство

$$|a_n|\leqslant na_1$$

имеет место при всех n. Гипотеза утверждает также, что равенство имеет место для некоторого n только тогда, когда f пропорциональна функции Кёбе.

Бибербах доказал свою гипотезу лишь для второго коэффициента. Несмотря на весьма скромные аргументы в её пользу, гипотеза привлекла внимание ведущих математиков. Памятным событием было доказательство гипотезы Бибербаха для третьего коэффициента К. Лёвнером [25] в 1923 г. В основе этого доказательства лежит общий метод, в принципе применимый ко всем коэффициентам. Но требуемые вычисления очень сложны. Когда в 1954 г. Гарабедян и Шиффер [25] дали доказательство гипотезы Бибербаха для четвёртого коэффициента, они использовали другой метод.

На протяжении следующего десятилетия гипотеза Бибербаха была предметом активных исследований. Доказательство гипотезы для шестого коэффициента было получено в 1968 г. Одзавой [34] и Педерсоном [35]. В 1972 г. Педерсону и Шифферу [36] удалось доказать гипотезу Бибербаха для пятого коэффициента.

Все применявшиеся методы становились для больших коэффициентов слишком сложными и в течение более десяти лет гипотеза Бибербаха не была подтверждена ни в каком другом случае. А потом автор получил доказательство гипотезы Бибербаха сразу для всех коэффициентов.

Доказательство было завершено в феврале 1984 г. с помощью Уолтера Готчи [20]. Возникла курьёзная ситуация — сперва не нашлось никого, кто подтвердил бы справедливость

приведённых рассуждений. Однако в рамках соглашения об обмене визитами между Национальной академией наук и Академией наук СССР был запланирован визит автора в Математический институт им. Стеклова в Ленинграде в апреле — июне 1984 г. Во время этого визита ленинградским Семинаром по геометрической теории функций корректность доказательства была подтверждена. Об этом сообщили трое. Одним из них был автор [10], двумя другими — старшие участники ленинградского семинара Г. В. Кузьмина [18] и И. М. Милин [27].

Автор выражает благодарность Национальной академии наук и Академии наук СССР, включившим упомянутый визит в свои планы.

ТЕОРИЯ ИНТЕРПОЛЯЦИИ

Причина, по которой гипотеза Бибербаха представляет интерес, заключена в теории интерполяции. Все математические вычисления, независимо от того, используются при этом машины или нет, по необходимости конечны. Следовательно, бесконечные процессы приходится аппроксимировать конечными. В теории аналитических функций это осуществляет теория интерполяции.

Функция, которая аналитична в единичном круге, представляется степенным рядом. Значения функции, особенно на границе круга, могут вести себя очень сложным образом. А коэффициенты степенного ряда дают естественный способ конечной аппроксимации.

Однако некоторые свойства функции непросто улавливаются по этим коэффициентам. Хороший пример тому — ограниченность. Как должны быть выбраны коэффициенты степенного ряда, чтобы получающаяся функция в единичном круге не превосходила бы по модулю единицы? Ответ на этот вопрос был дан в 1911 г. Каратеодори и Фейером [17].

Ограниченность является важным свойством аналитических функций ввиду связи между факторизацией ограниченных аналитических функций и инвариантными подпространствами — связи, которая впервые была обнаружена в фундаментальных статьях Бёрлинга [3] и Лившица и Потапова [24]. Интерполяционная теория ограниченных аналитических функций в контексте инвариантных подпространств была развита Сарасоном [40]. Она была трансформирована Сёкефальви-Надем и Фояшем в их формулировку теории инвариантных подпространств в работах [45, 46], где была обобщена на случай степенных рядов, коэффициенты которых являются операторами в гильбертовом пространстве.

Связанная с упомянутой, но отличная от неё теория инвариантных подпространств была предложена Джеймсом Ровняком и автором [16]. Эта теория доставляет естественный язык для формулировки и доказательства теоремы Каратеодори — Фейера [7]. Нам понадобятся некоторые предварительные сведения из теории квадратично-суммируемых степенных рядов [15]. Обозначим через $\mathscr{C}(z)$ гильбертово пространство квадратично-суммируемых степенных рядов $f(z) = \sum a_n z^n$,

$$\| f(z) \|^2_{\mathscr{C}(z)} = \sum | a_n |^2.$$

Степенной ряд $B(z)$ представляет функцию, не превосходящую по модулю единицы в единичном круге, тогда и только тогда, когда умножение на $B(z)$ является сжимающим преобразованием[1] в $\mathscr{C}(z)$.

Прямое использование свойства сжимаемости этого преобразования приводит к понятию дополнительности (комплементарности). Это — обобщение понятия ортогональности, оно было введено Джеймсом Ровняком и автором [16].

Если гильбертово пространство $\mathscr{P}$ содержится сжимающим образом в гильбертовом пространстве $\mathscr{H}$, то существует единственное гильбертово пространство $\mathscr{Q}$, которое сжимающим образом содержится в $\mathscr{H}$ и обладает следующими свойствами: если $c = a + b$, где a — из $\mathscr{P}$, а b — из $\mathscr{Q}$, то выполняется неравенство

$$\| c \|^2_{\mathscr{H}} \leqslant \| a \|^2_{\mathscr{P}} + \| b \|^2_{\mathscr{Q}};$$

каждый элемент c из $\mathscr{H}$ допускает разложение, для которого имеет место равенство.

Гильбертово пространство $\mathscr{Q}$ называется дополнительным (комплементарным) к пространству $\mathscr{P}$ в $\mathscr{H}$. Минимальное разложение является единственным. Элемент a получается из c действием оператора, сопряжённого с вложением $\mathscr{P}$ в $\mathscr{H}$. Элемент b получается из c действием оператора, сопряжённого с вложением $\mathscr{Q}$ в $\mathscr{H}$.

Если $B(z)$ — степенной ряд, который представляет функцию, не превосходящую единицы в единичном круге, то образ $\mathscr{M}(B)$ оператора умножения на $B(z)$ в $\mathscr{C}(z)$ рассматривается как гильбертово пространство с той единственной нормой, для которой умножение на $B(z)$ является частичной изометрией пространства $\mathscr{C}(z)$ на $\mathscr{M}(B)$. Пространство $\mathscr{H}(B)$, дополнительное к пространству $\mathscr{M}(B)$ в $\mathscr{C}(z)$, инвариантно относительно обратного сдвига[1], переводящего $f(z)$ в $[f(z) - f(0)]/z$. Справедливо неравенство

$$\| [f(z) - f(0)]/z \|^2_{\mathscr{H}(B)} \leqslant \| f(z) \|^2_{\mathscr{H}(B)} - | f(0) |^2.$$

Если $B'(z) = zB(z)$, то существует пространство $\mathscr{H}(B')$ и пространство $\mathscr{H}(B)$ сжимающим образом вложено в $\mathscr{H}(B')$. Пространство $\mathscr{B}(B)$, дополнительное к $\mathscr{H}(B)$ в $\mathscr{H}(B')$, является гильбертовым пространством размерности 0 или 1. Рассмотрим комплексные числа $\mathscr{C}$ как гильбертово пространство с модулем (абсолютной величиной) в качестве нормы. Тогда умножение на $B(z)$ является частичной изометрией пространства $\mathscr{C}$ на $\mathscr{B}(B)$.

Теорема Каратеодори — Фейера может быть сформулирована в рамках вышеописанной структуры. Пусть задано неотрицательное целое число r. Степенные ряды $f(z)$ и $g(z)$ будем называть r-эквивалентными, если коэффициенты при z^n у $f(z)$ равны коэффициентам при z^n у $g(z)$ для $n = 0, \ldots, r - 1$. Определим $\mathscr{C}_r(z)$ как конечномерное гильбертово пространство классов r-эквивалентных степенных рядов $f(z) = \sum a_n z^n$,

$$\| f(z) \|^2_{\mathscr{C}_r(z)} = \sum_{n=0}^{r-1} | a_n |^2.$$

Если $B(z)$ — степенной ряд, представляющий функцию, не превосходящую единицы в единичном круге, то оператор умножения на $B(z)$ является сжимающим в $\mathscr{C}_r(z)$.

Современная формулировка теоремы Каратеодори — Фейера утверждает, что это свойство характеризует класс r-эквивалентных степенных рядов, представляющих функции, не превосходящие единицы в единичном круге. Всякий степенной ряд $B(z)$, такой что умножение на $B(z)$ является сжимающим в $\mathscr{C}_r(z)$, r-эквивалентен степенному ряду, который не превосходит единицы в единичном круге.

Доказательство теоремы проводится индуктивным построением коэффициентов. Предположим, что r — положительное число и $A(z)$ — степенной ряд, такой что умножение на $A(z)$ является сжимающим в $\mathscr{C}_{r-1}(z)$. Тогда существует степенной ряд $B(z)$, r-эквивалентный $A(z)$ и такой, что умножение на $B(z)$ является сжимающим в $\mathscr{C}_r(z)$.

Шаг индукции производится путём вычисления соотношения между ассоциированными пространствами. Определим пространство $\mathscr{M}_{r-1}(A)$ как образ оператора умножения на $A(z)$

---

[1] У нас принят и термин «нерастягивающее преобразование». — *Прим. ред.*

[1] Автор использует термин «difference-quotient transformation». — *Прим. перев.*

в $\mathscr{C}_{r-1}(z)$. Оно рассматривается как гильбертово пространство с той единственной нормой, при которой умножение на $A(z)$ является частичной изометрией из $\mathscr{C}_{r-1}(z)$ на $\mathscr{M}_{r-1}(A)$. Тогда пространство $\mathscr{M}_{r-1}(A)$ сжимающим образом содержится в пространстве $\mathscr{C}_{r-1}(z)$. Определим $\mathscr{H}_{r-1}(A)$ как пространство, дополнительное к пространству $\mathscr{M}_{r-1}(A)$ в $\mathscr{C}_{r-1}(z)$.

Пусть $A'(z)=zA(z)$. Тогда умножение на $A'(z)$ является сжимающим преобразованием в $\mathscr{C}_r(z)$. Определим $\mathscr{M}_r(A')$ как образ оператора умножения на $A'(z)$ в $\mathscr{C}_r(z)$. Пространство $\mathscr{M}_r(A')$ рассматривается с той единственной нормой, при которой умножение на $A'(z)$ является частичной изометрией из $\mathscr{C}_r(z)$ на $\mathscr{M}_r(A')$. Пространство $\mathscr{M}_r(A')$ содержится сжимающим образом в $\mathscr{C}_r(z)$. Определим $\mathscr{H}_r(A')$ как пространство, дополнительное к пространству $\mathscr{M}_r(A')$ в $\mathscr{C}_r(z)$. Тогда $\mathscr{H}_r(A')$ — это множество классов r-эквивалентных степенных рядов $f(z)$, таких что $[f(z)-f(0)]/z$ принадлежит $\mathscr{H}_{r-1}(A)$. Для каждого элемента $f(z)$ из $\mathscr{H}_r(A')$ справедливо тождество

$$\|[f(z)-f(0)]/z\|^2_{\mathscr{H}_{r-1}(A)}=\|f(z)\|^2_{\mathscr{H}_r(A')}-|f(0)|^2.$$

Искомый степенной ряд $B(z)$ строится следующим образом. Мы строим любое гильбертово пространство $\mathscr{H}_r$, сжимающим образом содержащееся в $\mathscr{H}_r(A')$, такое что отображение, переводящее каждый элемент пространства $\mathscr{H}_r$ в его класс $(r-1)$-эквивалентности, является частичной изометрией пространства $\mathscr{H}_r$ на $\mathscr{H}_{r-1}(A)$. Пусть $\mathscr{B}_r$ — пространство, дополнительное к пространству $\mathscr{H}_r$ в $\mathscr{H}_r(A')$. Тогда $\mathscr{B}_r$ содержит некоторый степенной ряд $B(z)$ нулевой нормы, либо ряд, который $(r-1)$-эквивалентен $A(z)$. Умножение на $B(z)$ является сжимающим в $\mathscr{C}_r(z)$.

Эта формулировка теоремы Каратеодори — Фейера применима также и к степенным рядам, коэффициенты которых являются операторами в гильбертовом пространстве. Её можно использовать для получения информации о связи между факторизацией и инвариантными подпространствами [14].

Связь между теорией инвариантных подпространств Сёкефальви-Надя — Фояша и данной теорией квадратично-суммируемых рядов обсуждалась Никольским и Васюниным [33] и Сарасоном [41]. Эта связь может быть также сформулирована на языке теории систем [12].

Доказательство теоремы Каратеодори — Фейера вне рамок представленной здесь теории квадратично-суммируемых рядов было получено Розенблюмом и Ровняком [39]. Смотри также книгу Н. К. Никольского [32], где дана иная современная трактовка интерполяционной теории.

Неванлинна [28—31] ввёл в интерполяционную теорию индефинитное обобщение понятия гильбертова пространства. На эти пространства тоже можно обобщить понятие дополнительности [13].

Альтернативой методу Каратеодори — Фейера в интерполяционной теории ограниченных функций служит метод, основанный на алгоритме Шура. Элпей и Дим [1] показали, что в результате применения этой теории получаются аналогичные гильбертовы пространства аналитических функций.

ОГРАНИЧЕННЫЕ ОДНОЛИСТНЫЕ ФУНКЦИИ

Интерполяционная задача для ограниченных аналитических функций приобретает дополнительный интерес, если от функции требуется однолистность. Дело в том, что это условие совместимо с интерполяционной задачей для ограниченных функций.

Гильбертово пространство, подходящее для исследования ограниченных однолистных функций, — это множество степенных рядов

$$f(z)=\sum_{n=1}^{\infty} a_n z^n,$$

у которых нулевой коэффициент равен нулю и конечна норма Дирихле

$$\|f(z)\|^2_{\mathscr{D}}=\sum_{n=1}^{\infty} n|a_n|^2.$$

Элементами этого пространства служат степенные ряды, сходящиеся в единичном круге. Норма Дирихле важна потому, что её можно вычислить, интегрируя квадрат производной по единичному кругу:

$$\|f(z)\|^2_{\mathscr{D}}=\frac{1}{\pi}\iint\limits_{\Delta}|f'(z)|^2\,dx\,dy.$$

Если $B(z)$ — нормированная однолистная функция для подобласти единичного круга, то произведя замену переменных в кратном интеграле, получим

$$\|f(B(z))\|^2_{\mathscr{D}}=\frac{1}{\pi}\iint\limits_{\Delta}|f'(B(z))|^2|B'(z)|^2\,dx\,dy=$$
$$=\frac{1}{\pi}\iint\limits_{B(\Delta)}|f'(z)|^2\,dx\,dy,$$

поскольку $|B'(z)|^2$ совпадает с якобианом данного отображения в силу уравнений Коши — Римана. Так как по предположению образ $B(\Delta)$ единичного круга содержится в единичном круге Δ, выполняется неравенство

$$\iint\limits_{B(\Delta)} |f'(z)|^2 dx\, dy \leqslant \iint\limits_{\Delta} |f'(z)|^2 dx\, dy.$$

Следовательно, для любого элемента $f(z)$ пространства $\mathscr{G}$ справедливо неравенство

$$\| f(B(z)) \|^2_{\mathscr{G}} \leqslant \| f(z) \|^2_{\mathscr{G}}.$$

Это неравенство утверждает, что подстановка функции $B(z)$, т. е. преобразование, переводящее $f(z)$ в $f(B(z))$, является сжимающим преобразованием в норме Дирихле. Это наблюдение приводит к нескольким различным конструкциям, предваряющим доказательство гипотезы Бибербаха [5, 6, 8], которые и будут сейчас представлены.

Рассмотрим образ $\mathscr{N}(B)$ подстановки функции $B(z)$ как гильбертово пространство с той единственной нормой, при которой эта подстановка является изометрией из $\mathscr{G}$ на $\mathscr{N}(B)$. Тогда пространство $\mathscr{N}(B)$ сжимающим образом содержится в $\mathscr{G}$. Определим $\mathscr{G}(B)$ как пространство, дополнительное к пространству $\mathscr{N}(B)$ в $\mathscr{G}$. Тогда $\mathscr{G}(B)$ — гильбертово пространство с воспроизводящим ядром, которое поддается явному вычислению.

Действительно, функция

$$\log \frac{1}{1 - z\bar{w}} = z\bar{w} + \frac{1}{2}(z\bar{w})^2 + \frac{1}{3}(z\bar{w})^3 + \ldots$$

является воспроизводящим ядром пространства $\mathscr{G}$, функция

$$\log \frac{1}{1 - B(z)\overline{B(w)}}$$

— воспроизводящим ядром пространства $\mathscr{N}(B)$, а воспроизводящим ядром пространства $\mathscr{G}(B)$ будет функция

$$\log \frac{1 - B(z)\overline{B(w)}}{1 - z\bar{w}} = \log \frac{1}{1 - z\bar{w}} - \log \frac{1}{1 - B(z)\overline{B(w)}}.$$

Стоит сравнить это с вычислением воспроизводящего ядра для пространства $\mathscr{H}(B)$. Функция

$$\frac{1}{1 - z\bar{w}}$$

является воспроизводящим ядром пространства $\mathscr{C}(z)$, функция

$$\frac{B(z)\overline{B(w)}}{1 - z\bar{w}}$$

— воспроизводящим ядром пространства $\mathscr{M}(B)$, а воспроизводящим ядром пространства $\mathscr{H}(B)$ будет функция

$$\frac{1 - B(z)\overline{B(w)}}{1 - z\bar{w}} = \frac{1}{1 - z\bar{w}} - \frac{B(z)\overline{B(w)}}{1 - z\bar{w}}.$$

Заметим, что воспроизводящее ядро пространства $\mathscr{H}(B)$ является экспонентой от воспроизводящего ядра пространства $\mathscr{G}(B)$. Поэтому из теории воспроизводящих ядер вытекает следующее неравенство. Если f принадлежит $\mathscr{G}(B)$, то $\exp f(z)$ принадлежит $\mathscr{H}(B)$ и

$$\| \exp f(z) \|^2_{\mathscr{H}(B)} \leqslant \exp \| f(z) \|^2_{\mathscr{G}(B)}.$$

Равенство здесь имеет место тогда, когда $f(z)$ — воспроизводящее ядро пространства $\mathscr{G}(B)$, отвечающее некоторой точке единичного круга. Если $f(z)$ — элемент пространства $\mathscr{G}(B)$, для которого имеет место равенство, то справедливо тождество

$$\langle \exp f(z), \exp g(z) \rangle_{\mathscr{H}(B)} = \exp \langle f(z), g(z) \rangle_{\mathscr{G}(B)}$$

для любого элемента $g(z)$ из $\mathscr{G}(B)$.

Эти соотношения подтверждают, что норма Дирихле является подходящей для оценки однолистных функций. Далее, существует обобщение нормы Дирихле на случай рядов Лорана. Возникающие при этом квадратичные формы индефинитны, однако аналогичные методы всё ещё применимы. Действительно, можно рассмотреть обобщённые степенные ряды, у которых показатели не являются целыми.

Пусть ν — заданное вещественное число. Определим $\mathscr{G}^{\nu}$ как множество обобщённых степенных рядов

$$f(z) = \sum_{n=1}^{\infty} a_n z^{\nu+n},$$

таких что конечно выражение

$$\langle f(z), f(z) \rangle_{\mathscr{G}^{\nu}} = \sum_{n=1}^{\infty} (\nu + n) | a_n |^2.$$

Отметим, что квадратов с отрицательными коэффициентами возникает лишь конечное число, поэтому бесконечная сумма имеет смысл.

В случае когда $B(z)$ — нормированная однолистная функция, отвечающая подобласти единичного круга, ряд $f(B(z))$, получающийся после подстановки, допускает естественную интерпретацию как элемент $\mathcal{G}^\nu$, если $f(z)$ принадлежит $\mathcal{G}^\nu$, и справедливо неравенство

$$\langle f(B(z)),\ f(B(z))\rangle_{\mathcal{G}^\nu} \leqslant \langle f(z),\ f(z)\rangle_{\mathcal{G}^\nu}.$$

Это неравенство имеет долгую историю. При $\nu = -2$ оно представляет собой усиление теоремы площадей, открытой Гронуоллом [22] и использованной Бибербахом [4] при доказательстве его гипотезы для второго коэффициента. Теорема площадей — предельный случай приведённого неравенства, применяемого к неограниченным однолистным функциям.

В 1939 г. Грунский [23] получил обобщение теоремы площадей. Снова результат был сформулирован для неограниченных нормированных однолистных функций. Это предельный случай представленного неравенства, когда ν является отрицательным целым числом. Неравенство Грунского сыграло важную роль в доказательстве гипотезы Бибербаха для четвёртого, пятого и шестого коэффициентов. Другая причина интереса к неравенству Грунского — то, что оно является характеристическим для нормированных однолистных функций.

Послевоенное поколение аналитиков нашло более сильную форму неравенства, применимого к нормированным однолистным функциям, отвечающим подобластям единичного круга. В действительности неравенство характеризует эти функции. Эта информация была использована Тамми [47, 48] для описания первых четырёх коэффициентов ограниченной однолистной функции. Указанное неравенство эквивалентно сформулированному выше для неотрицательных целых ν, но записывается в другой форме. То наблюдение, что нормированные однолистные функции, отвечающие подобластям единичного круга, характеризуются свойством сжимаемости оператора подстановки относительно индефинитного скалярного произведения, было сделано автором [15].

В этих результатах пространство $\mathcal{G}^\nu$ используется лишь при целых значениях ν. Открытие того, что вариант теоремы площадей справедлив и для нецелых ν, принадлежит Правитцу [37].

Другой важный момент в доказательстве гипотезы Бибербаха — рассмотрение сопряжённых операторов. Всякое ограниченное линейное преобразование гильбертова пространства в себя обладает сопряжённым с ним преобразованием. Если исходное преобразование является в некотором смысле вычислимым и норма в гильбертовом пространстве «хорошо» с ним соотносится, то можно ожидать, что и сопряжённое преобразование может быть вычислено аналогичным образом.

Эти ожидания оправдываются для операторов подстановки при использовании нормы Дирихле и её индефинитных обобщений. Рассмотрим также пространства $\mathcal{G}^\nu$ и $\mathcal{G}^\mu$, для которых $\mu + \nu + 1$ является целым отрицательным числом. Для $g(z)$, принадлежащих $\mathcal{G}^\mu$, обозначим через $P_\nu g(1/z)$ тот единственный элемент $f(z)$ из $\mathcal{G}^\nu$, для которого $f(z) = \sum a_n z^{\nu+n}$ и $g(z) = \sum b_n z^{\mu+n}$, где $a_m = b_n$, если $\mu + \nu + m + n = 0$ для положительных чисел m и n, и $a_m = 0$ в противном случае. Если $f(z)$ принадлежит $\mathcal{G}^\nu$ и $g(z)$ принадлежит $\mathcal{G}^\mu$, то справедливо тождество

$$\langle f(z),\ P_\nu g(1/z)\rangle_{\mathcal{G}^\nu} = -\langle P_\mu f(1/z),\ g(z)\rangle_{\mathcal{G}^\mu}.$$

Предположим, что $B(z) = \sum B_n z^n$ — нормированная однолистная функция, отвечающая подобласти единичного круга. Тогда равенство $B^*(z) = \sum \overline{B}_n z^n$ задаёт нормированную однолистную функцию, также отвечающую некоторой подобласти единичного круга. Область, на которую $B^*(z)$ отображает единичный круг, получается отражением относительно вещественной оси той области, на которую единичный круг отображает функция $B(z)$.

Если $f(z)$ — элемент пространства $\mathcal{G}^\nu$, а $g(z)$ — элемент пространства $\mathcal{G}^\mu$, то справедливо тождество

$$\langle f(z),\ P_\nu(1/z)\rangle_{\mathcal{G}^\nu} = \langle f(B(z)),\ P_\nu g(B^*(1/z))\rangle_{\mathcal{G}^\nu}.$$

Оно доказывается с помощью формулы Коши.

В качестве следствия из этого тождества можно вычислить оператор, сопряжённый с оператором подстановки функции $B(z)$, по крайней мере на тех элементах из $\mathcal{G}^\nu$, которые имеют вид $P_\nu g(B^*(1/z))$, где $g(z)$ принадлежит $\mathcal{G}^\mu$. Применение этого сопряжённого к такому элементу из $\mathcal{G}^\nu$ даёт $P_\nu g(1/z)$. Таким образом, подстановке функции $B(z)$ в $\mathcal{G}^\nu$ соответствует оператор, обратный к подстановке функции $B^*(z)$ в $\mathcal{G}^\mu$. В силу произвольности μ этот результат задаёт действие сопряжённого с подстановкой $B(z)$ на плотном множестве элементов из $\mathcal{G}^\nu$.

Эти рассмотрения позволяют по-новому построить теорию однолистных функций. Интересно сравнить принятый здесь подход с подходом Шура [42—44], одного из основателей теории операторов, который провёл соответствующие вычисления с формальными степенными рядами и индефинитными скалярными произведениями. Для описания происходящего у Шура не было в распоряжении методов теории квадратично-суммируемых степенных рядов. Это концептуальное преимущество оказалось весьма существенным при доказательстве гипотезы Бибербаха.

ЛОКАЛИЗАЦИЯ ОЦЕНОК

Аппарат, который был построен для оценивания коэффициентов однолистных функций, страдает одним недостатком, а именно, в оценках используются все коэффициенты степенного ряда. По аналогии с теоремой Каратеодори — Фейера интересно найти соответствующие оценки, которые зависят лишь от первых r коэффициентов, для заданного положительного целого числа r. Однако простое усечение уже не даёт наилучшей возможной оценки. Доказательство гипотезы Бибербаха использует более тонкую технику локализации.

Источником новых оценок служит теория, с помощью которой К. Лёвнер доказал гипотезу Бибербаха для третьего коэффициента.

Ключевое понятие теории Лёвнера — подчиненность. Говорят, что степенной ряд $f(z)$ с нулевым свободным членом подчинён степенному ряду $g(z)$ с нулевым свободным членом, если $f(z) = g(B(z))$ для некоторого степенного ряда $B(z)$ с нулевым свободным членом, представляющего функцию, ограниченную единицей в единичном круге. Если $f(z)$ и $g(z)$ — однолистные функции, то равносильным будет условие, что область, на которую $f(z)$ отображает единичный круг, содержится в области, на которую отображает единичный круг функция $g(z)$. Полученный таким образом степенный ряд $B(z)$ является однолистной функцией, нормированной, если нормированными были $f(z)$ и $g(z)$.

Семейство Лёвнера однолистных функций — это максимальное семейство однолистных функций, которое является линейно упорядоченным в смысле подчинённости. Такое семейство обладает естественной параметризацией. Параметром служит коэффициент при z в степенном ряде. Он пробегает всё множество положительных чисел.

Лёвнер сделал интересное наблюдение. Он заметил, что всякое семейство Лёвнера однолистных функций $F(t, z)$ удовлетворяет дифференциальному уравнению

$$t\frac{\partial}{\partial t} F(t, z) = \varphi(t, z)\, z\frac{\partial}{\partial z} F(t, z).$$

Это уравнение утверждает, что производная по времени пропорциональна пространственной производной. Коэффициент пропорциональности $\varphi(t, z)$ является функцией Герглотца, т. е. степенным рядом со свободным членом, равным единице, представляющим функцию с положительной вещественной частью в единичном круге. Семейство функций Герглотца измеримо в том смысле, что коэффициент $\varphi(t, z)$ при z^n — измеримая функция от t для любого неотрицательного целого числа n. Производная по времени в уравнении Лёвнера берётся в смысле абсолютной непрерывности коэффициентов.

Верно и обратное. Предположим, что задано семейство функций Герглотца $\varphi(t, z)$. Тогда существует единственное семейство Лёвнера однолистных функций, для которого заданное семейство $\varphi(t, z)$ служит коэффициентом в уравнении Лёвнера.

В принципе всю информацию об однолистных функциях можно получить, проводя вычисления, основанные на уравнении Левнера. Однако если перед нами стоит конкретная задача, как например гипотеза Бибербаха, эти вычисления могут оказаться очень сложными. Необходима удобная форма записи, чтобы проследить зависимость оценок от времени. Естественным инструментом здесь являются квадратичные формы, «хорошо» зависящие от времени. Развитие во времени должно либо сохранять квадратичные формы, либо быть «диссипативным».

Сама гипотеза Бибербаха не является квадратичной оценкой, однако соответствующая квадратичная оценка была предложена в 1936 г. Робертсоном [38]. В формулировке этой гипотезы используется конструкция степеней однолистных функций, которая может быть проведена для любого вещественного показателя ν. Если $F(z)$ — нормированная однолистная функция, то формальное выражение $F(z)^\nu$ можно разложить в ряд с первым членом $F'(0)^\nu z^\nu$. Оставшиеся члены удобно записать следующим образом:

$$\frac{F(z)^\nu - F'(0)^\nu z^\nu}{\nu} = \sum_{n=1}^{\infty} a_n z^{\nu+n}.$$

В предельном случае $\nu = 0$ левая часть по непрерывности интерпретируется как

$$\log\frac{F(z)}{zF'(0)}.$$

Когда ν не является целым, смысл выражению придаётся делением каждой части уравнения на z^ν, после чего выбираем ту единственную ν-ю степень от $F(z)/z$, которая является степенным рядом с положительным свободным членом.

В этих обозначениях гипотеза Робертсона утверждает, что при $\nu = 1/2$ неравенство

$$|a_1|^2 + \ldots + |a_r|^2 \leqslant 4rF'(0)$$

справедливо для любых целых положительных r. Легко видеть, что, когда $F(z)$ пропорциональна функции Кёбе, имеет место равенство, и Робертсон высказал предположение, что равенство имеет место только в этом случае. Элементарная оценка показывает, что из этого неравенства вытекает гипотеза Бибербаха для $(r+1)$-го коэффициента. Робертсон проверил те случаи своей гипотезы, которые соответствуют гипотезе Бибербаха для второго и третьего коэффициентов. Таким образом, он показал, что все известные к тому времени свидетельства в пользу справедливости гипотезы Бибербаха свидетельствуют и в пользу справедливости более сильной гипотезы Робертсона. Метод его доказательства — уравнение Лёвнера. В ретроспективе эта комбинация квадратичных форм с уравнением Лёвнера представляется замечательным продвижением в теории оценивания однолистных функций. Уже в то время можно было бы получить больше информации о гипотезе Бибербаха, если бы этот метод был развит более глубоко. Однако тогда не было приложено необходимых усилий, поскольку не было достаточной веры в справедливость гипотезы Бибербаха. Позднее же, когда были получены новые свидетельства в её пользу, внимание исследователей было отвлечено от уравнения Лёвнера, так как новые результаты были получены другими методами.

Возвращение к квадратичным идеям Робертсона произошло в 1965 г., когда Лебедев и Милин [26] заметили, что в случае $\nu=0$ из неравенства

$$\sum_{n=1}^{r} n(r+1-n)|a_n|^2 \leqslant 4(r+1)\sum_{n=1}^{r}\frac{1}{n+1}$$

вытекает гипотеза Бибербаха для $(r+1)$-го коэффициента. На самом деле они показали, что из этого неравенства следует соответствующий случай гипотезы Робертсона. Гипотеза Милина утверждает, что неравенство выполняется всегда, а равенство имеет место только для функций, пропорциональных функции Кёбе.

Гипотезы Робертсона и Милина подсказывают конструкцию новых пространств, элементами которых являются обобщённые степенные ряды $f(z)=\sum\limits_{n=1}^{\infty} a_n z^{\nu+n}$ (точнее, классы эквивалентных рядов). Пусть σ_n — заданная вещественнозначния функция натурального аргумента n, которая начиная с некоторого места становится нулём. Положим

$$\langle f(z), f(z)\rangle_{\mathscr{F}_\sigma^\nu}=\sum_{n=1}^{\infty}(\nu+n)\,\sigma_n|a_n|^2.$$

Эквивалентность двух таких обобщённых степенных рядов $f(z)$ и $g(z)$ означает здесь, что коэффициент $f(z)$ при $z^{\nu+n}$ равен коэффициенту $g(z)$ при $z^{\nu+n}$, когда $(\nu+n)\sigma_n$ не равно нулю.

Теперь гипотезы Робертсона и Милина — оценки одного и того же вида: задача состоит в том, чтобы оценить выражение

$$\left\langle\frac{F(z)^\nu-F'(0)^\nu z^\nu}{\nu},\ \frac{F(z)^\nu-F'(0)^\nu z^\nu}{\nu}\right\rangle_{\mathscr{F}_\sigma^\nu},$$

где $F(z)$ — нормированная однолистная функция. Вопрос в том, достигает ли это выражение своего максимума при заданном значении $F'(0)$, когда $F(z)$ пропорциональна функции Кёбе?

Можно ожидать, что ответ зависит от выбора σ, и двумерные примеры показывают, что так оно и есть. Поэтому вопрос состоит в том, при каких весах проблема оптимизации имеет ожидаемое решение.

Вышеприведённая характеризация ограниченных однолистных функций предлагает и ответ. Можно ожидать, что искомая теория оценивания будет предельным случаем теории оценивания для однолистных функций, отвечающих подобластям единичного круга. Той характеристикой функции σ, которая делает её функцией с желаемыми свойствами, является свойство сжимаемости соответствующих операторов подстановки.

Но ответ усложняется неожиданной подкруткой, у которой нет аналога в теории Каратеодори — Фейера. Это означает, что весам σ надо разрешить изменяться в процессе подстановки. Свойство сжимаемости преобразования получается при переходе от одного пространства $\mathscr{G}_\sigma^\nu$ к другому. Необходимость провести такие изменения веса и точный способ, каким это следует, делать — одни из самых тонких моментов в доказательстве гипотезы Бибербаха.

Предположим, что 2ν не является целым отрицательным числом. Говорят, что семейство пространств $\mathscr{G}_{\sigma(t)}^\nu$, $t\geqslant 1$, является допустимым, если $(\nu+n)\sigma_n(t)$ — абсолютно непрерывная функция от t, которая не возрастает при $2\nu+n>0$ и не убывает при $2\nu+n<0$, и если для всех целых положительных n удовлетворяется дифференциальное уравнение

$$\sigma_n(t)+\frac{t\sigma_n'(t)}{2\nu+n}=\sigma_{n+1}(t)-\frac{t\sigma_{n+1}'(t)}{n+1}.$$

Рассматриваются решения, для которых $\sigma_n(t)$ начиная с некоторого номера тождественно равны нулю.

Эти условия аккуратно подбирались так, чтобы неравенство

$$\langle f(B(z)), f(B(z))\rangle_{\mathscr{F}_{\sigma(a)}^\nu}\leqslant\langle f(z), f(z)\rangle_{\mathscr{F}_{\sigma(b)}^\nu}$$

выполнялось для любого элемента f из $\mathscr{G}^{\nu}_{\sigma(b)}$, если B — нормированная однолистная функция, отвечающая подобласти единичного круга, и $1 \leqslant a = bB'(0)$. Вывод неравенства из сформулированных предположений осуществляется непосредственным применением дифференциального уравнения Лёвнера [11, 15].

После того как установлен этот результат, возникает вопрос, можно ли получить оценку требуемого выражения

$$\left\langle \frac{F(z)^{\nu} - F'(0)^{\nu} z^{\nu}}{\nu}, \frac{F(z)^{\nu} - F'(0)^{\nu} z^{\nu}}{\nu} \right\rangle_{\mathscr{G}^{\nu}_{\sigma(a)}},$$

когда $F(z)$ — нормированная однолистная функция. Мы хотим, чтобы максимум этого выражения достигался, когда $F(z)$ пропорциональна функции Кёбе, и равнялся

$$4F'(0)^{2\nu} \sum_{n=1}^{\infty} \frac{\Gamma(2\nu+n)^2}{\Gamma(2\nu+1)^2 \Gamma(n+1)^2} (\nu+n) \sigma_n(a).$$

Достаточно получить оценку для ограниченных функций. В этом случае можно ожидать более сильного неравенства. Действительно, опять применяя уравнение Лёвнера, приходим к неравенству

$$\left\langle \frac{B(z)^{\nu} - B'(0)^{\nu} z^{\nu}}{\nu} + f(B(z)), \frac{B(z)^{\nu} - B'(0)^{\nu} z^{\nu}}{\nu} + f(B(z)) \right\rangle_{\mathscr{G}^{\nu}_{\sigma(a)}}$$

$$\leqslant \langle f(z), f(z) \rangle_{\mathscr{G}^{\nu}_{\sigma(b)}} + \sum_{n=1}^{\infty} \frac{\Gamma(2\nu+n)^2}{\Gamma(2\nu+1)^2 \Gamma(n+1)^2}$$

$$\times (\nu+n) [a^{2\nu}\sigma_n(a) - b^{2\nu}\sigma_n(b)],$$

справедливому для любого элемента $f(z)$ из $\mathscr{G}^{\nu}_{\sigma(b)}$, если B — нормированная однолистная функция, отвечающая подобласти единичного круга, и $1 \leqslant \alpha = bB'(0)$. Далее, из доказательства видно, что равенство имеет место в том и только в том случае, когда существует комплексное число ω, равное по модулю единице, такое что

$$\frac{B(z)}{(1+\omega B(z))^2} = \frac{B'(0)z}{(1+\omega z)^2}$$

и коэффициенты $f(z)$ при $z^{\nu+n}$ равны коэффициентам при $z^{\nu+n}$ функции

$$\frac{z^{\nu}}{\nu(1+\omega z)^{2\nu}} - \frac{z^{\nu}}{\nu},$$

когда $\sigma_n(t)$ не равно тождественно нулю.

Таким образом случаи равенства связаны с функцией Кёбе. Желаемая оценка для неограниченных однолистных функций получается предельным переходом по большим b. Случай равенства имеет место только тогда, когда однолистная функция пропорциональна функции Кёбе.

Эти рассмотрения дают нам желаемые оценки однолистных функций, требуя для этого информацию о существовании монотонных решений дифференциальных уравнений.

Решение системы дифференциальных уравнений для весовых функций записывается в виде

$$\sigma_n(t) = \Delta_{2\nu+n}(t) + \frac{(2\nu+n)(2\nu+2n+2)}{(-1)(n+1)} \Delta_{2\nu+n+1}(t) +$$

$$+ \frac{(2\nu+n)(2\nu+n+1)(2\nu+2n+3)(2\nu+2n+4)}{(-1)(-2)(n+1)(n+2)} \Delta_{2\nu+n+2}(t) + \ldots,$$

где $\Delta_{2\nu+n}(t)$ — решение элементарного дифференциального уравнения

$$t\Delta'_{2\nu+n}(t) = -(2\nu+n)\Delta_{2\nu+n}(t).$$

Решением служит

$$\Delta_{2\nu+n}(t) = \Delta_{2\nu+n}(1) t^{-2\nu-n}.$$

Примеры решений можно дать в терминах гипергеометрических рядов. Для гипергеометрического ряда используется обозначение

$$F(a, b; c; z) = 1 + \frac{ab}{1 \cdot c} z + \frac{a(a+1)b(b+1)}{1 \cdot 2c(c+1)} z^2 + \ldots.$$

Обобщением гипергеометрического ряда является ряд

$$F(a, b, c; d, e; z) = 1 + \frac{abc}{1 \cdot de} z + \frac{a(a+1)b(b+1)c(c+1)}{1 \cdot 2d(d+1)e(e+1)} z^2 + \ldots$$

Принадлежащее Клаузену тождество утверждает, что

$$F(a, b; c, z)^2 = F(2a, 2b, a+b; c, 2c-1; z),$$

когда $a + b + 1/2 = c$.

При построении решений дифференциальных уравнений для весовых функций будет использован неотрицательный параметр λ. Пусть r — заданное положительное целое число. Выберем

$$\Delta_{2\nu+n}(1) = \frac{4^{-n}\Gamma(n+1)\Gamma(2\nu+2\lambda+r+n+1)}{\Gamma(\nu+n+1)\Gamma(\nu+\lambda+n+1)\Gamma(2\nu+n+1)\Gamma(r+1+n)}$$

для $n=1, \ldots, r$ и $\Delta_{2\nu+n}(1)=0$ для $n>r$. Тогда для $n=1, \ldots, r$ справедливо тождество

$$-t\sigma_n'(t)=\frac{4^{-n}\Gamma(n+1)\Gamma(2\nu+2\lambda+r+n+1)}{\Gamma(\nu+n+1)\Gamma(\nu+\lambda+n+1)\Gamma(2\nu+n)\Gamma(r+1-n)}\times$$
$$\times F(n-r,\ 2\nu+2\lambda+r+n+1,\ \nu+n+1/2;$$
$$\nu+\lambda+n+1,\ 2\nu+2n+1;\ t^{-1})t^{-2\nu-n}.$$

В силу неравенства, установленного Аски и Гаспером [2], полином

$$F(n-r,\ 2\nu+2\lambda+r+n+1,\ \nu+n+1/2;$$
$$\nu+\lambda+n+1,\ 2\nu+2n+1;\ x)$$

неотрицателен при $x\leqslant 1$, если $\nu>-3/2$. В случае $\lambda=0$ желаемое неравенство следует из тождества Клаузена. При $\lambda>0$ неравенство вытекает из разложения

$$\frac{\Gamma(2\nu+2\lambda+r+n+1)}{\Gamma(\nu+\lambda+n+1)\Gamma(r+1+n)}F(n-r,\ 2\nu+2\lambda+r+n+1,$$
$$\nu+n+1/2;\ \nu+\lambda+n+1,\ 2\nu+2n+1;\ z)$$
$$=\sum_{k=n}^{r}\frac{\Gamma(2\nu+k+n+1)}{\Gamma(\nu+n+1)\Gamma(k+1-n)}c_k$$
$$\times F(n-k,\ 2\nu+k+n+1,\ \nu+n+1/2;\ \nu+n+1,\ 2\nu+n+1;z),$$

где числа c_k неотрицательны. Действительно, $c_n=0$, когда n и r противоположной чётности, и

$$c_n=2^{2\lambda}\frac{\Gamma\left(\nu+\lambda+\frac{1}{2}r+\frac{1}{2}n+\frac{1}{2}\right)\left(\nu+n+\frac{1}{2}\right)\Gamma\left(\lambda+\frac{1}{2}r-\frac{1}{2}n\right)}{\Gamma\left(\nu+\frac{1}{2}r+\frac{1}{2}n+\frac{3}{2}\right)\Gamma(\lambda)\Gamma\left(1+\frac{1}{2}r-\frac{1}{2}n\right)}\,.$$

в случае когда n и r имеют одинаковую чётность.

Следовательно, допустимое семейство пространств получается, когда $\nu>-3/2$. Итоговые оценки сформулировать довольно сложно, за исключением случая $\lambda=1/2$. Предположим, что σ_n — произвольная вещественнозначная функция целого положительного аргумента n, такая что $\lim\limits_{n\to\infty}\sigma_n=0$ и функция

$$\rho_n=\frac{\Gamma(2\nu+n+1)}{\Gamma(n+1)}[\sigma_n-\sigma_{n+1}]$$

неотрицательна, не возрастает и стремится к нулю на бесконечности. Если $F(z)$ — нормированная однолистная функция и

$$\frac{F(z)^\nu-F'(0)^\nu z^\nu}{\nu}=\sum_{n=1}^{\infty}a_n z^{\nu+n},$$

то

$$\sum_{n=1}^{\infty}(\nu+n)\sigma_n|a_n|^2\leqslant 4F'(0)^{2\nu}\sum_{n=1}^{\infty}\frac{\Gamma(2\nu+n)^2}{\Gamma(2\nu+1)^2\Gamma(n+1)^2}(\nu+n)\sigma_n.$$

Если σ_n не равна тождественно нулю, то равенство имеет место тогда и только тогда, когда $F(z)$ пропорциональна функции Кёбе.

Этот результат в случае $\nu=0$ содержит гипотезу Милина и завершает доказательство гипотезы Бибербаха.

ЗАМЕЧАНИЯ

То, что представлено здесь как доказательство гипотезы Бибербаха, на самом деле есть общая теория оценивания однолистных функций. Гипотеза Бибербаха важна лишь как историческая веха. Она является тестом, применение которого к теории оценивания измеряет её силу. Результат интересен главным образом потому, что его трудно получить. Значение же работы скорее в созданной теории оценивания однолистных функций, чем в доказательстве самой гипотезы Бибербаха.

Предъявленное доказательство гипотезы Бибербаха обосновывает гипотезу Милина, из которой следует гипотеза Робертсона, которая в свою очередь влечёт справедливость гипотезы Бибербаха. Вполне возможно, что, исходя из той же теории оценивания, удастся найти другие пути, приводящие к доказательству гипотезы Бибербаха. Вероятно, полезными для этой цели будут свойства сжимаемости операторов подстановки в пространствах $\mathscr{G}_0^\nu$, где 2ν — целое отрицательное число. Эти случаи являются более сложными из-за наличия сингулярностей в дифференциальных уравнениях для весовых функций. Другая сложность заключается в том, что неравенства Аски — Гаспера неприменимы при $\nu<-3/2$. Для этих значений ν теорию оценивания надо строить заново.

Можно ожидать, что наступит время, когда утихнет интерес к самой гипотезе Бибербаха и снова будут рассматриваться более фундаментальные проблемы. Одной из них является интерполяционная задача для однолистных функций. Задача состоит в том, чтобы охарактеризовать первые r коэффициентов $a_1, \ldots, a_r$

нормированной однолистной функции $f(z)=\sum a_n z^n$. При проверке выполнения последовательности неравенств, которым должны удовлетворять коэффициенты, могут оказаться существенными компьютерные вычисления.

Из доказательства гипотезы Бибербаха следует, что менее традиционная задача может оказаться более фундаментальной. Такова задача описания первых r коэффициентов $a_1, \ldots, a_r$ нормированной однолистной функции $\sum a_n z^n$, отвечающей подобласти единичного круга. Если удастся охарактеризовать коэффициенты ограниченных функций, то как следствие получится описание коэффициентов неограниченных однолистных функций.

Вышеприведённая теория оценивания даёт большое количество неравенств, которым удовлетворяет нормированная однолистная функция, отвечающая подобласти единичного круга. Эти неравенства равносильны утверждениям о сжимаемости операторов подстановки в пространствах степенных рядов, снабжённых скалярным произведением. Естественно предположить, что неравенства, затрагивающие первые r коэффициенты ограниченной однолистной функции, характеризуют эти коэффициенты.

Сейчас трудно доказать эту гипотезу по причине слишком большого числа неравенств. Необходимо более глубокое проникновение в сущность проблемы, чтобы понять, какие неравенства являются важными. Выбор в качестве ν целого отрицательного числа, по-видимому, важен.

Скорее всего невозможно явно определить все решения дифференциальных уравнений для весовых функций, обладающих желаемым свойством монотонности. Однако описание коэффициентов наверняка должно учитывать все такие решения. Это и осложняет анализ, который должен быть проделан.

Можно ожидать, что главным препятствием при описании коэффициентов будет задача о продолжении типа той, что заключена в теореме Каратеодори — Фейера. Может оказаться необходимым компьютерное исследование, чтобы определить, является ли желаемое продолжение возможным при выполнении предложенных характеристических неравенств.

Наконец, необходимо исследовать связь интерполяционной задачи для ограниченных однолистных функций с задачей Каратеодори — Фейера. Как известно, существуют экспоненциальные соотношения между этими двумя теориями до локализации [15]. Необходимы экспоненциальные соотношения между локальными теоремами. С этой точки зрения нужно исследовать неравенство Лебедева — Милина [26], которое используется при доказательстве гипотезы Бибербаха. Быть может, удастся найти новое доказательство этого неравенства, которое обобщается на интересующую нас ситуацию.

Эта статья была написана в течение марта — июля 1986 г., когда автор был свободен от чтения лекций в Университете Пердью как степендиат фонда Александра фон Гумбольдта в Гейдельбергском университете. Он благодарит принимавших его профессоров А. Дольда и Э. Фрайтага.

ЛИТЕРАТУРА

1. D. Alpay and H. Dym, An application of reproducing kernel Hilbert spaces to the Schur algorithm and rational *J* unitary factorization, I. Schur Methods in Operator Theory and Signal Processing, Birkhäuser, Basel, 1986.
2. R. Askey and G. Gasper, Inequalities for polynomials, The Bieberbach Conjecture: Proceedings of the Symposium on the Occasion of the Proof, Math. Surveys Monogr., no. 21, Amer. Math. Soc., Providence, R. I., 1986, pp. 7—32.
3. A. Beurling, On two problems concerning linear transformations in Hilbert space, Acta Math. **81** (1949), 239—255.
4. L. Bieberbach, Über die Koeffizienten derjenigen Potenzreihen welche eine schlichte Abbildung des Einheitskreises vermitteln, Sitzungsbericht Preussische Akademie der Wissenschaften, 1916, pp. 940—955.
5. L. de Branges, Coefficient estimates, J. Math. Anal. Appl. **82** (1981), 420—450.
6. —, Grunsky spaces of analytic functions, Bull. Sci. Math. (2) **105** (1981), 401—416.
7. —, The Carathéodory-Fejér extension theorem, Integral Equations Operator Theory **5** (1982), 160—183.
8. —, Löwner expansions, J. Math. Anal. Appl. **100** (1984), 323—337.
9. —, A proof of the Bieberbach conjecture, Acta Math. **154** (1985), 137—152.
10. —, The story of the verification of the Bieberbach conjecture, The Bieberbach Conjecture: Proceedings of the Symposium on the Occasion of the Proof, Math. Surveys Monogr., no. 21, Amer. Math. Soc., Providence, R. I., 1986, pp. 199—203.
11. —, Powers of Riemann mapping functions, The Bieberbach Conjecture: Proceedings of the Symposium on the Occasion of the Proof, Math. Surveys Monogr., no. 21, Amer. Math. Soc., Providence, R. I., 1986, pp. 51—67.
12. —, Unitary linear systems whose transfer functions are Riemann mapping functions, Integral Equations Operator Theory (to appear).
13. —, Complementation in Krein spaces, Trans. Amer. Math. Soc. **305** (1988), 277—291.
14. —, Krein spaces of analytic functions, J. Funct. Anal. **81** (1988), 219—259.
15. —, Square summable power series, Springer-Verlag, Heidelberg (to appear).
16. L. de Branges and J. Rovnyak, Canonical models in quantum scattering theory, Perturbation Theory and its Applications in Quantum Mechanics, Wiley, New York, 1966, pp. 295—391.
17. C. Carathéodory and L. Fejér, Über den Zusammenhang der Extremen von Harmonischen Funktionen mit ihren Koffizienten und über den Picard Landauschen Satz, Rend. Circ. Mat. Palermo **32** (1911), 118—239.
18. O. M. Fomenko and G. V. Kuz'mina, The last hundred days of the Bieberbach conjecture, Math. Intelligencer **8** (1986), no. 1, 40—47.

19. P. R. Garabedian and M. Schiffer, A proof of the Bieberbach conjecture for the fourth coefficient, Arch. Rational Mech. Anal. **4** (1955), 427—465.

20. W. Gautschi, Reminiscences of my involvement in de Branges' proof of the Bieberbach conjecture, The Bieberbach Conjecture: Proceedings of the Symposium on the Occasion of the Proof, Math. Surveys Monogr., no. 21, Amer. Math. Soc., Providence, R. I., 1986, pp. 205—211.

21. Голузин Г. М. Геометрическая теория функций комплексного переменного. — М.: Наука, 1966.

22. T. H. Gronwall, Some remarks on conformal representation, Ann. of Math. **16** (1914—1915), 72—76.

23. H. Grunsky, Koeffizienten Bedingungen für schlicht abbildende meromorphe Funktionen, Math. Z. **45** (1939), 26—61.

24. Лившиц М. С., Потапов В. П. Теорема умножения характеристических матриц-функций. — ДАН СССР, 1950, т. 72, № 4, с. 625—628.

25. K. Löwner, Untersuchungen über schlichte konforme Abbildungen des Einheitskreises, Math. Ann. **89** (1923), 103—121.

26. Милин И. М. Однолистные функции и ортогональные системы. — М.: Наука, 1971.

27. —, Comments on the proof of the conjecture on logarithmic coefficients, The Bieberbach Conjecture: Proceedings of the Symposium on the Occasion of the Proof, Math. Surveys Monogr., no. 21, Amer. Math. Soc., Providence, R. I., 1986, pp. 109—112.

28. R. Nevanlinna, Erweiterung der Theorie des Hilbertschen Raumes, Comm. Sém. Math. Univ. Lund. Tome Supplémentaire (1952), 160—168.

29. —, Über metrische lineare Räume. II, Bilinearforme und Stetigkeit, Ann. Acad. Sci. Fenn. Sci. A I Math. **113** (1952).

30. —, Über metrische lineare Räume, III, Theorie der Orthogonalsysteme, Ann. Acad. Sci. Fenn. Sci. A I Math. **115** (1952).

31. —, Über metrische lineare Räume. IV, Zur Theorie der Unterräume, Ann. Acad. Sci. Fenn. Sci. A I Math. **163** (1954).

32. Никольский Н. К. Лекции об операторе сдвига. — М.: Наука, 1980.

33. N. K. Nikol'skii and V. I. Vasyunin, Notes on two functional models, The Bieberbach Conjecture: Proceedings of the Symposium on the Occasion of the Proof, Math. Surveys Monogr., no. 21, Amer. Math. Soc., Providence, R. I., 1986, pp. 113—141.

34. M. Ozawa, An elementary proof of the Bieberbach conjecture for the sixth coefficient, Kodai Math. Seminar Reports **21** (1969), 129—132.

35. R. N. Pederson, A proof of the Bieberbach conjecture for the sixth coefficient, Arch. Rational Mech. Anal. **31** (1968), 331—351.

36. R. Pederson and M. Schiffer, A proof of the Bieberbach conjecture for the fifth coefficient, Arch. Rational Mech. Anal. **45** (1972), 161—193.

37. H. Prawitz, Über Mittelwerte analytischer Funktionen, Arkiv für Mat. Astron. Fysik **20** (1927—1928), no. 6, 1—12.

38. M. S. Robertson, A remark on the odd schlicht functions, Bull. Amer. Math. Soc. **42** (1936), 366—370.

39. M. Rosenblum and J. Rovnyak, Hardy classes and operator theory, Clarendon Press, Oxford, 1985.

40. D. Sarason, Generalized interpolation in H^∞, Trans. Amer. Math. Soc. **127** (1967), 179—203.

41. —, Invariant subspaces from the Brangesian point of view, The Bieberbach Conjecture: Proceedings of the Symposium on the Occasien of the Proof, Math. Surveys Monogr., no. 21, Amer. Math. Soc., Providence, R. I., 1986, pp. 153—166.

42. I. Schur, On Faber polynomials, Amer. J. Math. **67** (1945), 33—41.

43. —, Ein Satz über quadratische Formen mit komplexen Koeffizienten, Amer. J. Math. **67** (1945), 472—480.

44. —, Identities in the theory of power series, Amer. J. Math. **69** (1947), 14—26.

45. B. Sz-Nagy and C. Foiaş, Analyse harmonique des opérateurs de l'espace de Hilbert, Masson, Paris, 1967.

46. —, On the structure of intertwining operators, Acta Sci. Math. (Szeged) **35** (1973), 225—254.

47. O. Tammi, Extremum problems for bounded univalent functions. I, Lecture Notes in Math., vol. 646, Springer-Verlag, Heidelberg, 1978.

48. —, Extremum problems for bounded univalent functions. II, Lecture Notes in Math., vol. 913, Springer-Verlag, Heidelberg, 1982.

Работы, добавленные переводчиком

49. L. de Branges, Das mathematische Erbe von Ludwig Bieberbach (1886—1982). — Address to Mathematisches Institut der Universität Basel, May 4, 1990 (in the occasion of receiving of Alexander Ostrowski Prize).

50. N. Nikolskii, V. Vasyunin, Quasi-orthogonal Hilbert space decompositions and estimates of univalent functions. I. — Proc. of Symposium on Functional Analysis and Applications, Sapporo, 1990 (Math. Reports of Sapporo University; to appear).

51. N. Nikolskii, V. Vasyunin. Quasi-orthogonal Hilbert space decompositions and estimates of univalent functions. II. — Proc. of the First USA-USSR Conference on Approximation Theory, Tampa, 1990.

52. J. Rovnyak, Coefficient estimates for Riemann mapping functions. — J. d'Anal. Math., **52** (1989), pp. 53—93.

53. J. Rovnyak, A vector extension of Loewner's differential equation. — OT: Advances and Appl., v. 43, Birkhäuser Verlag, Basel, 1990, pp. 301—308.

54. L. Weinstein, The Bieberbach Conjecture. — Preprint, 1989.

55. Васюнин В. И., Никольский Н. К. Квазиортогональные разложения по дополнительным метрикам и оценки одролистных функций. — Алгебра и анализ, 1990, т. 2, № 4, с. 1—81.

56. Васюнин В. И., Никольский Н. К. Операторные меры и коэффициенты однолистных функций. — Алгебра и анализ, 1991.

УНИВЕРСИТЕТ ПЕРДЬЮ, ЛАФАЙЕТ, ИНДИАНА 47906, США

ГЕОМЕТРИЯ ЧЕТЫРЁХМЕРНЫХ МНОГООБРАЗИЙ

Саймон К. Доналдсон[1]

I. ВВЕДЕНИЕ

Название этой статьи соответствует существу дела, поскольку, несмотря на то что результаты, которые мы будем излагать, принадлежат области дифференциальной топологии, применяемые методы — геометрические, основанные на использовании «инстантонов» или «полей Янга — Миллса», придуманных физиками. Прежде чем переходить к детальному изложению результатов и методов, мы хотели бы показать, как резко выделяются они на общем фоне топологии многообразий.

Топологическое n-мерное многообразие «собирается» из областей n-мерного эвклидова пространства, соединяемых гомеоморфизмами. Многообразие снабжено дифференциальной, или гладкой, структурой, если эти гомеоморфизмы дифференцируемы. Основное отношение эквивалентности между топологическими многообразиями — *гомеоморфизм*, между гладкими — *диффеоморфизм* (гомеоморфизм, определенный гладкими функциями). В 1960—1970-х годах топологи построили глубокую теорию многообразий в размерностях 5 и выше. Эта теория объясняет взаимосвязь между гладкой и топологической теориями [28], и для многих классов многообразий она даёт полную классификацию в терминах алгебро-топологических инвариантов [3, 26, 33]. Самые грубые среди последних — *гомотопические инварианты*, например гомологические группы. После них наиболее важными являются *классы Понтрягина* $p_i(X) \in H^{4i}(X; \mathbb{Z})$ гладкого многообразия X — характеристические классы касательного расслоения. Обратим внимание на четыре факта из теории для высоких размерностей:

(1) Односвязные гладкие многообразия размерности 5 и выше диффеоморфны, если они h-кобордантны.

(2) Гомотопический тип и классы Понтрягина компактного односвязного многообразия размерности 5 и выше определяют гладкую структуру с точностью до конечного числа возможностей.

(3) Приведения понтрягинских классов к $H^*(X, \mathbb{Q})$ являются топологическими инвариантами.

(4) Стягиваемое гладкое многообразие размерности 5 и выше имеет единственную гладкую структуру.

В больших размерностях техника h-кобордизма даёт эффективный метод для конструирования эквивалентностей между многообразиями, и классификация гладких и топологических многообразий различается лишь на «конечное число возможностей».

В 1982 г. Фридман [16] показал, что основные конструкции, используемые в больших размерностях, осуществимы для *топологических* четырёхмерных многообразий. Единственный классический инвариант компактного односвязного четырёхмерного многообразия — эта форма пересечений на его двумерных гомологиях. По теореме Хирцебруха первый класс Понтрягина $p_1(X^4)[X^4]$ есть утроенная сигнатура $b_2^+ - b_2^-$, где b_2^+ и b_2^- — размерности положительной и отрицательной частей этой квадратичной формы. Теория Фридмана утверждает, что с точностью до конечного числа возможностей топологическая классификация четырёхмерных многообразий повторяет алгебраическую классификацию форм.

Если же посмотреть на свежие результаты о гладких четырёхмерных многообразиях, то среди них имеются следующие:

(1)′ Существуют односвязные гладкие четырёхмерные многообразия, которые h-кобордантны, но не диффеоморфны.

(2)′ Имеется бесконечное счётное семейство гладких односвязных четырёхмерных многообразий, взаимно гомеоморфных, но с различными гладкими структурами.

(3)′ Существуют рациональные когомологические инварианты, которые (в отличие от классов Понтрягина) существенно зависят от гладкой структуры.

(4)′ Существует несчётное семейство гладких четырёхмерных многообразий, каждое из которых гомеоморфно $\mathbb{R}^4$, но со взаимно различными гладкими структурами.

По поводу более точных утверждений и ссылок на литературу (см. разд. III ниже). Эти факты, вытекающие из теории Янга — Миллса, подчёркивают, как необычна открывающаяся нашим глазам картина четырёхмерия.

---

[1] Simon K. Donaldson, The geometry of 4-manifolds, ICM86, pp. 43—54.

II. МЕТОДЫ

(i) Янг-миллсовские уравнения первого порядка. Эти уравнения в четырёхмерной геометрии в некотором смысле аналогичны уравнениям Коши — Римана в размерности два. Вместо функций на римановой поверхности в качестве основных геометрических объектов мы выбираем *связности* на расслоении E над ориентированным римановым четырёхмерным многообразием X. Структурная группа расслоения E — это некоторая компактная группа Ли G, например $SU(2)$ или $SO(3)$. Вместо расщепления производной от функции на голоморфную и антиголоморфную части мы имеем расщепление кривизны F_A связности A на автодуальную и антиавтодуальную части: $F_A = F_A^+ + F_A^-$. Эти части — компоненты в собственных подпространствах оператора Ходжа $*$, действующего на векторнозначных 2-формах. Вместо голоморфных функций мы имеем *антиавтодуальные связности* (или *инстантоны*) — решения уравнения $F_A^+ = 0$. Геометрически это условие означает, что кривизна F_A принимает противоположные по знаку значения на любых двух взаимноортогональных двумерных плоскостях в касательном пространстве к X.

Это уравнение антиавтодуальности есть уравнение в частных производных первого порядка на связность A. Оно имеет обширную группу симметрий — группу автоморфизмов, или «калибровочных преобразований», расслоения E. С учётом этого обстоятельства (мы отождествляем расслоения, которые различаются лишь калибровочным преобразованием) уравнение антиавтодуальности оказывается эллиптическим. Оно зависит лишь от конформного класса римановых метрик на X. Многие из его особых свойств протекают из основополагающего тождества, связывающего «энергию» решения с топологией расслоения E. Если X компактно и A антиавтодуально, то

$$\int_X |F_A|^2 d\mu = c(G) \cdot k(E), \tag{1}$$

где $c(G)$ — нормировочная константа, а $k(E)$ — целое число, некоторое характеристическое число расслоения E. Если, скажем, G есть $SO(3)$, то k — это первый класс Понтрягина E, взятый с обратным знаком и вычисленный на X. Опять же, аналогичные тождества имеют место для энергии голоморфных отображений.

(ii) Нелинейная фредгольмова теория. Дифференциальная топология бесконечномерных многообразий обеспечивает нас удобным языком для описания многих свойств янг-миллсовских инстантонов, прежде всего свойств, вытекающих из теоремы о неявных функциях для банаховых пространств. Она применяется к нелинейным дифференциальным операторам между подходящими соболевскими пространствами. Для данного расслоения $E \to X$ определено пространство $\mathscr{B}_E$ всех калибровочных классов связностей, орбит относительно действия группы калибровочных преобразований E. Некоторое открытое плотное подмножество $\mathscr{B}_E^*$ в $\mathscr{B}_E$ образует бесконечномерное многообразие; его дополнение, множество особенностей, состоит из *приводимых связностей* с группой голономии, представляющей собой подгруппу в G, централизатор которой собственным образом содержит центр G.

Другой фонд идей, которые могут быть применены здесь — это идеи, основанные на теореме Сарда и понятии трансверсальности. Как заметил Смейл [30], эти основополагающие конструкции дифференциальной топологии переносятся на бесконечномерные ситуации, в которых фигурируют *фредгольмовы* отображения — гладкие отображения, у которых производные имеют конечномерные ядра и коядра. В качестве иллюстрации (частично имеющей отношение к определению инвариантов в п. III (iv) ниже) рассмотрим фредгольмово отображение $\varphi\colon E \to F$ между банаховыми пространствами, индекс которого (целое число $\dim(\ker d\varphi)_x - \dim(\operatorname{coker} d\varphi)_x$, вычисленное для произвольной точки x в E) равен нулю. Типичные точки y в F являются регулярными значениями φ, и для них $\varphi^{-1}(y)$ есть дискретное подмножество в X. Если φ — собственное отображение, то это множество конечно. Мы можем присвоить знак каждой точке таким образом, что алгебраическая сумма по точкам слоя даст целочисленный инвариант, независимый от y. Доказательство в общем случае лишь незначительно отличается от конечномерного. Точно так же это целое число — «степень» φ — является деформационным инвариантом, не изменяющимся при собственных фредгольмовых гомотопиях. Более общо, если E заменить банаховым многообразием B, то подходящему фредгольмову отображению φ мы можем сопоставить гомологический класс в $H_d(B; \mathbb{Z})$, где d — индекс φ.

Уравнение антиавтодуальности вписывается в эту схему. *Пространство модулей* M_E (пространство классов эквивалентности решений уравнения $F_A^+ = 0$) определяется некоторым фредгольмовым отображением; индекс последнего был вычислен Атьёй, Хитчином и Зингером [2] (при помощи теоремы Атьи — Зингера об индексе, применённой к линеаризованному оператору). Они дали общую формулу

$$\dim M_E = 2a_G \cdot k(E) - \dim G\,(1 - b_1(X) + b_2^+(X)), \tag{2}$$

где a_G — целое число, зависящее лишь от G (скажем, для $G = \mathrm{SO}(3)$, равное 1). Это есть «виртуальная размерность» M_E, и обычно предполагают, что часть пространства модулей, лежащая в $\overset{*}{\mathscr{B}}_E$, является гладким многообразием этой размерности. Более точно, Фрид и Уленбек доказали в [15], что для нетривиальных $\mathrm{SU}(2)$- или $\mathrm{SO}(3)$-расслоений это действительно так для типичных метрик на X. В общем случае этого можно достичь путём малого «шевеления» уравнения антиавтодуальности. Заметим, что чётность $\dim M_E$ *не зависит от расслоения* E.

Того же рода идеи можно применить и к приводимым связностям, негладким точкам $\mathscr{B}$. Таким образом, мы можем описывать поведение в этих точках пространства модулей в типичных ситуациях. Здесь важны редукции с абелевой структурной группой S^1, и они могут быть изучены с помощью теории Ходжа. При вариации римановой метрики на X приводимые решения уравнения антиавтодуальности появляются на подмножествах коразмерности b_2^+. Следовательно, мы можем избежать их в семействах метрик размерности меньшей b_2^+. В семействах размерности b_2^+ мы сталкиваемся с некоторой фундаментальной особенностью в соответствующих пространствах модулей. Если, к примеру, X имеет отрицательно-определённую форму пересечений, т. е. $b_2^+ = 0$, особенности всегда имеют место и для типичных метрик являются конусами над комплексными проективными пространствами.

(iii) Компактификация. Пространство модулей Янга-Миллса, вообще говоря, некомпактно, но имеется теорема Уленбек, определяющая естественную компактификацию. Этот контроль «бесконечности» в пространстве $\mathscr{B}_E$ связностей помогает установить в некоторых специальных случаях собственность фредгольмова отображения, задающего пространство модулей.

Теорема Уленбекк [37] снабжает нас информацией о связностях с заданными ограничениями на их энергию. Для антиавтодуальных связностей эти оценки получаются из тождества (1). Ограничимся для простоты $\mathrm{SO}(3)$-расслоениями E, которые топологически определены характеристическими классами $p_1(E)$ из $H^4(X; \mathbb{Z}) \cong \mathbb{Z}$ и $w_2(E)$ из $H^2(X; \mathbb{Z}/2)$, удовлетворяющими условию $w_2^2 = p_1 \bmod 4$. Таким образом, зафиксировав $w_2 = U$, мы получаем семейство пространств модулей, скажем $M_j = M_{j,\,U}$, где $j \geqslant 0$. Можно определить топологию на

$$M_j \cup X \times M_{j-4} \cup S^2(X) \times M_{j-8} \cup \ \ldots,$$

такую что замыкание $\overline{M}_j$ пространства M_j компактно. Здесь $S^i(X)$ обозначает *симметрические произведения* i точек в X.

Точки низшего «страта» $S^i(X) \times M_{j-4i}$ соответствуют «воображаемым связностям», к плотности энергии $|F_A|^2$ которых добавлены δ-функции в i точках из X.

Благодаря исследованиям Таубса [34], продолженным в [7], мы хорошо осведомлены о структуре окрестностей низших стратов в $\overline{M}_k$, т. е. «концов» пространств модулей. Окрестности $S^i(X) \times M_{j-4i}$ в M_j описывают (проведя детальный анализ соответствующей теоремы о неявной функции) в терминах связностей в M_{j-4i}, i экземпляров «фундаментального инстантона» в точках X и «данных склейки», с помощью которых отождествляются эти составляющие части. Например, линк $X \times M_{j-4}$ в $\overline{M}_j$ обычно совпадает с экземпляром структурной группы $\mathrm{SO}(3)$.

Идеи этого рода, описывающие поведение дифференциальных операторов «на бесконечности» пространства функций, возникли недавно при изучении ряда дифференциально-геометрических проблем. В калибровочной теории Таубс использовал их для построения некоторого вариационного исчисления — см. сообщение Таубса на этом конгрессе[1].

(iv) Антиавтодуальные уравнения и голоморфная геометрия. Предположим, что базисное пространство X есть двумерная комплексная поверхность с эрмитовой метрикой. Если E — комплексное векторное расслоение над X (структурная группа которого — подгруппа унитарной группы), то любая связность определяет почти-комплексную структуру на E. Если связность антиавтодуальна, её кривизна имеет тип (1.1), откуда следует интегрируемость этой структуры. Мы получаем таким образом отображение из антиавтодуальных связностей на X в голоморфные расслоения, где последние уже зависят лишь от комплексной геометрии X.

Голоморфное расслоение, по определению, локально-тривиально, в то время как уравнение антиавтодуальности имеет много локальных решений. Однако глобально мы можем восстановить связность по его голоморфному расслоению. Размерность базисного пространства не играет здесь никакой особенной роли. Если (Y, ω) — произвольное кэлерово многообразие и $E \to Y$ — голоморфное расслоение со структурной группой, скажем $\mathrm{SL}(r, \mathbb{C})$, то любая метрика на E определяет редукцию структурной группы до $\mathrm{SU}(r)$, а также выделенную $\mathrm{SU}(r)$-связность. Будем искать метрики, для которых кривизна F этих связностей удовлетворяет уравнению

$$F \cdot \omega = 0 \qquad (3$$

---

[1] C. H. Taubes, Gauge theories and nonlinear partial differential equations, ICM86, pp. 1123—1132 (получасовое сообщение на секции «Дифференциальные уравнения в частных производных»). — *Прим. изд. ред.*

в каждой точке Y. Это — эллиптическое уравнение второго порядка для метрики на E. С другой стороны, в алгебраической (или головоморфной) геометрии существует понятие *стабильного* векторного расслоения, введённое геометрами в связи с проблемами модулей.

Справедливо

Предложение. *Голоморфное расслоение E стабильно тогда и только тогда, когда для него существует неприводимое решение дифференциального уравнения* (3). *В этом случае решение единственно.*

Это предложение было недавно доказано Уленбек и Яу [38]. В виде гипотезы этот результат был высказан (независимо) Хитчином и Кобаяси; в простейшем случае, когда Y — комплексная кривая, он эквивалентен теореме Нарасимхана — Сешадри и был получен в контексте теории Янга — Миллса Атьёй и Боттом [1]. Для алгебраической поверхности Y результат был доказан в [6].

Таким образом, для компактного кэлерова многообразия произвольной размерности эта теорема Уленбекк и Яу доставляет голоморфное описание унитарных связностей с кривизной типа $(1, 1)$, перпендикулярной к кэлеровой форме. Особое свойство, присущее лишь комплексным поверхностям, состоит в том, что для них эти связности суть *в точности* антиавтодуальные связности. Таким образом, для алгебраической поверхности пространство модулей M_E можно описать, используя алгебраическую геометрию. Это пространство есть квазипроективное комплексное многообразие. С этой точки зрения наилучшей алгебраической конструкцией является конструкция, предложенная Гизекером [20].

III. РЕЗУЛЬТАТЫ И ПРИЛОЖЕНИЯ

(i) Реализация форм пересечения. Здесь мы обсудим результаты, запрещающие существование гладких четырёхмерных многообразий с заданной формой пересечений. Эти результаты можно рассматривать ещё и как препятствия к сглаживанию топологических многообразий, сконструированных Фридманом. Из них также следует невозможность произвести гладкую *хирургию* по гомологическим классам во многих имеющихся примерах многообразий.

Первая теорема такого рода утверждала, что нестандартная (недиагонализуемая) определённая форма не может быть реализована гладким односвязным четырёхмерным многообразием [5]. В доказательстве использовалось пятимерное пространство модулей SU(2)-связностей. Этот результат был затем улучшен в двух различных направлениях.

С одной стороны, Финтушел и Стерн нашли более простое доказательство для отрицательно-определённых форм, которые представляют —2 или —3. Они рассмотрели SO(3)-связности, выбирая w_2 и p_1 таким образом, чтобы получить пространство модулей малых размерностей [12]. Их доказательство было проделано для многообразий без 2-кручения в H_1. В обоих доказательствах пространства модулей «обрезались», для того чтобы получить многообразия с известной границей. Вклады в границу приходят от линков особенностей в приводимых связностях и от низших стратов в компактификации пространства модулей. Утверждается, что многообразие модулей даёт кобордизм (а тем более и гомологическую эквивалентность) в $\overset{*}{\mathscr{B}}_E$ между этими границами.

С другой стороны, доказательство из работы [5] было обобщено Фурутой [18] и автором [8] с целью принять во внимание фундаментальную группу. Это потребовало более широкого использования понятия трансверсальности и детального исследования *ориентации* пространства модулей. В результате для определённых форм была получена следующая оптимальная

Теорема 1 [8]. *Если гладкое компактное ориентированное четырёхмерное многообразие имеет определённую форму пересечения, то эта форма может быть диагонализована над целыми числами.*

Имеются также результаты и для некоторых неопределённых форм [7]. Они доказываются сходным образом, но с использованием более сложного анализа и топологии. Разлагая четырёхмерный характеристический класс «универсального» расслоения над $\overset{*}{\mathscr{B}}_E \times X$, можно определить некоторое отображение

$$\mu\colon H_2(X;\mathbb{Z}) \to H^2(\mathscr{B}_E^*;\mathbb{Z}). \qquad (4)$$

Для неопределённых многообразий пространства модулей обычно не содержат приведённых связностей, так что, сужая μ, мы строим когомологический класс на пространстве модулей. По той же причине единственный вклад в границу происходит от низших стратов. Есть ещё дополнительные mod 2-когомологические классы, которые детектируют линки низших стратов в гомологиях $\overset{*}{\mathscr{B}}_E$. На настоящий момент наилучшим результатом является

Теорема 2 [7]. *Если гладкое компактное ориентированное четырёхмерное многообразие не имеет 2-кручения в H_1 и его* *форма пересечений чётна, с положительной частью ранга 2, то эта форма имеет вид*

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

С ростом b_2^+ эти методы теряют силу, из-за того что изменяются относительные размеры вкладов в концы от различных низших стратов.

(ii) Орбиобразия и реализация гомологических классов. Финтушел и Стерн начали изучение уравнений Янга — Миллса на четырёхмерных орбиобразиях — пространствах с дискретным множеством особенностей, смоделированных на конечных факторах $\mathbb{R}^4$. Эти орбиобразия естественно возникают как факторы гладких четырёхмерных многообразий по конечной группе или пятимерных многообразий по действию окружности. Орбиобразия являются рациональными гомологическими многообразиями, и анализ на них и на гладких многообразиях очень похож — главное отличие состоит в том, что благодаря наличию особенностей появляются дополнительные члены в формуле индекса (2). Используя разновидность своих рассуждений для многообразий, Финтушел и Стерн изучили условия существования орбиобразий с заданными формой пересечения и особенностями. Их результаты имеют многие применения, особенно к группе θ_H^3 гомологических трёхмерных сфер по модулю гомологического кобордизма. До работы Финтушела и Стерна казалось возможным, что эта группа достаточно мала, может быть даже порядка 2. На самом же деле справедлива

Теорема 3 [13]. *Гомологическая сфера Пуанкаре P имеет бесконечный порядок в θ_H^3 (т. е. никакая связная сумма $P\# \dots \#P$ не ограничивает ациклического гладкого четырёхмерного многообразия). Кроме того, $\theta_H^3/\langle P\rangle$ — ненулевая группа.*

(Ср. с тем, что сфера Пуанкаре ограничивает ациклическое топологическое четырёхмерное многообразие [16].)

Другим применением было новое доказательство теоремы Куги. Любой двумерный гомологический класс в четырёхмерном многообразии может быть реализован гладко вложенной поверхностью. Интересная общая проблема состоит в том, чтобы найти нижнюю границу на род такой поверхности. Теорема Куги имеет дело с классами в $H_2(S^2 \times S^2)$, которые записаны в стандартном базисе как пары (p, q) целых чисел.

Теорема 4 [23]. *Класс (p, q) в $H_2(S^2 \times S^2)$ может быть представлен гладко вложенной сферой, если и только если либо p, либо q есть 0, 1 или —1.*

(Ср. с тем фактом, что, если p и q взаимно просты, этот класс можно реализовать топологически плоско вложенной сферой.)

Первоначальное доказательство Куги было непрямым и использовало результаты п. III(*i*). (Похожие аргументы для других многообразий были предложены Лоусоном [25] и Сучу [31].) Используя орбиобразия, Финтушел и Стерн дали более простое доказательство. Когда двумерная сфера, вложенная в четырёхмерные многообразия с ненулевым числом самопересечения, стягивается в точку, образующее пространство будет орбиобразием (с сингулярностями, устроенными как конус над линзовым пространством). Позже Фурута [19] предложил ещё более прямое доказательство, используя другие пространства модулей на этих орбиобразиях. Действуя в ином направлении, Лоусон [24] использовал ту же технику для изучения вложенных проектных плоскостей.

Методы первой работы Финтушелла и Стерна были недавно обобщены Финтушелом, Лоусоном и Стерном. Одним из приложений их результатов являются утверждения об исключительных орбитах действия окружности на S^5, частично доказывающие гипотезу Монтгомери и Янга [14].

(ii) Экзотические структуры на $\mathbb{R}^4$. Теория Фридмана утверждает, что разложения формы пересечений на четырёхмерном многообразии в прямую сумму может быть реализовано, посредством хирургии, как топологическое разложение многообразий. Результаты п. III(i) показывают, что это нельзя сделать гладко. Из этой невозможности вытекает существование «экзотического $\mathbb{R}^4$» — гладкого многообразия, гомеоморфного, но не диффеоморфного эвклидову пространству [21]. Первые примеры такого рода были открытыми подмножествами в $S^2 \times S^2$ или $\mathbb{C}\mathbb{P}^2$. Доказательства их экзотической природы были непрямыми. Позже Гомпф, следуя этим путем, нашел счётное бесконечное семейство экзотических $\mathbb{R}^4$.

Впечатляющий прогресс был достигнут в дальнейшем Таубсом [35], осуществляющим программу, предложенную Фридманом. Таубс обобщил основы теории Янга — Миллса на четырёхмерные многообразия с «периодическими концами». Последние по определению суть некомпактные многообразия с концами, которые имеют периодическую форму $W_1 \cup W_2 \cup \dots \cup W_n \cup \dots$, где W_i — перекрывающиеся копии некоторого открытого многообразия W. Простейшие примеры — многообразия, у которых конец есть «труба» $Y^3 \times (0, \infty)$, а W — это $Y \times (0, 1)$. Таубс показал, что если выполнены некоторые условия

(1) на гомологии W и

(2) *на представления* $\pi_1(W) \to SU(2)$,

то четырёхмерные многообразия с периодическими концами ведут себя с точки зрения антиавтодуальных уравнений так же, как и компактные. Вводя полученные теоремы о формах пересечения на многообразиях с периодическими концами во фридмановский анализ неосуществимости гладкой хирургии. Таубс доказал следующий результат:

Теорема 5 [23, 25]. *Существует семейство $R_{s,t}$ гладких четырёхмерных многообразий, параметризованных $(s, t) \in \mathbb{R}^2$, каждое из которых гомеоморфно $\mathbb{R}^4$, но никакие два не диффеоморфны.*

Таким образом, у гладких структур на топологическом многообразии есть «модули». Более того, семейство Таубса содержит не все экзотические $\mathbb{R}^4$. Никакое из $R_{s,t}$ не может быть вложено в стандартное $\mathbb{R}^4$, но из того, что теорема об h-кобордизме неверна, следует, что примеры с таким свойством должны существовать.

В духе этой работы Таубса автор и Сулливан обобщили основания теории Янга — Миллса на квазиконформные четырёхмерные многообразия, у которых локальные координаты преобразуются квазиконформными заменами — отображениями областей в $\mathbb{R}^4$ [11]. Эта работа близка к работе Телемана о липшицевых многообразиях [36]. По существу, все результаты, доказанные для четырёхмерных многообразий и диффеоморфизмов с использованием уравнения антиавтодуальности, обобщаются на квазиморфные многообразия и квазиконформные отображения. (В частности, существуют экзотические квазиконформные структуры на $\mathbb{R}^4$.) Эти результаты и подавно обобщаются на липшицевы четырёхмерные многообразия. Мы имеем тут разительный контраст с теоремой Сулливана [32]: в больших размерностях любое топологическое многообразие имеет единственную липшицеву структуру.

(iv) Новые инварианты. Результаты этого пункта — другая сторона изложенного в п. III(i). Многие компактные топологические четырёхмерные многообразия нельзя сгладить; те же, которые могут быть сглажены, имеют много разных гладких структур. Это устанавливается посредством конструирования дифференциально-топологических инвариантов с помощью пространств модулей Янга — Миллса в соответствии с намётками п. II(i).

Сосредоточим внимание на односвязных четырёхмерных многообразиях. Для любого расслоения E над X рациональные когомологии порождены как кольцо классами c/α, где c — рациональный характеристический класс универсального расслоения $\mathscr{B}_E^* \times X$ и α — гомологический класс в X. В частности, все рациональные когомологии $\mathscr{B}_E^*$ лежат в *чётных размерностях.* Например, если G есть $SO(3)$, то $H^*(\mathscr{B}_E^*; \mathbb{Q})$ — это полиномиальная алгебра, порождённая образом отображения μ (см. (4)) и некоторым дополнительным классом из H^4. Поскольку инварианты, которые мы предполагаем получить на основе идей п. II(ii), лежат, грубо говоря, в *гомологиях* $\mathscr{B}_E^*$, хороших результатов можно ожидать лишь в случае, когда пространства модулей *чётномерны*, а согласно (2) так будет в точности тогда, когда $b_2^+(X)$ *чётно.*

Многообразия с $b_2^+ = 1$ образуют в этом контексте достаточно специальный класс, поскольку для них приводимые решения появляются в пространствах модулей для семейств метрик коразмерности 1. Рассмотрим двумерные пространства модулей $SU(2)$-связностей с классом Черна 1 ($SO(3)$-связности с классом Потрягина —4) на таком многообразии X. Они не всегда будут компактны, но есть возможность ввести член, компенсирующий наличие границы. Тогда можно сопоставить общей метрике на X гомологический класс в $\mathscr{B}_E^*$, который изменяется лишь вследствие появления приводимых решений. В результате мы получим дифференциально-топологический вариант X, имеющий вид отображения

$$\Gamma_X\colon \mathscr{C}_X \to H^2(X; \mathbb{Z}),$$

где $\mathscr{C}_X$ — некоторое множество «камер» в $H^2(X; \mathbb{R})$ [9].

Для алгебраических поверхностей X есть надежда вычислить этот инвариант, используя голоморфное описание антиавтодуальных связностей в терминах стабильных расслоений. Для любой рациональной поверхности $b_2^+ = 1$, но имеются и примеры иррациональных поверхностей с $b_2^+ = 1$, скажем семейство $D_{p,q}$ (p и q — взаимно простые целые), построенное Долгачёвым. Различие в комплексной геометрии рациональных и иррациональных многообразий отражается на стабильных расслоениях, а значит и на пространствах модулей и инварианте Γ. Это даёт следующую теорему:

Теорема 6 [9]. *Поверхность Долгачёва $D_{p,q}$ гомотопически эквивалентна (а следовательно, гомеоморфна и h-кобордантна), но не диффеоморфна связной сумме* $\mathbb{C}P^2 \# 9\overline{\mathbb{C}P}^2$.

Таким образом, теорема об h-кобордизме не распространяется на гладкие четырёхмерные многообразия. Следуя дальше, Р. Фридман и Морган и Оконек с Ван де Веном использовали Г-инвариант для доказательства такого результата:

Теорема 7 [17, 27]. *Среди многообразий, гомотопически эквивалентных $D_{p,q}$, существует бесконечно много диффеоморфных типов.*

Если b_2^+ нечётно и больше 1, то можно определить много других инвариантов. Для любого расслоения E пространства модулей имеют чётную размерность $2d(E)$. Классы $\mu(\alpha)$, $\alpha \in H_2(X)$, могут быть представлены коцепями с «малыми носителями» в $\overset{*}{\mathscr{B}}_E$. Наше описание концов пространства модулей делает возможным построение спаривания между степенями $\mu(\alpha)^d$ и фундаментальным классом пространства модулей M_E. Рассматривая SO(3)-связности, получаем:

Теорема 8 [10]. *Пусть X — односвязное гладкое ориентированное четырёхмерное многообразие с $b_2^+(X) = 2p+1$, $p > 0$. Зафиксируем какую-нибудь ориентацию максимального положительного подпространства формы пересечений на H^2. Тогда для любого u из $H^2(X; \mathbb{Z}/2)$ с $u^2 = \alpha \bmod 4$ и для $j > j_0(p)$ гомологическим классом пространства* SO(3)*-модулей $M_{u,j}$ определяется полином*

$$q_{u,j,X}\colon S^d(H_2(X)) \to \mathbb{Z}$$

степени $d = j - 3(1+p)$, не зависящий от выбора метрики на X.

(Здесь ориентация положительного подпространства определяет ориентацию пространства модулей M_E.) Таким образом, говоря грубо, для «половины» возможных односвязных четырёхмерных многообразий мы можем определить *бесконечно много* новых инвариантов. Пока вычислять их довольно трудно; их основное приложение связано с контрастом между двумя общими свойствами пространств модулей. С одной стороны, мы имеем «теорему об обращении в нуль» для связных сумм:

Теорема 9 [10]. *Если четырёхмерное многообразие X из теоремы 8 есть связная сумма $X_1 \# X_2$, где каждое $b_2^+(X_i) > 0$, то все инварианты $q_{u,j,X}$ нулевые.*

С другой стороны, если X — проективная алгебраическая поверхность, имеется выделенный класс «гиперплоскости» $[H]$ в $H_2(X)$. Для больших значений j пространства модулей являются квазипроективными многообразиями «правильной» размерности. Конструкция Гизекера показывает, что $\mu(H)$ есть первый класс Черна некоторого обильного линейного расслоения над пространством модулей. Таким образом, $\langle \mu(H)^d, M \rangle$ положительно, и мы получаем:

Теорема 10 [10]. *Если односвязная компактная комплексная алгебраическая поверхность может быть представлена в виде связной суммы, то форма пересечений одного из слагаемых отрицательно-определённа.*

Это немедленно даёт много примеров многообразий, имеющих одни и те же классические инварианты, но различаемых новыми инвариантами $q_{u,j,X}$.

IV. НЕРЕШЁННЫЕ ПРОБЛЕМЫ

Описанным здесь методам ещё весьма далеко до превращения в систематическую теорию. Один характерный момент состоит в том, что, хотя все известные доказательства десяти теорем раздела III используют инстантоны Янга — Миллса, в большинстве случаев имеется целый ряд параллельных доказательств (в которых привлекаются разные пространства модулей и т. д.). Похоже на то, что за всем этим стоит некоторый более фундаментальный принцип, связывающий топологию четырёхмерных многообразий и теорию Янга — Миллса, и указанные различные доказательства — лишь различные его проявления. Если бы мы нашли такой принцип, это могло бы указать путь к решению проблем, к которым с помощью методов, изложенных выше, подступиться не удаётся. Вот наиболее очевидные общие вопросы:

(1) Какие чётные неопределённые формы являются формами пересечений для гладких односвязных четырёхмерных многообразий? $\left(\text{Простейший пример, для которого ничего неизвестно, — форма } 4E_8 \oplus 3\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ ранга 38.}\right)$

(2) В каких гомотопических типах есть компактные односвязные четырёхмерные многообразия с различными гладкими структурами?

Для (ориентированных) многообразий с нечётным b_2^+ имеется много новых инвариантов, с помощью которых можно пытаться различить гладкие структуры. Поэтому мы приходим к вопросу

(3) Существуют ли гомотопические эквивалентные односвязные четырёхмерные многообразия с *чётным* b_2^+, обладающие различными гладкими структурами? Частный случай здесь — гладкая четырёхмерная гипотеза Пуанкаре.

С другой стороны, возникает много вопросов, касающихся наших новых инвариантов для многообразий с нечётным b_2^+, в частности:

(4) Есть ли универсальные сооотношения между инвариантами $q_{k,u,X}$?

(5) Есть ли способ систематически вычислять инварианты, имея какое-то стандартное описание четырёхмерного многообразия?

Два пути выглядят многообещающими. Во-первых, мы имеем голоморфное описание пространств модулей в случае, когда базисным многообразием служит комплексная поверхность. Быть может, имеются общие соотношения между нашими инвариантами и обычными алгебро-геометрическими инвариантами поверхностей. Один конкретный вопрос состоит в следующем:

(6) Пусть X — минимальная алгебраическая поверхность. Являются ли все инварианты $q_{k,u,X}$ из теоремы 8 полиномами от канонического класса $c_1(K_X)$ и формы пересечений на X?

Крайняя возможность здесь: эти полиномы суть универсальные полиномы от указанных двух переменных, с коэффициентами, зависящими от k.

Во-вторых, новый подход к изучению четырёхмерия может развиться из работы Кассона [4]. Кассон определил некоторый целочисленный инвариант для гомологической трёхмерной сферы Y^3, используя представления $\pi_1(Y^3)\to SU(2)$. Этот инвариант может быть вычислен с помощью описания Y, использующего хирургию Дена. Эти же представления выходят на передний план и в таубсовском обобщении теории Янга — Миллса на некомпактные (с периодическим концом) четырёхмерные многообразия. Недавно Таубс показал, что инвариант Кассона можно включить в ту же схему с фредгольмовыми отображениями банаховых многообразий, что была описана в п. II(ii). Далее, Кассон и Таубс, используя свои результаты, дали независимые доказательства следующего результата:

Теорема 11 [4, 35]. *Существует топологическое четырёхмерное многообразие, которое не гомеоморфно симплициальному комплексу.*

Эти два доказательства служат косвенным свидетельством существования каких-то связей между инвариантом Кассона и уравнениями антиавтодуальности на четырёхмерных многообразиях. Быть может, работа Кассона поможет найти способ, позволяющий дать определение наших новых инвариантов, привлекая более привычные методы геометрической топологии.

Благодарность. Автор весьма признателен Университету дель Валье (Кали, Колумбия) за гостеприимство, оказанное во время подготовки этой статьи.

ЛИТЕРАТУРА

1. M. F. Atiyah and R. Bott, The Yang-Mills equations over Riemann surfaces, Philos. Trans. Roy. Soc. London Ser. **A 308** (1982), 523—615.
2. M. F. Atiyah, N. J. Hitchin, and I. M. Singer, Self duality in four dimensional Riemannian geometry, Proc. Roy. Soc. London Ser. **A 362** (1978), 425—461.
3. W. Browder, Surgery on simply connected manifolds, Springer-Verlag, Berlin, 1971.
4. A. Casson (to appear).
5. S. K. Donaldson, An application of gauge theory to 4-dimensional topology, J. Differential Geom. **18** (1983).
6. —, Anti-self-dual Yang-Mills connections on complex algebraic surfaces and stable vector bundles, Proc. London Math. Soc. (3) **50** (1985), 1—26.
7. —, Connections, cohomology and the intersection forms of four manifolds, J. Differential Geom. **24** (1986), 275—341.
8. —, The orientation of Yang-Mills moduli spaces and four-manifold topology, J. Differential Geom. **26** (1987), 141—168.
9. —, Irrationality and the h-cobordism conjecture, J. Differential Geom. **25** (1987).
10. —, Polynomial invariants for smooth 4-manifolds, Topology, 1990.
11. S. K. Donaldson and D. P. Sullivan, Acta Matematica, **163**, 181—252.
12. R. Fintuschel and R. Stern, SO(3) connections and the topology of 4-manifolds, J. Differential Geom. **20** (1984), 523—539.
13. —, Pseudo-free orbifolds, Ann. of Math. (2) **122** (1985), **325**—364.
14. —, $O(2)$ actions on the 5-sphere, Invent. Math. (to appear).
15. D. S. Freed and K. K. Uhlenbeck, Instantons and 4-manifolds, Math. Sci. Res. Inst. Pub., Springer-Verlag, New York, 1984.
16. M. H. Freedman, The topology of 4-dimensienal manifolds, J. Differential Geom. **17** (1982), 357—453.
17. R. Friedman and J. W. Morgan, On the diffeomorphism types of certain algebraic surfaces, J. Differential Geom. **27** (1988), 297—369.
18. M. Furuta, Perturbation of moduli spaces of self-dual connections, J. Fac. Sci. Univ. Tokyo Sect. I A Math. **34** (198), 275—297.
19. —, On self-dual pseudo-connections on some orbifolds, Preprint, Univ. of Tokyo.
20. D. Gieseker, On the moduli of vector bundles over an algebraic surface, Ann. of Math. (2) **106** (1977), 45—60.

21. R. Gompf, Three exotic $\mathbf{R}^4$'s and other anomalies, J. Differential Geom. **18** (1983), 317—328.

22. —, An infinite set of exotic $\mathbf{R}^4$'s, J. Differential Geom. **21** (1985), 283—300.

23. K. Kuga, Representing homology classes in $S^2 \times S^2$, Topology **23** (1984), 133—137.

24. T. Lawson, Normal bundles for an embedded $\mathbf{RP}^2$ in a positive definite 4-manifold, J. Differential Geom. **22** (1985), 215—231.

25. —, Representing homology classes in an almost definite 4-manifold, Preprnit.

26. Новиков С. П. Гомотопически эквивалентные гладкие многообразия.— Изв. АН СССР, сер. мат., 1964, т. 28, с. 365—474.

27. C. Okonek and A. Van de Ven, Stable bundles and differentiable structures on certain elliptic surfaces, Invent. Math. **86** (1986), 357—370.

28. L. C. Siebenmann, Structures on topological manifolds, Proc. Internat. Congr. Math. (Nice, 1970), Vol. 2, Gauthier-Villars, Paris, 1971, pp. 133—163.

29. S. Smale, On the structure of manifolds, Amer. J. Math. **84** (1964), 387—399.

30. —, An infinite dimensional version of Sard's Theorem, Amer. J. Math. **87** (1965), 861—866.

31. A. I. Suciu, Immersed spheres in $\mathbf{CP}^2$ and $S^2 \times S^2$, Preprint, Yale University.

32. D. P. Sullivan, Hyperbolic geometry and homeomorphisms, Geometry Topology (Proc. Georgia Topology Conf., Athens, Ga., 1977), J. Cantrell, editor, Academic Press, New York, 1978, pp. 543—555.

33. —, Infinitesimal computations in topology, Inst. Hautes Études Sci. Publ. Math. **47** (1977), 269—332.

34. C. H. Taubes, On the existence of self-dual connections on manifolds with indefinite intersection matrix, J. Differential Geom. **19** (1984), 517—560.

35. —, Gauge theory on asymptotically periodic 4-manifolds, J. Differential Geom. **25** (1987), 363—430.

36. N. Teleman, The index of signature operators on Lipschitz manifolds, Inst. Hautes Études Sci. Publ. Math. **58** (1983), 39—79.

37. K. K. Uhlenbeck, Connections with L^p-bounds on curvature, Comm. Math. Phys. **83** (1982), 11—30.

38. K. K. Uhlenbeck and S.-T. Yau. On the existence of Hermitian Yang-Mills connections on stable bundles, Comm. on Pure and Appl. Math. **39** (198?), 257—293; errata *ibid* **42** (198?), 703—707.

МАТЕМАТИЧЕСКИЙ ИНСТИТУТ, ОКСФОРД, АНГЛИЯ

НОВЕЙШИЕ ДОСТИЖЕНИЯ В АРИФМЕТИЧЕСКОЙ АЛГЕБРАИЧЕСКОЙ ГЕОМЕТРИИ

Герд Фальтингз [1]

Я хотел бы представить в этом сообщении наиболее выдающиеся фрагменты развития арифметической алгебраической геометрии за последние четыре года. Разумеется, мой выбор совершенно субъективен и изложение не претендует на полноту. Мы сосредоточимся на следующих четырёх темах:

(a) формула Гросса — Цагира,
(b) гипотезы Тэйта, Шафаревича и Морделла,
(c) структуры Ходжа — Тэйта в p-адических когомологиях,
(d) гипотеза Бейлинсона.

В две из вышеназванных тем я сам внёс некоторый вклад и потому, быть может, имею определённое право рассказать об этих результатах перед столь многочисленной аудиторией. Что же касается двух других областей, то здесь я должен ограничиться изложением того, о чём узнал в последнее время из различных источников. Извинением заведомой неполноты моего изложения может служить то, что пленарные доклады рассчитаны всё же на более широкий круг слушателей, чем секционные. Желающим выслушать более компетентное изложение могу лишь рекомендовать соответствующие секционные доклады.

1. ФОРМУЛА ГРОССА — ЦАГИРА

В основе всего лежат две трудности, которые встречаются в диофантовой геометрии. Первая состоит в определении порядков нулей L-рядов. Эти ряды суть обобщения классических

---

[1] Gerd Faltings, Neuere Entwiklungen algebraischen Geometrie, ICM86, pp. 55—61.

L-рядов. Рассматриваются ряды вида $\Sigma a_n n^{-s}$, которые сходятся в некоторой полуплоскости $\mathrm{Re}(s) \geqslant s_0$. Такой L-ряд строится по некоторому арифметическому многообразию **X**, причём коэффициенты a_n вычисляют, исходя из поведения редукции многообразия **X** в конечных точках. Имеется предположение, что эти L-ряды продолжаются до мероморфных функций на всей комплексной плоскости и даже удовлетворяют некоторому функциональному уравнению. Это доказано для ряда арифметических многообразий **X**, прежде всего для некоторых многообразий Симуры. Кроме того, есть гипотеза о том, что порядки нулей и, соответственно, полюсов L-рядов в целых точках связаны с геометрическими инвариантами. Речь идёт о гипотезе Бёрча — Суиннертон-Дайера и, более общо, и гипотезе Тэйта.

Поскольку численные подсчёты проводятся с неизбежной погрешностью, с их помощью нельзя прямо установить, что некоторая аналитическая функция в данной точке обращается в нуль, хотя можно установить противное. Это делает сложной проверку указанных выше гипотез. Другая трудность относится к рациональным точкам на абелевых многообразиях, в частности на эллиптических кривых. Довольно просто найти точки кручения, но другие точки находятся значительно сложнее. При этом часто желательно показать, что группа Морделла — Вейля имеет положительный ранг, например в случае, когда это предсказывается гипотезой Бёрча — Суиннертон-Дайера. Формула Б. Гросса и Д. Цагира позволяет в некоторых случаях разрешить обе эти проблемы. Именно, она утверждает равенство между значением высоты Нерона — Тэйта некоторых рациональных точек на специальных эллиптических кривых (точек Хеегнера на кривых Вейля) и значением производной соответствующего L-ряда в точке $s = 1$. Это можно использовать в обоих направлениях: Если рациональная точка является точкой кручения (что просто установить), то производная L-ряда обращается в нуль (сложно). В обратную сторону: если производная не обращается в нуль (установить просто), то мы получаем некоторую точку бесконечного порядка (сложно).

Наиболее известное в настоящее время приложение формулы Гросса — Цагира — это её применение к получению эффективных нижних границ для чисел классов мнимых квадратичных полей. Они вычисляются на основе одного более позднего результата Гольдфельда, при условии что уже известен некоторый L-ряд указанного выше типа с нулём порядка, не меньшего трёх. Этот ряд как раз и находится с помощью формулы Гросса — Цагира.

Суммируя сказанное, можно сказать, что мы имеем дело с удивительным открытием, однако, к сожалению, не в состоянии «объяснить», почему это так.

Литература: [G1, G2, O, Zg].

2. ГИПОТЕЗА МОРДЕЛЛА

Обратимся к кругу вопросов, связанных с гипотезой Морделла. Она утверждает, что на кривой рода, большего единицы, над числовым полем лежит лишь конечное число рациональных точек. Это было предположено Л. Морделлом в 1922 г. в работе, где он установил, что рациональные точки на эллиптической кривой образуют конечно-порождённую абелеву группу. Между прочим, Морделл отмечал, что ему неизвестно никакого подхода к доказательству и даже соображений, показывающих, что высказанное предположение правдоподобно. Первые продвижения здесь принадлежат А. Вейлю и К.-Л. Зигелю: первый обобщил теорему Морделла на абелевы многообразия, а второй с помощью диофантовых приближений и теоремы А. Вейля сумел доказать гипотезу о конечности числа целочисленных точек. Впрочем, в дальнейшем эти результаты не привели к большим успехам, и любопытно, что окончательное доказательство опирается на совсем другие методы.

Это доказательство явилось развитием идей, приведших к доказательству гипотезы Морделла для функциональных полей. Дать это доказательство впервые удалось Манину и Грауэрту. Метод, сыгравший здесь важную роль, восходит к Паршину, Он позволяет свести искомое утверждение к гипотезе Шафаревича: над произвольным числовым полем имеется, с точностью до изоморфизма, лишь конечное число абелевых многообразий, имеющих хорошую редукцию вне заданного конечного множества точек поля.

Связь между этими двумя гипотезами осуществляется посредством конструкции некоторого накрытия данной кривой, связанного с фиксированной рациональной точкой, которое разветвлено только над этой точкой. Таким образом, получается новая кривая, которая может иметь плохую редукцию лишь в тех точках, где исходная кривая имела плохую редукцию, а также, возможно, ещё в некотором фиксированном множестве точек (например, во всех точках характеристики два). Класс изоморфизма данной кривой определяет данную рациональную точку, и затем нужно использовать теорему Торелли и гипотезу Шафаревича для якобиева многообразия накрывающей кривой. Отметим, что этот же метод даёт и теорему

Зигеля о целочисленных точках на аффинной эллиптической кривой.

Но как доказать гипотезу Шафаревича? Она могла вначале показаться неправдоподобной, поскольку над функциональными полями справедлива гипотеза Морделла, но не такое более сильное утверждение. Но над числовыми полями цели оказалось возможным достигнуть с помощью гипотезы Тэйта: она сводит данную проблему к некоторому вопросу об l-адических представлениях, разрешить который значительно легче, чем первоначальную проблему.

Для всего этого нужна определённая информация о пространстве модулей абелевых многообразий, снабжённых главной поляризацией, и в особенности о компактификациях этих пространств. Более точно, речь идёт об определении высоты абелева многообразия, т. е. высоты соответствующей точки пространства модулей. Для этого нужно компактификация пространства модулей, снабжённая обильным линейным расслоением с эрмитовой метрикой. Здесь приходится вначале применить ряд искусственных приёмов, но в итоге получается удовлетворительная теория. Сначала строится тороидальная компактификация над целыми числами. На эту компактификацию можно продолжить универсальное абелево многообразие до полуабелева многообразия, и мы получаем некоторое линейное расслоение ω, а именно пучок относительных дифференциалов этого полуабелева многообразия. Глобальные сечения пучка ω суть в точности зигелевы модулярные формы с целочисленными коэффициентами. С помощью тэта-рядов устанавливается, что некоторая степень ω глобально порождается своими сечениями и что образ тороидальной компактификации в проективном пространстве при соответствующем отображении является арифметической версией минимальной компактификации (или компактификации Сатакэ, или компактификации Бейли — Бореля). По построению, на ней имеется каноническое обильное линейное расслоение, и остаётся лишь определить подходящую метрику. Опять возможен только некоторый канонический выбор, а именно интегрирование квадрата дифференциала. К сожалению, на бесконечности данная метрика имеет особенности, однако эти особенности умеренны и не создают особых препятствий для проведения доказательства.

Таким образом получается простое определение высоты абелева многообразия, с которым можно работать. В существенном дело сводится к интегрированию квадрата образующей модуля целочисленных дифференциалов. Так получается некоторый численный инвариант абелева многообразия, такой что с точностью до изоморфизма имеется лишь конечное число

абелевых многообразий, инвариант которых не превосходит заданной величины. Отметим ещё, что арифметическая компактификация пространства модулей находит и другие приложения. В особенности полезной она оказывается в арифметической теории зигелевых модулярных форм.

Наконец, остаётся доказать гипотезу Тэйта. Более точно, требуется рассмотреть лишь специальный случай более общего предположения об эндоморфизмах абелевых многообразий. Доказательство восходит к Тэйту и Зархину, а дополнительно используется ещё компактификация пространства модулей плюс некоторые результаты Тэйта и Рэйно о p-делимых группах. Точнее говоря, рассматривается l-делимая подгруппа абелева многообразия, которая подразделяется на последовательные уровни. Это даёт последовательность абелевых многообразий, и надо установить, что среди них имеется бесконечное число изоморфных. Для этого рассматриваются соответственные численные инварианты этих многообразий, которые были определены выше. Достаточно доказать, что эта последовательность вещественных чисел стабилизируется. Таким образом, важно знать, как меняется определённая выше высота при изогениях. Можно показать, что изменение высоты в классе изогенности ограничено, и даже дать при этом вполне эффективно вычислимые границы.

Резюмируя, можно сказать, что эти результаты снова демонстрируют огромную силу методов Гротендика в алгебраической геометрии. Однако есть такое чувство, что эти теоремы в определенной степени завершают некоторый этап развития и что для изучения рациональных точек многообразий высших размерностей нужны новые методы. Конечно, разработка таких методов потребует многих попыток, так же, как это было и при изучении кривых.

Литература: [D, F2, F3, F4, FW, S1, S2, Z1, Z2].

3. p-АДИЧЕСКИЕ ПРЕДСТАВЛЕНИЯ

Здесь речь идёт о p-адических этальных когомологиях алгебраического многообразия над некоторым p-адическим полем. Связи этих когомологий с когерентными когомологиями позволяют получить некоторую p-адическую версию теории Ходжа. Кроме того, обнаруживаются связи этой теории с кристальными когомологиями. При этом вступает в игру «таинственный функтор» Гротендика.

Попытаемся объяснить, в чем тут дело. Пусть K — некоторое p-адическое поле, т. е. конечное расширение поля $\mathbb{Q}_p$. Обозначим через $\mathbb{C}_p$ пополнение алгебраического замыкания $\bar{K}$ поля K.

На нем действует непрерывными гомоморфизмами группа Галуа $\mathcal{G} = \mathrm{Gal}(\bar{K}/K)$. Пусть $\mathbb{C}_p(n)$ обозначает скручивание поля $\mathbb{C}_p$ с n-й степенью циклотомического характера $\mathcal{G} \to \mathbb{Z}_p^*$. Прежде всего отметим, что неимоверно огромное поле $\mathbb{C}_p$ поддаётся изучению: оно содержит промежуточное поле, порожденное корнями из единицы p-примарной степени, и почти неразветвлено над ним. Поэтому многие проблемы можно сводить к случаю этого меньшего поля, где они решаются прямым вычислением. Тэйт и Рэйно показали, что для собственного гладкого K-многообразия **X** справедливо соотношение

$$H^1(\mathbf{X}, \mathbb{Z}_p) \otimes \mathbb{C}_p \cong H^1(\mathbf{X}, \mathcal{O}_\mathbf{X}) \otimes \mathbb{C}_p \oplus H^0(\mathbf{X}, \mathcal{O}_\mathbf{X}) \quad (\text{как } \mathcal{G}\text{-модули}).$$

Следует также заметить, что различные скручивания $\mathbb{C}_p(m)$ не изоморфны и что всякое $\mathbb{C}_p$-векторное пространство с полулинейным $\mathcal{G}$-действием содержит некоторое наибольшее подпространство, изоморфное прямой сумме модулей $\mathbb{C}_p(m)$.

Было высказано предположение — и его удалось доказать, что, вообще, справедливо разложение:

$$H^n(\mathbf{X}, \mathbb{Z}_p) \otimes \mathbb{C}_p \cong \sum_{a+b=n} H^a(\mathbf{X}, \Omega_\mathbf{X}^b) \otimes \mathbb{C}_p(-b).$$

Тем самым можно, в частности, дать чисто алгебраическое доказательство вырождения спектральной последовательности Ходжа. Получается также и свойство симметрии чисел Ходжа, по крайней мере для проективных многообразий. Как известно, этих свойств уже достаточно для доказательства многих таких результатов, для получения которых обычно используют аналитические методы. Примером является теорема Кодаиры об обращении в нуль.

Способ доказательства основан на одном варианте промежуточных когомологий, в которые естественно отображаются обе части указанного выше соотношения. Отсюда вытекает, что обе эти части удовлетворяют обычным аксиомам, таким как формула Кюннета или двойственность Пуанкаре. Детали несколько сложны, но в существенном эту промежуточную теорию можно описать следующим образом.

Пусть R — аффинное кольцо многообразия, и пусть $\bar{R}$ обозначает p-адическое пополнение максимального неразветвлённого расширения R. Возьмём когомологии Галуа для $\bar{R}$. Когомологии Галуа для $\mathbb{Z}_p$ отображаются сюда, что даёт первое из естественных преобразований. Для получения другого рассматриваются дифференциалы. Вот некоторые детали. Примем, что R содержит достаточно много единиц. Присоединением корней из этих единиц получается хорошо контролируемое расширение

$\tilde{R}$ кольца R, которое в случае характеристики нуль неразветвлено. Кольцо $\bar{R}$ неразветвлено над $\tilde{R}$ в коразмерности $\leqslant 2$, что позволяет переносить различные трудные результаты с $\tilde{R}$ на $\bar{R}$. Пусть, к примеру, $\bar{V}$ обозначает пополнение целого замыкания кольца дискретного нормирования V поля K. Точная $\mathcal{G}$-линейная последовательность

$$0 \to \Omega_{\bar{V}/V} \otimes_{\bar{V}} \bar{R} \to \Omega_{\bar{R}/R} \to \Omega_{\bar{R}/RV} \to 0,$$

с учётом значений первых членов, вычисленных Ж. Фонтэном, приводит к искомому второму отображению, т. е. к соотношению между дифференциалами и когомологиями Галуа.

В некоторых специальных случаях можно сказать и больше: имеется соотношение между p-адическими этальными когомологиями и кристальными когомологиями, и они однозначно определяют друг друга. Это даёт пример «таинственного функтора», поисками которого занимался Гротендик. Доказательство Ж. Фонтэна и В. Мессинга, использующее «синтомическую топологию», наверняка будет представлено в одном из специальных докладов.

Стоит также отметить, что недавно Фонтэн доказал ещё одну гипотезу Шафаревича: не существует абелевых многообразий над $\mathbb{Q}$, имеющих всюду хорошую редукцию. В доказательстве применяются теория конечных групповых схем, и, естественно, p-адические представления.

Литература: [F5, F6, Fo].

4. АРИФМЕТИЧЕСКАЯ ТЕОРИЯ ПЕРЕСЕЧЕНИЙ

При изучении числовых полей выяснилось, что наилучшие результаты достигаются, если одновременно работать с конечными и с бесконечными точками. Если рассматривать многообразия над числовыми полями, то понятно, что надо делать в конечных точках: нужна хорошая модель над соответствующим кольцом дискретного нормирования. Что же отвечает бесконечным точкам? Опыт показывает, что алгебраические объекты необходимо снабжать метриками. Если, к примеру, рассматривается линейное расслоение на некоторой алгебраической кривой (над числовым полем), то требуется и продолжение этой модели над целыми числами. Так поступают в конечных точках. В бесконечных же точках линейное расслоение снабжается метрикой. На этом пути С. Аракелов развил теорию пересечений на арифметических поверхностях, и можно показать, что на этот случай переносится и часть теории пересечений на алгебраических поверхностях. Это относится и к теореме Римана — Роха, и к теореме Ходжа об индексе.

При этом используются следующие соображения. Сначала определяются формы объёма в когомологиях метризованного линейного расслоения. Этот результат чисто локален в бесконечных точках и относится к теории римановых поверхностей. Определение объёма подобрано так, что обычное доказательство теоремы Римана — Роха для алгебраических поверхностей переносится почти без изменений. Из теоремы Римана — Роха выводится связь между теорией пересечений и теорией высоты Нерона — Тэйта. Кроме того, справедливы ещё и другие результаты, например, о том, что для кривых рода, большего единицы, квадрат дуализирующего расслоения ω меньше либо равен нулю. Что, однако, при этом не проходит, так это аналог теории Ходжа и теории Фробениуса.

Далее встаёт вопрос, как быть в случае высших размерностей. Новые трудности возникают уже для конечных точек, поскольку отсутствует аналог теории минимальных моделей. Несмотря на это, А. Бейлинсону, А. Жилле и К. Суле удалось разработать теорию пересечений. Она основана на очень красивом сочетании идей из алгебры и анализа и, на мой взгляд, имеет очень широкие перспективы для дальнейших исследований.

Сходная область — гипотеза Бейлинсона об L-рядах. Мне представляется, что здесь очень важный момент — это вопрос о построении мотивов. Их существование было предположено А. Гротендиком и послужило лейтмотивом многих когомологических исследований. Они должны представлять собой, в определённом смысле, универсальную теорию когомологий. К настоящему моменту нам, собственно, хорошо известны лишь 1-мотивы, которые в существенном являются полуабелевыми алгебраическими группами. С помощью алгебраической K-теории Бейлинсон нашёл кандидатов для мотивов высших размерностей.

Литература: [B1 — B3, F1, GS].

ЛИТЕРАТУРА

[B1] Бейлинсон А. Высшие регуляторы и значения L-функций. Современные проблемы математики, Изд. ВИНИТИ, т. 24, 1984, с. 181—238.

[B2] —, Notes on absolute Hodge cohomology, preprint, 1984.

[B3] —, Height pairings between algebraic cycles, preprint, 1985.

[D] P. Deligne, Séminaire Bourbaki **616** (1984). [Имеется перевод: Делинь П. Доказательство гипотез Тейта и Шафаревича. — В сб.: Алгебра и теория чисел с приложениями. — М.: Мир, 1987, с. 100—124.]

[F1] G. Faltings, Calculus on arithmetic surfaces, Ann. of Math. **119** (1984), 387—424.

[F2] —, Endlichkeitssätze für abelsche Varietäten über Zahlkörpern, Invent. Math. **73** (1983), 349—366.

[F3] —, Die Vermutungen von Tate und Mordell, Jahresber. Deutsch. Math.-Verein. **86** (1984), 1—13.

[F4] —, Arithmetische Kompaktifizierung des Modulraums der abelschen Varietäten, Lecture Notes in Math., vol. 1111, Springer-Verlag, 1985, pp. 321—383.

[F5] —, Hodge-Tate structures and modular forms, submitted to Math. Ann.

[F6] —, P-adic Hodge-theory, Manuscript, Princeton University, Princeton, N. J., 1985.

[Fo] J. Fontaine, Il n'y a pas de variété abelienne sur **Z**, Invent. Math. **81** (1985), 515—538.

[FW] G. Faltings und G. Wüstholz, Rational points, Vieweg, Braunschweig, 1984.

[G1] D. Goldfeld, The class number of quadratic fields and the conjecture of Birch and Swinnerton-Dyer, Ann. Scuola Norm. Sup. Pisa **3** (1976), 623—663.

[G2] —, Gauss' class number problem for imaginary quadratic fields, Bull. Amer. Math. Soc. (N. S.) **13** (1985), 25—37.

[GS] H. Gillet und C. Soulé, Intersection sur les variétés d'Arakelov, C. R. Acad. Sci. Paris **299** (1984), 563—566.

[GZ] B. Gross und D. Zagier, Points de Heegner et dérivés de fonctions L, C. R. Acad. Sci. Paris Sér. I Math. **297** (1983), 85—87.

[O] J. Osterlé, Séminaire Bourbaki **631** (1984). [Имеется перевод: Остерле Ж. Кривые на абелевом многообразии (по Рейно). — В сб.: Алгебра и теория чисел с приложениями. — М.: Мир, 1987, с. 204—218.]

[S1] L. Szpiro, Séminaire Bourbaki **619** (1984). [Имеется перевод: Шпиро Л. Гипотеза Морделла. — В сб.: Алгебра и теория чисел с приложениями. — М.: Мир, 1987, с. 125—150.]

[S2] —, Séminaire sur les pinceaux arithmétiques: La conjecture de Mordell, Astérisque **127** (1985).

[Z1] Зархин Ю. Г. Изогении абелевых многообразий над полями конечной характеристики. — Мат. сб., 1974, т. 24, с. 451—461.

[Z2] —, Эндоморфизмы абелевых многообразий над полями конечной характеристики. — Изв. АН СССР, сер. мат., 1975, т. 9, с. 255—260.

[Zg] D. Zagier, Modular points, modular curves, modular surfaces and modular forms, Lecture Notes in Math., vol. 1111, Springer-Verlag, pp. 225—248.

ПРИНСТОНСКИЙ УНИВЕРСИТЕТ, ПРИНСТОН, НЬЮ-ДЖЕРСИ 08544, США

НЕКОТОРЫЕ ВОПРОСЫ ТЕОРИИ КВАЗИКОНФОРМНЫХ ОТОБРАЖЕНИЙ

Фредерик У. Геринг[1]

I. ВВЕДЕНИЕ

1. Обозначения. Для $n \geqslant 1$ через $\mathbb{R}^n$ обозначим n-мерное эвклидово пространство, при $x \in \mathbb{R}^n$ и $0 < r < \infty$ пусть $\mathbb{B}^n(x, r)$ — открытый n-мерный шар с центром x радиуса r, $\mathbb{S}^{n-1}(x, r) = \partial\, \mathbb{B}^n(x, r)$, $\mathbb{B}^n = \mathbb{B}^n(0, 1)$ и $\mathbb{S}^{n-1} = \mathbb{S}^{n-1}(0, 1)$. Через $\overline{\mathbb{R}}^n = \mathbb{R}^n \cup \{\infty\}$ обозначим одноточечную компактификацию $\mathbb{R}^n$, наделённую хордальной метрикой

$$q(x, y) = |\, p(x) - p(y)\, |, \tag{1.1}$$

где p — стереографическая проекция $\overline{\mathbb{R}}^n$ на сферу $\mathbb{S}^n$ в $\mathbb{R}^{n+1}$. В дальнейшем топология и сходимость рассматриваются относительно этой метрики.

Пусть D и D' — области в $\overline{\mathbb{R}}^n$, а $f: D \to D'$ — гомеоморфизм. Для $x \in D \setminus \{\infty, f^{-1}(\infty)\}$ положим

$$H_f(x) = \limsup_{r \to 0} H_f(x, r), \tag{1.2}$$

где при $0 < r < \operatorname{dist}(x, \partial D)$

$$H_f(x, r) = \frac{\max\{\, |f(x) - f(y)| : |x - y| = r\}}{\min\{\, |f(x) - f(z)| : |x - z| = r\}}, \tag{1.3}$$

и продолжим $H_f(x)$ в точки ∞ и $f^{-1}(\infty)$, полагая $H_f(\infty) = H_{f \circ g}(0)$ и $H_f(f^{-1}(\infty)) = H_{g \circ f}(f^{-1}(\infty))$, где $g(x) = x/|\, x\,|^2$. В случае $n > 2$ величину

$$K(f) = \begin{cases} \infty, & \text{если } \sup\limits_{x \in D} H_f(x) = \infty, \\ \operatorname*{ess\,sup}\limits_{x \in D} H_f(x), & \text{если } \sup\limits_{x \in D} H_f(x) < \infty, \end{cases} \tag{1.4}$$

---

[1] Frederick W. Gehring, Topics in quasiconformal mappings. ICM86, pp. 55—80.

назовём *линейной дилатацией* f в D. В этом докладе будем говорить, что f *квазиконформно*, если $K(f) < \infty$, и K-*квазиконформно*, если $K(f) \leqslant K$, где $1 \leqslant K < \infty$. Таким образом, гомеоморфизм квазиконформен, если для любой точки D «коэффициент искажения» формы бесконечно малой $(n-1)$-мерной сферы с центром в этой точке конечен; гомеоморфизм K-квазиконформен, если, кроме того, этот коэффициент не превосходит величины K почти во всех точках области D.

Следующий результат показывает, что класс квазиконформных отображений, в соответствии с их названием, является расширением класса конформных отображений.

1.5. **Теорема.** *Пусть D и D' — области в $\overline{\mathbb{R}}^n$, а f: $D \to D'$ — гомеоморфизм. Если $n = 2$, то отображение f 1-квазиконформно тогда и только тогда, когда оно само или комплексно-сопряжённое к нему представляет собой мероморфную функцию комплексного переменного в D. Если $n \geqslant 3$, то f 1-квазиконформно тогда и только тогда, когда f является ограничением на D мёбиусова преобразования, т. е. композицией конечного числа отражений относительно $(n-1)$-мерных сфер и плоскостей.*

Для $n = 2$ теорема 1.5 есть простая переформулировка одной теоремы Меньшова [M3]. Для $n \geqslant 3$ теорема 1.5 обобщает хорошо известный результат Лиувилля, освобождая его от априорных условий гладкости f [G3, R3].

2. Историческая справка. Плоские квазиконформные отображения изучаются уже почти шестьдесят лет. Они появились в конце 1920-х годов в работах Грёцша, который рассматривал задачу определения *возможно более близкого к конформному* гомеоморфизма между парой топологически эквивалентных плоских конфигураций с одним конформным инвариантом [G14]. Позднее они появились под названием *квазиконформных* в одной работе Альфорса о накрывающих поверхностях [A1].

В конце 1930-х годов Тайхмюллер далеко распространил исследования Грёцша на отображения замкнутых римановых поверхностей и получил весьма естественное пространство параметров для поверхностей фиксированного рода g, а именно пространство, гомеоморфное $\mathbb{R}^{6g-6}$ [T1]. Почти в то же самое время Лаврентьев и Морри обобщили классический результат Гаусса о существовании изотермических координат, доказав варианты теоремы, известной ныне как *измеримая теорема Римана об отображении для квазиконформных отображений* [L1, M4].

В последующие годы Альфорс, Берс и их последователи существенно обобщили результаты Тайхмюллера и с успехом использовали квазиконформные отображения во многих областях комплексного анализа, включая клейновы группы и топологию поверхностей [A5, B6, E1, K2]. Предложенное недавно Салливаном решение проблемы Фату — Жюлиа показывает, что этот класс отображений может быть весьма эффективно использован также при исследовании задач, касающихся итераций рациональных функций [S3, S4].

Квазиконформные отображения в высших размерностях рассматривались Лаврентьевым ещё в 1930-х годах [L2]. Однако систематического инструмента, пригодного для изучения этого класса отображений, не было вплоть до 1959 г., когда Лёвнер ввёл понятие *конформной ёмкости*, с тем чтобы показать, что пространство $\mathbb{R}^n$ нельзя квазиконформно отобразить на его правильную часть [L7].

Затем Геринг, Вяйсяля и многие другие, применяя метод Лёвнера и его эквивалентную формулировку в терминах экстремальных длин, получили первые результаты о квазиконформных отображениях в $\mathbb{R}^n$[G3, V1]. В конце 1960-х годов Решетняк и школа финских математиков опубликовали серию работ, в которых многомерная теория распространялась на неинъективные квазиконформные, или, как говорят, *квазирегулярные*, отображения [M1, R2, V4]. Это направление исследований недавно увенчалось замечательным обобщением теоремы Пикара, полученным Рикманом [R6].

3. Роль квазиконформных отображений. Плоские квазиконформные отображения — важный инструмент комплексного анализа. Они особенно ценны при исследовании римановых поверхностей и разрывных групп. Теорема Берса об одновременной униформализации [B3] служит примером красивого применения измеримой теоремы Римана об отображении, а решение обратной задачи теории Неванлинны, данное Дрейзином [D2], иллюстрирует, как ту же теорему можно применять при решении задач комплексного анализа в том же духе, в каком $\bar{\partial}$-уравнение используется в гармоническом анализе и теории функций многих комплексных переменных.

Геометрические доказательства, которые обычно нужны для установления квазиконформных аналогов результатов теории конформных отображений, иногда порождают новое понимание классических теорем и методов теории аналитических функций [L3]. Неожиданным и привлекательным образом квазиконформные отображения возникают в других разделах математики, например в гармоническом анализе при изучении функций с ограниченным средним колебанием и сингулярных интегралов [B1], в геометрии и теории упругости в связи с вопросом об инъективности и продолжении квазиизометрий.

Теория многомерных квазиконформных отображений представляет собой новое и нетривиальное обобщение комплексного анализа в $\mathbb{R}^n$, отличное от предложенного в [N2] и, пожалуй, более геометричное и гибкое, чем аналитическая теория, основанная на теории функций многих комплексных переменных. Эти отображения использовались при решении задач дифференциальной геометрии, они образуют замкнутый класс отображений, интерполирующий гомеоморфизмы и диффеоморфизмы, для которого многие результаты дифференциальной топологии справедливы независимо от размерности. Наконец, некоторые методы, развитые при изучении многомерных квазиконформных отображений, нашли важные приложения в других областях математики, например обратное неравенство Гёльдера в теории уравнений с частными производными [G13].

4. Комментарий к приведённому выше определению. Квазиконформные отображения, изучавшиеся Грёцшем и Тайхмюллером, предполагались непрерывно дифференцируемыми всюду, кроме, быть может, конечного числа точек. Позднее Альфорс [A2] и Берс [B2] заметили, что естественнее работать с отображениями $f: D \to D'$, для которых выполнено важное неравенство

$$K(f) \leqslant \liminf_{j \to \infty} K(f_j), \tag{4.1}$$

где $\{f_j\}$ — любая последовательность гомеоморфизмов, сходящаяся к f локально-равномерно в D. Мы определили $K(f)$ формулой (1.4), а не более простой формулой

$$K(f) = \sup_{x \in D} H_f(x) \tag{4.2}$$

как раз для того, чтобы выполнялось (4.1).

Неравенство (4.1) означает, что класс K-квазиконформных отображений замкнут относительно локально-равномерной сходимости. Более того, при $n=2$ измеримая теорема Римана об отображении позволяет заключить, что каждый гомеоморфизм f с $K(f) \leqslant K$ является локально-равномерным пределом непрерывно дифференцируемых гомеоморфизмов f_j с $K(f_j) \leqslant K$ [L3]. При $n=3$ совсем иные соображения приводят к тому же заключению, но с $K(f_j) \leqslant \tilde{K}$, где $\tilde{K}$ зависит только от K [K1]. В случае $n > 3$ вопрос, по-видимому, открыт.

Если $f: D \to D'$ — гомеоморфизм с $K(f) < \infty$, то теорема Радемахера — Степанова и соображения, аналогичные использованным Меньшовым, позволяют заключить, что f дифференцируемо почти всюду с якобианом $J_f \neq 0$ почти всюду в D, что f принадлежит классу Соболева $W^1_{n,\,\mathrm{loc}}(D)$ и что $K(f) = K(f^{-1})$

[G3]. Таким образом, отображение, обратное к квазиконформному, тоже квазиконформно; аналогично композиция K_1- и K_2-квазиконформных отображений есть $K_1 \cdot K_2$-квазиконформное отображение. Хотя 1-квазиконформные отображения вещественно-аналитичны, для любого $K > 1$ существует K-квазиконформное отображение f пространства $\mathbb{R}^n$ на себя, которое не дифференцируемо на множестве хаусдорфовой размерности n.

5. *Замечание.* Поскольку имеется целый ряд превосходных обзорных статей по плоским квазиконформным отображениям и их связям с пространствами Тайхмюллера [A6, B4, B5, B7], оставшаяся часть этой лекции будет посвящена менее развитой теории для высших размерностей. В разделе II мы обсудим некоторые основные результаты и открытые вопросы в теории квазиконформных отображений, сравнивая известное для $n = 2$ и $n > 2$. Затем в разделе III мы обсудим ряд примеров, когда такие отображения естественно возникают в других областях математики.

II. НЕКОТОРЫЕ РЕЗУЛЬТАТЫ И ОТКРЫТЫЕ ВОПРОСЫ

6. *Инструменты для изучения квазиконформных отображений.* Гомеоморфизм $f : D \to D'$ квазиконформен, если конечна *функция искажения* H_f, определённая соотношением (1.2). Это условие локальное, и нам надо найти некоторый способ проинтегрировать его по области D, чтобы получать глобальные свойства f. В классическом комплексном анализе это достигается посредством интегральной формулы Коши. Хотя аналогичная ей формула Помпею порой полезна при работе с плоскими квазиконформными отображениями, инструментом, чаще всего заменяющим формулу Коши, являются метод экстремальных длин, сформулированный Альфорсом и Бёрлингом [A8], и его обобщение на случай высших размерностей [F3, V3].

7. *Модуль семейства кривых.* Пусть Γ — семейство кривых в $\overline{\mathbb{R}}^n$, и пусть $\operatorname{adm}(\Gamma)$ — множество всех измеримых по Борелю функций $\rho : \mathbb{R}^n \to [0, \infty]$, таких что $\int_\gamma \rho\, ds \geqslant 1$ для любой локально-спрямляемой кривой γ из Γ. Числа

$$\operatorname{mod}(\Gamma) = \inf_{\rho \in \operatorname{adm}(\Gamma)} \int_{\mathbb{R}^n} \rho^n\, dm \quad \text{и} \quad \lambda(\Gamma) = \operatorname{mod}(\Gamma)^{1/(1-n)} \tag{7.1}$$

называются *конформным модулем* и *экстремальной длиной* Γ соответственно.

Как нетрудно видеть, $\operatorname{mod}(\Gamma)$ является внешней мерой на пространстве всех семейств кривых в $\overline{\mathbb{R}}^n$. Соответственно, если кривые из семейства Γ рассматривать как однородные провода, то $\lambda(\Gamma)$ можно считать сопротивлением семейства Γ. В частности, $\operatorname{mod}(\Gamma)$ велик, если кривые в Γ короткие и их много, и мал в противном случае.

Важность понятия конформного модуля в настоящем контексте определяется его квазиинвариантностью при квазиконформных отображениях.

7.2. **Теорема.** *Если отображение $f : D \to D'$ является K-квазиконформным, а Γ — семейство кривых, лежащих в D, то*

$$K^{1-n} \operatorname{mod}(\Gamma) \leqslant \operatorname{mod}(f(\Gamma)) \leqslant K^{n-1} \operatorname{mod}(\Gamma). \tag{7.3}$$

Неравенство (7.3) играет ключевую роль при изучении квазиконформных отображений. По этой причине величину

$$K^*(f) = \max\left(\sup_\Gamma \left(\frac{\operatorname{mod}(f(\Gamma))}{\operatorname{mod}(\Gamma)}\right),\ \sup_\Gamma \left(\frac{\operatorname{mod}(\Gamma)}{\operatorname{mod}(f(\Gamma))}\right)\right) \tag{7.4}$$

принято называть *максимальной дилатацией* (искажением) f и говорить, что f *K-квазиконформно*, если $K^*(f) \leqslant K$; при этом верхняя грань в (7.4) берётся по всем семействам кривых Γ из D, для которых $\operatorname{mod}(\Gamma)$ и $\operatorname{mod}(f(\Gamma))$ не равны одновременно 0 или ∞. Неравенства

$$K(f)^{n/2} \leqslant K^*(f) \leqslant K(f)^{n-1} \tag{7.5}$$

показывают, что это определение приводит к прежнему классу квазиконформных отображений и что $K^*(f) = K(f)$, если $n = 2$ или $K(f) = 1$.

Гомеоморфизм $f : \mathbb{R}^n \to \mathbb{R}^n$ квазиконформен тогда и только тогда, когда существует постоянная c, такая что

$$\limsup_{r \to 0} H_f(x, r) \leqslant c \tag{7.6}$$

для всех $x \in \mathbb{R}^n$. Проиллюстрируем применение соотношений (7.3), установив с их помощью глобальную форму последнего неравенства:

7.7. **Теорема.** *Если $f : \mathbb{R}^n \to \mathbb{R}^n$ K-квазиконформно, то*

$$H_f(x, r) \leqslant c \tag{7.8}$$

для всех $x \in \mathbb{R}^n$ и $0 < r < \infty$, где $c = c(K, n)$.

Доказательство опирается на две оценки конформных модулей конкретных семейств кривых [G2, G12, V1].

7.9. **Лемма.** *Если* $0 < a < b < \infty$ *и* Γ — *семейство открытых дуг в* $\overline{\mathbb{R}}^n$, *соединяющих* $\mathbb{S}^{n-1}(0, a)$ *с* $\mathbb{S}^{n-1}(0, b)$, *то*

$$\operatorname{mod}(\Gamma) \leqslant \omega_{n-1}\left(\log \frac{b}{a}\right)^{1-n},$$

где ω_{n-1} *есть* $(n-1)$*-мера* $\mathbb{S}^{n-1}$.

7.10. **Лемма.** *Если* C_1 *и* C_2 — *непересекающиеся континуумы в* $\overline{\mathbb{R}}^n$, *соединяющие соответственно* 0 *с* $\mathbb{S}^{n-1}(0, a)$ *и* ∞ *с* $\mathbb{S}^{n-1}(0, b)$, *а* Γ — *семейство всех открытых дуг, соединяющих* C_1 *с* C_2 *в* $\overline{\mathbb{R}}^n \setminus (C_1 \cup C_2)$, *то*

$$\operatorname{mod}(\Gamma) \geqslant \omega_{n-1}\left(\log\left(\lambda_n\left(\frac{b}{a}+1\right)\right)\right)^{1-n},$$

где λ_n *зависит только от* n.

7.11. **Следствие.** *Пусть* $n > 2$. *Если* C_1 *и* C_2 — *непересекающиеся зацепленные континуумы в* $\overline{\mathbb{R}}^n$ и Γ — *семейство всех открытых дуг, соединяющих* C_1 *и* C_2 *в* $\overline{\mathbb{R}}^n \setminus (C_1 \cup C_2)$, *то* $\operatorname{mod}(\Gamma) \geqslant c$, *где* $c = c(n) > 0$.

Доказательство теоремы 7.7. Выполнив предварительно сдвиг, можно считать, что $x = 0$ и $f(0) = 0$. Пусть m и M — соответственно минимум и максимум значений $|f|$ на $\mathbb{S}^{n-1}(0, r)$, и пусть $m < M$. Положим

$$C_1 = \{x \in \mathbb{R}^n : |f(x)| \leqslant m\}, \quad C_2 = \{x \in \mathbb{R}^n : |f(x)| \geqslant M\} \cup \{\infty\}$$

и рассмотрим семейство Γ открытых дуг, соединяющих C_1 с C_2 в $\overline{\mathbb{R}}^n \setminus (C_1 \cup C_2)$. Тогда приведённые выше оценки с учётом (7.3) дают

$$\omega_{n-1}(\log 2\lambda_n)^{1-n} \leqslant \operatorname{mod}(\Gamma) \leqslant K^{n-1} \operatorname{mod}(f(\Gamma))$$
$$\leqslant K^{n-1}\omega_{n-1}\left(\log \frac{M}{m}\right)^{1-n},$$

и мы получаем (7.8) с $c = (2\lambda_n)^K$.

8. Проблема существования отображения. Один из основных вопросов рассматриваемой тематики состоит в выяснении квазиконформной эквивалентности двух областей, т. е. выяснении того, можно ли одну из них квазиконформно отобразить на другую. Поскольку общий случай весьма сложен даже при $n = 2$, рассмотрим здесь более простую задачу описания областей D в $\overline{\mathbb{R}}^n$, квазиконформно-эквивалентных единичному шару $\mathbb{B}^n$. Теорема Римана о существовании конформного отображения и оценки из лемм 7.9 и 7.10 дают полный ответ на этот вопрос в случае $n = 2$:

8.1. **Теорема.** *Область* D *в* $\overline{\mathbb{R}}^2$ *квазиконформно-эквивалентна* $\mathbb{B}^2$ *тогда и только тогда, когда* ∂D — *невырожденный континуум.*

В высших размерностях подобной характеризации нет. Действительно, рассмотрение областей D_3 и D_4, определённых формулами (8.6) ниже, показывает, что при $n > 2$ по одной только границе области нельзя сказать, является ли область квазиконформным образом шара $\mathbb{B}^n$ при некотором гомеоморфизме $\overline{\mathbb{R}}^n$ на себя.

Следующее достаточное условие получается с помощью методов, используемых при решении многомерной проблемы Шёнфлиса [G5, M2].

8.2. **Теорема.** *Область* D в $\overline{\mathbb{R}}^n$ *квазиконформно-эквивалентна шару* $\mathbb{B}^n$, *если существуют замкнутые множества* $E \subset D, E' \subset \mathbb{B}^n$ *и квазиконформное отображение* $g : D \setminus E \to \mathbb{B}^n \setminus E'$, *такие что* $|g(x)| \to 1$ *при* $x \to \partial D$ *в* D.

Как и в топологическом случае, можно доказать локальный вариант теоремы 8.2, когда D — жорданова область в $\overline{\mathbb{R}}^n$, т. е. когда граница ∂D гомеоморфна $\mathbb{S}^{n-1}$ [B10, G1].

8.3. **Следствие.** *Если* D — *область в* $\mathbb{R}^n$, *граница* ∂D *которой диффеоморфна* $\mathbb{S}^{n-1}$, *то* D — *квазиконформно-эквивалентна* $\mathbb{B}^n$.

Легко построить область в $\mathbb{R}^n$, квазиконформно-эквивалентную шару $\mathbb{B}^n$ и не имеющую касательной плоскости ни в одной точке своей границы [G12]. Таким образом, приведённое в следствии 8.3 достаточное условие весьма далеко от необходимости.

Приводимое ниже необходимое условие квазиконформной эквивалентности шару $\mathbb{B}^n$ связано со следующим уточнением понятия локальной связности. Множество $E \subset \overline{\mathbb{R}}^n$ называют *линейно-локально-связным*, если существует постоянная c, $1 \leqslant c < \infty$, такая что для любых $x \in \mathbb{R}^n$ и $0 < r < \infty$

$$\begin{aligned} &E \cap \overline{\mathbb{B}}^n(x, r) \text{ содержится в некоторой компоненте } E \cap \overline{\mathbb{B}}^n(x, cr),\\ &E \setminus \mathbb{B}^n(x, r) \text{ содержится в некоторой компоненте } E \setminus \mathbb{B}^n(x, r/c). \end{aligned} \tag{8.4}$$

Соображения, основанные на тех же неравенствах (7.3), с использованием оценок из леммы 7.9 и следствия 7.11, приводят к такому результату [G7, G12]:

8.5. **Теорема.** *Если $n > 2$ и если область D в $\overline{\mathbb{R}}^n$ квазиконформно-эквивалентна $\mathbb{B}^n$, то $\overline{\mathbb{R}}^n \setminus D$ линейно-локально-связно.*

Эта теорема даёт много простых областей в $\mathbb{R}^n$, которые гомеоморфны, но не квазиконформно-эквивалентны шару $\mathbb{B}^n$. Например, пусть

$$\begin{aligned} D_1 &= \{x \in \mathbb{R}^n:\ r < 1,\ |x_n| < \infty\}, \\ D_2 &= \{x \in \mathbb{R}^n:\ r < \infty,\ |x_n| < 1\}, \\ D_3 &= \{x \in \mathbb{R}^n:\ x_n > \min(r^{1/2},\ 1)\}, \\ D_4 &= \{x \in \mathbb{R}^n:\ x_n < \min(r^{1/2},\ 1)\}, \end{aligned} \qquad (8.6)$$

где $x = (x_1, \ldots, x_n)$ и $r = (x_1^2 + \ldots + x_n^2)^{1/2}$. Тогда явно строятся квазиконформные отображения областей D_1 и D_3 на $\mathbb{B}^n$. С другой стороны, при $n > 2$ множества $\overline{\mathbb{R}}^n \setminus D_2$ и $\overline{\mathbb{R}}^n \setminus D_4$ не будут линейно-локально-связными, а поэтому области D_2 и D_4 не будут квазиконформно-эквивалентными $\mathbb{B}^n$.

Необходимое условие из теоремы 8.5 не является достаточным, и задача отыскания геометрических критериев (необходимых и достаточных) для проверки на квазиконформную эквивалентность шару $\mathbb{B}^n$ остаётся одним из наиболее интересных открытых вопросов.

9. Гомеоморфные и квазиконформное продолжения. Пусть D и D' — области в $\overline{\mathbb{R}}^n$, а f: $D \to D'$ — квазиконформное отображение. Обсудим вопрос, при каких условиях f допускает гомеоморфное продолжение на $\overline{D}$ или квазиконформное продолжение на $\overline{\mathbb{R}}^n$.

9.1. **Теорема.** *Пусть D и D' — односвязные области гиперболического типа в $\overline{\mathbb{R}}^2$. Для того чтобы каждое квазиконформное отображение $f: D \to D'$ гомеоморфно продолжалось на $\overline{D}$, необходимо и достаточно, чтобы области D и D' были жордановыми.*

Достаточность следует из одной теоремы Альфорса [A2], а необходимость — из результатов работы [E3]. В высших размерностях справедлива следующая теорема [V2]:

9.2. **Теорема.** *Если D и D' — жордановы области в $\overline{\mathbb{R}}^n$ и D квазиконформно-эквивалентна шару $\mathbb{B}^n$, то каждое квазиконформное отображение $f: D \to D'$ допускает гомеоморфное продолжение на $\overline{D}$.*

При $n = 2$ второе условие теоремы 9.2 излишне, поскольку любая жорданова область в $\overline{\mathbb{R}}^2$ конформно-эквивалентна кругу $\mathbb{B}^2$. Если же $n > 2$, то, как показывают примеры (8.6), это уже не так и теорема 9.2 неверна без этого дополнительного условия.

Теперь о задаче квазиконформного продолжения на $\overline{\mathbb{R}}^n$. Говорят, что множество E в $\overline{\mathbb{R}}^2$ есть *K-квазидиск* (соотв. *K-квазиокружность*), если оно является образом $\mathbb{B}^2$ (соотв. $\mathbb{S}^1$) при K-квазиконформном отображении $\overline{\mathbb{R}}^2$ на себя. По теореме Альфорса жорданова область D представляет собой квазидиск тогда и только тогда, когда существует такая постоянная c, что для любых $z_1, z_2 \in \partial D$

$$\min_{j=1,2} \operatorname{diam}(\gamma_j) \leqslant c\,|z_1 - z_2|, \qquad (9.3)$$

где γ_1 и γ_2 — компоненты $\partial D \setminus \{z_1, z_2\}$ [A3].

9.4. **Теорема.** *Пусть D и D' — жордановы области в $\overline{\mathbb{R}}^2$. Для того чтобы каждое квазиконформное отображение $f: D \to D'$ допускало квазиконформное продолжение на $\overline{\mathbb{R}}^2$, необходимо и достаточно, чтобы D и D' были квазидисками.*

Односвязная область D в $\overline{\mathbb{R}}^2$ является квазидиском в точности тогда, когда она линейно-локально-связна. Таким образом, последнее понятие возникает также в связи с задачей квазиконформного продолжения.

Достаточность в теореме 9.4 установлена Альфорсом [A2], а необходимость — Риккманом [R5]. Многомерный аналог этого результата звучит так [G4, V5]:

9.5. **Теорема.** *Пусть $n > 2$ и D — жорданова область в $\overline{\mathbb{R}}^n$. Для того чтобы каждое квазиконформное отображение $f: D \to \mathbb{B}^n$ допускало квазиконформное продолжение на $\overline{\mathbb{R}}^n$, необходимо и достаточно, чтобы область $D^* = \overline{\mathbb{R}}^n \setminus \overline{D}$ была квазиконформно-эквивалентна шару $\mathbb{B}^n$.*

Таким образом, проблема квазиконформного продолжения в высших размерностях отличается от плоского случая в двух отношениях. Во-первых, при $n = 2$ внешность D^* любой жордановой области квазиконформно-эквивалентна $\mathbb{B}^2$; при $n > 2$ это не так, что уже отмечалось выше. Во-вторых, при $n > 2$ любое квазиконформное отображение $f: D \to \mathbb{B}^n$ допускает квазиконформное продолжение на $\overline{\mathbb{R}}^n$, если D^* квазиконформно-эквивалентна $\mathbb{B}^n$; при $n = 2$ это не так, поскольку существуют жордановы области D, которые не удовлетворяют условию (9.3) и потому не являются квазидисками.

10. Соответствие границ. Переходим к задаче описания граничного отображения, индуцированного квазиконформным отображением шара или полупространства на себя. Для $n \geqslant 2$ пусть $\mathbb{H}^n$ обозначает верхнее полупространство $\{x \in \mathbb{R}^n : x_n > 0\}$. Тогда каждое квазиконформное отображение $f : \mathbb{H}^n \to \mathbb{H}^n$ допускает квазиконформное продолжение $\bar{f}$ на $\overline{\mathbb{R}}^n$, ограничение φ которого на $\partial\mathbb{H}^n$ является гомеоморфизмом $\overline{\mathbb{R}}^{n-1}$. Исследование этого граничного гомоморфизма было начато Бёрлингом и Альфорсом [B8].

10.1. **Теорема.** *Гомеоморфизм* φ: $\overline{\mathbb{R}}^1 \to \overline{\mathbb{R}}^1$, *удовлетворяющий условию* $\varphi(\infty) = \infty$, *является граничным для некоторого квазиконформного отображения* f *полуплоскости* $\mathbb{H}^2$ *на себя, подчинённого условию* $\bar{f}(\infty) = \infty$ *в том и только в том случае, если существует такая постоянная* c, *что*

$$\frac{1}{c} \leqslant \frac{\varphi(x+r) - \varphi(x)}{\varphi(x) - \varphi(x-r)} \leqslant c \qquad (10.2)$$

для любых $x \in \mathbb{R}^1$ *и* $0 < r < \infty$.

Неравенства (10.2) эквивалентны условию $H_\varphi(x, r) \leqslant c$. В многомерном аналоге теоремы 10.1 это условие заменяется его локальной формой $H_\varphi(x) \leqslant c$, что равносильно квазиконформности φ.

10.3. **Теорема.** *При* $n > 2$ *гомеоморфизм* φ: $\overline{\mathbb{R}}^{n-1} \to \overline{\mathbb{R}}^{n-1}$ *является граничным для некоторого квазиконформного отображения* f *полупространства* $\mathbb{H}^n$ *на себя тогда и только тогда, когда* φ *сам квазиконформен.*

Утверждение о необходимости в теоремах 10.1, 10.3 вытекает соответственно из неравенств (7.8), (7.6) и того факта, что

$$H_\varphi(x, r) \leqslant H_{\bar{f}}(x, r), \quad H_\varphi(x) \leqslant H_{\bar{f}}(x) \qquad (10.4)$$

для любых x и r.

Бёрлинг и Альфорс доказали достаточность условий теоремы 10.1, показав, что формула

$$f(x_1, x_2) = \frac{1}{2x_2} \int_0^{x_2} (\varphi(x_1 + t) + \varphi(x_1 - t))\, dt + \frac{i}{2x_2} \int_0^{x_2} (\varphi(x_1 + t) - \varphi(x_1 - t))\, dt \qquad (10.5)$$

определяет квазиконформное продолжение φ на $\mathbb{H}^2$.

Альфорс [A4], модифицировав эту конструкцию и используя возможность разложить любое квазиконформное отображение φ: $\overline{\mathbb{R}}^2 \to \overline{\mathbb{R}}^2$ в композицию отображений с малыми искажениями (см. следствие 11.4), построил квазиконформное продолжение φ на $\mathbb{H}^3$ и тем самым доказал достаточность условия теоремы 10.3 в случае $n = 3$. Затем Карлесон [C1], воспользовавшись совсем иным методом трёхмерной топологии, продолжил каждое квазиконформное отображение φ: $\overline{\mathbb{R}}^3 \to \overline{\mathbb{R}}^3$ на $\mathbb{H}^4$. Наконец, Тукиа и Вяйсяля [T5], отправляясь от идеи Карлесона и применяя результаты Сулливана [S1], установили достаточность условия теоремы 10.3 для любого n.

После взятия композиции с подходящими мёбиусовыми преобразованиями, (10.2) даёт характеризацию в терминах двойного отношения граничных отображений $\varphi : \partial D \to \partial D$, индуцированных произвольными квазиконформными отображениями диска или полуплоскости $D \subset \mathbb{R}^2$ на себя, а (10.5) определяет явное квазиконформное продолжение $T(\varphi) : D \to D$ каждого такого граничного отображения φ. Недавно Тукиа [T4] решил важную задачу теории Тайхмюллера, показав, что если G — подгруппа группы $\text{Möb}(D)$ всех мёбиусовых преобразований области D на себя, то любое *G-согласованное* граничное соответствие $\varphi : \partial D \to \partial D$ допускает G-согласованное квазиконформное продолжение на D. Дуади и Эрл [D1], развивая это достижение, явно предъявили *конформно-естественный* оператор T_0 квазиконформного распространения, такой что

$$g \circ T_0(\varphi) \circ h = T_0(g \circ \varphi \circ h) \qquad (10.6)$$

для любого гомеоморфизма $\varphi : \partial D \to \partial D$ и любых $g, h \in \text{Möb}(D)$. Этот красивый оператор, без сомнения, приведёт к многим новым результатам в рассматриваемой области (см. [E2]).

Пусть D — шар или полупространство в $\mathbb{R}^n$, $n > 2$. Метод Дуади и Эрла сопоставляет каждому гомеоморфизму $\varphi : \partial D \to \partial D$ его непрерывное продолжение $T_0(\varphi) : D \to D$, для которого выполнено соотношение (10.6). Однако отображение $T_0(\varphi)$ не является, вообще говоря, ни квазиконформным, ни инъективным, за исключением ситуации, когда $K(\varphi)$ достаточно мало, т. е. $K(\varphi) \leqslant K_n$, где K_n зависит лишь от n. Было бы интересно выяснить, любое ли квазиконформное φ допускает конформно-естественное квазиконформное распространение.

11. Измеримая теорема Римана об отображении и разложение отображения. Если отображение $f : D \to D'$ квазиконформно, то оно имеет невырожденный дифференциал $df : \mathbb{R}^n \to \mathbb{R}^n$ почти в каждой точке $x \in D$. В любой такой точке x дифференциал $df = df(x)$ отображает некоторый эллипсоид $E_f = E_f(x)$ с центром 0 и единичной минимальной осью на $(n-1)$-мерную сферу с центром 0. Тогда $H_f(x)$ равняется длине максимальной оси E_f, а наибольшее растяжение, осуществляемое отображением f в точке x, имеет место в направлении наименьшей оси $E_f(x)$. Если отображение $g: D' \to D''$ квазиконформно, то в силу теоремы 1.5 оно конформно тогда и только тогда, когда $E_{g \circ f} = E_f$ почти всюду в D, при этом E_f определяет f с точностью до композиции с действующим вслед за ним конформным отображением.

Если $n=2$ и f сохраняет ориентацию, то E_f определяется *коэффициентом Бельтрами* (или *комплексной дилатацией*, *комплексной характеристикой*)

$$\mu_f(x) = f_{\bar{x}}/f_x, \quad x = x_1 + ix_2 \tag{11.1}$$

отображения f в точке x. В частности, функция μ_f измерима, причём

$$|\mu_f(x)| = \frac{H_f(x)-1}{H_f(x)+1}, \quad \|\mu_f\|_{L^\infty} = \frac{K(f)-1}{K(f)+1} < 1, \tag{11.2}$$

и $\mu_{g \circ f} = \mu_f$ почти всюду тогда и только тогда, когда отображение $g: D' \to D''$ конформно. Более того, в размерности два можно априори задавать дилатацию μ_f, а значит и эллипс E_f, почти в каждой точке $x \in D$ [A7].

11.3. **Измеримая теорема Римана об отображении.** *Если функция μ на $\overline{\mathbb{R}}^2$ измерима и $\|\mu\|_{L^\infty} < 1$, то существует такое квазиконформное отображение $f = f_\mu$ плоскости $\overline{\mathbb{R}}^2$ на себя, что $\mu_f = \mu$ почти всюду. Если f нормировано тремя неподвижными точками, то f определено однозначно и f голоморфно зависит от μ.*

Теорема 11.3 играет фундаментальную роль при исследовании комплексной структуры пространства Тайхмюллера. Она является мощным средством и при решении других задач комплексного анализа. Примером может служить данное Дрейзином решение обратной задачи теории Неванлинны [D2], в котором тот сначала строит локально-квазиконформную функцию g с заданными дефектами, а затем, применяя теорему 11.3, получает квазиконформное отображение f плоскости $\mathbb{R}^2$ на себя, такое что $g \circ f$ является мероморфной функцией с теми же дефектами, что и g. Другой пример — недавнее решение Сулливаном [S3] задачи Фату — Жюлиа о блуждающих областях, где теорема 11.3 была использована для построения достаточно широкого вещественно-аналитического семейства квазиконформных деформаций заданной рациональной функции.

Важным следствием теоремы 1.3 является следующее утверждение:

11.4. **Следствие.** *Если $n=2$ и $\varepsilon > 0$, то любое квазиконформное отображение $f: D \to D'$ можно представить в виде $f = f_1 \circ \ldots \circ f_m$, где $K(f_j) < 1 + \varepsilon$ для $j = 1, \ldots, m$, причём $m = m(\varepsilon, K(f))$.*

Аналога теоремы 11.3 для высших размерностей нет. Более того, как показывают примеры, при $n > 2$ следствие 11.4 почти определённо неверно без дополнительных ограничений на область D. Важная открытая проблема состоит в выяснении того, сохраняется ли этот результат в какой-либо форме хотя бы для случая $D = D' = \overline{\mathbb{R}}^n$.

12. Квазиконформные группы. Группу G гомеоморфизмов пространства $\overline{\mathbb{R}}^n$ на себя называют *дискретной*, если в G нет последовательности элементов, сходящейся локально-равномерно в $\overline{\mathbb{R}}^n$ к тождественному отображению, и *K-квазиконформной*, если $K(g) \leqslant K$ для любого g из G. Хотя семейство квазиконформных групп содержит все мёбиусовы группы, можно, используя теорему 11.3, показать, что при $n=2$ это более широкое семейство никаких новых явлений не демонстрирует [S2, T2].

12.1. **Теорема.** *При $n=2$ каждая квазиконформная группа представима в виде $G = f^{-1} \circ H \circ f$, где H — мёбиусова группа, а f — квазиконформное отображение $\overline{\mathbb{R}}^2$ на себя.*

В высших размерностях ситуация иная, и для любого $n > 2$ существует квазиконформная группа, которая даже как топологическая группа не изоморфна никакой мёбиусовой [T3]. Тем не менее следующее свойство сходимости позволяет установить квазиконформные аналоги многих фундаментальных свойств мёбиусовых групп [G10]:

12.2. **Теорема.** *Если G — дискретная квазиконформная группа, то для любой последовательности различных элементов G существуют подпоследовательность $\{g_j\}$ и точки x_0, $y_0 \in \overline{\mathbb{R}}^n$, такие что $g_j \to y_0$ локально-равномерно в $\overline{\mathbb{R}}^n \setminus \{x_0\}$ и $g_j^{-1} \to x_0$ локально-равномерно в $\overline{\mathbb{R}}^n \setminus \{y_0\}$.*

Пусть G — группа гомеоморфизмов пространства $\overline{\mathbb{R}}^n$ на себя. Мы говорим, что *G — группа дискретной сходимости*, если она удовлетворяет заключению теоремы 12.2. Далее, мы называем элемент $g \in G$ *эллиптическим*, если он конечного порядка или периодичен, и *параболическим* (соотв. *локсодромическим*), если он бесконечного порядка и обладает одной неподвижной точкой (соотв. двумя неподвижными точками). *Предельное множество* $L(G)$ — это дополнение к множеству $O(G)$ обычных точек (тех точек $x \in \overline{\mathbb{R}}^n$, которые обладают такой окрестностью U, что $g(U) \cap U \neq \varnothing$ разве лишь для конечного множества $g \in G$). Наконец, группа G *собственно-разрывна* на открытом множестве O, если для любого компакта $F \subset O$ пересечение $g(F) \cap F$ непусто разве лишь для конечного множества элементов $g \in G$ [G10].

12.3. **Теорема.** *Пусть G — группа дискретной сходимости. Тогда каждый элемент G — эллиптический, параболический или локсодромический, а предельное множество $L(G)$ нигде не плотно или совпадает с $\overline{\mathbb{R}}^n$. Далее, если* $\operatorname{card}(L(G)) > 2$, *то множество $L(G)$ совершенно и лежит в замыкании любого непустого G-инвариантного множества, а множество пар неподвижных точек локсодромических элементов группы G плотно в $L(G) \times L(G)$.*

Хотя группы дискретной сходимости во многих отношениях напоминают мёбиусовы группы, существуют примеры, показывающие, что они не обязаны быть топологически сопряжённы с мёбиусовыми группами [F2, G10]. Они также естественным образом возникают в ситуациях, не имеющих ничего общего с мёбиусовыми или квазиконформными группами.

12.4. **Теорема.** *Группа G гомеоморфизмов пространства $\overline{\mathbb{R}}^n$ на себя является группой дискретной сходимости, если она собственно-разрывна на $\overline{\mathbb{R}}^n \setminus E$, где E — замкнутое вполне разрывное множество.*

Было бы интересно выяснить, в какой мере классическая теория клейновых групп переносится на этот более общий класс групп.

13. Непрерывность по Гёльдеру и интегрируемость. Теорему 12.2 можно получить, опираясь на соотношение (4.1) и следующую оценку изменения хордального расстояния q (см. (1.1)) при квазиконформном отображении [G7]:

13.1. **Теорема.** *Если отображение f: $D \to D'$ является K-квазиконформным и $\overline{\mathbb{R}}^n \setminus D \neq \varnothing$, то для $x, y \in D$*

$$q(f(x),\ f(y))\, q(\overline{\mathbb{R}}^n \setminus D') \leqslant c\,(q(x,\ y)/q(x,\ \partial D))^{1/K}, \qquad (13.2)$$

где $q(E)$ — хордальный диаметр E, а $c = c(n)$.

Теорема 13.1 в свою очередь есть следствие соотношения (7.3) и лемм 7.9, 7.10. Она показывает, что для D, $D' \subset \mathbb{R}^n$ каждое K-квазиконформное отображение $f: D \to D'$ локально удовлетворяет условию Гёльдера с показателем $1/K$. Таким образом, квазиконформные отображения в определённом смысле осуществляют интерполяцию между диффеоморфизмами и гомеоморфизмами при $1 \leqslant K < \infty$. Это обстоятельство проявляется также в степени интегрируемости якобиана J_f отображения f:

13.3. **Теорема.** *Если отображение $f: D \to D'$, где D, $D' \subset \mathbb{R}^n$, K-квазиконформно, то якобиан J_f локально L^p-интегрируем в D при $1 \leqslant p < p(K, n)$, где*

$$p(K,\ n) \leqslant K/(K-1), \quad \lim_{K \to 1} p(K,\ n) = \infty. \qquad (13.4)$$

Для случая $n = 2$ Боярский [B9] установил утверждаемую локальную интегрируемость с показателем $p(K, n)$, применяя неравенство Зигмунда — Кальдерона к преобразованию Бёрлинга (14.7). При $n > 2$ доказательство опирается на то обстоятельство, что $g = |J_f|$ удовлетворяет обратному неравенству Гёльдера:

$$\frac{1}{m(Q)} \int_Q g\, dm \leqslant c \left(\frac{1}{m(Q)} \int_Q g^{1/n}\, dm \right)^n, \qquad c = c(K,\ n) \qquad (13.5)$$

на любом n-кубе $Q \subset D$ с $\operatorname{diam}(f(Q)) < d(f(Q), D')$, а также на лемму о том, что если (13.5) выполняется на любом n-кубе Q, лежащем в кубе $\tilde{Q}$, то g принадлежит $L^p(\tilde{Q})$ при $1 \leqslant p < p(c, n)$ [G6]. Эти результаты были уточнены в работах [I2, R4], где было получено второе из соотношений (13.4). Пример

$$f(x) = |x|^{-a} x, \quad a = (K-1)/K, \qquad (13.6)$$

даёт первое из соотношений (13.4). Есть основания считать, что в теореме 13.3 можно взять $p(K,\ n) = K/(K-1)$, но это не доказано даже для $n = 2$.

III. СВЯЗИ С ДРУГИМИ ОБЛАСТЯМИ МАТЕМАТИКИ

14. Гармонический и функциональный анализ. Квазиконформные отображения встречаются в гармоническом анализе, ввиду их связей с функциями, имеющими ограниченное среднее колебание, и с сингулярными интегралами.

Говорят, что функция u имеет *ограниченное среднее колебание* в области $D \subset \mathbb{R}^n$ (или принадлежит $\mathrm{BMO}(D)$), если она локально-интегрируема и

$$\|u\|_{\mathrm{BMO}\,(D)} = \sup_B \frac{1}{m(B)} \int_B |u - u_B|\, dm < \infty, \qquad (14.1)$$

где верхняя грань берется по всем n-мерным шарам B, таким что $\bar{B} \subset D$, а

$$u_B = \frac{1}{m(B)} \int_B u\,dm. \tag{14.2}$$

Класс ВМО был введен Джоном и Ниренбергом [J3] в связи с исследованиями Джона по теории упругости и стал играть весьма заметную роль после того, как Фефферман показал, что класс $\mathrm{BMO}(\mathbb{R}^n)$ двойствен пространству Харди $H^1(\mathbb{R}^n)$ [F1].

Следующие связи между классом ВМО и классом квазиконформных отображений были установлены Райманном [R1]:

14.3. **Теорема.** *Если отображение $f\colon D \to D'$, D, $D' \subset \mathbb{R}^n$, K-квазиконформно, то $\|\log J_f\|_{\mathrm{BMO}(D)} \leqslant c$, где $c = c(K, n)$.*

14.4. **Теорема.** *Гомеоморфизм $f\colon D \to D'$ между областями D, $D' \subset \mathbb{R}^n$ квазиконформен тогда и только тогда, когда существует такая постоянная c, что*

$$\frac{1}{c}\|u\|_{\mathrm{BMO}(G')} \leqslant \|u \circ f\|_{\mathrm{BMO}(G)} \leqslant c\|u\|_{\mathrm{BMO}(G')} \tag{14.5}$$

для любой подобласти $G \subset D$ и любой непрерывной в $G' = f(G)$ функции u.

Теорема 14.3 и утверждение о необходимости в теореме 14.4 следуют из того, что $g = |J_f|$ удовлетворяет обратному неравенству Гёльдера (13.5). Утверждение о достаточности в теореме 14.4 — это принадлежащая Астале [A9] вариация исходного результата Райманна.

Теорема 14.4 характеризует квазиконформные отображения как гомеоморфизмы, сохраняющие класс ВМО. Следующий результат характеризует квазидиски в терминах свойства продолжаемости ВМО-функций.

14.6. **Теорема.** *Если D — односвязная область гиперболического типа в $\mathbb{R}^2$, то каждая функция $u \in \mathrm{BMO}(D)$ продолжается до функции класса $\mathrm{BMO}(\mathbb{R}^2)$ тогда и только тогда, когда D — квазидиск.*

Далее, лучший показатель степени интегрируемости якобиана, а также оценка искажения площади для плоских квазиконформных отображений тесно связаны с точными константами в двух неравенствах для преобразования Бёрлинга

$$Tg(x) = -\frac{1}{\pi} \int_{\mathbb{R}^2} \frac{g(y)}{(x-y)^2}\,dm. \tag{14.7}$$

Например, T является ограниченным оператором в $L^p(\mathbb{R}^2)$, причём для $1 < p < \infty$

$$\|T\|_p = \sup_g \frac{\|Tg\|_{L^p(\mathbb{R}^2)}}{\|g\|_{L^p(\mathbb{R}^2)}} \geqslant \max\left(p-1, \frac{1}{p-1}\right), \tag{14.8}$$

а $\|T\|_2 = 1$. Есть некоторые основания предположить, что

$$\lim_{p\to\infty} \frac{1}{p}\|T\|_p = 1. \tag{14.9}$$

Если это так, то отсюда получается точная верхняя граница $p(K, 2) = K/K - 1$ для степени интегрируемости якобиана плоского квазиконформного отображения, обсуждавшаяся в п. 13 [I1].

Далее, можно показать, что существуют постоянные a и b, такие что

$$\int_{\mathbb{B}^2} |T\chi_E(x)|\,dm \leqslant am(E)\log(\pi/m(E)) + bm(E) \tag{14.10}$$

для любого измеримого множества $E \subset \mathbb{B}^2$. Это неравенство в комбинации с теоремой 11.3 позволяет доказать, что для любого K-квазиконформного отображения $f\colon \mathbb{B}^2 \to \mathbb{B}^2$, нормированного условием $f(0) = 0$, и для любого измеримого множества $E \subset \mathbb{B}^2$ выполнено неравенство

$$\frac{m(f(E))}{\pi} \leqslant c\left(\frac{m(E)}{\pi}\right)^{K-a}, \tag{14.11}$$

причём $c = c(K) = 1 + O(K-1)$ при $K \to 1$ [G11]. Более того, эти рассуждения можно обратить и показать, что если (14.11) выполняется с некоторой постоянной a, то выполняется также и (14.10). Предположительно оба они справедливы с $a = 1$. Если это так, то отсюда снова следует, что $p(K, 2) = K/K - 1$.

Наконец, проблему квазиконформной эквивалентности областей можно переформулировать в терминах алгебр функций. Пусть D — область в $\mathbb{R}^n$. Обозначим через $A(D)$ алгебру функций $u \in C(D) \cap W_n^1(D)$ с нормой

$$\|u\| = \|u\|_{L^\infty(D)} + \|\nabla u\|_{L^n(D)}. \tag{14.12}$$

Это так называемая *алгебра Ройдена* в D. Справедлив следующий результат [L5, L6]:

14.13. **Теорема.** *Области D и D' в $\mathbb{R}^n$ квазиконформно-эквивалентны тогда и только тогда, когда алгебры $A(D)$ и $A(D')$ изоморфны.*

О структуре алгебр Ройдена известно мало, и может статься, что геометрические методы, используемые в теории квазиконформной эквивалентности, больше дадут для этих алгебр, чем наоборот.

15. Квазиизометрии и упругость. Отображение $f\colon E\subset \mathbb{R}^n\to\mathbb{R}^n$ называется *L-квазиизометрией* множества E, если

$$\frac{1}{L}|x_1-x_2|\leqslant|f(x_1)-f(x_2)|\leqslant L|x_1-x_2| \tag{15.1}$$

для любых $x_1,\ x_2\in E$, и *локальной L-квазиизометрией* E, если для любого $L'>L$ каждая точка $x\in E$ обладает окрестностью U, такой что f является L'-квазиизометрией $E\cap U$.

Если f — квазиизометрия области D, то f квазиконформно, в силу (1.2) и (1.4); пример отображения (13.6) показывает, что обратное неверно. Тем не менее квазиконформные отображения возникают в вопросах, связанных с продолжением и инъективностью таких отображений.

15.2. **Теорема.** *Если $n\neq 4$, то произвольная квазиизометрия f множества $E\subset\mathbb{R}^n$ продолжается до квазиизометрии $\mathbb{R}^n$ тогда и только тогда, когда f допускает квазиконформное продолжение на всё $\mathbb{R}^n$.*

Эта теорема [T6] дает критерий продолжимости в терминах отображения f. Существует и критерий в терминах множества E, когда E — жорданова кривая [G9]:

15.3. **Теорема.** *Пусть C — жорданова кривая в $\mathbb{R}^2$. Для того чтобы любая квазиизометрия C продолжалась до квазиизометрии $\mathbb{R}^2$, необходимо и достаточно, чтобы кривой C была квазиокружность.*

Для каждой области $D\subset\mathbb{R}^n$ обозначим через $L(D)$ верхнюю грань таких чисел $L\geqslant 1$, что любая локальная L-квазиизометрия f области D инъективна в D. Постоянной $L(D)$ можно дать физическую интерпретацию, если считать D упругим телом, а f — его деформацией под действием некоторого силового поля. Требование локальной L-квазиизометричности f — это ограничение на деформации, возникающие в D, и $L(D)$ есть мера критической деформации в D, предшествующей «коллапсу».

Об этой константе известно мало — лишь то, что $2^{1/4}\leqslant L(D)\leqslant 2^{1/2}$, когда D — шар или полупространство [J2]. Однако можно описать достаточно большой класс областей на плоскости, для которых $L(D)>1$ [G8]:

15.4. **Теорема.** *Если D — односвязная собственная подобласть в $\mathbb{R}^2$, то $L(D)>1$ тогда и только тогда, когда D — квазидиск.*

15.5. **Следствие.** *Если f локальная L-квазиизометрия ограниченной односвязной области в $\mathbb{R}^2$ и $L<L(D)$, то f продолжается до M-квазиизометрии $\mathbb{R}^2$, где $M=M(L,L(D))$.*

Следствие 15.5 говорит о том, что облик деформируемого односвязного плоского упругого тела, грубо говоря, остаётся таким, каким он был до деформирования, до тех пор пока деформации не достигнут некоторой критической величины. Было бы интересно получить аналогичный результат и в высших размерностях.

16. Комплексный анализ. Квазиконформные отображения порой возникают в таких теоретико-функциональных проблемах, которые на первый взгляд не имеют ничего общего с этим классом. Превосходным примером служит теорема Тайхмюллера [T1], связывающая экстремальные квазиконформные отображения между двумя римановыми поверхностями с квадратичными дифференциалами на этих поверхностях.

Чтобы привести более элементарный пример, возьмём функцию f, мероморфную в односвязной области $D\subset\overline{\mathbb{R}}^2$ гиперболического типа, и положим

$$S_f=\left(\frac{f''}{f'}\right)'-\frac{1}{2}\left(\frac{f''}{f'}\right)^2. \tag{16.1}$$

По теореме Нехари [N1], если D — диск или полуплоскость и $|S_f|\leqslant 2\rho_D^2$ в D, то f инъективна. Здесь ρ_D — гиперболическая метрика в D, задаваемая соотношением

$$\rho_D(z)=|g'(z)|(1-|g(z)|^2)^{-1}, \tag{16.2}$$

где $g\colon D\to\mathbb{B}^2$ — конформное отображение. Естественно спросить, для каких ещё областей D имеет место подобный результат. То есть для каких D верно, что $\sigma(D)>0$, где $\sigma(D)$ — супремум чисел $a\geqslant 0$, таких что функция f инъективна, если она мероморфна в D и $|S_f|\leqslant a\rho_D^2$ в D?

В следующем ответе на этот вопрос, дающем новую характеризацию предложенного Берсом универсального пространства Тайхмюллера [B5], фигурируют квазиконформные отображения:

16.3. **Теорема.** *$\sigma(D)>0$ тогда и только тогда, когда D — квазидиск.*

17. Дифференциальная геометрия и топология. Некоторые результаты, упомянутые в разделе II, имеют важные приложения в дифференциальной геометрии. Например, теорема 1.5 и необходимые условия из теоремы 10.3 являются ключевыми моментами первоначального доказательства теоремы Мостова о жесткости [M5], которая звучит так:

17.1. **Теорема.** *Если M и M' — диффеоморфные n-мерные компактные римановы многообразия постоянной отрицательной кривизны, то при $n > 2$ они конформно-эквивалентны.*

Аналогично свойство равностепенной непрерывности квазиконформных отображений, о котором идет речь в теореме 13.1, является важным инструментом при доказательстве следующей гипотезы Лихнеровича [L4]:

17.2. **Теорема.** *Если $n \geqslant 2$, а M — компактное риманово n-мерное многообразие, не являющееся конформно-эквивалентным сфере, то группа $C(M)$ конформных отображений M на себя компактна в топологии равномерной сходимости.*

Работа Эрла и Илза [E1] о группе диффеоморфизмов поверхности и данное Берсом доказательство [B6] теоремы Тёрстона о классификации отображений поверхности на себя иллюстрируют применения квазиконформных отображений в топологии поверхностей. Сулливан показал [S1], что для квазиконформных отображений теорема Шёнфлиса, гипотеза кольца (annulus conjecture) и теорема о компоненте (component problem) справедливы во всех размерностях без исключения. Результаты этой фундаментальной работы наводят на мысль, что квазиконформные отображения — важное промежуточное звено между гомеоморфизмами и диффеоморфизмами.

ЛИТЕРАТУРА

[A1] L. V. Ahlfors, Zur theorie der Überlagerungsflächen, Acta Math. **65** (1935), 157—194.

[A2] —, On quasiconformal mappings, J. Analyse Math. **3** (1953/54), 1—58.

[A3] —, Quasiconformal reflections, Acta Math. **109** (1963), 291—301.

[A4] —, Extension of quasiconformal mappings from two to three dimensions, Proc. Nat. Acad. Sci. U. S. A. **51** (1964), 768—771.

[A5] —, Finitely generated Kleinian groups, Amer. J. Math. **86** (1964), 413—429.

[A6] —, Quasiconformal mappings, Teichmüller spaces, and Kleinian groups, Proc. Internat. Congr. Math. (Helsinki, 1978), Acad. Sci. Fennica, Helsinki, 1980, pp. 71—84.

[A7] L. V. Ahlfors and L. Bers, Riemann's mapping theorem for variable metrics, Ann. of Math. **72** (1960), 385—404.

[A8] L. V. Ahlfors and A. Beurling, Conformal invariants and function-theoretic null sets, Acta Math. **83** (1950), 101—129.

[A9] K. Astala, A remark on quasi-conformal mappings and BMO-functions, Michigan Math. J. **30** (1983), 209—212.

[B1] A. Baernstein II and J. J. Manfredi, Topics in quasiconformal mapping, Topics in Modern Harmonic Analysis, Instituto Nazionale di Alta Mathematica, Roma, 1983, pp. 819—862.

[B2] L. Bers, Quasiconformal mappings and Teichmüller's theorem, Analytic Functions, Princeton Univ. Press, Princeton, N. J., 1960, pp. 89—119.

[B3] —, Uniformization by Beltrami equations, Comm. Pure Appl. Math. **14** (1961), 215—228.

[B4] —, Uniformization, moduli, and Kleinian groups, Bull. London Math. Soc. **4** (1972), 257—300.

[B5] —, Quasiconformal mappings, with applications to differential equations, function theory and topology, Bull. Amer. Math. Soc. **83** (1977), 1083—1100.

[B6] —, An extremal problem for quasiconformal mappings and a problem of Thurston, Acta Math. **141** (1978), 73—98.

[B7] —, Finite dimensional Teichmüller spaces and generalizations, Bull. Amer. Math. Soc. **5** (1981), 131—172.

[B8] A. Beurling and L. V. Ahlfors, The boundary correspondence under quasiconformal mappings, Acta Math. **96** (1956), 125—142.

[B9] Боярский Б. Обобщённые решения системы дифференциальных уравнений первого порядка эллиптического типа с разрывными коэффициентами.— Мат. сб., 1957, т. 43, с. 451—503.

[B10] M. Brown, Locally flat embeddings of topological manifolds, Ann. Math. **75** (1962), 331—341.

[C1] L. Carleson, The extension problem for quasiconformal mappings, Contributions to Analysis, Academic Press, New York, 1974, pp. 39—47.

[D1] A. Douady and C. J. Earle, Conformally natural extension of homeomorphisms of the circle, Acta Math. **157** (1986), 23—48.

[D2] D. Drasin, The inverse problem of the Nevanlinna theory, Acta Math. **138** (1977), 83—151.

[E1] C. J. Earle and J. Eells, A fibre bundle description of Teichmüller theory, J. Differential Geom. **3** (1969), 19—43.

[E2] C. J. Earle and S. Nag, Conformally natural reflections in Jordan curves with applications to Teichmüller spaces, *in*: Holomorphic functions and moduli, Vol. II (Berkeley, CA, 1986), Math. Res. Inst. Publ., II, Springer, New York — Berlin, 1988, pp. 179—194.

[E3] T. Erkama, Group actions and extension problems for maps of balls, Ann. Acad. Sci. Fenn. Ser. A I Math. **556** (1973), 1—31.

[F1] C. Fefferman, Characterizations of bounded mean oscillation, Bull. Amer. Math. Soc. **77** (1971), 587—588.

[F2] M. H. Freedman and R. Skora, Strange actions of groups on spheres, J. Differential Geom. **25** (1987), 75—98.

[F3] B. Fuglede, Extremal length and functional completion, Acta Math. **98** (1957), 171—219.

[G1] D. B. Gauld and J. Väisälä, Lipschitz and quasiconformal flattening of spheres and cells, Ann. Acad. Sci. Fenn. Ser. A I Math. **4** (1978/79), 371—382.

[G2] F. W. Gehring, Symmetrization of rings in space, Trans. Amer. Math. Soc. **101** (1961), 499—519.

[G3] —, Rings and quasiconformal mappings in space, Trans. Amer. Math. Soc. **103** (1962), 353—393.

[G4] —, Extension of quasiconformal mappings in three space, J. Analyse Math. **14** (1965), 171—182.

[G5] —, Extension theorems for quasiconformal mappings in n-space, J. Analyse Math. **19** (1967), 149—169.

[G6] —, The L^p-integrability of the partial derivatives of a quasiconformal mapping, Acta Math. **130** (1973), 265—277.

[G7] —, Quasiconformal mappings, Complex Analysis and its Applications. II, International Atomic Energy Agency, Vienna, 1976, pp. 213—268.

[G8] —, Injectivity of local quasi-isometries, Comment. Math. Helv. **57** (1982), 202—220.

[G9] —, Extension of quasiisometric embeddings of Jordan curves, Complex Variables Theory Appl. **5** (1986), 245—263.

[G10] F. W. Gehring and G. J. Martin, Discrete quasiconformal groups. I, Proc. London Math. Soc. (3) **55** (1987), 331—358.

[G11] F. W. Gehring and E. Reich, Area distortion under quasiconformal mappings, Ann. Acad. Sci. Fenn Ser. A I Math. **388** (1966), 1—15.

[G12] F. W. Gehring and J. Väisälä, The coefficients of quasiconformality of domains in space, Acta Math. **114** (1965), 1—70.

[G13] M. Giaquinta, Multiple integrals in the calculus of variations and nonlinear elliptic systems, Ann. of Math. Studies, Vol. 105, Princeton Univ. Press, Princeton, N. J., 1983.

[G14] H. Grötzsch, Über möglichst konforme Abbildungen von schlichten Bereichen, Ber. Verh. Sächs. Akad. Wiss. Leipzig **84** (1932), 114—120.

[I1] T. Iwaniec, Extremal inequalities in Sobolev spaces and quasiconformal mappings, Z. Anal. Anwendungen **1** (1982), 1—16.

[I2] —, On L^p-integrability in PDE's and guasiregular mappings for large exponents, Ann. Acad. Sci. Fenn. Ser. A I Math. **7** (1982), 301—322.

[J1] F. John, Rotation and strain, Comm. Pure Appl. Math. **14** (1961), 391—413.

[J2] —, On quasi-isometric mappings, II, Comm. Pure Appl. Math. **22** (1969), 265—278.

[J3] F. John and L. Nirenberg, On functions of bounded mean oscillation, Comm. Pure Appl. Math. **14** (1961), 415—426.

[J4] P. W Jones, Extension theorems for BMO, Indiana Univ. Math. J. **29** (1980), 41—66.

[K1] M. Kiikka, Diffeomorphic approximation of quasiconformal and quasisymmetric homeomorphisms, Ann. Acad. Fenn. Sci. Ser. A I Math. **8** (1983), 251—256.

[K2] I. Kra, On the Nielsen-Thurston-Bers type of some self-maps of Riemann surfaces, Acta Math. **146** (1981), 231—270.

[K3] T. Kuusalo, Quasiconformal mappings without boundary extensions, Ann. Acad. Sci. Fenn. Ser. A I Math. **10** (1985), 331—338.

[L1] Лаврентьев М. А. Об одном классе непрерывных отображений. — Мат. сб., 1935, т. 42, с. 407—423.

[L2] —, Об одном дифференциальном признаке гомеоморфных преобразований трёхмерных областей. — ДАН СССР, 1938, т. 20, с. 241—242.

[L3] O. Lehto and K. I. Virtanen, Quasiconformal mappings in the plane, Springer-Verlag, 1973.

[L4] J. Lelong-Ferrand, Transformations conformes et quasi-conformes des variétés riemanniennes compactes (Démonstration de la conjoncture de A. Lichnerowicz), Acad. Roy. Belg. Cl. Sci. Mém. Collect. **39** (1971), 1—44.

[L5] —, Étude d'une classe d'applications liées à des homomorphismes d'algèbres de fonctions, et généralisant les quasi conformes, Duke Math. J. **40** (1973), 163—186.

[L6] L. G. Lewis, Quasiconformal mappings and Royden algebras in space, Trans. Amer. Math. Soc. **158** (1971), 481—492.

[L7] C. Loewner, On the conformal capacity in space, J. Math. Mech. **8** (1959), 411—414.

[M1] O. Martio, S. Rickman, and J. Väisälä, Topological and metric properties of quasi-regular mappings, Ann. Acad. Sci. Fenn. Ser. A I Math. **488** (1971), 1—31.

[M2] B. Mazur, On embeddings of spheres, Bull. Amer. Math. Soc. **65** (1959), 59—65.

[M3] Меньшов Д. Е. Об одном обобщении теоремы М. Х. Бора. — Мат. сб., 1937, т. 44, с. 339—354.

[M4] C. B. Morrey, On the solutions of quasi-linear elliptic partial differential equations, Trans. Amer. Math. Soc. **43** (1938), 126—166.

[M5] G. D. Mostow, Quasi-conformal mappings in n-space and the rigidity of hyperbolic space forms, Inst. Hautes Études Sci. Publ. Math. **34** (1968), 53—104.

[N1] Z. Nehari, The Schwarzian derivative and schlicht functions, Bull. Amer. Math. Soc. **55** (1949), 545—551.

[N2] R. Nirenberg, On quasi-pseudoconformality in several complex variables, Trans. Amer. Math. Soc. **127** (1967), 233—240.

[R1] H. M. Reimann, Functions of bounded mean oscillation and quasiconformal mappings, Comment. Math. Helv. **49** (1974), 260—276.

[R2] Решетняк Ю. Г. Пространственные отображения с ограниченным искажением. — Сиб. мат. ж., 1967, т. 8, с. 629—658.

[R3] —, Теорема Лиувилля о конформных отображениях при минимальных предположениях регулярности. — Сиб. мат. ж., 1967, т. 8, с. 835—840.

[R4] —, Оценки устойчивости в теореме Лиувилля и L^p-интегрируемость производных квазиконформных отображений. — Сиб. мат. ж., 1976, т. 17, с. 868—896.

[R5] S. Rickman, Extension ver quasiconformally equivalent curves, Ann. Acad. Sci. Fenn. Ser. A I Math. **436** (1969), 1—12.

[R6] —, On the number of omitted values of entire quasiregular mappings, J. Analyse Math. 37 (1980), 100—117.

[S1] D. Sullivan, Hyperbolic geometry and homeomorphisms, Geometric Topology, Academic Press, New York, 1979, pp. 543—555.

[S2] —, On the ergodic theory at infinity of an arbitrary discrete group of hyperbolic motions, Riemann Surfaces and Related Topics: Proceedings of the 1978 Stony Brook Conference, Ann. of Math. Studies, No. 97, Princeton Univ. Press, Princeton, N. J., 1981, pp. 465—496.

[S3] —, Quasiconformal homeomorphisms and dynamics I, Solution of the Fatou-Julia problem on wandering domains, Ann. of Math. **122** (1985), 401—418.

[S4] —, Quasiconformal homeomorphisms and dynamics II: Structural stability implies hyperbolicity for Kleinian groups, Acta Math. **155** (1985), 243—260.

[T1] O. Teichmüller, Extremale quasikonforme Abbildungen und quadratische Differentiale, Abh. Preuss. Akad. Wiss. Mat.-Nat. Kl. **22** (1940), 1—197.

[T2] P. Tukia, On two-dimensional quasiconformal groups, Ann. Acad. Sci. Fenn. Ser. A I Math. **5** (1980), 73—78.

[T3] —, A quasiconformal group not isomorphic to a Möbius group, Ann. Acad. Sci. Fenn. Ser. A I Math. **6** (1981), 149—160.

[T4] —, Quasiconformal extension of quasisymmetric mappings compatible with a Möbius group, Acta Math. **154** (1985), 153—193.

[T5] P. Tukia and J. Väisälä, Quasiconformal extension from dimension n to $n+1$, Ann. of Math. **115** (1982), 331—348.

[T6] —, Bilipschitz extensions of maps having quasiconformal extensions, Math. Ann. **269** (1984), 561—572.

[V1] J. Väisälä, On quasiconformal mappings in space, Ann. Acad. Sci. Fenn. Ser. A I Math. **298** (1961), 1—36.

[V2] —, On quasiconformal mappings of a ball, Ann. Acad. Sci. Fenn. Ser. A I Math. **304** (1961), 1—7.

[V3] —, Lectures on n-dimensional quasiconformal mappings, Lectures Notes in Math., Vol. 229, Springer-Verlag, 1971.

[V4] —, A survey of quasiregular maps in $\mathbf{R}^n$, Proc. Internat. Congr. Math. (Helsinki, 1978), Acad. Sci. Fennica, Helsinki, 1980, pp. 685—691.

[V5] —, Quasimöbius maps, J. Analyse Math. **44** (1984/85), 218—234.

НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ МАТЕМАТИЧЕСКИЙ ИНСТИТУТ, БЕРКЛИ, КАЛИФОРНИЯ 94720, США
МИЧИГАНСКИЙ УНИВЕРСИТЕТ, АНН-АРБОР, МИЧИГАН 48109, США

ГИБКАЯ И ЖЁСТКАЯ СИМПЛЕКТИЧЕСКАЯ ГЕОМЕТРИЯ

Михаил Громов [1]

1. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ, ПРИМЕРЫ И ЗАДАЧИ

1.1. Симплектические формы и многообразия. Внешняя дифференциальная 2-форма ω на гладком многообразии V называется *невырожденной*, если ассоциированный гомоморфизм между касательным и кокасательным расслоениями V, обозначаемый через I_ω: $T(V)\rightarrow T^*(V)$ и определяемый формулой $I_\omega(\tau)=\omega(\tau,\cdot)$, является изоморфизмом. В этом случае размерность многообразия V обязательно чётная (мы предполагаем, что V конечномерно и все его связные компоненты имеют одинаковую размерность), скажем $\dim V=m=2n$, и внешняя степень ω^n (которая является формой старшей степени на V) не обращается в нуль на V. Обратно, если ω^n не обращается в нуль, то ω невырожденна. Например, каждая форма (ориентированной) площади на *поверхности* невырожденна.

Форма ω на V называется *симплектической*, если она невырожденна и замкнута, т. е. $d\omega=0$. Пара (V,ω) называется при этом *симплектическим многообразием*.

Примеры. Каждая поверхность с формой площади является симплектическим многообразием. Если (V_i,ω_i) — такие поверхности, $i=1,\ldots,n$, то декартово произведение $(V,\omega)=(V_1\times\times V_2\times\ldots\times V_n,\ \omega_1\oplus\omega_2\oplus\ldots\oplus\omega_n)$ есть $2n$-мерное симплектическое многообразие. *Симплектическая площадь* $\omega(S)=\int\limits_S\omega$ всякой поверхности S в этом многообразии V равна сумме (ориентированных!) площадей проекций $S\rightarrow V_i$.

Важный частный случай — *симплектическое пространство* $\left(\mathbb{R}^{2n},\ \omega=\sum\limits_{i=1}^{n}dx_i\wedge dy_i\right)$, т. е. сумма n экземпляров (x,y)-плоскости $\mathbb{R}^2$ с обычной формой площади $dx\wedge dy$.

---

[1] Michael Gromov, Soft and hard symplectic geometry, ICM86, pp. 81—98.

Менее очевидный пример — комплексное проективное пространство $\mathbb{C}P^n$, в котором имеется единственная (с точностью до скалярного множителя) 2-форма ω, инвариантная относительно действия унитарной группы $\mathrm{U}(n+1)$ в $\mathbb{C}P^n$. Легко видеть, что эта форма является симплектической, а симплектическая площадь каждой поверхности $S \subset \mathbb{C}P^n$ равна среднему числу точек пересечения S (считая с алгебраической кратностью) с гиперплоскостями $p \subset \mathbb{C}P^n$:

$$\omega(S) = \int_P \#(S \cap p)\, dp,$$

где P — (двойственное) проективное пространство $(\approx \mathbb{C}P^n)$ гиперплоскостей p в $\mathbb{C}P^n$ и dp есть $U(n+1)$-инвариантная мера на P.

1.2. Симплектические (диффео)морфизмы. Отображение $f\colon (V_1, \omega_1) \to (V_2, \omega_2)$ класса C^1 называется *симплектическим*, если поднятие формы ω_2 равно ω_1, т. е. $f^*(\omega_2) = \omega_1$. Такое отображение обязательно является *иммерсией*, т. е. дифференциал $D_f\colon T(V_1) \to T(V_2)$ инъективен на касательном пространстве $T_v(V_1)$ для всех $v \in V_1$.

Группа $\operatorname{Symp} V$ симплектических диффеоморфизмов симплектического многообразия $V = (V, \omega)$ бесконечномерна, если $\dim V \geqslant 2$. Действительно, пространство Ω замкнутых 2-форм на V по существу есть множество {1-формы/d (функции)}, в силу чего «функциональная размерность» Ω равна $m-1$ (здесь $m = \dim V$), тогда как «функциональная размерность» группы $\operatorname{Diff} V$ равна m, поскольку всякий диффеоморфизм $V \to V$ определяется m функциями на V. Следовательно, искомая «функциональная размерность» группы $\operatorname{Symp}(V, \omega)$, которая является подгруппой изотропии формы $\omega \in \Omega$ при естественном действии группы $\operatorname{Diff} V$ на Ω, равна единице. Это эвристическое рассуждение (которое на самом деле можно провести строго) наводит на мысль о некотором соответствии между функциями $V \to \mathbb{R}$ и симплектическими диффеоморфизмами многообразия V. (Заметим, что для «достаточно невырожденных» форм ω, степень которых заключена между 3 и $m-2$, группа автоморфизмов многообразия (V, ω) конечномерна и то же самое верно для *симметрических* дифференциальных форм степени $\geqslant 2$.)

Симплектические векторные поля. Векторное поле X на (V, ω) называется *симплектическим*, если производная Ли X_ω равна нулю. Так как

$$X_\omega = dI_\omega(X) \tag{1}$$

для всех полей X и замкнутых 2-форм ω, то условие $X_\omega = 0$ эквивалентно равенству $dI_\omega(X) = 0$. Поэтому для каждой *замкнутой* 1-формы l на V векторное поле $I_\omega^{-1}(l)$ является симплектическим. В частности, для любой гладкой функции (гамильтониана) $h\colon V \to \mathbb{R}$ поле $X = I_\omega^{-1}(dh)$, называемое *гамильтоновым*, симплектично. *Интегрируемые* гамильтоновы поля X (например, те, для которых h имеет компактный носитель) определяют однопараметрические подгруппы $X_t \subset \operatorname{Symp}(V, \omega)$. Этим подтверждается обширность группы $\operatorname{Symp} V$, предсказанная предыдущим рассуждением о размерности.

Другое свойство группы $\operatorname{Symp}(V, \omega)$, которое можно извлечь из рассмотрения гамильтоновых полей, состоит в *транзитивности* $\operatorname{Symp} V$ на k-кратных наборах попарно различных точек многообразия V для каждого $k = 1, 2, \ldots$, в предположении что V связно. В частности, группа $\operatorname{Symp} V$ транзитивна на V.

Следствие (Дарбу). *Каждые два симплектических многообразия V_1 и V_2 одинаковой размерности локально изоморфны.*

Доказательство. Если $U_i \subset V_i$, $i = 1, 2$, — достаточно малые окрестности, то, очевидно, существует *связное* симплектическое многообразие V', такое что U_i симплектически диффеоморфны некоторым окрестностям $U_i' \subset V'$.

Примеры симплектических диффеоморфизмов пространства $(\mathbb{R}^{2n}, \omega = \sum dx_i \wedge dy_i)$. (1) Любой параллельный перенос в $\mathbb{R}^{2n}$ является симплектическим.

(2) Положим $z_i = x_i + \sqrt{-1}\, y_i$ и отождествим $\mathbb{R}^{2n}$ с $\mathbb{C}^n$. Тогда форму ω можно выразить через эвклидово скалярное произведение формулой

$$\omega(\tau_1, \tau_2) = \langle \tau_1, \sqrt{-1}\, \tau_2 \rangle,$$

справедливой для всех (касательных) векторов τ_1 и τ_2 в $\mathbb{R}^{2n}$. Следовательно, ω инвариантна относительно унитарных преобразований пространства $\mathbb{C}^n$. На самом деле группа $\mathrm{Sp}(2n)$ *линейных* симплектических преобразований пространства $\mathbb{R}^{2n}$ *строго* содержит группу $\mathrm{U}(n)$, так как

$$\dim \mathrm{U}(n) = n^2 < n(2n+1) = \dim \mathrm{Sp}(2n).$$

(3) Разложим $\mathbb{R}^{2n}$ в сумму n экземпляров пространства $(\mathbb{R}^2, dx \wedge dy)$, и пусть f_i, $i = 1, \ldots, n$, — преобразования плоскости $\mathbb{R}^2$, сохраняющие площадь. Тогда декартова сумма преобразований f_i симплектична в $\mathbb{R}^{2n}$.

(4) Отождествим $\mathbb{R}^{2n}$ с тотальным пространством кокасательного расслоения пространства $\mathbb{R}^n$, снабжённого координа-

тами $x_1, \ldots, x_n$. Тогда естественное действие диффеоморфизмов пространства $\mathbb{R}^n$ на $T^*(\mathbb{R}^n) = \mathbb{R}^{2n}$ является симплектическим. Таким образом, группа $\operatorname{Diff} \mathbb{R}^n$ вкладывается в $\operatorname{Symp} \mathbb{R}^{2n}$.

Из диффеоморфизмов (1)—(4) и гамильтоновых диффеоморфизмов X_t, указанных выше, можно составлять всевозможные композиции. Это даёт симплектические диффеоморфизмы пространства $\mathbb{R}^{2n}$, которые могут показаться достаточно произвольными. Тем не менее не существует диффеоморфизма $f \in \operatorname{Symp} \mathbb{R}^{2n}$, переводящего U_1 в U_2, где U_1 и U_2 — открытые подмножества в $\mathbb{R}^{2n}$, такие что $\operatorname{Vol} U_1 > \operatorname{Vol} U_2$. Действительно, $\operatorname{Vol} U = \int\limits_U \omega^n$ — инвариант относительно группы $\operatorname{Symp} \mathbb{R}^{2n}$.

Задача. Имеются ли другие препятствия, кроме $\operatorname{Vol} U_1 \leqslant \operatorname{Vol} U_2$, для существования симплектических диффеоморфизмов пространства $\mathbb{R}^{2n}$, отображающих U_1 в U_2?

Легко видеть, что для диффеоморфизмов пространства $\mathbb{R}^{2n}$, *сохраняющих объём*, существенно новых препятствий нет. А именно, если U_1 — относительный компакт, множества $\mathbb{R}^{2n} \setminus U_1$ и U_2 связны и

$$\operatorname{Vol} U_1 < \operatorname{Vol} U_2 \qquad (*)$$

(строгое неравенство снимает не относящийся к делу вопрос о поведении отображений на границе), то найдётся гладкий (даже вещественно-аналитический) диффеоморфизм f пространства $\mathbb{R}^{2n}$, переводящий U_1 в U_2 и такой что $f^*(\omega^n) = \omega^n$. Однако для симплектических отображений новые препятствия есть, как это видно из следующего примера.

Пример. Пусть U_1 — стандартный шар радиуса r в $\mathbb{R}^{2n}$ и U_2 является ε-окрестностью линейного подпространства $L \subset \mathbb{R}^{2n}$, где $\dim L < n$. Тогда при $\varepsilon < r$ *не существует симплектического* диффеоморфизма, переводящего $U_1 \to U_2 \left(\omega = \sum\limits_{i=1}^n dx_i \wedge dy_i\right)$.

Доказательство (приведённое в § 4.1) опирается на геометрию *голоморфных кривых* относительно некоторой (неинтегрируемой) *почти-комплексной* структуры в $\mathbb{R}^{2n}$. «Гибкого» доказательства в настоящее время неизвестно.

Замечание. Если L — линейное n-мерное подпространство в $\mathbb{R}^{2n}$, заданное уравнениями $x_i = 0$, $i = 1, \ldots, n$, то единичный шар U_1 переходит в ε-окрестность U_2 множества L при симплектическом отображении $(x_i, y_i) \to (\varepsilon x_i, \varepsilon^{-1} y_i)$. Но если L задано уравнениями $x_1 = 0$ и $y_1 = 0$ (здесь $\dim L = 2n - 2$), то в случае $r > \varepsilon$ не существует симплектического диффеоморфизма $U_1 \to U_2$ (см. § 4.1).

1.3. Изотропные иммерсии и лагранжевы подмногообразия.

Отображение $f\colon W \to (V, \omega)$ класса C^1 называется *изотропным*, если $f^*(\omega) = 0$. Нас будет особенно интересовать случай, когда $\dim W = n$ $(2n = \dim V)$ и f — иммерсия. Такие отображения f называются *лагранжевыми иммерсиями*. Аналогично подмногообразие $W \subset V$ называется *лагранжевым, если* $\omega | W = 0$.

Примеры. (1) Пусть $f\colon (V_1, \omega_1) \to (V_2, \omega_2)$ — симплектическое отображение и $(V, \omega) = (V_1 \times V_2, \omega_1 \oplus -\omega_2)$. Тогда график $\Gamma_f\colon V_1 \to V$, где $\Gamma_f(v_1) = (v_1, f(v_1))$, является изотропной иммерсией. На самом деле график каждого отображения является вложением и $\Gamma_f(V_1)$ — лагранжево подмногообразие в V, если $\dim V_1 = \dim V_2$.

(2) Если $\dim V = 2$ и $W \subset V$ имеет размерность $\dim W = 1$, то, очевидно, W — лагранжево подмногообразие для любой формы площади ω на V. Декартово произведение n экземпляров таких многообразий

$$W_1 \times \ldots \times W_n \subset (V_1 \times \ldots \times V_n, \ \omega_1 \oplus \ldots \oplus \omega_n)$$

также лагранжево. В частности, тор $T^n = \{x_i, y_i \,|\, x_i^2 + y_i^2 = 1\}$ — лагранжево подмногообразие в $\left(\mathbb{R}^{2n}, \sum\limits_{i=1}^n dx_i \wedge dy_i\right)$.

(3) Пусть V — тотальное пространство кокасательного расслоения $V = T^*(X)$ некоторого n-мерного многообразия X. Обозначим через σ *каноническую* 1-форму на V, определённую тождеством

$$\alpha^*(\sigma) = \alpha, \qquad (**)$$

где $\alpha\colon X \to T^*(X) = V$ — произвольное C^1-сечение, которое в правой части $(**)$ рассматривается как 1-форма на X. Легко видеть, что форма $\omega = d\sigma$ симплектична и для $X = \mathbb{R}^n$ многообразие V симплектически изоморфно пространству $\left(\mathbb{R}^{2n}, \sum\limits_{i=1}^n dx_i \wedge dy_i\right)$. Далее, сечение $X \to T^*(X) = V$ лагранжево, если и только если соответствующая 1-форма на X замкнута. В частности, отображение $dh\colon X \to T^*(X)$ лагранжево для каждой гладкой функции h на X. (Заметим, что всякое лагранжево подмногообразие $W \subset T^*(X)$, проекция которого на X является *диффеоморфизмом* W *на* X, есть многообразие такого типа, причём $W=\alpha(X)$ для единственной 1-формы α: $X\to T^*(X)$ $=V$.)

(4) Вещественное проективное пространство $\mathbb{R}P^n\subset\mathbb{C}P^n$ является лагранжевым многообразием относительно указанной выше $U(n+1)$-инвариантной формы ω в $\mathbb{C}P^n$. Если $V\subset\mathbb{C}P^n$ — неособое комплексное алгебраическое подмногообразие, то, очевидно, индуцированная форма $\omega'=\omega|V$ невырожденна и, следовательно, симплектична на V. Более того, если V определено над полем $\mathbb{R}$, то его *вещественная часть* $W=V\cap\mathbb{R}P^n$ изотропна в (V,ω'). Многообразие W будет лагранжевым, если оно неособо и $\dim W=\dim_{\mathbb{C}}V$.

Как мы увидим в § 2.2, задача о существовании лагранжевых иммерсий φ: $W\to V$ для данных W и $V=(V,\omega)$ относится к гибкой геометрии. Что же касается похожей задачи о возможной топологии образа $\varphi(W)\subset V$, то (на данный момент) она представляется имеющей жёсткую природу.

Пример (см. § 3.5). Для каждой лагранжевой иммерсии φ *замкнутого* многообразия W в пространство $\mathbb{R}^{2n}=\mathbb{C}^n$:

$$\varphi: W\to\left(R^{2n},\ \omega=\sum_{i=1}^{n}dx_i\wedge dy_i\right)$$

существует непостоянный голоморфный диск в $\mathbb{C}^n$ с границей в $\varphi(W)$.

Отсюда легко следует, что относительный когомологический класс

$$[\omega]\in H^2(\mathbb{R}^{2n},\ W,\ \mathbb{R})$$

отличен от нуля. В частности, если $H^1(W;\mathbb{R})=0$, то W не допускает лагранжева *вложения* в $(\mathbb{R}^{2n},\omega)$.

2. СИМПЛЕКТИЧЕСКИЕ ИММЕРСИИ И ВЛОЖЕНИЯ

2.1. Топологические препятствия для симплектических иммерсий. Рассмотрим два симплектических многообразия (V,ω) и (W,ω') и выясним, когда данное непрерывное отображение φ: $W\to V$ гомотопно симплектическому отображению f: $W\to V$. Имеются два очевидных препятствия к существованию отображения f. Первое связано с тем, что отображение φ должно определять соответствие когомологических классов форм ω и ω'. То есть гомоморфизм φ^*: $H^2(V;\mathbb{R})\to H^2(W;\mathbb{R})$ должен переводить $[\omega]$ в $[\omega']$.

Для того чтобы предъявить второе препятствие, заметим, что дифференциал отображения f, представляющий собой послойно-линейное отображение касательных расслоений D_f: $T(W)\to T(V)$, *симплектичен*, поскольку симплектично отображение f. Здесь непрерывное послойно-линейное отображение Δ: $T(W)\to T(V)$ мы называем симплектическим, если $\Delta^*(\omega)$ $=\omega'$. (Отметим, что $f^*(\omega)=D_f^*(\omega)$ по определению $f^*(\omega)$.)

Теперь рассмотрим пространство $\{\Delta\}$ всех симплектических послойно-линейных отображений $T(W)\to T(V)$ и пространство $\{\varphi\}$ непрерывных отображений $W\to V$. Легко видеть, что проекция $\{\Delta\}\to\{\varphi\}$, которая сопоставляет каждому отображению Δ индуцированное отображение φ, является *расслоением в смысле Серра*. Таким образом, из существования гомотопии между φ и f (где f поднимается до $D_f\in\{\Delta\}$) следует существование поднятия отображения φ до некоторого $\Delta\in\{\Delta\}$. Заметим, что такое поднятие задаётся некоторым *симплектическим гомоморфизмом* расслоений над многообразием W

$$\delta: (T(W),\ \omega')\to\varphi^*(T(V),\ \omega)$$

(здесь через $\varphi^*(\ )$ обозначается индуцированное расслоение, а симплектичность отображения δ понимается в обычном смысле), причём гомоморфизмы δ являются сечениями расслоения над W, слой которого над точкой $w\in W$ состоит из линейных симплектических отображений $T_w(W)\to T_{\varphi(w)}(V)$. В частности, каждое отображение φ будет подниматься до некоторого отображения Δ, если, например, многообразие W стягиваемо и $\dim W\leqslant\dim V$. (Приведённое рассуждение основывается только на невырожденности форм ω и ω'. Первое же наше условие $\varphi^*[\omega]=[\omega']$ требует лишь замкнутости этих форм.)

2.2. **Теорема об иммерсии** (см. [Gr2], [Gr4]). *Пусть* φ: $V=(V,\omega)\to W=(W,\omega')$ — *непрерывное отображение, которое поднимается до симплектического отображения* $T(W)\to T(V)$ *и удовлетворяет условию* $\varphi^*[\omega]=[\omega']$. *Тогда в следующих двух случаях существует симплектическое отображение* f: $W\to V$, *гомотопное отображению* φ:

(i) открытый случай: *многообразие* W *открыто* (*т. е. не имеет связной компоненты, являющейся замкнутым многообразием*).

(ii) случай избыточной размерности: $\dim W$ $<\dim V$.

Замечания. Эта теорема, очевидно, неверна, если W замкнуто, V открыто и $\dim W=\dim V$. На самом деле в этом случае не существует даже топологической иммерсии $W\to V$.

Отображение f может быть выбрано столь же гладким, как и формы ω и ω'. Например, если ω и ω' являются C^∞-гладкими (вещественно-аналитическими), то найдётся и отображение f класса C^∞ (вещественно-аналитическое).

Следствие А. *Симплектическое $2n$-мерное многообразие (W, ω') можно симплектически отобразить в пространство $(\mathbb{R}^{2n}, \omega = \sum dx_i \wedge dy_i)$, если и только если выполнены следующие три условия:*

(a) *W открыто;*

(b) *форма ω' точна;*

(c) *касательное расслоение $(T(W), \omega')$ является тривиальным* $\operatorname{Sp}(2n)$*-расслоением. (Это эквивалентно существованию линейно-независимых векторных полей X_i и Y_i на многообразии W, $i = 1, \ldots, n$, таких что $\omega'(X_i, Y_i) = 1$ и $\omega'(X_i, X_j) = \omega'(Y_i, Y_j) = \omega'(X_i, Y_j) = 0$ при $i \neq j$.)*

Заметим, что условия (a)—(c) выполнены для любого стягиваемого (например, гомеоморфного пространству $\mathbb{R}^{2n}$) многообразия W.

Следствие В. *Гладкое n-мерное многообразие X можно лагранжево иммерсировать в пространство $\mathbb{R}^{2n}$, если и только если комплексификация $T(X) \oplus \sqrt{-1}\, T(X)$ является тривиальным $\mathrm{GL}_n(\mathbb{C})$-расслоением над X.*

Здесь (как и в следствии А) утверждение «только если» тривиально, а «если» вытекает из следствия А, применённого к многообразию $W = T^*(X)$.

О доказательстве теоремы 2.2. Прежде всего, используя симплектичность отображения Δ: $T(W) \to T(V)$, легко построить семейство локальных симплектических иммерсий f_w: $U_w \to V$, где U_w — маленький шар в W вокруг точки w и f_w непрерывно зависит от $w \in W$. Затем из отображений f_w можно склеить искомое отображение f после подходящего «изгибания» этих локально определённых отображений f_w, делающего их согласованными на пересечениях $U_{w_1} \cap U_{w_2}$. Процедура склейки использует (весьма гибкие) методы топологических пучков (см. [Gr4]). Кажется маловероятным, чтобы такое отображение f могло быть построено методами жёсткого анализа.

2.3. Вложения. Как мы упоминали ранее, каждое симплектическое отображение $W \to V$ является *иммерсией*, но не обязательно *вложением*, т. е. отображением, осуществляющим гомеоморфизм на свой образ. (Например, если W — компакт, то вложением будет каждая иммерсия без двойных точек.)

Рассмотрим *гладкое вложение* φ: $W \to V$ (т. е. φ — гладкая иммерсия и вложение) и попытаемся C^∞-*изотопией* перевести φ в *симплектическое вложение* f: $W \to V$. Заметим, что дифференциал такой изотопии является гомотопией послойно-линейных и послойно-инъективных отображений Δ_t: $T(W) \to T(V)$, где $\Delta_0 = D_\varphi$, а $\Delta_1 = D_f$ симплектично.

Предположим, что для данного вложения φ существует гомотопия послойно-инъективных отображений Δ_t: $T(W) \to T(V)$ (которые для $t > 0$ не обязаны быть дифференциалами отображений $W \to V$), таких что $\Delta_0 = D_\varphi$ и Δ_1 симплектично.

Теорема А. *Если $\varphi^*[\omega] = [\omega']$, то в следующих двух случаях вложение φ изотопно симплектическому вложению f: $W \to V$:*

(i) *многообразие W открыто и* $\dim W < \dim V$;

(ii) $\dim W \leqslant \dim V - 4$.

Заметим, что (i) и (ii) исключают следующие два случая, допускаемые теоремой об иммерсии:

(a) W открыто и $\dim W = \dim V$;

(b) W замкнуто и $\dim W = \dim V - 2$.

Следствие В. *Если W стягиваемо и* $\dim W < \dim V$, *то каждое вложение $W \to V$ изотопно симплектическому. В частности, если W диффеоморфно пространству $\mathbb{R}^{2n-2}$, то существует симплектическое вложение* $(W, \omega') \to (\mathbb{R}^{2n}, \omega = \sum dx_i \wedge dy_i)$.

Замечание. Предыдущая теорема о вложении справедлива для *собственных* симплектических вложений, если $\dim W \leqslant \dim V - 4$. Поэтому каждое многообразие W, гомеоморфное пространству $\mathbb{R}^{2n-4}$, допускает *собственное* симплектическое вложение $(W, \omega') \to (\mathbb{R}^{2n}, \omega = \sum dx_i \wedge dy_i)$.

О доказательстве теоремы А. Симплектические вложения $W \to V$ получаются из (гибкой) геометрической конструкции Нэша [N1] (используемой для изометрических C^1-иммерсий римановых многообразий), применённой в контексте топологических пучков (см. [Gr4]).

О случаях (i) *и* (ii). Ограничения (i) и (ii) в теореме А не могут быть ослаблены. Это следует из (жёсткого) анализа голоморфных кривых на многообразиях V и W (см. § 4.2).

3. ГОЛОМОРФНЫЕ КРИВЫЕ НА ПОЧТИ-КОМПЛЕКСНЫХ МНОГООБРАЗИЯХ

3.1. Напомним, что каждая комплексная линейная структура в $\mathbb{R}^{2n}$ определяется автоморфизмом J пространства $\mathbb{R}^{2n}$, для которого $J^2 = -\mathrm{Id}$, что соответствует умножению на $\sqrt{-1}$. *Почти-комплексная структура* на многообразии V задаётся автоморфизмом касательного расслоения $T(V)$, обозначаемым через J или $\sqrt{-1}$ и таким, что $J^2 = -\mathrm{Id}$. Почти-комплексные многообразия (V, J), где $\dim V = 2$, называются *римановыми поверхностями.*

Отображение f: $(V_1, J_1) \to (V_2, J_2)$ класса C^1 одного почти-комплексного многообразия в другое называется *голоморфным*, если его дифференциал D_f является *комплексным* линейным отображением $T_v(V_1) \to T_{f(v)}(V_2)$ для всех $v \in V_1$. Это эквивалентно тождеству $D_f \circ J_1 = J_2 \circ D_f$.

Если $\dim V_1 > 2$, то для типичных J_1 и J_2 не существует непостоянного голоморфного отображения $V_1 \to V_2$. Но если V_1 — риманова поверхность, то по крайней мере локально существует, грубо говоря, столько же голоморфных отображений $V_1 \to V_2$, сколько их имеется в случае $V_1 = \mathbb{C}$ и $V_2 = \mathbb{C}^n$ ($\dim V_2 = 2n$). Действительно, уравнение $D_f \circ J_1 = J_2 \circ D_f$ является тогда *эллиптическим.* То есть его линеаризация представляет собой эллиптическое уравнение, (главный) символ которого в каждой точке изоморфен символу уравнения Коши — Римана $\bar{\partial} f = 0$ для отображений $\mathbb{C} \to \mathbb{C}^n$.

Определение. *Голоморфной* (параметризованной) *кривой* на почти-комплексном многообразии V называется голоморфное отображение $f : S \to V$ римановой поверхности S в многообразие V. Иногда мы будем забывать параметризацию и рассматривать *непараметризованные голоморфные кривые* $f(S) \subset V$.

3.2. Ручные многообразия. Скажем, что *замкнутая* 2-форма ω на V *приручает* почти-комплексную структуру J многообразия V, если эта форма *J-положительна*, т. е.

$$\omega(\tau, J\tau) > 0 \qquad (*)$$

для всех ненулевых касательных векторов $\tau \in T(V)$. Так как каждая J-положительная форма (очевидно) невырожденна, то ω — симплектическая форма.

Примеры. (A) *Калибровочные формы и кэлеровы многообразия.* Пусть g — риманова метрика на многообразии V, инвариантная относительно J. Тогда 2-форма

$$\omega = \omega(\tau_1, \tau_2) \underset{\text{def}}{=} -g(\tau_1, J\tau_2)$$

удовлетворяет условию $\omega(\tau, J\tau) = g(\tau, \tau)$. Следовательно, форма ω является J-положительной. Такая форма называется *калибрующей* (ср. [H-L]), если она замкнута.

В случае когда структура J *комплексная* (т. е. интегрируемая), калибровочные формы называются *кэлеровыми.* Например, индуцированная форма ω' (см. пример 4 из § 1.3) на каждом комплексном подмногообразии в $\mathbb{C}P^n$ (очевидно) кэлерова.

(B) *Выпуклые функции.* Гладкая функция $h : V \to \mathbb{R}$ называется *строго J-выпуклой* (или плюрисубгармонической), если её ограничение на любую голоморфную кривую в (V, J) является субгармонической функцией. Очевидно, что J-выпуклость функции h эквивалентна положительности точной 2-формы $\omega = dJdh$ на V и что достаточно малая окрестность U каждой точки $v \in V$ допускает строго J-выпуклую функцию $U \to \mathbb{R}$. Таким образом, всякое многообразие (V, J) может быть локально приручено некоторой формой ω. (Существование же локальной *калибрующей* формы налагает на J нетривиальное условие в частных производных.)

Замечание. Дифференциальные формы (любой степени), приручающие уравнения с частными производными, — главный (если не единственный) источник интегрально-дифференциальных неравенств, необходимых для априорных оценок и предельных теорем. Эти формы определены на пространствах струй (решений уравнений), причём они часто (например, в формулах Бохнера — Вайценбёкка) точны и инвариантны относительно подходящих (инфинитезимальных) симметрических групп. Подобным же образом выпуклые (в соответствующем смысле) функции на пространствах струй отвечают за принципы максимума. Бо́льшая часть жёсткого анализа уравнений с частными производными станет излишней, когда прояснятся алгебраические и геометрические структуры приручающих форм и соответствующих выпуклых функций. (С точки зрения теории уравнений с частными производными симплектическая геометрия является средством приручения пространства 0-струй решений уравнения Коши — Римана.)

3.3. Замкнутые голоморфные кривые. Если S — замкнутая риманова поверхность, то пространство Σ голоморфных отображений $f : S \to V = (V, J)$ локально конечномерно, поскольку (эллиптический!) оператор Коши — Римана фредгольмов. В общем случае пространство Σ далеко не компактно (даже если V — компакт), но оно допускает вполне приемлемую компактификацию, если V — замкнутое многообразие (которое может быть приручено некоторой замкнутой 2-формой ω). Это следует (см. [Gr3]) из очевидного неравенства

$$\operatorname{area} f(S) \leqslant \operatorname{const} \int_S f^*(\omega), \qquad (*)$$

где площадь (area) вычисляется в фиксированной римановой метрике на V. Заметим, что правая часть неравенства $(*)$

зависит лишь от гомологического класса $[f(S)]\in H_2(V)$ и что неравенство, аналогичное $(*)$, справедливо для графика

$$\Gamma_f : S\to V\times S.$$

Примеры. (A) Пусть V и S — римановы сферы S^2. Тогда пространство Σ_1 голоморфных отображений $f : S^2\to S^2$ степени 1 некомпактно. (Это группа $\mathrm{PGL}_2\mathbb{C}$.) Рассмотрим (образы) графиков этих отображений, являющиеся голоморфными кривыми $S_f\subset S^2\times S^2$. Если последовательность отображений $f_i\in\Sigma_1$ расходится, то существует подпоследовательность f_j, такая что кривые $S_{f_j}\subset S^2\times S^2$ сходятся (в обычном смысле) к приводимой кривой в $S^2\times S^2$ вида $(S^2\times s_2)\cup(s_1\times S^2)$ для некоторых точек s_1 и s_2 в S^2.

(A′) Пусть S — риманова поверхность и V — произвольное проективное алгебраическое многообразие. Тогда пространство Σ голоморфных отображений $f : S\to V$ фиксированного гомологического класса представляет собой квазипроективное многообразие, которое можно дополнить до проективного присоединением к Σ (графиков) приводимых кривых, полученных стягиванием некоторых окружностей (нулевых циклов) в S.

(B) Пусть $V=(\mathrm{SL}_2\mathbb{C})/\Lambda$ для некоторой кокомпактной решётки Λ в группе $\mathrm{SL}_2\mathbb{C}$. Выберем в Λ какой-нибудь элемент λ без кручения и положим $S=C_\lambda/Z_\lambda$, где $C_\lambda\subset\mathrm{SL}_2\mathbb{C}$ — централизатор элемента λ (заметим, что $C_\lambda\approx\mathbb{C}^*$), а Z_λ — порождённая им бесконечная циклическая группа. Ясно, что S есть тор, голоморфно отображающийся в V, причём его площадь можно сделать сколь угодно большой, если (голоморфно) подействовать на V группой $\mathrm{SL}_2\mathbb{C}$. Последнее объясняется тем, что комплексная структура многообразия V не является ручной.

Лемма Шварца. *Оценка* $(*)$ *площади отображения* f *позволяет, в принципе, контролировать поточечную норму дифференциала* D_f. *Например, каждое голоморфное отображение* f *единичного диска* $B\to\mathbb{C}$ *удовлетворяет неравенству*

$$\left|\frac{df}{dz}(0)\right|^2\leqslant\pi^{-1}\int\limits_B f^*(\omega), \qquad (**)$$

при условии что форма площади в $\mathbb{C}$ *равна* $\omega=dx\wedge dy$. *(На самом деле*

$$\left|\frac{df}{dz}(0)\right|^2\leqslant\pi^{-1}A,$$

где A *— площадь минимального односвязного подмножества в* $\mathbb{C}$, *содержащего образ* $f(B)\subset\mathbb{C}$.)

Неравенство $(*\,*)$ обобщается на все многообразия (V,J), прирученные *точными* формами, в силу чего (см. [Gr3]) пространство замкнутых голоморфных кривых S в каждом замкнутом оснащённом многообразии (V,J) *может быть компактифицировано присоединением (графиков) особых кривых, полученных стягиванием окружностей в* S.

Замечания. (a) Стягивание окружностей в *минимальных поверхностях* было обнаружено Саксом и Уленбекк [S-U] (ср. [Sch-Y]).

(b) По-видимому, минимальные поверхности очень полезны (для «гибких целей») при наличии приручающих форм более высокой степени (форм кривизны). Приведём два примера.

(b_1) (Гипотеза Франкеля). *Всякое замкнутое кэлерово многообразие* V, *имеющее положительную кривизну по всем двумерным направлениям, биголоморфно эквивалентно пространству* $\mathbb{C}P^n$.

Эту гипотезу доказали Сю и Яу, которые начали с исследования некоторой подходящей минимальной поверхности S в V. Они показали, что поверхность S голоморфна и допускает столько же голоморфных деформаций, сколько $\mathbb{C}P^1\subset\mathbb{C}P^n$. (Возможно, это утверждение распространяется и на калиброванные почти-комплексные многообразия.)

Заметим, что справедлив (сформулированный Хартсхорном) алгебраический вариант гипотезы Франкеля (где «положительная кривизна» заменяется на «обильный касательный пучок», что согласно алгебраико-геометрическим стандартам даёт гибкое утверждение). Это доказал Мори, который использовал в качестве жёсткой техники действие Фробениуса на кривые в многообразии V (после сведения задачи к случаю конечной характеристики).

(b_2) Пусть V — риманово многообразие с положительным оператором кривизны. Тогда (см. [Mo] и [Mi]) у любой минимальной сферы в многообразии V индекс (Морса) $\geqslant m/2-3/2$, где $m=\dim V$, причём в случае замкнутого V из этого неравенства следует, что универсальное накрытие многообразия V гомотопически эквивалентно сфере S^m.

(c) Теоремы о компактификации получены (Уленбекк) для многих конформно-инвариантных эллиптических систем более высокой размерности.

(c′) Наиболее яркое (гибкое) приложение такой компактификации (относящееся к уравнению Янга — Миллса, где жёсткие результаты принадлежат Уленбекку и Таубсу) обнаружено Доналдсоном [D1]. См. также [D2 — D4, F-S, T1, T2]. (Отметим, что уравнение Янга — Миллса на данном главном расслоении

P над четырёхмерным многообразием приручено универсальной 4-формой Понтрягина в пространстве струй связностей на P.)

3.4. Теоремы о компактности и существовании для замкнутых голоморфных кривых. При отсутствии стягивания окружностей, описанного в § 3.3, можно говорить о *компактности* соответствующего пространства Σ голоморфных кривых.

3.4.A. **Пример.** Пусть $S = S^2$ и $f: S \to (V, J)$ — голоморфное отображение, такое что гомологический класс $f_*[S] \in H_2(V)$ порождает образ гомоморфизма Гуревича $\pi_2(V) \to H_2(V)$. Тогда при условии прирученности многообразия V некоторой формой ω указанное стягивание кривых в S невозможно (так как в противном случае $S \subset V$ разлагалось бы на меньшие голоморфные сферы). Следовательно, пространство Σ компактно по модулю конформных преобразований поверхности $S = S^2$. Другими словами, пространство Σ' соответствующих непараметризованных голоморфных кривых в V компактно для компактных многообразий V. Легко видеть (это вытекает из теории Фредгольма для нелинейных эллиптических операторов), что Σ' представляет некоторый гомологический класс $[\Sigma']$ в пространстве всех поверхностей в V (здесь V — замкнутое многообразие) и что класс $[\Sigma']$ инвариантен относительно гомотопий J_t структуры J, поскольку гомотопная структура J_t приручена (некоторой формой ω_t) для любого t. Так как пространство почти-комплексных структур, прирученных фиксированной формой ω, стягиваемо (это тривиально, см. [Gr3]), то *класс* $[\Sigma']$ *является инвариантом симплектической формы.*

Другие инварианты многообразия (V, ω) получаются из (гомологических классов) некоторых подмногообразий в Σ', таких как многообразие Σ'_0 голоморфных кривых $S \subset V$, проходящих через фиксированную точку $v_0 \in V$.

Замечание. Классы $[\Sigma']$ и $[\Sigma'_0]$ аналогичны инвариантам Доналдсона в калибровочной теории, однако непосредственная связь между двумя типами инвариантов не найдена до сих пор. (Ср. [Hi].)

3.4.B. Вычислим класс $[\Sigma'_0]$ в простейшем случае

$$(V, J) = (V_0 \times S^2, J_0 \oplus J'),$$

где (V_0, J_0) — замкнутое асферическое многообразие, прирученное некоторой формой ω_0, и (S^2, J') — стандартная (риманова) сфера. Здесь через любую точку $(v_0, s_0) \in V$ проходит *единственная* голоморфная сфера $S_0 \subset V$, гомологичная сфере $v_0 \times S^2 \in V$.

(В действительности $S_0 = v_0 \times S^2$.) Легко видеть, что соответствующий нульмерный класс $[\Sigma'_0]$ нетривиален. Следовательно, этот класс нетривиален для *каждой* почти-комплексной структуры J_1 многообразия V, прирученной формой $\omega = \omega_0 \oplus \omega'$ на V, где ω_0 приручает структуру J_0 многообразия V_0, а ω' — стандартная форма площади на S^2.

3.4.B$_1$. **Следствие.** *Для каждой почти-комплексной структуры J_1 на V, прирученной формой ω, существует голоморфное отображение $f: S^2 \to V$, проходящее через данную точку многообразия V и гомологичное сфере $v_0 \times S^2 \subset V$.*

3.4.B$_2$. **Замечание.** Это следствие аналогично *теореме Римана об отображении сфер*, которая утверждает (в наших обозначениях), что для любых двух почти-комплексных структур J_0 и J' на S^2 существует голоморфная сфера

$$S \subset (S^2 \times S^2, J_0 \oplus J'),$$

гомологичная диагонали в $S^2 \times S^2$ и проходящая через три данные точки многообразия $S^2 \times S^2$, которые не содержатся в $(s_1 \times S^2) \cup (S^2 \times s_2) \subset S^2 \times S^2$ для любой пары $(s_1, s_2) \in S^2 \times S^2$. На самом деле из нашего доказательства следствия 3.4.B$_1$ вытекает существование такой голоморфной сферы S для *каждой* (нерасщепляющейся) структуры J_1 многообразия $S^2 \times S^2$, прирученного формой $\omega_0 \oplus \omega'$ на $S^2 \times S^2$, если $\int_{S^2} \omega_0 = \int_{S^2} \omega'$, а рассматриваемые три точки не содержатся в объединении никаких двух голоморфных сфер S_1 и S_2 в $S^2 \times S^2$, гомологичных $s_1 \times S^2$ и $S^2 \times s_2$. (Доказательство следствия 3.4.B$_1$ даёт также голоморфные сферы S_1 и S_2, проходящие через произвольную точку многообразия $(S^2 \times S^2, J_1)$.) Это обобщение теоремы Римана подобно тому, что получили Шапиро [Sch] и Лаврентьев [L]. Однако результаты работ [Sch] и [L] (имеющие вполне общий характер, по «жёстким» стандартам) сформулированы и доказаны в неинвариантной форме. (Говоря геометрически, в [Sch] и [L] предполагается, что сферы $s_1 \times S^2$ и $S^2 \times s_2$ являются J_1-голоморфными для всех $(s_1, s_2) \in S^2 \times S^2$.) Кроме того, чрезвычайная сложность утверждений и доказательств делает результаты указанных работ непригодными для «гибких» приложений, хотя (в силу существования голоморфных сфер S_1 и S_2 в $(S^2 \times S^2, J_1)$, проходящих через произвольную точку (s_1, s_2) многообразия $S^2 \times S^2$) формулировка теоремы Римана в [Sch] и [L] по существу столь же общая, как и наша инвариантная версия. (По-видимому, жёсткие аналитические теоремы приобретают завершённый вид и становятся поистине полезными только тогда, когда они применяются к достаточно большому естественному классу объектов, определённому на гибком свободном от координат языке.)

3.4.С. *Решение неоднородного уравнения Коши — Римана.* Рассмотрим расслоения $H = \mathrm{Hom}(T(S),\ T(V))$ и $H' = \mathrm{Hom}_{\mathbb{C}}(T(S), T(V))$ над многообразием $S \times V$ с данными (почти-) комплексными структурами J' в $T(S)$ и J в $T(V)$ и обозначим через $h \to \bar{h}$ факторгомоморфизм $H \to \bar{H} = H/H'$. Тогда для каждого гладкого сечения $\varphi : S \times V \to \bar{H}$ можно составить уравнение

$$\overline{D}_f = \varphi = \varphi(s,\ f(s)) \qquad (*)$$

относительно гладких отображений $f : S \to V$.

Графики $S \to S \times V$ решений f уравнения $(*)$ голоморфны в некоторой (получаемой естественным путём) почти-комплексной структуре $J = J(J', J)$ на $S \times V$. Следовательно, предыдущее обсуждение компактности и существования распространяется и на решения уравнения $(*)$. Более того, доказано (см. [Gr3]), что стягивание окружностей у графика $S \to S \times V$ порождает голоморфные сферы в многообразии (V, J). Это приводит к следующему утверждению:

Альтернатива. *Если (V, J) — замкнутое ручное многообразие и $S = S^2$, то либо существует непостоянное голоморфное отображение $S \to V$, либо уравнение $(*)$ имеет для всех φ гомотопное нулю решение $f : S \to V$. Например, уравнение $(*)$ всегда разрешимо, если V — асферическое многообразие.*

3.5. Голоморфные кривые с границами. Подмногообразие $W \subset (V, J)$ называется *вполне вещественным*, если $\dim W = \frac{1}{2} \dim V$ и пересечение $T_w(W) \cap JT_w(W) \subset T_w(V)$ равно нулю для всех $w \in W$. Например, каждая кривая на римановой поверхности вполне вещественна. Будем говорить, что форма ω *приручает пару* (V, W), если она приручает многообразие (V, J) и $\omega | W = 0$, т. е. если W — лагранжево подмногообразие (см. § 1.3) в (V, ω).

Предположим, что S — компактная риманова поверхность с краем, и рассмотрим голоморфные отображения $S \to V$, переводящие ∂S в W. Структура пространства этих отображений $f : (S,\ \partial S) \to (V,\ W)$ по существу такая же, как и в случае замкнутой поверхности S. А именно, это пространство компактифицируется сингулярными голоморфными кривыми, полученными стягиванием некоторых окружностей и некоторых дуг на поверхности S с граничными точками в ∂S. Чтобы убедиться в этом, рассмотрим комплексное многообразие V с $\mathbb{R}$-структурой, заданной антиголоморфной инволюцией на V, у которой множество неподвижных точек (вещественная часть) $W \subset V$ имеет размерность $\dim W = \frac{1}{2} \dim V$ и поэтому вполне вещественно. Зафиксируем верхнюю полусферу $S \subset S^2$ и заметим, что каждое голоморфное отображение $(S, \partial S) \to (V, W)$ продолжается по симметрии до (единственного) голоморфного отображения $S^2 \to V$, коммутирующего с $\mathbb{R}$-структурами многообразий V и S^2 (где $\mathbb{R}$-инволюция на S^2 представляет собой отражение относительно экватора). Следовательно, пространство голоморфных отображений $(S, \partial S) \to (V, W)$ отождествляется с вещественной частью пространства голоморфных отображений $S^2 \to V$ и компактификация первого равна вещественной части второго.

Теоремы о компактности и существовании из § 3.4 легко обобщаются на кривые с границами (см. [Gr3]). В частности, альтернатива из § 3.4.С справедлива в классе лагранжевых подмногообразий $W \subset \mathbb{C}^n$, содержащем все *замкнутые* многообразия. В случае $V = \mathbb{C}^n$ уравнение $\overline{D}_f = \varphi$ равносильно уравнению $\bar{\partial} f = \varphi$ и, если φ — постоянное отображение $S \to \mathbb{C}^n$, то решения $f : S \to \mathbb{C}^n$ являются гармоническими. Следовательно, если норма $\|\varphi\|$ достаточно велика, то для компактного W таких f не существует. Поэтому согласно нашей альтернативе для каждого замкнутого лагранжева C^∞-подмногообразия $W \subset \mathbb{C}^n$ найдётся непостоянное голоморфное C^∞-отображение диска в пространство $\mathbb{C}^n$, переводящее границу диска в W. Подобное рассуждение применимо и к *иммерсированному* лагранжеву подмногообразию W в $\mathbb{C}^n$, что даёт голоморфные диски, упомянутые в § 1.3. Заметим, что голоморфные диски могут быть негладкими в граничных точках из-за наличия двойных точек у иммерсированного подмногообразия $W \subset \mathbb{C}^n$. Но если W имеет нормальные пересечения, то эти диски гёльдеревы.

4. ПРИМЕНЕНИЯ ГОЛОМОРФНЫХ КРИВЫХ В СИМПЛЕКТИЧЕСКОЙ ГЕОМЕТРИИ

4.1. Вложения открытых многообразий. Определим *симплектическую ширину* (width) многообразия (V, ω) как нижнюю грань чисел $a > 0$, таких что для каждой почти-комплексной структуры J на V, прирученной формой ω, и для каждой точки $v \in V$ существует непостоянная связная собственная J-голоморфная кривая $f : S \to V$, которая проходит через точку v и симплектическая площадь которой удовлетворяет условию

$$\int\limits_S f^*(\omega) \leqslant a.$$

Ясно (см. [Gr3]), что симплектическая ширина монотонна относительно симплектических вложений многообразий *одинаковой размерности*. А именно, если V_1 вкладывается в V_2, то

$$\operatorname{width} V_1 \leqslant \operatorname{width} V_2.$$

Примеры. (a) Пусть U_r — шар радиуса r в $\mathbb{C}^n$. Тогда у каждой голоморфной кривой в U_r, проходящей через центр шара, площадь $\geqslant \pi r^2$ (это хорошо известно и легко доказывается). Следовательно,

$$\operatorname{width}\left(U_r,\ \omega = \sum dx_i \wedge dy_i\right) \geqslant \pi r^2.$$

(На самом деле $\operatorname{width} U_r = \pi r^2$; см. [Gr3].)

(b) Пусть $(V, \omega) = (V_0 \times S^2,\ \omega_0 \oplus \omega')$, где (V_0, ω_0) — замкнутое $(2n-2)$-мерное асферическое многообразие и ω' — форма площади на S^2. Тогда в силу 3.4.B$_1$

$$\operatorname{width} V \leqslant \operatorname{area}(S^2,\ \omega') \underset{\text{def}}{=} \int_{S^2} \omega'.$$

(Очевидно, $\operatorname{width} V \geqslant \operatorname{area} S^2$.)

Следствие. *Если указанный выше шар U_r вкладывается в V, то* $\operatorname{area} S^2 \geqslant \pi r^2$.

Заметим, что сфера S^2 с выколотой точкой симплектически изоморфна ε-диску в $\mathbb{C}$, такому что $\pi\varepsilon^2 = \operatorname{area} S^2$, и каждое относительно компактное подмножество в $\mathbb{C}^{n-1} \times S^2$ вкладывается в замкнутое многообразие $(\mathbb{C}^{n-1}/\Lambda) \times S^2$ для некоторой решётки Λ в $\mathbb{C}^{n-1}$. Таким образом, следствие показывает, что шар U_r симплектически вкладывается в ε-окрестность подпространства $\{z_1 = 0\} \subset \mathbb{C}^n$ тогда и только тогда, когда $r \leqslant \varepsilon$. Отсюда вытекает результат о несуществовании симплектических вложений, изложенный в § 1.2 (ср. [Gr3]).

4.2. Вложения коразмерности 2. Пусть V — симплектическое многообразие и $V_0 \subset V$ — замкнутое симплектическое подмногообразие коразмерности два. Тогда, используя почти-комплексную структуру J на V, для которой подмногообразие V_0 является J-комплексным, можно показать, что инварианты Σ' многообразия V (см. § 3.4) дают при ограничении на V_0 (в некотором естественном смысле) аналогичные инварианты многообразия V_0. Типичным следствием этого может служить такое утверждение:

Пусть $V = \mathbb{C}P^2 \times \mathbb{R}^2$, где $\mathbb{C}P^2$ наделено стандартной (U(3)-*инвариантной) симплектической структурой и $\mathbb{R}^2 = (\mathbb{R}^2, dx \wedge dy)$. Тогда замкнутое четырёхмерное симплектическое многообразие V_0, допускающее симплектическое вложение в V, симплектически диффеоморфно пространству $\mathbb{C}P^2$, если $\pi_2(V_0)$ — циклическая группа.*

4.3. Задачи о существовании, продолжении и эквивалентности симплектических структур. Пусть ω_0 — невырожденная 2-форма на связном многообразии V.

Задача. Существует ли гомотопия невырожденных форм ω_t, $0 \leqslant t \leqslant 1$, такая что форма ω_1 замкнута (и, следовательно, симплектична)?

Если V — открытое многообразие, то утвердительный ответ на этот вопрос даётся теоретико-пучковым вариантом теоремы Смейла — Хирша об иммерсии в теории пучков (см. [Gr1]). Если же V замкнуто, то для существования гомотопии ω_t есть очевидные препятствия. А именно, на многообразии V должен быть двумерный когомологический класс α, такой что $\alpha^n \neq 0$, где $2n = \dim V$. До сих пор неизвестно, имеются ли здесь другие препятствия.

Аналогичная задача связана с продолжением симплектической формы с подмножества многообразия V на всё V. Нетривиальное препятствие в этом случае дают голоморфные кривые.

Пример (см. [Gr3]). Пусть V — открытое четырехмерное многообразие, такое что гомоморфизм Гуревича $\pi_2(V) \to H_2(V)$ равен нулю, и ω_0 — симплектическая форма, определённая в некоторой окрестности U бесконечности в V. Если окрестность (U, ω_0) симплектически диффеоморфна некоторой окрестности бесконечности в $(\mathbb{R}^4, dx_1 \wedge dy_1 + dx_2 \wedge dy_2)$ и ω_0 продолжается до симплектической формы на всём V, то многообразие V диффеоморфно пространству $\mathbb{R}^4$.

Рассмотрим теперь две симплектические формы ω_0 и ω_1 на *замкнутом* многообразии V, которые представляют один и тот же когомологический класс $[\omega_0] = [\omega_1] \in H^2(V; \mathbb{R})$ и могут быть соединены гомотопией невырожденных форм ω_t. Если форма ω_t симплектична для любого $t \in [0, 1]$, а когомологический класс $[\omega_t]$ не зависит от t, то по теореме Дарбу — Мозера многообразия (V, ω_0) и (V, ω_1) симплектически диффеоморфны. Если же $[\omega_t]$ меняется, то инварианты типа $[\Sigma']$ (см. § 3.4) тоже могут изменяться (см. [McD]) и тогда (V, ω_0) не будет симплектически диффеоморфно многообразию (V, ω_1).

4.4. C^0-пределы симплектических диффеоморфизмов. В («жёсткой») классической математике (и механике, где симплектические диффеоморфизмы называются *каноническими преобразованиями*) по-видимому предполагалось, что симплектические диффеоморфизмы могут отличаться от сохраняющих объём диффеоморфизмов некоторыми глобальными свойствами, устойчивыми относительно взятия равномерных пределов таких диффеоморфизмов. Это предположение (явно высказанное Арнольдом) было подтверждено Элиашбергом (см. (E2, E3]), который доказал, в частности (путём усовершенствования комбинаторного метода, восходящего к «доказательству» Пуанкаре его последней теоремы), что каждый диффеоморфизм пространства $\mathbb{R}^{2n}$, являющийся C^0-пределом симплектических диффеоморфизмов, симплектичен. Эту C^0-устойчивость можно также получить и из результатов § 4.1 о несуществовании симплектических вложений, если воспользоваться теоремой Нэша о неявной функции (см. [Gr4]).

4.5. Лагранжевы пересечения и неподвижные точки симплектических диффеоморфизмов. Как видно из § 1.2, каждый симплектический диффеоморфизм f, допускающий включение в однопараметрическую подгруппу, определяется некоторой (производящей) функцией (гамильтонианом) h на исходном многообразии V, если $H^1(V; \mathbb{R})=0$; неподвижные точки диффеоморфизма f соответствуют критическим точкам функции h. Следовательно, число неподвижных точек диффеоморфизма f можно оценить снизу с помощью теории Морса. Такая оценка (предложенная Арнольдом в виде гипотезы и обобщающая последнюю теорему Пуанкаре) для точных (имеется в виду естественное обобщение понятия точности замкнутых 1-форм, соответствующих симплектическим векторным полям) сохраняющих площадь диффеоморфизмов поверхностей была доказана Элиашбергом [E1] при помощи его комбинаторного метода.

Затем Конли и Цендер [C-Z] доказали другую гипотезу Арнольда: *Всякий точный симплектический диффеоморфизм f тора $T^{2n}=\mathbb{R}^{2n}/\mathbb{Z}^{2n}$ имеет по меньшей мере $2n+1$ неподвижных точек, причём если f — диффеоморфизм общего положения, то он имеет по меньшей мере 4^n неподвижных точек.* Доказательство сводится к изучению вспомогательной (производящей) функции L (лагранжиана) на пространстве стягиваемых отображений $S^1\to T^{2n}$; Конли и Цендер применили здесь вариационную технику, подобную той, что использовалась для нахождения периодических орбит гамильтоновых (т. е. симплектических) потоков (см. [Rab, W2, Ber]). Заметим, что функция L имеет бесконечный индекс Морса $\left(\text{она ведёт себя как квадратичная функция } \sum_{i=1}^{\infty} x_i y_i \text{ в } \mathbb{C}^\infty\right)$ и обычная теория Морса к ней не применима. (В действительности существует очень мало интересных вариационных задач, где теория Морса, базирующаяся на условии Пале — Смейла, может быть применена непосредственно.) Метод Конли — Цендера был очищен от «жёсткого» анализа Шапероном, который использовал вместо L некоторую функцию на конечномерном пространстве, аналогичном пространству ломаных геодезических на римановом многообразии. В подходе Шаперона жёсткая теорема о регулярности сводится к доказательству того, что каждая ломаная *экстремальная* кривая является на самом деле кривой без изломов.

Неподвижные точки симплектического диффеоморфизма f многообразия (V,ω) соответствуют точкам пересечения двух лагранжевых подмногообразий в $(V\times V, \omega\oplus-\omega)$, которые представляют собой график отображения f и диагональ в многообразии $V\times V$. Метод Конли — Цендера (в толковании Шаперона) распространяется и на некоторые лагранжевы многообразия, не являющиеся симплектическими графиками. Например, Лауденбах и Сикоров [L-S] определили этим методом морсовскую нижнюю грань для числа точек пересечения *точного* (в соответствующем смысле) лагранжева подмногообразия в $T^*(X)$ с нулевым сечением $X\subset T^*(X)$, где $T^*(X)$ наделено канонической (см. § 1.3) симплектической структурой. (ср. [Ch, F-W, W1, Z].)

Альтернативный подход к вопросу о (само)пересечениях иммерсированных лагранжевых подмногообразий $W\subset V$ дают голоморфные кривые $(S, \partial S)\to(V, W)$ (см. § 3.5 и [Gr3]). Подобные кривые (в рамках теории Морса) использовались Флёром [F1], который доказал гомологическую версию общей гипотезы Арнольда. По-видимому, метод голоморфных кривых обладает преимуществом большей общности, тогда как симплектическая теория Морса, в случае если она применима, приводит к более тонким неравенствам (Морса).

4.6. Контактная геометрия. Подрасслоение $\theta\subset T(X)$ коразмерности один называется *контактным*, если существует почти-комплексная структура J на $X\times\mathbb{R}$, такая что

$$\theta=T(X)\cap JT(X),$$

где $X=X\times 0\subset X\times\mathbb{R}$, причём многообразие X *строго J-выпукло* в $X\times\mathbb{R}$. То есть $X=X\times 0$ представляет собой линию уровня строго J-выпуклой функции (см. § 3.2) без критических точек. Гибкие свойства контактных многообразий (V,θ) вполне аналогичны гибким свойствам симплектических многообразий (см. [Gr4]). Основные жёсткие теоремы для контактных трёхмерных многообразий были доказаны Беннекеном [B1], использовавшим топологическую технику (теорию узлов). В частности, Беннекен доказал существование экзотических контактных структур в $\mathbb{R}^3$ и S^3. (Стандартная структура θ в S^3 есть $T(S^3) \cap \sqrt{-1}T(S^3)$ для обычного вложения $S^3 \subset \mathbb{C}^2$. Стандартное пространство $\mathbb{R}^3$ получается из этой сферы S^3 выкалыванием точки.)

Пусть (V, J) — почти-комплексное многообразие с J-выпуклым краем Y и $\theta = T(Y) \cap JT(Y)$. Если $W \subset Y$ — вполне вещественное подмногообразие в $V \supset Y$, то при соответствующем условии приручения голоморфные кривые в V с границами в W ведут себя аналогично кривым из § 3.5. Это даёт нетривиальные следствия, касающиеся контактной геометрии многообразия (Y, θ), и, в частности, показывает, что не каждое (Y, θ) является ручным J-выпуклым краем некоторого многообразия (V, J). Пользуясь этим, можно найти, например, экзотические контактные структуры в $\mathbb{R}^{2n-1}$ для всех $n \geqslant 1$ (см. [Gr3, B2]) и более изящным образом получить результаты Беннекена для трёхмерных многообразий (личное сообщение Элиашберга). Тем не менее к настоящему времени голоморфная контактная геометрия изучена меньше, чем симплектический случай.

Некоторые исторические замечания о гибкости и жёсткости. По-видимому, трудно (если вообще возможно) придать точный математический смысл понятиям относительной гибкости и жёсткости рассуждения или теории. Интуитивно «жёсткий» относится к строгой и чёткой структуре данного объекта, тогда как «гибкий» говорит о некотором более слабом общем свойстве целого класса объектов. Так, неравенства и оценки гибче тождеств, (алгебраическая) теория чисел жёстче анализа (над локально-компактными полями и аделями), вещественный анализ гибче комплексного. Полупростые группы Ли и симметрические пространства с некоторой точки зрения выглядят почти такими же жёсткими, как и целые числа, в то время как риманова геометрия в целом представляется приблизительно такой же гибкой, как и дифференциальная топология. Доказательство неравенства $x^3 + y^3 + z^3 \geqslant 3xyz$ для $x, y, z \geqslant 0$, основанное на выпуклости функции $\log t$, представляется более гибким, чем доказательство тождества $2(x^3 + y^3 + z^3 - 3xyz) = (x + y + z)((x - y)^2 + (x - z)^2 + (y - z)^2)$, хотя доказательства в алгебраической геометрии, в которых применяется теория эллиптических уравнений с частными производными, имеют тот же уровень жёсткости, что и доказательства, использующие действие Фробениуса.

В этом сообщении понятия «гибкий» и «жёсткий» ограничиваются рамками *глобального нелинейного анализа*, касающегося геометрии пространств отображений гладких многообразий. Современный подход к этим пространствам берёт начало с гибкого гомотопического метода Серра [S1, S2], после которого в исследованиях стали преобладать гибкие идеи и техника. До сих пор неизвестно, можно ли получить результаты Серра (например, конечность стабильных гомотопических групп сфер) жёстким образом, хотя жёсткие рассуждения и применяются время от времени в подобных задачах (например, теория Морса в теореме периодичности Ботта, действие Фробениуса в гипотезе Адамса, линейные эллиптические операторы в теореме о сигнатуре и гипотезе о высшей сигнатуре). Такая же гибкость превалирует и в дифференциальной топологии, поскольку эластичность диффеоморфизмов (Том [Th]), иммерсий (Смейл [Sm]) и хирургии сделала возможным существенное сведение Diff-задач к гомотопической теории. (Эта гибкая топологическая лавина была приостановлена тёрстоновой геометризацией теории трёхмерных многообразий и калибровочными доналдсоновыми инвариантами четырёхмерных многообразий.)

Открытая в топологии гибкость не смогла, однако, отбить охоту к поиску классических жёстких структур, основанных на нелинейных уравнениях с частными производными. Но наивная мечта о жёстком глобальном анализе была быстро развеяна Нэшем [N1, N2], продемонстрировавшим поразительную эластичность решений некоторых нелинейных уравнений, а именно изометрических иммерсий римановых многообразий. (Например, *каждое* вложение стандартной m-сферы в пространство $\mathbb{R}^n$, уменьшающее расстояние, допускает при $m < n$ равномерную аппроксимацию изометрическими C^1-вложениями; см. [N1, K].) Некоторое время (вплоть до работ Дональдсона) казалось, что явление Нэша (почти вездесущее в нелинейных уравнениях с частными производными, см. [Gr4]) полностью исключает любую жёсткую структуру уравнений с частными производными на гибком просторе нелинейных функциональных пространств. (Жёсткая структура *линейных* эллиптических уравнений с частными производными на сегодня прочно укоренилась в гибкой топологической почве, как и предвидел Атья [A].) В настоящее же время голоморфные кривые, которые связаны с наиболее элементарными уравнениями (и логически предшествуют калибровочным полям, хотя исторически дело обстоит наоборот), становятся первой ступенькой на пути к созданию исчерпывающей жёсткой нелинейной теории.

ЛИТЕРАТУРА

[A] M. Atiyah, Global aspects of the theory of elliptic differential operators. — В сб.: Труды Международного конгресса математиков (Москва 1966). — М.: Мир, 1968, с. 57—64.

[B1] D. Bennequin, Entrelacements et équations de Pfaff, Astérisque **107—108** (1982), 87—163. Имеется перевод: Беннекен Д. Зацепления и уравнения Пфаффа. — УМН, 1989, т. 44, № 3, с. 3—53.

[B2] —, Problems elliptiques, surfaces de Riemann et structures symplectiques, Sém. Bourbaki Févr., Soc. Math. de France, 1986, pp. 1—23.

[Ber] H. Berestycki, Solutions péridiques de systèmes hamiltoniens, Sém. Bourbaki Févr., Soc. Math. de France, 1983.

[Ch] M. Chaperon, Quelques questions de géometrie symplectique, Astérisque **105—106** (1983), 231—249.

[C-Z] C. Conley and E. Zehnder, The Birkhoff-Lewis fixed point theorem and a conjecture of V. I. Arnold, Invent. Math. **73** (1983), 33—49.

[D1] S. K. Donaldson, An application of gauge theory to four dimensional topology, J. Differential Geom. **18** (1983), 279—315.

[D2] —, Connection, cohomology and the intersection forms on 4-manifolds, J. Differential Geom. **24** (1986), 275—341.

[D3] —, The orientation of Yang-Mills moduli spaces and fundamental groups of 4-manifolds, Preprint, Oxford, Univ.

[D4] —, Gauge theory and smooth structures on 4-manifolds, Preprint, Oxford Univ.

[E1] Элиашберг Я. М. Оценка числа неподвижных точек преобразований, сохраняющих площадь. — Препринт, Сыктывкар, 1978.

[E2] Элиашберг Я. М. Жёсткость симплектических и контактных структур. — Препринт, 1981.

[F2] —, Rigidity of symplectic and contact structures, Preprint, 1981.

[E3] —, Combinatorial methods in symplectic geometry, Proc. Internat. Congr. Math. (Berkeley, Calif., USA, 1986).

[F1] A. Floer (in preparation).

[F-S] R. Fintushel and R. J. Stern, SO(3) connections and the topology of 4-manifolds, J. Differential Geom. **20** (1984), 523—539.

[F-W] B. Fortune and A. Weinstein, A symplectic fixed point theorem for complex projective spaces, Preprint, Univ. of Calif., Berkeley, 1984.

[Gr1] Громов М. Л. Стабильные отображения слоений в многообразия.— Изв. АН СССР, сер. мат., 1969, т. 33, № 4, с. 707—734.

[Gr2] —, A topological technique for the construction of solutions of differential equations and inequalities, Proc. Internat. Congr. Math. (Nice, 1970), Vol. 2, Gauthier-Villars, Paris, 1971, pp. 221—225.

[Gr3] —, Pseudo-holomorphic curves in symplectic manifolds, Invent. Math. **82** (1985), 307—347.

[Gr4] —, Partial differential relations, Springer-Verlag, Berlin — New York, 1986. [Имеется перевод: Громов М. Дифференциальные соотношения с частными производными. — М.: Мир, 1990.]

[Hi] N. J. Hitchin, The self-duality equation on a Riemann surface, Preprint, Oxford Univ.

[H-L] R. Harvey and B. Lawson, Calibrated geometries, Acta Math. **148** (1982), 48—156.

[K] N. H. Kuiper, On C^1-isometric embeddings. I, Nederl. Akad. Wetensch. Proc. Ser. **A58** (1955), 545—556.

[L] Лаврентьев М. А. Основная теорема теории квазиконформных отображений плоских областей. — Изв. АН СССР, сер. мат., 1948, т. 12, № 6, с. 513—554.

[L-S] F. Laudenbach et J.-C. Sikorov, Persistance d'intersections avec la section nulle au cours d'une isotopie hamiltonienne dans un fibré cotangent, Invent. Math. **82** (1985), no. 2, 349—358.

[McD] D. McDuff, Examples of symplectic structures, Preprint, SUNY at Stony Brook, 1985.

[Mi] M. J. Micallef, On the topology of positively curved manifolds, Preprint, Australian Nat. Univ., 1986.

[Mo] J. D. Moore, Minimal two-spheres on the topology of manifolds with positive curvature on totally isotropic two-planes, Preprint, 1986.

[N1] J. Nash, C^1-isometric embeddigs, Ann. of Math. **60** (1954), no. 3, 383—396. [Имеется перевод: Нэш Дж. C^1-изометрические вложения. — Математика, 1957, т. 1, № 2, с. 3—28.]

[N2] —, The embedding problem for Riemannian manifolds, Ann. of Math. **63** (1956), no. 1, 20—63. [Имеется перевод: Нэш Дж. Проблема вложения для римановых многообразий. — УМН, 1971, т. 26, № 4, с. 173—216.]

[Rab] P. H. Rabinowitz, Periodic solutions of Hamiltonian systems: A survey, MRC Tech. Report n° 2154, Univ. Wisconsin, Madison, 1980.

[S1] J.-P. Serre, Homologie singulière des espaces fibrés, Ann. of Math. **54** (1951), 425—505.

[S2] —, Groupes d'homotopie et classes de groupes abéliens, Ann. of Math. **58** (1953), 258—294.

[Sch] Шапиро З. Я. О существовании квазиконформных отображений.— ДАН СССР, 1941, т. 30, №8, с. 685—687.

[Sm] S. Smale, The classification of immersions of spheres in Euclidean spaces, Ann. of Math. (2) **69** (1959), 327—344.

[S-U] J. Sacks and K. Uhlenbeck, The existence of minimal immersions of two-spheres, Ann. of Math. **113** (1981), 1—24.

[Sch-Y] R. Schoen and S. T. Yau, Existence of incompressible minimal surfaces and the topology of three-dimensional manifolds of non-negative scalar curvature, Ann. of Math. **110** (1979), 127—142.

[T1] C. H. Taubes, Gauge theory on asymptotically period 4-manifolds, Preprint, Harvard University, 1986.

[T2] —, Gauge theories and non-linear partial differential equations, Proc. Internat. Congr. Math. (Berkeley, Calif., USA, 1986).

[Th] R. Thom, Quelques propriétés globales des variétés différentiables, Comment. Math. Helv. **28** (1954), 17—86.

[W1] A. Weinstein, Lectures on symplectic manifolds, Regional Conf. Ser. in Math., No. 29, Amer. Math. Soc., Providence, R. I., 1977.

[W2] —, Periodic orbits of convex hamiltonian systems, Ann. of Math. **108** (1977), 507—518.

[Z] E. J. Zehnder, Fixed points of symplectic mappings and periodic solutions of Hamiltonian systems, Proc. Internat. Congr. Math. (Berkeley, Calif., USA, 1986).

ИНСТИТУТ ВЫСШИХ НАУЧНЫХ ИССЛЕДОВАНИЙ, 91440 БЮР-СЮР-ИВЕТТ, ФРАНЦИЯ

ЭЛЛИПТИЧЕСКИЕ КРИВЫЕ И ТЕОРЕТИКО-ЧИСЛОВЫЕ АЛГОРИТМЫ

Х. В. Ленстра-мл. [1]

§ 1. ВВЕДЕНИЕ

В этом докладе мы обсудим проблему, которая на протяжении веков волновала стольких математиков, в том числе Эратосфена (~ —284 — ~ —202), Фибоначчи (~1180 — ~1250), Ферма (1601—1665), Эйлера (1707—1783), Лежандра (1752—1833) и Гаусса (1777—1855). Эта проблема состоит в разложении на простые множители заданного большого целого числа.

Обзоры методов, которые используются для этой цели, можно найти в недавней книге Риеселя [27] и в приложениях к сборнику [21]. Настоящий доклад посвящён новым событиям, которые произошли в этой области после выхода в свет книги Риеселя, точнее — использованию *эллиптических кривых*.

В большинстве способов нахождения простых сомножителей можно выделить два этапа. На первом из них (*проверка на простоту*) выясняется, является заданное число простым или составным. На втором этапе (разложение на множители или *факторизация*) осуществляется поиск нетривиального делителя числа, в случае если это число составное. Очевидно, что полное разложение на простые множители можно получить последовательным применением алгоритмов проверки на простоту и нахождения нетривиального делителя. Эллиптические кривые могут применяться как при проверке простоты, так и при разложении на множители; они дают превосходные алгоритмы и в теории, и на практике.

Считается, что проверка на простоту — более лёгкая задача, чем факторизация. Например, предположим, что два 100-значных числа p и q оказались простыми, это легко устанавливается современными методами проверки на простоту. Допустим теперь, что числа p и q случайно были потеряны, но сохранилось их произведение pq. Как восстановить p и q? Со стыдом приходится признать тот позорный для математики факт, что в этих обстоятельствах наиболее обещающие подходы — это либо поиски черновиков в корзине для мусора, либо мнемо-гипнотические приёмы.

До недавнего времени обе эти задачи — проверка на простоту и отыскание делителя — большинством математиков не воспринимались всерьёз. В наши дни такое отношение в значительной мере изменилось. Отчасти это вызвано внедрением более изысканных математических методов по сравнению с использовавшимися ранее. И на самом деле применение эллиптических кривых, которое является главной темой этого доклада, по сути — первый пример использования математики двадцатого столетия в проблеме разложения на простые множители.

Другая причина возрастающего интереса к этой области — это возможность применить теорию чисел к чему-то находящемуся вне неё. Существование и единственность разложения числа на простые множители составляют содержание *основной теоремы арифметики*, и эта теорема, действительно, играет фундаментальную роль. Скажем, известные теоретико-числовые вопросы о натуральном числе n — может ли число n быть представленным в виде суммы двух квадратов? или каков порядок мультипликативной группы кольца вычетов $(\mathbb{Z}/n\mathbb{Z})^*$? — считаются решёнными, если ответ формулируется в терминах разложения числа n на простые множители. Ввиду основополагающей роли в теории чисел теоремы о разложении на простые множители естественно предположить, что алгоритмы разложения на простые множители должны играть важную роль и в возможных приложениях теории чисел. К настоящему времени наиболее впечатляющей иллюстрацией этого является шифровальная схема, предложенная Ривестом, Шамиром и Адлеманом [28]. При использовании этой схемы существенно то, что проверка простоты есть *лёгкая* операция, а надёжность шифра основывается на том, что факторизация — это *трудная* процедура. Следует заметить, что до некоторой степени это — негативное применение, поскольку, в случае если будет найден лучший метод разложения на множители, схема может потерять своё значение. Это замечание предназначено для тех математиков, которых не увлекает возможность приложений теории чисел в других областях и кто хочет сохранить чистоту своей науки.

Для того чтобы проверить на простоту данное целое число $n \geqslant 1$, его обычно подвергают серии *тестов на псевдопростоту*. Большинство этих тестов основано на каком-нибудь варианте

---

[1] H. W. Lenstra, jr., Elliptic curves and number-theoretic algorithms, ICM86, pp. 99—120.

теоремы Ферма. (Эта теорема утверждает, что если n является простым числом, то для всех целых чисел a выполнено сравнение $a^n \equiv a \pmod{n}$.) Эти тесты на псевдопростоту обладают тем свойством, что любое простое число им удовлетворяет, а вероятность того, что им удовлетворяет составное число, очень мала. Таким образом, достаточно указать один-единственный тест, которому не удовлетворяет число n, чтобы сделать вывод о непростоте числа n. Такой тест, однако, не предъявляет явно нетривиального множителя n. С другой стороны, если число n удовлетворяет большому количеству тестов на псевдопростоту, оно с большой вероятностью будет простым. Возникает проблема: как *доказать*, что n является простым? Можно сказать, что реальная трудность алгоритмов проверки на простоту состоит не в том, чтобы *получить* ответ, «простое» или «составное», а в том, чтобы доказать *правильность* ответа в случае, когда этот ответ — «простое». По этой причине иногда говорят об алгоритмах, *доказывающих* простоту.

Как уже было отмечено, если тест на простоту определяет, что данное число является составным, он, как правило, не предъявляет какого-нибудь делителя этого числа. Для получения делителя применяется алгоритм факторизации. В отличие от тестирования на простоту трудность проблемы факторизации состоит именно в *получении* ответа (т. е. нетривиального делителя числа), а проверка правильности ответа после его получения уже не представляет труда. Полная свобода в выборе метода нахождения нетривиального делителя, по-видимому, является одной из причин гораздо большего разнообразия алгоритмов факторизации по сравнению с тестами на простоту. На самом деле априори неясно, почему методы математические должны быть лучше нематематических, в частности почему разложение на множители не может быть по силам профессиональным гадалкам или служителям культа.

Наилучший способ понять методы эллиптических кривых, являющиеся предметом этого доклада, — провести аналогию с некоторыми известными старыми алгоритмами, обсуждаемыми ниже в § 2. Эти более ранние алгоритмы опираются на свойства *мультипликативной группы* кольца вычетов, в частности на тот факт, что для простого числа p порядок мультипликативной группы $(\mathbb{Z}/p\mathbb{Z})^*$ равен $p-1$. Заметим, что алгоритмы, рассматриваемые в § 2, никоим образом не являются наилучшими среди тех, которые использовались до применения эллиптических кривых, мы обсуждаем их только потому, что они полезны для мотивировки и понимания новых методов.

Основные необходимые свойства эллиптических кривых излагаются в § 3. Лучшим руководством по этому предмету является недавно вышедший учебник Силвермана [35]. Как и в большинстве других источников, в этой книге рассматриваются лишь эллиптические кривые над полями. Как с концептуальной точки зрения, так и с точки зрения удобства изложения для наших целей является более естественным работать с эллиптическими кривыми, определёнными над кольцами. Общую теорию эллиптических кривых над коммутативными кольцами с единицей можно найти в [16, гл. 2]. В § 3 мы даём основные определения, правда только для случая, когда рассматриваемое кольцо удовлетворяет некоторому условию. Это условие выполняется, например, если кольцо является полем. Оно также выполнено, когда кольцо *конечно*, что как раз и имеет место в наших приложениях. Благодаря введению этого дополнительного условия мы можем определить эллиптические кривые непосредственно как кривые, задаваемые однородными кубическими многочленами от трёх переменных в некоторой нормальной форме. Для того чтобы эта нормальная форма была как можно более простой, мы предполагаем, что число 6 является обратимым элементом кольца. Тогда множество точек кривой над кольцом определяется как множество нулей рассматриваемого многочлена в подходящим образом определённой проективной плоскости. Основное свойство эллиптических кривых состоит в том, что множество точек на них имеет структуру абелевой группы. Следует отметить, что при описании и анализе обсуждаемых ниже алгоритмов в принципе возможно с помощью искусственных приёмов избежать использования эллиптических кривых над кольцами, не являющимися полями. Так оно и было в действительности сделано в первоначальных публикациях [14, 20, 30].

Как мы уже упомянули выше, часть старых алгоритмов проверки простоты основаны на том, что порядок мультипликативной группы $(\mathbb{Z}/p\mathbb{Z})^*$ кольца вычетов по модулю простого числа p равен $p-1$. Аналогичным образом, в методах, использующих эллиптические кривые, важную роль играет порядок группы точек $E(\mathbb{Z}/p\mathbb{Z})$ эллиптической кривой над $\mathbb{Z}/p\mathbb{Z}$ для простого числа p. По теореме Хассе 1934 года этот порядок равен $p+1-t$, где t — целое число, зависящее от E и p, причём $|t| \leqslant 2\sqrt{p}$. Можно сказать, что новый метод обязан успехом тому факту, что для фиксированного p число t изменяется, если варьировать кривую E. В § 4 мы обсуждаем несколько методов вычисления числа t.

В § 5 дано описание тестов проверки на простоту с помощью эллиптических кривых. В частности, рассмотрены алгоритмы Гольдвассера — Килиана [14] и Аткина [2]. Метод Аткина имеет большое практическое значение, и для большинства чисел, на которых он испытывался, он оказался гораздо более быстрым, чем прежний чемпион — принадлежащий Коэну и А. Ленстре вариант теста Адлемана — Померанца — Румели [1, 9, 10].

Наконец, в § 6 описан алгоритм факторизации с помощью эллиптической кривой [20]. Этот алгоритм в настоящее время является несомненным лидером среди методов факторизации для подавляющего большинства чисел. Алгоритм квадратного решета Померанца [26], который был прежним чемпионом, по-видимому, до сих пор остаётся наилучшим для чисел, которые составлены из двух простых одного порядка величины. Метод же эллиптических кривых имеет то весьма привлекательное свойство, что его скорость зависит от размера наименьшего простого делителя разлагаемого числа n и, таким образом, меньшие простые делители находить легче. Метод квадратичного решета и многие другие быстрые алгоритмы факторизации не обладают этим свойством, их время работы зависит лишь от величины самого числа n, а не от величины его простых сомножителей.

Обозначим через $\mathbb{F}_q$ конечное поле из q элементов. Далее предполагается, что все рассматриваемые кольца являются коммутативными кольцами с единицей, которая сохраняется при кольцевых гомоморфизмах. Группа обратимых элементов кольца R обозначается через R^*.

§ 2. МУЛЬТИПЛИКАТИВНЫЕ МЕТОДЫ

В этом параграфе мы обсудим два старых алгоритма — проверки на простоту и факторизации, которые опираются на свойства мультипликативной группы. На практике эти алгоритмы можно осуществить не для всех чисел, а только лишь для тех, для которых выполнены некоторые условия.

Начнём с проверки на простоту. Следующая теорема принадлежит Поклингтону [24]:

Теорема 1. *Пусть $n > 1$ — целое число, а s — натуральное число, делящее $n-1$. Предположим, что существует целое число a, удовлетворяющее условиям*

$$a^{n-1} \equiv 1 \bmod n$$

и

$$\text{н. о. д.}\,(a^{(n-1)/q} - 1,\ n) = 1$$

для каждого простого делителя q числа s. Тогда каждый простой делитель p числа n сравним с 1 по модулю s, и если $s > \sqrt{n} - 1$, то число n является простым.

Доказательство состоит в следующем. Пусть p — простой делитель n. Возьмём $b = (a^{(n-1)/s} \bmod n)$. Из равенства $a^{n-1} \equiv 1 \bmod n$ следует, что $b^s \equiv 1 \bmod p$. Следовательно, порядок элемента $b \bmod p$ в группе $\mathbb{F}_p^*$ делит s. Если q — простой делитель s, то $b^{s/q}$ не сравнимо с 1 по модулю p, так как по предположению $a^{(n-1)/q} - 1$ не делится на p. Поэтому порядок $b \bmod p$ в группе $\mathbb{F}_p^*$ не является делителем s/q ни для какого простого делителя q числа s, т. е. этот порядок равен s. Применяя теорему Лагранжа из элементарной теории групп, заключаем, что s делит $\# \mathbb{F}_p^* = p - 1$. Этим доказано первое утверждение теоремы. Если $s > \sqrt{n} - 1$, то $p > \sqrt{n}$ для всех простых делителей n, следовательно p просто. Теорема полностью доказана.

Теорема 1 используется при проверке на простоту следующим образом. Пусть $n > 1$ — целое число, о котором есть основания предполагать, что оно просто, например, потому, что оно прошло тесты на псевдопростоту, описанные в [17, p. 379; 27, p. 98]. Обозначим через s наибольший нетривиальный делитель числа $n - 1$, для которого известно полное разложение на простые множители; очевидно, $s > \sqrt{n} - 1$. Теперь выберем случайным образом ненулевое целое число $a \pmod n$ и проверим, удовлетворяет ли оно двум условиям теоремы 1. Заметим, что эти условия проверить легко: простые делители q числа s нам известны, степени $a^{n-1} \pmod n$ и $a^{(n-1)/q} \pmod n$ могут быть вычислены за $O(\log n)$ умножений и возведений в квадрат по модулю n, наибольшие общие делители можно вычислить с помощью алгоритма Эвклида. Если окажется, что все условия теоремы 1 выполнены, то n, действительно, является простым, как и предполагалось.

Следует заметить, что если n — простое число, то найти элемент $a \in \mathbb{Z}/n\mathbb{Z}$, удовлетворяющий условиям теоремы, нетрудно. Ясно, что в этом случае любой ненулевой элемент $a \in \mathbb{Z}/n\mathbb{Z}$ удовлетворяет первому условию. Нетрудно показать, что для фиксированного q второе условие удовлетворяется с вероятностью $1 - q^{-1}$, если n простое, а число $a \neq 0$ выбрано случайным образом. Вероятность того, что a удовлетворяет второму условию для всех q, может быть немного меньше, но в любом случае она не меньше $c_0/\log\log n$ для некоторой положительной константы c_0; нетрудно доказать чуть более общий вариант этой теоремы, в котором a выбирается в зависимости от q.

Основной недостаток теста проверки на простоту, основанного на теореме 1, состоит в том, что с его помощью можно доказать простоту лишь таких простых чисел n, для которых число $n-1$ имеет большой делитель с известным полным разложением на простые множители. Так обстоит дело, например, если $n-1$ имеет много малых простых множителей; это относится, скажем, к числам Ферма $n=2^{k+1}$. Теорема 1 полезна также, когда $n-1$ является произведением маленького числа и большого простого числа q. В этом случае можно попытаться доказывать простоту q, вновь основываясь на теореме 1.

Существует аналог теоремы 1, в котором мультипликативная группа заменяется *скрученной* мультипликативной группой. Например, если p — простое число, то $\mathbb{F}^*_{p^2}/\mathbb{F}^*_p$ — скрученная мультипликативная группа и её порядок равен $(p^2-1)/(p-1)=p+1$. Это приводит к тестам на простоту, которые могут быть использованы для тех чисел n, для которых $n+1$ имеет большой полностью разложимый делитель. Так обстоит дело, например, в случае чисел Мерсенна $n=2^k-1$. Классически эти тесты формулировались в терминах *последовательностей Лукаша.*

Мы отсылаем к [27, 38] за деталями этих и других обобщений теоремы 1 и за описанием тестов на простоту, основанных на комбинации $(n-1)$- и $(n+1)$-методов. Если число n обладает тем свойством, что по крайней мере одно из чисел $n\pm1$ может быть записано в виде произведения полностью разложимого числа и простого числа q, которое в свою очередь обладает тем же свойством, то простоту n можно доказать с помощью многократного применения двух вышеупомянутых методов. По этому пути пошли Селфридж и Вундерлих [32], которые эмпирически установили, что описанный способ может быть применён к большинству простых чисел, имеющих до 35 десятичных знаков, если под «полной разложимостью» подразумевать, что число представляет собой произведение простых чисел, не превосходящих 30030. Обобщение, предложенное Уильямзом и др. [38], пригодно для большинства простых чисел, имеющих до 80 знаков.

В данном контексте преимущество эллиптических кривых состоит прежде всего в большом их разнообразии. Каждая эллиптическая кривая порождает некоторую группу, и порядок этой группы разный в зависимости от выбора кривой. Вместо чисел $n\pm1$ используют, по существу, случайное число, близкое к n, являющееся порядком группы, отвечающей эллиптической кривой, и кривую можно варьировать до тех пор, пока этот порядок не будет иметь желаемого разложения на простые множители. Подробности будут обсуждаться в § 5.

Теперь рассмотрим метод разложения на простые множители, который также основан на свойствах мультипликативной группы. Этот метод придуман Поллардом [25] и известен под именем *$(p-1)$-метода Полларда.*

Поиск нетривиального делителя составного числа $n>1$ с помощью $(p-1)$-метода Полларда состоит в следующем. Выберем произвольно $a\in\mathbb{Z}/n\mathbb{Z}$ и некоторое положительное целое число k, которое делится на много степеней малых простых чисел. Например, в качестве k можно взять наименьшее общее кратное чисел $1, 2, \ldots, w$ для подходящего числа w. Затем вычисляется $a_k=(a^k \bmod n)$. Это вычисление может быть реализовано за $O(\log k)$ возведений в квадрат и умножений по модулю n. Наконец, с помощью алгоритма Эвклида вычисляется н.о.д. (a_k-1, n) в надежде, что он является нетривиальным делителем n.

Обычно $(p-1)$-метод Полларда приводит к успеху, только если n имеет простой делитель p, для которого $p-1$ есть произведение малых простых чисел. Предположим, например, что $p-1$ делит k, а p не делит a. Так как порядок группы $(\mathbb{Z}/p\mathbb{Z})^*$ равен $p-1$, получаем $a^k\equiv 1 \bmod p$. Следовательно, p делит н. о. д. (a_k-1, n). Во многих случаях оказывается, что $p=$ н. о. д. (a_k-1, n). Тем самым мы находим нетривиальный делитель n.

Можно доказать, что $(p-1)$-метод Полларда хорошо работает при поиске простых делителей p числа r в том случае, когда $p-1$ не имеет больших простых делителей. Можно также доказать, что в случае, если n не имеет таких простых делителей, вероятность получить результат в пределах обозримого времени очень мала.

В [25] описан усовершенствованный вариант этого метода, дающий лучшие практические результаты. В [35] предложен вариант, использующий скрученную мультипликативную группу; здесь уже не число $p-1$, а число $p+1$ должно состоять из малых простых делителей. Наконец, в [31] описано обобщение, имеющее, по-видимому, лишь теоретическое значение.

Преимущество эллиптических кривых в методах факторизации в том же, что и в случае тестов на простоту. Если вместо мультипликативной группы использовать эллиптическую кривую, то число $p\pm1$ заменяется некоторым числом в окрестности p, которое зависит от кривой. Эту кривую можно изменять до тех пор, пока алгоритм не приведет к успеху. Можно рассчитывать на то, что значительная часть чисел в окрестности p построена из одних только малых простых, поэтому перебрать понадобится не так много кривых. Подробности см. в § 6.

§ 3. ЭЛЛИПТИЧЕСКИЕ КРИВЫЕ НАД КОЛЬЦАМИ

Пусть R — некоторое кольцо. Будем называть конечный набор $(a_i)_{i\in I}$ элементов кольца R *примитивным*, если он порождает R как R-идеал $\left(\text{т. е. существуют такие элементы } b_i \in R,\ i\in I, \text{ что } \sum_{i\in I} b_i a_i = 1\right)$. Этот термин будет применяться, в частности, к векторам и к матрицам с коэффициентами в R. В случае когда R — поле, набор $(a_i)_{i\in I}$ примитивен тогда и только тогда, когда не все a_i равны нулю.

В дальнейшем мы предполагаем, что R удовлетворяет следующим двум условиям:

(i) $6 \in R^*$;

(ii) для всех положительных n, m и каждой примитивной матрицы $(a_{ij})_{1\leqslant i\leqslant n,\ 1\leqslant j\leqslant m}$ над R, все 2×2-миноры которой обращаются в нуль (т. е. $a_{ij}a_{kl} - a_{il}a_{kl} = 0$ для всех i, j, k, l, подчинённых условию $1 \leqslant i < k \leqslant n$, $1 \leqslant j < l \leqslant m$) существует примитивный элемент из R^m, являющийся R-линейной комбинацией строчек этой матрицы.

Если R — поле, то первое условие эквивалентно тому, что $\operatorname{char} R \neq 2, 3$. Нам это условие нужно лишь для упрощения изложения, поскольку в случае, когда $6 \notin R^*$, необходимо использовать более общие нормальные формы эллиптических кривых (см. [35, гл. 3]).

Второе же условие является существенным для того определения эллиптических кривых и закона сложения на них, которое мы дадим. Условие (ii) означает, что каждый *проективный R-модуль* ранга 1 является свободным, другими словами, что *группа Пикара* $\operatorname{Pic} R$ кольца R равна нулю [4]. Это условие, очевидно, выполняется для полей. Ниже мы увидим, что оно также выполнено для конечных колец. Более общим образом, ему удовлетворяют кольца с конечным числом максимальных идеалов. Если R — дедекиндово кольцо, например кольцо целых числового поля, то условие (ii) выполнено только в том случае, когда группа классов идеалов кольца R является тривиальной.

Легко доказать, что примитивный элемент в R^m, существование которого утверждается в (ii), в действительности определён однозначно с точностью до умножения на обратимый элемент кольца.

Пусть R — кольцо, удовлетворяющее условиям (i) и (ii). Группа R^* его обратимых элементов действует на множестве примитивных троек $(x, y, z) \in R^3$ по правилу $u(x, y, z) = (ux, uy, uz)$. Множество орбит относительно этого действия обозначим через $\mathbb{P}^2(R)$ и назовём *проективной плоскостью* над кольцом R. Орбиту элемента (x, y, z) будем обозначать через $(x : y : z)$.

Эллиптическая кривая над R — это пара элементов $a, b \in R$, для которых $4a^3 + 27b^2 \in R^*$. Можно рассматривать эту пару элементов как коэффициенты однородного уравнения Вейерштрасса

$$y^2 z = x^2 + axz^2 + bz^3.$$

Эллиптическую кривую (a, b) мы будем обозначать через $E_{a,b}$ или просто E. Если записанное выше уравнение умножить на u^6 для некоторого $u \in R^*$ и заменить u^2x, u^3y на x и y, то мы придём к уравнению для $E_{a',b'}$, где $a' = u^4 a$, $b' = u^6 b$. Две такие кривые $E_{a,b}$ и $E_{a',b'}$ называются *изоморфными* над R.

Пусть $E = E_{a,b}$ — эллиптическая кривая над R. Определим *множество точек* $E(R)$ кривой E над R формулой

$$E(R) = \{(x : y : z) \in \mathbb{P}^2(R) : y^2 z = x^3 + axz^2 + bz^3\}.$$

Точку $(0 : 1 : 0) \in E(R)$ назовём *нулевой точкой* кривой и обозначим через O. Заметим, что в случае, когда R — поле, она является единственным элементом множества $E(R)$ с нулевой координатой z.

Важным фактом является то, что $E(R)$ обладает естественной структурой *абелевой группы* с нейтральным элементом O. При этом групповой закон, записываемый аддитивно, удовлетворяет для всех $(x : y : z) \in E(R)$ условию $-(x : y : z) = (x : -y : z)$. Для того чтобы определить этот групповой закон, рассмотрим сначала случай, когда R — поле. Для этого случая формулы сложения и проверку аксиом группы для $E(R)$ можно найти в [35, гл. 3]. Кратко изложим необходимое.

Пусть R — поле и P_1, $P_2 \in E(R)$. Для того чтобы сложить P_1 и P_2, рассмотрим прямую, проходящую через P_1 и P_2 (или касательную к кривой $E(R)$, если $P_1 = P_2$). Эта прямая имеет с $E(R)$ три точки пересечения, если считать их с учётом кратности; две из них — это P_1 и P_2. Если Q — третья точка, то полагаем по определению $P_1 + P_2 = -Q$. При переводе этого геометрического описания в алгебраические формулы мы можем считать, что $P_1 \neq 0$, $P_2 \neq 0$ и $P_1 \neq -P_2$. Для $i = 1, 2$ можно представить P_i в виде $(x_i : y_i : 1)$, где (x_i, y_i) лежит на аффинной кривой $y^2 = x^3 + ax + b$. Наша прямая определяется уравнением $y = \lambda x + \nu$, где

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1} \quad \text{или} \quad \lambda = \frac{x_2^2 + x_2 x_1 + x_1^2 + a}{y_2 + y_1},$$

а $\nu = y_1 - \lambda x_1$. Заметим, что из условия $P_1 \neq -P_2$ следует, что по крайней мере одно из указанных значений для λ корректно определено, причём эти значения совпадают, если они корректно определены оба. Теперь сумма $P_3 = P_1 + P_2$ может быть представлена в виде $P_3 = (x_3 : y_3 : 1)$, где

$$x_3 = \lambda^2 - x_1 - x_2, \quad y_3 = -(\lambda x_3 + \nu).$$

Это уже даёт формулу сложения, если R — поле, но для дальнейшего желательно привести её к однородному виду. С этой целью заменяем x_i и y_i соответственно на x_i/z_i и y_i/z_i, а затем избавляемся от знаменателей. В результате находим, что сумма двух точек $P_1 = (x_1 : y_1 : z_1)$ и $P_2 = (x_2 : y_2 : z_2)$ на $E(R)$ определяется одной из формул $(q_1 : r_1 : s_1)$ или $(q_2 : r_2 : s_2)$, в зависимости от того, какая формула использовалась для λ. В этих формулах $q_1, \ldots, s_2$ — некоторые определенные многочлены от $x_1, y_1, z_1, x_2, y_2, z_2, a, b$ с целыми коэффициентами. Оказывается, что, за исключением пары (O, O), для каждой пары $(P_1, P_2) \in E(R) \times E(R)$ по крайней мере одна из двух формул сложения имеет смысл (т. е. не даёт набора $(0:0:0)$) и что любая из формул, имеющих смысл, действительно даёт сумму P_1 и P_2 в группе $E(R)$. Что касается оставшейся пары (O, O), то мы, конечно, знаем, что $O + O = O = (0:1:0)$, но наша формула является неудовлетворительной из-за того, что не обладает свойством давать значение суммы $P_1 + P_2$ для *всех* точек P_1, P_2, для которых эта сумма определена. Для исправления этого дефекта нужно преобразовать формулу сложения таким образом, чтобы она стала справедливой «в окрестности пары (O, O)». Это было сделано в [35, гл. IV, § 1]. Были найдены девять многочленов q_i, r_i, s_i $(i = 1, 2, 3)$ от переменных $x_1, y_1, z_1, x_2, y_2, z_2, a, b$ с целыми коэффициентами, удовлетворяющих такому условию: сумма *любых* двух точек $P_1 = (x_1 : y_1 : z_1)$ и $P_2 = (x_2 : y_2 : z_2)$ на $E(R)$ задаётся одним из трёх выражений $(q_i : r_i : s_i)$, $i = 1, 2, 3$, причём любое из этих выражений, имеющее смысл, подходит. Это эквивалентно девяти формальным тождествам $q_1 r_2 - q_2 r_1 = 0, \ \ldots, \ r_2 s_3 - r_3 s_2 = 0$ в кольце $\mathbb{Z}[\alpha, b, x_1, y_1, z_1, x_2, y_2, z_2]/I$, где $a, \ldots, z_2$ считаются переменными, а I обозначает идеал, порождённый двумя многочленами $y_i^2 z_i - x_i^3 - a x_i z_i^2 - b z_i^3$, $i = 1, 2$. Похожим способом с помощью набора формальных тождеств в этом же кольце выражается условие принадлежности $P_1 + P_2$ кривой $E(R)$ и условие выполнения аксиом группы с вышеуказанным нейтральным элементом и с введённой операцией взятия обратного элемента. Явный вид многочленов $q_1, \ldots, s_3$ с указанными свойствами можно найти в [19].

Пусть теперь кольцо R необязательно является полем. В этом случае для сложения двух точек $P_1 = (x_1 : y_1 : z_1)$ и $P_2 = (x_2 : y_2 : z_2)$ на кривой $E(R)$ нужно поступить следующим образом. С помощью тех же девяти многочленов вычислим 3×3-матрицу

$$\begin{pmatrix} q_1 & r_1 & s_1 \\ q_2 & r_2 & s_2 \\ q_3 & r_3 & s_3 \end{pmatrix}$$

с элементами из кольца R. Эта матрица является примитивной. Действительно, в противном случае существовал бы максимальный идеал $\mathfrak{m} \subset R$, содержащий все девять элементов этой матрицы. Но последнее противоречит тому, что по крайней мере одну из строчек матрицы можно использовать для сложения двух точек $P_1 \bmod \mathfrak{m}$, $P_2 \bmod \mathfrak{m}$ на эллиптической кривой $E_{a \bmod \mathfrak{m},\, b \bmod \mathfrak{m}}(R/\mathfrak{m})$ над полем $R/\mathfrak{m}$. Так как все 2×2-миноры нашей матрицы равны нулю, из условия (ii) на кольцо R следует, что существует R-линейная комбинация (q_0, r_0, s_0) строк матрицы, являющаяся примитивной. Далее, орбита точки (q_0, r_0, s_0) при действии группы R^* определена однозначно. Теперь берём $(q_0 : r_0 : s_0)$ в качестве суммы P_1 и P_2 на $E(R)$.

Проверить замкнутость $E(R)$ относительно введённой операции сложения и выполнение аксиом группы можно с помощью упоминавшихся выше формальных тождеств. Мы опускаем эти довольно скучные подробности.

Возникает естественный вопрос об *алгоритме* сложения двух точек на кривой $E(R)$. Из определения сложения мы немедленно получаем, что при использовании формул из [19] достаточно алгоритмизировать проверку условия (ii). Иначе говоря, для примитивной линейной комбинации, существование которой утверждается, нужен явный метод её *нахождения*. Перед тем как дать описание такого метода для случая *конечного* кольца R, заметим, что в данной ситуации он имеет чисто теоретическое значение. А именно, для достижения целей, которые мы имеем в виду (см. следующие параграфы), можно поступить гораздо более простым способом. Выберем любой ненулевой элемент из матрицы и посмотрим, обратим ли он в R. Если да, то строчка, содержащая выбранный элемент, примитивна, и мы достигли требуемого. В противном случае мы получаем ненулевой необратимый элемент R, и во всех случаях, с которыми мы будем иметь дело, этого уже достаточно. Например, предположим, что $R = \mathbb{Z}/n\mathbb{Z}$, где n — целое число, которое мы пытаемся разложить на множители. Тогда ненулевой необратимый элемент R приводит нас к нетривиальному делителю n. Это в точности то, что нам нужно.

Пусть теперь R — *конечное* кольцо. Предположим, что элементы кольца R представлены элементами некоторого конечного множества S. В качестве S можно взять, например, двоичные строчки, состоящие из нулей и единиц. Допускается возможность, что два различных элемента s, s' из S представляют один и тот же элемент кольца R, но требуется, чтобы для распознавания этого случая для данных $s, s' \in S$ существовал эффективный алгоритм. Здесь под словом «эффективный» подразумевается, что время, необходимое для работы алгоритма, ограничено полиномиальной функцией от $\log \# S$. Мы также потребуем существования эффективного алгоритма для операции *сложения*, позволяющего для данных $s, s' \in S$ найти элемент из S, соответствующий сумме элементов кольца R, представляемых элементами s и s'. Точно так же мы требуем эффективности в выполнении операций *вычитания*, *умножения* и в решении уравнений вида $cx = d$ (для данных c и d ищется x), если они разрешимы.

При этих предположениях существует эффективный алгоритм, который выдаёт примитивную линейную комбинацию строчек данной примитивной $n \times m$-матрицы (a_{ij}), удовлетворяющей условию (ii). Здесь опять «эффективность» означает, что необходимое время работы алгоритма ограничено полиномиальной функцией от n, m и $\log \# S$. Начнём с одной леммы.

Лемма. *Для введённых выше R и S обозначим через t наименьшее положительное целое число, для которого $2^{t+1} > \# S$. Тогда для каждого $c \in R$ существует элемент $x \in R$, удовлетворяющий равенству $c^{t+1}x = c^t$. Далее, элемент $c \in R$ является нильпотентным тогда и только тогда, когда $c^t = 0$.*

Доказательство. Рассмотрим последовательность идеалов

$$R \supset Rc \supset Rc^2 \supset \ldots \supset Rc^t \supset Rc^{t+1}.$$

Если любые два последовательных идеала в этой цепочке различны, то приходим к неравенствам $\# S \geqslant \# R \geqslant \operatorname{ind}[R : Rc^{t+1}] \geqslant 2^{t+1}$, что невозможно. Следовательно, $c^i = c^{i+1}x$ для некоторого $x \in R$ и некоторого номера i, где $0 \leqslant i \leqslant t$. Первое утверждение леммы получается теперь умножением на c^{t-i}.

Если целое число u больше t, то $c^u x = c^{u-1}$. Поэтому в случае нильпотентности элемента c наименьшее целое число u, для которого $c^u = 0$, не может быть больше t. Отсюда следует второе утверждение леммы.

Из леммы вытекает существование эффективного алгоритма, выясняющего, является ли данный элемент кольца нильпотентным.

Теперь опишем эффективный алгоритм, который для данной $n \times m$-матрицы $A = (a_{ij})$, удовлетворяющий условию (ii), находит примитивную комбинацию её строчек. Алгоритм работает рекурсивно относительно мощности R. Если R — нулевое кольцо, то любая строчка матрицы примитивна. Предположим теперь, что R — ненулевое кольцо. Так как матрица A примитивна, не все её элементы нильпотентны. Пусть c — её ненильпотентный элемент. Используя лемму, решим уравнение $c^{t+1}x = c^t$. Получим $c^{2t}x^t = c^t$. Следовательно, если мы положим $e = c^t x^t$, то e — идемпотент: $e^2 = e$. Из равенства $c^t e = c^t \neq 0$ следует, что $e \neq 0$. Если теперь $e = 1$, то c — обратимый элемент, строчка матрицы A, содержащая c, будет примитивной и дело сделано. Предположим поэтому, что $e \neq 1$. Тогда $R_1 = Re$ и $R_2 = R(1-e)$ — ненулевые коммутативные кольца с единичными элементами e и $1-e$ соответственно. Далее, отображение $R \to R_1 \times R_2$, переводящее $r \in R$ в $(re, r(1-e))$, является изоморфизмом колец. Матрица A определяет матрицу A_1 над R_1 и матрицу A_2 над R_2. Теперь заметим, что для каждого $i = 1, 2$ отображение $S \to R \to R_i$ показывает, что множество S можно снова использовать для представления элементов R_i, при котором выполнены те же условия, что и для R. Следовательно, применяя рекурсию, мы можем найти R_i-линейные комбинации строчек матрицы A_i, дающие примитивные элементы R_i^m для каждого $i = 1, 2$. Складывая эти две строчки в R^m, находим требуемую примитивную линейную комбинацию строчек A, чем описание алгоритма и завершено.

Заметим, что в этом алгоритме элемент $c \in R$ отображается в пару $(c_1, c_2) \in R_1 \times R_2$, в которой элемент c_1 обратим, а элемент c_2 нильпотентен. Следовательно, строчка A_1, содержащая c_1, уже является примитивной, и применять рекурсию нужно лишь для кольца R_2. Так как число нильпотентных элементов в A_2 по крайней мере на 1 больше, чем в матрице A, это рассуждение показывает, что глубина рекурсии ограничена числом nm. В интересующем нас случае имеем $nm = 9$.

§ 4. ЧИСЛО ТОЧЕК НА ЭЛЛИПТИЧЕСКОЙ КРИВОЙ

Пусть R — конечное кольцо, причём $6 \in R^*$, а $E = E_{a,b}$ — эллиптическая кривая над R. В этом параграфе мы обсудим вопрос о порядке группы $E(R)$.

Если f: $R \to R'$ — произвольный кольцевой гомоморфизм из R в некоторое кольцо R', которое также удовлетворяет условиям (i) и (ii) из § 3, то $E_{f(a), f(b)}$ — эллиптическая кривая над R'. Мы снова обозначим эту эллиптическую кривую через E.

Если R содержит элемент c, который не является ни обратимым, ни нильпотентным, то, как мы видели в предыдущем параграфе, R может быть записано в виде произведения двух ненулевых колец. Индукцией по мощности R получаем, что R изоморфно произведению конечного числа колец R_i, где каждое R_i состоит лишь из обратимых и нильпотентных элементов. В этом случае R_i — локальное кольцо. Это означает, что множество $\mathfrak{m}_i$ необратимых элементов кольца R_i образует идеал в R_i. Этот идеал должен быть максимальным, поэтому $R_i/\mathfrak{m}_i$ — поле. Теперь легко доказать, что $E(R)$ изоморфно произведению групп $E(R_i)$, откуда $\#E(R = \Pi_i \#E(R_i)$. Далее, из леммы Гензеля вытекает, что для каждого i естественный групповой гомоморфизм $E(R_i) \to E(R_i/\mathfrak{m}_i)$ является *сюръективным* и его ядро имеет ту же мощность, что и $\mathfrak{m}_i$, откуда $\#E(R_i) = \#E(R_i/\mathfrak{m}_i)\#\mathfrak{m}_i$. Резюмируя, имеем

$$\frac{\# E(R)}{\# R} = \prod_{m} \frac{\# E(R/\mathfrak{m})}{\# R/\mathfrak{m}},$$

где $\mathfrak{m}$ пробегает множество максимальных идеалов кольца R. Если эти максимальные идеалы известны, то выписанная выше формула сводит вычисление $\#E(R)$ к случаю, когда R — некоторое поле. Если $R = \mathbb{Z}/n\mathbb{Z}$ для некоторого положительного целого n, то наша формула принимает вид

$$\frac{\# E(Z/nZ)}{n} = \prod_{p} \frac{\# E(\mathbb{F}_p)}{p},$$

где p пробегает множество всех простых делителей числа n. Заметим, что аналогичная формула справедлива для порядка мультипликативной группы кольца вычетов, если заменить порядок группы точек на эллиптической кривой на значение функции Эйлера.

В оставшейся части этого параграфа будем предполагать, что R — конечное поле, характеристика которого не равна 2 или 3. Обозначим через q мощность R, так что можно записать $R = \mathbb{F}_q$. Как и в предыдущем параграфе, примем, что имеется явное представление для элементов R, причём каждая арифметическая операция в R может быть выполнена за время $O((\log q)^2)$.

Согласно теореме Хассе (1934 г.), $\#E(\mathbb{F}_q) = q + 1 - t$, где t — целое число, удовлетворяющее неравенству $|t| \leqslant 2\sqrt{q}$. Для вычисления $\#E(\mathbb{F}_q)$ или, что то же, числа t предлагались четыре способа.

Первый способ, применённый Ленгом и Троттером [18], опирается на формулу

$$\# E(\mathbb{F}_q) = 1 + \sum_{x \in \mathbb{F}_q} (1 + \chi(x)),$$

в которой $\chi(x)$ обозначает значение $(x^3 + ax + b)^{(q-1)/2}$, принадлежащее множеству $\{0, 1, -1\}$. (Чтобы доказать эту формулу, достаточно просто заметить, что для фиксированного $x \in \mathbb{F}_q$ количество элементов $y \in \mathbb{F}_q$, удовлетворяющих условию $y^2 = x^3 + ax + b$, равно $1 + \chi(x)$.) Непосредственное применение указанной формулы приводит к алгоритму вычисления $\#E(\mathbb{F}_q)$, который требует времени $O(q^{1+\varepsilon})$, где ε — сколь угодно малое положительное число.

Второй метод, который работает значительно быстрее, является *вероятностным*, поскольку в нём используются случайные выборки. Этот метод близок к алгоритму Шенкса [33] для вычисления числа классов мнимого квадратичного поля. Краткое описание его состоит в следующем.

Сначала случайным образом выбирается точка $P \in E(\mathbb{F}_q)$. Это можно осуществить, случайным образом выбирая элементы $x \in \mathbb{F}_q$, до тех пор пока не будет найден такой элемент x, для которого $x^3 + ax + b$ — квадрат в $\mathbb{F}_q$. Последнее проверяется выполнением неравенства $\chi(x) \neq -1$, где χ — функция, определённая выше. После того как требуемый элемент x найден, можно, применяя другой вероятностный алгоритм Шенкса [34] или рутинную технику отыскания корней многочленов в конечных полях, найти элемент $y \in \mathbb{F}_q$, удовлетворяющий равенству $y^2 = x^3 + ax + b$. Таким образом, на кривой получаем точку $P = (x : y : 1)$.

Следующий шаг — определение всех целых чисел m, для которых $|m - (q + 1)| \leqslant 2\sqrt{q}$ и $m \cdot P = O$. Очевидно, такие целые числа существуют; например, этим свойством обладает число $m = \#E(\mathbb{F}_q)$. Все такие целые m могут быть найдены за время $O(q^{(1/4)+\varepsilon})$ для сколь угодно малого $\varepsilon > 0$ с помощью стратегии «больших и малых шагов», за подробным описанием которой отсылаем читателя к [33].

Если число m с указанными свойствами единственно, то $m = \#E(\mathbb{F}_q)$, и требуемое число найдено. Если же m не является единственным, то разность между двумя такими последовательными числами равна порядку точки P в группе $E(\mathbb{F}_q)$, и нетрудно доказать, что P не может порождать всей группы $E(\mathbb{F}_q)$, если $q \geqslant 37$. В последнем случае выбирается другая случайная точка $P' \in E(\mathbb{F}_q)$ и аналогичным способом определяется порядок точки по модулю подгруппы, порожденной точкой P'. Этот процесс продолжается до тех пор, пока не будет найдена подгруппа, порядок которой k удовлетворяет неравенству $|k-(q+1)|\leqslant 2\sqrt{q}$. Если $q\geqslant 37$, то $\#E(\mathbb{F}_q)=k$.

Ожидаемое время работы этого алгоритма $O(q^{(1/4)+\varepsilon})$ для сколь угодно малого $\varepsilon>0$. При этом в результате определяется не только порядок группы $E(\mathbb{F}_q)$, но и её групповая структура. Это практически важно, если q имеет не более чем около 20 десятичных цифр.

Третий метод, который мы обсудим, принадлежит Скуфу [30]. Этот метод вполне детерминистический. Он опирается на свойства *эндоморфизма Фробениуса* φ кривой, который определяется следующим образом. Пусть K — некоторое алгебраическое замыкание поля $\mathbb{F}_q$. Тогда φ — это автоморфизм абелевой группы $E(K)$, задаваемый равенством

$$\varphi(x:y:z)=(x^q:y^q:z^q).$$

Заметим, что $E(\mathbb{F}_q)$ можно рассматривать как подгруппу в $E(K)$, причём $E(\mathbb{F}_q)=\{P\in E(K):\varphi(P)=P\}$. Основной факт состоит в том, что φ удовлетворяет квадратному уравнению $\varphi^2-t\varphi+q=0$ в кольце эндоморфизмов $E(K)$, где $t=q+1-\#E(\mathbb{F}_q)$. Заметим, что для нахождения t достаточно найти вычеты t по модулю всех нечётных простых $l<c_1\log q$, не равных характеристике поля $\mathbb{F}_q$, где положительная константа c_1 выбирается из условия $\Pi l>4\sqrt{q}$ для всех q. А именно, если известны все эти сравнения t по модулю l, то число t по модулю Πl можно определить с помощью китайской теоремы об остатках. Поскольку $|t|\leqslant 2\sqrt{q}$, этого достаточно для нахождения t и, следовательно, $\#E(\mathbb{F}_q)$.

Пусть теперь l — нечётное простое число, $l\neq\operatorname{char}\mathbb{F}_q$. Чтобы вычислить $t \bmod l$, вычисляют многочлен ψ_l, задаваемый формулой

$$\psi_l=l\cdot\Pi(X-x),$$

где x пробегает множество всех элементов поля K, для которых существует точка $(x:y:1)$ порядка l на кривой $E(K)$. Известно, что ψ_l имеет степень $(l^2-1)/2$ и принадлежит $\mathbb{F}_q[X]$. Многочлен ψ_l можно вычислить с помощью рекуррентных формул, которые приводятся, например, в [35, гл. III, упр. 3.7].

Определим кольцо T равенством

$$T=\mathbb{F}_q[X,Y]/(\psi_l,\ Y^2-X^3-aX-b).$$

Каждый элемент кольца T допускает единственное представление в виде

$$\sum_{i=0}^{(l^2-3)/2}\sum_{j=0}^{1}a_{ij}\bar{X}^i\bar{Y}^j \quad \text{с} \quad a_{ij}\in\mathbb{F}_q,$$

где $\bar{X}$, $\bar{Y}$ обозначают образы X, Y в T. Следовательно, T — конечное кольцо, в котором кольцевые операции могут быть выполнены эффективно в смысле § 3.

Пусть $Q=(\bar{X}:\bar{Y}:1)\in E(T)$. Определим эндоморфизм $\sigma:E(T)\to E(T)$ той же самой формулой, что и φ:

$$\sigma(x:y:z)=(x^q:y^q:z^q).$$

Как мы сейчас покажем, точки Q и $\sigma(Q)$ имеют порядок l и σ удовлетворяет уравнению $\sigma^2-t\sigma+q=0$ в кольце эндоморфизмов $E(T)$. Поэтому t по модулю l характеризуется равенством

$$\sigma^2(Q)+qQ=t\cdot\sigma(Q).$$

Таким образом, для нахождения t по модулю l можно просто вычислить левую часть этого равенства и сравнить её с $0\cdot\sigma(Q)$, $1\cdot\sigma(Q)$, $2\cdot\sigma(Q)$, … . Вычисления в $E(T)$ можно выполнить, следуя § 3.

Для доказательства указанных выше свойств Q и σ рассмотрим множество V, состоящее из всех точек $P\in E(K)$ порядка l. Для каждой такой точки $P=(x_P:y_P:1)$ существует единственный $\mathbb{F}_q$-линейный кольцевой гомоморфизм $T\to K$, переводящий $\bar{X}$ и $\bar{Y}$ соответственно в x_P и y_P. Непосредственная проверка показывает, что получающийся гомоморфизм колец $T\to\prod_{P\in V}K$ *инъективен*, поэтому $E(T)$ можно рассматривать как подгруппу в $\prod_{P\in V}E(K)$. Точка Q имеет порядок l, поскольку она соответствует $(P)_{P\in V}$. Далее, отображение σ является ограничением на $E(T)$ автоморфизма $\prod_{P\in V}E(K)$, который на каждом сомножителе определяется автоморфизмом Фробениуса φ. Следовательно, равенство $\sigma^2+t\sigma+q=0$ — следствие равенства $\varphi^2-t\varphi+q=0$. Очевидно, отображение σ инъективно, поэтому $\sigma(Q)$ имеет порядок l. Этим наш эскизный набросок алгоритма Скуфа завершен.

Алгоритм Скуфа полностью детерминистичен, и можно показать, что его время работы есть $O((\log q)^8)$. (Эта оценка немного лучше оценки $O((\log q)^9)$, приведённой у самого Скуфа в [30].) Однако этот алгоритм по всей видимости мало пригоден для практических вычислений.

Заметим, что структуру абелевой группы $E(\mathbb{F}_q)$ алгоритм Скуфа не вычисляет. Известно, что $E(F_q) \simeq \mathbb{Z}/d_1\mathbb{Z} \times \mathbb{Z}/d_2\mathbb{Z}$ для некоторых положительных целых чисел d_1, d_2, одно из которых, скажем d_1, делит другое, и что это число d_1 делит н. о. д. $(\#E(\mathbb{F}_q),\ q-1)$. В. Миллер показал, что, зная разложение на простые множители этого наибольшего общего делителя, можно при помощи вероятного алгоритма с ожидаемым временем работы $O\left((\log q)^{c_2}\right)$ (для некоторого $c_2 > 0$) найти d_1 и d_2. За описанием этого алгоритма, в котором используется *спаривание Вейля*, отсылаем к [22].

Четвёртый метод вычисления $\#E(\mathbb{F}_q)$ применяется лишь для кривых E, получаемых специальным образом. Для простоты ограничимся обсуждением случая, когда q — *простое число*.

Полем комплексного умножения эллиптической кривой E над $\mathbb{F}_q$ называется поле $L = \mathbb{Q}((t^2-4q)^{1/2})$, где $t = q+1 - \#E(\mathbb{F}_q)$. Это — мнимое квадратичное поле, и его кольцо целых A содержит некоторый корень π многочлена $X^2 - tX + q$. Имеем $\pi + \bar{\pi} = t$, $\pi \cdot \bar{\pi} = q$, $\# E(\mathbb{F}_q) = (\pi - 1)(\bar{\pi} - 1)$. Это даёт простой способ вычисления $\#E(\mathbb{F}_q)$ при условии, что поле L известно (выполняющемся для некоторых специальных кривых). Проиллюстрируем этот метод на двух примерах, которые по существу знал ещё Гаусс. По поводу доказательств см. [15, гл. 18], а также [12, § 7] и [5].

Сначала предположим, что $q \equiv 1 \bmod 3$ и для кривой $E = E_{a,b}$ число a равно 0. Тогда, как можно проверить, $L = Q(\sqrt{-3})$. Кольцо целых A в L имеет вид $A = \mathbb{Z}[(1 + \sqrt{-3})/2]$. Для нахождения элемента $\pi \in A$, удовлетворяющего условиям $\# F(\mathbb{F}_q) = (\pi - 1)(\bar{\pi} - 1)$ и $\pi \cdot \bar{\pi} = q$, начнём с того, что найдём идеал, для которого $Aq = \mathfrak{q} \cdot \bar{\mathfrak{q}}$. Это можно сделать следующим образом.

Подберём такое целое число d, что $d^2 \equiv -3 \bmod q$. Это можно выполнить одним из трёх способов. Первый из них состоит в применении техники нахождения корней для многочленов над конечными полями (см. [17, § 4.6.2]). Второй — в применении алгоритма извлечения квадратного корня [34]. Третий — в том, чтобы случайно выбирать элементы $u \in \mathbb{F}_q^*$ до тех пор, пока не попадётся элемент, для которого $u^{(q-1)/3} \neq 1$, а тогда взять $d \equiv 2u^{(q-1)/3} + 1 \pmod q$. Все эти три метода вероятностные и используются на практике.

Предположим, что требуемое d найдено. Добавляя при необходимости q к d, можно считать, что d нечётно. Тогда $\mathfrak{q} = \mathbb{Z}q + \mathbb{Z}(d + \sqrt{-3})/2$ — простой идеал в A, делящий q и $\mathfrak{g} \cdot \bar{\mathfrak{q}} = Aq$.

На следующем шаге отыскивается образующий элемент $\pi \in \mathfrak{q}$ для идеала $\mathfrak{q}$ (т. е. элемент, для которого $\mathfrak{q} = A\pi$). Это можно сделать, найдя в $\mathfrak{q}$ вектор наименьшей длины. Для этого существует стандартный алгоритм редукции. Другой возможный способ заключается в вычислении н. о. д. $(q, (d + \sqrt{-3}/2)$ с помощью алгоритма Эвклида, который работает в A. Заметим, что π определяется идеалом $\mathfrak{q}$ однозначно с точностью до обратимых элементов кольца A, каковых имеется шесть.

Пусть теперь ζ — единственный корень шестой степени из единицы в A, для которого $b^{(q-1)/6} \equiv \zeta \bmod \mathfrak{q}$; здесь b — число, определяющее кривую $E = E_{0,b}$. Умножая π на подходящий корень шестой степени из единицы, можно добиться, чтобы $\pi \equiv \bar{\zeta} \bmod 2\sqrt{-3}$. Тогда имеем

$$\# E(\mathbb{F}_q) = (\pi - 1)(\bar{\pi} - 1) = q + 1 - 2\,\mathrm{Re}(\pi).$$

Можно доказать, что группа $E(\mathbb{F}_q)$ изоморфна $A/(\pi - 1)A$ как абелева группа, поэтому описанный метод даёт также структуру группы.

Переходя ко второму примеру, предположим, что $q \equiv 1 \bmod 4$ и для кривой $E = E_{a,b}$ число b равно 0. Тогда, как можно проверить, $L = \mathbb{Q}(\mathrm{i})$, где $\mathrm{i}^2 = -1$. Для этого поля L кольцо целых $A = \mathbb{Z}[\mathrm{i}]$. Как и выше, можно найти такой простой идеал $\mathfrak{q} \subset A$, что $\mathfrak{q} \cdot \mathfrak{q} = Aq$, и такой элемент $\pi \in \mathfrak{q}$, что $\mathfrak{q} = A\pi$. Обозначим через ζ единственный корень четвёртой степени из единицы в A, удовлетворяющий условию $(-a)^{(q-1)/4} \equiv \zeta \bmod \mathfrak{q}$. Умножая, если надо, π на подходящий корень четвёртой степени из единицы, можно считать, что $\pi \equiv \bar{\zeta} \bmod 2(1 + \mathrm{i})$. В этом случае $\# E(\mathbb{F}_q) = (\pi - 1)(\bar{\pi} - 1)$.

Дадим краткий набросок того, каким образом можно обобщить эти результаты на случай произвольного мнимого квадратичного поля L. Пусть A — кольцо целых поля L, и пусть j_L обозначает j-инвариант эллиптической кривой $\mathbb{C}/A$ над полем $\mathbb{C}$ (см. [35, гл. VI]). Известно, что j_L служит нулём некоторого неприводимого многочлена $F_L \in \mathbb{Z}[X]$ со старшим коэффициентом 1 степени, равной числу классов поля L. Методы вычисления F_L приводятся, например, в [37], а также в последнем разделе работы [30]. Случаи $j = 0$ и $j = 1728$ соответствуют только что рассмотренным полям $L = \mathbb{Q}(\sqrt{-3})$ и $\mathbb{Q}(\mathrm{i})$, теперь мы их исключим из рассмотрения.

Пусть $q > 3$ — простое число, не делящее дискриминанта поля L. Существуют методы, аналогичные рассмотренным выше, которые позволяют ответить на вопрос о существовании элемента $\pi \in A$ со свойством $\pi \cdot \bar{\pi} = q$ и найти этот элемент, если он существует (заметим, что в последнем случае он единствен с точностью до знака и сопряжения). Предположим, что элемент π действительно существует. Тогда, как можно показать, многочлен $(F_L \bmod q) \in \mathbb{F}_q[X]$ разлагается на попарно различные линейные множители. Обозначим через j произвольный нуль этого многочлена в $\mathbb{F}_q$. Можно показать, что $j \neq 0$, 1728. Положим $k = j/(1728 - j) \in \mathbb{F}_q^*$ и рассмотрим две эллиптические кривые над $\mathbb{F}_q$

$$E = E_{3k,\,2k}, \quad E' = E_{3kc^2,\,2kc^3},$$

где c — произвольный ненулевой элемент $\mathbb{F}_q$, не являющийся квадратом. Тогда L — поле комплексного умножения для каждой из двух кривых E и E' и пара чисел $\#E(\mathbb{F}_q)$, $\#E'(\mathbb{F}_q)$ совпадает с парой чисел $(\pi - 1)(\bar{\pi} - 1)$, $(-\pi - 1)(-\bar{\pi} - 1)$. Вероятно, существует простой способ выяснить, какая кривая какому числу соответствует, но я его не знаю. На практике этот вопрос можно решить с помощью случайного выбора точки $P \in E(\mathbb{F}_q)$, используя то обстоятельство, что точка P аннулируется числом $\#E(\mathbb{F}_q)$.

На этом завершим наше обсуждение методов вычисления количества точек на эллиптической кривой над конечным полем.

Естественно поинтересоваться, каким образом распределены числа $\#E(\mathbb{F}_q)$, если q фиксировано, а E пробегает все (с точностью до изоморфизма) эллиптические кривые над $\mathbb{F}_q$. В частности, можно спросить, насколько часто данное число встречается в качестве $\#E(\mathbb{F}_q)$. Ответ на последний вопрос в терминах чисел классов мнимых квадратичных порядков получен в основном Дойрингом [13] (см. также [36, 31]). Из результатов Дойринга следует, что если q — *простое* число, то любое целое число вида $q + 1 - t$ с $|t| < 2\sqrt{q}$ встречается в качестве $\#E(\mathbb{F}_q)$ для некоторой эллиптической кривой E над $\mathbb{F}_q$. Далее, можно доказать, что числа $\#E(\mathbb{F}_q)$ распределены почти равномерно в окрестности $q + 1$, если в множестве всех эллиптических кривых над $\mathbb{F}_q$ равномерно распределены кривые $E = E_{a,\,b}$. Точнее говоря, справедливо следующее предложение, которое полезно при анализе алгоритмов, представленных в §§ 5 и 6.

Предложение. *Существуют такие эффективно вычисляемые константы c_3 и c_4, что для любого простого числа $q > 3$ и любого множества S целых чисел s, удовлетворяющих условию $|s - (q + 1)| < \sqrt{q}$ выполнены неравенства*

$$\frac{\# S - 2}{2[\sqrt{q}] + 1} \cdot c_3 (\log q)^{-1} \leqslant \frac{N}{q^2} \leqslant \frac{\# S}{2[\sqrt{q}] + 1} \cdot c_4 (\log q)(\log\log q)^2,$$

где N обозначает количество пар $(a, b) \in \mathbb{F}_q^2$, определяющих эллиптическую кривую $E = E_{a,\,b}$ над $\mathbb{F}_q$, для которой $\#E(\mathbb{F}_q) \in S$.

Заметим, что N/q^2 есть вероятность того, что случайная пара (a, b) обладает сформулированным свойством. Предложение утверждает, что с точностью до логарифмического множителя эта вероятность в существенном равна вероятности того, что случайное число в окрестности q лежит в S.

За доказательством предложения отсылаем к [20, prop. (1.16)].

§ 5. ПРОВЕРКА НА ПРОСТОТУ

То что для проверки на простоту можно использовать эллиптические кривые, впервые было указано в [5] и [8]. Гольдвассер и Килиан [14] показали по модулю разумных предположений, что эта идея приводит к некоторому вероятностному алгоритму проверки на простоту, ожидаемое время работы которого оценивается некоторой фиксированной степенью $\log n$, где n — число, подлежащее проверке. Алгоритм Гольдвассера—Килиана опирается на метод Скуфа для подсчёта числа точек на эллиптической кривой (см. § 4) и поэтому пока не представляет практической ценности. Аткин [2] разработал вариант этого алгоритма, использующий лишь специальные эллиптические кривые, к которым он применил четвёртый из описанных в § 4 методов подсчёта точек. Его алгоритм очень хорошо работает на практике, и для чисел, к которым он применялся (у этих чисел приблизительно 200 десятичных знаков), оставил позади метод Адлемана и др. [1] в реализации Коэна и А. Ленстры [10]. По-видимому, очень трудно дать точную оценку быстродействия алгоритма Аткина, но грубый эмпирический анализ показывает, что ожидаемое время работы опять-таки ограничено некоторой степенью $\log n$.

Все эти методы опираются на результат, близкий к следующей теореме, которая аналогична теореме 1.

Теорема 2. *Пусть $n > 1$ — целое число, удовлетворяющее условию* н. о. д. $(n, 6) = 1$, *E — эллиптическая кривая над кольцом вычетов $\mathbb{Z}/n\mathbb{Z}$, а m и s — положительные целые числа, причём m делится на s. Предположим, что существует точка $P \in E(\mathbb{Z}/n\mathbb{Z})$, такая что*

$$m \cdot P = O,$$

н. о. д. $(z_q, n) = 1$ *для каждого простого делителя q числа s, где $(m(m/q) \cdot P = (x_q : y_q : z_q.)$*

Тогда $\#E(\mathbb{Z}/p\mathbb{Z}) \equiv 0 \bmod s$ *для каждого простого делителя* p *числа* n, *причём если* $s < (n^{1/4}+1)^2$, *то* n — *простое.*

Доказательство этой теоремы, аналогичное доказательству теоремы 1, проводится так. Пусть p — простой делитель числа n, а $Q = (m/s) \cdot P \in E(\mathbb{Z}/n\mathbb{Z})$. Обозначим через Q_p образ Q в $E(\mathbb{Z}/p\mathbb{Z})$. Из равенства $m \cdot P = O$ следует, что $s \cdot Q = O$, поэтому порядок Q_p делит s. С другой стороны, если q — простой делитель s, то точка $s/q \cdot Q_p = (x_q \bmod p : y_q \bmod p : z_q \bmod p)$ не является нулём в $E(\mathbb{Z}/p\mathbb{Z})$, поскольку по условию теоремы z_q не делится на p. Следовательно, порядок Q_p не делит s/q ни для какого простого делителя q числа s. Таким образом, этот порядок равен s, и из теоремы Лагранжа следует, что $\#E(\mathbb{Z}/p\mathbb{Z})$ делится на s. Этим доказано первое утверждение теоремы. Если также выполнено неравенство $s > (n^{1/4}+1)^2$, то из неравенства Хассе $(p^{1/2}+1)^2 \geqslant \#E(\mathbb{Z}/p\mathbb{Z})$ вытекает, что $p > n^{1/2}$. Последнее может выполняться для всех простых делителей p числа n только в случае, когда n само является простым. Теорема доказана.

Для алгоритмов Гольдвассера—Килиана и Аткина нужен лишь тот случай теоремы 2, когда число s *просто*, так что во втором условии на точку P, фигурирующем в формулировке этой теоремы, достаточно рассматривать лишь случай $q = s$. Следующая схема пригодна для обоих алгоритмов.

Пусть n — большое положительное целое число, о котором есть основания предполагать, что оно просто (см. введение). Для доказательства его простоты поступают следующим образом.

(a) Выбираются эллиптическая кривая E над $\mathbb{Z}/n\mathbb{Z}$ и положительное целое число m, такие что

(i) $m < (\sqrt{n}+1)^2$, и *если* n просто,*то* $\#E(\mathbb{Z}/n\mathbb{Z}) = m$;

(ii) существуют такие целые числа $k > 1$ и $q > (n^{1/4}+1)^2$, что $m = kq$, а q является псевдопростым.

Здесь *псевдопростота* означает, что q удовлетворяет некоторому тесту на псевдопростоту, например тесту из [17, с. 379] (см. введение). Для того чтобы найти *одну* пару E, m, удовлетворяющую условиям (i) и (ii), алгоритмы Гольдвассера—Килиана и Аткина генерируют *много* пар E, m, удовлетворяющих (i); ниже мы увидим, как это делается. Есть надежда, что по крайней мере одна из этих пар удовлетворяет также условию (ii). Чтобы проверить, выполняется ли это условие для данной пары E, m, число m подвергают разложению на множители с помощью алгоритма, который эффективно находит его небольшие делители, например с помощью отдельных попыток деления, или $(p-1)$-методом Полларда (см. § 2), или методом эллиптических кривых (см. § 6). Затем выбирают k равным произведению этих найденных небольших делителей m и полагают $q = m/k$. Наконец, проверяют, выполнены ли неравенство $k > 1$ и условие псевдопростоты числа q в описанном выше смысле. (В действительности алгоритм Гольдвассера—Килиана требует равенства $k = 2$ в условии (ii), что облегчает проверку этого условия.)

(b) Теперь предположим, что E, m, k, q из этапа (a) уже найдены. Тогда выбирается случайная точка вида $(x_P : y_P : 1)$ в группе $E(\mathbb{Z}/n\mathbb{Z})$. Это делается так же, как и во втором алгоритме подсчёта, обсуждавшемся в § 4. (Заметим, что этот алгоритм успешно работает, если $\mathbb{Z}/n\mathbb{Z}$ — поле, на что мы надеемся, но для того чтобы алгоритм работал, вовсе не требуется иметь *доказательство* того, что $\mathbb{Z}/n\mathbb{Z}$ — поле!) Далее вычисляется $Q = k \cdot P$ в надежде, что $Q \neq O$. Можно доказать, что это соотношение выполнено для более чем половины выборов P, если n действительно является простым. Если же оказывается, что $Q = O$, то выбираем другую точку $P \in E(\mathbb{Z}/n\mathbb{Z})$, и так до тех пор, пока не выполнится условие $Q = kP \neq O$. Будем теперь считать, что $Q \neq O$. Проверяем, выполнено ли равенство $q \cdot Q = O$, как это должно быть в случае, если n — простое (ввиду равенства $q \cdot Q = m \cdot P$ и условия (i) выше). Наконец, проверяем справедливость равенства н. о. д. $(z, n) = 1$, где $Q = (x : y : z)$. Это свойство также должно быть выполнено, если n — простое число, поскольку $Q \neq O$.

(c) Заключительный этап алгоритма состоит в установлении простоты q. Это можно сделать рекурсивным применением алгоритма или некоторым прямым способом, если q не превосходит определённой границы. Заметим, что $q = m/k < (\sqrt{n}+1)^2/2$, поэтому глубина рекурсии есть $O(\log n)$.

Если все этапы (a)—(c) успешно пройдены, то из теоремы 2 для $s = q$ следует, что n действительно является простым числом.

Осталось объяснить, как найти достаточно много пар (E, m), удовлетворяющих условию (i). В алгоритме Гольдвассера—Килиана это делается следующим образом. Сначала случайно выбирают $a, b \in \mathbb{Z}/n\mathbb{Z}$ до тех пор, пока не окажется выполненным соотношение $4a^3 + 27b^2 \neq 0$. Если n, действительно, простое число, это соотношение имеет место с вероятностью $(n-1)/n$. Затем проверяется, справедливо ли равенство н. о. д. $(n, 4a^3+27b^2) = 1$, которое должно выполняться, если n простое. Теперь полагаем $E = E_{a,b}$ и с помощью алгоритма Скуфа вычисляем число m, для которого выполнено (i). Если алгоритм Скуфа не работает, то n не является простым. (В последнем случае хоть и мало вероятно, но возможно, что алгоритм Скуфа всё-таки вычислит некоторое число m. Интересен вопрос, какую информацию о числе n можно почерпнуть из этого и какова может быть роль этого числа m.)

Метод Аткина для нахождения пар E, m, удовлетворяющих условию (i), совсем другой. Рассмотрим последовательность

$$-3,\ -4,\ -7,\ -8,\ -11,\ -15,\ -19,\ -20,\ \ldots$$

дискриминантов мнимых квадратичных полей. Целое число принадлежит этой последовательности тогда и только тогда, когда оно отрицательно, не делится на квадрат нечётного простого числа и принадлежит одному из следующих классов вычетов: $1 \bmod 4$, $8 \bmod 16$, $12 \bmod 16$. Для каждого числа Δ из подходящего начального отрезка этой последовательности выясняется, содержит ли кольцо целых $A = \mathbb{Z}\left[(\Delta + \sqrt{\Delta})/2\right]$ мнимого квадратичного поля $L = \mathbb{Q}(\sqrt{\Delta})$ элемент π, удовлетворяющий условию $n = \pi \cdot \bar{\pi}$, и в случае, если такой элемент π существует, находят его. Позволяющие это сделать вероятностные методы, за описанием которых отсылаем к § 4, приводят к успеху, если n — простое число, но, как и выше, не нуждаются в *доказательстве* того, что n просто. Дискриминанты, для которых не существует вышеуказанного π, исключаются из рассмотрения, а каждый из оставленных дискриминантов Δ, как объяснялось в § 4, даёт либо шесть (если $\Delta = -3$), либо четыре (если $\Delta = -4$), либо две (если $\Delta \leqslant -7$) пары E, m, удовлетворяющие (i).

Для большинства значений Δ легче определить значения m, чем вычислить коэффициенты a, b, определяющие E. Следовательно, прежде чем вычислять значения a и b, разумно проверить, выполнено ли условие (ii) для m.

Этим завершается описание тестов на простоту Гольдвассера—Килиана и Аткина.

Быстродействие подходящей версии алгоритма Гольдвассера—Килиана может быть проанализировано с помощью предложения, сформулированного в § 4. Результат выражается двумя приводимыми ниже теоремами. Первая из них утверждает, что ожидаемое время работы алгоритма полиномиально, если справедлива некая стандартная гипотеза о распределении простых чисел. Вторая теорема утверждает, что в любом случае этот факт справедлив для почти всех входных простых чисел n.

Теорема 3. *Предположим, что существуют такие положительные константы* c_5 *и* c_6, *что для всех вещественных чисел* $x \geqslant 2$ *количество простых чисел* p, *удовлетворяющих условию*

$x \leqslant p \leqslant x + \sqrt{2x}$, *не меньше* $c_5 \sqrt{x}\,(\log x)^{-c_6}$. *Тогда для любого входного простого числа* n *алгоритм Гольдвассера—Килиана доказывает его простоту за ожидаемое время* $O\left((\log n)^{10+c_6}\right)$.

За доказательством отсылаем к [14]. (В связи с соответствующим усовершенствованием алгоритма Скуфа, показатель $10 + c_6$ здесь на 1 меньше, чем в [14].)

Теорема 4. *Существуют такие положительные константы* c_7 *и* c_8, *что для всех целых чисел* $k \geqslant 2$ *доля множества простых чисел* n, *которые имеют* k *двоичных цифр и для которых ожидаемое время работы алгоритма Гольдвассера — Килиана не больше* $c_7 (\log n)^{11}$, *составляет по крайней мере*

$$1 - c_8 2^{-k^{1/\log\log k}}.$$

Доказательство этой теоремы также приведено в [14]. В нём используется один результат Хит-Брауна, утверждающий, что предположение, фигурирующее в теореме 3, выполняется в некотором усредненном смысле.

§ 6. РАЗЛОЖЕНИЕ НА МНОЖИТЕЛИ

Теперь изложим метод разложения целых чисел на множители, использующий эллиптические кривые. Этот метод представляет собой аналог $(p-1)$-метода Полларда, описанного в § 2.

Пусть n — составное число, которое нужно разложить на множители. Предположим, что $n > 1$ и н. о. д. $(n, 6) = 1$. Выберем случайным образом пару (E, P), где E — эллиптическая кривая над $\mathbb{Z}/n\mathbb{Z}$, а $P \in E(\mathbb{Z}/n\mathbb{Z})$. Это можно сделать следующим образом: сначала выберем a, x, $y \in \mathbb{Z}/n\mathbb{Z}$ и положим $P = (x : y : 1)$, а затем определим E с помощью пары (a, b), где b выбрано так, чтобы выполнялось условие $P \in E(\mathbb{Z}/n\mathbb{Z})$, т. е. $b = y^2 - x^3 - ax$. Для того чтобы быть уверенным, что E — эллиптическая кривая, следует убедиться, что н. о. д. $(4a^3 + 27b^2, n) = 1$. Как и в $(p-1)$-методе Полларда, далее выбирается положительное целое число k, делящееся на много степеней небольших простых чисел, например $k =$ н. о. к. $(1, 2, \ldots, w)$, где w — некоторая подходящая граница. Затем вычисляется точка $k \cdot P \in E(\mathbb{Z}/n\mathbb{Z})$. Это вычисление требует $O(\log k)$ удвоений и сложений в группе $E(\mathbb{Z}/n\mathbb{Z})$. Если $k \cdot P = (x : y : z)$, то вычисляется н. о. д. (z, n). Алгоритм прекращает работу, если этот наибольший общий делитель является нетривиальным делителем числа n. В случае же, если он равен 1 или n, пара (E, P) заменяется другой и всё повторяется снова. Эта возможность нового выбора отсутствует в методе Полларда.

Что касается $(p-1)$-метода Полларда, то можно показать, что данная пара (E, P) с большой вероятностью приводит к успеху, если n имеет простой делитель p, для которого $\# E(\mathbb{Z}/n\mathbb{Z})$ состоит лишь из небольших простых сомножителей; вероятность последнего возрастает с ростом числа испытываемых пар (E, P).

За оценками времени работы алгоритма разложения на множители с помощью эллиптических кривых отсылаем к [20]. Верхнюю оценку ожидаемого времени работы алгоритма находят, используя предложение из § 4 и свойства модулярных кривых. Эта верхняя оценка выражается через вероятность того, что у случайно выбранного в интервале $(p+1-\sqrt{p},\ p+1+\sqrt{p})$ числа все простые делители не превосходят некоторой определённой константы; здесь p обозначает наименьший простой делитель n. Для оценки последней вероятности нам нужна следующая недоказанная гипотеза из аналитической теории чисел.

Для всякого вещественного числа $x > e$ положим

$$L(x) = e^{\sqrt{\log x \log \log x}}.$$

Из теоремы Канфилда, Эрдёша и Померанца [7, следствие теоремы 3.1] вытекает следующее утверждение. Пусть α — положительное вещественное число. Вероятность того, что у случайно выбранного положительного целого числа $m \leqslant x$ все простые делители не превосходят $L(x)^{\alpha}$, равна $L(x)^{-1/(2\alpha)+o(1)}$ при $x \to \infty$. Гипотеза, о которой идёт речь, состоит в том, что то же самое утверждение верно, если m случайно выбирать в интервале $(x-\sqrt{x},\ x+\sqrt{x})$.

Принимая эту гипотезу, мы приходим к следующей оценке времени работы алгоритма разложения на множители, использующего эллиптические кривые. Пусть $n \in \mathbb{Z}$, $n > 1$ — целое число, которое мы хотим разложить на множители. Предположим, что n не является степенью простого числа и не делится на 2 или на 3. Далее, пусть g — произвольное положительное целое число. Тогда метод разложения на множители с помощью эллиптических кривых, описанный в [20], с вероятностью $\geqslant 1-e^{-g}$ находит нетривиальный делитель n за время $gK(p)(\log n)^2$, где p обозначает наименьший простой делитель n, а $K: \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ — некоторая функция, удовлетворяющая условию

$$K(x) = e^{\sqrt{(2+o(1)) \log \log x}} \quad \text{при} \quad x \to \infty.$$

Алгоритм можно повторить для найденных делителей, пока не будет получено полное разложение n на простые множители.

Предполагаемая ты алгоритма будет содер-
жать также член вечающие другим простым
делителям p' чис ый большой. Во всех слу-
чаях можно ожид я для разложения на мно-
жители будет сам $^{1)}$ при $n \to \infty$, где L — вве-
дённая выше фун учай возникает, когда вто-
рой по величине исла n ненамного меньше
$\sqrt{n}$, т. е. n пред зведение какого-то числа
небольших и дву чисел одного порядка ве-
личины.

Предлагалось тодов разложения на мно-
жители, для кото вно $L(n)^{1+o(1)}$ при $n \to \infty$,
например метод] и метод квадратичного
решета [26] (см. того вопроса в [11]). Од-
нако для этих ме алгоритма в основном не
зависит от разме й n, в то время как метод
эллиптических кр твенно быстрее, если наи-
меньший простой меньше $\sqrt{n}$.

Требуемый об и для метода разложения
на множители с ческих кривых составляет
лишь $O(\log n)$. Т тода группы классов [29],
но все другие из разложения на множители
с гипотетической требуют объёма памяти,
представляющего ложительную степень $L(n)$.

Отсылаем чит исанием модификаций ме-
тода эллиптическ добных для практического
применения. Ока ми усовершенствованиями
метод эллиптичес тся одним из самых бы-
стрых методов р тели среди используемых
в настоящее врем ь алгоритм квадратичного
решета, по-видим е на целых числах, пред-
ставляющих собо простых одного порядка
величины; такие кную роль в криптогра-
фии [28].

Благодарность этой статьей была выпол-
нена в Чикагском выразить признательность
за оказанные мне ддержку.

1. L. M. Adlema . S. Rumely, On distinguishing
prime numbers from nn. of Math. (2) **117** (1983),
173—206.
2. A. O. L. Atkin

3. E. Bach and J. Shallit, Factoring with cyclotomic polynomials (26th Annual Sympos. Foundations of Comput. Sci. (FOCS), Portland, 1985) IEEE Comput. Sec. Press, Washington, 1985, pp. 443—450.

4. H. Bass, Algebraic K-theory, Benjamin, New York, 1968. [Имеется перевод: Басс X. Алгебраическая K-теория. — М.: Мир, 1973.]

5. W. Bosma, Primality testing using elliptic curves, Report 85-12, Math. Inst. Universiteit van Amsterdam, 1985.

6. R. P. Brent, Some integer factorization algorithms using elliptic curves, Research report CMA-R32-85, The Australian National Univ., Canberra, 1985.

7. E. R. Canfield, P. Erdös, and C. Pomerance, On a problem of Oppenheim concerning «Factorisatio Numerorum», J. Number Theory **17** (1983), 1—28.

8. D. V. Chudnovsky and G. V. Chudnovsky, Sequences of numbers generated by addition in formal groups and new primality and factorization tests, Research report RC 11262 (#50739), IBM Thomas J. Watson Res. Center, Yorktown Heights, N. Y., 1985.

9. H. Cohen and H. W. Lenstra, Jr., Primality testing and Jacobi sums, Math. Comp. **42** (1984), 297—330.

10. H. Cohen and A. K. Lenstra, Implementation of a new primality test, Math. Comp. **48** (1987), 103—121 and S1—S4.

11. D. Coppersmith, A. M. Odlyzko, and R. Schroeppel, Discrete logarithms in $GF(p)$, Algorithmica **1** (1986), 1—15.

12. H. Davenport and H. Hasse, Die Nullstellen der Kongruenzzetafunktionen in gewissen zyklischen Fällen, J. Reine Angew. Math. **172** (1934), 151—182.

13. M. Deuring, Die Typen der Multiplikatorenringe elliptischer Funktionenkörper, Abh. Math. Sem Hansischen Univ. **14** (1941), 197—272.

14. S. Goldwasser and J. Kilian, Almost all primes can be quickly certified (Proc. 18th Annual ACM Sympos. on Theory of Computing (STOC, Berkeley, 1986)), The Association for Computing Machinery, New York, 1986, pp. 316—329.

15. K. Ireland and M. Rosen, A classical introduction to modern number theory, Graduate Texts in Math., vol. 84, Springer-Verlag, New York, 1982.

16. N. M. Katz and B. Mazur, Arithmetic moduli of elliptic curves, Princeton Univ. Press, Princeton, N. J., 1985.

17. D. E. Knuth, The art of computer programming, vol. 2: Seminumerical algorithms, 2nd ed., Addison-Wesley, Reading, Mass., 1981. [Имеется перевод 1-го изд.: Кнут Д. Искусство программирования для ЭВМ. Т. 2. Получисленные алгоритмы. — М.: Мир, 1977.]

18. S. Lang and H. Trotter, Frobenius distrbutions in Gl_2-extensions, Lecture Notes in Math., vol. 504, Springer-Verlag, Berlin, 1976.

19. H. Lange and W. Ruppert, Complete systems of addition laws on abelian varieties, Invent. Math. **79** (1985), 603—610.

20. H. W. Lenstra, Jr., Factoring integers with elliptic curves, Ann. of Math. (to appear).

21. H. W. Lenstra, Jr., and R. Tijdeman (Editors), Computational methods in number theory, Math. Centre Tracts, no. 154/155, Mathematisch Centrum, Amsterdam, 1982.

22. V. S. Miller, Short programs for functions on curves, IBM Thomas J. Watson Res. Center, Yorktown Heights, N. Y., 1986.

23. P. L. Montgomery, Speeding the Pollard and elliptic curve methods ef factorization, Math. Comp. **48** (1987), 243—264.

24. H. C. Pocklington, The determination of the prime and composite nature of large numbers by Fermat's theorem, Proc. Cambridge Philos. Soc. **18** (1914—16), 29—30.

25. J. M. Pollard, Theorems on factorization and primality testing, Proc. Cambridge Philos. Soc. **76** (1974), 521—528.

26. C. Pomerance, Analysis and comparison of some integer factoring algorithms, in **21**, pp. 89—139.

27. H. Riesel, Prime numbers and computer methods for factorization, Progr. Math., vol. 57, Birkhaüser, Boston, 1985.

28. R. L. Rivest, A. Shamir, and L. Adleman, A method for obtaining digital signatures and public-key cryptosystems, Comm. ACM **21** (1978), 120—126.

29. C. P. Schnorr and H. W. Lenstra, Jr., A Monte Carlo factoring algorithm with linear storage, Math. Comp. **43** (1984), 289—311.

30. R. J. Schoof, Elliptic curves over finite fields and the computation of square roots mod p, Math. Comp. **44** (1985), 483—494.

31. R. J. Schoof, Nonsingular plane cubic curves over finite fields, J. Combin. Theory Ser. B (to appear).

32. J. L. Selfridge and M. C. Wunderlich, An efficient algorithm for testing large numbers for primality, Proc. Fourth Manitoba Conf. Numerical Math., University of Manitoba, Congressus Numerantium XII, Utilitas Math., Winnipeg, 1975, pp. 109—120.

33. D. Shanks, Class number, a theory of factorization, and genera, Proc. Sympos. Pure Math., vol. 20, Amer. Math. Soc., Providence, R. I., 1971, pp. 415—440.

34. —, Five number-theoretic algorithms, Proc. Second Manitoba Conf. Numerical Math., University of Manitoba, Congressus Numerantium VII, Utilitas Math., Winnipeg, 1973, pp. 51—70.

35. J. H. Silverman, The arithmetic of elliptic curves, Graduate Texts in Math., vol. 106, Springer-Verlag, New York, 1986.

36. W. C. Waterhouse, Abelian varieties over finite fields, Ann. Sci. École Norm. Sup. (4) **2** (1969), 521—560.

37. H. Weber, Lehrbuch der Algebra, Bd III, Friedrich Vieweg und Sohn, Braunschweig, 1908; reprinted by Chelsea Publishing Company, New York.

38. H. C. Williams, Primality testing on a computer, Ars Combin. **5** (1978), 127—185.

39. —, A $p+1$ method of factoring, Math. Comp. **39** (1982), 225—234.

МАТЕМАТИЧЕСКИЙ ИНСТИТУТ, АМСТЕРДАМСКИЙ УНИВЕРСИТЕТ, 1018 WB АМСТЕРДАМ, НИДЕРЛАНДЫ

О НЕДАВНИХ ДОСТИЖЕНИЯХ В ТЕОРИИ УРАВНЕНИЙ С ЧАСТНЫМИ ПРОИЗВОДНЫМИ, ВОЗНИКАЮЩИХ В ДИФФЕРЕНЦИАЛЬНОЙ ГЕОМЕТРИИ

Ричард Шён[1]

Уравнения, возникающие в дифференциальной геометрии, приводят к постановке многих естественных проблем в общей теории уравнений с частными производными. Примерами того служат уравнения Эйнштейна для метрического тензора, уравнение минимальных поверхностей, уравнения Янга — Миллса и уравнение гармонических отображений. В последние годы был достигнут существенный прогресс в решении важных геометрических проблем, обусловленный использованием глубоких аналитических методов. Но если анализ приводит к решению геометрических проблем, то в свою очередь уравнения, возникающие в геометрии, часто служат источником для развития новых методов анализа. Например, уравнение Монжа — Ампера послужило толчком для развития теории сильно нелинейных уравнений.

В этом докладе мы обсудим две геометрические проблемы. Первая из них — это уравнение Ямабэ, конформно-инвариантное скалярное уравнение, которое описывает конформные деформации римановых метрик. Вторая — система гармонических отображений, возникающая при нахождении экстремумов энергии отображений в некоторое заданное многообразие.

Общая характерная черта, присущая многим нелинейным процессам, изучаемым в разных науках, — наличие сингулярностей в их поведении. В геометрических задачах сингулярности решений очень важны и содержат полезную информацию. В случае уравнения Ямабэ сингулярности решений могут пониматься как точки, в которых обращается в бесконечность тензор полной римановой метрики постоянной скалярной кривизны на подобласти n-мерной сферы. В проблеме гармонических отображений сингулярное множество — это множество, вдоль которого отображение обрывается или разрывно. Определение сингулярностей решений нелинейных уравнений является важной и интригующей проблемой. Более детально мы обсудим это в п. 3.

---

[1] Richard Schoen, Recent progress in geometric partial differential equations, **ICM86, pp. 121—130.**

Решение проблемы Ямабэ (о которой мы поговорим более подробно ниже) — первый шаг в общей программе построения канонических римановых метрик на гладком многообразии. На самом деле первоначальная цель Ямабэ состояла в доказательстве трёхмерной гипотезы Пуанкаре. Если M — гладкое компактное многообразие, то эйнштейнова метрика g на M — это метрика с кривизной Риччи, пропорциональной g, т. е. метрика, удовлетворяющая условию $\mathrm{Ric}(g)=cg$ при некоторой константе c. Если M трёхмерно, то эйнштейнова метрика с необходимостью имеет постоянную кривизну. В частности, из существования эйнштейновой метрики на односвязном трёхмерном многообразии вытекает гипотеза Пуанкаре.

Один из наиболее эффективных подходов к решению нелинейных уравнений — это вариационный метод. Оказывается, уравнение Эйнштейна служит уравнением Эйлера — Лагранжа для естественного вариационного интеграла — действия Эйнштейна — Гильберта. Если мы договоримся обозначать через $\mathfrak{m}_1$ пространство римановых метрик g на M, удовлетворяющих условию на объём $\int\limits_M dv_g=1$, то действие Эйнштейна — Гильберта представляет собой функционал $\mathcal{R}: m_1 \to \mathbb{R}$, задаваемый формулой

$$\mathcal{R}(g)=\int\limits_M R_g\,dv_g,$$

где R_g — функция скалярной кривизны для g. Разумеется, функционал $\mathcal{R}(\cdot)$ инвариантен относительно действия группы диффеоморфизмов $\mathrm{Diff}(M)$, и, следовательно, он определён на $\mathfrak{m}_1/\mathrm{Diff}(M)$. Функционал $\mathcal{R}$ не ограничен ни сверху, ни снизу, поэтому бессмысленно искать его максимум или минимум. Анализ линеаризации этого функционала вблизи критической метрики g показывает, что если пренебречь конечномерными пространствами деформаций, то функционал $\mathcal{R}$ достигает минимума на множестве метрик, принадлежащих конформному классу метрики g (классу метрик $\bar{g}$ вида $\bar{g}=e^v g$, где $v\in C^\infty(M)$), и максимума на множестве направлений, ортогональных к конформному классу метрики g. Это наводит на мысль, что эйнштейновские метрики можно строить при помощи следующей двухэтапной процедуры. Прежде всего, зафиксируем метрику $g_0\in\mathfrak{m}_1$ и рассмотрим задачу минимизации

$$I(M;\ g_0)=\inf_{g\in[g_0]}\mathcal{R}(g),$$

где $[g_0]$ обозначает конформный класс метрики g_0. Затем возьмём максимум по g_0:

$$I(M) = \sup_{g_0 \in \mathfrak{m}_1} I(M;\ g_0).$$

Первый шаг указанной процедуры — это задача Ямабэ, т. е. задача о существовании гладкой метрики $g \in [g_0]$ с действием $\mathscr{R}(g) = I(M; g_0)$. Пожалуй, вполне естественно, что решение задачи Ямабэ (описываемое ниже) в значительной степени опирается на идеи общей теории относительности, в которой возникло и само действие $\mathscr{R}(\cdot)$. Другое следствие, получаемое при решении проблемы Ямабэ, состоит в том, что единственный максимум функционала $I(M; g_0)$ на множестве всех n-мерных многообразий M достигается на стандартной сфере S^n. В частности, $I(M) \leqslant I(S^n)$ для любого n-мерного многообразия M. Можно также отметить, что $I(M)$ положительно тогда и только тогда, когда M допускает метрику с положительной скалярной кривизной. Таким образом, из результата Хитчина [17] вытекает, что существуют экзотические сферы M, для которых $I(M) \leqslant 0$. Это показывает, что $I(\cdot)$ зависит от гладкой структуры на M.

Второй шаг заключается в том, чтобы реализовать $I(M)$ при помощи гладкой метрики. Для трёхмерного многообразия M представляется правдоподобным, что максимизирующая последовательность метрик должна сходиться к метрике, сингулярной вдоль двумерной поверхности. Это возвращает нас к вопросу о сингулярностях решений. Понимание возникающих в данном случае сингулярностей может послужить ключом к расшифровке тайн многообразий.

1. СКАЛЯРНАЯ КРИВИЗНА И КОНФОРМНО-ИНВАРИАНТНЫЕ СКАЛЯРНЫЕ УРАВНЕНИЯ

Мы будем рассматривать геометрические проблемы, связанные со следующим модельным уравнением:

$$\Delta u + u^p = 0, \tag{1.1}$$

где u — положительная функция, определённая на некотором открытом подмножестве в $\mathbb{R}^n$ при $n \geqslant 3$. Будем предполагать, что p больше единицы и не превосходит $p_0 = (n+2)/(n-2)$; на самом деле для геометрии прежде всего интересен случай $p = p_0$. Мы обсудим проблему существования как классического решения на компактном многообразии, так и слабого решения $u \in L^p_{\text{loc}}$, определённого на $\mathbb{R}^n$ (или S^n).

Уравнения типа (1.1) возникают также во многих физических вопросах.

Для случая гладкой ограниченной области[1] Ω многие авторы изучали краевую задачу с граничным условием $u \equiv 0$ на $\partial\Omega$. Как показал Похожаев [22], если область Ω является звёздной и $p = (n+2)/(n-2)$, то такая задача не имеет положительных решений. Брезис и Ниренберг [8] установили, что во многих случаях модифицированное уравнение $\Delta u + \lambda u + u^{(n+2)/(n-2)} = 0$ при $0 < \lambda < \lambda_1(\Omega)$ имеет положительное решение. Недавно Бахри и Корон [5] доказали существование решения в случае нестягиваемых областей Ω. Их доказательство использует тонкие вариационные соображения. Геометрически можно интерпретировать u как метрику $\bar{g}_{ij} = u^{4/(n-2)}\delta_{ij}$. При этом уравнение (1.1) означает, что скалярная кривизна метрики $\bar{g}$ тождественно равна $4(n-1)/(n-2)$. Из этой интерпретации легко вывести конформную инвариантность данного уравнения: поскольку $\gamma^*\bar{g}$ также должна иметь постоянную скалярную кривизну, то мы получаем, что если u удовлетворяет (1.1) (с $p = p_0$), то $u_\gamma = |\gamma'|^{(n-2)/2} u \circ \gamma$ тоже является решением, причем $|\gamma'|$ — коэффициент линейного растяжения конформного отображения γ.

В начале 1960-х годов Ямабэ [35] рассмотрел основные вопросы о существовании решений данного уравнения. Задача Ямабэ заключается в построении римановой метрики постоянной скалярной кривизны, конформно-эквивалентной произвольной заданной метрике g на компактном многообразии M^n размерности $n \geqslant 3$. Итак, задача состоит в построении положительного решения u следующего уравнения на M:

$$Lu + Ku^{(n+2)/(n-2)} = 0, \quad Lu = \Delta_g u - \frac{n-2}{4(n-1)} R_g u, \tag{1.2}$$

где K — константа, которую можно нормировать к 1, 0 или —1. Оператор L, входящий в (1.2), принято называть конформным лапласианом; он обладает естественной инвариантностью относительно конформных замен метрики. Дифференциальное уравнение (1.2) возникает как уравнение Эйлера — Лагранжа для некоторой вариационной задачи, а именно задачи о нахождении условных экстремумов действия Эйнштейна — Гильберта

$$\mathscr{R}(\bar{g}) = \int_M R_{\bar{g}}\, dv_{\bar{g}}$$

при ограничении на объём $\int_M dv_{\bar{g}} = 1$, где $\bar{g}$ берётся из класса

---

[1] То есть области с гладкой границей. — *Прим. перев.*

метрик, конформно-эквивалентных некоторой заданной римановой метрике g. Ямабэ рассматривал свою статью как первый шаг в направлении построения теории критических точек функционала $\mathscr{R}(\cdot)$ на пространстве римановых метрик на M. Как было сказано выше, $\mathscr{R}$ ведёт себя различно вдоль направлений (в пространстве метрик), конформных заданной метрике, и направлений, ортогональных к ним. Если записать $\bar{g}$ в виде $\bar{g}=u^{4/(n-2)}g$, то описанная вариационная проблема превращается в классическую задачу о нахождении экстремума функционала

$$\mathscr{R}(\bar{g})=\frac{4(n-1)}{n-2}\int\limits_{M}\left(|\nabla u|^2+\frac{n-2}{4(n-1)}R_g u^2\right)dv_g$$

при условии

$$\int\limits_{M}u^{2n/(n-2)}\,dv_g=1.$$

Это приводит к уравнению (1.2) для множителя u. В 1968 г. Традинжер [32] обнаружил, что в исходной работе Ямабэ имеется ошибка, и доказал существование решения задачи (1.2) для случая $R_g\leqslant 0$. (Предположение о постоянстве знака R_g не ограничивает общности.) Следующий значительный шаг вперёд сделал Обэн [3] в 1975 г., построив минимизирующую функцию u для случая, когда $n\geqslant 6$ и g не является локально конформно-плоской. Условия, фигурирующие в теореме Обэна, помогли обнаружить поправочный член в рассматриваемой задаче, возникающий при локальных конформных возмущениях метрики g. Из более раннего результата Похожаева [22] было ясно, что проблема существования не может иметь чисто локального характера, если g — локально конформно-плоская метрика. Возникло ощущение, что недостающие для решения задачи Ямабэ средства могут обнаружиться при глобальном анализе.

С другой стороны, задача изучения римановых многообразий неотрицательной скалярной кривизны возникает в общей теории относительности. Локальная положительность плотности материи приводит, в силу уравнения Эйнштейна, к неравенству для кривизны метрики пространства-времени. Ограничивая его на пространственноподобную гиперповерхность M, мы получаем скалярное неравенство на римановом многообразии M. В случае когда M — максимальная гиперповерхность в пространстве-времени, из этого неравенства вытекает неотрицательность скалярной кривизны. Стандартное граничное условие, налагаемое на пространство-время, состоит в том, что оно асимптотически плоское. Более точно, если гиперповерхность M максимальна, то разумно наложить асимптотические условия

$$g_{ij}=\left(1+\frac{E}{|x|}\right)\delta_{ij}+O\left(|x|^{-2}\right)$$

в координатной карте вблизи бесконечности на M. При этом число E является ADM-энергией [2] и имеет смысл полной энергии гравитационного поля, измеренной в пространственной бесконечности. Из специального варианта теоремы о положительности энергии вытекает, что $E\geqslant 0$, причём $E=0$, лишь если (M,g) изометрично $\mathbb{R}^3$ с эвклидовой метрикой.

Теорема о положительности энергии была доказана докладчиком и Яу [26] в конце семидесятых годов. Наше доказательство было основано на изучении минимальных поверхностей в M. Несколько лет спустя Виттен [34] нашёл интересное доказательство теоремы о положительности энергии, использующее оператор Дирака. Недавно мы применили некую модификацию нашего метода минимальных поверхностей для доказательства одной теоремы сингулярности в общей теории относительности [27]. Это — утверждение о том, что высокая концентрация материи с необходимостью приводит к особенностям пространства-времени. Существует естественное n-мерное обобщение упомянутой выше теоремы на случай, когда задаётся асимптотическое условие

$$g_{ij}=\left(1+\frac{E}{|x|^{n-2}}\right)\delta_{ij}+O\left(|x|^{1-n}\right).$$

Более общим образом, как показали Бартник [6] и Ли и Паркер [18], энергия корректно определена, если $g_{ij}=\delta_{ij}+o(|x|^{-(n-2)/2})$. При этом n-мерная теорема о положительности энергии утверждает, что $E\geqslant 0$, если $R_g\geqslant 0$ на M и g — асимптотически плоская метрика. Кроме того, $E=0$, лишь если (M,g) изометрично плоскому пространству $\mathbb{R}^n$. Эту теорему можно доказать, применяя методы, аналогичные использованным в трёхмерном случае.

Теперь обратимся к проблеме Ямабэ в случае компактного многообразия M с положительной скалярной кривизной. Поскольку $R_g>0$, то оператор L отрицательно-определён. Следовательно, для любой точки $x_0\in M$ существует положительное фундаментальное решение G_{x_0} оператора L с полюсом в x_0. Если мы возьмём $\bar{g}=G_{x_0}^{4/(n-2)}g$ — метрику, определяемую по G_{x_0}, то увидим, что $(M\setminus\{x_0\},\bar{g})$ является асимптотически плоским многообразием с $R_{\bar{g}}=0$. В случае $n<6$ или (при любом n) в случае, когда метрика g — локально конформно-плоская вблизи x_0, асимптотическое поведение $\bar{g}$ достаточно просто, так что энергия $E(x_0)$ корректно определена. Тогда по теореме о положительности энергии $E(x_0) \geqslant 0$, причём $E(x_0) = 0$, лишь если $(M \setminus \{x_0\}, \bar{g})$ изометрично эвклидову пространству. Этот второй вариант возможен, только если (M, g) конформно-эквивалентно стандартной сфере S^n. Докладчик показал в [24], что положительность $E(x_0)$ для некоторой точки x_0 обеспечивает наличие глобального поправочного члена, который удаётся использовать для построения минимизирующей функции u в задаче Ямабэ. Это позволяет охватить все возможные ситуации и, следовательно, даёт полное решение задачи Ямабэ.

Анализ конформного лапласиана L, и в частности его фундаментального решения, позволил получить информацию о целом классе локально конформно-плоских многообразий, в том числе многообразий с неотрицательной скалярной кривизной. Кратко опишем эти результаты, полученные докладчиком совместно с Яу [28]. Простейшим конформно-плоским многообразием является область $\Omega \subset S^n$ на n-сфере. Имеется также целый класс конформно-плоских многообразий, которые являются факторпространствами Ω по некоторой дискретной подгруппе Γ конформной группы $\mathcal{G}_n = O(n+1, 1)$ сферы S^n. Если Γ сохраняет Ω, действует собственно разрывно и не имеет неподвижных точек на Ω, то $M = \Omega/\Gamma$ — конформно-плоское многообразие. Мы будем называть такие дискретные группы Γ клейновыми группами.

Общая конформно-плоская структура на многообразии M задаётся погружением $\Phi: \tilde{M} \to S^n$, где $\tilde{M}$ — универсальное накрытие M, совместно с гомоморфизмом $\rho: \pi_1 M \to \mathcal{G}_n$, удовлетворяющим условию эквивариантности $\Phi \circ \gamma = \rho(\gamma) \circ \Phi$ при $\gamma \in \pi_1 M$. Отображение Φ называется развёртывающим, а гомоморфизм ρ — голономным представлением. Если отображение Φ инъективно, то, полагая $\Omega = \Phi(\tilde{M})$ и $\Gamma = \rho(\pi_1 M)$, мы получим, что группа Γ дискретна, а $M = \Omega/\Gamma$, как и выше. Для простоты предположим, что M компактно и локально конформно-плоское, и определим число $d(M)$ следующим образом. Выберем допустимую метрику g на M, возьмём $x_0 \in \tilde{M}$, и пусть G_{x_0} — минимальная функция Грина для L с полюсом в x_0. Положим

$$d(M) = \inf\left\{\frac{n-2}{2} p : G_{x_0} \in L^p(\tilde{M} \setminus B_1(x_0))\right\}.$$

(Отметим, что, как можно показать, G_{x_0} всегда существует.) Можно убедиться, что $d(M)$ лежит в интервале $[0, n]$ и не зависит от выбора метрики g, а следовательно, зависит лишь от конформно-плоской структуры. Существование допустимой метрики g на M с $R_g > 0$ эквивалентно неравенству $d(M) < (n-2)/2$. Основная структурная теорема утверждает, что если $d(M) < (n-2)^2/n$, то Φ инъективно и, следовательно, $M = \Omega/\Gamma$ для некоторой клейновой группы Γ. Кроме того, в случае клейновой группы число $d(M)$ совпадает с хаусдорфовой размерностью предельного множества $\Lambda = S^n \setminus \Omega$ группы Γ.

В частности, мы получаем, что локально конформно-плоские многообразия M с положительной скалярной кривизной — это в точности многообразия вида Ω/Γ, где предельное множество клейновой группы Γ имеет хаусдорфову размерность, меньшую чем $(n-2)/2$. Доказательство сформулированной выше основной теоремы включает в себя проверку инъективности накрывающего отображения Φ. Это делается при помощи (некоторой модификации) метода Бохнера. В качестве простого следствия из структурной теоремы получается прямое доказательство теоремы о положительности энергии для локально конформно-плоских многообразий размерности $n \geqslant 4$.

Вернёмся, наконец, обратно к модельным уравнениям (1.1) и (1.2). Функция $u \in L^{(n+2)/(n-2)}(S^n)$ является (глобальным) слабым решением уравнения (1.2), если для любой функции $\varphi \in C^\infty(S^n)$ справедливо тождество

$$\int_{S^n} [(L\varphi) u + \varphi u^{(n+2)/(n-2)}] \, dv_0 = 0.$$

Аналогично выглядит определение слабого решения класса L^p_{loc} уравнения (1.1). Слабые решения этих уравнений рассматривали многие авторы, в частности Серрин, Гидаш, Спракк, Авилес и Каффарелли (см. [29, 13]). Эти авторы получили полное описание асимптотического поведения слабого решения вблизи изолированной особенности.

Теория слабых решений уравнения (1.2) связана с геометрией посредством следующей теоремы. Предположим, что $\Omega \subset S^n$ и u — классическое решение уравнения (1.2) на Ω, для которого риманова метрика $u^{4/(n-2)} g_0$ геодезически полна на Ω. Тогда дополнение к Ω имеет хаусдорфову размерность, не превосходящую $(n-2)/2$, и функция u принадлежит $L^{(n+2)/(n-2)}(S^n)$ и является глобальным слабым решением уравнения (1.2). Разумеется, сингулярное множество функции u в точности совпадает с $S^n \setminus \Omega$.

В качестве следствия из этого результата и решения задачи Ямабэ на компактном локально конформно-плоском многообразии мы получаем конструкцию для построения большого семейства новых глобальных слабых решений уравнения (1.2). Они задаются полными метриками на универсальной накрывающей области Ω рассматриваемого многообразия. Эти слабые решения, как правило, имеют сингулярные множества с дробной хаусдорфовой размерностью, например канторовы множества или квазисферы. Мы думаем, что любое положительное глобальное решение уравнения (1.2) возникает из полной метрики на подобласти в S^n:

Гипотеза. *Если функция* $u \in L^{(n+2)/(n-2)}(S^n)$ *служит слабым решением уравнения* (1.2), *то она регулярна в области* $\Omega \subset S^n$ *и метрика* $u^{4/(n-2)}g_0$ *геодезически полна на* Ω.

Интригующий вопрос теории слабых решений — проблема существования. С точки зрения геометрии эта проблема эквивалентна построению полных конформных метрик постоянной положительной скалярной кривизны на заданной области $\Omega \subset S^n$. Это дало бы глобальные слабые решения уравнения (1.2), сингулярные в точности на $S^n \setminus \Omega$. При этом необходимое условие на Ω — чтобы хаусдорфова размерность дополнения $S^n \setminus \Omega$ была меньше, чем $(n-2)/2$.

Для полных метрик постоянной отрицательной скалярной кривизны проблема существования была решена в разумной общности в работе Лёвнера и Ниренберга [19]. Этот случай отличается тем, что решения на Ω не продолжаются до слабых решений соответствующих уравнений. Некоторая теорема единственности гладкого глобального решения уравнения (1.2) была получена в работах Обаты [21] и Гидаша, Ни и Ниренберга [12]. Обобщая эти методы, Гидаш и Спракк [13] дали доказательство несуществования глобального слабого решения задачи (1.2) с ровно одной особой точкой на S^n. Недавно докладчик построил глобальные слабые решения задачи (1.2), сингулярные точно на заданном конечном множестве точек S^n, содержащем по крайней мере две точки. Вообще интересно выяснить вопрос, имеет ли особое множество слабого решения какую-то специальную локальную геометрическую структуру.

2. ПРОБЛЕМА ГАРМОНИЧЕСКИХ ОТОБРАЖЕНИЙ

Очень кратко остановимся на аспектах проблемы гармонических отображений, связанных со слабыми решениями и вопросами регулярности решений. Пусть M^n и N^k — римановы многообразия, и пусть N изометрически вложено в $\mathbb{R}^K$. Пусть, далее, $\Psi(u) = (\psi_1(u), \ldots, \psi_{K-k}(u))$ — неособая вектор-функция, определяющая многообразие N, т. е. $N = \{u \in \mathbb{R}^K : \Psi(u) = 0\}$. Тогда отображения u из M в N задаются $\mathbb{R}^K$-значными вектор-функциями, удовлетворяющими тождеству $\Psi(u(x)) = 0$ при $x \in M$. Рассмотрим интеграл Дирихле

$$E(u) = \sum_{i=1}^{K} \int_M |\nabla u_i|^2 \, dv$$

для отображения $u : M \to N$. Критические точки этого функционала являются гармоническими отображениями и удовлетворяют системе уравнений Эйлера — Лагранжа

$$\Delta u^i + A^i_{u(x)}(\nabla u(x), \nabla u(x)) = 0, \quad i = 1, \ldots, K, \qquad (2.1)$$

где A — векторнозначная вторая фундаментальная форма подмногообразия N в $\mathbb{R}^K$. Проблема гармонических отображений интенсивно изучалась в геометрии. Она также постоянно затрагивается в литературе по математической физике. Недавно она обсуждалась в качестве упрощённой модели жидких кристаллов (см. [15]). Для краткости отсылаем читателя к обзорным статьям [9, 23], содержащим всестороннее обсуждение проблемы, а здесь приведём лишь сводку результатов о гладкости решений.

Аналитическая теория существования решений была построена в работах Морри [20], Иллза — Сэмпсона [10], Хильдебрандта — Кауля — Видмана [16], Джаквинты — Джусти [11], Шёна — Уленбекк [25] и Хардта — Киндерлерера — Лина [15]. Для слабых решений уравнения (2.1) никакой теоремы регулярности до сих пор неизвестно. Для отображений u, являющихся стационарными точками вариационной задачи (см. [23]), регулярность при $n = 2$ была доказана в работе Грютера [14] и докладчика [23]. Есть предположение, что при $n \geqslant 3$ для стационарных точек u имеет место частичная регулярность.

Случай точек минимума изучен гораздо лучше. Кратко опишем один из результатов работы Шёна — Уленбекк [25], который представляет собой единственную теорему о регулярности для гармонических отображений ($n \geqslant 3$), не содержащую специальных предположений о многообразии N. Прототипом сингулярностей служат сингулярности, возникающие из непостоянных гармонических отображений $u_0 : S^k \to N$ при $k \geqslant 2$. Полагая $u(x) = u_0(x/|x|)$, мы получаем однородное гармоническое отображение u из $\mathbb{R}^{k+1}$ в N. Если u локально минимизирует энергию в $\mathbb{R}^{k+1}$, то будем называть u минимизирующим касательным отображением. Для общего минимизирующего отображения u обозначим через $\mathcal{S}(u)$ его сингулярное множество. Теорема регулярности утверждает, что в общем случае $\mathcal{S}(u)$ имеет хаусдорфову размерность, не превосходящую $n-3$. Если никаких нетривиальных минимизирующих касательных отображений (МКО) в N при $k = 2, \ldots, l$ не существует, то $\dim S(u) \leqslant n - l - 1$. Кроме того, по этой теореме для любой точки $p_0 \in S(u)$ мы получаем нетривиальное МКО (возможно, сингулярное вне начала координат) как предел растянутых отображений $u_{\lambda_i}(x)=u(\lambda_i x)$ в нормальной системе координат x с центром в p_0 при $\lambda_i \to 0$. Этим объясняется употребление термина «касательное отображение».

3. НЕКОТОРЫЕ ЗАМЕЧАНИЯ ОБ ОСОБЕННОСТЯХ

В дифференциальной геометрии возникло несколько естественных нелинейных уравнений в частных производных, для которых неотъемлемой, характерной чертой является сингулярное поведение их решений. Помимо двух указанных выше примеров упомянем также о задаче Плато и гиперболическом уравнении Эйнштейна. В ряде случаев были высказаны гипотезы о природе таких особенностей, например гипотеза о космической цензуре Р. Пенроуза для уравнения Эйнштейна.

Имеется целый набор вопросов или гипотез относительно устройства сингулярностей минимальных поверхностей и решений задачи о гармонических отображениях. Я остановлюсь здесь на соображениях по поводу гармонических отображений, хотя и по поводу минимальных поверхностей можно было бы высказать много аналогичных соображений. Вопрос, который я собираюсь обсудить, был впервые сформулирован более двадцати лет назад в контексте минимальных поверхностей. Я рассмотрю этот вопрос применительно к гармоническим отображениям потому, что в этом случае теория слабых решений устроена проще. Вопрос об особенностях обсуждался в трёх основных аспектах: (1) структура сингулярного множества, (2) устойчивость сингулярностей и (3) решения с заданными особенностями.

Что касается вопроса (1), то основная гипотеза в случае гармонических отображений состоит в том, что $\mathscr{S}(u)$ есть объединение спрямляемых стратов целой размерности. Эта гипотеза представляется очень правдоподобной при условии, что M и N являются вещественными аналитическими римановыми многообразиями. Однако она ещё очень далека от своего обоснования, и среди экспертов имеются большие разногласия относительно её соответствия истине. Единственный пример, в котором была найдена структура сингулярного множества,— это пример, описанный в работе Тейлора [31] о двумерных множествах минимальной площади в трёхмерном пространстве. В этом случае сингулярное множество является одномерным комплексом. Главная трудность при исследовании свойств множества $\mathscr{S}(u)$ — это возможная неединственность касательного отображения в сингулярной точке $p_0 \in S$. Грубо говоря, при отсутствии единственности это отображение ведёт себя совершенно по-разному в разных масштабах и не дает никакой предельной конфигурации при бесконечном увеличении.

Недавно прорыв в этой области был осуществлен Л. Саймоном [30], доказавшим единственность касательного отображения (и касательного конуса для минимальных поверхностей) в случае изолированной особой точки при условии, что N — вещественно-аналитическое многообразие. Единственность при некоторых дополнительных предположениях была доказана Аллардом и Альмгреном [1]. Для неизолированных особенностей в вопросе о единственности до сих пор ничего неизвестно. Вопрос (2) состоит в том, меняется ли структура сингулярности коренным образом при изменении граничных условий или метрики в данной задаче. Бывают ситуации, когда было бы полезно знать, что при малом возмущении задачи сингулярность полностью исчезает. В этом направлении некоторый прогресс был достигнут Бёме и Тромба [7] и Уайтом [33] для случая двумерных минимальных многообразий. Однако остается множество открытых вопросов. В задаче о гармонических отображениях представляется неразумным пытаться задавать одновременно граничные значения и сингулярное множество, как мы это делали для скалярного уравнения в первой части доклада. Можно попытаться задать сингулярное множество $\mathscr{S}$ и рассмотреть вопрос о локальном существовании слабых решений, сингулярных на $\mathscr{S}$. Примеров сингулярностей известно сравнительно мало, и было бы полезно иметь лучшее представление о том, что может происходить локально. Некоторые результаты в этом направлении были получены Каффарелли, Хардтом и Л. Саймоном.

Если мы рассмотрим те же три вопроса применительно к слабым решениям конформно-инвариантного скалярного уравнения из первой части доклада, то увидим, что в этом случае свойства сингулярностей существенно отличаются от их ожидаемых свойств в случае гармонических отображений. Мы уже указали, что $\mathscr{S}(u)$ может быть множеством дробной хаусдорфовой размерности даже для глобального решения. Хотя это вполне естественно с точки зрения геометрии, но для столь простого аналитического эллиптического уравнения такое поведение его решений представляется весьма странным.

Частичное техническое объяснение указанного различия заключается в отсутствии *неравенства монотонности* для упомянутого скалярного уравнения. Теория регулярности гармонических отображений, а также результаты Л. Саймона о единственности касательного отображения существенным образом опираются на монотонность масштабно-инвариантной энергии для гармонического отображения. Грубо говоря, это означает, что в среднем поведение отображения обнаруживает тенденцию к улучшению при большом увеличении. С другой стороны, у нашего скалярного уравнения есть решения, которые инварианты относительно последовательности всё больших и больших увеличений, но не являются однородными. Представляется, что это несовместимо с неравенством монотонности.

В заключение подчеркнём, что имеется множество интересных явлений, связанных с сингулярностями, возникающими в геометрических задачах, и заслуживающих дальнейшего изучения. Несомненно, это — плодотворное направление для будущих исследований.

ЛИТЕРАТУРА

1. W. Allard and F. J. Almgren, On the radial behavior of minimal surfaces and the uniqueness of their tangent cones, Ann. of Math. **113** (1981), 215—265.

2. R. Arnowitt, S. Deser, and C. Misner, Energy and the criteria for radiation in general relativity, Phys. Rev. (2) **118** (1960), 1100.

3. T. Aubin, The scalar curvature, Differential Geometry and Relativity (Cahen and Flato, eds.), Reidel, Dordrecht, 1976.

4. P. Aviles, On isolated singularities in some nonlinear partial differential equations, Indiana Univ. Math. J. **32** (1983), 773—791.

5. A. Bahri and J. M. Coron, Une théorie des points critiques à l'infini pour l'equation de Yamabe et le problême de Kazdan-Warner, C. R. Acad. Sci. Paris Sér. I Math. **15** (1985), 513—516.

6. R. Bartnik, The mass of an asymptotically flat manifold, Preprint.

7. R. Böhme and A. Tromba, The index theorem for classical minimal surfaces, Ann. of Math. **113** (1981), 447—499.

8. H. Brezis and L. Nirenberg, Positive solutions of nonlinear elliptic equations involving critical Sobolev exponents, Comm. Pure Appl. Math. **36** (1983), 437—477.

9. J. Eells and L. Lemaire, A report on harmonic maps, Bull. London Math. Soc. **10** (1978), 1—68.

10. J. Eells and J. Sampson, Harmonic mappings of Riemannian manifolds, Amer. J. Math. **86** (1964), 109—160.

11. M. Giaquinta and E. Ciusti, On the gerularity of the minima of variational integrals, Acta Math. **148** (1982), 31—46.

12. B. Gidas, W. M. Ni, and L. Nirenberg, Symmetry and related properties via the maximum principle, Comm. Math. Phys. **68** (1979), 209—243.

13. B. Gidas and J. Spruck, Global and local behavior of positive solutions of nonlinear elliptic equations, Comm. Pure Appl. Math. **34** (1981), 525—598.

14. M. Grüter, Conformally invariant variational integrals and the removability of isolated singularities, Manuscripta Math. **47** (1984), 85—104.

15. R. Hardt, D. Kinderlehrer, and F. H. Lin, Existence and partial regularity of static liquid crystal configurations, Comm. Math. Phys. (to appear).

16. S. Hildebrandt, H. Kaul, and K. O. Widman, An existence theorem for harmonic mappings of Riemannian manifolds, Acta Math. **138** (1977), 1—16.

17. N. Hitchin, Harmonic spinors, Adv. in Math. **14** (1974), 1—55.

18. J. Lee and T. Parker, Preprint.

19. C. Löwner and L. Nirenberg, Partial differential equations invariant under conformal or projective transformations, Contributions to Analysis: A collection of Papers Dedicated to Lipman Bers, Academic Press, New York, 1974.

20. C. B. Morrey, The problem of Plateau on a Riemannian manifold, Ann. of Math. **49** (1948), 807—851.

21. M. Obata, The conjectures on conformal transformations of Riemannian manifolds, J. Differential Geom. **6** (1971), 247—258.

22. Похожаев С. Собственные функции уравнения $\Delta u+\lambda f(u)=0$. — ДАН СССР, 1965, т. 6, с. 1408—1411.

23. R. Schoen, Analytic aspects of the harmonic map problem, Seminar on Nonlinear Partial Differential Equations, MSRI Publications 2, Springer-Verlag, 1984.

24. —, Conformal deformation of a Riemannian metric to constant scalar curvature, J. Differential Geom. **20** (1984), 479—495.

25. R. Schoen and K. Uhlenbeck, A regularity theory for harmonic maps, J. Differential Geom. **17** (1982), 307—335.

26. R. Schoen and S. T. Yau, On the proof of the positive mass conjecture in General Relativity, Comm. Math. Phys. **65** (1983), 45—76.

27. —, The existence of a black hole due to condensation of matter, Comm. Math. Phys. **90** (1983), 575—579.

28. —, Conformally flat manifolds, scalar curvature, and Kleinian groups, Preprint.

29. J. Serrin, Removable singularities of solutions of elliptic equations. II, Arch. Rational Mech. Anal. **20** (1965), 163—169.

30. L. Simon, Asymptotics for a class of non-linear evolution equations, with applications to geometric problems, Ann. of Math. **118** (1983), 525—571.

31. J. Taylor, The structure of singularities in soap-bubble-like and soap-film-like minimal surfaces, Ann. of Math. **103** (1976), 489—539.

32. N. Trudinger, Remarks concerning the conformal deformation of Riemannian structures on compact manifolds, Ann. Sculoa Norm. Sup. Pisa Cl. Sci. (3) **22** (1968), 265—274.

33. B. White, Generic regularity of unoriented two-dimensional area minimizing surfaces, Ann. of Math. **121** (1985), 595—603.

34. E. Witten, A new proof of the positive energy theorem, Comm. Math. Phys. **80** (1981), 381—402.

35. H. Yamabe, On a deformation of Riemannian structures on compact manifolds, Osaka J. Math. **12** (1960), 21—37.

КАЛИФОРНИЙСКИЙ УНИВЕРСИТЕТ, САН-ДИЕГО, ЛА-ДЖОЛЛА, КАЛИФОРНИЯ 92093, США

РЕШЕНИЕ УРАВНЕНИЙ С ТОЧКИ ЗРЕНИЯ ВЫЧИСЛИТЕЛЬНОЙ СЛОЖНОСТИ

Арнольд Шёнхаге[1]

1

1.1. Введение. Вычислительная сложность — новая идея в математике, уходящая корнями в алгоритмические и конструктивные традиции нашей науки. Помимо теоретической информатики, где понятие вычислительной сложности играет ключевую роль, это понятие проникло и во многие другие области математики. Сложностные соображения тесно связаны с математической логикой и основаниями математики, с численными методами и их разнообразными приложениями, да и в других разделах математики — геометрии, теории чисел и алгебре — проявляют к ним всё возрастающий интерес.

Чтобы не впадать в излишнюю общность, ограничим наш очень обширный предмет вычислительной трактовкой *алгебраических* уравнений. Даже в этом узком смысле проблема «решения уравнений» исследовалась веками и сейчас является центральной в численных методах и компьютерной алгебре. Основной акцент в данном обзоре делается на вычислительной *сложности* таких проблем. После начального периода отдельных редких публикаций соответствующие исследования непрерывно продолжаются вот уже около двадцати лет. Поэтому, пожалуй, пора рассказать о некоторых результатах этих исследований.

Ввиду ограниченного времени мы обсудим только *последовательные* алгоритмы. Обзор важных сложностных результатов для моделей *параллельных* вычислений был недавно дан Куком [9]. Далее, в соответствии с личными вкусами и склонностями мы ограничимся *детерминированными вычислениями*, не упуская, однако, из виду то обстоятельство, что с практической точки зрения *вероятностные* алгоритмы могут оказаться

более предпочтительными для решения некоторых трудных задач. Более того, всегда имеется естественное взаимодействие между процессами решения уравнений и недетерминированного поиска решений. Соответственно иногда поучительно сравнить сложность *нахождения* решения уравнения и сложность *проверки* того, что данная величина действительно является решением.

Проблему решения алгебраических уравнений можно понимать по-разному. В простых полях $GF(p)$ и $\mathbf{Q}$, а также в их конечных расширениях применимы дискретные методы символьных вычислений. В этой ситуации основную задачу можно определить как построение соответствующих полей разложения. Укажем две работы: [23] и [6], в которых рассматривается сложность этой общей задачи. Однако с точки зрения приложений верхние оценки, полученные в этих работах, безнадёжно велики. Наверное, и преждевременно рассчитывать на полные ответы, когда мы не в состоянии получить хорошие оценки сложности решения для систем *линейных* уравнений (см. разд. 4). Значительный прогресс был достигнут в вопросе о разложении многочленов и в смежных вопросах. В своей пионерской работе [26] А. Ленстра, Х. Ленстра и Ловас показали, как за полиномиальное время разложить на множители многочлен над $\mathbf{Q}$ от одной переменной. С тех пор были опубликованы многочисленные модификации и обобщения этого результата и лежащего в его основе алгоритма редукции базиса (см. [5, 13, 21, 24, 25, 39] и обширную библиографию в [15]). В частности, за полиномиальное время проверяется *разрешимость* в самом классическом смысле теории Галуа (ср. [22]).

Но в случае уравнений над $\mathbf{R}$ и $\mathbf{C}$ нам приходится принять другую точку зрения. Наши сложностные результаты об основной теореме алгебры, представленные в разд. 3, будут относиться к вычислениям с динамической точностью в смысле рекурсивной зависимости, вполне в духе провидческой работы Г. Вейля [53] (1924 г.). Более подробно мы скажем о соответствующей модели вычислений в п. 1.2.

За последние годы вопрос о сложности нахождения корней и смежные вопросы независимо исследовались С. Смейлом и другими авторами, но совсем в другой постановке — так что их результаты несопоставимы с нашими. С учетом программы этого конгресса, по-видимому, будет достаточно упомянуть работу [44] и обзор [45], в которых содержатся дальнейшие ссылки на литературу.

В последующем изложении мы коснемся весьма важных проблем, связанных с решением уравнений. Материал разд. 2 о сложности умножения и деления целых чисел и многочленов

---

[1] Arnold Schönhage, Equations solving in terms of computational complexity, ICM86, pp. 131—153.

можно рассматривать как сложностную трактовку проблемы решения простейшего уравнения $ax = b$ над соответствующими областями. В разд. 3 мы обсуждаем основную теорему алгебры в терминах вычислительной сложности и, в частности, описываем одно изящное решение этой давно стоящей численной задачи. В заключительном разделе речь идет о системах линейных уравнений и сложности вычислений с матрицами, решения характеристических уравнений и вычисления собственных значений.

1.2. Модели вычислений. Вычислимость определяют многими эквивалентными способами. Один из них, основанный на классическом понятии машины Тьюринга, особенно удобен для введения интуитивно подходящих мер для количественного анализа алгоритмов. Если нас интересуют лишь сложностные классы, инвариантные относительно сводимостей с полиномиальной сложностью, то более детальная структура модели, на базе которой мы строим теорию, несущественна. Однако в случае сложности «нижнего уровня» следует проводить более тонкое различие. Многие конкретные алгоритмы, приведённые в литературе, возможно адекватно реализовать на *многоленточных* машинах Тьюринга. Для большей гибкости можно использовать машины с *произвольным доступом к памяти*, имеющие различные системы команд. Такие машины хорошо моделируют адресуемость памяти в современных компьютерах. В некотором смысле «истинную», «каноническую» основу для измерения временнóй сложности доставляет теоретическая модель, называемая *машиной с указателями* (ранее такие модели называли также «автоматами со связями» или «машинами с модифицируемой памятью», ср. [35]). Эти машины эквивалентны в смысле сводимости в реальное время очень простым машинам с произвольным доступом к памяти при единичном весе операций — машинам, выполняющим лишь косвенную адресацию, проверку на равенство и вычисление функции следования.

С точки зрения практической реализации было бы более реалистично рассматривать машины с произвольным доступом к памяти, имеющие *логарифмическую стоимость операций*, при которой обработка любого целого числа имеет вес, равный длине его двоичного представления. При этом любая верхняя оценка T требуемого числа шагов для рассматриваемой машины с указателями увеличится, но не более чем в $\log T$ раз по порядку. Однако нужно иметь в виду, что применимость такой модели ограничена размером внутренней памяти. Для больших задач с высокими требованиями к памяти многоленточная машина Тьюринга может опять оказаться более адекватной моделью.

При обсуждении вопросов временнóй сложности для таких моделей машин нужны некоторые соглашения о кодировании входных и выходных данных. Для простоты будем предполагать, что *целые числа* всегда представляются в стандартной *двоичной* форме, хотя не исключено, что при других кодированиях результаты могут быть иными (например, проверка свойства быть степенью числа 10 окажется, по-видимому, намного проще для десятичного представления). Элементы поля $GF(p)$ задаются как (приведённые) вычеты по модулю p, а рациональные числа — как пары целых; элементы конечных расширений этих полей представляются тогда очевидным образом как наборы двоичных чисел.

Для приближенных вычислений с вещественными или комплексными числами очень удобно использовать *двоично рациональные* числа, т. е. двоичные целые числа, масштабированные степенью двойки. Однако имеется значительная разница между придуманным ad hoc[1] понятием чисел с плавающей точкой (называемых в обычных языках программирования «вещественными») и нашим пониманием вещественного числа. Любое *входное* число $\alpha \in \mathbf{R}$ у нас (потенциально) доступно с любой желаемой точностью: при обращении с данным значением параметра N к некоторому *оракулу* тот выдаст двоично рациональное число a, такое что $|\alpha - a| < 2^{-N}$, без дополнительных затрат времени. Возможно, что различные оракулы определяют таким способом одно и то же вещественное число. Несмотря на то что входы α, β доступны со сколь угодно высокой точностью, равенство $\alpha = \beta$ рекурсивно неразрешимо. Аналогично выходы, получающиеся в результате приближенных вычислений, тоже будут доступны с некоторой заданной точностью.

В абстрактной постановке *теории алгебраической сложности* (см. [4] и более поздний обзор [50]) используется машинно-независимое понятие *неветвящейся программы*. Здесь обычно в качестве входов рассматриваются переменные $x_1, x_2, \ldots$ над некоторым основным полем F, а время (последовательного выполнения операций) измеряется как число арифметических операций, выполняемых такой программой в $F(x_1, \ldots)$ или в $F[x_1, \ldots]$, если в программе не допускаются деления. О *нескалярной сложности* говорят, когда принимается во внимание лишь количество *существенных* умножений и делений,— операции же сложения и умножения на скаляр не

---

[1] Для данного случая, по конкретному поводу (*лат.*). — *Прим. изд. ред.*

учитываются. Для задач, при решении которых требуются разветвления по сравнению (или проверка на нуль), используется более общая модель вычислительных деревьев (см. [49]).

2

2.1. *Сложность умножения и деления.* Возьмём для начала совсем простое уравнение $ax = b$. Если рассматривать его над целыми числами, то проверить его разрешимость и найти решение x можно с помощью целочисленного деления с нулевым остатком. Для проверки данного решения x достаточно и *умножения* целых чисел, а возможны ли более быстрые методы проверки — это интересный открытый вопрос. Решение уравнения $ax = b$ над рациональными числами (представляемыми парами целых) сводится к двум умножениям целых чисел, за которыми следует вычисление н. о. д., если результат надо представить в приведённой форме. Аналогично для решения уравнения $ax = b \bmod q$ (в предположении что b делится на $d = (a, q)$) требуется вычислить обобщённый н. о. д., т. е. найти н. о. д. числа d и *кофакторов* u, v, удовлетворяющих условиям

$$au + qv = d, \quad d \mid a, \ d \mid q. \tag{2.1}$$

Во всех этих задачах ключевую роль играет приближённое деление вещественных чисел. Его сложность оценивается через сложность умножения целых чисел, ибо n-битовое приближение к $1/a$ (скажем, при $1 \leqslant a < 2$) можно вычислить с помощью ньютоновских итераций без делений, использующих на каждом шаге 2 умножения; поэтому временна́я сложность n-битового деления ограничена сверху величиной

$$O(\mu(n) + \mu(n/2) + \mu(n/4) + \mu(n/8) + \ldots) \leqslant O(\mu(n)), \tag{2.2}$$

где $\mu(N)$ обозначает какую-нибудь верхнюю границу для временно́й сложности n-битового умножения (на эту границу налагаются некоторые условия регулярности, чтобы можно было обосновать оценку суммы в (2.2)). Поэтому сложность решения уравнения $ax = b$ в вещественных числах (когда a отделено от нуля) по существу не выше сложности проверки решения с помощью умножения. Обратно, равенство $ab = b/(1/a)$ показывает, как свести умножения к делениям.

В [18, § 4.3.3] Д. Кнут даёт довольно полный обзор того, что́ в настоящее время известно о временно́й сложности умножения целых чисел. Для многоленточных машин Тьюринга лучшей всё ещё остается указанная в нашей статье [41] 1971 года верхняя оценка, основанная на методе быстрого преобразования Фурье (БПФ) в кольце вычетов по модулю вида $2^L + 1$. Более рациональная версия этого подхода описана в [37]. С другой стороны, до сих пор никто не получил нелинейной *нижней* оценки, если не считать моделей со специальными дополнительными ограничениями. В действительности любое доказательство такой нижней оценки должно будет неизбежно использовать некоторые специфические свойства машин Тьюринга, поскольку, к примеру, на машинах с указателями можно умножать целые числа в *линейное время* (ср. [35]). Последний результат основан на численном БПФ подходящей точности. Таким образом, наши сегодняшние познания о временно́й сложности решения уравнения $ax = b$ над **Z**, **R** (а также и **C**) сводятся к следующим оценкам:

(2.3) Верхние оценки временно́й сложности N-битового умножения:

$\mu(N) = cN \log(N + 1) \log\log(N + 2)$ для многоленточных машин Тьюринга,

$\mu(N) = cN$ для машин с указателями (единичный вес),

$\mu(N) = cN \log(N + 1)$ для машин с указателями (логарифмический вес).

Для других упомянутых выше случаев, где фигурирует вычисление н. о. д., сложность, похоже, выше. В [31] на основе идеи Кнута, состоящей в комбинировании быстрого умножения целых чисел с принадлежащей Лемеру техникой «полу-н.о.д.», было показано, что вычисление н.о.д. двух чисел битовой длины не выше N (вместе с соответствующими кофакторами) возможно за время $O(\mu(N) \log N)$ (а не за время N^2 по порядку, как для алгоритма Эвклида). Представляется довольно правдоподобной гипотеза, что дополнительный множитель порядка $\log N$ при времени умножения неизбежен. Заметим, однако, что при данных кофакторах u, v *проверка* соотношения (2.1) на самом деле возможна за время порядка времени умножения.

Если при рассмотрении алгебраической сложности трактовать каждую арифметическую операцию как имеющую единичную стоимость, то вопрос о сложности решения уравнения $ax = b$ тривиален, поскольку для этого достаточно одного деления; но задача становится весьма нетривиальной, когда она трактуется как задача решения того же уравнения в конечномерных алгебрах над некоторым полем F. Например, рассмотрим поле комплексных чисел как алгебру над **R**. Совсем легко убедиться, что нескалярная сложность умножения двух комплексных чисел равна 3 (если учитывать только вещественные умножения и деления), но лишь недавно удалось решить соответствующую задачу для *деления* комплексных чисел — см. [27], где приведено довольно тонкое доказательство того, что известная верхняя граница 6 для числа требующихся умножений/делений действительно неулучшаема.

Много усилий было потрачено на изучение мультипикативной сложности алгебр. За подробностями и ссылками на литературу отсылаем к обзору [50]. Что касается «деления» в такой постановке вопроса, то здесь известно меньше. В следующем разделе кратко обсуждается специальный случай деления полиномов от одной переменной. Центральную же тему — обращение матриц — мы отложим до п. 4.3.

2.2. Основные вычисления с полиномами. Рассмотрим теперь уравнение $a(t)x(t)=b(t)$ для полиномов над некоторым полем F. Сначала обсудим алгебраическую модель, в которой все арифметические операции в F учитываются с единичным весом. Вход и выход задаются соответствующими коэффициентами, причём используется плотное представление полиномов. Точнее, предполагается, что F имеет вид $G(a_0, a_1, \ldots, b_0, \ldots)$, где коэффициенты входа трактуются как переменные над основным полем G. Мы опять можем выделить различные области, как то $F[t]$, $F(t)$ или $F[t]/(q(t))$ — по аналогии с рассмотренными ранее $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{Z}/q\mathbf{Z}$. Роль $\mathbf{R}$ здесь играет кольцо формальных степенных рядов над F. Увеличение точности по переменной в $F[[t]]$ означает, что мы рассматриваем $F[t]/(t^m)$ для возрастающих значений m. Поэтому сложность деления полиномов можно свести к сложности умножения полиномов (опять с точностью до постоянного множителя) с помощью приближённого вычисления обратных величин для единиц в $F[[t]]$, используя ньютоновские итерации; см. [42], [19]. На самом деле эти авторы разбирают более простой случай *нескалярной* сложности, которая для умножения полиномов по модулю t^m известна точно ($=2m-1$, в предположении что основное поле G содержит не менее $2m-2$ элементов). Верхнюю границу Кунга $4m$ для вычисления обратных по модулю t^m и гипотезу о её неулучшаемости можно заменить лучшей оценкой $3.75m$, однако точное значение мультипликативной константы пока неизвестно.

Алгебраическая сложность умножения полиномов по существу есть сложность дискретной свёртки и потому тесно связана с дискретным преобразованием Фурье. Если основное поле G содержит подходящие корни из единицы, например все корни из 1 степени 2^k или 3^k (для всех k), то можно непосредственно использовать БПФ, в противном случае нужны какие-то дополнительные приемы. Лучшие оценки, известные в настоящее время, очень похожи на оценки умножения целых чисел:

(2.4) Верхние оценки для числа арифметических операций при умножении полиномов степени m над основным полем G:

$M(m)=cm\log(m+1)$, если G допускает БПФ,

$M(m)=cm\log(m+1)\log\log(m+2)$ для произвольного поля.

Оценка для общего случая с дополнительным множителем $\log\log$ получается рекурсивным применением метода БПФ к кольцам полиномов $F[t]/(t^K+1)$ для различных значений $K=2^k$, аналогично технике быстрого умножения целых чисел. Этот подход не проходит для полей характеристики 2, но тут вместо упомянутых выше колец можно использовать

$$F[t]/(t^{2K}+t^K+1)$$

с $K=3^k$ (см. [34]). Отметим также, что для обсуждаемых алгебраических моделей никаких нелинейных нижних оценок для дискретного преобразования Фурье или умножения полиномов не известно.

Наилучшие верхние оценки, полученные в настоящее время для сложности решения уравнения $a(t)x(t)=b(t)\bmod q(t)$, где $q(t)$ — произвольный полином степени m, опять выше в $\log m$ раз по порядку. Довольно точные верхние и нижние оценки нескалярной сложности вычисления обобщённого н. о. д. полиномов получены Штрассеном (см. [49], где можно найти дальнейшие ссылки на литературу). Одна численная версия проблемы н. о. д. (над полями $\mathbf{C}$ или $\mathbf{R}$) изучалась в [40].

Это последнее замечание подводит нас к принципиальному вопросу, который рассматривается ниже: Каковы следствия результатов об *алгебраической* сложности для временно́й сложности (относительно машинных моделей) соответствующих *численных* вычислений с полиномами?

Подставляя числа (элементы $\mathbf{Z}$, $\mathbf{R}$ или $\mathbf{C}$) вместо коэффициентов полиномов, можно очевидным образом преобразовать быстрые алгебраические алгоритмы в быстрые численные процедуры (в предположении вычислительной устойчивости); но имеются также *машинные алгоритмы* другого типа, которые не получаются таким способом, и нельзя исключить, что некоторые из них значительно быстрее, чем какие бы то ни было построенные чисто алгебраически. Похоже, что численное умножение полиномов даёт пример такого феномена. Если просто скомбинировать оценки для БПФ из (2.4), которые основаны на единичной стоимости операций, с оценками (2.3) для умножения целых, то в лучшем случае получается оценка $O(m \log m\mu(N))$ — но можно-то ещё лучше!

Замена переменной t на 2^K для некоторого $K \geqslant 2N + \log(m+1) + 1$ сводит умножение двух полиномов степени m с коэффициентами длины не более N бит к одному умножению «длинных» целых размера $O(m(N + \log(m+1)))$, и для $N > \log(m+1)$ это приводит к улучшенной временно́й оценке $O(\mu(mN))$. Для машин с указателями мы таким образом получаем даже линейную оценку $O(mN)$. Линейность по m также показывает, что, по крайней мере асимптотически, знаменитое БПФ не является оптимальным численным методом для дискретного преобразования Фурье $\mathbf{C}^m \to \mathbf{C}^m$. Подробнее об этом подходе см. [37].

Если умножать полиномы с комплексными коэффициентами таким способом, основываясь на быстром умножении целых по модулю $2^{2L}+1$, то можно успешно сыграть на том, что $2^L = \mathrm{sqrt}(-1)$ служит мнимой единицей в этой области. Это обстоятельство существенно расширяет сферу приложений обсуждаемого метода.

Численное *деление* полиномов требует некоторых дополнительных забот об устойчивости. При делении на какой-либо (комплексный) полином F нужно, чтобы его старший коэффициент был отделен от нуля. Хороший способ выразить это количественно — задать верхнюю границу для *корневого радиуса* $\rho(F)$ (определяемого как максимум модулей корней F в $\mathbf{C}$) вместе с некоторой нормировкой *размера* F, в качестве которого берется l^1-норма $|F|$ вектора коэффициентов. Соответствующий результат из [37] гласит:

(2.5) Численное деление полинома $G \in \Pi_m$ на $F \in \Pi_n$ с ошибкой 2^{-N}, т. е. вычисление таких $Q \in \Pi_{m-n}$ и $R \in \Pi_{n-1}$, что $|G - QF - R| < 2^{-N}$, возможно на машине с указателями за время $O(m(N + m + m \log(1+r)))$, где r — *заданная* граница для $\rho(F)$, и предполагается, что $|G| \leqslant 1 \leqslant |F| \leqslant 2$.

Если ограничиться делениями с равномерно ограниченным корневым радиусом знаменателей, то данная оценка для машин с указателями опять оказывается порядка $O(mN)$, в предположении что точность N не меньше m по порядку.

3. ОСНОВНАЯ ТЕОРЕМА АЛГЕБРЫ С ТОЧКИ ЗРЕНИЯ ВЫЧИСЛИТЕЛЬНОЙ СЛОЖНОСТИ

В течение многих веков проблема решения полиномиальных уравнений в основном трактовалась в терминах *формул*. И после того как математики эпохи Возрождения «освоили» третью и четвертую степень, прошло ещё более 250 лет до того времени, когда Гаусс дал строгое доказательство *существования* корней у любого вещественного уравнения, а Абель смог показать, что в общем случае невозможно получить их, только применяя арифметические операции и последовательно решая «чистые» уравнения. Это узкое понятие «разрешимости» дошло и до наших дней, по крайней мере в устной традиции, несмотря на довольно ясные замечания в § 9 диссертации Гаусса о том, что следует чётко различать «resolutio aequationis» и «ipsius reductio ad aequationes puras»[1].

И лишь ещё столетие спустя удалось адекватно, а именно в терминах *алгоритмов*, понять, что же такое решение общего уравнения с комплексными коэффициентами. Данное Г. Вейлем [53] конструктивное доказательство основной теоремы алгебры по существу показывает, что нули комплексного полинома *рекурсивно* зависят от его коэффициентов (см. также [11] по поводу вклада Шпеккера в исследование этого вопроса). Другими словами, существует машина Тьюринга, которая, имея вход n и оракулы для коэффициентов полинома степени m, выдаёт приближение к m нулям этого полинома с погрешностью, не превосходящей 2^{-n}. С точки зрения *вычислительной сложности* ключевым является вопрос: *сколь быстро* это можно сделать?

3.1. Формулировка вычислительной задачи. Хорошо известен тот факт, что нули полинома, образующие сгусток, менее устойчивы к малым возмущениям коэффициентов, чем изолированные нули. Поэтому представляется разумным понимать любую заранее предписанную точность в смысле обратного анализа: набор приближённых нулей следует считать приемлемым, если элементарные симметрические функции от этого набора достаточно хорошо совпадают с коэффициентами данного полинома. Это применяется к полиному $P \in \Pi_m$, $P(z) = a_0 + a_1 z + \ldots + a_m z^m$ со старшим коэффициентом $a_m = 1$, но нужно допускать также и более общий случай $a_m \to 0$, т. е. когда некоторые нули уходят на бесконечность. Этот путь приводит нас к следующей формулировке задачи:

(3.1) Вычислительная задача *приближённой факторизации*: По данному целому $N > 0$ и данному полиному $P \in \Pi_m$ с нормой $|P| \leqslant 1$ вычислить линейные множители $L_j(z) = u_j z + v_j$ $(1 \leqslant j \leqslant m)$, такие что $|P - L_1 L_2 \ldots L_m| < 2^{-N}$.

---

[1] «Разрешение уравнений» и «непосредственное сведе́ние к чистым уравнениям» (*лат.*). — *Прим. изд. ред.*

Здесь и в дальнейшем $|\cdot|$ обозначает l^1-норму вектора коэффициентов полинома. Использование других обычных норм, как то l^2-нормы или $\max_{|z|=1}|P(z)|$, не очень меняет ситуацию, ибо множители, связывающие эти нормы, меньше $m+1$, что означает изменение N не более чем на $\log(m+1)$.

При анализе эффективности алгоритмов для этой задачи мы в основном интересуемся их сложностью *в худшем случае* (отметим, что «худший» случай на самом деле не обязательно «плохой»). Для любого такого алгоритма и для фиксированных значений параметров m, N имеется корректно определённое максимальное время работы алгоритма $T(m, N)$ — максимальное относительно всех входов (оракулов) $(a_0, \ldots, a_m)$ с l^1-нормой $\leqslant 1$. (Если бы этого максимума не существовало, то из леммы Кёнига следовало бы существование входов, на которых этот алгоритм никогда бы не останавливался.) Теперь мы можем переформулировать основную проблему этого раздела более точно: каков истинный порядок роста максимального времени работы $T(m, N)$ (почти) оптимальных алгоритмов приближённой факторизации, и каково асимптотическое поведение $T(m, N)$ при $N \to \infty$ или при возрастании степени полинома? Наиболее правдоподобно, что соответствующие ответы будут до определённой степени зависеть от рассматриваемой модели машины. В свете неполноты наших знаний о много более простых сложностных задачах, таких как умножение целых, здесь нельзя надеяться более чем на частичные ответы. Наш основной результат, излагаемый ниже, дает хорошую верхнюю границу сложности приближённой факторизации. Но перед тем, как углубляться в это, мы кратко обрисуем некоторые из частичных ответов, которые, по крайней мере неявно, содержатся в обширной литературе о методах нахождения корней.

В 1967 г. состоялся симпозиум о конструктивных аспектах основной теоремы алгебры. Труды этого симпозиума [11] хорошо отражают состояние дел в то время. Далее, в главе 6 книги Хенричи [14] дан обстоятельный обзор существующих методов. Во многих численных методах, используемых на практике, пытаются достичь компромисса между скоростью и универсальностью. Основная трудность связана с тем фактом, что все надёжные алгоритмы для приближённой факторизации необходимо должны использовать технику многократной точности, которая обычно работает медленно, и нужны строгие априорные оценки для ошибок округления, чтобы гарантировать надёжные результаты. То же верно и для всех итеративных методов, при применении которых обычно возникает дополнительная трудность нахождения подходящих начальных значений.

Применяя идеи, имеющиеся в литературе по численным методам, в строгом контексте, объясненном ранее, мы приходим к заключению, что приближённая факторизация заведомо возможна в *полиномиальное время*. На основе метода исключения Г. Вейля в виде, предложенном Хенричи [14, с. 517—522], мы можем получить, к примеру, оценку $T(m, N) = O(m^7 N^3)$. Похожие границы на время имеются для методов изоляции корней, развитых в компьютерной алгебре (см. [8] и указанные там работы). Проблема нахождения более точных оценок бросает вызов исследователям. Применение техники быстрого умножения целых чисел и полиномов (см. 2.1 и 2.2) понижает упомянутую границу до чего-то вроде $O(m^{5+\delta} N^{2+\delta})$ с произвольно малым $\delta > 0$, но похоже, что множитель N^2 этими методами не устранить.

3.2. Верхняя граница. Основной результат, который будет представлен здесь, — это верхняя граница для временно́й сложности приближённой факторизации; эта граница была найдена автором в 1981/82 годах. Она основывается на одном новом методе, называемом *методом разделяющего круга* (по-немецки Trennkreisverfahren) и описанным в предварительном сообщении [38]. Часть этого материала носит в высшей степени технический характер, и многие детали реализации метода нуждаются в более тщательной проработке. Полное изложение этих результатов появится в виде монографии. Далее мы дадим краткое описание основных идей и результатов.

Теорема. *Приближённая факторизация комплексных полиномов, как эта задача сформулирована в* (3.1), *возможна за максимальное время работы*

$$T(m, N) = O(m\mu(m^2 \log m) + m\mu(mN)) \tag{3.2}$$

для машинных моделей и соответствующих оценок времени умножения целых чисел, указанных в (2.3).

Чтобы упростить изложение, в дальнейшем ограничимся случаем, когда время умножения на машине с указателями линейно: $\mu(N) = cN$. (В случае других моделей добавляются дополнительные логарифмические множители.) Тогда верхняя оценка приобретает вид $T(m, N) = O(m^3 \log n + m^2 N)$. Полезно сравнить её с очевидными временны́ми границами для соответствующей задачи проверки. При этом можно принять как общее правило приближённого счета, что вычисления с полиномами степени порядка m следует выполнять по меньшей мере с m-битовой точностью, ввиду следующих теоретических соображений.

Прежде всего надо осознать тот факт, что у малых полиномов могут быть большие множители. При $m=2k$ для простого примера

$$P(z)=2^{-k}(z^2-i)^k=2^{-k}(z-a)^k(z+a)^k$$

(где $a^2=i$) мы имеем частичное разложение $P=FG$ с $|P|=|F|=1$, но $|G|=2^{m/2}$ — тем самым соответствующее увеличение ошибок, сделанных ранее при вычислении F, с необходимостью потребует дополнительной точности в $m/2$ битов. Точная верхняя оценка в случае многих сомножителей содержится в следующем *количественном дополнении к основной теореме алгебры*, которое обычно отсутствует в учебниках!

(3.3) Пусть $f\in\Pi_m$, $f=f_1\ldots f_k$. Тогда $|f|\leqslant|f_1|\ldots|f_k|\leqslant 2^{m-1}|f|$.

Эта верхняя оценка достигается для таких банальных полиномов, как $f(z)=z^m-1$ в случае его полной факторизации, т. е. при $k=m$.

Похожее увеличение ошибки может быть вызвано *сдвигами Тэйлора* (при фиксированном a операторная норма линейного отображения, определённого на Π_m формулой $g(z)=f(z+a)$, равна $(1+|a|)^m$) или численным делением полиномов, см. (2.5). При $N\geqslant m$ мы, таким образом, видим, что любой наивный метод, предназначенный только для *проверки* точности данной приближённой факторизации, потребует по меньшей мере $O(m^4)$ битовых операций, и даже с помощью техники быстрого умножения нам не удаётся сделать это быстрее, чем за время порядка $Nm\log m$ для машин с указателями; по всей видимости это и есть точный порядок роста для сложности этой задачи проверки.

Поэтому граница (3.2) представляется оптимальной с точностью до множителя порядка не более m. При возрастании точности второй член $O(m^2N)$ становится доминирующим, и для широкого класса полиномов, допускающих *сбалансированные разбиения* (см. ниже), его можно заменить меньшей величиной $O(Nm\log m)$. Основным достоинством всех этих оценок является их линейность по N. Для любого *фиксированного* m оценка (3.2) превращается просто в $O(\mu(N))$ (где неявные константы зависят от m), а это действительно оптимальная граница. Для машин с указателями мы получаем линейную оценку на время, причем для выдачи результатов требуется не более того же времени, разве что увеличенного в постоянное число раз. Но даже для моделей машин, для которых точная сложность умножения целых чисел ещё не известна, справедлив следующий относительный результат:

(3.4) **Теорема.** *При $N\to\infty$ и фиксированных $k,m\geqslant 2$ один и тот же порядок роста имеют оценки сложности следующих задач:*

(a) *умножение N-битовых целых чисел;*

(b) *вычисление $(1+x)^{1/k}$ для $|x|\leqslant 1/2$ с ошибкой не более 2^{-N};*

(c) *приближённая факторизация полиномов степени m с ошибкой не более 2^{-N}.*

Чтобы получить оценку для (c) с помощью (a), мы предполагаем выполненным некоторое условие регулярности, как и для (2.2), а затем применяем (3.2). Приближённая факторизация полинома $z^k-(1+x)$ сводит (b) к (c) при $k\leqslant m$, в общем же случае переходим от (b) к (c) через (a). Переход от (b) к (a) основан на одной идее X. Альта, состоящей в том, чтобы моделировать возведение в квадрат целого числа с помощью достаточно точного вычисления значения

$$\left(2-(1+x)^{1/k}-(1-x)^{1/k}\right)k^2/(k-1)=x^2+O(x^4)$$

для малых x.

В п. 3.4 мы приведём рассуждение, основанное на идее возмущения, которое нужно для получения из приближённой факторизации приближённых *решений* (нулей) полиномиального уравнения со старшим коэффициентом единица и, скажем, ограниченными коэффициентами. В случае когда степень многочлена фиксирована, это увеличит оценки времени исполнения алгоритма не более чем в ещё какое-то постоянное число раз. Таким образом, эквивалентность (b) и (c) показывает, что в отличие от результатов Абеля и Галуа о «разрешимости», с точки зрения вычислительной сложности нет значительной разницы между решением чистых или общих уравнений с комплексными коэффициентами. С вычислительной точки зрения приближённое решение общего полиномиального уравнения сводится к очевидно более фундаментальной задаче умножения целых чисел.

3.3. Метод разделяющего круга. Общая стратегия приближённой факторизации данного полинома $P\in\Pi_m$ основана на подстратегии приближённого расщепления P на два множителя $F\in\Pi_k$ и $G\in\Pi_{m-k}$, которая применяется рекурсивно. Случай $k=1$ просто означает приближённое определение одного нуля. Высокая точность может быть достигнута применением ньютоновских итераций, если нуль простой и хорошо изолирован и известно приемлемое начальное приближение. Аналогично случай $k=2$ «обслуживается» методом Бэрстоу, который квадратично сходится, если соответствующие два нуля хорошо отделены от остальных $m-2$ нулей. Метод разделяющего круга использует соответствующую общую ньютоновскую итерацию (ср. со статьей Шрёдера в сборнике [11]), где выбор подходящего k производится в зависимости от распределения нулей, так что k нулей P оказываются *внутри* некоторого подходящего *разделяющего круга*, а остальные $m-k$ нулей — *вне* этого круга, причём часть из них может быть близка к бесконечности. Более того, все нули должны находиться на некотором расстоянии от этого круга. Фундаментальным фактом является то, что такой круг можно определить всегда, за исключением единственного случая $P(z) \approx (az+b)^m$, когда приближённая факторизация уже и так выполнена.

Имеются устойчивые преобразования (перемасштабирование, сдвиги Тэйлора), с помощью которых любую окружность можно превратить в стандартную единичную окружность $E=\{z\colon |z|=1\}$. Ввиду упомянутого выше условия о наличии кольца вокруг E, не содержащего нулей, можно получить разумную нижнюю оценку для ключевой величины $\nu=\min_{|z|=1}|P(z)|$, и притом затратив на это не столь уж большие усилия (как с теоретической, так и с вычислительной точек зрения). При выборе разделяющих кругов детали можно обставить так, чтобы всегда была выполнена оценка $\log(1/\nu) \leqslant O(m)$. Если исходить из такого ν, то анализ общего метода Ньютона приводит к явным оценкам для его квадратичной сходимости. Это делается следующим образом.

Пусть дано некоторое приближённое разделение единичным кругом $|P-FG|<\varepsilon$ (т. е. k корней F и обратные к $m-k$ корням G ограничены величиной $1-\delta$). *Ньютоновской поправкой* назовём *пару* $(f, g) \in \Pi_{k-1} \oplus \Pi_{m-k-1}$, выбираемую так, чтобы члены более чем первого порядка взаимно сократились. Тогда новые приближённые множители $F+f$ и $G+g$ будут удовлетворять равенствам

$$P-(F+f)(G+g)=P-FG-fG-gF-fg=-fg,$$

причём новая ошибка ограничена величиной $|fg| \leqslant |f|\,|g|$, где f и g однозначно определяются (неполным) разложением на элементарные дроби $(P-FG)/(FG)=f/F+g/G$. Поэтому для получения верхней оценки для $|f|$ можно воспользоваться следующим интегральным представлением:

$$f(z)=\frac{1}{2\pi i}\int\limits_E \frac{(P-FG)(t)}{(FG)(t)}\,\frac{F(z)-F(t)}{z-t}\,dt,$$

и аналогично для $|g|$. Таким образом, в частности, доказывается, что достаточно начального приближённого разделения с точностью

$$|P-F_0G_0|<\varepsilon_0=\nu^4 2^{-cm},$$

где c — некоторая константа.

Такие начальные F_0 и G_0 можно найти с помощью контурного интегрирования (используемого также в [12], но без оценок ошибки). Суммы s_j степеней рассматриваемых k нулей полинома P, лежащих внутри E, представляются в виде

$$s_j=\frac{1}{2\pi i}\int\limits_E \frac{P'(z)}{P(z)}\,z^j\,dz.$$

Наличие кольца, не содержащего нулей, и нижней границы ν позволяет нам дать явные оценки точности, требующейся для одновременного вычисления значений этих интегралов с помощью дискретного преобразования Фурье. (Этот шаг вносит существенный вклад в первый член верхней оценки (3.2).) На основе приближений для $s_0=k$, $s_1, \ldots, s_k$ легко строится приближённый множитель F_0, после чего подходящий полином G_0 можно найти полиномиальным делением.

До сих пор мы молчаливо предполагали, что можно (заранее) достаточно хорошо локализовать нули данного полинома, чтобы обеспечить правильный выбор разделяющих кругов, но это, в действительности, и есть ключевая проблема всего метода. Наши алгоритмы для соответствующей *диагностики* основаны на методе Греффе[1], в котором *корни возводятся в квадрат*. Мы используем вариант, особенно удобный для применения быстрого умножения целых чисел. В отличие от обычных применений метода Греффе, в нашем случае мы прежде всего интересуемся не кругами, *в* которых расположены нули, а кругами, отделёнными от всех нулей, — для такой задачи достаточна умеренная точность. Далее, не следует пытаться одновременно вычислить все модули корней $\rho_1(f) \geqslant \rho_2(f) \geqslant \ldots \geqslant \rho_m(f)$ некоторого полинома f, что и в самом деле может потребовать неприемлемых (т. е. слишком трудоёмких) вычислений с длинными числами (см. обсуждение этого вопроса в работе Турана [52]). Вместо этого можно изменять *фокусировку точности* подходящим перемасштабированием после каждого шага возведения корней в квадрат, так чтобы сориентировать алгоритм на вычисление одного конкретного $\rho_k(f)$.

---

[1] В отечественной литературе чаще называемом «методом Лобачевского». — *Прим. перев.*

Типичным результатом, получаемым таким образом, является следующая оценка времени:

(3.5) Пусть дан комплексный полином $f(z) = a_0 + a_1 z + \ldots + a_m z^m$, у которого $|f| \leqslant 1$, и пусть заданы положительное $\delta < 1/2$ и индекс $k \leqslant m$. Тогда модуль k-го корня $\rho_k(f)$ можно вычислить с погрешностью, меньшей δ, на машине с указателями за время

$$O\,(m^2 (\log m + \log (1/\delta)) (\log \log m + \log (1/\delta))).$$

(Кроме того, должно выполняться что-то вроде $|a_0|, |a_m| \geqslant (\delta/m)^m$.)

При $\delta = m^{-O(1)}$ это даёт довольно благоприятную оценку времени $O(m^2 \log^2 m)$, но при много бо́льших требованиях к точности, когда доминирующим становится квадратичный рост по $\log(1/\delta)$, этот подход уже нельзя рекомендовать.

По поводу других аналогичных оценок и дальнейших подробностей сошлёмся на [38]. Полный анализ времени для метода разделяющего круга показывает, что предпочтительны сбалансированные разделения (с k и $m-k$ одинакового размера, идеально $k = m/2$). Даже в случае, когда m простых корней хорошо изолированы, принцип отделения линейных множителей одного за другим проигрывает по сравнению с принципом *разделяй и властвуй* из теории сложности вычислений, но существуют определённые трудные полиномы, которые в силу специфического расположения их нулей будут вынуждать использование несбалансированных разделений. Целое семейство таких примеров (с произвольными параметрами $1 < u_j < 1.01$) даётся формулой

$$f(z) = (z + u_1/4)(z + u_2/16)(z + u_3/64) \ldots (z + u_m/4^m).$$

Здесь $N = 2m^2$ описывает точность, хорошо настроенную относительно размера коэффициентов. Это свойство *сильной вложенности* представляет собой новый феномен в численной трактовке полиномов. В настоящее время трудно сказать, являются ли такие полиномы плохо обусловленными с точки зрения приближённой факторизации или они всего лишь демонстрируют специфическую слабость метода разделяющего круга. Во всяком случае эту вложенность следует должным образом отличать от случая сгустков нулей, который сам по себе не вызывает дополнительных трудностей для приближённой факторизации полиномов благодаря технике растаскивания с помощью подходящего масштабирования (см. [38]). Однако наличие сгустков влияет на устойчивость нулей, о которой пойдёт речь ниже.

3.4. Некоторые приложения. Любую приближённую факторизацию $|P - L_1 \ldots L_m| < \varepsilon$ с $0.5 \leqslant |P| \leqslant 1$ и известными линейными множителями $L_j(z) = u_j z + v_j$ можно использовать для приближённого определения нулей полинома P. Естественными кандидатами для приближений к нулям являются числа $w_j = -v_j/u_j$; в случае $|v_j| > |u_j|$ может оказаться предпочтительнее рассматривать их обратные в качестве приближений к обратным значениям корней P.

Соответствующие неравенства $|P(w_j)| < \varepsilon$ (или $|P^*(1/w_j)| < \varepsilon$ для обращённого полинома $P^*(z) = z^m P(1/z)$) в сочетании с хорошо известным принципом гомотопии дают $W = \{z\colon |P(z)| < \varepsilon \max(1, |z|^m)\}$ в качестве *множества включения* нулей, а их возмущения ограничены диаметрами компонент W. Сейчас мы рассмотрим только априорные границы (для худшего случая), в практических же приложениях следует, конечно, использовать возможность сокращения вычислений на основе апостериорного улучшения границ. Сохраняя предыдущие обозначения, имеем следующую оценку возмущения (теорема 2.7 в [40]):

(3.6) Нули $z_1, \ldots, z_m$ полинома P можно занумеровать так, что (при $\varepsilon < 2^{-7m}$)

$$|z_j - w_j| < 9\varepsilon^{1/m} \quad \text{при} \quad |w_j| \leqslant 1,$$
$$|1/z_j - 1/w_j| < 9\varepsilon^{1/m} \quad \text{при} \quad |w_j| \geqslant 1.$$

Эти оценки оптимальны с точностью до мультипликативной константы и лучше по порядку в m раз, чем соответствующие оценки Островского (ср. [29, приложения A, B]). Если поставить целью найти корни некоторого полинома степени m с погрешностью 2^{-n}, то, очевидно, будет достаточно приближённой факторизации с $\varepsilon = 2^{-N}$, $N = (n + 4)m$, отсюда и из оценки времени (3.2) следует, что

(3.7) На машине с указателями можно вычислить нули (или их обратные) любого полинома P степени m (с нормой, скажем, $0.5 \leqslant |P| \leqslant 1$) с погрешностью, меньшей 2^{-n}, за время $O(m^3 \log m + m^3 n)$.

В некоторых приложениях достаточно определить лишь один нуль. Применительно к этой задаче можно так модифицировать метод разделяющего круга, чтобы второй член в оценке времени (3.7) можно было заменить меньшей величиной $O(m^2 n)$, поскольку после первого приближённого разбиения P на два множителя F и G нужно дальше работать только с одним из них (степени $\leqslant m/2$) и так далее, рекурсивно. Соответственно, в этом случае наиболее желательны несбалансированные разбиения.

В качестве такого рода примера мы хотели бы упомянуть довольно быстрый метод факторизации полиномов от одной переменной с целыми коэффициентами, который основан на вычислении одного нуля с высокой точностью и на диофантовом приближении для нахождения его минимального полинома в качестве множителя (см. [26] и [39]). На этом пути становятся реально возможными разнообразные численные операции над *алгебраическими числами.*

Другим важным следствием из оценок времени, приведённых выше, является то, что *все алгебраические функции можно вычислять быстро.* В своей работе [20], носящей в точности такое название, Кунг и Трауб исследуют сложность несколько иной задачи, а именно задачи вычисления начального отрезка одного из разложений алгебраической функции в (дробно-) степенной ряд, причём они считают число арифметических операций, приписывая им единичный вес, и трактуют нахождение нуля полинома с числовыми коэффициентами (с любой заданной точностью) как примитивную операцию.

В нашей модели такая задача в общем случае неразрешима, так как невозможно распознавание равенства. Вместо этого рассмотрим более прямой подход. Пусть даны полином $F(z, v) = a_0(v) + a_1(v)z + \ldots + a_m(v)z^m$ и точка $v = (v_1, \ldots, v_k)$, т. е. коэффициенты полиномов $a_j \in \mathbf{C}[v_1, \ldots, v_k]$ и точка v опять доступны как оракулы. Тогда значения m ветвей алгебраической функции $z(v)$, определяемой равенством $F(z, v) = 0$, можно найти в точке v за время, ограниченное по порядку так же, как в (3.7), в предположении что v и все a удовлетворяют некоторым условиям нормировки. Самая сильная сторона этого утверждения состоит в его *равномерности* по v; можно так выбрать скрытые константы, чтобы, например, утверждение было верно при $|v| \leqslant 1$.

4. ЛИНЕЙНЫЕ УРАВНЕНИЯ И СЛОЖНОСТЬ МАТРИЧНЫХ ВЫЧИСЛЕНИЙ

Одним из пионерских результатов на заре теории вычислительной сложности было открытие Штрассеном [46] того, что «гауссов метод исключения не оптимален». Он показал, как заменить классический метод умножения и обращения матриц со сложностью $O(n^3)$ алгоритмом, требующим менее $O(n^{2.81})$ арифметических операций. Основная идея этого результата совершенно элементарна и теперь широко известна, хотя многие

специалисты по численному анализу сохранили скепсис по поводу практической применимости таких методов.

Тем временем показатель 2.81 был существенно улучшен усилиями нескольких авторов (см. [3, 36, 10] и обзор Пана [30]). Наилучшая оценка, в настоящее время известная автору, чуть меньше 2.4785, она была сообщена Ф. Штрассеном на одном неформальном семинаре в 1985 г. (см. также [51])[1]. В действительности эти оценки касаются *сложности умножения матриц.* В пп. 4.1 и 4.2 мы дадим краткий обзор некоторых основных идей и результатов, относящихся к этой привлекательной нерешённой математической проблеме; может быть, этот набросок стимулирует других математиков внести свой вклад в её окончательное решение.

В п. 4.3 мы обсудим вопрос о сложности решения линейных систем и других процедур над матрицами, таких как обращение матриц, вычисление определителей и т. д., в основном в рамках теории алгебраической сложности, где элементы матриц и векторов рассматриваются как переменные над некоторым основным полем F. Имея дело с вещественными или комплексными числами, мы должны снова обращаться к численным аспектам, особенно в части оценок сложности приближенного определения собственных значений.

Что касается других областей, то здесь следует упомянуть, что линейные системы диофантовых уравнений можно решать за полиномиальное время, как было показано Й. фон цур Гатеном и М. Зивекингом [45, с. 57—65]. Описание соответствующих алгоритмов и дальнейшие ссылки на литературу можно найти в [7] и [16].

4.1. Показатель умножения матриц. Алгебраическую сложность умножения матриц можно измерять разными способами. Рассмотрим $n \times n$-матрицы $A = (a \ldots)$ и $B = (b \ldots)$, элементами которых являются переменные над некоторым полем скаляров F (которое на некоторое время зафиксируем). Пусть $L(n)$ обозначает минимальную длину неветвящейся программы в $F(a \ldots, b \ldots)$, вычисляющей элементы матрицы AB (при подсчёте этой длины мы учитываем все арифметические операции), и пусть $L^{*/}(n)$ есть соответствующая нескалярная сложность либо $L^*(n)$, в случае кольца $F[a \ldots, b \ldots]$, если не используются деления. Для бесконечного F имеем $L^{*/}(n) = L^*(n)$ (ср. [47]).

---

[1] Сейчас известна оценка 2.37 ... — см. Coppersmith D., Winograd S. Proc. 19th Annu ACM Symp. on Theory of Comput., 1987, pp. 1—6. — *Прим. перев.*

Последней введём (не последнюю по важности) *билинейную сложность* $L^{\otimes}(n)$, относящуюся к модели, в которой допускаются лишь умножения $\xi(a\ \ldots)*\eta(b\ \ldots)$ F-линейных форм от a на F-линейные формы от b.

Показатель умножения матриц (над F) определяется как следующая точная нижняя грань:

$$\omega(F) = \inf\{\beta \colon L(n) = O(n^{\beta})\}. \tag{4.1}$$

Очевидно, он удовлетворяет неравенству $2 \leqslant \omega(F) \leqslant 3$. Первая нетривиальная оценка Штрассена $\omega(F) \leqslant \log 7/\log 2 < 2.81$ была выведена из неравенства $L^{\otimes}(2) \leqslant 7$. Точно так же можно рекурсивно применить любой билинейный алгоритм для какого-либо другого малого n, чтобы умножать $n \times n$-*блочные матрицы*, поэтому из $L^{\otimes}(n) \leqslant r$ следует $L(N) = O(N^{\beta})$ при $\beta = \log r/\log n$. Вместе с элементарными неравенствами

$$L^{*/}(n) \leqslant L^{*}(n) \leqslant L^{\otimes}(n) \leqslant 2L^{*}(n)$$

это показывает, что мы можем, не изменяя $\omega(F)$, заменить $L(n)$ в (4.1) на любую другую меру в последней цепочке неравенств. В дальнейшем мы сосредоточим внимание на мере $L^{\otimes}(n)$ и её обобщениях на случай умножения прямоугольных матриц. Технические детали и доказательства см. в [36].

Пусть $\langle k, m, n\rangle$ обозначает *тензор* билинейного отображения, соответствующего умножению $k \times m$- на $m \times n$-матрицы. Его *ранг* $\operatorname{rk}\langle k, m, n\rangle$ равняется минимальной длине представлений следа ассоциированной трилинейной формы ABC, где C имеет размер $n \times k$, в виде суммы произведений линейных форм, которые (произведения) *имеют ранг* 1. В терминах координат это означает, при $A = (a\ \ldots)$, $B = (b\ \ldots)$, $C = (c\ \ldots)$, что неравенство $\operatorname{rk}\langle k, m, n\rangle \leqslant r$ эквивалентно существованию линейных форм ξ_j от a, η_j от b и ζ_j от c, таких что выполнены равенства

$$\sum_{j=1}^{r} \xi_j(a\ \ldots)\,\eta_j(b\ \ldots)\,\zeta_j(c\ \ldots) = \operatorname{tr}(ABC) = \sum_{\varkappa,\mu,\nu} a_{\varkappa,\mu} b_{\mu,\nu} c_{\nu,\varkappa}, \tag{4.2}$$

которые в свою очередь эквивалентны разрешимости большой системы из $k^2m^2n^2$ нелинейных уравнений над F с $kmr + mnr + nkr$ неизвестными коэффициентами упомянутых линейных форм. Устрашающий размер таких систем может дать некоторое представление о трудностях на пути определения значения любого $\operatorname{rk}\langle k, m, n\rangle$ даже в случае небольших размерностей (несмотря на тот факт, что для полей вроде $\mathbf{R}$ или $\mathbf{C}$ с разрешимой теорией первого порядка функция $\operatorname{rk}\langle k, m, n\rangle$ *вычислима*). Одно из редких известных значений таково: $\operatorname{rk}\langle 2, 2, 2\rangle = 7$ для любого поля F, поскольку всегда $\operatorname{rk}\langle n, n, n\rangle = L^{\otimes}(n) \geqslant L^{*}(n) \geqslant 2n^2 - 1$; последнее неравенство является частным случаем одной общей нижней оценки для сложности ассоциативных алгебр (см. [1, 50]). Следующий случай более высокой размерности, а именно случай 3×3-матриц, уже приводит к системе из 729 уравнений с 567 неизвестными (для $r = 21$); пока что ничего лучше $\operatorname{rk}\langle 3, 3, 3\rangle \leqslant 23$ неизвестно. Обобщая эти наблюдения, можно надеяться получить лучшие оценки для $\omega(F)$ из хороших верхних оценок ранга для *прямоугольного* случая с помощью следующей леммы:

(4.3) Функция $\operatorname{rk}\langle k, m, n\rangle$ симметрична по k, m, n и субмультипликативна, т. е.

$$\operatorname{rk}(\langle k', m', n'\rangle \otimes \langle k'', m'', n''\rangle) \leqslant \operatorname{rk}\langle k', m', n'\rangle \operatorname{rk}\langle k'', m'', n''\rangle.$$

Отправляясь от некоторой оценки $\operatorname{rk}\langle k, m, n\rangle \leqslant r$, мы применяем симметризацию и получаем $L^{\otimes}(kmn) = \operatorname{rk}\langle kmn, mnk, nkm\rangle \leqslant r^3$; поэтому

$$\operatorname{rk}\langle k, m, n\rangle \leqslant r = (kmn)^{\tau} \Rightarrow \omega(F) \leqslant 3\tau. \tag{4.4}$$

Долгое время, однако, ни одна из оценок ранга для малых размеров не оказывалась достаточно хорошей, чтобы привести к оценке, лучшей чем 2.81. В 1978 г. этот барьер преодолел В. Пан. С помощью своей техники *трилинейного агрегирования* он смог показать, что $L^{*}(n) \leqslant n^3/3 + O(n^2)$ с такими константами, что случай $n = 48$ привёл к улучшенной оценке $\omega(F) < 2.781$.

Аналогичная оценка была получена Бини, Каповани, Лотти и Романи [3], но они использовали совершенно новый подход, основанный на понятии граничного ранга. Пусть ε — ещё одна переменная над F. *Граничный ранг* $\operatorname{brk}\langle k, m, n\rangle$ можно определить как минимальную длину представлений в виде сумм (4.2) с точностью до $O(\varepsilon)$, причём теперь в качестве коэффициентов линейных форм допускаются *мероморфные формальные степенные ряды* по ε (определение на языке алгебраической геометрии см. в §§ 13, 14 работы [50]).

Свойства (4.3) и (4.4) верны и для граничного ранга, поэтому

$$\operatorname{brk}\langle k, m, n\rangle \leqslant r = (kmn)^{\tau} \Rightarrow \omega(F) \leqslant 3\tau. \tag{4.4b}$$

В [3] эти оценки были применены к $\operatorname{brk}\langle 2, 2, 3\rangle \leqslant 10$; другой пример: $\operatorname{brk}\langle 3, 3, 3\rangle \leqslant 21$, что даёт $\omega(F) < 2.772$. Для эталонного случая 2×2-матриц известны оценки $6 \leqslant \operatorname{brk}\langle 2, 2, 2\rangle \leqslant 7$.

При доказательстве утверждения (4.4b) используется тот факт, что тензорные степени любого тензора t, для которого $\mathrm{brk}\, t \leqslant r$, удовлетворяют неравенству $\mathrm{rk}\,(t^{\otimes s}) \leqslant O(s^2) \cdot r^s$; дополнительный множитель полиномиального роста не играет роли, когда (4.4) применяется при $s \to \infty$. Применяя аналогичные соображения, можно показать, что $\omega(F) = \omega(F_0)$, где F_0 обозначает простое поле поля F (см. [36, теорема 2.8]), поэтому показатель умножения матриц над F может зависеть только от характеристики поля F. Ниже мы будем писать просто ω, поскольку все дальнейшие рассуждения «равномерны» относительно F.

4.2. *Субаддитивность.* При *массовом решении* некоторых вычислительных проблем можно добиться существенной экономии ресурсов. Важным примером является вычисление значений общего полинома степени m в m различных точках (см. [4]). Аналогичные эффекты наблюдаются в области матричных вычислений. Умножение $n \times n$-матрицы A на вектор-столбец b использует в точности n^2 умножений, а умножение той же матрицы A на n различных векторов $b_1, \ldots, b_n$ возможно одним умножением матриц — менее чем за $O(n^{2.8})$ шагов. Можно возразить, что в этих примерах мы имеем дело с определённым количеством общей, совместной информации (коэффициенты полинома, матрица A). Поэтому введём теперь очень строгие правила. Рассмотрим две совершенно *непересекающиеся* (дизъюнктные) задачи умножения матриц и спросим, нельзя ли получить AB и UV одним составным вычислением с меньшими затратами, чем при (оптимальном) выполнении по отдельности двух умножений матриц.

При использовании билинейной меры это эквивалентно вопросу о *субаддитивности* функции rk (по отношению к прямым суммам), т. е. вопросу, может ли нестрогое неравенство

$$\mathrm{rk}\,(\langle k', m', n'\rangle \oplus \langle k'', m'', n''\rangle) \leqslant \mathrm{rk}\,\langle k', m', n'\rangle + \mathrm{rk}\,\langle k'', m'', n''\rangle \quad (4.5)$$

обращаться в строгое. Для тензоров порядка 2 (матриц) всегда имеет место равенство, для тензоров же порядка 3 соответствующая *гипотеза аддитивности* (см. [47]) всё ещё остаётся открытой проблемой.

Субаддитивность имеет место и для граничного ранга, но в этом случае мы знаем определённо, что аддитивности в общем случае нет. Уже для специального класса тензоров $\langle k, m, n\rangle$ левая и правая части (4.5) (с заменой rk на brk) могут отличаться сколь угодно сильно. Приведём пример из [36], который сыграл важную роль при нахождении оценки показателя матричного умножения. По соображениям размерности, $\mathrm{brk}\langle 3, 1, 3\rangle = 9$ и $\mathrm{brk}\langle 1, 4, 1\rangle = 4$, но

$$\mathrm{brk}\,(\langle 3,\ 1,\ 3\rangle \oplus \langle 1,\ 4,\ 1\rangle) = 10. \quad (4.6)$$

Оказывается, что трилинейная форма $\mathrm{tr}\,(ABC) + \mathrm{tr}\,(UVW)$, ассоциированная с прямой суммой этих дизъюнктных проблем, и в самом деле обладает приближённым разложением (4.2) длины 10 с коэффициентами из $F((\varepsilon))$ и с ошибкой не более $O(\varepsilon)$, а именно:

$$\sum_{i=1}^{3} \sum_{j=1}^{3} (a_i + \varepsilon u_{i,\,j})(b_j + \varepsilon v_{i,\,j})(c_{j,\,i} + \varepsilon^{-2} w) - \Big(\sum_i a_i\Big)\Big(\sum_j b_j\Big)\varepsilon^{-2} w =$$
$$= \sum_{i=1}^{3} \sum_{j=1}^{3} (a_i b_j c_{j,\,i} + u_{i,\,j} v_{i,\,j} w) + O(\varepsilon),$$

где u и v с индексами $(i, j) = (1, 1), (1, 2), (2, 1), (2, 2)$ являются переменными компонентами двух векторов длины 4, а остальные u и v подбираются так, чтобы обеспечить надлежащее взаимное сокращение, а именно

$$u_{i,\,3} = v_{3,\,j} = 0, \quad u_{3,\,j} = -u_{1,\,j} - u_{2,\,j}, \quad v_{i,\,3} = -v_{i,\,1} - v_{i,\,2}.$$

Аналогично тому как оценка Штрассена $\mathrm{rk}\langle 2, 2, 2\rangle \leqslant 7$ в сопоставлении с тривиальной оценкой (8 умножений) приводит к оценке $\omega \leqslant 3 \log 7/\log 8$, равенство (4.6) можно рассматривать как оценку 10 на граничный ранг для задачи *частичного* умножения матриц с тривиальной оценкой 13 на число умножений, а это уже влечёт $\omega \leqslant 3 \log 10/\log 13 < 2.7$, если воспользоваться соответствующим обобщением импликации (4.4b) [36, теорема 4.1]. Между тем имеется ещё более продуктивный способ оценивания ω на основе таких примеров, в котором дополнительно используется *дизъюнктность* частей. Таким более сильным обобщением утверждения (4.4b) является следующая

τ-теорема (название, данное де Грооте для одного специального случая теоремы 7.1 из [36]).

$$\mathrm{brk}\,(\langle k_1,\ m_1,\ n_1\rangle \oplus \ \ldots\ \oplus \langle k_p,\ m_p,\ n_p\rangle) \leqslant r = \sum_{i=1}^{p} (k_i m_i n_i)^{\tau} \Rightarrow \omega \leqslant 3\tau.$$

Применительно, например, к (4.6) решаем уравнение $9^{\tau} + 4^{\tau} = 10$ и получаем улучшение оценки: $\omega < 2.6$.

Сообщение об этой теореме на конференции в Обервольфахе (1979 г.) инициировало энергичное соревнование между специалистами, которые стремились найти всё лучшие и лучшие оценки на основе всё более и более изощрённых схем, аналогичных (4.6), и оценки для ω стали подходить ближе и ближе к 2.5. Предположения о такой нижней оценке были, однако, очень скоро опровергнуты Копперсмитом и Виноградом (см. [10]). Они нашли общий метод итеративного построения таких примеров и, отправляясь от (4.6), добились оценки $\omega < 2.5$. Более того, важным теоретическим следствием их изысканий явилось утверждение, что любой конкретный случай τ-теоремы допускает некоторое дальнейшее улучшение, что влечёт чуть-чуть лучшую оценку для ω, т. е. мы всегда имеем строгое неравенство $\omega < 3\tau$.

Новый метод Штрассена [51] основан на глубоком исследовании понятий ранга и граничного ранга в более общей алгебраической трактовке. Его конструкции позволили применить τ-теорему в сочетании с некоторым предельным процессом, что дало оценку $\omega < \log((h+1)^3/4)\log h$ (для $h = 2, 3, \ldots$). Наилучшая оценка достигается при $h = 5$, а именно $\omega < \log 54/\log 5 < 2.4785$.

4.3. Линейные системы и обращение матриц. Нахождение решения общей системы n линейных уравнений с n неизвестными $Ax = b$ можно трактовать как задачу вычисления компонент вектора x как рациональных функций от переменных элементов матрицы A и вектора b. Оптимальные алгоритмы для решения этой задачи являются неветвящимися программами в $F(a \ldots, b \ldots)$ минимальной длины, и именно относительно этой модели сначала было показано, что гауссов метод исключения не оптимален. (Кстати, для $n = 2$ гауссов метод исключения всё-таки *оптимален* — в смысле нескалярной сложности; см. [28]). Решение нашей линейной системы можно записать просто как $x = A^{-1}b$. Поэтому возможно решение такой общей системы с верхней оценкой сложности $I(n) + 2n^2 - n$, где $I(n)$ обозначает минимальное число арифметических операций, достаточное для вычисления матрицы, обратной к общей $n \times n$-матрице. В [46] для того, чтобы свести обращение матрицы к (быстрому) умножению матриц, было рекурсивно применено 2×2-*блочное обращение*. Соответствующая оценка имеет вид

$$I(2n) \leqslant 2I(n) + 6L(n) + n^2 + n.$$

Аналогичные соображения показывают, что сложность $D(n)$ вычисления $n \times n$-определителей удовлетворяет рекурсивному неравенству

$$D(2n) \leqslant I(n) + 2L(n) + 2D(n) + n^2 + 1.$$

Поэтому из оценки $L(n) = O(n^{\beta})$ (для любого $\beta > \omega$) следует, что $I(n)$ и $D(n)$ имеют один и тот же порядок роста. Обратно, можно показать, что сложность обращения матриц ненамного меньше, чем сложность умножения матриц. Однако, когда речь идёт о таких доказательствах, нужно иметь в виду, что неветвящиеся программы, годные для модели с переменными, как правило, непригодны для отдельных конкретных невырожденных матриц, поскольку подстановка алгебраически зависимых элементов может привести к делению на нуль. Ниже мы обсудим другие модели, у которых нет этого недостатка.

Равенство $X^2 = (X^{-1} - (I + X)^{-1})^{-1} - X$ показывает, что сложность $Q(n)$ *возведения в квадрат* общей $n \times n$-матрицы удовлетворяет неравенству $Q(n) \leqslant 3I(n) + 2n^2 + n$. Умножение общих $2n \times 2n$-матриц

$$\begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}\begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix},$$

состоящих из $n \times n$-блоков A_i, B_j, сводится к возведению в квадрат следующим образом. Полагая $S = B_1 + B_2$, мы можем получить первые n столбцов результата из соотношения

$$\begin{pmatrix} A_3 - S & B_1 \\ A_1 & S \end{pmatrix}^2 - \begin{pmatrix} A_4 - S & -B_2 \\ A_2 & S \end{pmatrix}^2 = \begin{pmatrix} \ldots\ldots & A_3B_1 + A_4B_2 \\ \ldots\ldots & A_1B_1 + A_2B_2 \end{pmatrix},$$

и аналогично для второй половины матрицы. Это показывает, что $L(2n) \leqslant 4Q(2n) + 12n^2$, откуда $L(2n) \leqslant 12I(2n) + O(n^2)$. Для бесконечных полей F похожую оценку (с другими константами) можно найти в [47].

Возможно также свести общую задачу обращения матриц к вычислению *определителей*. Такое сведéние было проведено Бауром и Штрассеном, которые доказали следующий более общий результат (см. [2]):

(4.7) Если функция $f \in F(x_1, x_2, \ldots, x_k)$ вычислима неветвящейся программой длины s, то сложность вычисления этой функции вместе со всеми её частными производными $\partial f/\partial x_j$ ограничена величиной $4s$.

Ещё одной иллюстрацией использования этой техники сведéния служит следующее короткое доказательство (неопубликованного) результата К. Калоркоти, гласящего, что вычисление *следа обратной матрицы* столь же трудоёмко, как и умножение матриц, с точностью до мультипликативной константы (для конечных полей), и тем самым имеет по существу ту же сложность, что и вычисление самóй обратной матрицы. Для неветвящейся программы длины s, вычисляющей $\mathrm{tr}(X^{-1})$, достаточно сделать $4s$ шагов, чтобы получить X^{-2} (с использованием частных производных); применив это ещё раз, получим верхнюю оценку $8s$ для сложности вычисления X^4. После удаления делений из такой программы (что обойдётся в ещё одну мультипликативную константу, см. [47]) мы можем взять в качестве X подходящую $3n \times 3n$-матрицу, такую что вычисление X^4 смоделирует умножение общих матриц размера $n \times n$.

До сих пор открыт вопрос, является ли исходная проблема решения $n \times n$-систем, т. е. проблема вычисления $A^{-1}b$ для произвольных A и b, достаточно мощной для моделирования умножения матриц или любой другой эквивалентной проблемы. Тем не менее совсем недавно Т. Ликтейгу удалось показать, что более общая задача вычисления *всех* решений *прямоугольной* системы $Ax = b$, которую можно понимать, скажем, как задачу нахождения базиса ядра общей $n \times 2n$-матрицы, и в самом деле имеет тот же порядок сложности, что и умножение матриц.

До сих пор мы рассматривали лишь задачи с алгебраически независимыми входами, но с точки зрения приложений более предпочтительными представляются другие модели. В случае простых полей и их конечных расширений более адекватной моделью для рассмотрения процесса решения линейных систем, вычисления ранга матриц или же вычисления обратной матрицы (если таковая существует) и т. д. являются *деревья вычислений*, в которых разветвления определяются проверкой на равенство нулю. Большинство сведений, упомянутых ранее, применимы по аналогии и к этой модели. По поводу дальнейших приложений отсылаем к [17] — в том числе и по поводу доказательства того факта, что алгебраическая сложность вычисления коэффициентов *характеристического полинома* имеет тот же порядок, что и сложность умножения матриц, с точностью до логарифмических множителей.

При возможности использовать проверку на равенство нулю предстаёт в новом свете открытая проблема, может ли сложность решения системы $Ax = b$ быть значительно меньше сложности умножения матриц. Соответствующую проблему *проверки* данного решения можно, очевидно, решить за $O(n^2)$ шагов. Также интересен вопрос, можно ли проверить правильность матричного произведения, т. е. равенство $AB = C$ при данном C, быстрее, чем вычислить произведение.

Предыдущая модель, использующая проверку на равенство нулю, неприменима к полям $\mathbf{R}$ или $\mathbf{C}$ с оракульными входами (см. п. 2), ибо там равенство алгоритмически неразрешимо. Например, невозможно вычислить ранг произвольной матрицы или выяснить, равен ли нулю определитель данной $n \times n$-матрицы A. Но если допустить регулярность A, то можно применить быстрые алгебраические методы умножения матриц для приближённого вычисления обратной матрицы. Привлечение эрмитово-сопряжённых матриц дает $A^{-1} = (A^H A)^{-1} A^H$, а применительно к обращению положительно-определённых матриц блочный метод исключения Штрассена (см. [46]) устойчив (в предположении что используемый метод умножения матриц не содержит делений). Время работы таких машинных алгоритмов будет, конечно, зависеть от требуемой точности и от обусловленности матрицы A.

Обсуждая методы приближённого решения $Ax = b$, следует должным образом учитывать, известна ли какая-нибудь априорная оценка на обусловленность A или нет. В последнем случае довольно правдоподобно, что основная часть времени уйдёт на получение информации об обусловленности, позволяющей гарантировать определённую точность решения.

4.4. *Характеристические уравнения.* В заключение нам хотелось бы дать краткий набросок того, как, комбинируя оценки времени, потребного на нахождение корней (см. п. 3), с описанными выше результатами об алгебраической сложности умножения матриц, получить хорошую верхнюю оценку сложности приближённого вычисления *собственных чисел* произвольной комплексной $m \times m$-матрицы. Это — лишь самое начало исследований широкого круга смежных проблем, таких как определение инвариантных подпространств или получение улучшенных оценок времени для специальных классов матриц и т. д., — исследований, которые ещё только надлежит выполнить.

Пусть A — произвольная $m \times m$-матрица, имеющая операторную l^1-норму $|A| \leqslant 1$ (или $\leqslant 1/2m$, после надлежащего масштабирования). Предположим, что последующие вычисления выполняются с N-битовой точностью. Мы опишем метод вычисления коэффициентов характеристического полинома матрицы A, который численно имитирует алгебраический подход, предложенный в [17]. Пусть ω обозначает показатель матричного умножения над полями характеристики нуль, и пусть β — произвольное число, такое что $\beta > \omega$. Умножение комплексных $m \times m$-матриц с N-битовой точностью возможно на машине $m \times m$-матриц с N-битовой точностью возможно на машине с указателями за время $O(m^\beta N)$. Первым шагом в вычислении характеристического полинома является приближённое *унитарное* преобразование матрицы A в *верхнюю* (правую) *форму Хессенберга* B ($b_{j+d,j} = 0$ при $d \geqslant 2$). С помощью методов из [32, 33] это можно сделать на машине с указателями за время $O(m^\beta N)$. Аналогично оценке (3.6), возмущения собственных чисел, вызванные ошибками $|B - U^H A U| < \varepsilon$, ограничены в худшем случае величиной $O(\varepsilon^{1/m})$.

Следующим шагом является *масштабирование* при помощи аналогичного преобразования с диагональной матрицей D, такой что в матрице $M = D^{-1}BD$ на поддиагонали стоят только единицы (не умаляя общности, можно считать, что все поддиагональные элементы матрицы B ненулевые, иначе задача расщепляется на несколько меньших задач о собственных значениях). Таким образом, мы получаем $M = S + R$, где R обозначает верхний треугольник матрицы M, а поддиагональ S представляет оператор сдвига, который переводит j-й единичный вектор e_j в вектор e_{j+1} $(Se_m = 0)$. Новые элементы

$$r_{k,\,k+d} = b_{k,\,k+d} \sum_{0 \leqslant j < d} b_{k+j+1,\,k+j} \quad (1 \leqslant k \leqslant k + d \leqslant m)$$

можно вычислить устойчивым способом с помощью $O(m^2)$ умножений комплексных чисел, модуль которых не превосходит 1. Далее, $|R| \leqslant |B| \leqslant 1/2m$. Характеристический полином $f(z) = v_0 + v_1 z + \ldots + v_{m-1}z^{m-1} + z^m$ матрицы M удовлетворяет условию $f(M) = 0$. Будучи применено к первому единичному вектору e_1, это уравнение даёт $v_0x_1 + v_1x_2 + \ldots + v_{m-1}x_m = -x_{m+1}$, где $x_j = M^{j-1}e_1$. Поскольку S доминирует над R при $|R| \leqslant 1/2m$, легко получается оценка $|x_j - e_j| < e^{0.5} - 1 < 0.65$. Поэтому матрица X, построенная из векторов-столбцов $x_1, \ldots, x_m$, удовлетворяет неравенству $|X - I| < 0.65$, а обратная к ней матрица ограничена: $|X^{-1}| < 3$, откуда следует, что вектор коэффициентов v можно получить как $v = X^{-1}(-x_{m+1})$, что занимает при численной реализации на машине с указателями время $O(m^\beta N)$. Основная идея работы [17] касается эффективного вычисления итераций x_j при $j \leqslant m + 1$ с помощью $O(\log m)$ матричных умножений: сначала формируются M^2, M^4, ..., а затем $M^q(x, \ldots, x_q) = (x_{q+1}, \ldots, x_{2q})$ для $q = 1, 2, 4, 8, \ldots$, по рекурсивной схеме.

В целом мы видим, что на машине с указателями времени $O(m^\beta N)$ достаточно для приближённого вычисления характеристического полинома матрицы A, при котором возмущения собственных чисел ограничены величиной $O(2^{-N/m})$. Поэтому заключительное применение утверждения (3.7) даёт следующий результат (ср. [38]):

Теорема. *Приближённое нахождение собственных чисел произвольной комплексной $m \times m$-матрицы (имеющей норму $\leqslant 1$) с ошибкой 2^{-n} возможно за время*

$$O(m^\beta mn) + O(m^3 \log m + m^2 mn) \quad (\text{для любого } \beta > \omega). \quad (4.8)$$

Сравнивая два члена в этой оценке времени, мы видим, что (даже если β близко к 2) нахождение собственных чисел как нулей характеристического полинома всегда быстрее, чем вычисление коэффициентов этого полинома, хотя последнее не намного дороже умножения матриц, если говорить о показателях.

Множитель mn соответствует порядку точности N, требуемой при анализе худшего случая. На практике, однако, сначала естественно попытаться добиться успеха с меньшей точностью, надеясь на лучшие апостериорные оценки, которые, возможно, и оправдают эту малую точность.

ЛИТЕРАТУРА

1. A. Alder and V. Strassen, On the algorithmic complexity of associative algebras, Theoret. Comput. Sci. **15** (1981), 201—211.

2. W. Baur and V. Strassen, The complexity of partial derivatives, Theoret. Comput. Sci. **22** (1983), 317—330. [Имеется перевод: Баур В., Штрассен Ф. Сложность частных производных. — Киберн. сб., н. с. — М.: Мир, 1985, с. 3—18.]

3. D. Bini, M. Capovani, G. Lotti, and F. Romani, $O(n^{2.7799})$ complexity for matrix multiplication, Inform. Process Lett. **8** (1979), 234—235.

4. A. Borodin and I. Munro, Computational complexity of algebraic and numeric problems, American Elsevier, New York, 1975.

5. A. L. Chistov and D. Yu. Grigoryev, Polynomial-time factoring of the multivariable polynomials over a global field, LOMI preprint E-5-82, Leningrad, 1982.

6. —, Subexponential-time solving systems of algebraic equations. I, II, LOMI preprints E-9-83, E-10-83, Leningrad, 1983.

7. T. J. Chou and G. E. Collins, Algorithms for the solution of systems of linear diophantine equations, SIAM J. Comput. **11** (1982), 687—708.

8. G. E. Collins and R. Loos, Real zeros of polynomials, Computing, Suppl. **4** (1982), 83—94.

9. S. Cook, A taxonomy of problems with fast parallel algorithms, Inform. and Control **64** (1985), 2—22.

10. D. Coppersmith and S. Winograd, On the asymptotic complexity of matrix multiplication, SIAM J. Comput. **11** (1982), 472—492.

11. B. Dejon and P. Henrici (Editors), Constructive aspects of the fundamental theorem of algebra (Proc. Sympos., Zürich-Rüschlikon, 1967), Wiley, London, 1969.

12. L. M. Delves and J. N. Lyness, A numerical method for locating zeros of an analytic function, Math. Comp. **21** (1967), 543—560.

13. J. von zur Gathen, Hensel and Newton methods in valuation rings, Math. Comp. **42** (1984), 637—661.

14. P. Henrici, Applied and computational complex analysis, vol. 1, Wiley-Interscience, New York, 1974.

15. E. Kaltofen, Polynomial-time reductions from multivariate to bi- and univariate integral polynomial factorization, SIAM J. Comput. **14** (1985), 469—489.

16. R. Kannan, Solving systems of linear equations over polynomials, Theor. Comput. Sci. **39** (1985), 69—88.

17. W. Keller-Gehrig, Fast algorithms for the characteristic polynomial, Theor. Comput. Sci. **36** (1985), 309—317.

18. D. E. Knuth, The art of computer programming, Vol. 2, Seminumerical Algorithms, 2nd ed., Addison-Wesley, Reading, Mass., 1981. [Имеется перевод 1-го изд.: Кнут Д. Искусство программирования для ЭВМ. Т. 2. Получисленные алгоритмы. — М.: Мир, 1977.]

19. H. T. Kung, On computing reciprocals of power series, Numer. Math. **22** (1974), 341—348.

20. H. T. Kung and J. F. Traub, All algebraic functions can be computed fast, J. Assoc. Comput. Mach. **25** (1978), 245—260.

21. S. Landau, Factoring polynomials over algebraic number fields, SIAM J. Comput. **14** (1985), 184—195.

22. Landau and G. L. Miller, Solvability by radicals is in polynomial time, Proc. 15th Annual ACM Sympos. on Theory of Comp., Ass. Comp. Mach., New York, 1983, pp. 140—151.

23. D. Lazard, Résolution des systèmes d'équations algébriques, Theor. Comput. Sci. **15** (1981), 77—110.

24. A. K. Lenstra, Factoring multivariate integral polynomials, Theor. Comput. Sci. **34** (1984), 207—213.

25. —, Polynomial-time algorithms for the factorization of polynomials, Ph. D. Thesis, Univ. of Amsterdam, 1984.

26. A. K. Lenstra, H. W. Lenstra, and L. Lovász, Factoring polynomials with rational coefficients, Math. Ann. **261** (1982), 515—534.

27. T. Lickteig, The computational complexity of division in quadratic extension fields, SIAM J. Comput. **16** (1987), 278—311.

28. —, Gaussian elimination is optimal for solving linear equations in dimension two, Inform. Process Lett. **22** (1986), 277—279.

29. A. Ostrowski, Solution of equations and systems of equatons, 2nd ed., Pure and Applied Mathematics, vol. 9, Academic Press, New York, 1966.

30. V. Ya. Pan, How can we speed up matrix multiplication?, SIAM Rev. **26** (1984), 393—415.

31. A. Schönhage, Schnelle Berechnung von Kettenbruchentwicklungen, Acta Inform. **1** (1971), 139—144.

32. —, Unitäre Transformationen grosser Matrizen, Numer. Math. **20** (1973), 409—417.

33. —, Fast Schmidt orthogonalization and unitary transformations of large matrices, Complexity of Sequential and Parallel Numerical Algorithms (J. F. Traub, ed.), Academic Press, New York, 1973, pp. 283—291.

34. —, Schnelle Multiplikation von Polynomen über Körpern der Charakteristik 2, Acta Inform. **7** (1977), 395—398.

35. —, Storage modification machines, SIAM J. Comput. **9** (1980), 490—508.

36. —, Partial and total matrix multiplication, SIAM J. Comput. **10** (1981), 434—455.

37. —, Asymptotically fast algorithms for the numerical multiplication and division of polynomials with complex coefficients, Computer Algebra (Marseille, 1982), Lecture Notes in Comput. Sci., Vol. 144, Springer, Berlin — New York, 1982, pp. 3—15.

38. —, The fundamental theorem of algebra in terms of computational complexity, Technical Report, Univ. Tübingen, 1982, 74 pp.

39. —, Factorization of univariate integer polynomials by diophantine approximation and an improved basis reduction algorithm, Automata, Languages and Programming (ICALP, Antwerp, 1984), Lecture Notes in Comput. Sci., Vol. 172, Springer, Berlin — New York, 1984, pp. 436—447.

40. —, Quasi-GCD computations, J. of Complexity **1** (1985), 118—137.

41. A. Schönhage and V. Strassen, Schnelle Multiplikation grosser Zahlen, Computing **7** (1971), 281—292. [Имеется перевод: Шёнхаге А., Штрассен Ф. Быстрое умножение больших чисел. — Киберн. сб., н. с. — М.: Мир, 1973, с. 87—98.]

42. M. Sieveking, An algorithm for division of power series, Computing **10** (1972), 153—156.

43. S. Smale, The fundamental theorem of algebra and complexity theory, Bull. Amer. Math. Soc. (N. S.) **4** (1981), 1—36.

44. —, On the efficiency of algorithms of analysis, Bull. Amer. Math. Soc. (N. S.) **13** (1985), 87—121.

45. E. Specker and V. Strassen, Komplexität von Entscheidungsproblemen, Lecture Notes in Comput. Sci., Vol. 43, Springer, Berlin — New York, 1976.

46. V. Strassen, Gaussian elimination is not optimal, Numer. Math. **13** (1969), 354—356. [Имеется перевод: Штрассен Ф. Алгоритм Гаусса не оптимален. — Киберн. сб., н. с. — М.: Мир, 1970, с. 67—70.]

47. —, Vermeidung von Divisionen, J. Reine Angew Math. **264** (1973), 184—202.

48. —, Die Berechnungskomplexität von elementarsymmetrischen Funktionen und von Interpolationskoeffizienten, Numer. Math. **20** (1973), 238—251. [Имеется перевод: Штрассен Ф. Сложность вычисления элементарных симметрических функций и коэффициентов интерполяционного полинома. — Киберн. сб., н. с. — М.: Мир, 1978, с. 5—21.]

49. —, The computational complexity of continued fractions, SIAM J. Comput. **12** (1983), 1—27.

50. —, Algebraische Berechnungskomplexität, Perspectives in Mathematics, Anniversary of Oberwolfach 1984 (W. Jäger, J. Moser, and R. Remmert, eds.), Birkhäuser, Basel, 1984, pp. 509—550.

51. —, Relative bilinear complexity and matrix multiplication, Manuscript, Universität Zürich, 1986.

52. P. Turán, Power sum method and an approximative solution of algebraic equations, Math. Comp. **29** (1975), 311—318.

53. H. Weyl, Randbemerkungen zu Hauptproblemen der Mathematik. II, Fundamentalsatz der Algebra und Grundlagen der Mathematik, Math. Z. **20** (1924), 131—150.

ТЮБИНГЕНСКИЙ УНИВЕРСИТЕТ, D-7400 ТЮБИНГЕН, ФРГ

СИСТЕМАТИКА УНИВЕРСАЛЬНЫХ И ПРОЧИХ КЛАССОВ

Сахарон Шелах[1]

1. ПРОБЛЕМАТИКА

Математическая наука привлекла меня своей общностью и способностью внести порядок туда, где царит хаос, а не возможностью прояснять с её помощью детали там, где в принципе и так всё ясно. Поэтому вряд ли покажется удивительным сделанный мною, ещё когда я приступал к работе над диссертацией, выбор темы — основной для моих интересов в математике. (Мы рассматриваем «универсальные классы» лишь как пример; изучаемое нами понятие не связано с логикой, см. определение.)

1.1. Первая проблема: проблема таксономии = проблема классификации для универсальных классов. Требуется указать главные разделяющие линии в семействе универсальных классов; каждая линия должна выражать нечто существенное в том смысле, что по одну сторону от неё расположены классы, являющиеся в определённом смысле «простыми» или «поддающимися анализу», а по другую сторону — сложные или не поддающиеся анализу.

1.2. Универсальные классы. (1). Примерами служат: класс всех групп, класс колец или любое многообразие колец, класс локально-конечных групп.

(2) Вообще, пусть τ — словарь, содержащий список функциональных символов и предикатов (т. е. символов отношений), каждый символ — с соответствующей «арностью» ($n(F)$ или $n(R)$[2]); τ-структура M образуется непустым множеством $|M|$ (универсум структуры), в котором интерпретированы все функциональные символы и предикаты из τ, т. е. функциональный символ F интерпретируется как $n(F)$-местная функция из $|M|$ в $|M|$, а символ отношения R — как $n(R)$-местное отношение на $|M|$.

(3) *Универсальным классом* называется всякий класс K, составленный из τ-структур для соответствующего $\tau = \tau(K)$ и обладающий тем свойством, что какая-либо τ-структура M принадлежит K в том и только в том случае, когда каждая её конечно-порождённая подструктура принадлежит K.

Я люблю в математике ещё и то, что у решившего задачу (обычно) не возникает сомнений относительно того, действительно ли она решена. Достаточно найти правильное решение, а способность довести его до сознания других играет второстепенную роль. Мне нравятся точно поставленные задачи, желательно предполагающие ответ «да» или «нет», и я считаю, что правильный подход к ускользающим от понимания проблемам заключается в выработке точной постановки вопроса. Поэтому нужно уточнить и нашу задачу.

1.3. Вторая проблема: проблема структурности — бесструктурности. (1) Построить для некоторой совокупности объектов K (например, универсальных классов) структурную теорию (см. ниже) и доказать теоремы бесструктурности для прочих объектов такого же типа, т. е. продемонстрировать для них невозможность структурной теории.

(2) *Структурной теорией для* K является теория, снабжающая каждую структуру $M \in K$ полным множеством её инвариантов; каждый инвариант должен вполне определяться типом изоморфности структуры M, и обратно, M должна определяться с точностью до изоморфности полным набором своих инвариантов.

Разумеется, инварианты не должны быть слишком сложными; например, нет смысла брать в качестве инварианта сам тип изоморфности. Конструкция инвариантов структуры и, обратно, воссоздание структуры по набору инвариантов должны быть достаточно ясными. Однако мы не будем здесь обсуждать эти тонкости. Подробнее см. [Sh 86].

Какие, однако, объекты могли бы служить инвариантами структур заданной мощности λ? Знаменитая теорема Штейница для алгебраически замкнутых полей фиксированной характеристики — прототип структурных теорем — предлагает в этом качестве мощность (= трансцендентную размерность). Заметим, что если каждый класс K_i, $i \in I$, имеет структурную теорию, то и «сумма»

$$\sum_{i \in I} K_i \overset{\mathrm{df}}{=} \left\{ \sum_{i \in I} M_i : M_i \in K_i \text{ для } i \in I \right\}$$

---

[1] Saharon Shelah, Taxonomy of universal and other classes, ICM86, pp. 154—162.

[2] Для функциональных символов и символов отношений соответственно. — *Прим. изд. ред.*

(с любым разумным определением $\sum_i M_i$, при котором $\sum_i K_i$ является достаточно хорошим классом) также обладает структурной теорией. Аналогично если мы имеем структурную теорию для класса K, то таковая возможна и для класса

$$\sum K \overset{\mathrm{df}}{=} \left\{ \sum_{i \in I} M_i : I \text{ — произвольное множество и } M_i \in K \text{ для всех } i \in I \right\}.$$

Мы приходим к тому, что в качестве инвариантов для структур мощности λ надо взять λ-значения рода (α, χ), определяемые так:

1.4. Определение. (1) Пусть λ и χ — кардиналы, а α — ординал. Понятие *λ-значения рода* (α, χ) вводится индукцией по α:

$\alpha = 0$. λ-значение рода $(0, \chi)$ есть кардинал $\leqslant \lambda$.

$\alpha = \beta + 1$. λ-значением рода $(\beta + 1, \chi)$ может быть либо λ-значение рода (β, χ), либо последовательность длины χ, каждый член которой есть функция из множества

$$\{x : x \text{ есть } \lambda\text{-значение рода } (\beta, \chi)\}$$

в множество кардиналов $\leqslant \lambda$.

α предельное. λ-значением рода (α, χ) является любое λ-значение рода (β, χ), $\beta < \alpha$.

(2) Скажем, что инвариант рода $(\beta, 1)$ имеет глубину β.

(3) *Инвариантом рода* (α, χ) для класса K является функция, сопоставляющая каждой структуре $M \in K$ мощности λ некоторое λ-значение рода (α, χ), вполне определяемое типом изоморфности M.

Теперь, если мы не ограничиваем α, то даже при $\chi = 1$ любая структура мощности λ может быть закодирована с точностью до изоморфизма, поскольку для любой пары α, χ найдется β, такое что любое λ-значение рода (α, χ) может быть естественным образом закодировано подходящим λ-значением глубины β (т. е. рода $(\beta, 1)$). Это даёт основание сформулировать

1.5. Первый тезис. Класс K обладает структурной теорией в том и только в том случае, когда для некоторого β имеется набор инвариантов глубины β для K, однозначно классифицирующий структуры $M \in K$ с точностью до изоморфизма.

Нам представляется весьма сильным положение о том, что полный набор инвариантов дает структурную теорию. Например, справедливо

1.6. Утверждение. Если класс K обладает структурной теорией в силу тезиса 1.5, посредством какого-либо инварианта рода (α, χ), то для любого кардинала $\aleph_\gamma$ выполнено неравенство $I(\aleph_\gamma, K) \leqslant \beth_\alpha(|\gamma| + \chi)$. Следовательно, если дополнительно предполагается справедливой обобщённая континуум-гипотеза (GCH), то $I(\aleph_\gamma, K) \leqslant (|\gamma| + \chi)^{+\beta}$.

1.7. Определение. (1) $|\gamma|$ обозначает мощность ординала γ.

(2) $I(\aleph_\gamma, K)$ обозначает число всех типов изоморфности структур из K, имеющих мощность $\aleph_\gamma$.

(3) GCH есть утверждение о том, что $2^{\aleph_\gamma} = \aleph_{\gamma+1}$ для каждого индекса γ.

4) $(\aleph_\alpha)^{+\beta} = \aleph_{\alpha+\beta}$, а кардиналы $\beth_\alpha(\lambda)$ определяются при помощи индукции по α равенством

$$\beth_\alpha(\lambda) = \aleph_0 + \sum_{\beta<\alpha} 2^{\beth_\beta(\lambda)}.$$

(Отметим, что даже при $\alpha = 2$ и $\chi = 1$ может быть выполнено неравенство $\beth_\alpha(|\gamma| + \chi) > 2^{\aleph_\gamma}$.)

Функция $I(\lambda, K)$ привлекает внимание ввиду большой важности проблемы определения числа типов изоморфности структур из рассматриваемого класса K в данной мощности. Обычно простым классам K соответствуют малые значения $I(\lambda, K)$. Однако мне представляется, что действительно существенную роль играет следующее положение (а не точное вычисление значений $I(\lambda, K)$):

1.8. Второй тезис: (первая) главная разделяющая линия = главный разрыв. Для достаточно «хороших» семейств классов (как например, для семейства универсальных классов), разделяющая линия, основанная на принципе «найдётся α, такое что $I(\aleph_\gamma, K) \leqslant \beth_\alpha(|\gamma|)$ для всех γ», является достаточно разумной. А именно:

(a) Она коррелирует со свойством «иметь структурную теорию в соответствии с тезисом 1.5».

(b) Классы по «сложную» сторону от этой линии вряд ли могут обладать структурной теорией из-за несоответствия в оценках для $I(\lambda, K)$. Точнее, имеется большой разрыв между верхней оценкой 1.5 и следующей нижней оценкой для $I(\lambda, K)$:

(1) $I(\lambda, K) = 2^\lambda$ для достаточно больших λ (это вообще — наибольшее возможное значение при $\lambda \geqslant |\tau(K)|$),

или по крайней мере

(2) $I(\lambda, K) > \lambda$ для достаточного числа кардиналов λ.

Мы предпочитаем доказывать бесструктурность в рамках теории ZFC, однако стоит отметить, что для демонстрации невозможности структурной теории вполне достаточно установить непротиворечивость гипотезы о бесструктурности.

(c) Мне представляется, что если удастся разрешить проблему (b), то мы сможем получить достаточно разветвлённую систематику, с достаточно богатым инструментарием, и узнаем много о классах по каждую сторону от соответствующих разделительных линий; иными словами, мы предлагаем 1.8(b) в качестве контрольного вопроса.

Уместно отметить, что отнюдь не каждая разделяющая линия действительно представляет интерес в плане классификации. Например, класс колец — весьма важная часть класса структур с двуместными функциями, однако вряд ли имеет смысл рассматривать дополнение класса колец в этом более широком классе. С другой стороны, разделяющие линии, помимо чисто классификационного назначения, могут быть полезны в доказательствах посредством разбора случаев.

1.9. Тезис. (d) Рассматривая какой-либо класс структур, имеет смысл особо обратить внимание на большие мощности, в которых исчезают все «сингулярности».

Отметим, что, к примеру, теории, имеющие единственную счётную модель, могут иметь много сложно устроенных моделей в несчётных мощностях. Отметим также, что следующий вопрос не нашёл пока ещё исчерпывающего ответа.

1.10. Вопрос. Существует ли достаточно обширное семейство классов, для которых тезис 1.8 может быть доказан?

Ниже мы вернемся к этому вопросу.

2. ПРЕДЫСТОРИЯ И ФОН

Почему мы вообще обращаемся здесь к универсальным классам? Дело в том, что в теории моделей наиболее простые классы структур имеют вид Mod(T), где T — счётная теория первого порядка.

2.1. Определение. Mod(T) есть класс всех моделей теории T. (Иногда для простоты мы пишем T вместо Mod(T).)

Разумеется, можно изучать и более сложные классы — несчётные теории, теории в рамках инфинитарной логики, языки с обобщёнными кванторами, универсальные теории и т. д.; всё это представляет несомненный интерес и не нуждается в дополнительных подтверждениях своей важности.

Обратимся теперь к предыстории рассматриваемых вопросов (читатель может не обращать внимания на понятия, с которыми он не знаком). Примерно в начале 60-х годов исследования в области математической логики приобрели новый уровень глубины и математической сложности. В то время желанной целью многих специалистов по теории моделей считалось продвижение к решению гипотезы Лося. Дадим её формулировку:

Гипотеза Лося. *Если счётная (первого порядка) теория T категорична в некоторой мощности $\lambda > \aleph_0$ (т. е. $I(\lambda, T) = 1$), то T категорична в каждой несчётной мощности.*

Разумеется, если мы желаем доказать что-либо новое относительно $I(\lambda, T)$, лучше всего начать именно с этой проблемы; было бы наивно браться за значительно более сложные задачи вроде 1.8(b), не заложив некоторый фундамент. Изыскания по гипотезе Лося нашли своё завершение в работе Морли [Mo], где было дано положительное её решение. Эта теорема, в доказательстве которой были использованы различные разработанные ранее (в частности, Вотом, Эренфойхтом, Мословским) технические средства, рассматривается многими (в том числе и автором этих строк) как одно из наиболее значительных достижений математической логики за 60-е годы.

Затем в течение какого-то времени почти ничего не происходило. На это были свои причины — и среди них не в последнюю очередь мнение Морли и Кейслера о том, что теория моделей для теорий первого порядка завершена или почти завершена (я слышал от них об этом в 1969 г.). Однако со временем эта область теории моделей вернула себе былую популярность. Появилась масса журнальных статей, вышли монография [Sh78] и носящие в основном обзорный характер книги Пиллея [Pi], Ласкара [La], Пуаза [Po], Болдуина [Ba]; несколько позже был организован ряд конференций. [О более ранней предыстории см. статью [Sh74], частично перекрывающуюся с [Sh85] (где речь идёт о счётных теориях первого порядка), а также введение к сборнику [Ba 1] (где теория представлена во всей общности).]

Нельзя, однако, не отметить, что бóльшая часть публикаций связана с несколько иной постановкой вопроса, которую можно назвать «тонкоструктурной», т. е. речь идёт о доскональном исследовании там, где принципиальные результаты уже получены, или об изучении семейств классов, обладающих некоторой структурной теорией, с целью уточнить какие-то моменты. В качестве примеров можно назвать теорему Болдуина — Лаклана (о счётной первого порядка теории T, категоричной в мощности $\aleph_1$, но не категоричной в мощности $\aleph_0$), работы Зилбера и Черлина, Харрингтона и Лаклана (о теории, категоричной в мощностях $\aleph_1$ и $\aleph_0$, или о тотально трансцендентной теории, категоричной в мощности $\aleph_0$). Я думаю, что некоторые из этих результатов получат своё отражение в сообщениях Лаклана и Перетятькина на этом конгрессе[1].

Всё же реальных оснований для противопоставления нет: решения проблемы 1.1 ведут или могут привести к тонкоструктурным исследованиям, снабжая их некоторым начальным инструментарием. Но я продолжу излагать свою собственную точку зрения. Начиная работать над проблемой 1.1 в 1967 г., я рассмотрел в качестве разделяющих линий свойства стабильности и суперстабильности и предположил в [Sh71, p. 283, (13)] (см. также [Sh74]), что аналогичную роль могут исполнить функции $I(\lambda, T)$ в области достаточно больших λ (T — счётная теория первого порядка); в частности, был рассмотрен «главный разрыв» 1.8(b). Первым классом, для которого наша концепция нашла подтверждение (с незначительными поправками), стала совокупность всех $\aleph_\varepsilon$-насыщенных моделей теории T первого порядка (результат анонсирован в [Sh74], см. также [Sh83; Sh83a; Sh77, гл. X, пример 3.3]), но, вводя понятие $\aleph_\varepsilon$-насыщенности, я чувствовал, что здесь не всё в порядке. Второй пример доставили универсальные теории первого порядка (см. [Sh86]), т. е. теории, аксиомы которых имеют вид

$$\forall x_1 \ldots \forall x_n \bigvee_i \bigwedge_j \varphi_{ij},$$

где каждая из формул φ_{ij} — атомарная или отрицание атомарной. Остановимся на нём подробнее:

2.2. Теорема. *Для любой универсальной первого порядка теории T может быть выполнена одна и только одна из трёх указанных ниже возможностей* (A) — (C):

(A) $I(\lambda, T) = 2^\lambda$ *для всех* $\lambda > |T|$ (*мы опускаем некоторые другие условия типа «сложности», см.* [Sh85]).

(B) *Снова* $I(\lambda, T) = 2^\lambda$ *для всех* $\lambda > |T| + \aleph_1$, *но при этом*

(*) *для любой модели M теории T мощности λ существует индексированное множество* $\langle M_\eta : \eta \in I \rangle$, *такое что*

---

[1] A. H. Lachlan, Homogenenous structures, ICM86, pp. 314—321; М. Г. Перетятькин, Конечно-аксиоматизируемые теории, *там же*, pp. 322—330 (сообщения на секции «Математическая логика и основания математики»). — *Прим. изд. ред.*

(i) *I есть непустое множество конечных последовательностей ординалов $<\lambda$, замкнутое относительно взятых начальных сегментов;*

(ii) *M_η — подмодель M мощности $\leqslant |T|$, а если ν является начальным сегментом η, то M_ν — подмодель M_η;*

(iii) *Модель M свободно порождается моделями M_η, $\eta \in I$, в том смысле, что*

(α) *замыкание объединения $\bigcup_{\eta \in I} |M_\eta|$ (где $|M_\eta|$ — универсум модели M_η) относительно функций из M совпадает с $|M|$;*

(β) *если $\eta \in I$ имеет длину $n+1$, $\nu = \eta \restriction n$, то для любой конечной последовательности $\bar{c}$ элементов $|M_\eta|$ и любого конечного множества Φ бескванторных формул с параметрами из совокупности*

$$\bigcup \{ |M_\rho| : \rho \in I \text{ и } \eta \text{ не является начальным сегментом } \rho \},$$

удовлетворяющихся на $\bar{c}$, найдётся конечная последовательность $\bar{c}'$ элементов $|M_\nu|$, на которой все формулы из Φ также удовлетворяются.

(C) *Выполняется требование* (*) *из* (B), *и, кроме того, найдётся ординал* $\mathrm{Dp}(T)$, *счётный, если T счётна, а в общем случае имеющий мощность $\leqslant 2^{|T|}$ и такой, что для любой модели M существует удовлетворяющее* (*) *индексированное семейство $\langle M_\eta : \eta \in I \rangle$ глубины $< \mathrm{Dp}(T)$. Последнее означает, что можно построить функцию $d : I \to \alpha$, обладающую тем свойством, что если ν — собственный начальный сегмент η, то $d(\nu) > d(\eta)$.*

Таким образом, $I(\aleph_\alpha, T) \leqslant \beth_{Dp(T)}(\alpha)$ для достаточно больших α (на самом деле в этом случае имеется более полная информация относительно значений функции $I(\lambda, T)$).

Первая разделяющая линия проходит между (A)+(B) и (C). Но заслуживает внимания и вторая, между (A) и (B)+(C) (см. [Sh85]); здесь, однако, нет возможности говорить об этом подробнее.

Теорема предлагает достаточно разумный ответ на вопрос 1.10 — обширное семейство, в котором мы имеем решение. Однако с теоретико-модельной точки зрения я не мог быть удовлетворенным, пока не получил решение для произвольных счётных теорий первого порядка (см. [Sh85], [Sh87]). Считая, что существо вопроса содержится в тезисе 1.8, я тем не менее чувствовал себя обязанным разобраться с доставшейся в наследство от предшествующего поколения логиков гипотезой

Морли. Если мы имеем разумный ответ на вопрос 1.10, то неизбежно встаёт еще один вопрос:

2.3. Вопрос. Является ли предлагаемая в качестве ответа теория теорией классов первого порядка, или же это в действительности некая совокупность таких теорий, которую, быть может, надлежит заключить в какие-то общие рамки?

Некоторые успехи были достигнуты в связи с другими семействами классов и несколько иной проблематикой. Болдуин и Шелах [BaSh] (см. также [Sh86a, Sh85a]), отправляясь от идей, высказанных в работе [Sh78], исследовали сложность класса $K = \mathrm{Mod}(T)$ для теорий T первого порядка в монадической логике, финитарной или инфинитарной.

Была получена достаточно полная (в некотором подходящем смысле) классификация. Эти результаты объяснили, почему значительная часть работ по монадической логике (скажем, исследования Бучи, Рабина, Шелаха и Гуревича, см. [Gu]) связаны с линейными порядками и деревьями. Изучены и некоторые новые случаи, а также обобщённые кванторы. Отсюда можно сделать вывод, что классификация, данная в [Sh78], распространяется на весьма широкую группу проблем, о которых первоначально речь и не шла. О классификации кванторов см. [Sh86c].

Рассматривая классы K двухуровневых структур, имеет смысл выяснить, до какой степени структура $M \in K$ может быть определена с точностью до изоморфности над первым уровнем. Примером может служить изучение структуры векторного пространства V над полем F (как V, так и F не фиксированы), где достаточно одного инварианта — размерности.

Таким образом, бесструктурность должна в этом случае означать, что для любого (или, по крайней мере любого достаточно большого) кардинала λ найдётся структура M_0, такая что для большого числа сложно устроенных $N \in K$ первый уровень N будет тождествен M_0 (с точностью до изоморфизма). По-видимому, для классов $K = \mathrm{Mod}(T)$ (где T — счётная теория первого порядка) можно достигнуть полного решения, *если* допустить формулировки с независимостью в «бесструктурной» части. Мы говорим «по-видимому», поскольку здесь еще не всё ясно, см. [PiSh, Sh87], а также лекционные записки [Sh85b]. Однако я уверен, что результат будет получен.

Для универсальных классов ситуация аналогична [Sh87a].

Гроссберг и Харт доказали теорему о главном разрыве и некоторые другие факты для семейства превосходных классов. Последние были введены в [Sh83c], где установлено, что если предложение $\psi \in L_{\omega_1\omega}$ имеет несчётную модель и не имеет большого числа (точнее, ни для какого $n > 0$ не имеет $2^{\aleph_n}$) неизоморфных моделей, то $\mathrm{Mod}(\psi)$ есть объединение нескольких превосходных классов.

3. ВЗАИМОДЕЙСТВИЕ С «ВНЕШНИМ МИРОМ»

В общем, наша теория разрешила почти все связанные с ней проблемы теории моделей 60-х годов и ряд проблем из таких областей, которые до того не казались имеющими к ней отношение (как, скажем, исследование упорядочения по Кейслеру теорий первого порядка).

Имеются любопытные следствия для универсальных алгебр. В частности, теорема 2.2 даёт возможные значения функции $I(\lambda, K)$, где K — многообразие, хотя мы и не знаем в точности, действительно ли могут быть реализованы все разрешённые значения параметров (например, может ли быть бесконечной глубина теории; также не всё ясно с небольшими значениями мощностей). К сожалению, алгебраисты, кажется, не торопятся изучать этот материал, хотя проблематика и результаты должны представлять для них интерес.

Мы занимаемся этой теорией ради её самой; приложения же рассматриваются исключительно как средство убедить «неверующих». Тем не менее мы искренне считаем, что она должна помочь и в исследовании конкретных классов.

Нельзя не отметить определённой асимметрии между «структурной» и «бесструктурной» сторонами. Развивать структурную теорию для конкретного класса можно, вообще не располагая никаким формальным понятием структурной теории. Но необходимость в таком понятии (или в чём-то подобном) неизбежно возникает, если мы желаем установить невозможность структурной теории.

Структурная теория, если она есть, делает естественным и успешным изучение конкретных классов, которые входят в данную классификацию. Исследования в этом направлении были открыты работой Макинтайра [Mc], где речь шла о теориях (первого порядка) полей в духе статьи [Mo]. В работах Черлина [Ch] и Черлина и Шелаха [ChSh] эти исследования были продолжены для полей и колец с делением. В результате получено такое

Следствие. Если T — теория первого порядка, а $\mathrm{Mod}(T)$ — некоторый класс бесконечных полей (или колец с делением), но не все алгебраически замкнутые поля, то T не является суперстабильной теорией; следовательно, T имеет большое число сложно устроенных моделей и модулей.

Имеется обширная литература по теориям (первого порядка) колец, модулей и групп. Весьма плодотворна теория T^p_{dcf} дифференциально замкнутых полей. Робинсон, основываясь на результатах Зейденберга [Se], доказал, что T^0_{dcf} — теория первого порядка. Ленор Блам [B] предложила для неё конкретную аксиоматику и установила, что эта теория тотально трансцендентна. Ей также удалось показать, опираясь на результаты Морли, что над любым дифференциальным полем F характеристики 0 можно построить простое дифференциально замкнутое поле, расширяющее F (т. е. такое, что его можно вложить в любое другое дифференциально замкнутое поле, расширяющее F). Мы установили, при помощи одной из теорем теории классификации (см. [Sh78, гл. IV, § 4]), что такое простое дифференциально замкнутое расширение единственно. Используя некоторые алгебраические факты, Вуд [W] доказал, что T^p_{dcf} — также теория первого порядка, однако не являющаяся тотально трансцендентной. Шелах [Sh] и Вуд [W1] независимо установили существование простого дифференциально замкнутого расширения для любого дифференциального поля. Шелах [Sh73] доказал стабильность теории T^p_{dcf}; тем самым, согласно теореме из [Sh78, гл. IV, § 5], упомянутое простое дифференциально замкнутое расширение единственно. Вместе с тем теория $T^p_{\mathrm{dcf}}(p > 0)$ не является суперстабильной [Sh73], и потому структурная теория для $\mathrm{Mod}(T^p_{\mathrm{dcf}})$ невозможна.

Используя существующую теорию для $\mathrm{Mod}(\psi)$ (где ψ — предложение инфинитарной логики $L_{\omega_1\omega}$), Меклер и Шелах [MSh] доказали (в предположении, что выполнена аксиома конструктивности $V = L$), что для любого многообразия либо $L_{\infty\omega}$-свободность влечёт свободность, либо для любого λ имеются λ-свободные, но не свободные многообразия (более ранние результаты в этом направлении были получены Эклофом и Меклером [EM]).

В статье Гроссберга и Шелаха [GSh] изложено опирающееся на одну доказанную там общую теорему решение проблемы Фукса и Зальце (см. [FS]) о возможности структурной теории кручения модулей над универсальными кольцами. С помощью этой теоремы можно также дать прямое доказательство одного старого результата из [Sh74] (решение проблемы Фукса [F]), утверждающего, что существует большое число сложно устроенных сепарабельных редуцированных абелевых p-групп в любой мощности $\lambda > \aleph_0$.

Понятно, что во многих случаях ценность для приложений представляют не теоремы структурной теории сами по себе, а методы, заложенные в их доказательствах. Мы попытались в [Sh83b] дать теоремы, пригодные для «общего пользования»; в качестве приложения там были построены булевы алгебры, обладающие достаточно необычными свойствами (например, не имеющие нетривиальных автоморфизмов или взаимно однозначных эндоморфизмов, полные и/или УСЦ).

Еще одна попытка приспособить теоремы для приложений была предпринята в [Sh85c]; эта статья нашла разнообразные применения в работах Корнера, Гобела и автора по представлениям колец как групп эндоморфизмов абелевых групп.

В целом мы полагаем, что описанный метод может быть полезен для построения структур особого типа, которые можно назвать «сложными», например структур, не обладающих нетривиальными автоморфизмами или эндоморфизмами, либо же неразложимых в том или ином смысле, и т. п.

ЛИТЕРАТУРА

[B] L. Blum, Generalized algebraic structures, a model theoretic approach, P. D. Thesis, M. I. T., Cambridge, Mass., 1968.

[Ba] J. Baldwin, Springer-Verlag.

[Ba1] J. Baldwin (ed.), Introduction, USA — Israel, Proc. Conference on Classification Theory (Chicago, December 1985), Lecture Notes in Math., Springer-Verlag, Berlin and New York (to appear).

[Ba2] J. Baldwin, Definable second order quantifiers, Model-Theoretic Logics (J. Barwise and S. Feferman, eds.), Springer-Verlag, Berlin and New York, 1985, pp. 445—478.

[BaSh] J. Baldwin and S. Shelah, Classification of theories by second order quantifiers, Proc. 1980/Jerusalem Model Theory Year, Notre Dame J. Formal Logic **26** (1985), 229—303.

[Ch] G. Cherlin, Superstable division rings, Logic Colloquium **77** (Proc. Conf., Wroclaw, 1977), North-Holland, Amsterdam, 1978, pp. 99—112.

[ChSh] G. Cherlin and S. Shelah, Superstable fields and groups, Ann. of Math. Logic **18** (1980), 227—280.

[EM] P. Eklof and A. Mekler, Categoricity results for $L_{\infty\varkappa}$-free algebras, Ann. Pure Appl. Logic (to appear).

[F] L. Fuchs, Infinite Abelian groups, Academic Press, 1970, 1973.

[FS] L. Fuchs and L. Salce, Modules over valuation domains, Lecture Notes in Pure Appl. Math., vol. 97, Marcel Dekker, New York, 1985.

[Gu] Y. Gurevich, Monadic second order theories, Model-Theoretic Logics (J. Barwise and S. Feferman, eds.), Springer-Verlag, Berlin and New York, 1985, pp. 479—506.

[GSh] R. Grossberg and S. Shelah, A nonstructure theorem for an infinitary theory which has the unsuperstability property, Illinois J. Math. **30** (1986), 364—390.

[K] E. R. Kolchin, Differential algebra and algebraic groups, Academic Press, New York, 1973.

[La] D. Lascar, Introduction to stability.

[Mc] A. Macintyre, On ω_1-categorical theories of fields, Fund. Math. **71** (1971), 1—25.

[Mo] M. D. Morley, Categoricity in power, Trans. Amer. Math. Soc. **114** (1965), 514—538.

[MSh] A. Mekler and S. Shelah, For which varieties $L_{\infty,\omega}$-freeness implies freeness and excellent classes, in preparation.

[Pi] A. Pillay, An introduction to stability theory, Clarendon Press, Oxford, 1983.

[PiSh] A. Pillay and S. Shelah, Classification over a predicate. I, Notre Dame J. Formal Logic **26** (1985), 361—376.

[Po] B. Poizat, Cours de théorie des modèles, nur al-mantiq wal-ma'rifah, 1985.

[Ro] A. Robinson, On the concept of a differentially closed fields, Bulletin Research Council of Israel, Section F (later: Israel J. Math.) **8F** (1959), 113—128.

[Se] A. Seidenberg, An elimination theory for differential fields, Univ. Calif. Publ. Math. (N. S.) **3** (1956), 31—65.

[Sh71] S. Shelah, Stability, the f. c. p. and superstability, model theoretic properties of formulas in first order theory, Ann. of Math. Logic **3** (1971), 271—362.

[Sh73] —, Differentially closed fields, Israel J. Math. **16** (1973), 314—328.

[Sh74] —, Categoricity of uncountable theories, Proc. Sympos. Pure Math., vol. 25, Amer. Math. Soc., Providence, R. I., 1974, pp. 187—204.

[Sh74a] —, Infinite abelian groups, Whitehead problem and some constructions, Israel J. Math. **18** (1974), 243—256.

[Sh78] —, Classification theory and the number of nonisomorphic models, North-Holland, Amsterdam, 1978.

[Sh83] —, The spectrum problem. I, $\aleph_\varepsilon$-saturated models the main gap, Israel J. Math. **43** (1982), 324—356.

[Sh83a] —, The spectrum problem. II, Totally transcendental theories and the infinite depth case, Israel J. Math. **43** (1982), 357—364.

[Sh83b] —, Construction of many complicated uncountable structures and Boolean algebras, Israel J. Math. **45** (1983), 100—146.

[Sh83c] —, Classification theory for non-elementary classes. I. The number of uncountable models, models of $\psi \in L_{\omega_1,\omega}$, Israel J. Math. **46** (1983), 212—273.

[Sh85] —, Classification of first order theories which have a structure theory, Bull. Amer. Math. Soc. (N. S.) **12** (1985), 227—232.

[Sh85a] —, Monadic Logic: Lowenheim numbers, Ann. Pure Appl. Logic **28** (1985), 203—216.

[Sh85b] —, Classification over a predicate, Notes from Lectures in Simon Fraser University, Summer 1985.

[Sh85c] —, A combinatorial principle and endomorphism rings of abelian groups. II, Proc. of the Conference on Abelian Groups Indine 4/1984, CISM cources and Lecture ∞, No. 287, International Center for Mechanical Sciences, Abelian Groups and Modules, R. Gobel, C. Metelli, A. Orsatti, and L. Salce, eds., 1985, pp. 37—86.

[Sh86] —, Spectrum problem. III, Universal theories, Israel J. Math. **55** (1986), 229—252.

[Sh86a] —, Monadic logic: Hanf numbers, Around Classification Theory, Lecture Notes in Math., vol. 1182, Springer-Verlag, Berlin and New York, 1986, pp. 203—223.

[Sh86b] —, Classification over a predicate. II, Around Classification Theory, Lecture Notes in Math., vol. 1182, Springer-Verlag, Berlin and New York, 1986, pp. 47—90.

[Sh86c] —, Classifying generalized quantifiers, Around Classification Theory, Lecture Notes in Math., vol. 1182, Springer-Verlag, Berlin and New York, 1986, pp. 1—46.

[Sh87] —, Classification theory: completed for countable theories, North-Holland, Amsterdam (to appear).

[Sh87a] —, Universal classes, Proc. of the USA — Israel Sympos. on Classification Theory, Chicago 12/85, Springer-Verlag.

[W] C. Wood, The model theory of differential fields of characteristic $p \neq 0$, Proc. Amer. Math. Soc. **40** (1973), 577—584.

[W1] —, Prime model extensions for differential fields of characteristic $p \neq 0$, J. Symbolic Logic **39** (1974), 469—477.

ЕВРЕЙСКИЙ УНИВЕРСИТЕТ, ИЕРУСАЛИМ, ИЗРАИЛЬ

УНИВЕРСИТЕТ РАТДЖЕРЗА, НЬЮ-БРУНСУИК, НЬЮ-ДЖЕРСИ 08903, США

СЛУЧАЙНЫЕ ПРОЦЕССЫ В БЕСКОНЕЧНОМЕРНЫХ ПРОСТРАНСТВАХ

А. В. Скороход [1]

В последние годы в теории случайных процессов всё больше внимания уделяется изучению случайных процессов в бесконечномерных пространствах. Интерес к таким процессам вызван не только желанием распространить известные методы исследования конечномерных процессов, но и тем, что бесконечномерные процессы возникают и при решении ряда задач, естественных для конечномерных процессов. Те особенности и трудности, которые возникают при рассмотрении бесконечномерных процессов, связаны в первую очередь с геометрической структурой пространства. Поэтому такие трудности исчезают, если удаётся найти подходящее отображение исходного пространства в некоторое другое. Заметим, что упомянутые трудности появляются при рассмотрении вопросов конструктивного плана (изучение свойств выборочных функций, исследование сходимости случайных процессов, построение стохастических интегралов, определение и исследование решений стохастических дифференциальных уравнений). Общие вопросы теории меры безразличны к топологии.

В работе мы отвлекаемся от топологии фазового пространства и исследуем случайный процесс в измеримом пространстве со счётнопорожденной σ-алгеброй. Такие пространства составляют достаточно обширный класс, только такие пространства могут использоваться в приложениях. Выясняется структура измеримого процесса. Определяются некоторые инварианты процесса, которые при весьма широких ограничениях определяют процесс с точностью до неупреждающих обратимых преобразований. Среди этих инвариантов — ранг процесса, который может быть равен или $+\infty$, или натуральному числу. Ранг и есть существенная размерность процесса. Случайной заменой времени и неупреждающим преобразованием процесс можно свести к процессу в R^∞, удовлетворяющему линейному стохастическому дифференциальному уравнению типа Ито. Ранг процесса — это число входящих в уравнение независимых винеровских процессов.

В работе без специальных ссылок используются результаты Мейера, Куниты, Ватанабэ, Деллашери по теории мартингалов и общей теории случайных процессов. Эти результаты широко известны и имеются в книгах этих авторов и во многих других книгах по теории случайных процессов.

1. ТЕОРЕМА КОЛМОГОРОВА

Случайный процесс определяется такими тремя объектами: (1) вероятностным пространством $(\Omega, \mathscr{F}, P)$, (2) фазовым пространством $(X, \mathscr{B})$, которое является измеримым пространством, (3) отображением $x(t, \omega): R_+ \times \Omega \to X$, обладающим тем свойством, что для всех $t \in R_+$ отображение $x(t, \cdot): R_+ \to X$ измеримо относительно σ-алгебр $\mathscr{F}$ и $\mathscr{B}$. Функция $x(t, \omega)$ и есть случайный процесс в фазовом пространстве X, определённый на неотрицательной полупрямой R_+. Вероятностное пространство играет существенную роль лишь в том случае, когда процесс определяется с помощью некоторой конструкции. При общем рассмотрении случайных процессов основной характеристикой процесса являются конечномерные распределения. Это — семейство функций

$$\mu_{t_1, \ldots, t_n}(B_1, \ldots, B_n) = P\{x(t_1, \omega) \in B_1, \ldots, x(t_n, \omega) \in B_n\},$$
$$n = 1, 2, \ldots, \quad t_k \in R_+, \quad B_k \in \mathscr{B}. \tag{1}$$

Эти функции удовлетворяют следующим условиям согласования:

(1) $\mu_{t_{i_1}, \ldots, t_{i_n}}(B_{i_1}, \ldots, B_{i_n}) = \mu_{t_1, \ldots, t_n}(B_1, \ldots, B_n)$,

$\{i_1, \ldots, i_n\} = \{1, 2, \ldots, n\}, \quad n = 1, 2, \ldots, t_k \in \mathbb{R}_+, \quad B_k \in \mathscr{B}$,

(2) $\mu_{t_1, \ldots, t_{n-1}, t_n}(B_1, \ldots, B_{n-1}, X) = \mu_{t_1, \ldots, t_{n-1}}(B_1, \ldots, B_{n-1})$,

(3) $\mu_{t_1, \ldots, t_n}(B, B_2, \ldots, B_n)$ есть мера по B, $\mu_t(X) = 1$.

Теорема Колмогорова утверждает, что для всякого набора согласованных конечномерных распределений существует случайный процесс $x(t, \omega)$, определённый на некотором вероятностном пространстве $\{\Omega, \mathscr{F}, P\}$, для которого выполнено (1).

В качестве Ω можно взять пространство X^{R_+} — пространство всех X-значных функций, определённых на R_+; при этом $\mathscr{F}$

---

[1] ICM86, pp. 163—171. [В оригинальном издании статья опубликована на русском языке. — *Изд. ред.*]

будет цилиндрической σ-алгеброй в X^{R_+}, а мера P однозначно определяется конечномерными распределениями равенством

$$P(\{x(\cdot):x(t_1)\in B_1,\ \ldots,\ x(t_n)\in B_n\})=\mu_{t_1,\ldots,t_n}(B_1,\ \ldots,\ B_n).$$

Сам процесс $x(t,\ \omega)$ задан соотношением

$$x(t,\ \omega(\cdot))=\omega(t),\quad \omega(\cdot)\in X^{R_+}.$$

2. ИЗМЕРИМОСТЬ

Пусть $x(t,\ \omega)$ и $\tilde{x}(t,\ \omega)$ — два случайных процесса, заданных на одном и том же вероятностном пространстве. Они называются стохастическими эквивалентными, если для всех $t\in\mathbb{R}_+$

$$P\{x(t,\ \omega)=\tilde{x}(t,\ \omega)\}=1.$$

Тогда ещё $x(t,\ \omega)$ и $\tilde{x}(t,\ \omega)$ называются модификациями друг друга. При задании процесса конечномерными распределениями естественно не различать модификации. Дальнейшее изучение случайного процесса также не зависит от выбора модификации.

Будем называть процесс измеримым, если он имеет измеримую модификацию, т. е. такую модификацию $\tilde{x}(t,\ \omega)$, которая является измеримым отображением пространства $(R_+\times\Omega,$ $\mathscr{B}_{R_+}\otimes\mathscr{F})$ в $(X,\ \mathscr{B})$, $\mathscr{B}_{R_+}$ — борелевская σ-алгебра на R_+.

Для дальнейшего нам потребуются некоторые пространства случайных величин. Обозначим через $R(\Omega)$ пространство всех числовых величин на вероятностном пространстве $\{\Omega,\ \mathscr{F},\ P\}$. Если $\xi,\ \eta\in R(\Omega)$, то

$$\rho(\xi,\ \eta)=M(1-e^{-|\xi-\eta|})$$

есть метрика в $R(\Omega)$, относительно которой $R(\Omega)$ — полное пространство.

Пусть H_t — замыкание в этой метрике множества величин вида $g(x(s_1,\ \omega),\ \ldots,\ x(s_n,\ \omega))$, где $n=1,\ 2,\ \ldots,\ s_k\in R_+,\ s_k\leqslant t$, $g(x_1,\ \ldots,\ x_n)$ есть B^n-измеримая функция из X^n в R.

H_t — величины, определяемые течением процесса до момента t включительно.

Через H обозначим замыкание $\cup_t H_t$.

Теорема 1. *Пусть σ-алгебра $\mathscr{B}$ счётно порождена. Процесс $x(t,\ \omega)$ измерим тогда и только тогда, когда: (1) H сепарабельно, (2) $\mu_{s,t}$, $(B_1,\ B_2)$ — борелевская функция s при любых $t\in R_+$, $B_1,\ B_2\in\mathscr{B}$.*

В случае счётно порождённого $(X,\ \mathscr{B})$ это пространство можно измеримо взаимно-однозначно отобразить в отрезок $[0,\ 1]$. Это тем не менее не даёт основания утверждать, что по существу все процессы с такими фазовыми пространствами одномерны. В дальнейшем будем предполагать, что σ-алгебра $\mathscr{B}$ счётно порождена, а процесс $x(t,\ \omega)$ является измеримым.

3. ОПРЕДЕЛЯЮЩИЙ ПОТОК σ-АЛГЕБР

Эволюция процесса во времени определяет изменение пространств H_t. Заметим, что эти пространства не меняются при взаимно-однозначных неупреждающих преобразованиях случайного процесса, в частности взаимно-однозначных отображениях фазовых пространств. Вместо семейства пространств H_t можно рассмотреть поток σ-алгебр $(\mathscr{F}_t)$, где $\mathscr{F}_t$ есть σ-алгебра событий, порождённая случайными величинами из H_t. Так как $\mathscr{F}_t$ и H_t определяют друг друга, то будем изучать строение потока $(\mathscr{F}_t)$. Обозначим через $\mathscr{M}_2$ пространство квадратично-интегрируемых $(\mathscr{F}_t)$-согласованных мартингалов.

Теорема 2. (1) *$\mathscr{M}_2$ определяет поток $(\mathscr{F}_t)$;* (2) *существует счётное множество мартингалов $\{\eta_n(t),n=1,\ 2,\ \ldots,\}$ в $\mathscr{M}_2$, такое что для каждого $t\in R_+$ значения мартингалов плотны в H_t.*

Для дальнейшего нам потребуются следующие условия на поток $(\mathscr{F}_t)$:

(1) $\mathscr{F}$ полно относительно P, F_t содержит все множества меры 0;

(2) $\mathscr{F}_t=\bigcap\mathscr{F}_{t+}=\bigcap\limits_{s>t}\mathscr{F}_s$.

При выполнении этих условий мартингалы из $\mathscr{M}_2$ имеют непрерывные справа модификации без разрывов второго рода. Можно считать, что $\mathscr{M}_2$ состоит только из таких мартингалов. Будем обозначать через $\mathscr{W}$ и $\mathscr{P}$ σ-алгебры вполне измеримых и предсказуемых множеств в $R_+\times\Omega$. Первая порождается непрерывными справа, а вторая — непрерывными согласованными процессами.

Для $\eta_1(t)$ и $\eta_2(t)$ из $\mathscr{M}_2$ обозначим через $\langle\eta_1,\eta_2\rangle_t$ их взаимную характеристику — такой $\mathscr{P}$-измеримый процесс, что $\eta_1(t)\eta_2(t)-\langle\eta_1,\eta_2\rangle_t$ есть мартингал. Величина $\langle\eta,\eta\rangle_t=\langle\eta\rangle_t$ — характеристика мартингала $\eta(t)$. Два мартингала $\eta_1(t)$ и $\eta_2(t)$ из $\mathscr{M}_2$ ортогональны $(\eta_1\perp\eta_2)$, если $\langle\eta_1,\eta_2\rangle_t=0$ для всех $t\in R_+$.

Теорема 3. $\mathscr{M}_2=\mathscr{M}_2^0\oplus\mathscr{M}_2^1\oplus\mathscr{M}_2^2$ *(ортогональная сумма в смысле введённого понятия ортогональности);*

$\mathscr{M}_2^0$ — пространство непрерывных мартингалов;

$\mathscr{M}_2^1$ — пространство чисто разрывных компенсированных мартингалов с непрерывными характеристиками;

$\mathscr{M}_2^2$ — пространство чисто разрывных мартингалов с чисто разрывными характеристиками.

Это разложение имеет аналогию в известном разложении Леви процесса с независимыми приращениями на чисто разрывную $(\mathscr{M}_2^2)$, стохастически непрерывную скачкообразную $(\mathscr{M}_2^1)$ и непрерывную составляющие. Как мы увидим дальше, эта аналогия может быть продолжена более глубоко.

4. СУЩЕСТВЕННОЕ ВРЕМЯ. РАНГ

Рассмотрим пространство $\mathscr{M}_2^0 \oplus \mathscr{M}_2^1$ всех мартингалов из $\mathscr{M}_2$ с непрерывными характеристиками. Выбирая некоторую плотную последовательность мартингалов из $\mathscr{M}_2^0 \oplus \mathscr{M}_2^1$ и рассматривая их характеристики, можно построить такой мартингал $\bar{\eta}(t) \in \mathscr{M}_2^0 \oplus \mathscr{M}_2^1$, что для всякого другого мартингала $\eta(t) \in \mathscr{M}_2^0 \oplus \oplus \mathscr{M}_2^1$ возрастающая функция $\langle \eta \rangle_t$ будет абсолютно непрерывна относительно возрастающей функции $\langle \bar{\eta} \rangle_t = \delta_t$. Всякую такую функцию δ_t (она представима как характеристика некоторого мартингала, и относительно неё абсолютно непрерывны характеристики всех других мартингалов из $\mathscr{M}_2^0 \oplus \mathscr{M}_2^1$) будем называть существенным временем для потока $(\mathscr{F}_t)$ (или исходного процесса). Если на некотором интервале δ_t не растёт, то на этом интервале исходный процесс эволюционирует детерминированно, хотя эволюция зависит от того, что происходило до этого интервала времени. Очевидно, что существенное время определяется с точностью до эквивалентности: если δ_t — существенное время и $\hat{\delta}_t$ — непрерывный возрастающий согласованный процесс, для которого

$$P\{0 < d\hat{\delta}_t/d\delta_t < \infty\} = 1,$$

то $\hat{\delta}_t$ — также существенное время.

Рассмотрим теперь пространство $\mathscr{M}_2^0$. Пусть $\{\eta_n(t)\}$ плотно в $\mathscr{M}_2^0$. Ортогонализуем $\{\eta_n(t)\}$ следующим образом:

$$\xi_1(t) = \eta_1(t), \quad \xi_n(t) = \eta_n(t) - \sum_{k=1}^{n-1} \int_0^t \alpha_{nk}(s)\, d\xi_k(s), \quad n > 1,$$

$$\alpha_{nk}(s) = \frac{d\langle \eta_n, \xi_k \rangle}{d\langle \xi_k \rangle_s}.$$

Построенная последовательность $\{\xi_k(t)\}$ обладает следующим свойством: для всякого мартингала $\eta(t) \in \mathscr{M}_2^0$ справедливо представление

$$\eta(t) = \sum_{k=1}^{\infty} \int_0^t \alpha_k(s)\, d\xi_k(s), \quad \alpha_k(s) = \frac{d\langle \eta, \xi_k \rangle_s}{d\langle \xi_k \rangle_s}.$$

Всякую последовательность, обладающую этим свойством, будем называть базисом в $\mathscr{M}_2^0$.

Теорема 4. *Пусть $\delta(t)$ — существенное время, $\{\xi_k(t)\}$ и $\{\tilde{\xi}_k(t)\}$ — два базиса в $\mathscr{M}_2^0$. Тогда если*

$$\nu(t) = \sum_k I_{\{d\langle \xi_k \rangle_t / d\delta(t) > 0\}}, \quad \tilde{\nu}(t) = \sum_k I_{\{d\langle \tilde{\xi}_k \rangle_t / d\delta(t) > 0\}},$$

то

$$P\left\{ \int I_{\{\nu(s) \neq \tilde{\nu}(s)\}}\, d\delta(s) = 0 \right\} = 1.$$

Таким образом, существует не зависящая от выбора базиса такая $\mathscr{P}$-измеримая функция $r(t)$, что

$$P\{\nu(t) = r(t)\} = 1;$$

$r(t)$ называется рангом потока $(\mathscr{F}_t)$ в момент t. А величина

$$r = \inf\left\{ k : P\left\{ \int I_{\{\nu(s) > k\}}\, d\delta(s) = 0 \right\} = 1 \right\}$$

(считаем $\inf \varnothing = +\infty$) называется рангом потока $(\mathscr{F}_t)$. Ранг потока — это размерность минимального базиса в $\mathscr{M}_2^0$; это вытекает из следующего утверждения:

Теорема 5. *Пусть $\delta(t)$ — существенное время, r — ранг $(\mathscr{F}_t)$. Существуют последовательность $\mathscr{P}$-измеримых множеств $A_1 \supset A_2 \supset \ldots \supset A_r$ (эта последовательность бесконечна при $r = \infty$) и базис $\{\xi_k(t), k < r + 1$, такие что*

$$\langle \xi_i \rangle_t = \int_0^t I_{A_i}(s)\, d\delta(s).$$

5. ПРИМЕРЫ

Мартингалы из $\mathscr{M}_2^1 \oplus \mathscr{M}_2^2$ имеют довольно простую структуру, хотя для $\mathscr{M}_2^1$ мы укажем далее некоторый определяющий инвариант. Сейчас наша цель показать, что ранг процесса может быть различным, причём процесс при этом может быть довольно регулярным процессом в R.

Пример 1. Пусть $x(t, \omega)$ — процесс в R^m, его компоненты $x_1(t, \omega), \ldots, x_m(t, \omega)$ — независимые винеровские процессы.

Покажем, что его ранг $r=m$. Так как для всякого мартингала, согласованного с потоком $(\mathscr{F}_t)$, порождённым $x(t,\omega)$, справедливо представление

$$\eta(t)=\sum_{k=1}^{m}\int_0^t \varphi_k(s)\,dx_k(s,\,\omega),$$

то $r\leqslant m$. Если бы существовал базис из $r<m$ мартингалов $\eta_1(t),\,\ldots,\,\eta_r(t)$ и

$$x_k(t,\,\omega)=\sum_{i=1}^{k}\int_0^t \alpha_{ki}(s)\,d\eta_i(s),$$

то

$$\delta_{lk}(t)=\sum_{i=1}^{r}\int_0^t \alpha_{li}(s)\,\alpha_{kj}d\langle\eta_i\rangle_s,\qquad \delta_{lk}=\sum_{i=1}^{r}\alpha_{li}\alpha_{ki}(t)\frac{d\langle\eta_i\rangle_t}{dt},$$

$l,\ k\leqslant m$, что невозможно (ранг матрицы $\|\alpha_{ki}(d\langle\eta_i\rangle/ds)^{1/2}\|$ не превосходит $r\leqslant m$).

Пример 2. Пусть $\{w_k(t)\}$ — последовательность независимых винеровских процессов. Положим

$$\eta_{n,\,n+1}(t)=\int_0^t(\operatorname{arctg} w_{n+1}(s)+\pi/2)\,dw_n(s),$$

$$\eta_{n,\,m}(t)=\int_0^t(\operatorname{arctg}\eta_{n+1,\,m}(s)+\pi/2)\,dw_n(s),\quad m>n+1.$$

Поток $(\mathscr{F}_t^{n,\,m})$, порождаемый мартингалом $\eta_{n,\,m}(t)$, совпадает с потоком, порождаемым винеровскими процессами $w_n(t)$, $w_{n+1}(t),\,\ldots,\,w_m(t)$, и его ранг равен $m-n+1$. Легко убедиться, что существует предел в среднеквадратической сходимости $\lim_{m\to\infty}\eta_{n,\,m}(t)=\eta_n(t)$ и поток $\mathscr{F}_t^n$, порождённый предельным процессом, совпадает с потоком, порождённым винеровскими процессами $w_n(t)$, $w_{n+1}(t)$, ... Он имеет ранг $+\infty$. В то же время процессы $\eta_{n,\,m}(t)$ и $\eta_n(t)$ — одномерные непрерывные мартингалы с дифференцируемыми характеристиками.

6. КВАЗИНЕПРЕРЫВНЫЙ ПОТОК $(\mathscr{F}_t)$ СЛУЧАЙНАЯ ЗАМЕНА ВРЕМЕНИ

Поток называется квазинепрерывным, если $\mathscr{M}_2^2$ содержит только 0. В этом случае характеристики всех мартингалов из $\mathscr{M}_2$ непрерывны. Пусть $\delta(t)$ — существенное время и $\delta(t)\uparrow\infty$ при

$t\uparrow\infty$. Положим $\tau_t=\delta^{-1}(t)$, $\tilde{x}(t,\,\omega)=x(\tau_t,\,\omega)$, $(\tilde{\mathscr{F}}_t)$ $\tilde{\mathscr{M}}_2^0$, $\tilde{\mathscr{M}}_2^1$, $\tilde{\mathscr{M}}_2^2$ отвечает потоку $(\tilde{\mathscr{F}}_t)$.

Теорема 6. *$\tilde{\mathscr{M}}_2=\{\eta(\tau_t),\ \eta\in\mathscr{M}_2\}$, $\tilde{\mathscr{M}}_2^i=\{\eta$ $i=0,\ 1,\ 2$. Заметим, что процесс $\tilde{x}(t,\omega)$ не обяза дает $(\tilde{\mathscr{F}}_t)$, он вообще может стать неслучайным.*

Поток называется абсолютно непрерывным, е рывен и существенное время абсолютно непрерывн цируемо по t). Приведённая выше замена врем к тому, что поток $(\mathscr{F}_t)$ абсолютно непрерывен.

Теорема 7. *Для абсолютно непрерывного поток (1) существуют последовательность винеровск $\{w_k(t)\}$ и пуассоновская мера $\pi(dx\times dt)$ на R_+^2, $E\pi(ds\times dx)=dxdt$, такие что всякий мартингал η ставим в виде*

$$\eta(t)=\sum_{k=1}^{\infty}\int_0^t\alpha_k(s)\,d\omega_k(s)+\int_0^t\int_R\gamma(s,\ x)\,[\pi(ds\times dx)$$

где $\alpha_k(s)$ и $\gamma(s,\ x)$ — некоторые функции, измер тельно $\mathscr{P}$ и $\mathscr{P}\otimes\mathscr{B}_{R_+}$ соответственно, для которых

$$P\left\{\sum\int_0^t\alpha_k^2(s)\,ds+\int_0^t\int_R\gamma^2(s,\ x)\,ds\,dx<\infty\right.$$

(2) существуют такие $\mathscr{P}$-измеримые функции принимающие значения: первая — 0, 1, 2, ..., + из R_+ или $+\infty$, такие что $\mathscr{F}_t$ совпадает с поток ным процессами

$$\left\{\int_0^t I_{\{r(s)\geqslant k\}}\,dw_k(s),\ k=1,\ 2,\ \ldots;\ \int_0^t\int_R I_{\{x\leqslant\lambda(t)\}}\frac{1}{1+x^2}\right.$$

$r(t)$ — ранг потока. Функция $\lambda(t)$ имеет следующий $\nu(t)-\nu(s)$ при $s<t$ — число скачков всех марти на отрезке $[s,t]$; тогда $\int_s^t\lambda(u)\,du$ есть компенсат $\lambda(t)$ называется шириной пуассоновского спект тоже величина, инвариантная при преобразован сохраняющих поток.

7. НЕУПРЕЖДАЮЩИЕ ПРЕОБРАЗОВАНИЯ И СТОХАСТИЧЕСКИЕ УРАВНЕНИЯ

Пусть $\{\eta_n(t)\}$ — такая последовательность мартингалов из $\mathscr{M}_2$, что для всех t линейная оболочка величин $\{\eta_n(t),\ n \geqslant 1\}$ плотна в H_t. Будем предполагать, что поток абсолютно непрерывен. Тогда для каждого мартингала $\eta_n(t)$ можно записать формулу (2) с функциями $\alpha_{nk}(s)$ и $\gamma_n(s, x)$. При фиксированном x — это $\mathscr{F}_s$-измеримые случайные величины, входящие в H_s. Значит, они могут быть выражены как некоторая линейная функция от $\{\eta_n(s),\ n = 1, 2, \ldots\}$. Более того, поскольку они предсказуемы, а предсказуемая проекция $\eta_n(s)$ есть $\eta_n(s-)$, то их можно рассматривать как линейные функции от $\eta_n(s-)$. Будем формально записывать эти линейные функции в виде бесконечных линейных комбинаций. (На самом деле это пределы конечных линейных комбинаций.) Если

$$\alpha_{nk}(s) = \sum \alpha^n_{km}(s)\,\eta_m(s-),$$
$$\gamma_n(s) = \sum g^n_m(s, x)\,\eta_m(s-)$$

(функции $\alpha^n_{km}(s)$ и $g^n_m(s, x)$ уже неслучайны), то для последовательности $\{\eta_n(t)\}$ получаем следующую систему линейных стохастических дифференциальных уравнений:

$$\alpha\eta_n(t) = \sum_{k=1}^{\infty}\left(\sum_{m=1}^{\infty} \alpha^n_{km}(s)\,\eta_m(s-)\right) dw_k(s) + \int\limits_{R_+} \sum_{m=1}^{\infty} g_m(s, x)\,\eta_m(s-)\,[\pi(ds \times dx) - ds\,dx]. \quad (3)$$

Уравнение (3) можно рассматривать как линейное уравнение в $\mathbb{R}^\infty$ с неограниченными операторными коэффициентами. Уточнение смысла уравнений (3), а также исследование их решений (в частности, условий единственности) — задача теории линейных стохастических дифференциальных уравнений.

Поскольку процесс $\eta_n(t)$ $\mathscr{F}_t$-измерим, а $\mathscr{F}_t$ порождено значениями процесса $x(s, \omega)$ при $s \leqslant t$, то существует такая $\mathscr{B}_{R_+} \otimes \mathscr{B}^{R_+}$- измеримая функция $f_n(t, x(\cdot))$ из $R_+ \times X^{R_+}$ в R, что

$$\eta_n(t) = f_n(t, x(\cdot, \omega));$$

при этом функция f_n неупреждающая: если $y_1(s) + y_2(s)$ при $s \leqslant t$, $y_i(\cdot) \in X^{R_+}$, то $f_n(t, y_1(\cdot)) = f_n(t, y_2(\cdot))$. Отображение процесса $x(t, \omega) \to \{f_n(t, x(\cdot, \omega)),\ n = 1, 2, \ldots\}$ в R^∞ сохраняет поток $(\mathscr{F}_t)$, и поэтому оно обратимо в обобщённом смысле. Это и есть то обратимое неупреждающее преобразование в R^∞, которое после случайной замены времени приводит квазинепрерывный процесс к процессу, удовлетворяющему системе линейных стохастических дифференциальных уравнений.

8. НЕКОТОРЫЕ ВЫВОДЫ И ЗАДАЧИ

1. Вполне понятно строение квазинепрерывного потока. Он определяется такими тремя характеристиками: (а) существенным временем $\delta(t)$, которое определяется с точностью до эквивалентности для мер, порождаемых возрастающими непрерывными функциями; (б) рангом процесса $r(t)$, определённым однозначно почти всюду относительно существенного времени; (в) шириной пуассоновского спектра $\lambda(t)$, которая определяется однозначно почти всюду относительно существенного времени, если это время фиксировано.

Заметим, что функции $r(t)$ и $\lambda(t)$, определяющие поток при $\delta(t) = t$, должны быть предсказуемы относительно определяемого потока. Это накладывает на них определённые ограничения, которые следует изучить.

2. Выясняется особая роль линейных стохастических дифференциальных уравнений с неограниченными операторными коэффициентами. Изучение таких уравнений проводилось в рамках теории линейных систем при весьма жестких ограничениях; оказывается, имеет смысл рассматривать такие системы без всяких ограничений.

3. Рассмотрены только некоторые преобразования процессов и потоков σ-алгебр. Весьма интересен вопрос о необратимых преобразованиях процессов (например, с помощью преобразований фазового пространства), при которых сохраняются потоки σ-алгебр.

4. Было бы полезно иметь способ доказывать квазинепрерывность потока, порождённого случайным процессом, без рассмотрения пространства мартингалов, а используя лишь его конечномерные распределения. (Это возможно, например, для марковских процессов.)

5. Представляет интерес изучение общих потоков (с предсказуемыми разрывами). Здесь можно было бы использовать операцию предельного перехода для потоков с конечным числом предсказуемых разрывов.

ЛИТЕРАТУРА

1. Скороход А. В. О локальном строении непрерывных марковских процессов. — ТВП, 1966, т. 1, № 13. с. 381—423.

2. — Операторные стохастические дифференциальные уравнения и стохастические полугруппы. — УМН, 1982, т. 37, № 6, с. 157—183.

3. Гихман И. И., Скороход А. В. Стохастические дифференциальные уравнения и их приложения. — Киев: Наукова думка, 1982.

4. Скороход А. В. Стохастические уравнения для сложных систем. — М.: Наука, 1983.

ИНСТИТУТ МАТЕМАТИКИ АКАДЕМИИ НАУК УССР, КИЕВ-4, СССР

АЛГОРИТМЫ РЕШЕНИЯ УРАВНЕНИЙ

Стив Смейл[1]

Основная цель этой работы — попытаться осмыслить эффективность алгоритмов решения систем уравнений. Возьмём, например, отображение f комплексного декартова пространства $\mathbb{C}^n$ в $\mathbb{C}^n$, заданное многочленами. Насколько быстро можно найти хорошие приближения одного или всех нулей f? Для ясности и в то же время достаточной общности изложения мы будем обычно «плясать» вокруг этого примера, рассматривая два крайних случая, а именно будем доказывать теоремы, с одной стороны, для одного-единственного комплексного многочлена, а с другой — для аналитического отображения $f : \mathbf{E} \to \mathbf{F}$ одного банахова пространства в другое (оба вещественные или оба комплексные).

Эту тему, несомненно, следовало бы отнести к численному анализу. Однако моя точка зрения сформировалась в основном под влиянием теории информационной сложности. Поэтому упор будет сделан на самих алгоритмах и глобальном изучении их быстродействия. Соответственно меньше уделяется внимания результатам, получаемым с помощью алгоритмов, и асимптотическим критериям эффективности, которые часто обсуждаются в литературе по численному анализу.

Такое глобальное изучение, основанное на теории сложности, придаёт нашему подходу бо́льшую систематичность, бо́льшую абстрактность и бо́льшую теоретическую направленность; в частности, мы меньше заботимся о непосредственном создании быстрых методов решения систем уравнений. Мы хотим проникнуть в суть понятий надёжности и эффективности. Глобальное исследование алгоритма требует привлечения топологии и геометрии.

Если какой-нибудь алгоритм и зарекомендовал себя, как пригодный для решения нелинейных систем, так это метод Ньютона

---

[1] Steve Smale, Algorithms for solving equations, ICM86, pp. 172—195.

и его многочисленные модификации. По существу уже греки применяли этот метод для вычисления квадратных корней; и сегодня он остаётся лучшим для этой цели. И в теории нелинейных функциональных уравнений в банаховых пространствах метод Ньютона также занимает центральное место.

Метод Ньютона и будет предметом настоящего сообщения. Мы докажем теоремы, касающиеся этого метода, для случая одной переменной (используя основную теорему алгебры) и для отображений банаховых пространств; кроме того, мы аппроксимируем с помощью метода Ньютона симплекс-метод Данцига [8] для задач линейного программирования.

На современную трактовку метода Ньютона решающим образом повлияли работы и точка зрения Канторовича. Подробности см. у Бергера [2], Хенричи [19], Канторовича — Акилова [23] и Островского [38]. Эта трактовка привлекает своей простотой, общностью и минимальностью предположений. В случае банаховых пространств все предположения состоят в ограниченности первой и второй производных в области определения рассматриваемого отображения.

Наше исследование мы начнем с аналитических отображений и *получим оценки в одной точке.*

Заметим, что при выполнении итерационного алгоритма мы на каждом его шаге находимся в некоторой точке области определения и можем вычислять производные в этой точке. По информации о производных требуется вывести заключение о необходимости следующего шага или о завершении работы алгоритма. Эти соображения мотивировали нашу статью [52], упоминаемую далее как [point estimates] или просто [P. E.]. Так как результаты из [P. E.] ниже многократно используются, мы дадим частичный обзор этой теории.

Важным моментом теоретического изучения проблем, где возможны лишь приближённые решения, является наличие вездесущего малого параметра $\varepsilon > 0$, при помощи которого оценивается расстояние до точного решения. Однако для хорошо-обусловленного решения уравнения $f(\zeta)=0$, где f — комплексный многочлен, можно избежать такого произвола, используя понятие приближённого нуля.

В случае одной переменной метод Ньютона для решения уравнения $f(\zeta)=0$ с начальной точкой $z_0 \in \mathbb{C}$ состоит в вычислении значений индуктивно определённой последовательности $z_k = z_{k-1} - f(z_{k-1})/f'(z_{k-1})$. Будем называть точку z_0 *приближённым нулём*, если

$$|z_k - z_{k-1}| \leqslant \left(\frac{1}{2}\right)^{2^{k-1}-1} |z_1 - z_0|, \quad k = 1, 2, 3, \ldots .$$

Таким образом, скажем, для $k=5$ коэффициент в правой части уже равен $(1/2)^{15}$. Это явление с удвоением точности на каждом шаге можно наблюдать при вычислениях на ЭВМ. Однако нужен критерий, который по информации в точке z_0 гарантировал бы такую оценку для всех k. Этим мотивируются введение нашего инварианта $\alpha(z,f)$ (см. ниже) и приводимые далее теоремы А и В.

Пусть $\beta(z, f)$ — длина «вектора Ньютона»:

$$\beta(z, f) = \left| \frac{f(z)}{f'(z)} \right|$$

и

$$\gamma(z, f) = \max_{k>1} \left| \frac{f^{(k)}(z)}{k! f'(z)} \right|^{1/(k-1)} .$$

Здесь через $f^{(k)}(z)$ обозначена k-я производная многочлена f в точке z. Тогда инвариант $\alpha(z,f)$ определяется как произведение β на γ:

$$\alpha(z, f) = \beta(z, f)\,\gamma(z, f) = \beta(z)\,\gamma(z).$$

Теорема А (частный случай). *Существует константа* α_0, *равная приблизительно* 0.130707, *такая что если* $\alpha(z,f) < \alpha_0$, *то* z *есть приближённый нуль* f.

В общем случае теорема А формулируется аналогично, только f и определение приближённого нуля обобщаются следующим образом.

Отображение f является теперь аналитическим отображением $f: \mathbf{E} \to \mathbf{F}$ одного банахова пространства в другое (скажем, $\mathbf{E} = \mathbb{C}^n = \mathbf{F}$). Метод Ньютона с начальной точкой $z_0 \in E$ задаётся формулой

$$z_k = z_{k-1} - Df(z_{k-1})^{-1} f(z_{k-1});$$

точка z_0 называется *приближённым нулем* f, если

$$\| z_k - z_{k-1} \| \leqslant \left(\frac{1}{2}\right)^{2^{k-1}-1} \| z_1 - z_0 \|, \quad k = 1, 2, 3, \ldots .$$

Положим

$$\beta(z, f) = \| Df(z)^{-1} f(z) \|,$$

$$\gamma(z, f) = \sup_{k>1} \left\| Df(z)^{-1} \frac{D^k f(z)}{k!} \right\|^{1/(k-1)} ,$$

где через $D^k f(z)$ обозначена k-я производная f, представляющая собой k-линейный функционал (по поводу дифференциального исчисления в банаховых пространствах см. Дьёдонне [10] или Ленг [30]). Если $Df(z)^{-1}$ не определено, то считаем α, β и γ

бесконечными. Теперь теорема А приобретает смысл для аналитических $f : \mathbf{E} \to \mathbf{F}$ и остаётся при этом верным утверждением; доказательство см. в [Р. Е].

Для одной переменной и $\alpha_0 = 1/54$ теорему А независимо доказал Ким [27]. В [Р. Е.] я высказал предположение, что подобные теоремы можно получить, исходя из теории Канторовича. Впоследствии Ройден [42] показал, что это так и есть, немного улучшив при этом константу α_0 в теореме А. Недавно Карри [7] распространил одномерный вариант теоремы А на обобщённые методы Ньютона высших порядков.

В теореме А может показаться неудобным то обстоятельство, что для вычисления входящего в α множителя γ требуется знание *всех* производных f. На самом деле нужна лишь первая производная вместе с некоторой грубой информацией об f. Чтобы сформулировать это точно, определим функцию $\varphi_d : \mathbb{R} \to \mathbb{R}$ формулой $\varphi_d(r) = \sum_{i=0}^{d} r^i$, и пусть φ_d' обозначает её производную.

Если f — многочлен, задаваемый формулой $f(z) = \sum_{i=0}^{d} a_i z^i$, то положим $\|f\| = \max_i |a_i|$.

Теорема В (частный случай). *Для всякого многочлена f и каждой точки $z \in \mathbb{C}$*

$$\gamma(z, f) \leqslant \frac{\|f\|}{|f'(z)|} \frac{\varphi_d'(|z|)^2}{\varphi_d(|z|)}.$$

В общем случае, когда $f : \mathbf{E} \to \mathbf{F}$ — аналитическое отображение банаховых пространств, его можно представить в виде $f(z) = \sum_{k=0}^{\infty} a_k z^k$, где a_k — симметричные k-линейные отображения и $a_k z^k$ — значение a_k на k-наборе $(z, \ldots, z)$ элементов из $\mathbf{E}$. Пусть $\|a_k\|$ — обычная норма (см., например, Ленг [30]) и $\|f\| = \sup_k \|a_k\|$.

Теорема В. *Для указанного отображения $f : \mathbf{E} \to \mathbf{F}$ и любого $z \in \mathbf{E}$*

$$\gamma(z, f) \leqslant \|f\| \left\| Df(z)^{-1} \right\| \frac{\varphi_d'(\|z\|)^2}{\varphi_d(\|z\|)}.$$

Теорема В анонсирована в [Р. Е.] и доказывается в § 1 ниже. Она используется в наших доказательствах формулируемых далее теорем о вычислительной сложности случайного алгоритма и средней площади множества приближённых нулей.

Теорема А позволяет ввести простой случайный алгоритм для нахождения нуля многочлена, поскольку предоставляет в наше распоряжение критерий окончания. Этот алгоритм формулируется так: для данного комплексного многочлена f

(1) Выбрать случайное $z \in \mathbb{C}$, такое что $|z| < 3$.

(2) Проверить, выполнено ли условие $\alpha(z, f) < \alpha_0$? Если да, то закончить процедуру (или применить метод Ньютона, скажем 5 раз, а затем закончить). Если нет, перейти к (1).

Естественно, для конкретного f возникает вопрос: сколько шагов в среднем потребуется этому случайному алгоритму? Чтобы придать вопросу точный смысл, рассмотрим множество Ω всевозможных последовательностей $Z = (z_1, z_2, z_3, \ldots)$, $|z_k| < 3$. Снабдим диск радиуса 3 нормированной мерой Лебега и наделим Ω вероятностной мерой, представляющей собой бесконечное произведение вероятностных мер, как в работе Шуба — Смейла [44].

Определим функцию $\sigma : \Omega \to \mathbb{Z}^+$, положив $\sigma(Z)$ равным первому σ, такому что $\alpha(z_\sigma, f) < \alpha_0$. Тогда *среднее число шагов*, необходимых для завершения нашего случайного алгоритма, определяется формулой

$$\sigma_f = \operatorname*{average}_{Z \in \Omega} \sigma(Z).$$

Теорема (о вычислительной сложности случайного алгоритма). *Множество многочленов f в $P_d(1)$, для которых среднее число шагов больше σ, имеет меру, меньшую cd^5/σ, где c — некоторая константа.*

Эта теорема доказана в § 2. Можно надеяться, что со временем будет получена более точная оценка.

Через $P_d(1)$ здесь обозначено множество всех комплексных многочленов $f(z) = \sum_{i=0}^{d} a_i z^i$ с $a_d = 1$ и $|a_i| \leqslant 1$. Мы рассматриваем на $P_d(1)$ равномерно распределённую вероятностную меру (т. е. нормированную меру Лебега), используя тот факт, что $P_d(1)$ — ограниченное множество в $\mathbb{C}^d$. Множество многочленов $f \in P_d(1)$ с $\sigma_f = \infty$ имеет меру 0. Эти f единственные, для которых в некотором разумном смысле нахождение нуля является некорректной задачей.

Предположим, далее, что $z_0 \in \mathbb{C}$, f — комплексный многочлен и

$$|z_n - \zeta| \leqslant \left(\frac{1}{2}\right)^{2^n - 1} |z_0 - \zeta|, \qquad (*)$$

где $f(\zeta)=0$ и $z_n = z_{n-1} - f(z_{n-1})/f'(z_{n-1})$. Неравенство (*) определяет еще одну форму «сверхсходимости», когда нет зависимости от малого параметра. Пусть Ω_f — множество точек $|z_0| < 1$, удовлетворяющих (*), и A_f — его площадь.

Теорема (о средней площади множества приближённых нулей).

$$\underset{f \in P_d(1)}{\mathrm{average}}\, A_f > c > 0,$$

где c не зависит от d.

Этот результат верен и для случая усреднения по множеству

$$\left\{ f \mid f(z) = \sum_{i=0}^{d} a_i z^i, \quad |a_i| \leqslant 1, \quad i = 0,\ 1,\ \ldots,\ d \right\}.$$

Доказательство теоремы см. в § 3.

Точки z_0, удовлетворяющие неравенству (*), называются приближенными нулями второго рода. Вообще, если $f: \mathbf{E} \to \mathbf{F}$ — аналитическое отображение, $z_0 \in \mathbf{E}$, $z_n = z_{n-1} - Df(z_{n-1})^{-1} f(z_{n-1})$ и

$$\| z_n - \zeta \| \leqslant \left(\frac{1}{2}\right)^{2^n - 1} \| z_0 - \zeta \|, \quad f(\zeta) = 0,$$

то z_0 будет называться *приближённым нулём второго рода*. Для доказательства предыдущих теорем мы используем следующую теорему C, которая была доказана в [P. E.]:

Теорема C. *Пусть $f: \mathbf{E} \to \mathbf{F}$ — аналитическое отображение, $\zeta \in \mathbf{E}$, $f(\zeta) = 0$ и точка $z \in \mathbf{E}$ удовлетворяет неравенству*

$$\| z - \zeta \| < \frac{3 - \sqrt{7}}{2} \frac{1}{\gamma(\zeta, f)}.$$

Тогда z есть приближённый нуль второго рода.

Рассмотрим аналитическое отображение $f: \mathbf{E} \to \mathbf{F}$ одного банахова пространства в другое. Мы хотим изучить один алгоритм аппроксимации решений уравнения $f(\zeta) = 0$, основанный на методе Ньютона. Будет исследован вопрос о сложности этого алгоритма.

Пусть f — заданное отображение и $z_0 \in \mathbf{E}$. Для наших целей потребуется важное понятие *поднятого пути* $\sigma = \sigma(z_0, f) \subset \mathbf{E}$. Оно определяется так. Если производная $Df(z_0): \mathbf{E} \to \mathbf{F}$ является изоморфизмом, то существует отображение $f_{z_0}^{-1}$ окрестности точки $f(z_0)$ в пространство $\mathbf{E}$, локально обратное к f и переводящее $f(z_0)$ в z_0. Для неотрицательных t_0, достаточно близких к единице, $f_{z_0}^{-1}$ отображает множество $\{ t f(z_0) \mid t_0 < t \leqslant 1 \}$ в $\mathbf{E}$.

Может оказаться, что $f_{z_0}^{-1}$ продолжается до отображения луча $\{ t f(z_0) \mid 0 \leqslant t \leqslant 1 \} \to \mathbf{E}$, оставаясь обратным к f и таким, что производная по направлению этого луча является изоморфизмом. В этом случае мы скажем, что $\sigma = \sigma(z_0, f)$ определено и представляет собой множество $f_{z_0}^{-1} \{ t f(z_0) \mid 0 \leqslant t \leqslant 1 \}$.

Для случая комплексного многочлена f от одной переменной кривая σ обсуждается и играет важную роль в работах Смейла [50, 51] и Шуба — Смейла [43, 44].

Далее, $\sigma(z_0, f)$ можно определить для типичных систем $f: \mathbb{C}^n \to \mathbb{C}^n$, заданных многочленами.

Для указанных выше z_0, f введём следующий инвариант $M(z_0, f)$. Если $\sigma(z_0, f)$ не определено, то $M(z_0, f) = \infty$. В противном случае

$$M(z_0,\ f) = \max_{z \in \sigma} \alpha(z,\ f) / \| f(z) \|.$$

Описание упомянутого выше алгоритма решения уравнения $f(\zeta) = 0$ содержится в следующей теореме:

Теорема (о быстродействии глобального метода Ньютона). *Существуют (малые) положительные константы — вещественное число c и целое число l со следующим свойством. Пусть $f: \mathbf{E} \to \mathbf{F}$ — аналитическое отображение, $z_0 \in \mathbf{E}$ и $M(z_0, f) < \infty$. Пусть далее n — целое число,*

$$n > c \| f(z_0) \| M(z_0,\ f), \qquad \Delta = 1/n.$$

Положим $w_i = (1 - i\Delta) f(z_0)$, $i = 0,\ \ldots,\ n$. Тогда корректно определены все точки $z_i = N_{f-w_i}^{l}(z_{i-1})$, $i = 1,\ \ldots,\ n$, причём z_n есть приближённый нуль f.

Здесь $N_{f-w_i}(z_{i-1})$ обозначает результат преобразования точки z_{i-1}, за один шаг метода Ньютона при решении уравнения $f(\zeta) - w_i = 0$, а $N_{f-w_i}^{l}(z_{i-1})$ — то же преобразование, применённое повторно l раз. Константы c и l близки к 4. Доказательство дано в § 4.

Алгоритм в этой теореме является версией глобального метода Ньютона, который рассматривали Смейл [47], Хирш — Смейл [20], Келлер [26], Гарсиа — Гоулд [15], Абади — Герреро [1]. Близкие результаты можно найти у Кунга [28] и в сборнике [9] под редакцией Дежона и Хенричи (особенно в статье Дежона — Никкеля в этом сборнике). Отметим, что сформулированная теорема является обобщением результатов для одномерного случая (Шуб — Смейл [43, 44] и Смейл [50, 51]).

Идеи, на которых основано доказательство теоремы, пригодны и в других ситуациях, например к гомотопическим задачам (см. Келлер [25] и Чжоу — Малле-Паре — Йорк [4]). Далее, если рассматриваемая задача нахождения нуля некорректна, то может оказаться корректной соответствующая возмущённая задача $|f(z)| < \varepsilon$, а значит применима предыдущая теорема (ср. Смейл [50, 51] и Шуб — Смейл [43, 44]).

И метод Ньютона, и кусочно-линейные алгоритмы можно считать методами последовательных приближений. Этот вопрос разбирается в работах Ивза — Скарфа [11], Смейла [47], Хирша — Смейла [20]. Самодвойственный вариант симплекс-метода Данцига [8] тоже можно интерпретировать как метод последовательных приближений (Смейл [48]). Поэтому упомянутая выше связь между симплекс-методом линейного программирования и методом Ньютона не является неожиданной. В действительности эта связь очень проста и естественна в рамках задачи о линейном дополнении (LCP[1]).

Задача LCP с исходными данными (M, q), где M — квадратная матрица N-го порядка и $q \in \mathbb{R}^N$, ставится следующим образом.

Рассмотрим функции $x^{\pm} = (x \pm |x|)/2$ вещественной переменной x (т. е. $x^{+}(x) = (x + |x|)/2$ и аналогично для знака минус). Пусть $\Phi_M : \mathbb{R}^N \to \mathbb{R}^N$ — отображение, заданное формулой $\Phi_M(x) = x^{+} + Mx^{-}$, где $x^{\pm} = (x_1^{\pm}, \ldots, x_N^{\pm})$. Тогда задача LCP состоит в том, чтобы для данных (M, q) решить относительно x уравнение

$$\Phi_M(x) = q. \qquad \text{(LCP)}$$

Подробности см. у Коттла — Данцига [6], Смейла [48, 49] и в других работах, на которые они ссылаются.

Мы интересуемся задачей LCP в основном потому, что она включает в себя как частный случай задачу линейного программирования (LPP[2]). Задача LPP с исходными данными (A, b, c), где A — матрица размера $m \times n$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ формируется так:

$$\begin{array}{l} \text{минимизировать } c \cdot x \text{ по } x \in \mathbb{R}^n \\ \text{при условии } x \geqslant 0,\ Ax \geqslant b. \end{array} \qquad \text{(LPP)}$$

Здесь через $c \cdot x = \langle c, x \rangle$ обозначено скалярное произведение.

Задача LPP с исходными данными (A, b, c) порождает задачу LCP с исходными данными (M, q), где $N = n + m$,

$q = (c, -b)$ и $M = \begin{bmatrix} 0 & -A^T \\ A & 0 \end{bmatrix}$. В дальнейшем будет предполагаться, что M имеет именно такой вид.

Задача LPP разрешима тогда и только тогда, когда разрешима соответствующая задача LCP, причём их решения соответствуют друг другу естественным образом.

Ниже мы даём аналитическую аппроксимацию отображения Φ_M, которая одновременно устраняет особенности в задаче LPP и делает аналитические методы, в частности, метод Ньютона, пригодными для её решения. Это позволяет нам аппроксимировать с помощью метода Ньютона самодвойственный вариант симплекс-метода Данцига [8].

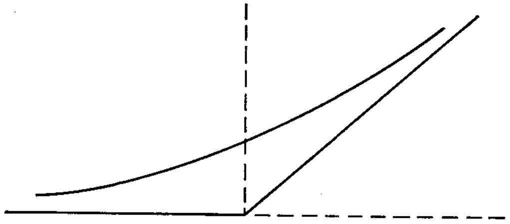


Рис. 1.

Начнём с аппроксимации функций одной переменной $x^{\pm}$. Пусть $\varphi_a^{\pm} : \mathbb{R} \to \mathbb{R}$ — функции, определённые формулой

$$\varphi_a^{\pm}(x) = \frac{x \pm (x^2 + a^2)^{1/2}}{2}, \quad \text{где} \quad a \geqslant 0.$$

Графиком функции φ_a^{+} для $a > 0$ служит гипербола (рис. 1).

При $a \to 0$ функции $\varphi_a^{\pm}$ равномерно приближаются к $x^{\pm}$, и эта аппроксимация только улучшается, когда $x \to \pm\infty$.

Пусть $\Phi_a : \mathbb{R}^N \to \mathbb{R}^N$ — отображение, задаваемое формулой

$$\Phi_a(x) = \Phi_a^{+} + M\Phi_a^{-}(x),$$

где

$$\Phi_a^{\pm}(x) = (\varphi_a^{\pm}(x_1), \ldots, \varphi_a^{\pm}(x_N)).$$

Согласно предыдущим замечаниям отображение Φ_a стремится к Φ_M равномерно на всём пространстве $\mathbb{R}^N$ при $a \to 0$; кроме того, $\Phi_a(x)$ для каждого a стремится к $\Phi_M(x)$, когда $\|x\| \to \infty$.

Наша задача — изучать отображение Φ_a при $a > 0$. Для этого определим множество $\mathcal{U}_M \subset \mathbb{R}^N$ как конус всех положительных линейных комбинаций вектор-столбцов матрицы $-M$ и координатных векторов $e_1, \ldots, e_N$.

---

[1] Сокращение от linear complementary problem. — *Прим. изд. ред.*

[2] Linear programming problem. — *Прим. изд. ред.*

Напомним, что диффеоморфизмом называется отображение, имеющее обратное и дифференцируемое вместе с ним. Диффеоморфизм называется аналитическим, если в окрестности каждой точки он может быть представлен сходящимся степенным рядом.

Теорема (о регуляризации задачи LPP). *Для каждого* $a > 0$ *образ отображения* $\Phi_a : \mathbb{R}^N \to \mathbb{R}^N$ *равен* $\mathcal{U}_M$, *причём* $\Phi_a : \mathbb{R}^N \to \mathcal{U}_M$ — *аналитический диффеоморфизм.*

Эта теорема доказана в § 5.

Замечание 1. На самом деле специальный вид матрицы M в доказательстве не используется. Доказательство и теорема справедливы для любой квадратной матрицы, удовлетворяющей условию $\langle Mx, x\rangle \geqslant 0$. То же самое верно и в отношении предложения из § 5.

Замечание 2. Из теоремы следует, что для данной матрицы M при любых $q \in \mathcal{U}_M$ и $a > 0$ уравнение $\Phi_M(x) = q$ имеет единственное решение. Оно аналитически зависит от исходных данных (A, b, c) задачи LPP и параметра a при $q \in \mathcal{U}_M$ и $a > 0$.

Скажем, что задача LPP с исходными данными (A, b, c) «крайне некорректна», если $q \in \partial\mathcal{U}_M$ $(\partial\mathcal{U}_M = \bar{\mathcal{U}}_M \setminus \mathcal{U}_M)$. Такая терминология имеет следующее объяснение. Задача LPP разрешима тогда и только тогда, когда $q \in \bar{\mathcal{U}}_M$ (см. Коттл [5]). Поэтому если $q \in \partial\mathcal{U}_M$, то малое возмущение исходных данных может привести к тому, что задача не будет иметь решения. Из предыдущей теоремы следует, что, выбрав произвольное $a > 0$, можно разрешить особенности во всех задачах LPP, не являющихся крайне некорректными.

Метод Ньютона для отображения $\Phi_a - q$ состоит в построении кривых $\Phi_a^{-1}(q\hat{q})$, где $q\hat{q}$ — отрезок с начальным значением $\hat{q}$ и конечным q. Метод Данцига состоит в построении кривой $\Phi_a^{-1}(q\hat{q})$ для $a = 0$. Таким образом, при $a \to 0$ метод Ньютона с подходящим шагом стремится к алгоритму Данцига.

Удобно ввести в рассмотрение отображения $\hat{\Phi}_a : \mathbb{R}^N \to \mathbb{R}^N$, где $\hat{\Phi}_a(x) = \Phi_a(x) - \Phi_a(0)$. При подходе a к нулю отображение $\hat{\Phi}_a$ аппроксимирует задачу LPP. Кроме того, как показано в § 5, теорему А можно применять для решения уравнения $\hat{\Phi}_a(x) = q$ при всех a, таких что $a > c\|q\|$. Поэтому первоначальную задачу LPP можно получить, устремив a к нулю.

С предыдущими результатами так или иначе связаны работы Мангасаряна [31], Вержбицкого [53], Кармаркара [24], Гилла — Марри — Сондерза — Томлина — Райта [17], Блам [3], Ренегара [41] и Мегидо — Шуба [36].

Имеется много других работ, связанных с темой настоящей статьи. В частности, по поводу информационной теории сложности см. недавний обзор Вожьняковского [55]. В статье Смейл [51] приведён большой список литературы. Из работ, не указанных там, но содержащих интересные близкие результаты, отметим статьи Пана [39], Вонга [54], Гао [13, 14] и Райта [56].

Я признателен многим математикам за помощь в работе над этой статьей. В частности, я извлёк большую пользу из бесед с Диком Коттлом о задаче LCP. Ленор Блам своей MSRI-лекцией о числе обусловленности в линейном программировании с привлечением задачи LCP помогла мне осознать роль этой задачи. Вообще её замечания, а также замечания Джима Карри и Фенга Гао были очень полезными.

Особенно важными для меня на всех этапах работы были беседы с Джимом Ренегаром и Майком Шубом. С Майком Шубом такое сотрудничество продолжается у нас уже многие годы.

§ 1

Пусть α — определённый выше инвариант и $\varphi_d(r) = \sum_{j=0}^{d} r^j$. Для доказательства теоремы В покажем, что верна

Теорема 1. $\alpha(r, \varphi_d) \leqslant 1$ для всех $r \geqslant 0$.

Доказательство. Рассмотрим сначала предельный случай $d = \infty$: $\varphi_\infty(r) = \varphi(r) = 1/(1 - r)$. Тогда

$$\frac{\varphi^{(k)}(r)}{k!} = \left(\frac{1}{1-r}\right)^{k+1},$$

и из определения α следует, что $\alpha(r, \varphi) \equiv 1$.

Пусть, далее,

$$Q_p(d) = \sum_{j=p}^{d} \binom{j}{p},$$

где p, d, j — неотрицательные целые числа, причём $d \geqslant p$ (напомним, что биномиальный коэффициент $\binom{j}{p} = 0$, если $j < p$).

Лемма 1.

$$Q_p(d) = \frac{(d+1)\, d\, (d-1) \,\ldots\, (d-p+1)}{(p+1)!}.$$

Доказательство. Заметим, что $Q_p(d)$ — многочлен от d степени $p+1$. Поэтому достаточно проверить формулу для $p+2$ значений d. Но $Q_p(d)=0$ при $d=0, 1, \ldots, p-1$, $Q_p(p)=1$ и $Q_p(p+1)=p+2$, ч. т. д.

Лемма 2. $(m+1)Q_p(d)\geqslant(d+1)Q_p(m)$ *для* $0\leqslant m\leqslant d$.

Используя лемму 1, можно переписать доказываемое неравенство в виде

$$\frac{(m+1)(d+1)\ldots(d-p+1)}{p!}\geqslant\frac{(d+1)(m+1)\ldots(m-p+1)}{p!},$$

а тогда оно очевидно.

Введём для каждого $k=0, 1, 2, \ldots, d$ функцию

$$\psi_k(x)=\frac{\varphi^{(k)}}{k!}(x)=\sum_{l=0}^{d-k}\binom{l+k}{k}x^l=\sum_{j=0}^{d}\binom{j}{k}x^{j-k}.$$

В этих обозначениях теорема 1 сводится к оценкам $\psi_0^{k-1}\psi_k\leqslant\leqslant\psi_1^k$, $k=1, 2, \ldots$. В случае $k=1$ оценка тривиальна. Применим индукцию по k. Таким образом, нам надо показать, что $\psi_0\psi_k\leqslant\psi_1\psi_{k-1}$ при всех $k=2, 3, \ldots$ и $x>0$.

Умножив обе части последнего неравенства на x^k, перепишем его в виде

$$\sum_0^d x^i\sum_0^d\binom{j}{k}x^j\leqslant\sum_0^d ix^i\sum_0^d\binom{j}{k-1}x^j.$$

Произведение в левой части неравенства равно $\sum\limits_{m=0}^{2d}c_mx^m$, в правой $\sum\limits_{m=0}^{2d}\sigma_mx^m$. Остаётся доказать, что $c_m\leqslant\sigma_m$ для любого m.

Лемма 3. $c_m=\sigma_m$ *для* $m\leqslant d$.

Это следует из тождества $\alpha(r, \varphi)\equiv 1$. (Заметим, что можно считать $d=\infty$.)

Вычислим коэффициенты c_m и σ_m. Если $m\leqslant d$, то

$$c_m=\sum_{j=0}^{m}\binom{j}{k}=Q_k(m),\quad \sigma_m=\sum_{j=0}^{m}(m-j)\binom{j}{k-1}.$$

Если $d<m\leqslant 2d$, то

$$c_m=\sum_{j=m-d}^{d}\binom{j}{k},\quad \sigma_m=\sum_{j=m-d}^{d}(m-j)\binom{j}{k-1}.$$

Пусть $P_{k-1}(q)=\sum\limits_{j=0}^{q}j\binom{j}{k-1}$. Из леммы 3 и предыдущих вычислений следует формула $Q_k(m)=mQ_{k-1}(m)-P_{k-1}(m)$, справедливая для *всех* m, так как она не зависит от d.

Чтобы закончить доказательство теоремы 1, достаточно показать, что при $d<m\leqslant 2d$ имеет место неравенство $\sigma_m-c_m\geqslant 0$, т. е.

$$mQ_{k-1}(d)-P_{k-1}(d)-mQ_{k-1}(m-d-1)\\+P_{k-1}(m-d-1)-Q_k(d)+Q_k(m-d-1)\geqslant 0.$$

Используя указанную выше формулу, перепишем это неравенство в виде

$$(m-d)Q_{k-1}(d)-(d+1)Q_{k-1}(m-d-1)\geqslant 0,$$

или

$$(m_1+1)Q_{k-1}(d)\geqslant(d+1)Q_{k-1}(m_1),\quad\text{при}\quad 0\leqslant m_1\leqslant d-1,$$

где $m_1=m-d-1$.

Но последняя оценка верна в силу леммы 2. Это и доказывает теорему 1.

Докажем теперь теорему В:

Теорема 2.

$$\gamma(z, f)\leqslant\|Df(z)^{-1}\|\|f\|\varphi_d'(\|z\|)^2/\varphi_d(\|z\|).$$

Доказательство. Сначала заметим, что

$$D^kf(z)=\sum_{i=0}^{\infty}i(i-1)\ldots(i-k+1)a_iz^{i-k},$$

где a_iz^{i-k} есть k-линейная форма, полученная из i-линейной формы a_i подстановкой z вместо $i-k$ её аргументов. Отсюда следует, что

$$\left\|\frac{D^kf(z)}{k!}\right\|\leqslant\|f\|\frac{\varphi_d^{(k)}(\|z\|)}{k!}.$$

Поскольку $\|Df(z)^{-1}\|\|Df(z)\|\geqslant\|Df(z)^{-1}Df(z)\|=1$, то, используя предыдущее неравенство в случае $k=1$, получим $\|Df(z)^{-1}\|$

$\times \|f\| \varphi_d'(\|z\|) \geqslant 1$. Для $k > 1$ в силу теоремы 1 имеем

$$\frac{\varphi_d^{(k)}}{k!}(r) \leqslant \frac{\varphi_d'(r)^k}{\varphi_d(r)^{k-1}}, \quad \left\|\frac{D^k f(z)}{k!}\right\| \leqslant \|f\| \frac{\varphi_d'(\|z\|)^k}{\varphi_d(\|z\|)^{k-1}}.$$

Теперь воспользуемся этими оценками:

$$\begin{aligned}\gamma(z, f) &\leqslant \sup_{k>1} \left\| Df(z)^{-1} \frac{D^k f(z)}{k!} \right\|^{1/(k-1)} \\ &\leqslant \sup_{k>1} \left(\| Df(z)^{-1} \| \left\| \frac{D^k f(z)}{k!} \right\| \right)^{1/(k-1)} \\ &\leqslant \sup_{k>1} \left(\| Df(z)^{-1} \| \|f\| \frac{\varphi_d'(\|z\|)^k}{\varphi_d(\|z\|)^{k-1}} \right)^{1/(k-1)} \\ &\leqslant \frac{\varphi_d'}{\varphi_d}(\|z\|) \sup_{k>1} \left(\| Df(z)^{-1} \| \|f\| \varphi_d'(\|z\|) \right)^{1/(k-1)}.\end{aligned}$$

Супремум здесь берётся по корням $(k-1)$-й степени из числа, которое не зависит от k и не меньше 1. Поэтому он достигается при $k = 2$, что и доказывает теорему 2.

Замечание. Я пробовал доказать «L^2-версию» теоремы 1, но это мне не удалось.

Задача. Верно ли, что $\alpha(r, \psi_d) \leqslant 1$ при всех $r \geqslant 0$, где $\psi_d(r) = \left(\sum_{i=0}^{d} r^{2i} \right)^{1/2}$?

§ 2

Цель этого параграфа — доказать теорему о вычислительной сложности случайного алгоритма, дав вероятностную оценку функции σ_f (теорема сформулирована во введении).

Пусть B_f — площадь множества точек $z \in D_3$, таких что $\alpha(z, f) < \alpha_0$. Следующее (элементарное) утверждение аналогично предложению 3 из § 2 работы Шуба — Смейла [44]:

Предложение 1. $\sigma_f = 9\pi/B_f$. (*Заметим, что площадь* D_3 *равна* 9π.)

Далее мы докажем предложение, которое показывает, что наш критерий для приближённых нулей второго рода (теорема C) справедлив также и для приближённых нулей (первого рода):

Предложение 2. *Существует константа* $c > 0$ *со следующим свойством. Пусть* f: $\mathbf{E} \to \mathbf{F}$ — *аналитическое отображение одного банахова пространства в другое и* ζ — *элемент пространства* $\mathbf{E}$, *такой что* $f(\zeta) = 0$. *Тогда если* $z \in \mathbf{E}$ *удовлетворяет неравенству* $\|z - \zeta\| < c/\gamma(\zeta, f)$, *то* $\alpha(z, f) < \alpha_0$ *и* z *есть приближённый нуль* f.

Доказательство. Пусть $c < (1 - \sqrt{2}/2)/2$ и z удовлетворяет условию предложения. Тогда, согласно лемме 2c из § 3 работы [P. E.], выполняется неравенство $\gamma(z) < c_1 \gamma(\zeta)$ для подходящей константы c_1. Используя предложение 3 из § 4 той же работы, получим, что $\beta(z) < c_2 \|z - \zeta\|$ для подходящей константы c_2. Выберем

$$c < \min\left(\frac{1}{2}\left(1 - \frac{\sqrt{2}}{2}\right), \frac{\alpha_0}{c_1 c_2} \right).$$

Тогда $\alpha(z, f) = \beta(z)\gamma(z) < c_1 c_2 \|z - \zeta\| \gamma(\zeta) < \alpha_0$, ч. т. д.

Введём для каждого многочлена f число

$$\rho_f = \min_{\substack{\theta \\ f'(\theta)=0}} |f(\theta)|.$$

Пусть $\tau_f = \min(1, \rho_f)$. Справедлива (см. Смейл [50, с. 29])

Лемма 1. $\operatorname{meas}\{f \in P_d(1) \mid \tau_f < \rho\} < d\rho^2$.

Следующее утверждение было доказано в статье Шуба — Смейла [43, p. 124]. Здесь даётся новое доказательство, не использующее теорию функций Шлихта.

Предложение 3. *Существует константа* $K > 0$, *такая что* $B_f > K\tau_f^2/d^4$ *для всех* $f \in P_d(1)$.

При доказательстве этого предложения нам потребуется несколько лемм. Но сначала отметим тот хорошо известный факт, что если $f \in P_d(1)$ и $f(\zeta) = 0$, то $|\zeta| < 2$. В конечном счёте именно это явилось основанием для выбора числа 3 в теореме, которую мы доказываем в этом параграфе.

Лемма 2. *Если* $f \in P_d(1)$, *то*

$$B_f > c \sum_{\substack{\zeta \\ f(\zeta)=0}} \left(\frac{|f'(\zeta)|}{\varphi_d'(|\zeta|)} \right)^2 \left(\frac{\varphi_d(|\zeta|)}{\varphi_d'(|\zeta|)} \right)^2.$$

Доказательство. Из предложения 2 следует, что

$$B_f > c \sum_{\substack{\zeta \\ f(\zeta)=0}} \frac{1}{\gamma(\zeta, f)^2}$$

для некоторой константы $c > 0$ (ср. с леммой 1 из § 3). Теперь применим теорему B. Так как $f \in P_d(1)$, то $\|f\| = 1$. Это и доказывает лемму 2.

Лемма 3. *$\varphi_d'(r)/\varphi_d(r) \leqslant d$ при всех $r \geqslant 0$.*

Доказательство. $\sum\limits_{i=0}^{d} i r^{i-1} \leqslant d \sum\limits_{i=0}^{d} r^{i-1} \leqslant d \sum\limits_{i=0}^{d} r^{i}$.

Следующее утверждение получается применением неравенства, связывающего между собой среднее арифметическое и среднее геометрическое.

Лемма 4.

$$\frac{1}{d} \sum_{\substack{\zeta \\ f(\zeta)=0}} \left| \frac{f'(\zeta)}{\varphi_d'(|\zeta|)} \right|^2 \geqslant \prod_{\substack{\zeta \\ f(\zeta)=0}} \left(\frac{|f'(\zeta)|}{\varphi_d'(|\zeta|)} \right)^{2/d}.$$

Лемма 5. *Существует положительная константа c, такая что*

$$\prod_{\substack{\zeta \\ f(\zeta)=0}} \varphi_d'(|\zeta|)^{1/d} \leqslant c d^{5/2} \quad \text{для всех} \quad f \in P_d(1).$$

Доказательство. Воспользуемся одной теоремой Шпехта (см. Марден [32, p. 129]), из которой легко вытекает следующее утверждение. Если многочлен $f \in P_d(1)$ имеет корни $\zeta_1, \ldots, \zeta_d$, то для любого подмножества индексов $i_1, \ldots, i_k$, из $\{1, \ldots, d\}$ выполняется неравенство $|\zeta_{i_1}\zeta_{i_2} \ldots \zeta_{i_k}| \leqslant \sqrt{d+1}$.

Отсюда следует, что

$$\prod_{\substack{\zeta \\ f(\zeta)=0}} \varphi_d'(|\zeta|) = \prod_{i=1}^{d} \sum_{j=1}^{d} j |\zeta_i|^{j-1}$$

$$= \sum_{\substack{(i_1, \ldots, i_d) \\ 1 \leqslant i_k \leqslant d}} i_1 |\zeta_1|^{i_1-1} i_2 |\zeta_2|^{i_2-1} \ldots i_d |\zeta_d|^{i_d-1}$$

$$\leqslant (\sqrt{d+1})^d \sum i_1 i_2 \ldots i_d = (\sqrt{d+1})^d \sum_1^d i_1 \sum_1^d i_2 \ldots \sum_1^d i_d$$

$$\leqslant (\sqrt{d+1})^d \left(\frac{d(d+1)}{2} \right)^d.$$

Извлекая корень d-й степени, получим

$$\prod_{\substack{\zeta \\ f(\zeta)=0}} (\varphi_d'(|\zeta|))^{1/d} \leqslant c d^{5/2}, \text{ ч. т. д.}$$

Лемма 6. *Пусть $f \in P_d(1)$ и D_f — дискриминант многочлена f* (Ленг [29]). *Тогда $|D_f|^{1/d} \geqslant d\tau_f$.*

По поводу доказательства см. Шуб — Смейл [44, предложение 7 из § 2].

Леммы 2—6 позволяют доказать предложение 3. Действительно, в силу лемм 2 и 3

$$B_f > c \frac{1}{d^2} \sum_{\substack{\zeta \\ f(\zeta)=0}} \left(\frac{|f'(\zeta)|}{\varphi_d'(|\zeta|)} \right)^2.$$

Но тогда, согласно леммам 4 и 5,

$$B_f > c_1 \frac{1}{d} \frac{1}{d^5} \prod_{\substack{\zeta \\ f(\zeta)=0}} |f'(\zeta)|^{2/d}.$$

Поскольку

$$D_f = \prod_{\substack{\zeta \\ f(\zeta)=0}} (f'(\zeta))$$

(Ленг [29]), то, применив лемму 6, получим оценку $B_f > c_2 \tau_f^2 / d^4$, доказывающую предложение 3.

Теперь докажем теорему о вычислительной сложности случайного алгоритма. Пусть G_d: $\mathbb{Z}^+ \to [0, 1]$ — функция, определённая формулой

$$G_d(\sigma) = \text{meas}\{f \in P_d(1) \mid \sigma_f > \sigma\}.$$

Тогда нашу теорему можно переформулировать в виде $G_d(\sigma) < cd^5/\sigma$. Но ввиду предложений 1 и 3

$$G_d(\sigma) = \text{meas}\{f \in P_d(1) \mid 9\pi/B_f > \sigma\},$$

$$G_d(\sigma) \leqslant \text{meas}\{f \in P_d(1) \mid cd^4/\tau_f^2 > \sigma\},$$

или

$$G_d(\sigma) \leqslant \text{meas}\{f \in P_d(1) \mid \tau_f < (cd^4/\sigma)^{1/2}\}.$$

Отсюда и из леммы 1 следует $G_d(\sigma) \leqslant cd^5/\sigma$, ч. т. д.

§ 3

Цель этого параграфа — доказать теорему о средней площади множества приближённых нулей.

По определению,

$$\underset{f \in P_d(1)}{\text{average}}\, A_f = \left(\frac{1}{\pi} \right)^d \int\limits_{|a_{d-1}| \leqslant 1} \ldots \int\limits_{|a_0| \leqslant 1} A_f \, da_0 \ldots da_d.$$

Нашей основной задачей будет получение оценки

$$\int\limits_{|a_0|\leqslant 1} A_f\, da_0 \geqslant K \int\limits_{\substack{|\zeta|<1/2\\ \zeta\in\mathbb{C}}} |f'(\zeta)|^4\, d\zeta. \tag{I}$$

После этого для доказательства теоремы останется лишь показать, что

$$\int\limits_{|a_{d-1}|\leqslant 1} \cdots \int\limits_{|a_1|\leqslant 1} \int\limits_{|\zeta|<1/2} |f'(\zeta)|^4\, d\zeta > L(\pi)^d. \tag{II}$$

Здесь K и L — положительные константы, не зависящие от d. Некоторые оценки этих констант можно извлечь из доказательства.

Займёмся сначала неравенством (I). Пусть $\gamma_\zeta = \max(1, \gamma(\zeta, f))$, где $\gamma(\zeta, f)$ — величина, определённая во введении.

Лемма 1.

$$A_f > \left(\frac{3-\sqrt{7}}{2}\right)^2 \pi \sum_{\substack{\zeta\\ f(\zeta)=0\\ |\zeta|<1/2}} \gamma_\zeta^{-2}.$$

Доказательство. Пусть $\hat{\Omega}_f(\zeta)$ — множество приближённых нулей z второго рода, отвечающих фактическому нулю ζ и таких, что $|z-\zeta|<1/2$. Множества $\hat{\Omega}_f(\zeta)$ не пересекаются и

$$\Omega_f \supset \bigcup_{\substack{\zeta\\ |\zeta|<1/2}} \hat{\Omega}_f(\zeta).$$

Из теоремы C следует, что если $|z-\zeta|<(3-\sqrt{7})/2\gamma_\zeta$, то $z\in\hat{\Omega}_f(\zeta)$. Поэтому площадь множества $\hat{\Omega}_f(\zeta)$ больше или равна $\pi\left((3-\sqrt{7})/2\gamma_\zeta\right)^2$. Все это вместе и даёт лемму 1.

Теперь докажем следующую формулу замены переменной:

Лемма 2. *Пусть* $f_0 = \sum_{i=1}^{d} a_i z^i$ *(так что* $f_0(0)=0$*), где* $|a_i|\leqslant 1$, *и* $f=f_0+a_0$, *где* $a_0\in\mathbb{C}$, $|a_0|\leqslant 1$. *Тогда*

$$\int\limits_{|a_0|\leqslant 1} \sum_{\substack{\zeta\\ f(\zeta)=0\\ |\zeta|<1/2}} \gamma_\zeta^{-2}\, da_0 = \int\limits_{\substack{\zeta\in f_0^{-1}(D_1)\\ |\zeta|<1/2}} \gamma_\zeta^{-2} |f_0'(\zeta)|^2\, d\zeta.$$

Доказательство. Заметим, что если $|\zeta|<1/2$, то $|f_0(\zeta)|<1$. Пусть $V=f_0(D_{1/2})$ и $g_i\colon V\to D_{1/2}$, $i=1, \ldots, d$ — сужения на V ветвей f_0^{-1}. Тогда $U_i = g_i(V)$ — непересекающиеся измеримые подмножества в $D_{1/2}$, объединение которых равно $D_{1/2}$, причём $f_0 g_i(w) = w$ для любого $w\in V$.

Ясно, что $\zeta = g_i(-a_0)$ тогда и только тогда, когда $f_0(\zeta) = -a_0$, т. е. $f(\zeta)=0$, где $f=f_0+a_0$. Следовательно,

$$\begin{aligned}\int\limits_{\zeta\in D_{1/2}} \gamma_\zeta^{-2}|f_0'(\zeta)|^2\, d\zeta &= \sum_{i=1}^{d} \int\limits_{\zeta\in U_i} \gamma_\zeta^{-2}|f_0'(\zeta)|^2\, d\zeta\\ &= \sum_{i=1}^{d} \int\limits_{a_0\in V} (\gamma(g_i(-a_0)))^{-2}\, da_0\\ &= \int\limits_{a_0\in V} \sum_{f(\zeta)=0} \gamma(\zeta)^{-2}\, da_0\\ &= \int\limits_{|a_0|\leqslant 1} \sum_{\substack{f(\zeta)=0\\ |\zeta|<1/2}} \gamma(\zeta)^{-2}\, da_0.\end{aligned}$$

Лемма 3. *Существует константа* $K_1>0$, *такая что если* $|\zeta|<1/2$, *то* $\gamma_\zeta^{-2}\geqslant K_1|f'(\zeta)|^2$.

Доказательство. Предположим, что $\gamma_\zeta = \gamma(\zeta, f)$. Тогда по теореме B

$$\gamma_\zeta \leqslant \frac{\varphi_d'(|\zeta|)^2}{\varphi_d(|\zeta|)} \|f\| \frac{1}{|f'(\zeta)|}.$$

Отсюда

$$\gamma_\zeta^2 \leqslant \frac{\varphi_d'(1/2)^4}{\varphi_d(1/2)^2} \frac{1}{|f'(\zeta)|^2}$$

для $f\in P_d$ и $|\zeta|\leqslant 1/2$.

Случай $\max(1, \gamma(\zeta, f)) = 1$ легко разобрать непосредственно. Это доказывает лемму 3 и позволяет оценить константу K_1 явно.

Из лемм 1—3 вытекает неравенство (I) (заметим, что $f'=f_0'$). Неравенство (II) получается при помощи следующей оценки:

Лемма 4. *Для любого комплексного многочлена* g

$$\int\limits_{\substack{|z|<R\\ z\in\mathbb{C}}} g(z)\overline{g(z)}\, dz \geqslant |g(0)|^2 \pi R^2.$$

Доказательство. Воспользуемся полярными координатами. Нечётные степени r дают нулевой вклад, чётные — положительный, откуда и следует наше неравенство.

Теперь докажем неравенство (II). Здесь $g(\zeta)=f'(\zeta)^2$ и $g(0)=(a_1)^2$. Сначала применим лемму 4 к самому правому выполняемому первым интегралу в левой части (II). Второе интегрирование проводится по множеству $|a_1|\leqslant 1$ в полярных координатах. Остальные интегрирования тривиальны. Тем самым доказано неравенство (II), а с ним и теорема о средней площади множества приближенных нулей.

Другое, может быть даже более простое доказательство аналогичной теоремы для приближённых нулей первого рода можно дать, используя теорему А, леммы 1, 3 и следующую лемму:

Лемма 5. *Существуют универсальные положительные константы* δ *и* ε, *такие что если* $f(z)=\sum_{i=0}^{d} a_i z^i$, $|a_i|\leqslant 1$, $|a_0|/|a_1|^2<\varepsilon$ *и* $|z|<\delta$, *то* $\alpha(z, f)<\alpha_0$ *и* z *есть приближённый нуль* f *(первого рода).*

§ 4

Здесь мы докажем теорему о быстродействии глобального метода Ньютона.

Лемма 1. *Существует положительная константа* K *со следующим свойством. Пусть заданы любые* (z_0, f), *такие что* $\alpha(z_0, f)<\alpha_0$. *Положим,* $z_l=N_f^l(z_0)$, $l=1, 2, \ldots$, *и* $\zeta=\lim\limits_{l\to\infty} z_l$. *Тогда* $\gamma(\zeta, f)\leqslant K\gamma(z_0, f)$.

Доказательство. Пусть $\alpha_l=\alpha(z_l)$, $\psi_l=\psi(\alpha_l)$, $l=1$, $2, \ldots$, где $\psi(r)=2r^2-4r+1$. Тогда α_l убывает при увеличении l и $\alpha_l\leqslant(1/2)^{2^{l-1}}\alpha(z_0)$ (см. [P. E.], § 4, предложение 1 и § 3, предложение 2). Следовательно, $1/\psi_l$ убывает к 1 с ростом l, причём очень быстро (см. начало § 4 в [P. E.]).

Далее, положим

$$K_0=\frac{1}{(1-\alpha(z_0))\,\psi(\alpha(z_0))},\qquad K_l=\frac{1}{(1-\alpha_l)\,\psi_l},\quad l=1, 2, \ldots.$$

Согласно лемме 2с из § 3 работы [P. E.], справедливо неравенство $\gamma(z_l)\leqslant K_{l-1}\gamma(z_{l-1})$, откуда $\gamma(z_l)\leqslant\gamma(z_0)\prod\limits_{j=0}^{l-1}K_j$. Поэтому достаточно показать, что величина $\prod\limits_{j=0}^{l-1}K_j$ возрастает при $l\to\infty$ до некоторой конечной константы K. Но это следует из равенств

$$\log\prod_0^{l-1}K_j=\sum_0^{l-1}\log K_j=\sum_0^{l-1}-\log\psi_j-\sum_0^{l-1}\log(1-\alpha_j)$$

и приведённой выше оценки инварианта α_l. Лемма доказана.

Выберем константу $K_2<1\frac{3}{4}$, как в предложении 1 из § 2 статьи [P. E.], и положим $L_0=KK_2\alpha_0$, где K — константа из леммы 1.

Лемма 2. *Как и в лемме* 1, *пусть* $\alpha(z_0, f)<\alpha_0$, $z_l=N_f^l(z_0)$ *и* $\zeta=\lim\limits_{l\to\infty} z_l$. *Тогда* $\|z_l-\zeta\|\,\gamma(\zeta)<(1/2)^{2^{l-1}}L_0$.

Доказательство. Согласно предложению 1 из § 3 [P. E.],

$$\|z_l-\zeta\|<K_2\left(\frac{1}{2}\right)^{2^{l-1}}\|z_1-z_0\|.$$

Поэтому, по лемме 1,

$$\|z_l-\zeta\|\,\gamma(\zeta)<K_2\left(\frac{1}{2}\right)^{2^{l-1}}\|z_1-z_0\|\,\gamma(z_0)\,K.$$

Утверждение леммы следует теперь из определения L_0 и того факта, что $\|z_1-z_0\|\gamma(z_0)=\alpha(z_0)<\alpha_0$.

Теперь докажем теорему. Достаточно показать, что $\alpha(z_i, f-w_{i+1})<\alpha_0$, $i=0, 1, \ldots, n-1$; это мы сделаем по индукции. Прежде всего заметим, что

$$\alpha(z_i, f-w_{i+1})\leqslant\gamma(z_i)\,\|Df(z_i)^{-1}(f(z_i)-w_i)\|+\gamma(z_i)\,\|Df(z_i)^{-1}(w_i-w_{i+1})\|.$$

Поэтому достаточно убедиться в справедливости для некоторого фиксированного l следующих неравенств:

(a) $\gamma(z_i)\,\|Df(z_i)^{-1}(f(z_i)-w_i)\|\leqslant\alpha_0\left(\frac{1}{2}\right)^{2^{l-1}}$,

(b) $\gamma(z_i)\,\|Df(z_i)^{-1}(w_i-w_{i+1})\|\leqslant\alpha_0\left(1-\left(\frac{1}{2}\right)^{2^{l-1}}\right)$.

При $i=0$ неравенство (a) верно, так как $f(z_0)=w_0$. Пусть $i>0$. Используем предложение 1b из § 4 [P. E.] с $a=1/2$. Получим

$$\alpha(z_i, f-w_i)\leqslant(1/2)^{2^{l-1}}\alpha(z_{i-1}, f-w_i),$$

что по предположению индукции меньше $(1/2)^{2^{l-1}}\alpha_0$. Этим неравенство (a) доказано.

Для доказательства неравенства (b) положим $\zeta_i = \lim\limits_{k\to\infty} N^k_{f-w_i} \times \times (z_{i-1})$ (так что $f(\zeta_i) = w_i$) и $\tau = \|z_i - \zeta_i\| \gamma(\zeta_i)$. В силу леммы 2 справедлива оценка $\tau \leqslant (1/2)^{2^{l-1}} L_0$, и можно показать, что для l, бо́льших некоторого малого l_0 ($=4$, например), мы имеем $\tau \leqslant 1 - \sqrt{2}/2$. По лемме 2c из § 3 [P. E.]

$$\gamma(z_i) \leqslant \gamma(\zeta_i) \left(\frac{1}{1-\tau}\right) \frac{1}{\psi(\tau)}.$$

Кроме того, согласно лемме 2b из § 2 той же статьи

$$\begin{aligned} &\| Df(z_i)^{-1}(w_i - w_{i+1})\| \\ &\quad \leqslant \| Df(z_i)^{-1} Df(\zeta_i)\| \| Df(\zeta_i)^{-1}(w_i - w_{i+1})\| \\ &\quad \leqslant \frac{(1-\tau)^2}{\psi(\tau)} \beta(\zeta_i, f) \frac{\| w_i - w_{i+1}\|}{\| w_i \|}. \end{aligned}$$

Поэтому для установления неравенства (b) достаточно показать, что

$$\gamma(\zeta_i) \frac{1-\tau}{\psi(\tau)^2} \beta(\zeta_i, f) \frac{\| w_i - w_{i+1}\|}{\| w_i \|} < \alpha_0 \left(1 - \left(\frac{1}{2}\right)^{2^{l-1}}\right),$$

или

$$\frac{\alpha(\zeta_i, f)}{\| w_i \|} \frac{(1-\tau)}{\psi(\tau)^2} \Delta \| f(z_0)\| < \alpha_0 \left(1 - \left(\frac{1}{2}\right)^{2^{l-1}}\right),$$

или

$$M\Delta \| f(z_0)\| \frac{(1-\tau)}{\psi(\tau)^2} < \alpha_0 \left(1 - \left(\frac{1}{2}\right)^{2^{l-1}}\right),$$

или

$$\frac{1}{c}\left[\frac{1-\tau}{\psi(\tau)^2}\right]_{\tau = (1/2)^{2^{l-1}} L_0} \leqslant \alpha_0 \left(1 - \left(\frac{1}{2}\right)^{2^{l-1}}\right)$$

при некоторых $c > 0$ и l. Но последнее неравенство очевидно и определяет выбор констант c и l в теореме о быстродействии глобального метода Ньютона. Доказательство завершено.

§ 5

Здесь мы докажем теорему о регуляризации задачи LPP. Затем обсудим связь метода Ньютона с алгоритмами решения этой задачи.

Для доказательства теоремы нам потребуется ряд лемм.

Лемма 1. *Производная $D\Phi_a(x)$: $\mathbb{R}^N \to \mathbb{R}^N$ есть обратимое отображение, и норма обратного отображения ограничена константой, не зависящей от M, а именно*

$$\| D\Phi_a(x)^{-1}\| \leqslant 4 \max_i \left(\frac{x_i^2 + a^2}{a^2}\right).$$

Замечание. Уже это показывает непрерывную зависимость решения возмущённой задачи LPP от исходных данных (A, b, c).

Доказательство. Прежде всего заметим, что матрица M кососимметрична, т. е. $M^T x = -Mx$, где M^T — матрица, транспонированная к M. Это следует из того, что $M = \begin{bmatrix} 0 & -A^T \\ A & 0 \end{bmatrix}$. Таким образом, $\langle Mv, v\rangle = 0$ для любого $v \in \mathbb{R}^N$, где $\langle \cdot, \cdot \rangle$ — обычное скалярное произведение в $\mathbb{R}^N$. Этот факт будет играть решающую роль во всех последующих рассуждениях (хотя достаточно и более слабого условия $\langle Mv, v\rangle \geqslant 0$).

Пусть до конца этого параграфа

$$\Delta = \operatorname{diag\,matrix} \left(\frac{x_i}{(x_i^2 + a^2)^{1/2}}\right)_{i=1}^N.$$

Тогда $2D\Phi_a(x) = I + \Delta + M(I - \Delta)$.

Согласно предыдущему замечанию,

$$\langle 2D\Phi_a(x)(u), (I-\Delta)u\rangle = \langle (I+\Delta)u, (I-\Delta)u\rangle.$$

Применив неравенство Шварца, получим

$$|\langle (I+\Delta)u, (I-\Delta)u\rangle| \leqslant 2\| D\Phi_a(x)(u)\| \| (I-\Delta)(u)\|.$$

Левая часть этого неравенства равна $\sum\limits_i u_i^2 (a^2/(x_i^2 + a^2))$; кроме того, $\|(I-\Delta)u\| \leqslant 2\|u\|$. Поэтому

$$\min_i \left(\frac{a^2}{x_i^2 + a^2}\right) \| u \| \leqslant 4 \| D\Phi_a(x)(u)\|,$$

откуда и следует утверждение леммы.

Лемма 2. *Для всех вещественных чисел a, x, y имеет место равенство*

$$\begin{aligned} &(x-y)^2 - ((x^2+a^2)^{1/2} - (y^2+a^2)^{1/2})^2 \\ &\qquad = 2((xy+a^2)^2 + a^2(x-y)^2)^{1/2} - 2(xy+a^2). \end{aligned}$$

Доказательство сводится к раскрытию скобок и сравнению полученных членов. Заметим, что правая часть этого равенства положительна при $x \neq y$. Поэтому то же самое верно и для его левой части.

Лемма 3. *При* $a > 0$ *отображение* Φ_a: $\mathbb{R}^N \to \mathbb{R}^N$ *является взаимно-однозначным.*

Доказательство. Пусть $\Lambda_x = ((x_1^2+a^2)^{1/2}, \ldots, (x_N^2+a^2)^{1/2})$. Тогда $2\Phi_a(x) = x + \Lambda_x + M(x - \Lambda_x)$ и

$$\begin{aligned} &2\langle \Phi_a(x) - \Phi_a(y), (x - \Lambda_x) - (y - \Lambda_y)\rangle \\ &= \langle x + \Lambda_x - (y + \Lambda_y), (x - \Lambda_x) - (y - \Lambda_y)\rangle \\ &= \sum_{i=1}^{N} (x_i - y_i)^2 - ((x_i^2 + a^2)^{1/2} - (y_i^2 + a^2)^{1/2})^2. \end{aligned}$$

Применение леммы 2 завершает доказательство.

Лемма 4. *Образ отображения* Φ_a: $\mathbb{R}^N \to \mathbb{R}^N$ *лежит в* U_M.

Доказательство. Так как $\Phi_a(x) = \Phi_a^+(x) + M\Phi_a^-(x)$, то $\Phi_a(x)$ является положительной линейной комбинацией координатных векторов $e_1, \ldots, e_N$ и вектор-столбцов матрицы $-M$. Более явно:

$$\varphi_a^+(x_i)\,\varphi_a^-(x_i) = -a^2/4, \quad i = 1, \ldots, N,$$

поэтому если через M_k обозначить k-й столбец матрицы M, то

$$\Phi_a(x) = \sum_k \left(\varphi_a^+(x_k)\, e_k + \frac{a^2}{4\varphi_a^+(x_k)}(-M_k)\right).$$

Теперь рассмотрим задачу LCP, порождённую задачей LPP. Один из результатов Коттла [5, теорема на с. 663] можно сформулировать следующим образом:

Лемма 5. *Для данной матрицы* M *уравнение* $\Phi_M(x) = q$ *имеет решение* x *тогда и только тогда, когда вектор* q *принадлежит замыканию* $\bar{U}_M$ *множества* U_M.

Лемма 6. *Для данных* M *и* $q \in U_M$ *найдётся число* a^*, *такое что при любом* $0 < a < a^*$ *существует решение* x *уравнения* $\Phi_a(x) = q$.

Доказательство леммы 6 использует понятие степени отображения и лемму 5. Пусть $V = V_\delta(q)$ — замкнутый шар радиуса δ с центром q в $\mathbb{R}^N$, где δ выбрано так, что $V_\delta(q) \subset U_M$. Тогда из стандартной теории линейного программирования следует, что ограничение $\Phi_M : \Phi_M^{-1}(V) \to V$ определено и является собственным, а его степень равна 1. Так как Φ_a равномерно аппроксимирует Φ_M, отсюда следует утверждение леммы 6.

Чтобы закончить доказательство теоремы, достаточно показать, что образ отображения Φ_a: $\mathbb{R}^N \to U_M$ совпадает со всем U_M. Это делается следующим образом. Зафиксируем M, q и рассмотрим решение $x(a)$ уравнения $\Phi_a(x(a)) = q$, задаваемое, по лемме 6, формулой $x(a) = \Phi_a^{-1}(q)$ для достаточно малых a. Очевидно, $x(a)$ определено при всех $a < a^*(q)$, где $a^*(q)$ можно считать максимальным возможным. Докажем, что $a^*(q) = \infty$.

Действительно, для каждого q (при фиксированном M) решение $x(a)$ удовлетворяет обыкновенному дифференциальному уравнению, полученному дифференцированием равенства $\Phi_a(x(a)) = q$ по a. Это уравнение имеет вид

$$\frac{dx}{da} = -D\Phi_a(x)^{-1}\frac{\partial \Phi_a}{\partial a}.$$

Здесь

$$2\frac{\partial \Phi_a}{\partial a} = (I - M)\left(\frac{a}{(x_i^2 + a^2)^{1/2}}\right)_{i=1}^{N}$$

и $2D\Phi_a = (I + \Delta) + M(I - \Delta)$, как и выше.

Пусть $u = dx/da$. Тогда u удовлетворяет уравнению

$$(I + \Delta)u + \frac{a}{(x_i^2 + a^2)^{1/2}}\Bigg|_{i=1}^{N} + M\left\{(I - \Delta)u - \frac{a}{(x_i^2 + a^2)^{1/2}}\Bigg|_{i=1}^{N}\right\} = 0.$$

Умножив его скалярно на величину в фигурных скобках, получим

$$\left\langle (I + \Delta)u + \frac{a}{(x_i^2 + a^2)^{1/2}}\Bigg|_i,\ (I - \Delta)u - \frac{a}{(x_i^2 + a^2)^{1/2}}\Bigg|_i\right\rangle = 0,$$

или $\|u\|^2 = \|\Delta u + a/(x_i^2 + a^2)^{1/2}|_i\|^2$. Отсюда

$$\sum_{i=1}^{N} \frac{(au_i - x_i)^2 - (x_i^2 + a^2)}{(x_i^2 + a^2)} = 0,$$

т. е.

$$\sum_{i=1}^{N} \frac{(u_i - x_i/a)^2}{1 + (x_i/a)^2} = N.$$

Таким образом, для каждого i

$$|u_i - x_i/a| \leqslant \sqrt{N}\,(1 + (x_i/a)^2)^{1/2}$$

и, значит, справедлива оценка $|u_i| \leqslant K_1 + K_2|x_i/a|$. Она гарантирует равенство $a^*(q) = \infty$, из которого вытекает сюръективность отображения Φ_a: $\mathbb{R}^N \to U_M$, а тем самым и теорема о регуляризации задачи LPP.

Далее мы покажем, как можно использовать теоремы о приближённых нулях для инициализации алгоритмов решения задач, аппроксимирующих задачу LPP.

Прежде всего заметим, что

$$\Phi_a(0) = \frac{a(I-M)}{2} q_0, \quad q_0 = (1, \ldots, 1),$$

и рассмотрим отображение $\hat{\Phi}_a(x) = \Phi_a(x) - \Phi_a(0)$. Если $\hat{\Phi}_a(x(a)) = q$, то при $a \to 0$ функция $x(a)$ стремится к решению уравнения $\Phi_M(x) = q$ и, следовательно, к решению первоначальной задачи LPP. Кроме того, $x(a)$ удовлетворяет обыкновенному дифференциальному уравнению

$$\frac{dx}{da} = -D\hat{\Phi}_a(x)^{-1} \frac{\partial \hat{\Phi}_a}{\partial a}(x).$$

Чтобы определить начальное условие для этого уравнения (а потом устремить a к 0), достаточно решить уравнение $\hat{\Phi}_a(x) = q$ при каком-нибудь конкретном значении a. Этим мотивируется следующее утверждение:

Предложение. *Если $a \geqslant 8\|q\|/\alpha_0$, то $0 \in \mathbb{R}^N$ является приближённым нулем отображения $\hat{\Phi}_a - q$.*

Отсюда следует, что при $a = 8\|q\|/\alpha_0$ можно с большой точностью решить уравнение $\hat{\Phi}_a(x) = q$, выполнив несколько (например, 5) ньютоновых итераций с начальной точкой в 0. Напомним, что $\alpha_0 > 1/8$.

Для доказательства предложения воспользуемся теоремой А и покажем, что $\alpha(0, \hat{\Phi}_a - q) < \alpha_0$.

Лемма 7. *Пусть*

$$\varphi(x) = \varphi_a^{\pm}(x) = \frac{x \pm (x^2 + a^2)^{1/2}}{2}.$$

Тогда $\alpha(x, \varphi) \leqslant 4$ и для всех $k > 1$

$$\left(\frac{2|\varphi^{(k)}(0)|}{k!}\right)^{1/(k-1)} \leqslant \frac{4}{a}.$$

Доказательство. Применим теорию функций Шлихта (в частности, результаты Лёвнера, касающиеся гипотезы Бибербаха). Оценить высшие производные можно и элементарным способом, но это не тривиально. Ниже рассматривается случай функции φ_a^+; для φ_a^- рассуждения аналогичны.

Заметим, что функция, обратная к φ: $\mathbb{R} \to \mathbb{R}^+$, задаётся формулой $g(y) = y - a^2/4y$. Поэтому радиус сходимости обратной функции относительно точки y равен y, т. е. $\varphi(x)$, в терминах переменной x.

Далее, согласно Смейлу [50], Шубу — Смейлу [43] или Смейлу [51, p. 105], для функций φ довольно общего вида имеет место неравенство $\alpha(x, \varphi) \leqslant 4|\varphi(x)|/r(\varphi_x^{-1})$ (это доказывается при помощи результатов Лёвнера). Здесь $r(\varphi_x^{-1})$ — радиус сходимости функции φ^{-1} относительно точки $\varphi(x)$. В нашем случае отсюда следует, что $\alpha(x, \varphi) \leqslant 4$. Последнее утверждение леммы вытекает из равенств $\varphi(0) = a/2$, $\varphi'(0) = 1/2$.

Недавно Элиз Коли показала, что, непосредственно вычисляя k-ю производную, можно получить оценку $1/4a$, а не $4/a$. Это позволяет усилить сформулированное предложение, заменив 8 на 1/2.

Лемма 8. *Если матрица M кососимметрична, то $\|(I+M)^{-1} \cdot (I-M)\| = 1$.*

Доказательство. Пусть $(I+M)^{-1}(I-M)u = v$. Тогда $(u-v) - M(u+v) = 0$ и $\langle u-v, u+v \rangle = 0$, откуда $\|u\| = \|v\|$, ч. т. д.

Лемма 9. *Пусть G: $\mathbb{R}^N \to \mathbb{R}^N$ — аналитическое отображение вида $G(x) = (g_1(x_1), \ldots, g_N(x_N))$. Тогда*

$$\|D^k G(x)\| = \sum_1^N \lambda_i^2 \Big/ \left(\sum_1^N \lambda_i^{2/k}\right)^{k/2},$$

где $\lambda_i = g_i^{(k)}(x_i)$.[1]

Доказательство. Если $v \in \mathbb{R}^N$, то

$$\|D^k G(x)(v^k)\| = \left(\sum (\lambda_i v_i^k)^2\right)^{1/2}, \qquad D^k G(x)(v^k) = (\lambda_1 v_1^k, \ldots, \lambda_N v_N^k)$$

и максимум последней нормы на множестве $\|v\| = 1$ достигается при $v_i^k = K\lambda_i$ для некоторого $K > 0$. Таким образом,

$$\|D^k G(x)(v^k)\| = K\left(\sum \lambda_i^2\right)^{1/2}, \quad K = 1/\left(\sum \lambda_i^{2/k}\right)^{k/2}, \text{ ч. т. д.}$$

---

[1] Здесь небольшая неточность. Правильная формула

$$\|D^k G(x)\| = \max\{|\lambda_1|, \ldots, |\lambda_N|\}.$$

— *Прим. перев.*

Теперь легко оценить величину $\alpha(0, \widehat{\Phi}_a - q)$. Имеем $\widehat{\Phi}_a(0)=0$, $D\widehat{\Phi}_a(0) = (I+M)/2$, $\|D\widehat{\Phi}_a(0)^{-1}\| \leqslant 2$, откуда $\beta(0, \widehat{\Phi}_a - q) \leqslant 2\|q\|$. Далее,

$$\begin{aligned}\gamma(0, \widehat{\Phi}_a - q) &\leqslant \sup_{k>1} \left\| D\Phi_a^{-1}(0)\frac{D^k\Phi_a(0)}{k!} \right\|^{1/(k-1)} \\ &\leqslant \left\| (I+M)^{-1}(I-M)\frac{D^{k-1}\Delta}{k!} \right\|^{1/(k-1)} \\ &\leqslant \left| \frac{\varphi_a^{(k)}(0)}{k!} \right|^{1/(k-1)} \leqslant \frac{4}{a}.\end{aligned}$$

Здесь мы воспользовались леммами 7—9. Следовательно, $\alpha = \beta\gamma \leqslant 8\|q\|/a < \alpha_0$, чем наше предложение и доказано.

§ 6

В статье Смейл [51] был приведён список нерешённых задач, касающихся алгоритмов численного анализа. С тех пор как была написана эта статья, для части из них получено решение или достигнут некоторый прогресс. В этом параграфе мы сообщим об этих результатах.

Задача 1 связана с обобщением результатов Смейла [50] и Шуба — Смейла [43, 44] о нахождении нулей многочленов. В указанных работах даны некоторые вероятностные оценки для алгоритмов типа метода Ньютона и в конечном счёте доказано, что число итераций пропорционально степени многочлена. Используемые там методы основаны на теории функций Шлихта (в той её части, которая имеет отношение к гипотезе Бибербаха) и не распространяются на задачу нахождения нулей полиномиальных систем $f\colon \mathbb{C}^n \to \mathbb{C}^n$, где $n > 1$. В задаче 1 предлагалось получить результаты для этого случая. Недавно Ренегар [40] статистическими методами нашёл полиномиальные границы для каждого n. Эта важная работа даёт некоторое решение нашей задачи. Тем не менее остаются открытыми многие серьёзные вопросы, в том числе вопросы концептуального порядка. Границы Ренегара, полиномиальные по степени многочленов, очень грубые. Например, в случае одной переменной число итераций у Ренегара — порядка d^{26} (ср. с d у Шуба — Смейла [44]). Результаты Ренегара послужили одним из стимулов для написания настоящей работы, и я надеюсь, что сформулированные здесь теоремы внесут некоторый вклад в теорию сложности полиномиальных систем.

Задача 6. Здесь речь идет о систематическом изучении множества многочленов от одной переменной, для которых метод Ньютона может зацикливаться на открытых множествах. Коли доказала, что для квадратных многочленов такого не бывает. Джанет Хед [18] из Корнелла исследовала кубические многочлены и показала, что зацикливание в этом случае — довольно редкое явление. Многочлены f степени выше 3 в её работе не рассматриваются.

Задача 7 ставит вопрос о том, как часто (обычный) метод Ньютона сходится для многочленов от одной переменной. Пусть A_f — нормированная площадь множества точек D_2, для которых метод Ньютона сходится к нулю многочлена f. Положим

$$A_d = \min_{f \in P_d(1)} A_f.$$

Джоуэл Фридман [12] из Калифорнийского университета (Беркли) доказал мою гипотезу о том, что $A_d > 0$ для всех d и дал некоторые оценки на A_d как функцию от d, решив тем самым задачу 7А. Относительно задачи 7В см. ниже обсуждение задачи 9.

Задача 8. Каждому многочлену f сопоставляется некоторый график Γ_f, отражающий топологию f как отображения и тесно связанный с методом Ньютона. В задаче спрашивается, какие при этом возможны графики. Исследование этого вопроса уже было начато Джонгеном — Джонкером — Туилтом [21]. В двух статьях этих авторов изложены их результаты и приведена соответствующая библиография. Кроме того, обстоятельное изучение проблемы провели Шуб — Тишлер — Уильямз [46], так что можно считать задачу по существу решённой.

Задача 9 об ограниченности снизу средней площади множества приближённых нулей полностью решается теоремой, доказанной в § 3 настоящей работы (между прочим, эта теорема решает и задачу 7В).

Задача 10. В ней было высказано предположение, что для нахождения нулей многочлена в $\mathbb{C}$ не существует чисто итеративных алгоритмов, сходящихся при достаточно общих предположениях. Эту задачу полностью решил Кёрт Макмаллен [33].

Теорема. *Пусть $d > 3$ и $T\colon P_d \times S \to S$ — произвольное отображение рациональное по f и z над $\mathbb{C}$. Не существует открытого множества $U \subset P_d \times S$ полной меры со следующим свойством: если $(f, z) \in U$, то последовательность $T_f^k(z) = z_k$ сходится к корню многочлена f при $k \to \infty$.*

Здесь P_d — пространство всех многочленов степени $\leqslant d$, а S — сфера Римана. Рациональность отображения T над $\mathbb{C}$ означает, что T можно выразить с помощью комплексных рациональных операций $(+, -, \times, \div)$ через z и коэффициенты f. Как доказала Коли, для $d=2$ таким алгоритмом является метод Ньютона $T_f(z)=z-f(z)/f'(z)$. Для $d=3$ Макмаллен [33] нашёл новый алгоритм, чисто итеративный и сходящийся в общем положении.

В работах [34, 35] Макмаллен усилил свои результаты. Шуб и я [45] показали, что если дополнить список рациональных операций комплексным сопряжением, то построить чисто итеративный алгоритм, сходящийся в общем положении, уже можно. Наш результат распространяется и на случай полиномиальных отображений нескольких переменных $f\colon \mathbb{C}^n \to \mathbb{C}^n$.

ЛИТЕРАТУРА

1. J. Abadie and G. Guerrero, Méthode du GRG, méthode de Newton globale et application a la programmation mathématique, RAIRO Rech. Opér. **18** (1984), 319—351.

2. M. Berger, Non-linearity and functional analysis, Academic Press, New York, 1974.

3. L. Blum, Towards an asymptotic analysis of Karmakar's algorithm, Inform. Process. Lett. **23** (1986), 189—194.

4. S. Chow, J. Mallet-Paret, and J. Yorke, Finding zeros of maps: homotopy methods that are constructive with probability one, Math. Comp. **32** (1978), 887—899.

5. R. Cottle, Note on a fundamental theorem in quadratic programming, J. SIAM **12** (1964), 663—665.

6. R. Cottle and G. Dantzig, Complementary pivot theory of math. programming in math. of the decision sciences, G. Dantzig and A. Veisott, Jr., eds., Amer. Math. Soc., Providence, R. I., 1968, pp. 115—136.

7. J. Curry, On zero finding methods of higher order from data at one point, MSRI Preprint, 1986.

8. G. Dantzig, Linear programming and extensions, Princeton Univ. Press, Princeton, N. J., 1963.

9. B. Dejon and P. Henrici, Constructive aspects of the fundamental theorem of algebra, Wiley, New York, 1969.

10. J. Dieudonné, Foundations of modern analysis, Academic Press, New York, 1960. [Имеется перевод: Дьёдонне Ж., Основы современного анализа.— М.: Мир, 1964.]

11. C. Eaves and H. Scarf, The solution of systems of piecewise linear equations, Math. Oper. Res. **1** (1976), 1—27.

12. J. Friedman, Rough draft of a theorem on Newton's method, Univ. of California Press, Berkeley, Calif., 1985.

13. F. Gao, Nonasymptotic error of numerical integration — an average analysis, Univ. of California Press, Berkeley, Calif., 1986 (to appear).

14. —, Probabilistic analysis of automatic integration (to appear).

15. C. Garcia and F. J. Gould, Relations between several path following algorithms and local and global Newton methods, SIAM Rev. **22** (1980), 263—274.

16. G. Ghellinck and M. P. Vial, A. polynomial Newton method for linear programming, CORE, Louvain, Belgium, 1986.

17. P. Gill, W. Murray, M. Saunders, J. Tomlin, and M. Wright, On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method, Tech. Report SOL 85-11, Dept. of Op. Res., Stanford Univ., Stanford, Calif., 1985.

18. J. Head (to appear).

19. P. Henrici, Applied and computational complex analysis, Wiley, New York, 1977.

20. M. Hirsch and S. Smale, On algorithms for solving $f(x)=0$, Comm. Pure Appl. Math. **32** (1979), 281—312.

21. H. Th. Jongen, P. Jonker, and F. Twilt, The continuous desingularized Newton's method for meromorphic functions, Memorandum No. 501, Dept. of Appl. Math., Twente Univ. of Technology, 1985.

22. —, Non-linear optimization theory in $\mathbf{R}^n$ from a global point of view. VIII, Memorandum No. 559, Dept. of Appl. Math., Twente Univ. of Technology, 1986.

23. Канторович Л. В., Акилов Г. П. Функциональный анализ в нормированных пространствах. — 2-е изд. — М.: Наука, 1977.

24. N. Karmarkar, A new polynomial-time algorithm for linear programming, Combinatorica **4** (1984), 373—395.

25. H. Keller, Numerical solution of bifurcation and nonlinear eigenvalue problems, Application of Bifurcation Theory, Academic Press, New York, 1977.

26. —, Global homotopic and Newton methods, Recent Advances in Numerical Analysis, Academic Press, New York, 1978.

27. M. Kim, Ph. D. Thesis, Graduate School, CUNY, 1985 (to appear).

28. H. Kung, The complexity of obtaining starting points for solving operator equations by Newton's method, Analytic Computational Complexity (J. Traut, ed.), Academic Press, New York, 1976.

29. S. Lang, Algebra, Addison-Wesley, Reading, Mass., 1963. [Имеется перевод: Ленг С. Алгебра. — М.: Мир, 1968.]

30. —, Real analysis, Addison-Wesley, Reading, Mass., 1983.

31. O. Mangasarian, Equivalence of the complementarity problem to a system of nonlinear equations, SIAM J. Appl. Math. **31** (1976), 89—92.

32. M. Marden, Geometry of polynomials, Math. Surveys, no. 3, Amer. Math. Soc., Providence, R. I., 1966.

33. C. McMullen, Families of rational maps and iterative root-finding algorithm, Ann. Math. (2) **125** (1987), 467—493.

34. —, Automorphisms of rational I: Nielsen realization and dynamics on the ideal boundary, MSRI, Berkeley, Calif., 1986.

35. —, Automorphisms of rational maps II: braiding of the attractor and the failure of iterative algorithms, MSRI, Berkeley, Calif., 1986.

36. N. Megido and M. Shub, The boundary behavior of interior methods for linear programming (to appear).

37. J. L. Nazareth, Homotopy technniques in linear programming, Algorithmica **1** (1986), 529—535.

38. A. Ostrowski, Solutions of equations in Euclidean and Banach spaces, Academic Press, New York, 1973.

39. V. Pan, Algebraic complexity of computing polynomial zeros, Tech. Report 85-27, SUNY, Albany, N. Y., 1985.

40. J. Renegar, On the efficiency of Newton's method in approximating all zeros of a system of complex polynomials, Math. Oper. Res. **12** (1987), 121—148.

41. —, A polynomial-time algorithm based on Newton's method for linear programming, MSRI, Berkeley, Calif., 1986.

42. H. Royden (te appear).
43. M. Shub and S. Smale, Computational complexity: on the geometry of polynomials and a theory of cost. I, Ann. Sci. École Norm. Sup (4) **18** (1985), 107—142.
44. —, Computational complexity: on the geometry of polynomials and a theory of cost. II, SIAM J. Comput. **15** (1986), 145—161.
45. —, On the existence of generally convergent algorithms, J. Complexity **2** (1986), 2—11.
46. M. Shub, D. Tischler, and R. Williams, The Newtonian graph of a complex polynomial, SIAM J. Math. Anal. **19** (1988), 246—256.
47. S. Smale, A convergent process of price adjustment and global Newton methods, J. Math. Econom. **3** (1976), 107—120.
48. —, On the average number of steps in the simplex method of linear programming, Math. Programming **27** (1983), 241—262.
49. —, The problem of the average speed of the simplex method, Mathematical Programming: The State of the Art (Bonn, 1982), Springer-Verlag, Berlin and New York, 1983, pp. 530—539.
50. —, The fundamental theorem of algebra and complexity theory, Bull. Amer. Math. Soc. (N. S.) **4** (1981), 1—36.
51. —, On the efficiency of algorithms of analysis, Bull. Amer. Math. Soc. (N. S.) **13** (1985), 87—121.
52. —, Newton's method estimates from data at one point, *in*: The merging of disciplines: new directions in pure, applied, and computational mathematics (Laramie, Wyo., 1985), Springer, New York — Berlin, 1986, pp. 185—196. (В тексте мы ссылаемся на эту работу как на [P. E.].)
53. A. Wierzbicki, Note on the equivalence of Kuhn-Tucker complementarity conditions to an equation, J. Optim. Theory Appl. **37** (1982), 401—405.
54. S. Wong, Newton's method and symbolic dynamics, Proc. Amer. Math. Soc. **91** (1984), 245—253.
55. H. Woźniakowski, A survey of information-based complexity, J. Complexity **1** (1985), 11—44.
56. Paul Wright, Statistical complexity of the power method for Markov chains, Univ. Calif., Berkeley, 1986 (to appear).

КАЛИФОРНИЙСКИЙ УНИВЕРСИТЕТ, БЕРКЛИ, КАЛИФОРНИЯ 94720, США

НЕКОТОРЫЕ ПРОБЛЕМЫ ГАРМОНИЧЕСКОГО АНАЛИЗА, СВЯЗАННЫЕ С ПОНЯТИЕМ КРИВИЗНЫ И ОСЦИЛЛЯТОРНЫМИ ИНТЕГРАЛАМИ

Илайас М. Стейн[1]

ВВЕДЕНИЕ

Ряд результатов, полученных за последние десять лет в области гармонического анализа, помогли нам осознать всё возрастающую важность некоторых первоначальных геометрических понятий вроде понятия «кривизны», а также некоторых «метрик», связанных с анализом векторных полей и нильпотентных групп. Зачастую использование этих понятий и метрик тесно связано с осцилляторными интегралами. Указанные результаты — и как побудительные мотивы для дальнейших исследований, и с точки зрения приложений — имеют прямое отношение к таким разделам математики, как анализ на полупростых группах Ли и на симметрических пространствах, теория функций многих комплексных переменных, теория дифференциальных уравнений в частных производных и теория интегральных операторов Фурье. Из-за недостатка места мы не станем подробно останавливаться на взаимосвязи этих разделов и на приложениях[2], а сконцентрируем внимание на результатах, касающихся взаимосвязи кривизны, осцилляторных интегралов, нильпотентных групп и теории функций вещественной переменной.

Объект исследования. Мы будем иметь дело с тремя взаимосвязанными аналитическими конструкциями, которые в рамках настоящего обзора можно считать центральными объектами исследования. Это «средние значения», «сингулярные интегралы» и «осцилляторные интегралы». Каждая из этих конструкций имеет свою долгую и богатую историю, которую мы не можем здесь излагать. Нас больше всего интересуют

---

[1] Elias M. Stein, Problems in harmonic analysis related to curvature and oscillatory integrals, ICM86, pp. 196—221.

[2] См., однако, некоторые работы из приводимого в конце списка литературы.

установление более тесных связей между ними и новые подходы к ним. Сначала несколько вводных слов о каждом из этих понятий.

Средние значения. Основным примером использования понятия среднего значения, относящимся к самой ранней стадии развития анализа, служит теорема о среднем — одна из основных теорем дифференциального исчисления. Другие классические примеры возникают при решении уравнения теплопроводности $\partial^2 u/\partial x^2 = \partial u/\partial t$, при решении задачи Дирихле с помощью интеграла Пуассона и т. д. Основные свойства таких средних в теории функций вещественной переменной были выражены в 30-е годы с помощью понятия максимальной функции Харди — Литтлвуда — Винера, определяемой в $\mathbb{R}^n$ формулой

$$(Mf)(x) = \sup_r \frac{1}{v_n r^n} \left| \int_{|y| \leqslant r} f(x-y)\, dy \right|. \qquad (0.1)$$

На наш взгляд, наиболее существенным изменением точки зрения, происшедшим за последнее время, является то, что сейчас возрос интерес к использованию средних по множествам малой размерности. Например, в выражении (0.1) шар радиуса r заменяется подходящим семейством многообразий меньшей размерности. Изучение таких средних важно не только само по себе, но и потому, что оно оказывается полезным при решении других проблем, таких как исследование поведения классических средних (0.1) при $n \to \infty$, исследование поведения решений волновых уравнений и т. д.

Сингулярные интегралы. Классический сингулярный интегральный оператор (восходящий к интегралу Коши и теории эллиптических уравнений) может быть записан в стандартном виде

$$T(f)(x) = \int_{\mathbb{R}^n} f(y) K(x-y)\, dy, \qquad (0.2)$$

где ядро K однородно степени $-n$, гладко всюду, кроме начала координат, и имеет нулевое среднее значение. Однако для теории функций многих комплексных переменных и для широкого класса «субэллиптических» уравнений представляют бо́льшой интерес операторы с более общими ядрами $K(x, y)$ (вместо $K(x-y)$). Здесь можно выделить два случая: (a) ядро $K(x, y)$ по-прежнему имеет сингулярность на диагонали, но эта сингулярность не является более «изотропной», хотя поведение K вблизи диагонали регулируется некоторыми функциями «расстояния» и соответствующими формами объёма; (b) бывает необходимо также рассматривать ситуации, когда сингулярность имеет место не только на диагонали, но и на более широком подмногообразии.

Осцилляторные интегралы. Существует целое множество различных типов осцилляторных интегралов, и их нелегко классифицировать. Здесь достаточно привести в качестве основного примера преобразование Фурье

$$f \to \int_{\mathbb{R}^n} e^{-2\pi i x \cdot \xi} f(x)\, dx = \hat{f}(\xi). \qquad (0.3)$$

Другие примеры получаются заменой $f(x)dx$ какой-либо фиксированной плотностью (тогда основное внимание уделяется изучению поведения $\hat{f}(\xi)$ при $|\xi| \to \infty$) или заменой $2\pi x \cdot \xi$ более общим показателем (тогда изучают свойства ограниченности построенного оператора).

Некоторые общие принципы. Мы сейчас сформулируем три достаточно общих (но несколько расплывчатых) принципа, которые помогут прояснить используемый ниже подход. Как и все утверждения подобного рода, предлагаемые принципы можно считать, с одной стороны, интерпретацией уже полученных результатов, а с другой — эвристическими соображениями для будущих исследований.

(i) При изучении операторов, ядра которых обладают сингулярностями неклассического типа, ключевую роль играют такие геометрические понятия, как кривизна, ассоциированные квазиметрики и их формы объёма.

(ii) Чтобы использовать свойства кривизны, надо привлекать осцилляторные интегралы.

(iii) Часто решающим свойством кривизны в рассматриваемых вопросах является то, что она не имеет нулей бесконечного порядка. Поэтому изучаемые явления можно моделировать в $\mathbb{R}^n$ с помощью полиномиальных функций или, в крайнем случае, с помощью анализа на нильпотентных группах Ли.

Структура настоящего обзора. Материал в данном обзоре организован в соответствии с упомянутыми выше общими принципами. А именно, часть I посвящена рассмотрению ситуации в $\mathbb{R}^n$, где связь между кривизной и осцилляторными интегралами наиболее ясна. Затем в части II мы обсуждаем ситуацию для нильпотентных групп. Этот случай можно считать модельным для «общей» ситуации, которая изучается в части III (и о которой мы говорим как о случае «переменных коэффициентов»). Наконец, в части IV кратко изложены некоторые близкие результаты.

ЧАСТЬ I. $\mathbb{R}^n$

1. Максимальные сферические средние. Наиболее наглядный пример того, какую роль играет понятие кривизны в теории функций вещественной переменной и гармоническом анализе, мы получаем при рассмотрении средних значений функций, когда эти средние строятся не по шарам, как в классической теории (см. выше (0.1)), а по сферам; другой поучительный случай — изучение поведения стандартных средних в $\mathbb{R}^n$ при $n \to \infty$.

Для заданной функции f на $\mathbb{R}^n$ определим её среднее значение на сфере с центром в точке x радиуса t как

$$M_t(f)(x) = \int\limits_{|y|=1} f(x - ty)\, d\sigma(y). \tag{1.1}$$

(Здесь $d\sigma$ обозначает нормированную равномерно распределенную меру на единичной сфере.)

Для исследования предела $\lim\limits_{t\to 0} M_t(f)(x)$ удобно ввести *максимальную функцию*

$$\mathcal{M}(f)(x) = \sup_{t>0} |M_t(f)(x)|, \tag{1.2}$$

которая корректно определена по крайней мере для случая, когда функция f непрерывна. Имеет место следующий фундаментальный результат:

Теорема 1. *Пусть $n \geqslant 2$ и $p > n/(n-1)$. Тогда справедлива априорная оценка*

$$\|\mathcal{M}(f)\|_{L^p} \leqslant A \|f\|_{L^p}. \tag{1.3}$$

(См. Stein [99], Stein — Wainger [105] для случая $n \geqslant 3$, Bourgain [4] для случая $n = 2$.)

Замечания. (a) Прежде всего следует подчеркнуть, что мы не получим никакого неравенства типа (1.3), если в определении (1.1) заменим сферу на другую поверхность, которая гладка, но о кривизне которой ничего заранее не предполагается. Для того чтобы убедиться в этом, заметим, что в случае $n = 1$ и для сферической максимальной функции нельзя получить нетривиального результата, так как в этом случае $M_t(f)(x) = (f(x+t) + f(x-t))/2$ и для того, чтобы было исключено любое L^p-неравенство, достаточно, чтобы функция f была неограничена в окрестности какой-нибудь одной точки. Аналогичное соображение показывает, что для сферической максимальной функции неравенство (1.3) не может выполняться при $p \leqslant n/(n-1)$, а если «распрямить» сферу около некоторой точки, то (1.3) не будет выполняться при любом $p < \infty$.

(b) Кратко коснёмся ещё нескольких вопросов анализа, с которыми связана теорема 1. Во-первых, отметим удивительную аналогию с поведением классического преобразования Радона

$$f \to R(f)(\sigma, t) = \int\limits_{\langle x,\, \sigma\rangle = t} f(x)\, dx$$

($|\sigma| = 1$, $t > 0$), для которого инструментом использования служит $\sup\limits_{t>0} |R(f)(\sigma, t)|$ при $p > n/(n-1)$. Парадоксально, однако, что эта аналогия прослеживается лишь для $n \geqslant 3$; из рассмотрения множества Безиковича — Какэя явствует, что аналог теоремы 1 для преобразования Радона в случае $n = 2$ не имеет места. См. Marstrand [59], Falconer [27], Strichartz [108], Oberlin — Stein [75].

Во-вторых, существует тесная связь теоремы 1 с вопросом о сходимости к своим начальным значениям решений волнового уравнения или, вообще, уравнений гиперболического типа. См. Stein [99], Greenleaf [37], Ruiz [86], Sogge [91].

Для доказательства результатов типа теоремы 1 помимо прочего используются осцилляторные интегралы и их связь с кривизной, а также квадратичные функции. Кратко обсудим эти вопросы.

2. Осцилляторные интегралы и кривизна. Поскольку семейство средних значений, задаваемых операторами M_t из формулы (1.1), представляет собой свёртку, естественно прибегнуть к преобразованию Фурье. Итак, $(M_t(f))\hat{}\,(\xi) = \hat{f}(\xi)\, \hat{\sigma}(t\xi)$, где

$$\hat{\sigma}(\xi) = \int\limits_{|x|=1} e^{-2\pi i x\cdot\xi}\, d\sigma(x).$$

Оказывается, $\hat{\sigma}$ можно выразить через функции Бесселя (а именно $\hat{\sigma}(\xi) = c_n |\xi|^{-(n-2)/2} J_{(n-2)/2}(2\pi|\xi|)$). Отсюда легко следует ключевой факт убывания $\hat{\sigma}$ на бесконечности:

$$|\hat{\sigma}(\xi)| \leqslant A |\xi|^{-(n-1)/2}. \tag{2.1}$$

(См., например, Stein — Weiss [106, гл. 4].)

Именно на этом убывании и основана теорема 1. Для того чтобы лучше понять роль кривизны, рассмотрим аналогичную ситуацию, но только вместо сферы в $\mathbb{R}^n$ возьмём, скажем, гладкую гиперповерхность S и заменим $d\sigma$ на $d\mu = \psi d\sigma$, где ψ — некоторая срезающая функция из C_0^∞, а $d\sigma$ — индуцированная лебегова мера на S. Определим $\hat{\mu}(\xi)$ следующим образом:

$$\hat{\mu}(\xi) = \int_S e^{-2\pi i x \cdot \xi} d\mu. \tag{2.2}$$

Известно (для $n = 2$ этот результат восходит к Ван дер Корпуту, а для $n \geqslant 3$ — к Хлавке [47]; см. также Herz [46], Hörmander [51]), что

$$\hat{\mu}(\xi) = O\left(|\xi|^{-(n-1)/2}\right) \quad \text{при} \quad |\xi| \to \infty$$

по существу тогда и только тогда, когда S обладает ненулевой кривизной в каждой точке носителя функции ψ.

Приведём ещё два результата о связи необращения в нуль кривизны и убывания преобразования Фурье, на которые мы будем ссылаться в дальнейшем. Для того чтобы сформулировать первый результат, введём следующие обозначения. Пусть S^k обозначает k-мерное гладкое многообразие в $\mathbb{R}^n$, $d\mu = \psi d\sigma$, где $\psi \in C_0^\infty$, а $d\sigma$ — индуцированная мера на S^k. Скажем, что S^k — *финитного типа* в точке $x_0 \in S^k$, если S^k имеет касание конечного порядка с любой аффинной гиперплоскостью, проходящей через x_0. Справедлива (см. Stein [103])

Лемма 1. *Для указанных выше S^k и $d\mu$ существует $\varepsilon > 0$, $\varepsilon = \varepsilon(S^k)$, такое что $\hat{\mu}(\xi) = O\left(|\xi|^{-\varepsilon}\right)$ при $|\xi| \to \infty$.*

В частности, этот результат справедлив, когда S^k есть вещественно-аналитическое многообразие, не лежащее ни в какой аффинной гиперплоскости, — факт, установленный ранее Бьёрком [2].

Предположим, однако, что мы хотим получить полную скорость убывания $O\left(|\xi|^{-(n-1)/2}\right)$, по крайней мере для гиперповерхностей в $\mathbb{R}^n$. Это можно сделать, введя «демпфирующий» сомножитель, куда входит гауссова кривизна. Пусть S обозначает теперь произвольную гладкую гиперповерхность в $\mathbb{R}^n$, $K(x)$ — её гауссова кривизна в точке $x \in S$, а $d\mu = \psi d\sigma$, как и ранее. Наш второй результат звучит так:

Лемма 2. *Для достаточно больших N*

$$\int_S e^{-2\pi i x \cdot \xi} |K(x)|^N d\sigma(x) = O\left(|\xi|^{-(n-1)/2}\right). \tag{2.3}$$

Фактически формула (2.3) *справедлива для* $N \geqslant 2n - 2$.

Эта лемма приведена в работе Sogge — Stein [92]. Оценку другого типа для интеграла (2.2), дающую $O(|\xi|^{-(n-1)/2})$ (при $|\xi| \to \infty$) для «большинства» направлений, можно найти в Randol [80] и Svenson [112]. См. также оценку в п. 16, полученную с помощью других геометрических понятий. В связи с леммой 2 сформулируем следующую задачу:

Задача. (a) Для какого наименьшего N имеет место формула (2.3)?

(b) Интереснее другая постановка вопроса: для какой «наибольшей» функции (вместо $|K(x)|^N$) интеграл (2.3) есть $O(|\xi|^{-(n-1)/2})$?

3. Квадратичные функции. Свойства убывания преобразования Фурье меры, сосредоточенной на поверхности, можно использовать, вводя определённые квадратично-интегрируемые функции, которые мы далее будем называть *квадратичными*. Наиболее простым примером таких функций, возникающим при рассмотрении сферических средних, служит функция, определяемая формулой

$$S(f)(x) = \left(\int_0^\infty \left|\frac{dM_t(f)(x)}{dt}\right|^2 t\,dt\right)^{1/2}. \tag{3.1}$$

С помощью формулы Планшереля и оценки, аналогичной (2.1), можно показать, что $\|S(f)\|_{L^2} \leqslant c\|f\|_{L^2}$ при $n \geqslant 4$. Поскольку

$$t^n M_t = \int_0^t \frac{d(s^n M_s)\,ds}{ds} = n\int_0^t M_s s^{n-1}\,ds + \int_0^t \frac{dM_s}{ds} s^n\,ds,$$

отсюда легко следует, что $\mathscr{M}(f) \leqslant c\{S(f) + M(f)\}$. Теперь видно, как доказать теорему 1 для случая $p = 2$, $n \geqslant 4$. Для $p = 2$, $n \geqslant 3$ требуются более тонкие рассуждения. Вот пример результата, который может быть получен с помощью таких более тонких рассуждений. Предположим, что $m(\xi)$ — функция класса C^1 на $\mathbb{R}^n$, удовлетворяющая для некоторого $\varepsilon > 0$ условию

$$|m(\xi)| + |\nabla m(\xi)| \leqslant c\,(1 + |\xi|)^{-1/2-\varepsilon}. \tag{3.2}$$

Пусть $M_t(f)$ задано соотношением $M_t(f)^{\wedge}(\xi) = m(t\xi)\hat{f}(\xi)$.

Лемма. $\|\sup\limits_{t>0}|M_t(f)|\|_{L^2}\leqslant A\|f\|_{L^2}$.

Доказательство см. в Sogge — Stein [92]. Другие случаи рассмотрены в работах Bourgain [3], Carbery [10], Rubio de Francia [85].

Вышеприведенный результат является точным в том смысле, что он перестаёт быть верным при $\varepsilon = 0$.

С помощью квадратичных функций случай L^p, $p\neq 2$, фактически сводится при таком подходе к случаю L^2. Более пристальное исследование показывает, что на самом деле имеется некоторый «запас» и несколько более «широкий» или «грубый» вариант оператора $\mathscr{M}$ всё ещё будет ограниченным в L^2. Используя этот «L^2-кредит», мы можем комбинировать его с тем фактом, что некую «улучшенную» версию оператора $\mathscr{M}$ можно исследовать (и притом для всех L^p) стандартным способом. Отметим, однако, что случай $n = 2$ более тонкий. Как показал Бурген, вместо стандартных L^p-оценок здесь надо воспользоваться хитрым геометрическим рассуждением.

Таким образом, общая схема доказательства максимальных L^p-неравенств (к которым мы еще вернёмся) выглядит следующим образом: сначала доказательство проводится для L^2 с помощью квадратичных функций; затем полученные результаты интерполируются на случай L^p, $p\neq 2$. Фактически этот подход следует образцу, развитому в более абстрактном контексте общей максимальной теоремы для симметрических диффузионных однопараметрических полугрупп (см. Stein [95, 96]). В той или иной форме квадратичные функции используются для получения и других результатов, которые мы обсудим в последующих разделах. См. также п. 9 ниже.

4. Некоторые другие результаты для средних по поверхностям. Очертив основные идеи доказательства теоремы 1, перейдём теперь к некоторым её разновидностям и следствиям. Первый случай касается обобщения M_t, когда сфера заменяется более общей гладкой гиперповерхностью S. Положив $d\mu(x) = \psi(x)\,d\sigma$, как в (2.2), можем написать

$$M_t(f)(x) = \int\limits_S f(x-ty)\,d\mu(y), \qquad t>0,$$

и $\mathscr{M}(f) = \sup\limits_{t>0}|M_t(f)|$.

Теорема 2. *Предположим, что гауссова кривизна поверхности S не имеет нуля бесконечного порядка ни в одной точке S. Тогда существует $p_0 = p_0(S)$, $p_0 < \infty$, такое что*

$$\|\mathscr{M}(f)\|_{L^p}\leqslant A_p\|f\|_{L^p} \quad \text{при} \quad p_0<p\leqslant\infty. \tag{4.1}$$

Заметим, что условия теоремы заведомо выполнены, если S — компактная вещественно-аналитическая гиперповерхность. Доказательство теоремы проводится с помощью оценки (2.3) и интерпретации L^∞-случаем (где и используется отсутствие у кривизны нулей бесконечного порядка). Следует отметить, что наличие у кривизны S нуля бесконечного порядка хотя бы в одной точке может привести к тому, что оценка (4.1) не будет выполняться ни для одного $p<\infty$. По поводу подробностей доказательства теоремы 2 см. Sogge — Stein [92] (случай $n\geqslant 3$) и Bourgein [5] (случай $n = 2$). Те же идеи развивались ранее в работах Greenleaf [37] и Cowling — Mauceri [21]; см. также Cowling — Mauceri [22].

В этой связи естественно поставить следующий вопрос.

Задача. (a) Верно ли такое предположение: максимальное неравенство (4.1) справедливо для нетривиального диапазона изменения показателя p тогда и только тогда, когда поверхность S в каждой своей точке имеет касание конечного порядка с касательной гиперповерхностью.

(b) К этой задаче примыкает проблема поиска наименьшего значения $p_0 = p_0(S)$, для которого выполняется (4.1).

5. Результаты для $\mathbb{R}^n$ при $n\to\infty$. Полученные результаты и методы их получения, описанные выше для нестандартных максимальных функций, применимы также для обычных максимальных функций и других стандартных операторов гармонического анализа в $\mathbb{R}^n$, при изучении вопроса о поведении норм этих операторов при $n\to\infty$. Обратимся вначале к обычным (центрированным) максимальным функциям, определённым следующим образом:

$$M^{(n)}(f)(x) = \sup_{r>0}\frac{1}{v_n r^n}\left|\int\limits_{|y|\leqslant r} f(x-y)\,dy\right|, \tag{5.1}$$

где v_n — объём единичного шара.

Теорема 3. (a) *Справедливо неравенство*

$$\|M^{(n)}(f)\|_p\leqslant A_p\|f\|_p, \qquad 1<p\leqslant\infty,$$

где константа A_p не зависит от n.

(b) *Что касается неравенств слабого типа, то можно утверждать, что*

$$m\{x\mid M^{(n)}f(x)>\alpha\}\leqslant\frac{c_n}{\alpha}\|f\|_{L^1}, \quad \text{где} \quad c_n = O(n). \tag{5.2}$$

Доказательство утверждения (a) основано на возможности перенесения сферического максимального неравенства (1.3) (для заданного p и фиксированной размерности) на случай большего числа измерений без увеличения нормы. (Для стандартной максимальной функции этот перенос не проходит!) Для доказательства утверждения (b) требуется дополнительное рассуждение, в котором используется общая максимальная эргодическая теорема. Подробности можно найти в Stein — Stromberg [104]; см. также Stein [100—102].

Другой результат того же типа касается базисных сингулярных интегралов — преобразований Рисса R_j, $j = 1, \ldots, n$, определяемых формулой как $R_j(f)^\frown(\xi) = (i\xi_j/|\xi|)\hat{f}(\xi)$. По аналогии со сказанным выше, можно рассмотреть задачу о получении оценок для $|R(f)| = \left(\sum_{j=1}^{n} |R_j(f)|^2\right)^{1/2}$ в терминах подходящих квадратичных функций, в данном случае n-мерных обобщений g-функций Литтлвуда — Пэли.

Речь идёт о функциях

$$g_1(f)(x) = \left(\int_0^\infty \left|\frac{\partial u}{\partial t}(x, t)\right|^2 t\,dt\right)^{1/2},$$

$$g(f)(x) = \left(\int_0^\infty |\nabla u(x, t)|^2 t\,dt\right)^{1/2},$$

где $u(x, t)$ — интеграл Пуассона от функции f.

Предложение. *Если $1 < p \leqslant \infty$, то L^p-нормы функций f, $R(f)$, $g(f)$, и $g_1(f)$ эквивалентны, и притом с постоянными, которые не зависят от n.*

Эквивалентность (с константами, не зависящими от размерности) норм функций f, $g(f)$ и $g_1(f)$ по существу доказана в работах Stein [96, 97], посвященных этим квадратичным функциям; см. также Meyer [62]. Соответствующие неравенства для преобразований Рисса следуют из оценки

$$g_1(Rf)(x) \leqslant g(f)(x). \qquad (5.3)$$

(Другой подход к получению оценок для $R(f)$, использующий «метод вращений», развит Дуоандикоэтксеа и Рубио де Франсиа [25].)

После этого отступления вернёмся к теореме 3 и зададимся вопросом: можно ли систему центрированных шаров заменить гомотетиями фиксированного выпуклого симметричного тела B? Итак, для некоторого такого фиксированного B рассмотрим оператор M_B, определённый формулой

$$(M_B f)(x) = \sup_{r>0} \frac{1}{m(rB)} \left| \int_{rB} f(x-y)\,dy \right|,$$

где $rB = \{rx, x \in B\}$, а $m(rB)$ — объем тела rB. Вопрос заключается в том, можно ли получить L^p-оценки для M_B, независимые от B и размерности n.

Теорема 4. (a) *$\|M_B(f)\|_p \leqslant A_p\|f\|_p$, для $3/2 < p \leqslant \infty$, причём A_p не зависит от B и n.*

(b) *$m\{x | M_B f(x) > \alpha\} \leqslant \bar{c}_n\|f\|_1/\alpha$, где $\bar{c}_n = O(n \log n)$ и $\bar{c}_n$ не зависит от B.*

Для $p \geqslant 2$ утверждение (a) было доказано Бургеном [3]. Доказательство основано на его ключевой лемме (см. ниже). Используя эту лемму, Бурген [6] и Карбери [10] независимо обобщили этот результат на случай $p > 3/2$. (Откуда взялось число 3/2, вскоре выяснится.) Утверждение (b) доказывается совершенно другими методами, см. Stein — Stromberg [104].

Основная идея доказательства утверждения (a) заключается в использовании следующей замечательной универсальной оценки убывания для преобразования Фурье характеристической функции χ_B выпуклого симметричного тела B и $\mathbb{R}^n$ единичного объема. Пусть $\hat{\chi}_B(\xi) = \int_B e^{-2\pi i x\cdot\xi}\,dx$.

Лемма. *Существует $L \in \mathrm{GL}(n, \mathbb{R})$, $L = L(B)$, такое что если $m(\xi) = \hat{\chi}_B(L(\xi))$, то*

$$\begin{aligned} &|m(\xi)| \leqslant c/|\xi|,\\ &|m(\xi) - 1| \leqslant c|\xi|,\\ &|\langle\xi, \nabla m(\xi)\rangle| \leqslant c, \end{aligned} \qquad (5.4)$$

где c — некоторая универсальная константа.

Для $p \geqslant 2$ теорема 4 легко вытекает из этой леммы и (соответственно модифицированной) леммы п. 3. Отметим, что для $p < 2$ оценки (5.4) совпадают с оценками для преобразования Фурье поверхностной меры на единичной сфере в $\mathbb{R}^3$! Так как в силу теоремы 1 результат в этом случае неверен, если $p \leqslant 3/2$, то и нельзя ожидать выполнения оценок (5.4) «по ту сторону» числа 3/2. Однако, с другой стороны, из этого наблюдения можно сделать вывод, что L^2-оценки выполняются «с запасом», и можно надеяться, как в случае сферической максимальной функции, получить оценки для $p > 3/2$. Для этого приходится использовать сложную «бутстрэп»-технику, включающую в себя интерполяцию, подобно тому как это будет сделано ниже при доказательстве теоремы для «лакунарных максимальных функций» (в п. 14) и в «плоском случае» для кривых (в п. 15). Здесь уместно зафиксировать следующую задачу:

Задача. (а) Справедливы ли для $M^{(n)}$ (и, вообще, для M_B) оценки слабого типа с постоянными, не зависящими от n?

(b) Справедливы ли для преобразования Рисса оценки слабого типа с постоянными, не зависящими от n?

(c) Можно ли утверждение (a) теоремы 4 распространить на значения p из интервала $1 < p \leqslant 3/2$?

6. Случай фиксированного подмногообразия. В пп. 1 и 4 мы имели дело со средними на гомотетических образах фиксированной поверхности, т. е. со средними на множествах вида $\{x+\varepsilon S\}_{\varepsilon>0}$; в этом случае, как было показано, свойства кривизны играли решающую роль при использовании осцилляторных интегралов и квадратичных функций. В данном пункте мы будем иметь дело со случаем, когда максимальные функции, сингулярные интегралы и т. п. берутся по отношению к некоторому фиксированному подмногообразию. Хотя имеются и некоторые аналогии с изложенным выше, тем не менее, как мы сейчас убедимся, есть и существенные различия. Теория для случая фиксированного подмногообразия началась с изучения ситуации, когда в роли такого подмногообразия выступает кривая. Эта теория достигла своего расцвета в работах Нагела, Ривьера и Уэйнджера [66, 67]; она возникла из стремления лучше понять, с одной стороны, «метод вращений» для неизотропных сингулярных интегралов, а с другой — граничное поведение интеграла Пуассона на симметрических пространствах (см. Stein — Wainger [105]; некоторые более ранние связанные с этим идеи обсуждаются в Stein [98]).

Наши результаты будут представлены в двух формах. Сначала в виде «глобальных версий», а затем — в виде «локальных». Глобальные теоремы являются, по существу, следствиями локальных; они играют роль моделей для общих теорем. Мы приводим глобальные теоремы отдельно для того, чтобы подчеркнуть ранее сформулированный принцип, гласящий, что свойства, связанные с кривизной, не имеющей нулей бесконечного порядка, должны моделироваться в рамках полиномиального поведения.

Зафиксируем полиномиальное отображение P: $\mathbb{R}^k \to \mathbb{R}^n$, и пусть K — стандартное ядро Кальдерона — Зигмунда в $\mathbb{R}^k$, т. е. K однородно степени $-k$, гладко везде, кроме начала координат, и обладает нулевым средним значением. (Здесь роль нашего подмногообразия играет образ P в $\mathbb{R}^n$.) Определим сингулярный интеграл T от соответствующих функций на $\mathbb{R}^n$ формулой

$$T(f)(x) = \text{p.v.} \int\limits_{\mathbb{R}^k} f(x - P(u))\, K(u)\, du. \qquad (6.1)$$

Аналогично определим максимальную функцию M равенством

$$M(f)(x) = \sup_{0<r<\infty} r^{-k} \left| \int\limits_{|u|<r} f(x - P(u))\, du \right|. \qquad (6.2)$$

Для того чтобы сформулировать «локальные версии», заменим полиномиальное отображение P каким-нибудь гладким вложением ρ единичного шара $B_1 \subset \mathbb{R}^k$ в $\mathbb{R}^n$. Будем считать, что $\rho(0) = 0$. Наше ключевое предположение состоит в том, что $\rho(B_1)$ — конечного типа (в смысле п. 2) в начале координат.

По аналогии с предыдущим положим

$$T_1(f)(x) = \text{p. v.} \int\limits_{|u|\leqslant 1} f(x - \rho(u))\, K(u)\, du, \qquad (6.1')$$

$$M_1(f)(x) = \sup_{0<r\leqslant 1} r^{-k} \left| \int\limits_{|u|\leqslant r} f(x - \rho(u))\, du \right|. \qquad (6.2')$$

Теорема 5. (a) *Оператор T ограничен на L^p, $1 < p < \infty$, причём константы в оценке нормы не зависят от коэффициентов полиномиального отображения P, а зависят лишь от его полной степени.*

(b) *Оператор T_1 ограничен на L^p, $1 < p < \infty$, если ρ — конечного типа в начале координат.*

Аналогичные результаты имеют место для M и M_1 в интервале $1 < p \leqslant \infty$.

Ниже в п. 8 приведены обобщения и модификации этого результата для случая нильпотентных групп Ли.

Замечания. По-видимому, некоторая степень кривизны (в виде условия «конечности типа» или ослабленного условия выпуклости) так или иначе необходима для справедливости утверждения (b) теоремы 5. Действительно, существует пример гладкой кривой на плоскости — а именно кривой, имеющей в начале координат касание бесконечного порядка с некоторой прямой, — для которой рассматриваемые неравенства не выполняются ни для одного $p < \infty$. Однако, в резком контрасте с ситуацией, рассмотренной в пп. 1 и 4, существуют (помимо прямых линий) кривые, которые имеют в начале координат касание бесконечного порядка с некоторой прямой и для которых положительный результат верен для всех p. Ниже в п. 15 этот вопрос рассмотрен более подробно.

Набросаем кратко основные идеи доказательства теоремы 5. Применяя квадратичные функции, можно убедиться в том, что локальный максимальный оператор M_1 ограничен на L^2, подобно тому как это было сделано в п. 3. Для того чтобы определить требуемую квадратичную функцию, поступим следующим образом. Пусть $\rho(u) \sim \sum a_\alpha u^\alpha$ — разложение ρ в ряд Тэйлора в начале координат; здесь a_α — векторы в $\mathbb{R}^k$. Используя эти векторы, построим попарно непересекающиеся (возможно, пустые) подпространства $V_1, V_2, \ldots$ пространства $\mathbb{R}^k$ так, чтобы их сумма $V_1+V_2+\ldots+V_m$ совпадала с линейной оболочкой системы векторов $\{a_\alpha\}_{|\alpha|\leqslant m}$. Поскольку ρ — конечного типа в начале координат, то для некоторого N имеем $\sum\limits_{j=1}^{N} V_j = \mathbb{R}^k$. Определим теперь растяжение δ_t формулой $\delta_t(v) = t^j v$, $v \in V_j$, $t > 0$. Тогда $\delta_t^{-1}\rho(tu) \to \rho_0(u)$ при $t \to 0$, где ρ_0 — снова полином конечного типа в начале координат. Выберем функции $\varphi \in C_0^\infty(\mathbb{R}^k)$ и $\psi \in C_0^\infty(\mathbb{R}^n)$, $\varphi \geqslant 0$, $\psi \geqslant 0$, для которых $\int\limits_{\mathbb{R}^k} \varphi\, du = \int\limits_{\mathbb{R}^n} \psi\, dy = 1$, и положим

$$M^j(f) = \int\limits_{\mathbb{R}^k} f(x-\rho(2^{-j}u))\,\varphi(u)\,du,$$

$$N^j(f) = \int\limits_{\mathbb{R}^n} f(x-\delta_{2^{-j}}(y))\,\psi(y)\,dy.$$

Требуемой квадратичной функцией будет

$$S(f)(x) = \Bigl(\sum_{j\geqslant 0} |M^j(f)(x) - N^j(f)(x)|^2\Bigr)^{1/2}.$$

Оценка $\|S(f)\|_{L^2} \leqslant A\|f\|_{L^2}$ без труда следует из (слегка видоизмененной) оценки, полученной в лемме 1 п. 2. Далее L^p-оценки для M_1 проводятся таким же образом, как и в случае кривой; существенным здесь является наличие «ε-люфта», гарантируемое леммой 1.

Следует заметить, что для оператора T из (6.1) L^2-результат эквивалентен, в силу теоремы Планшереля, следующей оценке для осцилляторных интегралов. Пусть P_0 — произвольный вещественнозначнынй полином на $\mathbb{R}^k$. Тогда

$$\left|\text{p. v.} \int\limits_{\mathbb{R}^k} e^{iP_0(u)} K(u)\,du\right| \leqslant A, \tag{6.3}$$

где константа может быть выбрана зависящей лишь от степени полинома P_0, но не от коэффициентов. См. по этому поводу Stein [103].

ЧАСТЬ II. НИЛЬПОТЕНТНЫЕ ГРУППЫ ЛИ

7. Нильпотентные группы Ли: некоторые предварительные сведения. Кратко обсудим некоторые стандартные аспекты анализа на нильпотентных группах Ли, имеющие отношение к интересующей нас теме[1].

Пусть N — односвязная нильпотентная группа Ли, которую с помощью экспоненциального отображения мы будем всегда отождествлять с её алгеброй Ли n. Предположим, что фиксированы разложение n в прямую сумму: $n = \sum\limits_{j=1}^{m} n_j$ и последовательность строго положительных чисел $a_1, \ldots, a_m$, с помощью которых определим растяжения $\{\delta_t\}$, $t > 0$, на n по формуле $\delta_t(v) = t^{a_j} v$ для $v \in n_j$. В общем случае такие растяжения не будут автоморфизмами алгебры n; в случае же, когда это так, мы будем называть их *автоморфными растяжениями*. Группу N вместе с этими растяжениями назовём однородной группой. Примером служит гейзенбергова группа H^1, отождествлённая с $\mathbb{R}^3 = \{(x, y, z)\}$; её закон умножения имеет вид

$$(x,\ y,\ z)\cdot(x',\ y',\ z') = (x+x',\ y+y',\ z+z'+x'y-xy'). \tag{7.1}$$

Укажем здесь два различных типа растяжений на H^1:

$$\begin{aligned}&(\text{а})\ \delta_t\colon (x,\ y,\ z) \to (tx,\ ty,\ t^2z),\\ &(\text{в})\ \delta_t\colon (x,\ y,\ z) \to (tx,\ ty,\ tz).\end{aligned} \tag{7.2}$$

Растяжения первого типа автоморфны; они тесно связаны с геометрией группы H^1, реализованной как граница некоторой области в $\mathbb{C}^2$. О растяжениях второго типа (которые не являются автоморфизмами) можно сказать, что это — «следы» применения теории эллиптических уравнений, насколько она применима к $\bar\partial$-задаче Неймана для этой области.

Рассмотрим сначала случай однородных нильпотентных групп. Более общая ситуация будет обсуждена в п. 8. Кстати,

---

[1] Кроме работ, на которые мы будем ссылаться ниже, читатель может ознакомиться также с обзором Goodman [36].

в примерах, указанных после (7.3), все группы являются однородными.

Структура однородной группы естественным образом приводит к некоторым основным геометрическим конструкциям. Это — (левоинвариантное) квазирасстояние $\rho(x, y)$, однородное относительно растяжений и определяемое формулой $\rho(x, y) = |y^{-1}x|$, где $|\cdot|$ — подходящая однородная норма на N, а также функция объёма $V(t)$, задающая объём шара радиуса t (при этом расстояние измеряется с помощью функции ρ). В силу условия однородности, $V(t) = ct^Q$, $t > 0$, где Q — однородная размерность N.

Теперь мы можем определить «стандартные сингулярные интегралы» на N в тесной аналогии с классическим случаем пространства $\mathbb{R}^n$ (см. (0.2)). Итак, рассмотрим операторы свертки T: $f \to f * K$, где ядро (в смысле главного значения) K удовлетворяет следующим условиям:

(a) K гладко всюду, кроме начала координат;

(b) $K(\delta_t x) = t^{-Q}K(x), t > 0$; (7.3)

(c) среднее значение K равно нулю.

Вот примеры таких операторов: (1) сплетающие операторы (см. Knapp — Stein [54]); (2) оператор Коши — Сегё для обобщённого верхнего полупространства (см. Koranyi — Vagi [55]); (3) операторы, появляющиеся в критических оценках для лапласиана Кона $\Box_b$ (см. Folland — Stein [34]) и, более общим образом, при изучении классов Хёрмандера [48] (см. Rothschild — Stein [84]).

Однако мы не собираемся подробно останавливаться здесь на этих «стандартных» операторах, которые к настоящему времени изучены достаточно хорошо. Обратимся к операторам на нильпотентных группах, носителями ядер которых служат подмногообразия малой размерности, по аналогии с рассмотрениями части I. При проведении детального анализа $\bar{\partial}$-задачи Неймана мы по существу вынуждены рассматривать такие обобщения. Дело в том, что получающиеся явные формулы (по поводу разрешающих операторов для оператора $\bar{\partial}$ см. Henkin [45], Lieb [57], а также обзор Krantz [56]; по поводу $\bar{\partial}$-задачи Неймана см. Phong [76], Harvey — Polking [43], Lieb — Range [58], Stanton [94] и обзор Beals — Fefferman — Grossman [1]) приводят к ядрам, являющимся произведениями ядер двух типов однородности (7.2). Каким же образом удаётся «сопрячь» два столь различных вида однородности? Оказывается, ядро

$$\tilde{K}(x, y, z) = L(x, y)\Delta(z) \qquad (7.4)$$

на H^1, где $L(x, y)$ — стандартное ядро Кальдерона — Зигмунда на $\mathbb{R}^2$, а $\Delta(z)$ — дельта-функция Дирака, одновременно однородно относительно обоих этих типов растяжений, причем для каждого — в критической степени! Поэтому неудивительно, что анализ оператора $\tilde{T}$, задаваемого формулой $\tilde{T}f = f * \tilde{K}$, столь важен для решения $\bar{\partial}$-задачи Неймана.

С учётом особенностей строения ядра $\tilde{K}$ (см. формулу (7.4)) L^2-теорию оператора $\tilde{T}$ проще всего строить с помощью преобразования Фурье по переменной z. Тогда L^2-теория оператора $\tilde{T}$ сводится к L^2-теории некоторых осцилляторных операторов, которые можно задать следующим образом. Пусть m фиксировано, $B(\cdot, \cdot)$ — вещественнозначная билинейная форма на $\mathbb{R}^m \times \mathbb{R}^m$ и K — стандартное ядро Кальдерона — Зигмунда на $\mathbb{R}^m$. Рассмотрим оператор T, определяемый равенством

$$(Tf)(x) = \text{p. v.} \int_{R^m} e^{iB(x, y)}K(x - y) f(y)\, dy. \qquad (7.5)$$

Для гейзенберговой группы $\mathbf{H}^n$ возьмём $m = 2n$, $B(x, y) = \lambda\langle x, y\rangle$, где $\langle\cdot, \cdot\rangle$ — симплектическая форма на $\mathbb{R}^{2n} \times \mathbb{R}^{2n}$, определяющая законы умножения в $\mathbf{H}^n$, $-\infty < \lambda < \infty$. В этом случае операторы (7.5) представляют собой «сплетающие свёрточные» операторы на гейзенберговой группе — операторы, представляющие интерес и сами по себе (см., например, Segal [89], A. Grossman — Loupias — Stein [40], Howe [52], Mauceri — Picardello — Ricci [60]).

Подведём итог всему вышеизложенному. По аналогии с частью I мы выделили некоторую группу свёрточных операторов на подмногообразиях низкой размерности и порождаемые ими осцилляторные интегралы; в этом контексте условие кривизны заключается в невырожденности формы $\langle\cdot, \cdot\rangle$. Более геометрическая трактовка этого понятия кривизны состоит в следующем. Каждой точке $P = (x, y, z)$ в H^1 сопоставляется двумерная (аффинная) плоскость M_P, где M_P — левый групповой сдвиг M_0 с помощью P, $M_0 = \{(x, y, 0)\}$. Значение оператора $\tilde{T}(f)$ в P задаётся интегрированием по M_P; кривизна здесь означает «поворот» плоскости M_P при изменении точки P. Теория операторов $\tilde{T}$ и операторов, задаваемых формулой (7.5), развита в работах Geller — Stein [35] и Phong — Stein [77—79]. Ниже в пп. 8 и 11 представлены некоторые другие относящиеся сюда результаты.

8. Некоторые общие результаты для нильпотентных групп Ли. Перейдём теперь к изложению ряда результатов, которые позволяют перенести многое из теории для $\mathbb{R}^n$, развитой в § 6, на нильпотентные группы Ли, рассмотренные в п. 7.

Ниже N — произвольная односвязная нильпотентная группа, а $\{\delta_t\}$, $t>0$, — фиксированное семейство растяжений, не обязательно являющихся автоморфизмами. Фиксируем какое-нибудь вещественно-аналитическое подмногообразие V в $N/\{0\}$, которое ради простоты будем считать связным, а также предположим, что V однородно в том смысле, что $\delta_t(V)=V$ для всех $t>0$. Пусть задано распределение (обобщённая функция) K на N, инвариантное относительно растяжений (т. е. такое что $\langle K, f(x)\rangle = \langle K, f(\delta_t x)\rangle$ для всех $t>0$) и вне начала координат представляющее собой меру на V с гладкой плотностью (по отношению к индуцированной лебеговой мере $d\sigma$ на V), т. е. имеющее вид $\psi(x)d\sigma(x)$, где ψ — гладкая функция, ограничение которой на множество $\{|x|=1\}\cap V$ имеет компактный носитель. Определим T формулой

$$Tf=f*K. \tag{8.1}$$

При тех же условиях можно рассмотреть аналогичную максимальную функцию. Пусть $d\mu=\chi_\Gamma d\sigma$, где χ_Γ — характеристическая функция компактного подмножества Γ в V. Определим $d\mu_t$ равенством $\langle d\mu_t, f(x)\rangle = \langle d\mu, f(\delta_t x)\rangle$, $t>0$, и положим

$$(Mf)(x)=\sup_{t>0}|(f*d\mu_t)(x)|. \tag{8.2}$$

Теорема 6. *Операторы T и M, определённые по формулам* (8.1) *и* (8.2), *являются ограниченными операторами в пространстве $L^p(N)$, $1<p$.*

Стоит отдельно сформулировать один частный случай этого результата. Предположим, что K — ядро в смысле главного значения, удовлетворяющее условиям (7.3), носитель которого совпадает со всем N, и пусть $Tf=f*K$. Аналогично если Γ — открытое подмножество в N с компактным замыканием, положим

$$(Mf)(x)=\sup_{t>0}\left|\int_\Gamma f\left(x\cdot(\delta_t(y))^{-1}\right)dy\right|. \tag{8.3}$$

Следствие 1. *Определённые таким образом операторы T и M ограничены в $L^p(N)$, $1<p$.*

Отсюда видно, что требование, чтобы группа имела автоморфные расширения, раньше считавшееся существенным при доказательстве таких результатов, на самом деле не является необходимым. Заметим, однако, что шары, о которых идёт речь в определении (8.3), не удовлетворяют условию покрытия Витали (а именно, если N не является однородной группой, то для $|y^{-1}\cdot x|=\rho(x,y)$ не выполняется квазинеравенство треугольника. Следовательно, обычные методы здесь неприменимы.

Другим следствием (доказательства) теоремы является такое обобщение утверждения (а) теоремы 5. Пусть P — полиномиальное отображение $\mathbb{R}^k$ в N и K — обычное ядро Кальдерона — Зигмунда на $\mathbb{R}^k$.

Следствие 2. *Операторы*

$$f\mapsto \text{p.v.}\int_{\mathbb{R}^k} f\left(x\cdot(P(u))^{-1}\right)K(u)\,du$$

и

$$f\mapsto \sup_{0<r} r^{-k}\left|\int_{|u|\leqslant r} f\left(x\cdot(P(u))^{-1}\right)du\right|$$

ограничены в $L^p(N)$, $1<p<\infty$. Константы в оценках их норм не зависят от коэффициентов полинома P, а зависят лишь от его степени.

Наконец, следствие 3 «объединяет» следствие 1 с осцилляторными интегралами из (6.3) и (7.5).

Следствие 3. *Пусть $P(x,y)$ — фиксированный вещественный полином на $N\times N$. Тогда оператор*

$$f\mapsto \text{p. v.}\int_N e^{iP(x,y)}K(y^{-1}\cdot x)f(y)\,dy \tag{8.4}$$

ограничен в $L^p(N)$, $1<p<\infty$, и константы в оценке его нормы зависят лишь от степени полинома P (постольку, поскольку речь идёт о зависимости от P).

Кроме уже указанной литературы можно сослаться на работы Strichartz [109], Müller [63, 64], Christ [14] и Greenleaf [39], в которых рассмотрены некоторые случаи операторов вида (8.1), (8.2) и (8.4). Общая теорема, как она сформулирована выше, и следствия из нее содержатся в недавней работе Риччи и автора [81—83].

Кратко изложим теперь три основные идеи, используемые в доказательстве теоремы. Первая идея заключается в переходе от нильпотентной группы N с неавтоморфными растяжениями δ_t к более широкой нильпотентной группе G, обладающей соответствующими автоморфными растяжениями и такой, что N является её факторгруппой. По сути дела в качестве G мы выбираем «свободную» нильпотентную группу достаточно высокой ступени, алгебра Ли g которой порождена базисом алгебры n. В этом случае свойства ограниченности операторов T и M можно свести к свойствам ограниченности соответствующих операторов T' и M' на G (строящихся аппроксимативно). В действительности эта процедура представляет собой упрощенный вариант метода «лифтинга», к рассмотрению которого мы ещё вернемся в п. 10.

Вторая идея основана на связи свойства «кривизны» нашего многообразия V с фактом его вещественной аналитичности, а также с возможностью предполагать, что оно порождает группу N. Эта ситуация обыгрывается в следующей лемме:

Лемма. *Пусть V есть k-мерное вещественно-аналитическое подмногообразие в связной группе Ли G. Предположим, что V порождает G. Пусть, далее, $d\mu$ — мера, сосредоточенная на V, плотность которой является гладкой функцией с компактным носителем. Тогда $d\nu = d\mu * d\mu * \dots * d\mu$ ($2^{\dim(G)-k}$ множителей) как мера на G абсолютно непрерывна по отношению к мере Хаара: $d\nu = h dx$, где h имеет L^1-модуль непрерывности и удовлетворяет условию Гёльдера с показателем $\varepsilon > 0$.*

Третья идея заключается в использовании подходящих свойств ортогональности. Обсудим эту идею более подробно.

9. Квадратичные функции и ортогональность. Имеются три вариации на тему ортогональности.

(*i*) *Метод TT^*.* Речь идёт об очень простой вещи: для того чтобы доказать ограниченность оператора T в L^2, достаточно сделать это для TT^* (или T^*T).

Это наблюдение полезно для приложений по той причине, что если оператор T представляется ядром K, т. е. $(Tf)(x) = \int K(x, y) f(y)\,dy$, то TT^* представляется ядром

$$\int K(x, z)\bar{K}(y, z)\,dz, \tag{9.1}$$

и это ядро часто лучше, чем K, потому, что интегрирование в (9.1) может одновременно обладать сглаживающим действием и проявлять свойства «погашения» (cancellation) ядра K. Классическими примерами служат унитарные операторы, такие как преобразование Фурье или преобразование Гильберта. Менее тривиальный пример доставляет теорема Колмогорова — Селиверстова — Плеснера (см. Zygmund [115], из которой, между прочим, видно, что вычисления, необходимые для доказательства ограниченности TT^*, могут существенно отличаться от вычислений, проводимых при доказательстве ограниченности T^*T. Другие примеры возникают в теории полугрупп (см. Stein [96]) и в теории интегральных операторов Фурье (см. Hörmander [49]). Ближе к нашей теме это — осцилляторные интегралы вида (7.5), (8.4) и рассмотренные ниже интегралы вида (11.4) и (12.3), а также исследование преобразования Гильберта и максимальные функции для «переменных кривых» (см. п. 11), как оно приведено в работе Nagel — Stein — Wainger [69].

(*ii*) *Почти-ортогональность.* Не всегда для получения желаемого результата достаточно рассмотреть оператор TT^*. В некоторых случаях приходится прибегать к более тонким построениям. Например, можно рассуждать следующим образом. Оператор T ограничен в L^2, если он представлен в виде суммы ряда $\sum T_j$, у которого нормы членов равномерно ограничены, а сами члены приближённо взаимно ортогональны; последнее условие выполняется, если потребовать, чтобы нормы как $T_iT_j^*$, так и $T_i^*T_j$ стремились к нулю при $|i-j| \to \infty$. Это лемма о почти-ортогональности, принадлежащая Котлару и автору. Она была использована для доказательства ограниченности в L^2 стандартных сингулярных интегралов на нильпотентных группах, о которых шла речь выше (см. Knapp — Stein [54]), и затем было найдено много её приложений и модификаций. Один такой уместный здесь вариант леммы (полученный Кристом, а в другой форме — Гринлифом [38]) состоит в наблюдении, что поскольку нормы операторов T_j ограничены, контроль за поведением $T_iT_j^*$ можно свести (путём последовательного применения принципа TT^*) к контролю за поведением $(T_i^*T_i)^k T_j^*$ для любого фиксированного (большого) целого k. Это особенно важно в связи с доказательством леммы п. 8.

(*iii*) *Новые квадратичные функции.* Мы уже рассмотрели несколько примеров квадратичных функций (в пп. 3, 4) и обсудили их полезность при доказательстве неравенств в L^p. Представим ещё один вариант квадратичных функций, имеющий непосредственное отношение к почти-ортогональным разложениям, о которых только что шла речь. Предположим, что можно написать $T = \sum T_j$, где T_j представляется в виде произведения $T_j = B_j^* A_j$. Построим две квадратичные функции S_1 и S_2 следующим образом:

$$S_1(f) = \Big(\sum_j |A_j f|^2\Big)^{1/2}, \quad S_2(f) = \Big(\sum_j |B_j f|^2\Big)^{1/2}.$$

Оказывается, ограниченность оператора T в L^p прямо следует из ограниченности преобразований $f \to S_1(f)$ и $g \to S_2(g)$ в L^p и $L^{p'}$ соответственно. Этот вид разложений и использование квадратичных функций играют важную роль в теории сингулярных интегралов Койфмана — Макинтоша — Мейера [16] и Давида — Журнэ [24]; читателю полезно также ознакомиться с работой David [23].

ЧАСТЬ III. ПЕРЕМЕННЫЕ КОЭФФИЦИЕНТЫ

10. Геометрия и анализ векторных полей. Переход к переменным коэффициентам через посредство нильпотентных групп лучше всего иллюстрируется примерами из области анализа и геометрии векторных полей. Идея использовать нильпотентные группы первоначально возникла при изучении оператора $\overline{\partial}_b$ и лапласиана Кона, а затем была перенесена на случай операторов, исследованных Хёрмандером [48]. Коротко опишем её.

Пусть $X_1, \ldots, X_m$ — гладкие вещественные векторные поля в (окрестности начала координат) $\mathbb{R}^n$. Предположим, что эти векторные поля и их коммутаторы порядка не выше k порождают всё $\mathbb{R}^n$ (вблизи начала координат). Тогда можно «поднять» эти векторные поля, поместив их в некоторое более широкое пространство $\mathbb{R}^n \times \mathbb{R}^{n'}$ (с дополнительными переменными $t \in \mathbb{R}^{n'}$), так, чтобы поднятые векторные поля $\tilde{X}_i$ имели вид

$$\tilde{X}_i = X_i + \sum_{j=1}^{n'} a_{ij}(x, t) \frac{\partial}{\partial t_j}, \qquad i = 1, \ldots, m, \tag{10.1}$$

и обладали двумя свойствами:

(i) поля $\tilde{X}_i$ и их коммутаторы длины $\leqslant k$ порождают всё $\mathbb{R}^n \times \mathbb{R}^{n'}$ (вблизи) начала координат;

(ii) поля $\tilde{X}_i$ «свободны», т. е. они сами и их коммутаторы удовлетворяют тем же соотношениям, что и элементы свободной нильпотентной алгебры Ли $\mathcal{F}_{m,k}$ (эта алгебра Ли получается как факторалгебра свободной алгебры Ли, порождённой m элементами $Y_1, Y_2, \ldots, Y_m$, по идеалу, порожденному всеми коммутаторами длины $>k$).

Помимо этого лифтинга, имеется также важное свойство аппроксимации. Запишем векторные поля $Y_1, \ldots, Y_m$ в канонических координатах $(y_1, \ldots, y_{n+n'})$:

$$Y_i = \sum_j b_{ij}(y)\, \partial/\partial y_j.$$

Для каждой фиксированной точки $(x_0, t_0) \in \mathbb{R}^n \times \mathbb{R}^{n'}$ вблизи начала можно найти систему координат $(y_1, \ldots, y_{n+n'})$ с центром в (x_0, t_0), такую что

$$\tilde{X}_j = Y_j + R_j, \qquad j = 1, \ldots, m, \tag{10.2}$$

где R_j убывает достаточно быстро при $(x, t) \to (x_0, t_0)$ (т. е. при $y_1, \ldots, y_{n+n'}) \to 0$).

Описанная выше двойная процедура позволяет свести изучение, например, оператора $\sum X_j^2$ к изучению оператора $\sum \tilde{X}_j^2$, который в свою очередь моделируется левоинвариантным (и однородным) оператором $\sum Y_j^2$ на нильпотентной группе G, алгеброй Ли которой служит $\mathcal{F}_{m,k}$. Это обращение к анализу на нильпотентных группах Ли восходит к работам Folland — Stein [34] и Rothschild — Stein [84]; с тех пор эти идеи получили существенное развитие (в этой связи см., скажем, обзор Helffer — Nourigatt [44]).

Далее, можно попытаться прийти к лучшему пониманию геометрии векторных полей, например «размеров» фундаментального решения оператора $\sum X_j^2$. Для этого требуются, во-первых, подходящая «метрика» (или квазирасстояние) и, во-вторых, формула для объёма шаров в этой метрике.

Если заданы векторные поля $X_1, \ldots, X_m$, мы можем определить метрику естественным образом — взять в качестве расстояния $\rho(x, y)$ кратчайшее время, которое требуется для того, чтобы, двигаясь с «единичной скоростью», пройти путь от x до y вдоль кривой, идущей в направлениях $X_1, \ldots, X_m$. А именно, $\rho(x, y) = \inf T$, где точная нижняя грань берётся по множеству спрямляемых кривых $\gamma(t)$, $0 \leqslant t \leqslant T$, таких что $\gamma(0) = x$, $\gamma(T) = y$ и

$$\dot{\gamma}(t) = \sum_{j=1}^{m} a_j(t) X_j(\gamma(t)), \quad |a_j(t)| \leqslant 1.$$

Для того чтобы было удобнее работать с таким расстоянием, весьма полезно описать его в нескольких эквивалентных формах. Здесь мы рассмотрим только одну такую форму, связанную с экспоненциальным отображением, а кроме того получим формулу для объёма $V_x(\delta)$ шара $\{y \mid \rho(x, y) < \delta\}$.

Для каждого заданного набора l целых чисел $I(i_1, \ldots, i_l)$, $|I| = l$, где $1 \leqslant i_j \leqslant m$ и $l \leqslant k$, определим векторное поле X_I формулой

$$X_I = [X_{i1} [X_{i2} \ldots [X_{i_{l-1}}, X_{i_l}] \ldots]].$$

Далее, для каждого набора J из n таких полей, $J = \{I_1, I_2, \ldots, I_n\}$ положим $|J| = |I_1| + |I_2| + \ldots + |I_n|$ и определим λ_J

равенством

$$\lambda_I(x) = \det\left(X_{I_1}, X_{I_2}, \ldots, X_{I_n}\right)(x).$$

Величины λ_I можно рассматривать как обобщения инвариантов Леви, к которым они и сводятся в некоторых случаях.

Теорема 7. (a) *Для достаточно малых δ неравенство $\rho(x, y) < \delta$ имеет место тогда и только тогда, когда*

$$y = \exp\left(\sum_I a_I X_I\right)(x),$$

где a_I — константы, удовлетворяющие условию $|a_I| < c\delta^{|I|}$.

(b) $V_x(\delta) \approx \sum_I |\lambda_I(x)| \delta^{|I|}$.

Из утверждения (b) видно, в частности, что объем шаров обладает важным свойством удвоения.

Приведём теперь оценки для параметрикса $K(x, y)$ оператора $\sum_{j=1}^{m} X_j^2$:

Теорема 8.

$$|X_{i_1} \ldots X_{i_r} K(x, y)| \leqslant c_r \frac{(\rho(x, y))^{2-r}}{V_x(\delta)}, \tag{10.3}$$

где $\delta = \rho(x, y)$.[1]

Замечание. Отметим, что для однородных нильпотентных групп вид таких оценок определяется масштабной инвариантностью.

Две приведённые выше теоремы, а также излагаемые ниже результаты можно найти в работах Nagel — Stein — Wainger [70, 71]; см. также Nagel [65]. К теореме 8 можно прийти другим способом, получив при этом и другие результаты. Этот подход развит в работах Fefferman — Phong [31], Fefferman [30], Sanchez-Calle [87] и Fefferman — Sanchez-Calle [32]. У него то преимущество, что с его помощью можно рассматривать общие субэллиптические операторы вида $\sum a_{ij}(x)\partial^2/\partial x_i \partial x_j$ с полуопределёнными и гладкими $\{a_{ij}(x)\}$. С другой стороны, описанный выше подход применим к общему оператору Хёрмандера $\sum X_j^2 + X_0$, а поэтому и к любому субэллиптическому оператору, являющемуся «однородным» полиномом от векторных полей, и это позволяет получить оценку вида (10.3) для параметриксов таких операторов.

11. *Сингулярные преобразования Радона.* Эту тему можно рассматривать как кульминацию перехода от рассмотренной в п. 6 теории сингулярных интегралов в трансляционно-инвариантном случае пространств $\mathbb{R}^n$ к случаю нильпотентных групп, разобранному в пп. 7 и 8, а тем самым и к общему случаю «переменных коэффициентов». Однако полной картины здесь, как в отношении оптимальных формулировок общих положений, так и в отношении деталей предполагаемых доказательств, пока ещё нет. Поэтому мы остановимся лишь на одном важном случае, который представляет собой интерес, поскольку имеет непосредственное отношение к $\bar{\partial}$-задаче Неймана.

Пусть M — (компактное) многообразие размерности $m+1$, $m > 1$. Имитируя ситуацию, рассмотренную в конце п. 7, предположим, что для каждой точки $P \in M$ существует подмногообразие M_P размерности m, такое что $P \in M_P$; далее, для каждой точки P зададим ядро Кальдерона — Зигмунда плотностью $K(P, \cdot)$ на M_P с сингулярностью в P так, чтобы отображения $P \to M_P$ и $P \to K(P)$ были гладкими. Определим сингулярное преобразование Радона T (отображающее $C^\infty(M)$ в себя) по формуле

$$(Tf)(P) = \int_{M_P} K(P, \cdot) f_P(\cdot), \tag{11.1}$$

где f_P обозначает сужение f на M_P.

Оказывается, условие кривизны, появляющиеся в этом случае, связано с аналогичным условием, возникающим (см. Guillemin — Sternberg [41]) в задаче об обращении для преобразования Радона более стандартного вида (например, с везде гладким ядром $K(P, \cdot)$ в (11.1)).

Пусть $\Phi(P, Q) = 0$ — определяющее уравнение для соотношения $Q \in M_P$, причём функция невырожденна (т. е. $d_P\Phi \neq 0$, $d_Q\Phi \neq 0$). Тогда форма «вращательной кривизны» $L_\Phi = L_\Phi(P)$ — это билинейная форма, определённая для $(v_1, v_2) \in T_P(M_P) \times T_P(M_P)$ равенством

$$L_\Phi(v_1, v_2) = \langle d^2_{PQ}\Phi|_{P=Q} v_1, v_2 \rangle,$$

где гессиан $d^2_{PQ}\Phi$ представляет собой дифференциал отображения $\Phi \to d_P\Phi(P, Q)$ (этот дифференциал есть линейное

---

[1] Конечно, особняком стоит случай $n = 2$, когда поля $X_1, \ldots, X_m$ уже порождают рассматриваемое пространство; в этом случае справедливость оценки (10.3) утверждается только для $r \geqslant 1$.

отображение $T_Q(M)$ в $T_P(M)^*$). Условие кривизны выглядит следующим образом:

(11.2) Форма вращательной кривизны L невырожденна в каждой точке $P \in M$.

Теорема 9. *Оператор T, заданный формулой* (11.1), *ограничен в $L^p(M)$, $1 < p < \infty$, если выполнено условие кривизны* (11.2).

(См. работы Phong — Stein [77, 78]; описание более раннего подхода для случая переменных кривых на плоскости можно найти в работе Nagel — Stein — Wainger [69].)

Замечания. 1. Получены и L^p-неравенства для соответствующей максимальной функции, которая так же связана с T, как M_1 связана с T_1 (см. (6.1′) и (6.2′)).

2. Ульманн [114] указал на связь операторов типа (11.1) и интегральных операторов Фурье с сингулярными символами, ассоциированными с парой лагранжианов. По поводу его класса операторов см. Guillemin — Uhlmann [42], а также Melrose — Uhlmann [61].

3. Для более ясного понимания условия кривизны следует остановиться на двух случаях. Во-первых, в трансляционно-инвариантном случае пространства $\mathbb{R}^{n+1}$ это условие означает, что M_0 обладает ненулевой гауссовой кривизной в начале координат. Во-вторых, в случае когда $\mathscr{D}$ — гладкая область в $\mathbb{C}^{n+1}$, а M — её граница ($m = 2n$), для некоторого естественным образом присоединённого семейства $\{M_P\}$ это условие эквивалентно невырожденности формы Леви.

Идея доказательства теоремы заключается прежде всего в том, чтобы локализовать систему координат $(t, x) \in \mathbb{R} \times \mathbb{R}^m$ таким образом, чтобы если $P = (t, x)$, $Q = (s, y)$, то условие $Q \in M_P$ задавалось в виде $s = t + S(t, x, y)$, где $S(t, x, x) \equiv 0$, а значит (11.2) можно записать так:

$$\det\left(\frac{\partial^2 \Phi(x, y)}{\partial x_i \, dy_j}\right) \neq 0, \tag{11.3}$$

где $\Phi(x, y) = S(t, x, y)$.

После преобразования Фурье по переменной t можно в конечном счёте свести проблему L-ограниченности формы (11.1) к рассмотрению осцилляторных интегралов вида

$$I_\lambda : f \to \int_{R^n} e^{i\lambda\Phi(x, y)} K(x, y) f(y)\, dy. \tag{11.4}$$

Здесь K — ядро с компактным носителем некоторого стандартного псевдодифференциального оператора порядка μ, где $-m < \mu \leqslant 0$.

Лемма. *Пусть Φ удовлетворяет условию* (11.3) *и $-m < \mu \leqslant 0$. Тогда для нормы оператора I_λ справедлива оценка*

$$\| I_\lambda \|_{L^2 \to L^2} \leqslant A(1 + |\lambda|)^{\mu/2}. \tag{11.5}$$

Отметим связь между осцилляторным интегралом (11.4) и осцилляторными интегралами в (7.5), (8.4) и (12.3) (см. ниже). Частный случай, когда K является ядром оператора из класса $\mathrm{Op}\left(S_{1,0}^{-\infty}\right)$, рассмотрен в работе Hörmander [50]; ничего лучше, чем $(1 + |\lambda|)^{-m/2}$ в (11.5) получить нельзя. Связь теоремы 9 и $\bar{\partial}$-задачи Неймана обсуждается в работе Phong — Stein [79].

12. Средние по переменным гиперповерхностям. Вернёмся к нашей первой теме — максимальным средним по поверхностям, которая обсуждалась в пп. 1 и 4, и попытаемся получить диффеоморфно-инвариантные версии приведённых там результатов. Примем следующие допущения.

Для любого $x \in \mathbb{R}^n$ и любого $\varepsilon \geqslant 0$ задана гладкая гиперповерхность $S_{\varepsilon, x} \subset \mathbb{R}^n$. Семейство поверхностей $\{S_{\varepsilon, x}\}$ будем описывать как образцы некоторой фиксированной гиперповерхности $\bar{S}$, т. е. предположим, что $S_{\varepsilon, x} = \rho_{\varepsilon, x}(\bar{S})$, где $\rho_{\varepsilon, x}$ для каждой пары (ε, x) — отображение вложения, причём $\rho_{\varepsilon, x}(u)$ — гладкая функция от $(\varepsilon, x, u) \in [0, \infty) \times \mathbb{R}^n \times \bar{S}$. Зафиксируем срезающую функцию $\psi \in C_0^\infty(\mathbb{R}^n)$ и определим оператор среднего значения формулой

$$M_\varepsilon(f)(x) = \int_{\bar{S}} f(x + \varepsilon\rho_{\varepsilon, x}(u))\, \psi(x)\, d\sigma(u), \tag{12.1}$$

где $d\sigma$ — некоторая мера с гладкой плотностью и компактным носителем на $\bar{S}$. Условия кривизны в данном случае выглядят следующим образом:

(12.2) (a) Для любой точки $x \in \mathbb{R}^n$ гиперповерхность $S_{0, x}$ обладает гауссовой кривизной, не имеющей нуля бесконечного порядка ни в одной точке..

(b) То же, что и в (a), только гауссова кривизна предполагается вообще всюду отличной от нуля.

Заметим. что эти условия диффеоморфно-инвариантны в том смысле, что если $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ — диффеоморфизм, то

$$\Phi(x + \varepsilon\rho_{\varepsilon, x}(u)) = y + \varepsilon\tilde{\rho}_{\varepsilon, y}(u), \quad \text{если} \quad \Phi(x) = y,$$

где $\tilde{\rho}$ удовлетворяет тем же условиям (12.2) ((a) или (b)), что и ρ.

Теорема 10. *Пусть*

$$\mathscr{M}(f)=\sup_{0<\varepsilon\leqslant\varepsilon_0}|M_{\varepsilon_0}(f)|.$$

Если ε_0 достаточно мало, то отображение $f\to\mathscr{M}(f)$ ограничено в L^p в следующих случаях:

(i) когда выполняется (12.2(a)), $n\geqslant 6$ и $p>p_0$ для некоторого конечного $p_0=p_0(\{S_{\varepsilon,x}\})$;

(ii) когда выполняется (12.2(b)), $n\geqslant 3$, $p>n/(n-1)$.

Утверждение (ii) теоремы доказано в работе Greenleaf [37], однако применённое там рассуждение не обобщается на случай утверждения (i). Что касается доказательства этого последнего (см. Sogge — Stein [93]), то, хотя оно, строго говоря, и не основано на использовании нильпотентных групп, тем не менее у него много общего с доказательством теоремы 9 для сингулярных преобразований Радона. По сути дело сводится к тому, чтобы после выбора подходящей системы координат, положив $n=m+1$, с помощью преобразования Фурье по одной переменной исследовать ключевую квадратичную функцию. (Именно на этом этапе доказательства используется условие $n\geqslant 6$.) Снова приходим к осцилляторным интегральным операторам

$$\tilde{I}_\lambda:\ f\to\int_{\mathbb{R}^m}e^{i\lambda\Phi(x,y)}|H(x,y)|^N\psi(x,y)f(y)\,dy, \tag{12.3}$$

где Φ — некоторая вещественная функция, $H(x,y)$ — её гессиан, т. е. $\det\left(\partial^2\Phi/\partial x_i\,\partial x_j\right)$ и $\psi\in C_0^\infty\left(\mathbb{R}^m\times\mathbb{R}^m\right)$. В свойствах функции Φ и её гессиана фактически выражены требуемые условия кривизны.

По аналогии с леммой 2 п. 2 и леммой п. 1 можно доказать следующую L^2-оценку для нормы операторов I_λ:

Лемма. *Для достаточно больших N*

$$\|\tilde{I}_\lambda\|_{L^2\to L^2}\leqslant A(1+|\lambda|)^{-m/2}. \tag{12.3}$$

Фактически (12.3) *выполняется для* $N\geqslant 5m/2$.

Теоремой 10 подсказывается следующая

Задача. Можно ли и в (i) и в (ii) ослабить условие на n до $n\geqslant 2$?

ЧАСТЬ IV. НЕКОТОРЫЕ ДАЛЬНЕЙШИЕ РЕЗУЛЬТАТЫ

Кратко остановимся на некоторых направлениях гармонического анализа, в которых осцилляторные интегралы (и кривизна) играют важную роль.

13. Теоремы о сужении и суммируемость по Бохнеру — Риссу. Термин «теорема о сужении» относится к априорным неравенствам для сужения преобразования Фурье в $\mathbb{R}^n$ на некоторое подмногообразие S, имеющим вид

$$\left(\int_S|\hat{f}(\xi)|^q\,d\sigma(\xi)\right)^{1/q}\leqslant A\|f\|_{L^p} \tag{13.1}$$

и восходящим к (неопубликованному) наблюдению автора, что такие оценки имеют место независимо от того, убывает преобразование Фурье меры $d\sigma$ на бесконечности как $O(|\xi|^{-\varepsilon})$ или же нет. В силу леммы 1 п. 2 такие теоремы о сужении справедливы для всех S конечного типа.

Суммируемость по Бохнеру — Риссу тесно связана с проблемой получения L^p-оценок мультипликаторов Фурье, на $\mathbb{R}^n$ задаваемых формулами

$$\begin{gathered}m(\xi)=(1-|\xi|^2)_+^\lambda.\\ S_R^\lambda(f)\hat{}=m(\xi/R)\hat{f}(\xi).\end{gathered} \tag{13.2}$$

Результаты, полученные в работах Feffermann [28], Carleson — Sjölin [11] и Tomas [113], устанавливают связь между теоремами о сужениях на сферы и поведением мультипликатора (13.2) и дают при $n=2$ точные L^p-оценки для обоих рассматриваемых в этом пункте проблем; при $n\geqslant 3$ полученные результаты носят частичный характер. Самые первые доказательства некоторых таких результатов были заново изложены в работе Hörmander [50] в терминах свойств ограниченности некоторых осцилляторных интегралов как отображений из $L^p(\mathbb{R}^{n-1})$ в $L^q(\mathbb{R}^n)$, где $q=((n+1)/(n-1))p'$.

Рассмотрим операторы

$$(T_\lambda f)(x)=\int_{\mathbb{R}^{n-1}}e^{i\lambda\Phi(x,y)}\psi(x,y)f(y)\,dy,\quad x\in\mathbb{R}^n, \tag{13.3}$$

где $\psi\in C_0^\infty\left(\mathbb{R}^n\times\mathbb{R}^{n-1}\right)$, а Φ удовлетворяет некоторому условию, аналогичному условию необращения в нуль гауссовой кривизны. Тогда справедливы неравенства

$$\|T_\lambda(f)\|_{L^q(\mathbb{R}^n)}\leqslant A\lambda^{-u/q}\|f\|_{L^q(\mathbb{R}^{n-1})}. \tag{13.4}$$

Как показали Карлесон и Шёлин, эти неравенства имеют место при $1 \leqslant p < 4$, в случае когда $n = 2$, а в случае $n \geqslant 3$ при $1 \leqslant p \leqslant 2$ (см. Stein [103]). Для диапазона $2 < p < 2n/(n-1)$ проблема остается открытой.

Другой подход для случая $n = 2$ (с акцентом на геометрии прямоугольников в пространстве фурье-переменных) был предложен Фефферменом [29]; используя этот подход, Кордоба [17] получил соответствующие L^2-оценки для прямоугольных максимальных функций в $\mathbb{R}^2$ (см. также п. 14 ниже). Далее, Карбери [8] применил этот подход при доказательстве точных оценок для ключевой квадратичной функции

$$\left(\int\limits_0^\infty |S_R^\lambda - S_R^{\lambda-1}|^2 \, dR/R\right)^{1/2}$$

в $\mathbb{R}^2$ при $2 \leqslant p \leqslant 4$ и получил максимальную теорему для суммируемости по Бохнеру — Риссу. Ещё раньше эта квадратичная функция была использована, например, в книге Stein — Weiss [106, гл. 7]. Родственные результаты имеются также в работах: Igari [53], Cordoba, López-Melero [20], Cordoba [18], Carbery [9], Christ [13], Seeger [88], Sogge [90].

14. Направленные максимальные функции и, в частности, лакунарные направления. Для произвольного единичного вектора θ в $\mathbb{R}^n$ определим направленную максимальную функцию M_θ формулой

$$M_\theta(f)(x) = \sup_{h>0} \frac{1}{2h} \left| \int\limits_{-h}^{h} f(x - t\theta)\, dt \right|.$$

В некоторых задачах представляют интерес свойства ограниченности оператора M_Γ, где $M_\Gamma(f) = \sup_{\theta\in\Gamma} M_\theta(f)$, а Γ — заданное множество направлений. (Если $\Gamma = S^{n-1}$, $n \geqslant 2$, то L^p-неравенств нельзя получить ни для какого $p < \infty$ ввиду наличия множества Безиковича — Какэя.)

Оператор M_Γ исследован в двух случаях. Во-первых, когда Γ *равномерно распределено*, т. е. $\exists\varepsilon > 0$, такое что если θ, $\theta' \in \Gamma$, $\theta \neq \theta'$, то $|\theta - \theta'| \geqslant \varepsilon$. В этом случае можно считать, что для некоторого $N = N(p, n)$

$$\| M_\Gamma(f) \|_p \leqslant A (\log(1/\varepsilon))^N \| f \|_p$$

при $p \geqslant n$. Для $n = 2$ это было доказано Кордобой [17] при изучении мультипликатора (13.2); см. также Stromberg [111].

Второй случай — это когда множество Γ *лакунарно*, например, в размерности два, когда $\Gamma = \{\theta_k\}$, где направление θ_k образуют с осью абсцисс угол 2^{-k}. Тогда

$$\| M_\Gamma(f) \|_p \leqslant A_p \| f \|_p, \quad 1 < p < \infty. \tag{14.1}$$

Это неравенство можно доказать используя квадратичные функции и свойства убывания преобразования Фурье; см. Nagel — Stein — Wainger [68]. Кроме того, при доказательстве используется неравенство Литтвуда — Пэли

$$\| (\sum | P_k f |^2)^{1/2} \|_p \leqslant A_p \| f \|_p, \quad 1 < p < \infty,$$

где $P_k(f)^\frown(\xi) = \chi_{S_k}(\xi) \hat{f}(\xi)$, а S_k обозначает сектор

$$S_k = \{\xi \,|\, |(\xi, \theta_k)| - |\xi|| \leqslant 2^{-k} |\xi|\},$$

в сочетании с бутстрэп-методом подлиннее: если (14.1) справедливо для p_0, то

$$\left\| \left(\sum | M_{\theta_k} f_k |^2\right)^{1/2} \right\|_p \leqslant A \left\| \left(\sum | f_k |^2\right)^{1/2} \right\|_p$$

при $1/p < [1 + 1/p_0]/2$.

Более ранние результаты для $p \geqslant 2$ были получены в работах Stromberg [110] и Cordoba — R. Fefferman [19] с помощью метода покрытий. Относительно случая оператора M^Γ с равномерно распределённым Γ см. R. Fefferman [33] и Christ — Duoandikoetxea — Rubio de Francia [15]. Будстрэп-метод, аналогичный использованному выше, применён в работе Duoandikoetxea — Rubio de Francia [26].

15. Преобразование Гильберта на плоских выпуклых кривых. Вернёмся к п. 6 и покажем, что даже в случае, когда подмногообразие является «бесконечно плоским» в начале координат, результаты, аналогичные теореме 5, всё ещё могут быть справедливыми. Рассмотрим кривые $\gamma(t) = (t, \gamma_2(t))$ в $\mathbb{R}^2$, где функция $t \to \gamma_2(t)$ выпукла и непрерывна для $0 \leqslant t < \infty$, удовлетворяет условиям $\gamma_2(0) = \gamma_2'(0) = 0$ и нечётна (т. е. $\gamma_2(-t) = -\gamma_2(t)$).

Рассмотрим преобразование Гильберта и максимальную функцию вдоль γ:

$$H(f)(x) = \text{p. v.} \int\limits_{-\infty}^{\infty} f(x - \gamma(t))\, dt/t, \tag{15.1}$$

$$M(f)(x) = \sup_{h>0} \left| \frac{1}{h} \int\limits_0^h f(x - \gamma(t))\, dt \right|. \tag{15.2}$$

Для того чтобы оператор (15.1), был ограничен в $L^p(\mathbb{R}^2)$, необходимо и достаточно, чтобы кривая $\gamma(t)$ удовлетворяла условию «удвоения»: для $h(t)=t\gamma_2'(t)-\gamma_2(t)$ существует константа $D>1$, такая что $h(Dt)\geqslant 2h(t)$, $t\geqslant 0$. Для $p=2$ этот результат получен в работах Nagel — Vance — Wainger — Weinberg [72—74] (равно как и соответствующие результаты для «чётных» кривых, некоторые утверждения для случая большего числа измерений и для оператора (15.2) при $p=2$). Случай произвольного p исследован в работе Carlsson и др. [12], где получены также L^p-неравенства для оператора M. При доказательстве этих L^p-результатов используется бутстрэп-метод, аналогичный рассмотренному выше в п. 14.

16. Преобразование Фурье мер, сосредоточенных на поверхностях — (заключительная виньетка). Завершая наш обзор, возвратимся к проблеме, поставленной в п. 2, и кратко изложим недавно полученный элегантный результат Бруны, Нагела и Уэйнджера [7], который является впечатляющей иллюстрацией взаимодействия многих понятий, рассмотренных нами выше, как-то: убывание преобразования Фурье, кривизна, квазиметрики, ассоциированные с ними формы объёма.

Пусть S обозначает граничную поверхность гладкого выпуклого компактного множества в $\mathbb{R}^n$. Вопрос состоит в том, как убывает на бесконечности преобразование Фурье (2.2) гладкой меры, сосредоточенной на этой поверхности. Предположим, что S является в каждой точке поверхностью конечного типа, т. е., более точно, что для каждой точки $x_0\in S$ любая касательная к S, проходящая через x_0, имеет с S касание лишь конечного порядка.

Для любой точки $x_0\in S$ и любого $\delta>0$ рассмотрим шар

$$B(x_0,\ \delta)=\{x\in S\,|\,(x-x_0,\ \nu_{x_0})<\delta\},$$

где ν_{x_0} — направленная вовнутрь нормаль к S в точке x_0. (Шар $B(x_0,\delta)$ состоит из точек поверхности S, лежащих между гиперплоскостями, определяемыми с помощью $T_{x_0}(S)$ и $T_{x_0}(S)+\delta\nu_{x_0}$).

Можно показать, что это семейство шаров удовлетворяет основным условиям леммы Витали о покрытиях: т. е. величина $\rho(x,y)=\inf\{\delta\,|\,y\in B(x,\delta)\}$ есть квазиметрика на S, а форма объёма $V_x(\delta)=\operatorname{vol}B(x,\delta)$ обладает свойством удвоения.

Далее, для $\xi\neq 0$ определим $x^+(\xi)$ и $x^-(\xi)$ и $x^-(\xi)$ как точки на S, для которых

$$\nu_{x^+}=\xi/|\xi|,\qquad \nu_{x^-}=-\xi/|\xi|.$$

Тогда требуемая оценка имеет вид

$$|\hat{\mu}(\xi)|\leqslant c\,(V_{x^+}(\delta)+V_{x^-}(\delta)),\quad \text{где}\quad \delta=1/|\xi|. \tag{16.1}$$

ЛИТЕРАТУРА

1. M. Beals, C. Fefferman, and R. Grossman, Bull. Amer. Math. Soc. (N. S.) **8** (1983), 125—322.

2. J. E. Björck, On Fourier transforms of smooth measured carried by real-analytic sub-manifolds of **R**^{n}, Preprint, 1973.

3. J. Bourgain, On high dimensional maximal functions associated to convex bodies, Preprint, 1985.

4. —, On the spherical maximal function in the plane, Preprint, 1985.

5. —, Averages in the plane over convex curves and maximal operators, Preprint, 1986.

6. —, On the L^p bounds for maximal functions associated to convex bodies in **R**^{n}, Preprint, 1986.

7. J. Bruna, A. Nagel, and S. Wainger, personal communication, 1986.

8. A. Carbery, Duke Math. J. **50** (1983), 409—416.

9. —, Recent progress in Fourier analysis, Proceedings of El Escorial Seminar, 1983, North-Holland Math. Stud. No. 111, North-Holland, Amsterdam — New York, 1985, pp. 49—56.

10. —, Bull. Amer. Math. Soc. (N. S.) **14** (1986), 269—273.

11. L. Carleson and P. Sjölin, Studia Math. **44** (1972), 287—299.

12. H. Carlsson et al., Bull. Amer. Math. Soc. (N. S.) **14** (1986), 263—267.

13. M. Christ, Proc. Amer. Math. Soc. **95** (1985), 16—20.

14. —, Ann. of Math. **122** (1985), 575—596.

15. M. Christ, J. Duoandikoetxea, and J. L. Rubio de Francia, Duke Math. J. **53** (1986), 189—209.

16. R. Coifman, A. MacIntosh, and Y. Meyer, Ann. of Math. **116** (1982), 361—387.

17. A. Cordoba, Amer. J. Math. **99** (1977), 1—22.

18. —, Ann. Inst. Fourier (Grenoble) **32** (1982), 215—226.

19. A. Cordoba and R. Fefferman, Proc. Nat. Acad. Sci. U. S. A. **74** (1977), 2211—2213.

20. A. Cordoba and B. López-Melero, Ann. Inst. Fourier (Grenoble) **31** (1981), 147—152.

21. M. Cowling and G. Mauceri, personal communication, 1982.

22. —, Trans. Amer. Math. Soc. **287** (1985), 431—455.

23. G. David, Proc. Internat. Congr. Math. (Berkeley, California, U. S. A., 1986), 1987.

24. G. David and J. L. Journé, Ann. of Math. **120** (1984), 371—397.

25. J. Duoandikoetxea and J. L. Rubio de Francia, personal communication, 1984.

26. —, Invent. Math. (1986).

27. K. J. Falconer, Math. Proc. Cambridge Philos Soc. **87** (1980), 221—226.

28. C. Fefferman, Acta Math. **124** (1970), 9—36.

29. —, Israel J. Math. **15** (1973), 44—52.

30. —, Bull. Amer. Math. Soc. (N. S.) **9** (1983), 129—206.

31. C. Fefferman and D. H. Phong, Conference on Harmonic Analysis in Honor of Antoni Zygmund, 1981, Vol. 2, Wadsworth, Calif., 1983, pp. 590—606.

32. C. Fefferman and A. Sanchez-Calle, 1986 (to appear).

33. R. Fefferman, Proc. Nat. Acad. Sci. U. S. A. **80** (1983), 3877—3878.

34. G. Folland and E. M. Stein, Comm. Pure Appl. Math. **27** (1974), 429—522.

35. D. Geller and E. M. Stein, Bull. Amer. Math. Soc. (N. S). **6** (1982), 99—103; Math. Ann. **267** (1984), 1—15.

36. R. Goodman, Nilpotent Lie groups, Lecture Notes in Math., Vol. 562, Springer-Verlag, Berlin and New York, 1976.

37. A. Greenleaf, Indiana Math. J. **30** (1981), 519—537.
38. —, personal communication, 1983.
39. —, Singular integral operators with conical singularities, Preprint, 1985.
40. A. Grossman, G. Loupias, and E. M. Stein, Ann. Inst. Fourier (Grenoble) **18** (1969), 343—368.
41. V. Gullemin and S. Sternberg, Geometric asymptotics, Math. Surveys, No. 14, Amer. Math. Soc., Providence, R. I., 1977. [Имеется перевод: Гийемин В., Стернберг С. Геометрические асимптотики. — М.: Мир, 1973.]
42. V. Guillemin and G. Uhlmann, Duke Math. J. **48** (1981), 251—267,
43. R. Harvey and J. Polking, Proc. Sympos. Pure Math., Vol. 41, Amer. Math. Soc., Providence, R. I., 1985, pp. 117—136.
44. B. Helffer and J. Nourigatt, Hypoellipticité maximale pour des opérateurs polynoms de champs de vecteurs, Birkhäuser, Basel, 1985.
45. Хенкин Г. М. — Мат. сб., 1969, т. 7, с. 597—716.
46. C. S. Herz, Ann. of Math. **75** (1962), 81—92.
47. E. Hlawka, Monatsh. Math. **54** (1950), 1—36.
48. L. Hörmander, Acta Math. **119** (1967), 147—171.
49. —, Acta Math. **127** (1971), 79—183.
50. —, Ark. Mat. **11** (1973), 1—11.
51. —, The analysis of linear partial differential operators. I, Springer-Verlag, Berlin — New York, 1983. [Имеется перевод: Хёфмандер Л. Анализ линейных дифференциальных операторов с частными производными. Т. 1. — М.: Мир, 1986.]
52. R. Howe, J. Funct. Anal. **38** (1980), 188—254.
53. S. Igari, Tôhoku Math. J. **33** (1981), 413—419.
54. A. Knapp and E. M. Stein, Ann. of Math. **93** (1971), 489—578.
55. A. Koranyi and I. Vagi, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **25** (1971), 575—648.
56. S. Krantz, Beijing lectures on harmonic analysis, Ann. of Math. Stud. (to appear).
57. I. Lieb, Math. Ann. **190** (1970), 6—44.
58. I. Lieb and M. Range, Math. Ann. **265** (1983), 221—251.
59. J. M. Marstrand, Mathematika **26** (1979), 180—183.
60. G. Mauceri, M. Picardello, and F. Ricci, Supp. Rend. Circ. Mat. Palermo **1** (1981), 191—202.
61. R. Melrose and G. Uhlmann, Comm. Pure Appl. Math. **32** (1979), 483—519.
62. P. A. Meyer, Demonstrations probabilste de certaines inegalités de Littlewood-Paley, Lecture Notes in Math., vol. 511, Springer-Verlag, Berlin — New York, 1976, pp. 125—183; Retour sur la théorie de Littlewood-Paley, Lecture Notes in Math., vol. 850, Springer-Verlag, Berlin — New York, 1980, pp. 151—166.
63. D. Müller, Invent. Math. **73** (1983), 467—489.
64. —, J. Reine Angew. Math. **356** (1985), 90—118.
65. A. Nagel, Beijing lectures on harmonic analysis, Ann. of Math. Stud. (to appear).
66. A. Nagel, N. M. Riviere, and S. Wainger, Amer. J. Math. **98** (1976), 395—403.
67. —, Proc. Nat. Acad. Sci. U. S. A. **73** (1976), 1416—1417.
68. A. Nagel, E. M. Stein, and S. Wainger, Proc. Nat. Acad. Sci U. S. A. **75** (1978), 1060—1062.
69. —, Hilbert transforms and maximal functions related to variable curves, Proc. Sympos. Pure Math., Vol. 35, Part 1, Amer. Math. Soc., Providence, R. I., 1979, pp. 95—98.

70. —, Proc. Nat. Acad. Sci. U S. A. **78** (1981), 6596—6599.
71. —, Acta Math. **155** (1985), 103—147.
72. A. Nagel, J. Vance, S. Wainger, and D. Weinberg, Duke Math. J. **50** (1983), 735—744.
73. —, Duke Math. J. **52** (1985), 715—722.
74. —, Amer. J. Math. (to appear).
75. D. M. Oberlin and E. M. Stein, Indiana Univ. Math. J. **31** (1982), 641—650.
76. D. H. Phong, Proc. Nat. Acad. Sci. U. S. A. **76** (1979), 1554—1558.
77. D. H. Phong and E. M. Stein, Proc. Nat. Acad. Sci. U. S. A. **80** (1983), 7697—7701.
78. —, Hilbert integrals, singular integrals, and Radon transforms. Part I, Acta Math. **157** (1986), 99—157.
79. —, ibid., Part II, Invent Math. **86** (1986), 75—113.
80. B. Randol, Trans. Amer. Math. Soc. **139** (1969), 279—285.
81. F. Ricci and E. M. Stein, Proc. Nat. Acad. Sci. U. S. A. **83** (1986), 1—3.
82. —, Harmonic analysis on nilpotent-groups and singular integrals. Part I, J. Funct. Anal. (to appear).
83. —, *ibid.*, Part II, Preprint, 1986.
84. L. P. Rothschild and E. M. Stein, Acta Math. **137** (1976), 247—320.
85. J. L. Rubio de Francia, Maximal functions and Fourier transforms, Preprint, 1986.
86. A. Ruiz, Trans. Amer. Math. Soc. **287** (1985), 167—188.
87. A. Sanchez-Calle, Invent. Math. **78** (1984), 143—160.
88. A. Seeger. Über Fouriermultiplikatoren und die ihnen Zugeordneten Maximalfunktionen, Thesis, Darmstadt, 1985.
89. I. Segal, Math. Scand. **13** (1963), 31—43.
90. C. Sogge, Duke Math. J. **53** (1986), 43—65.
91. —, On almost everywhere convergence to L^p data for higher order hyperbolic operators, Preprint, 1986.
92. C. Sogge and E. M. Stein, Invent. Math. **82** (1985), 543—556.
93. —, Averages over hypersurfaces. II, Invent. Math. **86** (1986), 133—242.
94. N. K. Stanton, Invent. Math. **65** (1981), 137—174.
95. E. M. Stein, Proc. Nat. Acad. Sci. U. S. A. **47** (1961), 1894—1897.
96. —, Topics in harmonic analysis related to the Littlewood-Paley theory, Princeton Univ. Press, Princeton, N. J., 1970.
97. —, Singular integrals and differentiability properties of functions, Princeton Univ. Press, Princeton, N. J., 1970. [Имеется перевод: Стейн И. Сингулярные интегралы и дифференциальные свойства функций. — М.: Мир, 1973.]
98. —, Some problems in harmonic analysis suggested by symmetric spaces and semi-simple groups, Actes du Congrès International des Mathématiciens (Nice, 1970), Vol. 1, Gauthier-Villars, Paris, 1971, pp. 173—189.
99. —, Proc. Nat. Acad. Sci. U. S. A. **73** (1976), 2174—2175.
100. —, Bull. Amer. Math. Soc. (N. S.) **7** (1982), 359—376.
101. —, Bull. Amer. Math. Soc. (N. S.) **9** (1983), 71—73.
102. —, Recent progress in Fourier analysis, Proceedings of El Escorial Seminar, 1983, North-Holland Math. Stud. No. 111, North-Holland, Amsterdam — New York, 1985, pp. 229—244.
103. —, Beijing lectures on harmonic analysis, Ann. of Math. Stud. (to appear).
104. E. M. Stein and J. O. Stromberg, Ark. Mat. **21** (1983), 259—269.
105. E. M. Stein and S. Wainger, Bull. Amer. Math. Soc. **84** (1978), 1239—1295.

106. E. M. Stein and G. Weiss, Introduction to Fourier analysis on Euclidean spaces, Princeton Univ. Press, Princeton, N. J., 1971. [Имеется перевод: Стейн И., Вейс Г. Введение в гармонический анализ на евклидовых пространствах. — М.: Мир, 1974.]
107. R. S. Strichartz, Duke Math. J. **44** (1977), 705—713.
108. —, Duke Math. J. **48** (1981), 699—727.
109. —, Studia Math. **74** (1982), 137—151.
110. J. O. Stromberg, Ark. Mat. **15** (1977), 229—240.
111. —, Ann. of Math. **107** (1978), 399—402.
112. I. Svenson, Ark. Mat. **9** (1971), 11—22.
113. P. Tomas, Proc. Sympos. Pure Math., Vol. 35, Part 1, Amer. Math. Soc., Providence, R. I., 1979, pp. 111—114.
114. G. Uhlmann, personal communication, 1986.
115. A. Zygmund, Trigonometric series, Cambridge Univ. Press, London and New York, 1959. [Имеется перевод: Зигмунд А. Тригонометрические ряды. — В двух томах. — М.: Мир, 1965.]

ПРИНСТОНСКИЙ УНИВЕРСИТЕТ, ПРИНСТОН, НЬЮ-ДЖЕРСИ 08544, США

АЛГЕБРАИЧЕСКАЯ *К*-ТЕОРИЯ ПОЛЕЙ

А. А. Суслин [1]

Построение высшей алгебраической K-теории было завершено фундаментальной работой Куиллена [24]. После этого основные усилия были сконцентрированы в области вычислений и применений K-теории к конкретным алгебраическим проблемам. Наиболее интригующими являются гипотезы, связывающие алгебраическую K-теорию с этальными когомологиями. Такие гипотезы в некоторых специальных случаях были выдвинуты Куилленом и Лихтенбаумом [14, 9, 23]. В настоящее время все гипотезы такого типа обычно называются гипотезами Куиллена — Лихтенбаума. Одним из наиболее важных свойств алгебраической K-теории является наличие точной последовательности локализации (если $Y \subset X$ — замкнутая подсхема, то имеется длинная точная последовательность

$$\ldots \to K_i'(Y) \to K_i'(X) \to K_i'(X-Y) \xrightarrow{\partial} K_{i-1}'(Y) \to \ldots$$

($K' = K$ для регулярных схем) и соответствующей спектральной последовательности

$$E_1^{pq} = \coprod_{\operatorname{codim} x = p} K_{-p-q}(k(x)) \Rightarrow K'_{-p-q}(X).$$

Другим важным свойством является гипотеза Герстена, доказанная Куилленом, которая позволяет идентифицировать второй член этой спектральной последовательности: $E_2^{pq} = H^p(X, K_{-q})$. Эти свойства часто позволяют сводить общие проблемы алгебраической K-теории к частному случаю полей, в котором эти проблемы особенно ясно формулируются и выглядят крайне интригующе.

Высшая K-теория поля F (так же как и K-теория любого кольца) может быть определена в терминах плюс-конструкции

---

[1] A. A. Suslin, Algebraic K-theory of fields, ICM86, pp. 222—244. [Текст доклада был зачитан на конгрессе Э. Фридландером. — *Изд. ред.*]

Куиллена: $K_i(F)=\pi_i(\mathrm{BGL}(F)^+)$, где $\mathrm{BGL}(F)^+$ есть H-пространство, имеющее те же гомологии, что и $\mathrm{BGL}(F)$, т. е. совпадающие с гомологиями дискретной группы $\mathrm{GL}(F)$. Таким образом, K-теория оказывается тесно связанной с теорией гомологий группы $\mathrm{GL}(F)$.

Этот доклад касается некоторых из недавних достижений в K-теории полей и близких областях математики. К сожалению, я лишь кратко коснулся такой важной темы, как этальная K-теория Дуайера — Фридландера; идеи и методы, используемые в этой теории, очень далеки от того, что обсуждается в основной части этой работы.

1. ГОМОМОРФИЗМ НОРМЕННОГО ВЫЧЕТА

Ввиду теоремы Мура — Мацумото группа $K_2(F)$ может быть описана как группа с образующими $\{a, b\}$ $(a,\ b\in F^*)$ и соотношениями $\{a_1a_2, b\}=\{a_1, b\}+\{a_2, b\}$, $\{a, b_1b_2\}=\{a, b_1\}+\{a, b_2\}$, $\{a, 1-a\}=0$ $(a\neq 1)$. Предположим, что n — натуральное число, не делящееся на $\operatorname{char} F$; тогда теория Куммера определяет изоморфизм χ: $F^*/F^{*n}\xrightarrow{\sim}H^1(F,\ \mu_n)$. Несложно проверить, что $\chi(a)\cup\chi(1-a)=0\in H^2(F,\ \mu_n^{\otimes 2})$, и, следовательно, мы получаем канонический гомоморфизм

$$R_n=R_{n,\,F}\colon\ K_2(F)/n\to H^2(F,\ \mu_n^{\otimes 2})\colon\ \{a,\ b\}\longmapsto\chi(a)\cup\chi(b),$$

который называется гомоморфизмом норменного вычета. Если $F\supset\mu_n$, то выбор примитивного корня n-й степени из единицы ξ позволяет отождествить G_F-модули μ_n и $\mu_n^{\otimes 2}$ и, следовательно, отождествить $H^2(F,\ \mu_n^{\otimes 2})$ с $H^2(F,\ \mu_n)$. После этого отождествления R_n превращается в гомоморфизм, заданный посредством циклических алгебр: $\{a,\ b\}\longmapsto[A_\xi(a,\ b)]$ (см. [19]). Так что в этом случае вопрос о сюръективности R_n равносилен классической проблеме Альберта: всякая ли алгебра экспоненты n подобна тензорному произведению циклических алгебр?

Теорема 1.1 [17, 30]. *Для любого поля F и любого n, не делящегося на $\operatorname{char} F$, гомоморфизм R_n: $K_2(F)/n\to H^2(F,\ \mu_n^{\otimes 2})$ является изоморфизмом.*

Общий случай теоремы легко сводится к случаю (который мы и будем рассматривать в дальнейшем), когда $n=p$ есть простое число и $F\supset\mu_p$. Имеется два различных, но тесно связанных между собой подхода к доказательству теоремы 1.1. Оба подхода существенно используют вычисление некоторых групп K-когомологий многообразий Севери — Брауэра.

Первый метод, исходный метод Меркурьева [16], до сих пор использовался только при $p=2$. Положим временно $k_2=K_2/2$. Предположим, что $E=F(\sqrt{a})$ — квадратичное расширение поля F, и обозначим через $\chi(a)\in H^1(F,\mu_2)$ класс когомологий, соответствующий a при изоморфизме Куммера. Точная последовательность когомологий

$$H^1(F,\ \mu_2)\xrightarrow{\chi(a)}H^2(F,\ \mu_2)\to H^2(E,\ \mu_2)\xrightarrow{N_{E/F}}H^2(F,\ \mu_2)$$

показывает, что справедливость теоремы 1.1 влечёт точность последовательности

$$F^*/F^{*2}\xrightarrow{a}k_2(F)\to k_2(E)\xrightarrow{N_{E/F}}k_2(F). \tag{1.1.1}$$

Наоборот, если последовательность (1.1.1) точна для любого квадратичного расширения, то несложное индуктивное рассуждение доказывает теорему 1.1. Более того, как показано в [16], достаточно уметь доказывать точность лишь последовательности

$$k_2(F)\to k_2(E)\xrightarrow{N_{E/F}}k_2(F),$$

чтобы завершить доказательство теоремы 1.1. Любой элемент из $k_2(E)$ может быть записан в виде $\sum\limits_{i=1}^{n}\{x_i+\sqrt{a}y_i,\ z_i\}$, где $x_i,\ y_i,\ z_i\in F$. Норма этого элемента в $k_2(F)$ равна $\sum\limits_{i=1}^{n}\{x_i^2-ay_i^2,\ z_i\}$. Несложно выписать в явном виде условия, необходимые и достаточные для обращения в нуль последнего элемента: для этого необходимо и достаточно, чтобы (после некоторых косметических изменений, включающих возможное увеличение n) существовали (для каждого непустого $S\subset\{1,\ \dots,\ n\}$) элементы u_S, v_S, такие что, полагая $z_S=\prod\limits_{i\in S}z_i$, будем иметь следующие соотношения:

$$x_i^2-y_i^2a=\prod_{S\ni i}(u_S^2-z_Sv_S^2).$$

Обозначим через F_0 простое подполе в F и положим $F_1=F_0(a)$. Уравнения (1.1.2) определяют аффинное многообразие T над F_1, а элементы x_i, y_i, z_i, u_S, v_S определяют F-значную точку этого многообразия. Обозначая соответствующие координатные функции через X_i, Y_i, Z_i, U_S, $V_S\in F_1(T)$, мы получаем в $k_2(F_1(T))$ «универсальный» элемент с нулевой нормой

$\sum_{i=1}^{n} \{X_i + \sqrt{a}Y_i,\ Z_i\}$. Достаточно доказать, что этот универсальный элемент приходит из $k_2(F_1(T))$,— специализация завершит доказательство. Чтобы доказать последнее утверждение, достаточно показать, что R_2 является изоморфизмом для полей $F_1(T)$ и $F_1(T)(\sqrt{a})$. Это тривиально для второго поля, поскольку это поле чисто трансцендентно над $F_1(\sqrt{a})$ (как ядро, так и коядро R_n не меняются при чисто трансцендентных расширениях [4]). Поле $F_2 = F_1(Z_i, U_S, V_S)$ чисто трансцендентно над F_1, а $F_1(T)$ получается из F_2 при помощи нескольких переходов к полю функций на конике, заданной уравнением вида $X^2 - Y^2a = *$. Таким образом, теорема из следующего предложения:

Предложение 1.2 [27]. *Предположим, что* $\operatorname{char} k \neq 2$, a, $b \in k^*$, *и обозначим через* F *поле функций на конике, заданной уравнением* $X^2 - aY^2 = b$. *Если гомоморфизмы* $R_{2,k}$ *и* $R_{2,k(\sqrt{a})}$ *являются изоморфизмами, то изоморфизмом будет и* $R_{2,F}$.

Доказательство основано на вычислении некоторых групп K-когомологий коники, оно также интенсивно использует теорию квадратичных форм, что затрудняет использование этого метода при $p \neq 2$.

Второй подход к доказательству теоремы 1.1, развитый в [17, 28, 30], является в некотором смысле противоположным подходу, обсуждавшемуся выше. Основной технический результат при этом подходе таков:

Предложение 1.3. *Предположим, что* $\operatorname{char} F \neq p$ *и* F *содержит первообразный корень* p*-й степени из единицы* ξ. *Пусть* a, $b \in F^*$ *и через* X *обозначено многообразие Севери — Брауэра, соответствующее циклической алгебре* $D = A_\xi(a, b)$. *Естественные гомоморфизмы* $\ker R_{p,F} \to \ker R_{p,F(X)}$ *и* $\operatorname{coker} R_{p,F} \to \operatorname{coker} R_{p,F(X)}$ *инъективны.*

Используя теорему 1.3, доказательство теоремы 1.1 можно завершить следующим образом. Хорошо известно, что гомоморфизмы $\ker R_{p,F} \to \ker R_{p,E}$, $\operatorname{coker} R_{p,F} \to \operatorname{coker} R_{p,E}$ инъективны, если E является алгебраическим расширением поля F степени, взаимно простой с p. Таким образом, применяя предложение 1.3, можно построить расширение $\tilde{F}/F$, такое что

(a) все циклические p-алгебры над $\tilde{F}$ тривиальны;

(b) $\tilde{F}$ не имеет расширений степени, взаимно простой с p;

(c) $\ker R_{p,F} \hookrightarrow \ker R_{p,\tilde{F}}$, $\operatorname{coker} R_{p,F} \hookrightarrow \operatorname{coker} R_{p,F}$.

Классический результат Милнора [19] показывает, что (a) равносильно равенству $K_2(\tilde{F})/p = 0$. Таким образом, $\ker R_{p,\tilde{F}} = 0$. Кроме того, легко видеть, что свойства (a) и (b) влекут обращение в нуль группы Брауэра поля $\tilde{F}$: $\operatorname{Br}(\tilde{F}) = 0$ и, следовательно, $\operatorname{coker} R_{p,\tilde{F}} = 0$. Теперь свойство (c) показывает, что $\ker R_{p,F} = \operatorname{coker} R_{p,F} = 0$.

Доказательство предложения 1.3, так же как и доказательство предложения 1.2, основано на вычислении групп K-когомологий многообразий Севери — Брауэра.

Предложение 1.4. *В условиях предложения* 1.3 *имеем* $H^1(X,\ K_2) = N = \operatorname{Nrd} D^* \subset F^*$; *естественный гомоморфизм* $K_2(F) \to H^0(X, K_2)$ *сюръективен.*

Для доказательства надо рассмотреть спектральную последовательность $E_2^{ij} = H^i(X, K_{-j}) \Rightarrow K_{-i-j}(X)$. Теория классов Чжэня и теорема Римана — Роха позволяют показать, что все дифференциалы этой спектральной последовательности, начинающиеся или кончающиеся в членах E^{ij} этой последовательности с $i + j = 0,\ -1$, аннулируются числом $(\dim X)!$. В нашем случае $\dim X = p - 1$, и мы знаем также, что все дифференциалы аннулируются числом p (поскольку D обладает полями расщепления степени p над F). Это показывает, что все дифференциалы, входящие или выходящие из членов E^{ij} с $i + j = 0,\ -1$, равны нулю и, следовательно, $H^1(X, K_2) = E_2^{1,-2} = E_\infty^{1,-2} = K_1(X)^{1/2}$. K-теория многообразий Севери — Брауэра вычислена Куилленом [24]:

$$K_i(X) = K_i(F) \oplus K_i(D) \oplus \ldots \oplus K_i(D^{\otimes(p-1)}).$$

Таким образом, для завершения доказательства первого утверждения достаточно вычислить топологическую фильтрацию на $K_1(X) = F^* \oplus N \oplus \ldots \oplus N$, что делается без труда. Тривиальность всех дифференциалов, начинающихся в $E^{0,-2}$, означает, что краевой гомоморфизм

$$K^2(F) \oplus K_2(D) \oplus \ldots \oplus K_2(D^{\otimes(p-1)}) = K_2(X) \to H^0(X,\ K_2) = E_2^{0,-2}$$

сюръективен. Для завершения доказательства второго утверждения мы должны теперь показать, что образ $K_2(D^{\otimes(i)})$ в $H^0(X, K_2) \subset K_2(F(X))$ содержится в образе $K_2(F)$. Это требует дополнительной информации о группе K_2 для алгебр простого индекса (см. теорему 3.1 ниже).

Предложение 1.4 всё же недостаточно для доказательства предложения 1.3; необходима более точная информация, что $K_2(F) = H^0(X, K_2)$. Доказательство этого равенства требует информации о кручении в K_2. Основным результатом в этом направлении служит теорема Гильберта 90 для K_2.

Теорема 1.5. *Пусть E/F — циклическое расширение простой степени p, и пусть σ — образующая группы Галуа $\mathrm{Gal}(E/F)$. Следующая последовательность точна*:

$$K_2(E) \xrightarrow{1-\sigma} K_2(E) \xrightarrow{N_{E/F}} K_2(F). \qquad (1.5.1)$$

Точность последовательности (1.5.1) легко проверяется при условии, что гомоморфизм нормы $N\colon E^* \to F^*$ сюръективен — в этом случае удаётся в явном виде построить гомоморфизм $K_2(F) \to K_2(E)/(1-\sigma)K_2(E)$, обратный к $N_{E/F}$, при помощи формулы $\{a, b\} \mapsto \{\alpha, b\} \bmod (1-\sigma)K_2(E)$, где $N(\alpha) = a$. Используя тот же трюк, что и выше, мы убеждаемся, что для рассмотрения общего случая достаточно уметь доказывать, что если X есть многообразие Севери — Брауэра, отвечающее циклической алгебре $(E/F, \sigma, a)$, то отображение

$$\ker N_{E/F}/(1-\sigma)K_2(E) \to \ker N_{E(X)/F(X)}/(1-\sigma)K_2(E(X))$$

инъективно. Эта проблема упрощается за счёт того, что рассматриваемая алгебра распадается над E и, следовательно, $X_E \cong \mathbb{P}_E^{p-1}$. Благодаря этому доказательство требует лишь вычисления группы $H^1(X, K_2)$, проведённого выше.

Применяя теорему 1.5 к универсальному расширению Куммера $F(\sqrt[p]{T})/F(T)$ (T трансцендентно над F) или к универсальному расширению Артина — Шрайера, мы получаем следующий результат, высказанный в качестве гипотезы Тэйтом [37]:

Следствие 1.6. *Если F содержит примитивный корень n-й степени из единицы ξ, то ${}_nK_2(F) = \{\xi, F^*\}$; $K_2\{F\}$ не имеет p-кручения при $p = \mathrm{char}\, F$.*

Чтобы завершить описание кручения в K_2, необходимо дать описание тех элементов $x \in F^*$, для которых $\{\xi, x\} = 0$. Это делается в следующей теореме:

Теорема 1.7 [28, 30]. *Предположим, что F содержит примитивный корень n-й степени из единицы ξ, и обозначим через F_0 подполе констант в F (т. е. алгебраическое замыкание простого подполя). Для $x \in F^*$ следующие условия равносильны*:

(a) $\{\xi, x\} = 0 \in K_2(F)$;

(b) $x = x_0 y^n$, *где* $y \in F^*$, $x_0 \in F_0^*$ *и* $\{\xi, x_0\} = 0 \in K_2(F_0)$.

Следствие 1.8. *Если E/F — такое расширение, что F алгебраически замкнуто в E, то $K_2(F) \hookrightarrow K_2(E)$.*

Последнее следствие показывает, что в условиях предложения 1.4 $K_2(F) \xrightarrow{\sim} H^0(X, K_2)$. Используя этот факт и предложение 1.4, легко завершаем доказательство предложения 1.3.

Доказательство теоремы 1.7 основано на изучении некоторых групп l-адических когомологий. Решающее значение имеет следующий результат, связанный с теоремой Вейля о собственных значениях оператора Фробениуса, действующего на модуле Тэйта абелевого многообразия.

Предложение 1.9. *Пусть поле F конечно-порождено и $l \neq \mathrm{char}\, F$. Тогда $H^1(F_0, \mathbb{Z}_l(2)) \xrightarrow{\sim} H^1(F, \mathbb{Z}_l(2))$ и $H^2(F_0, \mathbb{Z}_l(2)) \hookrightarrow H^2(F, \mathbb{Z}_l(2))$.*

Приведённые выше результаты имеют много важных приложений в алгебре и алгебраической геометрии, некоторые из которых можно найти в [30, 39, 40, 41]. Мы отметим для дальнейшего использования только следующий факт:

Предложение 1.10 [28, 30]. *Пусть X/F — полное рациональное многообразие. Тогда $H^0(X, K_2) = K_2(F)$.*

2. АЛГЕБРАИЧЕСКИ ЗАМКНУТЫЕ И ЛОКАЛЬНЫЕ ПОЛЯ

Поскольку этальные когомологии алгебраически замкнутого поля тривиальны, естественно ожидать, что K-группы подобного поля также устроены достаточно просто. Вот одна из гипотез Куиллена — Лихтенбаума (см. [9, 23]).

(2.1) Если поле F алгебраически замкнуто, то группы $K_i(F)$ делимы при $i \geqslant 1$; при этом подгруппа кручения в $K_i(F)$ тривиальна, если i чётно, и изоморфна $\coprod_{l \neq \mathrm{char}\, F} \mathbb{Q}_l/\mathbb{Z}_l(n)$, если $i = 2n-1$.

Эта гипотеза очевидно верна при $i = 1$ и легко доказывается при $i = 2$ [2]. Помимо этих тривиальных случаев справедливость гипотезы была известна в случае, когда F является алгебраическим замыканием конечного поля [22]. Для полей положительной характеристики эта гипотеза доказана в [31]. Основной результат работы [31] таков:

Теорема 2.2. *Если F/F_0 есть расширение алгебраически замкнутых полей, то для любого n индуцированные гомоморфизмы*

$$K_i(F_0)/n \to K_i(F)/n, \quad {}_nK_i(F_0) \to {}_nK_i(F), \quad K_i(F_0, \mathbb{Z}/n) \to K_i(F, \mathbb{Z}/n)$$

биективны.

Эта теорема является частным случаем некоторого простого общего принципа. Пусть V — контравариантный функтор из подходящей категории схем в категорию периодических абелевых групп. Предположим, далее, что для любого конечного плоского морфизма $X\to Y$ задан гомоморфизм переноса $N_{X/Y}$: $V(X)\to V(Y)$, обладающий обычными свойствами. Предположим наконец, что V гомотопически инвариантен, т. е. $V(X\times\mathbb{A}^1)=V(X)$ для любого X.

Предложение 2.3 (теорема о жёсткости). *Пусть* X/F — *связное многообразие над алгебраически замкнутым полем. Для любых двух точек* x, y: $\mathrm{Spec}\,F\to X$ *индуцированные гомоморфизмы* $V(X)\rightrightarrows V(\mathrm{Spec}\,F)=V(F)$ *совпадают.*

Очевидно, достаточно рассмотреть случай гладкой аффинной кривой. Рассмотрим билинейное отображение $\mathrm{Div}(X)\times V(X)\to V(F)$, заданное формулой $x\times u\mapsto x^*(u)$. Мы должны доказать, что ограничение этого отображения на $\mathrm{Div}^0(X)\times V(X)$ тривиально. Обозначим через $\bar{X}$ гладкую проективную модель X и положим $X_\infty=\bar{X}-X$. Если f есть рациональная функция на $\bar{X}$, определённая и равная единице на X_∞, то главный дивизор (f) лежит в ядре рассматриваемого спаривания: f определяет накрытие $X_0\to\mathbb{A}^1_F=\mathbb{P}^1_F-1$, где X_0 получается из X выбрасыванием точек, в которых f равна единице. Обычные свойства переноса показывают, что образ $(f)\times u$ в $V(F)$ совпадает с образом $(0-\infty)\times N_{X_0}/\mathbb{A}^1_F\left(u\,|_{X_0}\right)$, который равен нулю в силу гомотопической инвариантности. Таким образом, наше спаривание пропускается через $\mathrm{Pic}^0(\bar{X},X_\infty)\otimes V(X)$. Группа $\mathrm{Pic}(\bar{X},X_\infty)$ совпадает с группой точек якобиана Розенлихта многообразия $\bar{X}$ и, следовательно, делима. Поскольку группа $V(X)$ периодична, мы заключаем, что $\mathrm{Pic}^0(\bar{X},X_\infty)\otimes V(X)=0$.

Следствие 2.3.1. *Пусть* F/F_0 — *расширение алгебраически замкнутых полей, и пусть* X_0/F_0 — *связное многообразие. Если* x, y: $\mathrm{Spec}\,F\to X_0$ — *любые* F_0-*точки, то индуцированные гомоморфизмы* $V(X_0)\rightrightarrows V(F)$ *совпадают.*

Следствие 2.3.2. *В условиях следствия* 2.3.1 *для любой* F_0-*точки* x: $\mathrm{Spec}\,F\to X_0$ *образ соответствующего гомоморфизма* $V(X_0)\to V(F)$ *содержится в образе* $V(F_0)$.

Выберем рациональную точку $\mathrm{Spec}\,F_0\to X_0$ и применим следствие 2.3.1 к x и y: $\mathrm{Spec}\,F\to\mathrm{Spec}\,F_0\to X_0$.

Следствие 2.3.3. *Предположим дополнительно, что* V *коммутирует с прямыми пределами*:

$$V(\mathrm{Spec}\varinjlim A_i)=\varinjlim V(\mathrm{Spec}\,A_i).$$

Тогда $V(F)=V(F_0)$ *для любого расширения алгебраически замкнутых полей* F/F_0.

F может быть записано в виде $\varinjlim A$, где A пробегает все конечно-порождённые F_0-подалгебры F. Наши условия гарантируют, что $V(F)=\varinjlim V(A)$. Гомоморфизм $V(\mathrm{Spec}\,A)\to V(F)$ индуцирован F_0-точкой $\mathrm{Spec}\,F\to\mathrm{Spec}\,A$, отвечающей вложению $A\hookrightarrow F$. Ввиду (2.3.2), образ этого гомоморфизма содержится в образе $V(F_0)$. Поскольку это справедливо для всех A, мы заключаем, что гомоморфизм $V(F_0)\to V(F)$ сюръективен. Инъективность этого гомоморфизма тривиальна.

Приведённое доказательство предложения 2.3, которое является незначительной модификацией первоначального доказательства автора [31], принадлежит Габберу, Жилле и Томасону. Использование относительных групп Пикара вместо абсолютных позволило этим авторам доказать следующее важное обобщение предложения 2.3.

Предложение 2.4. *Пусть* O — *гензелево кольцо с полем частных* F *и полем вычетов* k, *и пусть* $X/\mathrm{Spec}\,O$ — *гладкая аффинная кривая. Пусть, далее,* x, y: $\mathrm{Spec}\,O\to X$ — *два сечения, совпадающие в замкнутой точке* $\mathrm{Spec}\,O$. *Предположим дополнительно, что*

(a) $n\cdot V(X)=0$, *где* $(n,\mathrm{char}\,k)=1$;

(b) $V(O)\hookrightarrow V(F)$.

Тогда индуцированные гомоморфизмы x^*, y^*: $V(X)\to V(O)$ *совпадают.*

Выберем проективное замыкание $\bar{X}$ кривой X и положим $X_\infty=\bar{X}-X$. Сечения x, y определяют относительные дивизоры D_x, D_y на $\bar{X}$ (относительно X_∞). Их разность делится на n в $\mathrm{Pic}(\bar{X},X_\infty)$, поскольку $\mathrm{Pic}(\bar{X},X_\infty)/n\hookrightarrow H^2_{\mathrm{et}}(X,j_!(\mu_n))$ (где j: $X\hookrightarrow\bar{X}$) и $H^2_{\mathrm{et}}(X,j_!(\mu_n))=H^2_{\mathrm{et}}(X_0,(j_0)_!(\mu_n))$, где X_0 — замкнутый слой X, ввиду теоремы о собственной замене базы в этальных когомологиях. Согласно условию (b), достаточно проверить совпадение отображений $V(X_F)\rightrightarrows V(F)$, где X_F — общий слой X. Последний факт доказывается точно так же, как и при доказательстве предложения 2.3, поскольку разность соответствующих точек делится на n в соответствующей относительной группе Пикара многообразия X_F.

Замечание 2.5. Условие (b) предложения 2.4 часто выполняется в алгебраической K-теории в силу теоремы Куиллена

[24]. Более того, от этого условия можно отказаться во многих интересных случаях.

Используя индукцию и приём, аналогичный применённому при доказательстве следствия 2.3.3, мы выводим из предложения 2.4 следующую важную теорему:

Теорема 2.6 (Габбер (неопубликовано), Жилле и Томасон [10]). *Пусть V/F — гладкое многообразие и $v \in V$ — рациональная точка. Обозначим через O_v^h гензелизацию локального кольца O_v. Для любого натурального m, не делящегося на* $\operatorname{char} F$, *канонический гомоморфизм* $K_*(O_v^h, \mathbb{Z}/m) \to K_*(F, \mathbb{Z}/m)$ *биективен.*

Обозначим через I_v^h максимальный идеал локального кольца O_v^h. Из теоремы 2.6 несложно вывести, что $H_i(\mathrm{GL}(O_v^h, I_v^h), \mathbb{Z}/m) = 0$ $(i \geqslant 1)$ (см. [32]). Рассмотрим теперь симплициальную схему BGL_n/F и обозначим через $X_{n,i}^h$ гензелизацию $(\mathrm{BGL}_n)_i = (\mathrm{GL}_n)^i$ в единице; обозначим, далее, через $O_{n,i}^h$ координатное кольцо $X_{n,i}^h$ и через $I_{n,i}^h$ его максимальный идеал. Поскольку операторы граней и вырождения симплициальной схемы BGL_n сохраняют единицу, мы видим, что $X_{n,i}^h$ также образуют симплициальную схему. Очевидные отображения $X_{n,i}^h \to \mathrm{GL}_n^i \xrightarrow{\mathrm{pr}_k} \mathrm{GL}_n$ задают матрицы $a_k \in \mathrm{GL}_n(O_{n,i}^h, I_{n,i}^h)$. Через $u_{n,i}$ мы будем обозначать цепь $[a_1, \ldots, a_i] \in C_i(\mathrm{GL}_n(O_{n,i}^h, I_{n,i}^h), \mathbb{Z}/m)$. Используя индукцию по i и то, что $H_i(\mathrm{GL}(O_{n,i}^h, I_{n,i}^h), \mathbb{Z}/m) = 0$, легко построить цепи $c_{n,i} \in C_{i+1}(\mathrm{GL}(O_{n,i}^h, I_{n,i}^h), \mathbb{Z}/m)$, такие что

$$d(c_{n,i}) = u_{n,i} - \sum_{j=0}^{i} (-1)^j (d_j)^* (c_{n,i-1}).$$

Эти рассмотрения позволяют далее обобщить теорему 2.6:

Следствие 2.7. *Пусть (R, I) — гензелева пара, причём R есть F-алгебра. Тогда $\tilde{H}_*(\mathrm{GL}(R, I), \mathbb{Z}/m) = 0$ и $K_*(R, \mathbb{Z}/m) \xrightarrow{\sim} K_*(R/I, \mathbb{Z}/m)$.*

Второе утверждение следует из первого. Для доказательства первого утверждения достаточно проверить, что вложение комплексов $\tilde{C}_*(\mathrm{GL}_n(R, I), \mathbb{Z}/m) \hookrightarrow \tilde{C}_*(\mathrm{GL}(R, I), \mathbb{Z}/m)$ гомотопно нулю. Рассмотрим матрицы $b_1, \ldots, b_i \in \mathrm{GL}_n(R, I)$. Эти матрицы определяют морфизм $\operatorname{Spec} R \to \mathrm{GL}_n$, переводящий $\operatorname{Spec} R/I$ в единицу. Поскольку пара (R, I) — гензелева, этот морфизм факторизуется через морфизм f_b: $\operatorname{Spec} R \to X_{n,i}^h$. Желаемая нульгомотопия может быть определена теперь посредством формулы $s([b_1, \ldots, b_i]) = f_b^*(c_{n,i})$.

Тот же метод вычисления значений «универсального оператора гомотопии» $c_{n,i}$ может быть использован во многих других случаях. Следующие результаты доказаны в [32]:

Теорема 2.8. *Пусть R — гензелево кольцо дискретного нормирования с максимальным идеалом I, полем частных F и полем вычетов k. Для любого m, не делящегося на характеристику поля F, имеются канонические изоморфизмы прогрупп*

$$H_*(\mathrm{GL}(R), \mathbb{Z}/m) \to \{H_*(\mathrm{GL}(R/I^n), \mathbb{Z}/m\}_n,$$
$$K_*(R, \mathbb{Z}/m) \to \{K_*(R/I^n), \mathbb{Z}/m\}_n.$$

Чтобы вывести второе утверждение из первого, надо использовать теорему Гуревича для пропространств, доказанную Паниным [43].

Следствие 2.8.1. *В условиях теоремы* 2.8 *предположим дополнительно, что* $(m, \operatorname{char} k) = 1$. *Тогда* $K_*(R, \mathbb{Z}/m) \xrightarrow{\sim} K_*(k, \mathbb{Z}/m)$.

Следствие 2.8.2. Пусть k — алгебраически замкнутое поле положительной характеристики p, и пусть F есть алгебраическое замыкание поля частных кольца векторов Витта над k. Тогда для любого m, не делящегося на p, имеем канонические изоморфизмы $K_*(k, \mathbb{Z}/m) = K_*(F, \mathbb{Z}/m)$.

Это следствие вместе с теоремой 2.2 показывает, что группы $K_i(F, \mathbb{Z}/m)$ не зависят от алгебраически замкнутого поля F (при условии, что m не делится на $\operatorname{char} F$). Это позволяет завершить доказательство гипотезы Куиллена — Лихтенбаума для полей нулевой характеристики. Более естественно, однако, применить метод универсальных операторов гомотопии и доказать следующую теорему:

Теорема 2.9 [32]. *Пусть F обозначает либо поле $\mathbb{R}$ вещественных чисел, либо поле $\mathbb{C}$ комплексных чисел. Естественный морфизм* $\mathrm{BGL}(F)^+ \to \mathrm{BGL}(F)^{\mathrm{top}}$ *индуцирует изоморфизмы на гомологиях и гомотопиях с конечными коэффициентами.*

Используя дополнительно теорему о стабилизации (см. теорему 4.6 ниже), получаем следующий результат, дающий частичный ответ на гипотезу Фридландера — Милнора об изоморфизме [21]:

Следствие 2.9.1. $\mathrm{BGL}_n(F) \to \mathrm{BGL}_n(F)^{\mathrm{top}}$ *индуцирует изоморфизмы на $H_i(-, \mathbb{Z}/m)$ при $i \leqslant n$.*

Следствие 2.9.2. *По модулю однозначно делимых групп K-теория полей $\mathbb{R}$ и $\mathbb{C}$ имеет вид, указанный в следующей таблице:*

$i \bmod 8$	0	1	2	3	4	5	6	7
$K_i(\mathbb{R})$	0	$\mathbb{Z}/2$	$\mathbb{Z}/2$	$\mathbb{Q}/\mathbb{Z}$	0	0	0	$\mathbb{Q}/\mathbb{Z}$
$\downarrow$	0	incl	0	2	0	0	0	iso
$K_i(\mathbb{C})$	0	$\mathbb{Q}/\mathbb{Z}$	0	$\mathbb{Q}/\mathbb{Z}$	0	$\mathbb{Q}/\mathbb{Z}$	0	$\mathbb{Q}/\mathbb{Z}$

Замечание 2.10. Другое и более алгебраическое доказательство гипотезы Куиллена — Лихтенбаума было позже предложено Жардином [42]. Его метод также основан на использовании теоремы 2.6.

3. K-ТЕОРИЯ АЛГЕБР С ДЕЛЕНИЕМ

K_2-теория алгебр с делением уже была использована выше при доказательстве предложения 1.4. Результат, который был нам необходим, таков. Если D/F — центральная простая алгебра и E/F — её поле разложения конечной степени, то мы можем рассмотреть канонический гомоморфизм g_E: $K_2(E) = K_2(D_E) \xrightarrow{N_{E/F}} K_2(D)$.

Теорема 3.1 [17]. *Если индекс алгебры D бесквадратен, то $K_2(D)$ порождается образами $K_2(E)$, когда E пробегает всевозможные поля расщепления D конечной степени над F.*

Замечание 3.1.1. Представляется возможным, что ограничение на индекс несущественно для справедливости теоремы 3.1. Это весьма интересная проблема. Аналогичный вопрос можно задать для высших K-групп.

Пусть X/F — многообразие Севери — Брауэра, отвечающее алгебре D. Согласно предложению 1.10, $H^0(X, K_2) = K_2(F)$, и мы получаем канонический гомоморфизм Nrd: $K_2(D) \to K_2(X) \to H^0(X, K_2) = K_2(F)$.

Теорема 3.2 [17]. *Пусть D — алгебра бесквадратного индекса над полем F. Тогда*

(a) *если* c. d. $F = 2$, *то* Nrd: $K_2(D) \xrightarrow{\sim} K_2(F)$;

(b) *если F — глобальное поле, то имеется точная последовательность* $0 \to K_2(D) \xrightarrow{\mathrm{Nrd}} K_2(F) \to \coprod_v \mathbb{Z}/2 \to 0$, *где v пробегает все вещественные точки поля F, в которых D не тривиальна.*

Теорема 3.2 легко следует из теоремы 3.1 и теоремы Гильберта 90 для K_2. Я уверен, что ограничение на индекс излишне для справедливости этой теоремы. Следующий важный результат Меркурьева [18] намного глубже:

Теорема 3.3. *Для любой кватернионной алгебры D/F гомоморфизм редуцированной нормы* Nrd: $K_2(D) \to K_2(F)$ *инъективен.*[1]

Для доказательства Меркурьев использует свой метод универсальных проблем (см. § 1). В данном случае поле определения универсальной проблемы является полем функций на произведении трёхмерных квадрик. Чтобы доказать аналог предложения 1.2, необходимо вычислить некоторые D-операторные группы K-когомологий этих квадрик. Первый шаг в этом направлении обеспечивается теоремой Суона [36], который вычислил операторную K-теорию произвольной квадрики. Далее необходимо изучить дифференциалы в BGQ-спектральной последовательности. Теория классов Чжэня больше не помогает, но Меркурьев изобрёл прямой метод доказательства тривиальности этих дифференциалов.

Замечание 3.4. (a) Меркурьев дал также описание образа гомоморфизма редуцированной нормы.

(b) Естественно ожидать, что инъективность гомоморфизма редуцированной нормы имеет место для всех алгебр бесквадратного индекса, но в настоящий момент я не вижу, как подойти к этой проблеме.

В локальном случае для вычисления K-теории алгебр с делением можно использовать видоизменение методов предыдущего параграфа.

Теорема 3.4 [35]. *Пусть R — гензелево кольцо дискретного нормирования с полем частных F, и пусть D — алгебра с делением над F. Обозначим через A максимальный порядок в D и через I — его максимальный идеал. Для любого m, не делящегося на* char F, *имеется канонический изоморфизм прогрупп* $K_i(A, \mathbb{Z}/m) \to \{K_i(A/I^n, \mathbb{Z}/m)\}_n$.

Следствие 3.4.1. *Если m не делится на* char $\bar{R}$, *то*

$$K_i(A, \mathbb{Z}/m) \xrightarrow{\sim} K_i(A/I, \mathbb{Z}/m).$$

---

[1] Теорема 3.3 независимо доказана М. Ростом [47]. (Это и все последующие подстрочные примечания, за исключением самого последнего, добавлены при корректуре английского издания.)

Следствие 3.4.2. *Пусть F — обычное локальное поле (т. е. конечное расширение поля p-адических чисел), и пусть D/F — алгебра с делением степени, не делящейся на p. При всех $i \geqslant 1$ имеются канонические изоморфизмы* $\mathrm{Nrd}\colon K_i(D) \xrightarrow{\sim} K_i(F)$.

Следствие 3.4.3 [11]. *В условиях следствия* 3.4.2, $\mathrm{Nrd}\colon K_2(D) \xrightarrow{\sim} K_2(F)$ *для любой алгебры с делением D.*

Действуя, как и при доказательстве (2.9), мы сверх того получаем

Предложение 3.5. *Обозначим через $\mathbb{H}$ классическую алгебру кватернионов над $\mathbb{R}$. Естественное отображение* $\mathrm{BGL}(\mathbb{H})^+ \to \mathrm{BGL}(\mathbb{H})^{\mathrm{top}}$ *индуцирует изоморфизмы на гомологиях и гомотопиях с конечными коэффициентами.*

4. К-ТЕОРИЯ МИЛНОРА

Для любого поля F определим его кольцо Милнора $K_*^{\mathrm{M}}(F)$ как факторкольцо тензорной алгебры $T(F^*)$ по однородному идеалу, порождённому тензорами $a \otimes (1-a) \in T_2(F^*) = F^* \otimes F^*$ [2]. Образ тензора $a_1 \otimes \ldots \otimes a_n$ в $K_n^{\mathrm{M}}(F)$ будет обозначаться $\{a_1, \ldots, a_n\}$. Имеется канонический гомоморфизм колец $K_*^{\mathrm{M}}(F) \to K_*(F)$, который является изоморфизмом в размерностях $\leqslant 2$. Пример конечного поля показывает, что в размерностях $n \geqslant 3$ отображение $K_n^{\mathrm{M}}(F) \to K_n(F)$ не обязательно сюръективно; однако я не знаю примеров, когда это отображение не инъективно (см. следствие 4.7.1 ниже)[1]. Следующая гипотеза принципиальна для понимания структуры $K_*(F)$:

Гипотеза 4.1. *Обозначим через F_0 подполе констант в F. Кольцо $K_*(F)$ порождается $K_1(F) = F^*$ и $K_*(F_0)$*[2].

В случае положительной характеристики гипотеза 4.1 означает, что гомоморфизм $K_*^{\mathrm{M}}(F) \to K_*(F)$ сюръективен по модулю кручения.

Следующая гипотеза является частным случаем общих гипотез Бейлинсона (см. § 7):

---

[1] Ч. Вибель обратил моё внимание на то, что гомоморфизм $K_n^{\mathrm{M}}(\mathbb{Q}) \longrightarrow K_n(\mathbb{Q})$ не инъективен при $n \geqslant 4$.

[2] Формулировка этой гипотезы должна быть модифицирована; легко видеть, что в настоящем виде гипотеза неверна уже в размерности 4.

Гипотеза 4.2. *Для любого поля F и любого n, не делящегося на* $\mathrm{char}\, F$, *гомоморфизм норменного вычета*

$$K_*^{\mathrm{M}}(F)/n \to \coprod_{i \geqslant 0} H^i(F, \mu_n^{\otimes i})\colon \{a_1, \ldots, a_i\} \mapsto \chi_n(a_1) \cup \ldots \cup \chi_n(a_i)$$

является изоморфизмом колец.

Другая интересная гипотеза относительно кольца $K_*^{\mathrm{M}}(F)$ — это гипотеза Милнора о квадратичных формах [20]. Предположим, что $\mathrm{char}\, F \neq 2$, и обозначим через $W(F)$ кольцо Витта невырожденных квадратичных форм над F. Пусть $I(F)$ обозначает максимальный идеал $W(F)$, состоящий из форм чётной размерности. Для любого $a \in F^*$ положим $\langle\langle a \rangle\rangle = 1 \perp -a \in I(F)$ и $\langle\langle a_1, \ldots, a_n \rangle\rangle = \langle\langle a_1 \rangle\rangle \cdot \ldots \cdot \langle\langle a_n \rangle\rangle$. Идеал $I(F)$ аддитивно порождается формами $\langle\langle a \rangle\rangle$, и, следовательно, $I^n(F)$ аддитивно порождается n-кратными формами Пфистера $\langle\langle a_1, \ldots, a_n \rangle\rangle$. Легко видеть, что дискриминант определяет изоморфизм $I(F)/I^2(F) \xrightarrow{\sim} F^*/F^{*2}$. Поскольку двукратные формы Пфистера $\langle\langle a, 1-a \rangle\rangle$ тривиальны, мы получаем канонический гомоморфизм колец

$$K_*^{\mathrm{M}}(F)/2 \to \coprod_{n \geqslant 0} I^n(F)/I^{n+1}(F)\colon \{a_1, \ldots, a_n\} \mapsto$$
$$\mapsto \langle\langle a_1, \ldots, a_n \rangle\rangle \bmod I^{n+1}(F),$$

который сюръективен согласно вышеприведённым замечаниям.

Гипотеза 4.3 (Милнор [20]). $K_n^{\mathrm{M}}(F)/2 \xrightarrow{\sim} I^n(F)/I^{n+1}(F)$.

При $n \leqslant 2$ справедливость этой гипотезы установлена в [20].

Среди общих свойств K-групп Милнора отметим следующие:

(4.4) Пусть O — кольцо дискретного нормирования с нормированием v, полем частных F и полем вычетов k. Тогда имеются канонические гомоморфизмы $\partial\colon K_n^{\mathrm{M}}(F) \to K_{n-1}^{\mathrm{M}}(k)$, которые полностью характеризуются следующей формулой:

$$\partial(\{x_1, \ldots, x_n\}) = v(x_1) \cdot \{\bar{x}_2, \ldots, \bar{x}_n\} \quad (x_2, \ldots, x_n \in O^*)$$

(см. [2]).

Замечание 4.4.1. Следующая диаграмма коммутативна:

$$\begin{tikzcd}
K_n^{\mathrm{M}}(F) \arrow[r] \arrow[d, "\partial"] & K_n(F) \arrow[d, "\partial"] \\
K_{n-1}^{\mathrm{M}}(k) \arrow[r] & K_{n-1}(k)
\end{tikzcd}$$

(4.5) *Гомоморфизмы переноса* (Басс и Тэйт [2], Като [12]). Если E/F — конечное расширение полей, то можно определить канонические гомоморфизмы $N_{E/F}$: $K_n^{\mathrm{M}}(E) \to K_n^{\mathrm{M}}(F)$, которые обладают обычными свойствами гомоморфизмов переноса и полностью характеризуются следующей формулой взаимности:

(4.5.1) Если C/F — полная регулярная кривая, то для любого $u \in K_{n+1}^{\mathrm{M}}(F(C))$ имеем

$$\sum_{x \in C} N_{F(x)/F}(\partial_x(u)) = 0.$$

Замечание 4.5.2. Следующая диаграмма коммутативна:

$$\begin{tikzcd}
K_n^{\mathrm{M}}(E) \arrow[r, "N_{E/F}"] \arrow[d] & K_n^{\mathrm{M}}(F) \arrow[d] \\
K_n(E) \arrow[r, "N_{E/F}"] & K_n(F)
\end{tikzcd}$$

K-теория Милнора тесно связана с гомологическими свойствами GL_n.

Теорема 4.6 [29]. *Пусть F — бесконечное поле. Тогда гомоморфизмы*

$$H_n(\mathrm{GL}_n(F), \mathbb{Z}) \to H_n(\mathrm{GL}_{n+1}(F), \mathbb{Z}) \to \ldots \to H_n(\mathrm{GL}(F))$$

являются изоморфизмами. Сверх того, гомологическое произведение

$$F^* \otimes \ldots \otimes F^* = H_1(\mathrm{GL}_1(F)) \otimes \ldots \otimes H_1(\mathrm{GL}_1(F)) \to H_n(\mathrm{GL}_n(F))$$

определяет изоморфизм

$$K_n^{\mathrm{M}}(F) \to H_n(\mathrm{GL}_n(F))/H_n(\mathrm{GL}_{n-1}(F)) = H_n(\mathrm{GL}(F))/H_n(\mathrm{GL}_{n-1}(F))$$

Последняя теорема позволяет определить гомоморфизм

$$\begin{aligned} \mathfrak{f}\colon K_n(F) &= \pi_n(\mathrm{BGL}(F)^+) \to H_n(\mathrm{BGL}(F)^+) \\ &= H_n(\mathrm{GL}(F)) \to H_n(\mathrm{GL}(F))/H_n(\mathrm{GL}_{n-1}(F)) = K_n^{\mathrm{M}}(F). \end{aligned}$$

Предложение 4.7. [29]. (a) *Композиция* $K_n^{\mathrm{M}}(F) \to K_n(F) \to K_n^{\mathrm{M}}(F)$ *совпадает с умножением на* $(-1)^{n-1}(n-1)!$.

(b) *Композиция* $K_n(F) \to K_n^{\mathrm{M}}(F) \to K_n(F)$ *совпадает с классом Чжэня* $c_{n,n}$.

Следствие 4.7.1. *Ядро гомоморфизма* $K_n^{\mathrm{M}}(F) \to K_n(F)$ *аннулируется числом* $(n-1)!$.

Следствие 4.7.2. *Предположим, что O — кольцо дискретного нормирования с полем частных F и полем вычетов k. Тогда диаграмма*

$$\begin{tikzcd}
K_3(F) \arrow[r, "\mathfrak{f}"] \arrow[d, "\partial"'] & K_3^{\mathrm{M}}(F) \arrow[d, "\partial"] \\
K_2(k) \arrow[r, "2"] & K_2(k)
\end{tikzcd}$$

коммутативна.

В связи с гипотезой Милнора 4.3 отметим следующее

Предложение 4.8 [29]. *Образ гомоморфизма* $\mathfrak{f}$: $K_3(F) \to K_3^{\mathrm{M}}(F)$ *совпадает с ядром гомоморфизма Милнора*

$$K_3^{\mathrm{M}}(F) \to I^3(F)/I^4(F).$$

5. K_3 И ГРУППА БЛОХА

Для любого поля F обозначим через $D(F)$ свободную абелеву группу с базисом $[x]$ $(x \in F^* - 1)$ и через r: $D(F) \to F^* \otimes F^*$ — гомоморфизм, заданный формулой $[x] \mapsto x \otimes (1-x)$. На группе $F^* \otimes F^*$ действует инволюция s по формуле $s(a \otimes b) = -(b \otimes a)$. Легко проверить, что индуцированный гомоморфизм $D(F) \to (F^* \otimes F^*)_s$ тривиален на элементах вида

$$[x] - [y] + \left[\frac{y}{x}\right] - \left[\frac{1-x^{-1}}{1-y^{-1}}\right] + \left[\frac{1-x}{1-y}\right] \quad (x \neq y \in F^* - 1).$$

Будем обозначать через $T(F)$ факторгруппу $D(F)$ по подгруппе, порождённой этими элементами. Ядро индуцированного гомоморфизма $T(F) \to (F^* \otimes F^*)_s$ обозначается $B(F)$ и называется группой Блоха поля F. Таким образом, мы имеем точную последовательность

$$0 \to B(F) \to T(F) \to (F^* \oplus F^*)_s \to K_2(F) \to 0.$$

При $x \neq 1$ положим $\langle x \rangle = [x] + [x^{-1}]$; положим также $\langle 1 \rangle = 0$.

Лемма 5.1. (a) $x \mapsto \langle x \rangle$ *определяет гомоморфизм* $F^* \to {}_2T(F)$; *в частности,* $\langle x^2 \rangle = 0$.

(b) *Элемент* $c = [x] + [1-x] \in T(F)$ *не зависит от выбора* $x \in F^* - 1$.

(c) $3c = \langle -1 \rangle$.

(d) *Если уравнение* $x^2 + 1 = 0$ *имеет решение в* F, *то* $3c = 0$; *если уравнение* $x^2 - x + 1 = 0$ *имеет решение в* F, *то* $2c = 0$.

Группа $B(F)$ следующим образом связана с $K_3(F)$. Обозначим через $\mathrm{GM}(F)$ подгруппу $\mathrm{GL}(F)$, состоящую из мономиальных матриц. Эта группа квазисовершенна, так что к $\mathrm{BGM}(F)$ можно применить плюс-конструкцию Куиллена. Ввиду теоремы Баррата—Придди—Куиллена гомотопические группы пространства $\mathrm{BGM}(F)^+$ совпадают со стабильными гомотопическими группами BF^* и, следовательно, более или менее вычислимы. Вложение групп $\mathrm{GM}(F)\hookrightarrow \mathrm{GL}(F)$ индуцирует морфизм пространств $\mathrm{BGM}(F)^+\to \mathrm{BGL}(F)^+$ и, следовательно, гомоморфизмы $\pi_i^s(BF^*)=\pi_i(\mathrm{BGM}(F)^+)\to K_i(F)$. Эти гомоморфизмы сюръективны в размерностях $\leqslant 2$.

Теорема 5.2 [33]. *Если поле F бесконечно, то*

$$\operatorname{coker}(\pi_3(\mathrm{BGM}(F)^+)\to K_3(F))=B(F)/2c.$$

Доказательство использует гомологическую технику. Сначала доказывается, что

$$\operatorname{coker}(\pi_3(\mathrm{BGM}(F)^+)\to K_3(F))=\operatorname{coker}(H_3(\mathrm{GM}(F))\to H_3(\mathrm{GL}(F))).$$

Затем вычисляется группа $H_3(\mathrm{GL}_2(F))/H_3(\mathrm{GM}_2(F))$. Этот шаг очень близок к доказательству теоремы Блоха [6]. Рассмотрим комплекс $C_*(F)$, где $C_i(F)$ есть свободная абелева группа, порождённая всевозможными наборами $(x_0, \ldots, x_i)$ различных точек $\mathbb{P}^1(F)$. Легко видеть, что все гомологии этого комплекса тривиальны, за исключением группы H_0, которая равна $\mathbb{Z}$. Естественное действие группы $\mathrm{GL}_2(F)$ на комплексе $C_*(F)$ даёт спектральную последовательность

$$H_p(\mathrm{GL}_2(F),\ C_q(F))\Rightarrow H_{p+q}(\mathrm{GL}_2(F),\ \mathbb{Z}).$$

Действие $\mathrm{GL}_2(F)$ на базисе $C_i(F)$ транзитивно при $i=0,1,2$ и стабилизаторы наборов (0), $(0,\infty)$, $(0,\infty,1)$ соответственно равны $B_2=\begin{pmatrix} F^* & * \\ 0 & F^* \end{pmatrix}$, $T_2=\begin{pmatrix} F^* & 0 \\ 0 & F^* \end{pmatrix}$ и F^*. Таким образом, член E^1 этой спектральной последовательности выглядит следующим образом:

$$H_*(B_2)\quad H_*(T_2)\quad H_*(F^*)\qquad \overset{*}{\coprod_{x\in\mathbb{P}^1(F)-\{0,\infty,1\}}}\mathbb{Z}\cdot[x]\qquad \overset{*}{\coprod_{x\neq y\in\mathbb{P}^1(F)-\{0,\infty,1)}}\mathbb{Z}\cdot[x,y],$$

где символом $[x]$ (соотв. $[x,y]$) обозначена орбита набора $(0,\infty,1,x)$ (соотв. $(0,\infty,1,x,y)$). Используя то, что $H_*(B_2)=H_*(T_2)$, легко вычислить дифференциал d^1. Представляющие для нас интерес E^2-члены выглядят так (через s обозначена инволюция на $H_*(T_2)$, индуцированная перестановкой диагональных членов):

$$\begin{array}{llll}
H_3(T_2)_s & & & \\
H_2(T_2)_s=H_2(F^*)\oplus(F^*\otimes F^*)_s & (F^*\otimes F^*)^s & 0 & \\
F^* & 0 & 0 & \\
\mathbb{Z} & 0 & 0 & T(F)
\end{array}$$

Единственный нетривиальный дифференциал, исходящий из $T(F)$, — это d^3: $T(F)\to H_2(F^*)\oplus(F^*\otimes F^*)_s$, который задан формулой

$$d^3([x])=x\wedge(1-x)-x\otimes(1-x)\in\Lambda^2(F^*)\oplus(F^*\otimes F^*)_s.$$

Ясно, что $E^\infty_{0,3}=\ker d^3=B(F)$. Наша спектральная последовательность определяет фильтрацию на $H_3(\mathrm{GL}_2(F))$. На группе $H_3(\mathrm{GM}_2(\mathrm{F}))$ также имеется фильтрация (возникающая из спектральной последовательности Хохшильда—Серра, отвечающей расширению $1\to T_2\to\mathrm{GM}_2(F)\to S_2\to 1$). В обоих случаях нулевой член фильтрации совпадает с образом $H_3(T_2)$. Таким образом, гомоморфизм $H_3(\mathrm{GM}_2(F))\to H_3(\mathrm{GL}_2(F))$ переводит $H_3(\mathrm{GM}_2(F))^0$ на $H_3(\mathrm{GL}_2(F))^0$. Сравнивая члены $E^2_{2,1}$ рассматриваемых спектральных последовательностей, несложно показать, что $H_3(\mathrm{GM}_2(F))^1$ отображается на $H_3(\mathrm{GL}_2(F))^1$. Наконец, $H_3(\mathrm{GM}_2(F))=H_3(\mathrm{GM}_2(F))^1+H_3(S_2)$, и образ $H_3(S_2)$ в $H_3(\mathrm{GL}_2(F))$ очевидно содержится в $H_3(T_2)$. Таким образом, $H_3(\mathrm{GL}_2(F))/H_3(\mathrm{GM}_2(F))=H_3(\mathrm{GL}_2(F))/H_3(\mathrm{GL}_2(F))^1=E^\infty_{0,3}=B(F)$.

Затем надо проверить, что ядро гомоморфизма $H_3(\mathrm{GL}_2(F))\to H_3(\mathrm{GL}_3(F))$ содержится в образе $H_3(\mathrm{GM}_2)$. После этого остаётся только вычислить пересечение $H_3(\mathrm{GL}_2(F))$ и $H_3(\mathrm{GM}(F))$ в $H_3(\mathrm{GL}(F))$. Ввиду изоморфизма $H_3(\mathrm{GL}(F))/H_3(\mathrm{GL}_2(F))=K_3^M(F)$, это равносильно вычислению ядра $H_3(\mathrm{GM}(F))\to K_3^M(F)$. Ответ таков: рассмотрим на $H_3(\mathrm{GM}(F))$ фильтрацию, возникающую из спектральной последовательности Хохшильда—Серра; тогда $H_3(\mathrm{GM}(F))^2\cap H_3(\mathrm{GL}_2(F))=H_3(\mathrm{GM}_2(F))$. Поскольку $H_3(\mathrm{GM}(F))=H_3(\mathrm{GM}(F))^2+H_3(S)$, мы заключаем, что

$$H_3(\mathrm{GL}(F))/H_3(\mathrm{GM}(F))=B(F)/\mathrm{Im}(H_3(S)),$$

и теперь достаточно проверить, что $\mathrm{Im}(H_3(S))=2c$.

Для применения теоремы 5.2 необходимо знать группу $\pi_3(\mathrm{BGM}(F)^+)$ и её образ в $K_3(F)$. Используя спектральную последовательность $H_i(F^*,\ \pi_j^s(\mathrm{pt}))\Rightarrow\pi_{i+j}^s(BF^*)$, легко доказать следующее предложение:

Предложение 5.3. *Обозначим через* $\overline{K}_3^{\mathrm{M}}(F)$ *образ* $K_3^{\mathrm{M}}(F)$ *в* $K_3(F)$ *и через* G *(соотв.* G_μ*) подгруппу* GM, *состоящую из мономиальных матриц, ненулевые члены которых равны* ± 1 *(соотв. входят в группу* μ *корней из единицы).*

(a) $\mathrm{Im}\left(\pi_3\left(\mathrm{BGM}(F)^+\right) \to K_3(F)\right)$

$$= \overline{K}_3^{\mathrm{M}}(F) + \mathrm{Im}\left(\pi_3(BG\mu)^+ \to K_3(F)\right).$$

(b) *Имеется канонический сюръективный гомоморфизм* $\mathrm{Tor}(\mu, \mu) = \mathrm{Tor}(F^*, F^*) \to \mathrm{Im}\left(\pi_3\left(\mathrm{BGM}(F)^+\right)/K_3^{\mathrm{M}}(F) + \mathrm{Im}\left(\pi_3\left(BG^+\right)\right)\right.$. [1]

Следствие 5.3.1. *Группа* $B(F)$ *не меняется при чисто трансцендентных расширениях.*

Следствие 5.3.2. *Обозначим через* F_0 *подполе констант в* F. *Тогда* $K_3(F)/K_3(F_0) + K_3^{\mathrm{M}}(F) = B(F)/B(F_0)$.

Применительно к группе K_3 гипотеза 4.1 принимает такой вид:

Гипотеза 5.4. $B(F) = B(F_0)$.

Ниже нам потребуется несколько модифицированное описание $B(F)$.

(5.5) Обозначим через $\overline{F^* \otimes F^*}$ факторгруппу $F^* \otimes F^*$ по подгруппе, порождённой тензорами $a \otimes (-a)$, и через $T'(F)$ — факторгруппу $T(F)$ по подгруппе, порождённой элементами $\langle a \rangle$. Поскольку $r(\langle a \rangle) = a \otimes (-a)$, мы получаем индуцированный гомоморфизм $T'(F) \to \overline{F^* \otimes F^*}$, ядро которого будем обозначать $B'(F)$.

Лемма 5.5.1. $B'(F) = B(F)/\langle -1 \rangle$.

(5.6) Можно также дать описание группы $B(F)$ в терминах соотношений на элементы $a \otimes (1 - a)$ непосредственно в $F^* \otimes F^*$. Легко видеть, что

$$r\left([x] - [y] + \left[\frac{y}{x}\right] - \left[\frac{1 - x^{-1}}{1 - y^{-1}}\right] + \left[\frac{1 - x}{1 - y}\right]\right)$$
$$= x \otimes \frac{1 - x}{1 - y} + \frac{1 - x}{1 - y} \otimes x = r\left(\left\langle x \cdot \frac{1 - x}{1 - y}\right\rangle - \langle x \rangle - \left\langle \frac{1 - x}{1 - y} \right\rangle\right).$$

---

[1] Более точная связь между $K_3(F)$ и $B(F)$ устанавливается точной последовательностью $0 \to \mathrm{Tor}(F^*, F^*)^\sim \to K_3(F)_{\mathrm{ind}} \to B(F) \to 0$, где $K_3(F)_{\mathrm{ind}} = K_3(F)/K_3^M(F)$ и $\mathrm{Tor}(F^*, F^*)^\sim$ есть единственное нетривиальное расширение $Z/2$ при помощи $\mathrm{Tor}(F^*, F^*)$.

Таким образом, ядро r содержит элементы

$$[x] - [y] + \left[\frac{y}{x}\right] - \left[\frac{1 - x^{-1}}{1 - y^{-1}}\right]$$
$$+ \left[\frac{1 - x}{1 - y}\right] - \left\langle x \cdot \frac{1 - x}{1 - y}\right\rangle + \langle x \rangle + \left\langle \frac{1 - x}{1 - y} \right\rangle_{(x \neq y \in F^* - 1)},$$
$$\langle xyz \rangle - \langle xy \rangle - \langle xz \rangle - \langle yz \rangle + \langle x \rangle + \langle y \rangle + \langle z \rangle,$$
$$\langle x^2 \rangle - 4\langle x \rangle$$

(где, как всегда, $\langle 1 \rangle = 0$). Обозначим через $T''(F)$ факторгруппу $D(F)$ по подгруппе, порождённой вышеприведёнными элементами, и через $B''(F)$ — ядро гомоморфизма $T''(F) \to F^* \otimes F^*$.

Лемма 5.6.1. $B''(F) = B(F)$.

6. ДЕЛИМОСТЬ В ГРУППЕ БЛОХА [1]

Все поля, рассматриваемые в этом разделе, содержат алгебраически замкнутое подполе.

Предположим, что поле F дискретно нормировано с кольцом нормирования O и полем вычетов k. Выберем локальный параметр π. Этот выбор позволяет определить гомоморфизм s_π: $F^* \to k^* : x \longmapsto \overline{x/\pi^{v(x)}}$ и индуцированные гомоморфизмы $F^* \otimes F^* \to k^* \otimes k^*$, $\overline{F^* \otimes F^*} \to \overline{k^* \otimes k^*}$. Положим $\bar{x} = \infty$ при $x \notin O$ и определим элементы $[0]$, $[1]$, $[\infty] \in T'(k)$ как нулевые.

Лемма 6.2. $[x] \longmapsto [\bar{x}]$ *определяет гомоморфизм* $T'(F) \xrightarrow{s} T'(k)$. *При этом следующая диаграмма коммутативна:*

$$\begin{tikzcd}
T'(F) \arrow[r] \arrow[d, "s"'] & \overline{F^* \otimes F^*} \arrow[d, "s_\pi"'] \\
T'(k) \arrow[r] & \overline{k^* \otimes k^*}
\end{tikzcd}$$

Следовательно, s *переводит* $B'(F)$ *в* $B'(k)$.

Если E/F — конечное расширение, то $N_{E/F} : K_3(E) \to K_3(F)$ определяет ввиду (5.3) и (4.5) гомоморфизм переноса $N_{E/F} : B(E) \to B(F)$. Незначительная модификация доказательства (2.3) позволяет установить

Предложение 6.2. *Пусть* C — *гладкая связная кривая над алгебраически замкнутым полем* F. *Для любых двух точек*

---

[1] Другой и гораздо более продуктивный подход к изучению делимости в $K_3(F)_{\mathrm{ind}}$ (а следовательно и в $B(F)$) развит в [46].

$x, y \in C$ гомоморфизмы специализации $s_x, s_y : B(F(C)) \to B(F)$ совпадают на B/n и ${}_nB$.

Теорема 6.3. *Если поле F алгебраически замкнуто, то группа $B(F)$ однозначно делима.*

Поскольку $\overline{F^* \otimes F^*}$ и $K_2(F)$ однозначно делимы, достаточно установить однозначную делимость группы $T'(F) = T(F)$. Делимость $T(F)$ установлена в [6]. Она следует из формул

$$[x^p] = p\Bigl(\sum_{\xi \in \mu_p} [\xi x]\Bigr) \quad (p \neq \operatorname{char} F), \quad [x^p] = p^2[x] \quad (p = \operatorname{char} F),$$

которые справедливы для любого поля F. Чтобы доказать эти формулы, рассмотрим элемент $[t^p] - p\Bigl(\sum_{\xi \in \mu_p} [\xi t]\Bigr) \in B(F(t))$. Поскольку $B(F(t)) = B(F)$, этот элемент равен любой своей специализации. Специализируя в нуле, мы получаем нуль. Чтобы доказать однозначную делимость, определим гомоморфизм $T'(F) \to T'(F)$, обратный к умножению на p, посредством формулы $[x] \longmapsto \sum_{y^p = x} [y]$. Необходимо проверить, что определяющие соотношения на $[x]$ переходят в нуль, т. е. проверить следующую формулу (где $u, v, w \notin \mu_p \cup 0$ и $w^p = (1 - u^p)/(1 - v^p)$):

$$\sum_{\xi \in \mu_p} [\xi u] - \sum_{\xi \in \mu_p} [\xi v] + \sum_{\xi \in \mu_p} [\xi \cdot v/u] - \sum_{\xi \in \mu_p} [\xi \cdot wv/u] + \sum_{\xi \in \mu_p} [\xi \cdot w] = 0.$$

Отметим, что рассматриваемый элемент лежит в ${}_pB(F)$. Зафиксируем w и рассмотрим кривую C, заданную уравнением $(1 - V^p)w^p = 1 - U^p$. Мы можем рассмотреть далее универсальный элемент, лежащий в ${}_pB(F(C))$, специализацией которого является рассматриваемый элемент. Специализируя универсальный элемент в точке $U = 1$, $V = 1$, мы получим нуль, и нам осталось лишь воспользоваться предложением 6.2.

Будем предполагать в дальнейшем, что $\operatorname{char} F \neq 2$. Пусть E/F — квадратичное расширение: $E = F(\alpha)$, $\alpha^2 = a \in F^*$. Обозначим через A образ гомоморфизма $E^* \otimes F^* \to E^* \otimes E^*$. Легко видеть, что $N_{E/F} \otimes \mathrm{id}$: $E^* \otimes F^* \to F^* \otimes F^*$ индуцирует гомоморфизм N: $A \to F^* \otimes F^*$. Пусть $T''(E/F)$ обозначает обратный образ A в $T''(E)$. Нетрудно указать образующие и определяющие соотношения для этой группы. Важным является то, что соотношения имеют «рациональный характер» (т. е. параметризуются F-рациональными многообразиями).

Предложение 6.4. *Существует канонический гомоморфизм (заданный рациональными формулами) $L_{E/F}$: $T''(E/F) \to T''(F)$, делающий следующую диаграмму коммутативной:*

$$\begin{tikzcd}
T''(E/F) \arrow[r] \arrow[d, "L_{E/F}"'] & A \arrow[d, "N"] \\
T''(F) \arrow[r] & F^* \otimes F^*
\end{tikzcd}$$

и, следовательно, индуцирующий гомоморфизм $L_{E/F}$: $B(E) \to B(F)$.

На образующих $L_{E/F}$ задаётся явными формулами. Чтобы проверить, что определяющие соотношения переходят в нуль, надо лишь заметить, что образ любого соотношения будет рационально параметризованным элементом в $B(F)$ и, следовательно, обязан равняться нулю.

Теорема 6.5. *Обозначим через t образующую группы* $\mathrm{Gal}(E/F)$. *Следующая последовательность точна:*

$$B(E) \xrightarrow{1-t} B(E) \xrightarrow{L_{E/F}} B(F).$$

Как и при доказательстве теоремы 1.5, мы прежде всего сводим общий случай к случаю, когда гомоморфизм $N_{E/F}$: $E^* \to F^*$ сюръективен. В настоящей ситуации это тривиально: чтобы сделать элемент $b \in F^*$ нормой, достаточно перейти к полю функций на конике C, заданной уравнением $X^2 - aY^2 = b$. Поле $E(C)$ чисто трансцендентно над E, и, следовательно, $B(E(C)) = B(E)$. Таким образом,

$$\ker L_{E/F}/(1 - t)\,B(E) \hookrightarrow \ker L_{E(C)/F(C)}/(1 - t)\,B(E(C)).$$

Предполагая теперь, что $N_{E/F} : E^* \to F^*$ — сюръекция, определим отображение f: $T''(F) \to T''(E/F)_t$ при помощи формулы

$$f([x]) = [-z] + [(1 + z^t)\,z/(1 + z)] - [-(1 + z^t)/(1 + z)],$$

где элемент $z \in E^*$ таков, что $N_{E/F}(z) = x$ и $\operatorname{Tr}(z) \neq -2$. Легко видеть, что f и $L_{E/F}$ взаимно обратны. Положим $M = \operatorname{Im}(T''(E/F) \to A)$. Мы получаем две короткие точные последовательности t-модулей:

$$0 \to B_t(E) \to T''(E/F) \to M \to 0, \quad 0 \to M \to A \to K_2(E) \to 0.$$

Гомологии группы $G = \mathrm{Gal}(E/F)$ с коэффициентами в A и $K_2(F)$ легко вычисляются (во втором случае существенно используются результаты § 1). Это позволяет вычислить гомологии с коэффициентами в M: $H_i(G, M) = \mathbb{Z}/2$. Теперь легко проверить, что гомоморфизм $H_1(G, T''(E/F)) \to H_1(G, M)$ сюрьективен, и, следовательно, $(1-t)T''(E/F) \cap B(E) = (1-t)B(E)$.

Применяя теорему 6.5 к универсальному расширению Куммера, получаем

Следствие 6.6. *Для любого поля F (содержащего алгебраически замкнутое подполе) характеристики, не равной двум, группа $B(F)$ не имеет 2-кручения.*

Теорема 6.7. *В условиях следствия* 6.6. *группа $B(F)$ однозначно 2-делима*[1].

(6.7.1) Пусть C — коника над F. Если $B(F)$ 2-делима, то то же верно для $B(F(C))$.

Пусть E — квадратичное расширение F, расщепляющее C. Если $u \in B(F(C))$, то $zu = N_{E(C)/F(C)}(u_{E(C)})_{F(C)}$ (поскольку $B(E(C)) = B(E)$). Запишем $N_{E/F}(u_{E(C)}) = 4v$; тогда $u = 2v_{F(C)}$.

Следствие 6.7.2. *Если F есть поле функций на произведении коник, определённых над алгебраически замкнутым полем, то группа $B(F)$ 2-делима.*

Используя описание кручения в K_2, несложно проверить точность последовательности $0 \to B(F)/2 \to T'(F)/2 \to \wedge^2(F^*/F^{*2})$. Это позволяет выписать универсальные элементы для $B/2$. Поля определения этих элементов являются полями функций на произведении коник, и мы заключаем, что эти элементы равны нулю согласно (6.7.2). Специализация завершает доказательство.

Следствие 6.8. *Пусть F такое же, как выше, и пусть F_0 обозначает подполе констант в F. Тогда $K_3(F) = K_3(F_0) + K_3^{\mathrm{M}}(F) + 2K_3(F)$.*

Следствие 6.9. *Если F такое же, как выше, то $K_3^{\mathrm{M}}(F)/2 \xrightarrow{\sim} I^3(F)/I^4(F)$.*

Действительно, $\ker(K_3^{\mathrm{M}}(F) \to I^3/I^4) \to \mathrm{Im}(K_3(F) \to K_3^{\mathrm{M}}(F))$, но образы всех трёх слагаемых, возникающих в (6.8), очевидно содержатся в $2K_3^{\mathrm{M}}(F)$.

---

[1] Группа $B(F)$ однозначно делима для любого F, содержащего алгебраически замкнутое подполе [46].

Следствие 6.10. *Пусть F такое же, как выше, и пусть X/F — гладкое многообразие. Тогда образы гомоморфизмов*

$$K_3(F(X)) \to \coprod_{\operatorname{codim} x = 1} K_2(F(x)), \quad K_3^{\mathrm{M}}(F(X)) \to \coprod_{\operatorname{codim} x = 1} K_2(F(x))$$

совпадают[1].

Замечание 6.11. Похоже, что следствие 6.10 вместе с теоремой Меркурьева 3.3 позволяют доказать теорему Гильберта 90 для K_3^{M} (для квадратичных расширений) и, в частности, доказать, что $K_3^{\mathrm{M}}(F)/2 = H^3(F, \mu_2)$, но мы ещё не проверили все детали[2].

7. ВЫСШИЕ ГРУППЫ ЧЖОУ

Чтобы сделать связь между K-теорией и этальными когомологиями более осязаемой, Бейлинсон [3] предположил наличие некоторой «универсальной» теории когомологий на категории схем, которая непосредственно связана как с K-теорией, так и с этальными когомологиями (эта теория должна быть аналогом теории целочисленных сингулярных когомологий в топологии). Близкий список гипотез был предложен Лихтенбаумом [15]. Согласно Бейлинсону должны существовать комплексы пучков $\Gamma(i)$ на большом ситусе Зарисского, удовлетворяющие (среди прочего) следующим требованиям:

(a) $\Gamma(i) = 0$ при $i < 0$, $\Gamma(0) = \mathbb{Z}$, $\Gamma(1) = O^*[-1]$.

(b) При $i \geqslant 1$ комплекс $\Gamma(i)$ ацикличен вне $1, \ldots, i$; для гладкого X пучок групп $H^i(\Gamma(i))$ совпадает с пучком K-групп Милнора K_i^{M}.

(c) Для любого n, обратимого на «хорошем» гладком X, имеем $\Gamma(i) \otimes^L \mathbb{Z}/n = \tau_{\leqslant i} R\pi_* \mathbb{Z}/n(i)$, где π: $X_{\mathrm{et}} \to X_{\mathrm{Zar}}$ — канонический морфизм.

(d) Имеется спектральная последовательность, расщепляющаяся с точностью до стандартных факториалов посредством классов Чжэня $H^i(X, \Gamma(j)) \Rightarrow K'_{2j-i}(X)$. Возникающая на K'-теории фильтрация совпадает с γ-фильтрацией.

Недавно С. Блох построил теорию, которая обладает свойствами (a) и (d). Нет сомнений, что это и есть желаемая теория, но доказать оставшиеся свойства представляется очень трудным. Мы будем работать в категории квазипроективных многообразий над полем. Определим стандартный симплекс Δ^n

---

[1] Утверждения 6.9 и 6.10 справедливы для любого поля F [46].

[2] Теорема Гильберта 90 для K_3^{M} и изоморфизм $K_3^{\mathrm{M}}(F)/2 = H^3(F, \mu_2)$ доказаны в [45, 48].

как гиперплоскость в $\mathbb{A}^{n+1}$, заданную уравнением $t_0 + \ldots + t_n = 1$. Симплексы Δ^n образуют косимплициальное многообразие. Для любого многообразия X обозначим через $z^i(X, n)$ подгруппу группы $Z^i(X \times \Delta^n)$, состоящую из тех циклов, которые собственно пересекают $X \times \Delta^m$ для любой грани $\Delta^m \subset \Delta^n$. Очевидно, $z^i(-, -)$ является комплексом (степени -1) пучков в любой разумной топологии; ожидаемый комплекс $\Gamma(i)$ получается из $z^i(-, -)$ посредством переиндексации. Блох определил $\mathrm{CH}i(X, n)$ как n-ю группу гомологий комплекса $z^i(X, -)$ (группа $\mathrm{CH}^i(X, 0)$ очевидно совпадает с обычной группой Чжоу циклов коразмерности i по модулю рациональной эквивалентности). Группы $\mathrm{CH}^i(X, n)$ контравариантно зависят от X, если ограничиться плоскими морфизмами; на $\mathrm{CH}^i(X, n)$ можно определить гомоморфизм обратного образа по отношению к произвольным морфизмам, если ограничиться рассмотрением гладких многообразий. Группы $\mathrm{CH}^i(X, n)$ ведут себя ковариантно относительно собственных морфизмов. Блох доказал также следующие свойства $\mathrm{CH}^i(X, n)$:

(7.1) *Гомотопическая инвариантность*: $\mathrm{CH}^i(X, n) = \mathrm{CH}^i(X \times \mathbb{A}^1, n)$.

(7.2) *Локализация.* Если $Y \subset X$ — замкнутое подмногообразие чистой коразмерности d, то имеется точная последовательность

$$\mathrm{CH}^i(X - Y, n+1) \to \mathrm{CH}^{i-d}(Y, n) \to \mathrm{CH}^i(X, n) \to \mathrm{CH}^i(X - Y, n) \to \ldots \to \mathrm{CH}^i(X - Y, 0) \to 0.$$

(7.3) *Умножение.* Для любых X, Y имеются канонические спаривания

$$\mathrm{CH}^i(X, n) \otimes \mathrm{CH}_j(Y, m) \to \mathrm{CH}^{i+j}(X \times Y, n+m).$$

Комбинируя эти спаривания с гомоморфизмом обратного образа относительно диагонального вложения, получаем на $\mathrm{CH}^*(X, *)$ (для гладкого X) структуру биградуированного кольца.

(7.4)
$$\mathrm{CH}^1(X, q) = \begin{cases} \operatorname{Pic} X, & q = 0, \\ \Gamma(X, O_X^*), & q = 1, \\ 0, & q \geqslant 2. \end{cases}$$

(7.5) *Связь с К-теорией.* Блох показал, что $\mathrm{CH}^*(X, *)$ удовлетворяют аксиомам Жилле [44] и, следовательно, существует теория классов Чжэня со значениями в $\mathrm{CH}^*(X, *)$. Он доказал также, что эти классы Чжэня определяют изоморфизмы

$$\mathrm{CH}^i(X, n) \otimes Q = \mathrm{gr}^i K'_n(X) \otimes Q.$$

(7.6) Гипотеза Герстена верна для $\mathrm{CH}^i(X, n)$.

Спектральная последовательность, связывающая высшие группы Чжоу с *К*-теорией, была построена ранее Ландсбургом [13] [1].

Рассмотрим случай поля. Ясно, что $\mathrm{CH}^i(\operatorname{Spec} F, n) = 0$ при $n < i$. Легко также проверить, что $\mathrm{CH}^n(\operatorname{Spec} F, n) = K_n^{\mathrm{M}}(F)$. Эти факты в точности соответствуют свойствам из пункта (b) гипотез Бейлинсона (см. выше). Однако остающаяся часть пункта (b) означает, что $\mathrm{CH}^i(\operatorname{Spec} F, n) = 0$ при $n \geqslant 2i$. Вопрос о справедливости этого равенства при $i > 1$ представляется очень сложным. Наконец, отметим, что группа $\mathrm{CH}^2(\operatorname{Spec} F, 3)$ совпадает с $K_3(F)_{\mathrm{ind}}$ и, тем самым, очень близка к группе Блоха $B(F)$.

8. ЭТАЛЬНАЯ *К*-ТЕОРИЯ

Для любой симплициальной схемы X можно построить определённое пропространство X_{et} — её эталь-топологический тип [1, 8]; $X \longmapsto X_{\mathrm{et}}$ есть функтор из категории схем в категорию пропространств. Основным свойством X_{et} является то, что его фундаментальная группа совпадает с фундаментальной группой X, определённой Гротендиком, а когомологии с конечными коэффициентами совпадают с этальными когомологиями X.

Для многообразия над полем $\mathbb{C}$ его этальная *К*-теория может быть определена как топологическая *К*-теория пропространства X_{et}. В общем случае можно поступить следующим образом [7]. Зафиксируем простое число l и обозначим $\mathbb{Z}[l^{-1}]$ через R. Будем рассматривать только схемы над R. Для любого X имеем морфизмы пропространств

$$X_{\mathrm{et}} \to (\operatorname{Spec} R)_{\mathrm{et}} \leftarrow (\mathrm{BGL}_n)_{\mathrm{et}}.$$

Рассмотрим теперь пространство относительных l-адических функций [7, 8]

$$\mathrm{Hom}_l\left(X_{\mathrm{et}}, (\mathrm{BGL}_n)_{\mathrm{et}}\right)_{R_{\mathrm{et}}}$$

и положим

$$K_i^{\mathrm{et}}(X) = \varinjlim_n \pi_i\left(\mathrm{Hom}_l\left(X_{\mathrm{et}}, (\mathrm{BGL}_n)_{\mathrm{et}}\right)_{R_{\mathrm{et}}}\right),$$

$$K_i^{\mathrm{et}}(X, \mathbb{Z}/l^\nu) = \varinjlim_n \pi_i\left(\mathrm{Hom}_l\left(X_{\mathrm{et}}, (\mathrm{BGL}_n)_{\mathrm{et}}\right)_{R_{\mathrm{et}}}, \mathbb{Z}/l^\nu\right).$$

---

[1] *Примечание при переводе.* В работе Ландсбурга [13] имеются неисправимые ошибки, работа Блоха также содержит ряд ошибок.

Этальная K-теория легко вычислима ввиду наличия спектральной последовательности, связывающей её с этальными когомологиями (которая сильно сходится, если X имеет конечную l-когомологическую размерность):

$$E_2^{p,q} = H^p_{\mathrm{cont}}(X_{\mathrm{et}}, \mathbb{Z}_l(q/2)) \Rightarrow K^{\mathrm{et}}_{q-p}(X),$$

$$E_2^{p,q} = H^p(X_{\mathrm{et}}, \mathbb{Z}/l^\nu(q/2)) \Rightarrow K^{\mathrm{et}}_{q-p}(X, \mathbb{Z}/l^\nu)$$

$$(E_2^{p,q} = 0, \text{ если } q \text{ нечётно}).$$

Если X есть квазипроективное многообразие над нётеровой R-алгеброй, то имеются канонические гомоморфизмы $K_i(X) \to K_i^{\mathrm{et}}(X)$, $K_i(X, \mathbb{Z}/l^\nu) \to K_i^{\mathrm{et}}(X, \mathbb{Z}/l^\nu)$.

Теорема 8.1 [7]. *Пусть A есть кольцо целых в поле алгебраических чисел, и пусть l — простое число. Если $l = 2$, предположим дополнительно, что $F \ni \sqrt{-1}$. Тогда канонические гомоморфизмы*

$$K_i(A) \otimes \mathbb{Z}_l \to K_i^{\mathrm{et}}(A[l^{-1}])$$

сюръективны.

Следует отметить, что гипотеза Куиллена—Лихтенбаума для числовых полей равносильна тому, что гомоморфизм, рассматриваемый в теореме 8.1, биективен.

Предположим, что A есть R-алгебра, содержащая примитивный корень степени l^ν из единицы ξ. Группа $\pi_2(BA^*, \mathbb{Z}/l^\nu)$ совпадает с группой корней степени l^ν из единицы в A. Образ ξ при каноническом гомоморфизме $\pi_2(BA^*, \mathbb{Z}/l^\nu) \to K_2(A, \mathbb{Z}/l^\nu)$ (индуцированном вложением $BA^* \to \mathrm{BGL}(A)^+$) обозначается β и называется элементом Ботта. Пусть X — схема конечной l-когомологической размерности над A. Легко видеть, что её этальная K-теория $K_*^{\mathrm{et}}(X, \mathbb{Z}/l^\nu)$ два-периодична и периодичность задаётся умножением на β. Таким образом, получаем индуцированное отображение $K_*(X, \mathbb{Z}/l^\nu)[\beta^{-1}] \to K_*^{\mathrm{et}}(X, \mathbb{Z}/l^\nu)$ (при $l = 2$ или 3 надо быть аккуратнее, так как в этих случаях возникают трудности с умножением в $K_*(X, \mathbb{Z}/l^\nu)$). Фундаментальный результат, связывающий алгебраическую K-теорию с этальной, даётся следующей теоремой Томасона [38]:

Теорема 8.2. *При незначительных дополнительных ограничениях (по поводу точных формулировок см.* [38]) *индуцированное отображение*

$$K_*(X, \mathbb{Z}/l^\nu)[\beta^{-1}] \to K_*^{\mathrm{et}}(X, \mathbb{Z}/l^\nu)$$

является изоморфизмом.

ЛИТЕРАТУРА

1. M. Artin and B. Mazur, Etale homotopy, Lecture Notes in Math., Vol. 100, Springer-Verlag, Berlin — New York, 1969.

2. H. Bass and J. Tate, The Milnor ring of a global field, Lecture Notes in Math., Vol. 342, Springer-Verlag, Berlin — New York, 1973, pp. 349—446.

3. A. Beilinson, Letter te C. Soulé, November 1, 1982.

4. S. Bloch. Torsion algebraic cycles, K_2 and Brauer groups of function fields, Bull. Amer. Math. Soc. **80** (1974), 941—945.

5. —, Algebraic cycles and higher K-theory, Adv. in Math. **61** (1986), 267—304.

6. J. Dupont and C. Sah, Scissor congruences. II, J. Pure Appl. Algebra **25** (1982), 159—195.

7. W. Dwyer and E. Friedlander, Algebraic and etale K-theory, Trans. Amer. Math. Soc. **292** (1985), 247—280.

8. E. Friedlander, Etale homotopy of simplicial schemes, Ann. of Math. Studies, No. 104, Princeton Univ. Press, Princeton, N. J., 1982.

9. S. Gersten, Problems about higher K-functors, Lecture Notes in Math., Vol. 341, Springer-Verlag, Berlin — New York, 1973, pp. 43—57.

10. H. Gillet and R. Thomason, The K-theory of strict hensel local rings and a theorem of Suslin, J. Pure Appl. Algebra **34** (1984), 241—254.

11. Юфряков А. В. О группе K_2 для локальных алгебр с делением, ДАН СССР, 1986, т. 291, № 1, с. 54—56.

12. K. Kato, A generalisation of local class field theory by using K-groups, J. Fac. Sci. Univ. Tokyo 1A Math. **26** (1979), 303—376; **27** (1980), 603—683.

13. S. Landsburg, Relative cycles and algebraic K-theory, Preprint, 1983.

14. S. Lichtenbaum, Values of zeta-function, etale cohomology and algebraic K-theory, Lecture Notes in Math., Vol. 342, Springer-Verlag, Berlin — New York, 1973, pp. 489—501.

15. —, Values of zeta-function at non-negative integers, Lecture Notes in Math., Vol. 1068, Springer-Verlag, Berlin — New York, 1984, pp. 127—138.

16. Меркурьев А. С. О гомоморфизме норменного вычета степени два. — ДАН СССР, 1981, т. 261, с. 542—547.

17. Меркурьев А. С., Суслин А. А. K-когомологии многообразий Севери — Брауэра и гомоморфизм норменного вычета. — Изв. АН СССР, сер. мат., 1982, т. 46, с. 1011—1046.

18. —, О группе K_2 для кватернионных алгебр. — ДАН СССР, 1988, т. 298, № 3, с. 548—550.

19. J. Milnor, Algebraic K-theory, Ann. of Math. Studies, No. 72, Princeton Univ. Press, Princeton, N. J., 1971.

20. —, Algebraic K-theory and quadratic forms, Invent. Math. **9** (1970), 318—344.

21. —, On the homology of Lie groups made discrete, Comment. Math. Helv. **58** (1983), 72—85.

22. D. Quillen, On the cohomology and K-theory of the general linear group over finite fields, Ann. of Math. **96** (1972), 552—586.

23. —, Higher algebraic K-theory of the general linear group over finite fields, Ann. of Math. **96** (1972), 552—586.

24. —, Higher algebraic K-theory, Proc. Internat. Congr. Math. (Vancouver, B. C., 1974), Vol. 1, Canad. Math. Congress, Montreal, Que., 1975, pp. 171—177.

25. Шехтман В. В. Теорема Римана — Роха и спектральная последовательность Атьи — Хирцебруха, УМН, 1980, т. 35, с. 179—180.

26. J.-P. Serre, Groups algêbriques et corps de classes, Hermann, Paris, 1959. [Имеется перевод: Серр Ж.-П. Алгебраические группы и поля классов. — М.: Мир, 1968.]

27. Суслин А. А. Кватернионный гомоморфизм для поля функций на конике. — ДАН СССР, 1982, т. 265, с. 292—296.

28. —, Torsion in K_2 of fields, K-theory **1** (1987), 5—31.

29. —, Homology of GL_n, characteristic classes and Minor K-theory, Lecture Notes in Math., Vol. 1046, Springer-Verlag, Berlin — New York, 1984, pp. 357—375.

30. —, Алгебраическая K-теория и гомоморфизм норменного вычета. — Итоги Науки и Техники, Совр. Пробл. Матем., 1984, т. 25, с. 115—209.

31. —, On the K-theory of algebraically closed fields, Invent. Math. **73** (1983), 241—245.

32. —, On the K-theory of local fields, J. Pure Appl. Algebra **34** (1984), 301—318.

33. —, K_3 от поля и группа Блоха. — Труды МИАН, 1990, т. 183, с. 180—199.

34. —, Divisibility in Bloch's group and Milnor's conjecture on quadratic forms (not to appear).

35. Суслин А. А., Юфряков А. В. K-теория локальных алгебр с делением. — ДАН СССР, 1986, т. 288, с. 832—836.

36. R. Swan, K-theory of quadratic hypersurfaces, Ann. of Math. **122** (1985).

37. J. Tate, On the torsion in K_2 of fields, Proc. Internat. Sympos. Algebr. Number Theory (Kyoto, 1976), Japan Soc. for the Prom. of Science, Tokyo, 1977.

38. R. Thomason, Algebraic K-theory and etale cohomology, Ann. Sci. École Norm. Sup. (4) **18** (1985), 437—552.

39. J.-L. Colliot-Thélène, Hilbert's theorem 90 for K_2, with application to the Chow groups of rational surfaces, Invent. Math. **71** (1983), 1—20.

40. J.-L. Coillot-Thélène, J.-J. Sansuc, and C. Soulé, Torsion dans le groupe de Chow de codimension deux, Duke Math. J. **50** (1983), 763—801.

41. J. Murre, Applications of algebraic K-theory to the theory of algebraic cycles, Lecture Notes in Math., Vol. 1124, Springer-Verlag, Berlin — New York, 1985.

42. J. Jardine, Simplicial objects in a Grothendieck topos.

43. Панин И. А. Теорема Гуревича и K-теория полных дискретно нормированных колец. — Изв. АН СССР, сер. мат., 1986, т. 50, № 4, с. 763—775.

44. H. Gillet, Riemann-Roch theorems for higher algebraic K-theory, Adv. in Math. **40** (1981), 203—289.

45. Меркурьев А. С., Суслин А. А. О гомоморфизме норменного вычета степени три. — Изв. АН СССР, сер. мат., 1990, № 2, с. 339—356.

46. —, Группа K_3 для поля. — Изв. АН СССР, сер. мат., 1990, т. 54, № 3, с. 522—546.

47. M. Rost, Injectivity of $K_2D \to K_2F$ for quaternion algebras, Preprint, Regensburg, May 1986.

48. —, Hilbert 90 for K_3 for degree two extensions, Preprint, Regensburg, May 1986.

СССР 191011, ЛЕНИНГРАД, ЛОМИ

ПРЕДСТАВЛЕНИЯ РЕДУКТИВНЫХ ГРУПП ЛИ

Дейвид А. Воуган-мл. [1]

1. ВВЕДЕНИЕ

Теория представлений строится вокруг двух фундаментальных идей математики — симметрии и линеаризации. К сожалению, только одну из них мы могли бы с гордостью привести в дом и представить родителям. Эти двойственные начала отражаются и в развитии предмета; это история озарений поразительной силы и красоты, запрятанных в лабиринте технических деталей. Я хотел бы дать краткий обзор развития одного из таких озарений (философии коприсоединенных орбит Кириллова — Константа) сквозь призму её приложений к теории представлений редуктивных групп Ли. Технические подробности с их собственной, более трудной для восприятия красотой, я буду в основном опускать.

Итак, начнём. Пусть G — группа. Рассматривать G как группу симметрий означает задать множество X, на котором действует G. То есть каждому элементу g группы G надо поставить в соответствие перестановку $x \to g \cdot x$ множества X так, чтобы произведению двух элементов группы соответствовало произведение перестановок, отвечающих сомножителям. Формально:

Определение 1.1. *Действие* (или *представление перестановками*) группы G — это пара (σ, X), где X — множество, а σ — гомоморфизм G в группу перестановок X. Если это не может привести к недоразумению, мы будем писать $g \cdot x$ вместо $\sigma(g)(x)$. Множество X будем иногда называть *G-пространством.*

Вот другая формулировка той же идеи:

---

[1] David A. Vogan, jr., Representations of reductive Lie groups, ICM86, pp. 245—266.

Определение 1.1′. *Действием* группы G на множестве X называется отображение $G \times X \to X$, $(g, x) \to g \cdot x$, удовлетворяющее следующим условиям:

(a) для всех g и h из G и x из X

$$g \cdot (h \cdot x) = (gh) \cdot x;$$

(b) если e — единичный элемент группы G, то $e \cdot x = x$.

Наиболее важный пример: если H — некоторая подгруппа в G, G действует на множестве G/H левых смежных классов в G по H по правилу $g \cdot (xH) = (gx)H$.

Кое-что интересное можно сказать о действиях даже в столь общей постановке.

Определение 1.2. Предположим, группа G действует на множестве X. *Орбитой* точки x относительно этого действия называется подмножество

$$G \cdot x = \{g \cdot x \mid g \in G\} \subset X.$$

Группой изотропии действия в точке x называется подгруппа

$$G(x) = \{g \in G \mid g \cdot x = x\} \subset G.$$

Говорят, что действие *транзитивно*, если X состоит ровно из одной орбиты. В этом случае мы называем X *однородным пространством* для G.

На пустом множестве любая группа действует единственным способом. Так как число орбит в этом случае равно нулю, действие не является транзитивным.

Лемма 1.3. *Пусть G действует на X. Тогда X есть дизъюнктное объединение всех орбит в X относительно G. Это также — единственное разложение X в объединение однородных пространств для G.*

Эта лемма показывает, что транзитивные действия имеют особое значение. Чтобы все их описать, нам необходимо решить, какие два действия считать эквивалентными.

Определение 1.4. Пусть G действует на двух множествах X и Y. Отображение f из X в Y называется *G-эквивариантным*, если оно *уважает*[1] действие G в том смысле, что $f(g \cdot x) = g \cdot f(x)$. Действия называются *эквивалентными*, если существует G-эквивариантная биекция из X в Y.

---

[1] Обычно говорят «сохраняет», но выбранный переводчиком буквальный перевод английского «respects», по нашему мнению, ярче и удачнее. — *Прим. ред.*

Лемма 1.5. *Пусть G действует на X и $x \in X$. Тогда существует корректно определенное отображение из факторпространства $G/G(x)$ в орбиту $G \cdot x$ (определение 1.2), задаваемое формулой $gG(x) \to g \cdot x$. Это отображение является эквивариантной биекцией из $G/G(x)$ на $G \cdot x$.*

Пусть G действует транзитивно на X и Y; зафиксируем точки x и y в этих множествах. Действия G на X и на Y эквивалентны тогда и только тогда, когда группы изотропии $G(x)$ и $G(y)$ сопряжены как подгруппы в G.

В некотором смысле эти результаты описывают действия групп достаточно полно. Однако было бы заблуждением придавать им слишком большое значение. Недостаточно знать о сфере лишь то, что она эквивалентна $SO(3)/SO(2)$.

В определение 1.1′ можно ввести дополнительную структуру почти любого вида. Например, предположим, G — группа Ли, а X — многообразие. Тогда действие G в X называется *гладким*, если отображение из $G \times X$ в X является гладким. Как правило, интересующие нас вопросы, относящиеся к действию группы, зависят от какой-либо такой дополнительной структуры. Например, если G и X конечны, нас может интересовать мощность X. Если G и X — топологические пространства, можно исследовать топологический тип пространства X. Если G и X — алгебраические многообразия, могут представлять интерес особенности X.

Все эти вопросы и теория действий групп вообще относятся к нелинейному анализу в том наивном смысле, что в них не фигурируют векторные пространства. Поэтому они довольно сложны. Некий намек на это можно было увидеть в лемме 1.5: нахождение всех подгрупп данной группы — чрезвычайно сложная задача, даже для такого вполне безобидного примера, как группа перестановок n символов. Идея (линейной) теории представлений состоит в том, что иногда полезно заменить действие группы на множестве некоторым связанным с ним действием группы на векторном пространстве.

Определение 1.6. Пусть G действует на множестве X. Обозначим через $\mathbb{C}X$ комплексное векторное пространство всех функций на X со значениями в $\mathbb{C}$ и определим действие G в $\mathbb{C}X$ формулой

$$(g \cdot f)(x) = f(g^{-1} \cdot x) \quad (g \in G,\ f \in \mathbb{C}X,\ x \in X).$$

Назовём это действие *регулярным представлением* G в $\mathbb{C}X$. Иногда мы будем записывать операторы этого действия как $\lambda(g)$: $\lambda(g)f = g \cdot f$.

Свойства регулярного представления в функциях на G-пространстве обобщаются в следующем определении:

Определение 1.7. (*Линейным*) *представлением* группы G называется векторное пространство V, в котором G действует линейными преобразованиями. Другими словами, это пара (π, V), где V — векторное пространство, а π — гомоморфизм из G в группу автоморфизмов V. Иногда мы будем говорить, что представление группы G *реализовано* в пространстве V[1].

Нас будут интересовать почти исключительно комплексные представления, т. е. тот случай, когда V является векторным пространством над $\mathbb{C}$. Однако в качестве вспомогательных объектов появятся и некоторые вещественные представления.

Возможна и часто полезна такая переформулировка определения 1.7 в духе определения (1.1)′:

Определение 1.7′. *Представлением* группы G в векторном пространстве V называется отображение $G \times V \to V$, $(g, v) \to g \cdot v$, удовлетворяющее следующим условиям:

(a) для всех g и h из G и v из V

$$g \cdot (h \cdot v) = (gh) \cdot v;$$

(b) если e — единичный элемент G, то $e \cdot v = v$;

(c) если v и w — элементы пространства V, а a и b — скаляры, то

$$g \cdot (av + bw) = a(g \cdot v) + b(g \cdot w).$$

Точно так же как в случае общего действия группы, в это определение могут быть введены дополнительные структуры. Например, если G — топологическая группа и V — топологическое векторное пространство, мы скажем, что представление является *непрерывным*, если отображение из $G \times V$ в V непрерывно.

Аналог понятия транзитивного действия (определение 1.2) для представлений имеет некоторую специфику.

Определение 1.8. Пусть (π, V) — представление группы G. *Инвариантным подпространством* в V называется линейное подпространство W в V, которое сохраняется действием группы:

$$\pi(g) \cdot w \in W \text{ (для всех } g \in G, \ w \in W).$$

Говорят, что представление *неприводимо*, если имеется ровно два инвариантных подпространства.

---

[1] А также что V *несёт* (*на себе*) представление группы G. — *Прим. изд. ред.*

Непрерывное представление называют *неприводимым*, если существуют ровно два замкнутых инвариантных подпространства. (Это не вполне совпадает с нетопологическим определением.)

Подпространства $\{0\}$ и V всегда инвариантны. Неприводимость представления означает, что пространство V ненулевое и других инвариантных подпространств нет.

Наиболее очевидный способ построить инвариантное подпространство состоит в том, чтобы взять какой-нибудь ненулевой вектор v и рассмотреть подпространство $\langle G \cdot v \rangle$, порождённое его орбитой. Однако эти подпространства ведут себя не так хорошо, как орбиты. Дело в том, что другой ненулевой вектор w из $\langle G \cdot v \rangle$ может порождать строго меньшее подпространство. Если V бесконечномерно, этот процесс может продолжаться до бесконечности; минимальных инвариантных подпространств может и не быть. Другой аспект той же проблемы заключается в том, что даже неприводимое представление не характеризуется группой изотропии одного-единственного ненулевого вектора. Для линейных представлений не существует простых аналогов лемм 1.3 и 1.5. Наша цель, тем не менее, — найти такого рода аналоги. Более точно, нас интересует

Абстрактная общая проблема гармонического анализа 1.9.

I. Для разумного представления (π, V) разумной группы G показать, что (π, V) есть (в разумном смысле) «прямая сумма» неприводимых представлений G.

II. Для разумной группы G найти все разумные неприводимые представления G.

Собирая всё вместе (т. е. симметрию и линеаризацию), приходим к следующей постановке вопроса:

Конкретная общая проблема гармонического анализа 1.10. Предположим, имеется некая математическая задача P.

I. Перевести P в задачу GXP о пространстве X, в котором действует G.

II. Перевести GXP в задачу GVP о векторном пространстве V (чего-то вроде) функций на X.

III. Разложить V как представление группы G на неприводимые представления V_i.

IV. Решить задачу для каждого неприводимого представления V_i.

V. Соединить эти решения в решение задачи GVP.

VI. Перевести последнее обратно с языка задачи GVP на язык задачи P.

Блестящий пример задачи, решаемой в соответствии с этой схемой, — это задача о нахождении собственных значений оператора Лапласа на $(n-1)$-мерной сфере. В этом случае в качестве X можно взять сферу S^{n-1}, в качестве G — ортогональную группу $O(n)$ и в качестве V — пространство $L^2(S^{n-1})$. Результат — теория сферических гармоник. Мы хотим привести ещё два примера — один совсем глупый, но простой, другой очень технический, но хорошо иллюстрирующий успешное применение общей схемы в контексте редуктивных групп.

Первая задача состоит в нахождении числа N всех p-элементных подмножеств n-элементного множества S. Пусть X — множество всех таких подмножеств, V — пространство всех функций на X. Симметрическая группа G перестановок S действует в пространстве V. Ясно, что $N = \dim V$. Неприводимые представления G хорошо изучены; они нумеруются разбиениями числа n. Предположим, p не превосходит $[n/2]$ (другой случай рассматривается аналогично). Пространство V разлагается в прямую сумму $p+1$ неприводимых представлений V_i $(i=0, 1, \ldots, p)$. Здесь V_i — неприводимое представление, отвечающее разбиению $(n-i, i)$. Размерности неприводимых представлений известны, и

$$\dim V_i = [n!\,(n-2i+1)]/[i!\,(n-i)!\,(n-i+1)].$$

Индукцией по p легко показать, что

$$(\dim V_0) + \ldots + (\dim V_p) = n!/[p!\,(n-p)!].$$

Такова размерность пространства V, а следовательно, и мощность множества X.

Следующий пример — формула Мацусимы (подробные сведения об этой формуле можно найти в [3] и [14]). Пусть M — связное компактное локально-симметрическое риманово многообразие. («Локально-симметрическое» означает, что для каждой точки p отображение касательного пространства в этой точке, переводящее v в $-v$, порождает с помощью экспоненциального отображения изометрию θ_p некоторой окрестности точки p. Например, римановы поверхности всегда обладают этим свойством.) Рассмотрим задачу нахождения когомологий де Рама многообразия M. В этой задаче нет группы, которая сразу давалась бы в руки: многообразие M может не иметь изометрий. Пусть, однако, $M^{\sim}$ — универсальное накрытие для M. Локальные изометрии θ_p продолжаются на всё $M^{\sim}$; тем самым, возникает большая группа G изометрий $M^{\sim}$. Фундаментальная группа Γ многообразия M, рассматриваемая как группа скольжений $M^{\sim}$, является дискретной подгруппой G. Положим $X = G/\Gamma$. Это — компактное однородное пространство для G. В качестве представления возьмём $V = L^2(X)$. Оказывается, гильбертово пространство V разлагается в прямую сумму счётного числа неприводимых представлений V_i $(i = 1, 2, \ldots)$.

С любым хорошим непрерывным представлением G в W можно связать некоторый набор векторных пространств $\mathrm{H}^j(G, W)$, занумерованный неотрицательными целыми числами j. Они называются *непрерывными когомологиями группы G с коэффициентами в W*. А именно:

$$H^0(G, W) = \{w \in W \mid g \cdot w = w \quad \text{для всех} \quad g \in G\},$$

а следующие H^j определяются как производные функторы. Оказывается, что

$$H^i(M, \mathbb{C}) \cong H^i(G, V) \cong \bigoplus H^i(G, V_i).$$

Это и есть формула Мацусимы. Наконец, группы $\mathrm{H}^j(G, W)$ могут быть явно вычислены для любого неприводимого представления W, содержащегося среди V_i.

Только одна часть этой программы не является эффективной, а именно точное определение представлений V_i. Тем не менее наше точное знание непрерывных когомологий для всех возможных V_i накладывает довольно строгие априорные ограничения на то, какими могут оказаться когомологии M.

Имея в виду такого рода приложения, мы рассмотрим в этой статье следующий вопрос: как могут выглядеть разумные представления редуктивных групп Ли? (Мы не рассчитываем дать на этот вопрос исчерпывающий ответ.) В разд. 2 обсуждается, что означает «разумные представления», и в общих чертах предлагается ответ, подсказанный методом коприсоединенных орбит Кириллова — Костанта.

В разд. 3 показывается, как с помощью разложения Жордана для матриц можно «разделить» метод орбит на три части — гиперболическую, эллиптическую и нильпотентную. Несколько более точно, это разложение наводит на мысль, что должны быть три основные конструкции разумных представлений и что все разумные представления получаются их последовательным применением.

В следующих трёх разделах изучаются эти три основные конструкции. Гиперболический шаг (дающий непрерывные семейства представлений) осуществляется с помощью параболической индукции, разработанной в 50-х годах Гельфандом и Наймарком, а также Хариш-Чандрой. Здесь используется довольно несложный вещественный анализ; пространства представлений являются L^2-пространствами функций.

Эллиптический шаг (дающий дискретные семейства представлений) впервые был глубоко разработан Хариш-Чандрой в 60-х годах в его теории квадратично-интегрируемых представлений. Эти результаты были значительно обобщены за последние десять лет Цуккерманом и другими авторами. Используемые методы — комплексно-аналитические (по крайней мере по духу); представления «живут» в пространствах типа пространств когомологий Дольбо.

Что касается нильпотентного шага, то можно ожидать, что он даст лишь конечное число представлений для каждой группы. Этот шаг ещё систематически не исследован, за исключением простых случаев; общей конструкции представлений также не существует; однако уже можно сделать некоторые предположения о том, какого сорта представления таким образом возникают.

Я бы хотел поблагодарить Берта Костанта, обучившего меня методу коприсоединенных орбит (во всяком случае, позволившего мне черпать из его колодца). Мишель Дюфло объяснил мне большое количество технических моментов.

2. УНИТАРНЫЕ ПРЕДСТАВЛЕНИЯ И КОПРИСОЕДИНЕННЫЕ ОРБИТЫ

Вот класс «разумных» представлений, который мы рассмотрим в связи с проблемой 1.9.

Определение 2.1. Пусть G — топологическая группа. Представление (π, V) группы G называется *унитарным*, если

(a) π непрерывно (см. абзац после определения 1.7′);

(b) V — комплексное гильбертово пространство;

(c) для каждого элемента g группы G оператор $\pi(g)$ унитарен.

Множество классов эквивалентности неприводимых унитарных представлений группы G называется *унитарным сопряжённым* к G и записывается как $\hat{G}$.

Даже и унитарные представления, вообще говоря, не разлагаются в прямую (гильбертову) сумму неприводимых представлений. Примером может служить регулярное представление $\mathbb{R}$ в $L^2(\mathbb{R})$ (определение 1.6). В этом примере преобразование Фурье разлагает пространство в некоторого рода непрерывную прямую сумму неприводимых представлений. Входящие в неё «слагаемые» — это пространства V_t, состоящие из скалярных кратных функции χ_t, где $\chi_t(x) = e^{itx}$. (V_t не является подпространством в V. Однако оно служит представлением $\mathbb{R}$, является пространством функций и сохраняется под действием регулярного представления.) О прямой сумме можно говорить в том смысле, что каждый вектор v из $L^2(\mathbb{R})$ записывается как непрерывная комбинация $v = \int a_t\chi_t\,dt$. Таким образом, основные L^2-свойства преобразования Фурье можно суммировать в виде равенства $V = \int^{\oplus} V_t\,dt$. Будем говорить, что V есть *прямой интеграл* представлений V_t.

Мы оставляем творческому воображению читателя общее определение прямого интеграла (семейства (π_x, V_x) унитарных представлений, параметризованного точками некоторого пространства с мерой X). После того как оно должным образом сформулировано, мы можем записать следующий результат:

Теорема 2.2 (см. [10]). *Пусть π — унитарное представление группы G в сепарабельном гильбертовом пространстве V. Тогда (π, V) эквивалентно некоторому прямому интегралу неприводимых унитарных представлений. Если G является сепарабельной локально-компактной группой типа* I, *то это разложение единственно.*

«Типа I» — это еще один термин, который мы оставляем не определённым. Дискретные группы относятся к типу I, только если они близки к абелевым. Все ваши любимые группы Ли (например редуктивные, которые определяются ниже в разд. 3) принадлежат типу I. По первому пункту проблемы 1.9, мы тем самым нашли разумный класс групп вместе с разумным классом представлений.

Метод коприсоединенных орбит представляет собой попытку описать неприводимые унитарные представления групп Ли в более или менее геометрических терминах. Подробное введение в этот метод можно найти в [7, 9, 5]. Мы же опишем метод в общих чертах, концентрируя внимание на целях, а не на средствах, которыми они достигаются. Для этого потребуется ввести некоторые обозначения.

Пусть G — группа Ли. Будем всегда предполагать, что G имеет конечное число компонент связности. Всюду ниже

$$\begin{aligned} &G_0 \text{ — компонента единицы группы } G,\\ &\mathfrak{g} \text{ — алгебра Ли группы } G, \qquad\qquad (2.3)\\ &\mathfrak{g}^* \text{ — векторное пространство вещественнозначных}\\ &\qquad \text{линейных функционалов на } \mathfrak{g}. \end{aligned}$$

Мы рассмотрим $\mathfrak{g}$ как касательное пространство к G в единице. Тогда $\mathfrak{g}^*$ — кокасательное пространство в единице.

Предположим, G гладко действует на некотором многообразии X и x принадлежит X. Напомним, что $G(x)$ обозначает группу изотропии в точке x (определение 1.2); обозначим через $\mathfrak{g}(x)$ её алгебру Ли. Отображение, определяющее действие, даёт при ограничении гладкое отображение

$$G \times \{x\} \to X, \quad g \to g \cdot x. \tag{2.4a}$$

Дифференциал этого отображения в единице есть отображение из $\mathfrak{g}$ в $T_x(X)$ (касательное пространство к X); его ядром служит в точности $\mathfrak{g}(x)$. Мы получаем:

$$\mathfrak{g}/\mathfrak{g}(x) \hookrightarrow T_x(X). \tag{2.4b}$$

Если X — однородное пространство, то это вложение — изоморфизм.

Зафиксируем теперь g из G и рассмотрим другое ограничение

$$\lambda_g\colon \{g\} \times X \to X, \quad x \to g \cdot x. \tag{2.5a}$$

Дифференциал этого отображения есть изоморфизм из $T_x(X)$ в $T_{g \cdot x}(X)$. В частности, мы получаем *представление изотропии*

$$(\tau_x, T_x(X)) \qquad \text{(2.5b}$$

группы $G(x)$, определённое формулой

$$\tau_x(g) = d\lambda_g(x) \quad (g \in G(x)). \tag{2.5c}$$

Это — конечномерное вещественное представление.

Группа G действует на самой себе сопряжением:

$$g \cdot x = gxg^{-1}. \tag{2.6a}$$

Группа изотропии в единице e группы G совпадает со всей G. Представление изотропии группы G в касательном пространстве $\mathfrak{g}$ к G в единице называется *присоединённым представлением* и обозначается Ad:

$$\mathrm{Ad}(g)\colon \mathfrak{g} \to \mathfrak{g}. \tag{2.6b}$$

Из любого представления π в векторном пространстве V можно построить *контрагредиентное представление* π^* в пространстве V^* линейных функционалов на V. Мы просто рассмотрим π как действие группы, возьмём связанное с ним регулярное представление G в функциях на V и ограничим его на линейные функционалы. Более точно,

$$(\pi^*(g)\xi)(v) = \xi(\pi(g^{-1})v) \tag{2.7}$$

(для ξ из V^*, g из G и v из V). Представление, контрагредиентное к присоединённому представлению, называется *коприсоединенным* и обозначается

$$(\mathrm{Ad}^*, \mathfrak{g}^*). \tag{2.8}$$

Нас коприсоединенное представление будет больше интересовать как действие группы, чем как представление. То есть мы будем рассматривать не инвариантные подпространства в $\mathfrak{g}^*$, а орбиты. Поэтому будем говорить о *коприсоединенном действии* и *коприсоединенных орбитах*.

Теперь мы приступаем, наконец, к изложению метода орбит Кириллова — Костанта.

Философия коприсоединенных орбит (первое приближение) 2.9. Если G — группа Ли, то множество $\widehat{G}$ классов эквивалентности её неприводимых унитарных представлений находится в некотором соответствии с множеством орбит G в $\mathfrak{g}^*$.

Это поразительная идея. Два множества, о которых идёт речь, на первый взгляд совершенно не связаны друг с другом. У опытного математика может возникнуть подозрение, что здесь просто некий трюк и, должным образом понятые, неприводимые унитарные представления и коприсоединенные орбиты окажутся одним и тем же объектом почти по определению. Возможно, такой трюк и существует, однако вот уже более двадцати лет он ускользает от группы специалистов, работающих в теории представлений. Я предпочитаю считать, что здесь кроется настоящее волшебство.

Чтобы лучше понять эту философию, необходимо её уточнить. Для этой цели полезно рассмотреть случай абелевых групп. Заметим, что унитарный оператор в одномерном комплексном гильбертовом пространстве — это просто оператор умножения на скаляр, по модулю равный 1. Обозначим через $\mathbb{T}$ группу комплексных чисел, равных по модулю единице (*группа окружности*). Как мы только что видели, одномерные унитарные представления представляют собой гомоморфизмы в $\mathbb{T}$.

Лемма 2.10. *Пусть H — абелева группа. Тогда любое её неприводимое унитарное представление одномерно. Если H — связная абелева группа Ли, то любой такой гомоморфизм χ определяется своим дифференциалом $d\chi$. Если отождествить* $\mathrm{Lie}(\mathbb{T})$ *с $i\mathbb{R}$, то $d\chi$ будет соответствовать некоторому линейному функционалу if на $\mathfrak{h}$. Это даёт включение*

(a) $\widehat{H} \subset \mathfrak{h}^*$, $\chi \to f$, *определяемое по правилу*

(b) $\chi(\exp X) =$ *умножение на* $e^{if(X)}$.

Обратно, пусть f — произвольный линейный функционал на $\mathfrak{h}$. Тогда f отождествляется с некоторым характером χ в соот-

ветствии с формулой (b), *если и только если* f *переводит ядро экспоненциального отображения в* $2\pi\mathbb{Z}$:

$$\text{(c) } f(X) \in 2\pi\mathbb{Z} \quad \textit{для всех} \quad X \in \ker(\exp).$$

Для абелевой группы H всякий элемент $f \in \mathfrak{h}^*$, удовлетворяющий условию (c), называют *целочисленным*. В случае когда H односвязна, это условие тривиально и выполняется для всех f. Если H — тор, то ядро экспоненциального отображения представляет собой некоторую решётку в $\mathfrak{h}$; целочисленные веса образуют в 2π раз растянутую дуальную решётку в $\mathfrak{h}^*$.

Лемма 2.10 наводит на мысль, что философия коприсоединенных орбит должна иметь своим объектом лишь орбиты, удовлетворяющие некоторого рода условию целочисленности. Вот соответствующая общая формулировка:

Определение 2.11. Пусть G — группа Ли и $f \in \mathfrak{g}^*$ (см. (2.3)). Напомним, что группа изотропии присоединённого действия обозначается через $G(f)$ (определение 1.2), а её компонента единицы — через $G(f)_0$. Мы скажем, что элемент f — *целочисленный*, если выполнено какое-нибудь из следующих двух эквивалентных условий:

(a) существует унитарный характер $\pi_0(f)$ группы $G(f)_0$ с дифференциалом if;

(b) существует конечномерное неприводимое унитарное представление $\pi(f)$ группы $G(f)$ с дифференциалом

$$[d\pi(f)](X) = if(X) \cdot \mathrm{Id} \quad (X \in \mathfrak{g}).$$

Будем называть пару $(f, \pi(f))$ *целочисленными данными* для G.

Теперь мы можем сформулировать улучшенную версию метода орбит:

Философия коприсоединенных орбит (второе приближение) 2.12. Если G — группа Ли, $\widehat{G}$ находится во взаимно-однозначном соответствии с множеством целочисленных данных для G (определение 2.11). В частности, каждой целочисленной коприсоединенной орбите соответствует конечное число элементов $\widehat{G}$.

Лемма 2.10 утверждает, что это справедливо, если G — связная абелева группа. Нетрудно проверить, что это верно также и в том случае, когда компонента единицы в G является абелевой. В работе [6] это утверждение доказано для односвязных нильпотентных групп G; отсюда следует, что оно справедливо и для случая, когда компонента единицы в G нильпотентна.

В качестве следующего класса групп естественно рассмотреть разрешимые группы. Существуют разрешимые группы Ли, не принадлежащие типу I (см. теорему 2.2). Они не удовлетворяют философии 2.12. Однако ничего страшного в этом нет, так как неприводимые унитарные представления не являются основным инструментом гармонического анализа в случае групп не типа I. Ауслендер и Костант показали в [2], что философия 2.12 применима к односвязным разрешимым группам типа I. Так же как и в нильпотентном случае, хотелось бы опустить предположение об односвязности группы G. Однако для этого есть формальные препятствия. Хотя условия целочисленности в определении 2.11 очень приятные, теперь они оказываются не вполне правильными. Необходимые уточнения довольно тонкие, но с ними связаны некоторые важные идеи; поэтому мы их кратко обрисуем.

Пусть V — конечномерное вещественное векторное пространство, и пусть ω — невырожденная симплектическая форма на V, т. е. кососимметрическая билинейная форма на V с нулевым ядром. Иначе говоря, можно рассматривать ω как 2-форму на V. (Невырожденность в этой интерпретации означает, что для любого ненулевого вектора v из V существует вектор w, такой что $\omega(v \wedge w)$ не равно нулю.) *Симплектическая группа* пространства V определяется как

$$\mathrm{Sp}(V) = \{g \in \mathrm{GL}(V) \mid \omega(gv, gw) = \omega(v, w) \text{ для всех } v, w\}. \quad (2.13\text{a})$$

(Мы будем также писать $\mathrm{Sp}(\omega)$ или $\mathrm{Sp}(V, \omega)$.) Её алгеброй Ли будет

$$\mathfrak{sp}(V) = \{X \in \mathfrak{gl}(V) \mid \omega(Xv, w) + \omega(v, Xw) = 0\}. \quad (2.13\text{b})$$

Мы хотим определить некоторое двулистное накрытие $\mathrm{Mp}(V)$ для $\mathrm{Sp}(V)$. Чтобы сделать это должным образом, нам понадобилось бы слишком много времени. Однако описать это накрытие нетрудно. Если пространство V нулевое, то группа $\mathrm{Sp}(V)$ тривиальна, и мы полагаем

$$\mathrm{Mp}(0) = \mathbb{Z}/2\mathbb{Z}. \quad (2.14)$$

Если V ненулевое, то определим $\mathrm{Mp}(V)$ как единственное (связное) двулистное накрытие $\mathrm{Sp}(V)$. В любом случае мы получаем короткую точную последовательность

$$\{1, \varepsilon\} \to \mathrm{Mp}(V) \to \mathrm{Sp}(V). \quad (2.15)$$

$\mathrm{Mp}(V)$ называется *метаплектической группой*.

Пусть M — многообразие. *Симплектическая структура* на M — это замкнутая 2-форма ω на M, которая невырожденна как билинейная форма на каждом касательном пространстве $T_m(M)$. Если задана такая структура, то говорят, что M — *симплектическое многообразие.*

Теорема 2.16. *Пусть G — группа Ли, F — коприсоединенная орбита. Тогда F обладает естественной G-инвариантной симплектической структурой ω_F, определённой следующим образом. Зафиксируем $f \in F$ и обозначим через ω_f симплектическую форму на касательном пространстве $T_f(F)$. Отождествим это касательное пространство с $\mathfrak{g}/\mathfrak{g}(f)$ (ср. (2.4b)). Для данных касательных векторов x и y выберем представителей X и Y в $\mathfrak{g}$. Тогда $\omega_f(x, y) = f([X, Y])$. Представление изотропии τ_f группы $G(f)$ (ср. (2.5)) сохраняет ω_f.*

Тот факт, что указанная формула определяет симплектическую форму на $T_f(F)$, почти тривиален. То что ω_F — замкнутая 2-форма на F, доказывается лишь ненамного сложнее; это следует из тождества Якоби в $\mathfrak{g}$.

Определение 2.17. Пусть G — группа Ли и $f \in \mathfrak{g}^*$. По лемме 2.16 существует гомоморфизм τ_f: $G(f) \to \mathrm{Sp}(\omega_f)$. *Метаплектической накрывающей* группы $G(f)$ называется прообраз метаплектической накрывающей группы $\mathrm{Sp}(\omega_f)$ относительно τ_f (см. (2.15)). Она обозначается $G(f)^{\mathrm{mp}}$. Имеем точную последовательность

$$1 \to \{1, \varepsilon\} \to G(f)^{\mathrm{mp}} \to G(f) \to 1.$$

Представление $\pi(p)^{\mathrm{mp}}$ группы $G(f)^{\mathrm{mp}}$ называется *подлинным*, если

(1) нетривиальный элемент ε ядра накрывающего отображения действует в $\pi(f)^{\mathrm{mp}}$ умножением на -1,

и *допустимым*, если оно подлинно и его дифференциал удовлетворяет условию

$$d\pi(f)^{\mathrm{mp}}(X) = if(X) \cdot \mathrm{Id}. \qquad (2)$$

В этом случае пару $(f, \pi(f)^{\mathrm{mp}})$ называют *допустимыми данными* для G. Элемент f, а также орбита $G \cdot f$ называются *допустимыми*, если существует допустимое представление группы $G(f)^{\mathrm{mp}}$.

Используя это определение, Дюфло показал, что все вопросы, касающиеся накрывающих групп, должным образом связываются воедино; неприводимые унитарные представления разрешимых групп Ли типа I параметризуются классами сопряжённости допустимых данных относительно G. Это позволяет получить для метода орбит ещё одно приближение, которое читатель легко сформулирует сам.

Когда эта философия применяется к полупростым группам (определение см. в разд. 3 ниже), некоторые утверждения перестают быть верными. Первая проблема возникает, когда G есть $\mathrm{SO}(3)$, группа вращений трёхмерного пространства. В этом случае тривиальному представлению G соответствуют целых две орбиты, одна из которых — это $\{0\}$, а вторая — наименьшая ненулевая допустимая орбита. Теперь уже нельзя рассчитывать получить биекцию между представлениями и допустимыми данными.

Более фундаментальная проблема возникает, когда G есть $\mathrm{SL}(2, \mathbb{R})$. У неё существует семейство неприводимых унитарных представлений, называемое *дополнительной серией*, которая параметризуется открытым единичным интервалом $(0, 1)$. За исключением значения параметра, равного 1/2, эти представления не соответствуют никакой коприсоединенной орбите. Тем самым соответствие между орбитами и представлениями не является сюръективным.

Кроме этого, для $\mathrm{SL}(2, \mathbb{R})$ существуют представления, отвечающие объединению двух орбит, но не отвечающие никакой отдельной орбите. (Это сферическая основная серия с нулевым нормированным параметром.) Орбиты могут оказаться незамкнутыми, и их замыкания могут иметь непустое пересечение.

Для более сложных полупростых групп возникают и другие мелкие проблемы того же плана: представления, отвечающие допустимым данным, могут оказаться приводимыми или нулевыми, либо их вообще может не существовать. Однако философии не так чувствительны к контрпримерам, как теоремы или даже гипотезы;

Философия коприсоединенных орбит (третье приближение) 2.18. Пусть G — группа Ли типа I. Существуют унитарные представления G, отвечающие конечным наборам допустимых данных для G (определение 2.17) и некоторым граничным условиям на замыканиях соответствующих орбит. Все хорошие неприводимые унитарные представления возникают таким способом.

В оставшейся части этой статьи речь будет идти о реализации одной части этой философии (сопоставление гильбертовых пространств и операторов многообразиям и групповым действиям). Мы полностью игнорируем две увлекательные задачи, связанные со второй ее частью: дать априорное определение «хорошего» представления и указать орбиты, соответствующие (хорошим) представлениям.

3. КОПРИСОЕДИНЕННЫЕ ОРБИТЫ ДЛЯ РЕДУКТИВНЫХ ГРУПП

Первое что надо знать о редуктивных группах (ещё до определения редуктивности), это что такой группой является $\mathrm{GL}(n, \mathbb{R})$ — группа всех обратимых вещественных $n \times n$-матриц. Её алгеброй Ли служит

$$\mathfrak{gl}(n, \mathbb{R}) = \text{все } n \times n \text{ вещественные матрицы.} \tag{3.1a}$$

Присоединённое действие задаётся сопряжением матриц:

$$\mathrm{Ad}(g)(X) = gXg^{-1}. \tag{3.1b}$$

На алгебре Ли $\mathfrak{gl}(n, \mathbb{R})$ имеется невырожденная симметрическая билинейная форма, называемая *формой следа* и обозначаемая $\langle\ ,\ \rangle$:

$$\langle X, Y\rangle = \mathrm{tr}\, XY. \tag{3.1c}$$

Она сохраняется присоединённым действием. Форма следа позволяет отождествить $(\mathfrak{gl}(n, \mathbb{R}))^*$ с пространством всех $n \times n$-матриц. При этом отождествлении линейному функционалу f отвечает матрица $X(f)$, определяемая формулой:

$$f(Y) = \langle X(f), Y\rangle. \tag{3.1d}$$

Так как форма следа $\langle\ ,\ \rangle$ инвариантна относительно Ad, это отождествление переводит коприсоединенные орбиты в присоединенные. То есть:

> коприсоединенные орбиты для $\mathrm{GL}(n, \mathbb{R})$ находятся во взаимно-однозначном соответствии с классами сопряжённости вещественных $n \times n$-матриц. (3.2)

Напомним теперь некоторые элементарные сведения о классах сопряжённости для матриц.

Определение 3.3. Пусть X — вещественная матрица размера $n \times n$. Говорят, что X *нильпотентна*, если $X^k = 0$ для некоторого k, или, эквивалентно, если все (комплексные) собственные значения матрицы X равны нулю. Матрицу X называют *полупростой*, если она диагонализуема над $\mathbb{C}$. Полупростая матрица X называется *эллиптической*, если все её собственные значения чисто мнимые, и *гиперболической*, если все её собственные значения вещественны; последнее эквивалентно диагонализуемости матрицы над $\mathbb{R}$.

Предложение 3.4 (разложение Жордана). *Пусть X — вещественная $n \times n$-матрица. Существуют единственные матрицы X_h, X_e* и *X_n со следующими свойствами:*

(a) $X = X_h + X_e + X_n$;

(b) X_h — *гиперболическая,* X_e — *эллиптическая,* X_n — *нильпотентная;*

(c) *X_h, X_e и X_n попарно коммутируют между собой.*

Они обладают также следующим дополнительным свойством:

(d) *любая матрица, коммутирующая с X, коммутирует также с X_h, X_e и X_n.*

В обычном разложении Жордана на полупростую и нильпотентную части фигурирует полупростая часть X_s матрицы X:

$$X_s = X_h + X_e. \tag{3.5}$$

Инволюцией Картана для группы $\mathrm{GL}(n, \mathbb{R})$ называется её автоморфизм θ, определённый по правилу:

$$\theta g = {}^t g^{-1} \tag{3.6a}$$

для g из $\mathrm{GL}(n, \mathbb{R})$. Дифференциал θ, также обозначаемый через θ, есть автоморфизм алгебры $\mathfrak{gl}(n, \mathbb{R})$, действующий по правилу

$$\theta X = -{}^t X. \tag{3.6b}$$

(Поскольку и группа и алгебра состоят из матриц, эта система обозначений несогласованна!) Для этого автоморфизма θ алгебра Ли кососимметричных матриц служит собственным подпространством с собственным значением $+1$. Она состоит из эллиптических элементов, и форма следа на ней отрицательно-определённа. Симметричные матрицы образуют собственное подпространство для θ с собственным значением -1; они гиперболические, и форма следа на этом подпространстве положительно-определённа.

В этом месте удобно ввести произвольные редуктивные группы. Используемое нами определение заимствовано из [8].

Определение 3.7. Группа Ли G (имеющая конечное число компонент) называется *редуктивной*, если существует гомоморфизм η: $G \to \mathrm{GL}(n, \mathbb{R})$ со следующими свойствами:

(1) ядро η конечно;

(2) образ η является θ-инвариантным.

Группу G называют *полупростой*, если она редуктивна и центр группы G_0 **конечен.**

Обозначим той же буквой θ единственное поднятие θ на G, которое тривиально на ядре η, и назовём его *инволюцией Картана* для G. Дифференциал гомоморфизма η позволяет отождествить $\mathfrak{g}$ с некоторой матричной алгеброй Ли. Элементы $\mathfrak{g}$ называются *полупростыми*, *нильпотентными*, *гиперболическими* или *эллиптическими*, если таковы соответствующие матрицы. Введем следующие обозначения:

K — неподвижные точки автоморфизма θ в G,
$\mathfrak{k}$ — $\mathrm{Lie}(K)$-неподвижные точки θ в $\mathfrak{g}$,
$\mathfrak{s}$ — собственное подпространство для θ в $\mathfrak{g}$,
отвечающее собственному значению -1.

Мы получаем *разложение Картана* $\mathfrak{g}=\mathfrak{k}+\mathfrak{s}$. В соответствии с замечаниями, сделанными после формул (3.6), форма следа $\langle\ ,\ \rangle$ положительно-определённа на $\mathfrak{s}$ и отрицательно-определённа на $\mathfrak{k}$. Ввиду разложения Картана эта форма, тем самым, невырожденна. Используя это, как и в (3.1d), отождествим $\mathfrak{g}^*$ с $\mathfrak{g}$: линейный функционал f на $\mathfrak{g}$ соответствует элементу $X(f)$, удовлетворяющему условию $f(Y)=\langle X(f), Y\rangle$ для всех Y из $\mathfrak{g}$.

Приведём теперь некоторые важные структурные факты. Для случая $\mathrm{GL}(n, \mathbb{R})$ они сводятся к предложению 3.4 и тому факту, что любая вещественная эллиптическая матрица сопряжена с кососимметрической.

Предложение 3.8. *Пусть* G — *вещественная редуктивная группа и* X — *элемент из* $\mathfrak{g}$.

(a) *Компоненты* X_h, X_e *и* X_n *разложения Жордана для* X *все лежат в* $\mathfrak{g}$.

(b) *Если элемент* X — *гиперболический, то он сопряжён относительно* $\mathrm{Ad}(G)$ *с некоторым элементом из* $\mathfrak{s}$.

(c) *Если элемент* X — *эллиптический, то он сопряжён относительно* $\mathrm{Ad}(G)$ *с некоторым элементом из* $\mathfrak{k}$.

Определение 3.9. Пусть G — редуктивная группа и $f\in\mathfrak{g}^*$. Обозначим через $X(f)$ соответствующий элемент из $\mathfrak{g}$ (определение 3.7), и пусть

$$X(f)=X(f)_h+X(f)_e+X(f)_n$$

— его разложение Жордана (предложение 3.4). Соответствующие линейные функционалы на $\mathfrak{g}$ обозначим через f_h, f_e и f_n. Тогда $f=f_h+f_e+f_n$ — *разложение Жордана* для f. Назовём f *гиперболическим*, *нильпотентным* и т. д., если таков элемент $X(f)$. *Полупростая часть* f — это $f_s=f_h+f_e$.

Покажем теперь, как можно воспользоваться разложением Жордана для решения проблемы построения представлений по орбитам. Пусть нам задан функционал f. Мы будем постоянно использовать тот факт, что

$$G(f)=\text{централизатор элемента } X(f) \text{ в } G, \tag{3.10a}$$

$$\mathfrak{g}(f)=\{Y\in\mathfrak{g}\mid[X(f), Y]=0\}. \tag{3.10b}$$

Предложение 3.8 позволяет нам заменить f на сопряжённый к нему и получить

$$\theta f_h=-f_h, \quad \theta f_e=f_e. \tag{3.11a}$$

Отсюда следует, что группы изотропии

$$G(f_h), \quad G(f_e) \quad \text{и} \quad G(f_s)=G(f_h)\cap G(f_e) \tag{3.11b}$$

сохраняются при действии θ, а значит, все они редуктивны (это устанавливается с помощью ограничений отображения η, используемого для самой группы G). Элементы X_e и X_n коммутируют с X_h и, следовательно, принадлежат $\mathfrak{g}(f_h)$. Поэтому мы можем отождествить f_e и f_n (посредством ограничения) с элементами из $\mathfrak{g}(f_h)^*$. В результате получаем цепочку

$$G(f_h)\supset[G(f_h)](f_e)\supset[[G(f_h)](f_e)](f_n). \tag{3.11c}$$

Это те же самые группы, что и в цепочке

$$G(f_h)\supset G(f_s)\supset G(f). \tag{3.11c}$$

Данные, о которых шла речь выше, определяют (более или менее) представление группы $G(f)$. Из этого мы предполагаем получить представление G в три шага. Вначале мы получим представление группы $G(f_s)$ (нильпотентный шаг), затем представление группы $G(f_h)$ (эллиптический шаг) и, наконец, представление группы G (гиперболический шаг). Эта терминология объясняется тем, что, например, группа $\mathrm{G}(f_s)$ (с которой мы начинаем на эллиптическом шаге) является группой изотропии для эллиптического элемента f_e в группе $G(f_n)$ (к которой мы переходим).

Как было отмечено во введении, порядок этих шагов в точности противоположен историческому, и к выполнению первого шага фактически еще только предстоит приступить. Однако, отважно преодолевая сомнения, что на пустом месте что-то может возникнуть (и отгоняя мысли о бездарно потраченных детстве и юности), мы все-таки исследуем последние два шага в последующих двух разделах. В заключительном разделе мы обсудим перспективы для первого шага.

4. ПАРАБОЛИЧЕСКАЯ ИНДУКЦИЯ И ГИПЕРБОЛИЧЕСКИЙ ШАГ

Все что мы знаем до сих пор о коприсоединенных орбитах — это что они являются однородными симплектическими пространствами. (Для полупростых групп это и все, что надо знать: Кириллов, Костант и Сурьо независимо показали, что любое такое пространство является конечнолистным накрытием некоторой орбиты.) В соответствии с методом орбит требуется построить представление по орбите $X = G\cdot f$. Единственное что сразу приходит в голову — это регулярное представление на X (определение 1.6) или что-то тесно с ним связанное. Небольшой опыт работы с этим представлением показывает, что оно слишком велико — например, оно почти всегда приводимо.

Проведя более внимательное изучение ситуации, мы могли бы предложить построить расслоение на X, отвечающее допустимым данным $(f, \pi(f)^{\mathrm{mp}})$. Трудность здесь состоит в том, что $\pi(f)^{\mathrm{mp}}$ должно быть сперва преобразовано в представление группы $G(f)$ (а не только $G(f)^{\mathrm{mp}}$). Для этого, по-видимому, требуется нечто вроде тензорного произведения этого представления на метаплектическое представление группы $\mathrm{Mp}(\omega_f)$. У нас получается бесконечномерное расслоение над пространством, которое и так было слишком большим с самого начала. Это выглядит достаточно безумно и потому обнадёживает, однако в этом направлении не было достигнуто успеха.

Если симплектической структуры недостаточно, чтобы построить представление, то разумно спросить себя, какие дополнительные структуры могут пригодиться. Полезно этот вопрос поставить так: какие более сложные объекты могут иметь естественную структуру симплектического многообразия?

Первый ответ — это кокасательные расслоения. Пусть Y — многообразие. Тогда T^*Y обладает естественной симплектической структурой. Симплектическому многообразию T^*Y можно сопоставить унитарное представление группы G в $L^2(Y)$. Более точно, следует рассмотреть представление в квадратично-интегрируемых сечениях расслоения полуплотностей на Y. Это имеет смысл, даже когда Y не обладает инвариантной мерой.

Более общо, пусть $\mathscr{L}$ — эрмитово линейное расслоение над Y. Как показал Р. Урвин, существует *расслоение связностей* $\mathscr{C}_{\mathscr{L}}$ над Y, такое что самосопряжённые связности на $\mathscr{L}$ — это в точности сечения $\mathscr{C}_{\mathscr{L}}$. Пространство $\mathscr{C}_{\mathscr{L}}$ имеет симплектическую структуру, и если $\mathscr{L}$ — однородное линейное расслоение, то G действует на $\mathscr{C}_{\mathscr{L}}$. Симплектическому G-пространству $\mathscr{C}_{\mathscr{L}}$ мы сопоставим представление G в квадратично-интегрируемых сечениях расслоения $\mathscr{L}$ (подкрученных на полуплотности).

Первый способ построения представления по орбите X состоит поэтому в том, чтобы попытаться реализовать X как (грубо говоря) кокасательное расслоение на некотором однородном пространстве Y или (точнее) как расслоение связностей для однородного линейного расслоения над Y.

В случае редуктивных групп мы можем применить этот подход к гиперболическим элементам. Зафиксируем гиперболический функционал f_{h} в $\mathfrak{g}^*$ и обозначим через X_{h} соответствующий элемент алгебры Ли (определение 3.7). Запишем разложение на собственные подпространства

$$\mathfrak{g} = \sum_{r \in \mathrm{R}} \mathfrak{g}^r. \qquad (4.1\mathrm{a})$$

Здесь

$$\mathfrak{g}^r = \{Y \in \mathfrak{g} \mid [X_{\mathrm{h}}, Y] = rY\}. \qquad (4.1\mathrm{b})$$

В соответствии с (3.10), $G(f_{\mathrm{h}})$ сохраняет это разложение, и

$$\mathfrak{g}^0 = \mathfrak{g}(f_{\mathrm{h}}). \qquad (4.1\mathrm{c})$$

В силу тождества Якоби,

$$[\mathfrak{g}^r, \mathfrak{g}^s] \subset \mathfrak{g}^{r+s}. \qquad (4.1\mathrm{d})$$

Из инвариантности формы следа $\langle\ ,\ \rangle$ относительно $\mathrm{ad}(X_{\mathrm{h}})$ вытекает, что

$$\langle \mathfrak{g}^r, \mathfrak{g}^s \rangle = 0, \quad \text{если} \quad r + s \neq 0. \qquad (4.1\mathrm{e})$$

Положим

$$\mathfrak{n}_{\mathrm{h}} = \sum_{r > 0} \mathfrak{g}^r. \qquad (4.2\mathrm{a})$$

Согласно (4.1), $\mathfrak{n}_{\mathrm{h}}$ является нильпотентной подалгеброй в $\mathfrak{g}$, сохраняющейся при действии $G(f_{\mathrm{h}})$. Положим

$$N_{\mathrm{h}} = \exp(\mathfrak{n}_{\mathrm{h}}) \qquad (4.2\mathrm{b})$$

$$P_{\mathrm{h}} = G(f_{\mathrm{h}})\, N_{\mathrm{h}}. \qquad (4.2\mathrm{c})$$

Группа P_{h} есть то, что называют *параболической подгруппой* в G. Мы пытаемся сделать пространство $G/G(f_{\mathrm{h}})$ похожим на что-то вроде расслоения над меньшим однородным пространством. Этим меньшим пространством будет G/P_{h}.

Вот схематическое описание гиперболического шага метода орбит. Напомним, что предыдущие шаги (которые еще будут обсуждаться) должны были дать нам (грубо говоря) унитарное представление π_{h} группы $G(f_{\mathrm{h}})$. Расширим это представление на всю группу P_{h}, полагая его тривиальным на N_{h}. Образуем индуцированное эрмитово расслоение $\mathscr{V}_{\mathrm{h}}$ над G/P_{h}. Представление группы G, которое мы хотим получить, реализуется в пространстве глобальных квадратично-интегрируемых сечений $\mathscr{V}_h$. Более точно, следует взять сечения $\mathscr{V}_h$, подкрученные расслоением полуплотностей над G/P_h.

Специалисты заметят, что это может быть доведено до частичной теоретико-орбитальной классификации унитарных представлений в духе теоремы Дюфло [4] для произвольных алгебраических групп Ли. Данные для такой частичной классификации — это пара (f, π^{mp}) (с точностью до сопряжения относительно G). Здесь f — гиперболический элемент, π^{mp} — неприводимое унитарное подлинное представление $G(f)^{mp}$ (определение 2.17) и мнимая часть инфинитезимального характера π^{mp} совпадает с $i \cdot f$.

5. КОГОМОЛОГИЧЕСКАЯ ИНДУКЦИЯ И ЭЛЛИПТИЧЕСКИЙ ШАГ

Сделав всё что можно с кокасательными расслоениями, мы снова спрашиваем себя: какие еще объекты являются симплектическими многообразиями? Следующий ответ, который мы рассмотрим, — кэлеровы многообразия. Это многообразия, которые имеют согласованные между собой комплексную и симплектическую структуры.

Более точно, пусть V — вещественное векторное пространство. Напомним, что *комплексной структурой* на V называется отображение

$$J\colon V \to V, \tag{5.1a}$$

такое что

$$J^2 = -\operatorname{Id}. \tag{5.1b}$$

Вот ещё одна формулировка. Положим

$$V_{\mathbb{C}} = V \otimes_{\mathbb{R}} \mathbb{C} = \{v + iw \mid v,\ w \in V\}. \tag{5.2a}$$

Комплексное векторное пространство $V_{\mathbb{C}}$ называется *комплексификацией* V. *Комплексное сопряжение* на $V_{\mathbb{C}}$ — это **линейный** автоморфизм σ, действующий по правилу

$$\sigma(v + iw) = v - iw. \tag{5.2b}$$

Задание комплексной структуры на V эквивалентно заданию комплексного подпространства

$$V^{0,1} \subset V_{\mathbb{C}}, \tag{5.3a}$$

удовлетворяющего условию

$$V_{\mathbb{C}} = \sigma(V^{0,1}) \oplus V^{0,1} = V^{1,0} \oplus V^{0,1}. \tag{5.3b}$$

Назовём $V^{0,1}$ *антиголоморфным* подпространством. (Эквивалентность этого определения приведённому выше устанавливается сопоставлением $V^{0,1}$ собственному подпространству отображения J на $V_{\mathbb{C}}$ с собственным значением $-i$.)

Определение 5.4. Задать *кэлерову структуру* на вещественном векторном пространстве V означает задать на нём

(1) комплексную структуру J и

(2) симплектическую структуру ω

(см. (5.1) и (2.13)). При этом требуется, чтобы они удовлетворяли условию

(a) $\omega(Jv, w) = -\omega(v, Jw)$.

Эквивалентное условие:

(a)′ $\omega(Jv, Jw) = \omega(v, w)$.

Если комплексная структура задаётся подпространством $V^{0,1}$, то это требование равносильно тому, что

(a)″ $\omega_{\mathbb{C}}(v, w) = 0$ для всех v, $w \in V^{0,1}$. Здесь $\omega_{\mathbb{C}}$ обозначает комплексно-линейное продолжение ω на $V_{\mathbb{C}}$.

Эквивалентная формулировка: кэлерова структура на V определяется заданием

(1) комплексной структуры J и

(2) невырожденной эрмитовой формы на (комплексном) векторном пространстве V.

(Напомним, что эрмитовой формой называется комплекснозначная полуторалинейная форма на V, удовлетворяющая условию $\langle v, w\rangle = \overline{\langle w, v\rangle}$.) Эти два определения связаны формулами

$$\omega(v, w) = \operatorname{Im}\langle v, w\rangle, \quad \langle v, w\rangle = \omega(Jv, w) + i\omega(v, w).$$

Будем называть $\langle\ ,\ \rangle$ *кэлеровой формой* на V. Её сигнатура (p, q) называется *сигнатурой* кэлеровой структуры.

Наконец, мы можем под кэлеровой структурой понимать структуру, определяемую заданием

(1) невырожденной симметричной билинейной формы B на V и

(2) невырожденной симплектической формы ω на V.

Эти две формы должны быть связаны условием: если J — автоморфизм V, действующий по правилу

(a) $B(v, w) = \omega(Jv, w)$, то $J^2 = -\operatorname{Id}$.

Определение 5.5. Пусть M — многообразие. *Кэлеровой структурой* на M называется пара, состоящая из комплексной и симплектической структур, которые определяют кэлерову структуру на каждом касательном пространстве T_mM.

Итак, кэлерово многообразие является одновременно комплексным, симплектическим и римановым (возможно, с неопределенной индефинитной метрикой), и любые две из этих структур определяют третью.

Напомним, что в свете метода орбит для нас важно найти уменьшенный вариант пространства функций (или сечений линейного расслоения) на M. Один из естественных выборов такого пространства — голоморфные функции. Точно так же как функции на заданном пространстве Y имеют (в сущности) половину степеней свободы по сравнению с функциями на T^*Y, так же и голоморфные функции на комплексном многообразии имеют половину степеней свободы по сравнению с гладкими.

Здесь мы, однако, ступаем на тонкий аналитический лёд. Отсутствие функций-«шапочек» с малым носителем среди голоморфных функций приводит к глобальным ограничениям, не имеющим аналога в вещественном анализе. Чтобы лучше понять проблему, рассмотрим пример.

Пример 5.6. В качестве G возьмём $\mathrm{GL}(2n, \mathbb{R})$. Напомним, что $\mathfrak{g}^*$ состоит из $2n \times 2n$-матриц, т. е. линейных преобразований $\mathbb{R}^{2n}$. Зафиксируем отождествление $\mathbb{R}^{2n}$ в $\mathbb{C}^n$, и пусть

(a) f — матрица умножения на i.

Если отождествление сделано надлежащим образом,

(b) $f = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$.

Этот элемент эллиптический. Минутное размышление показывает, что его централизатор выглядит как

(c) $G(f) \cong GL(n, \mathbb{C})$.

Орбита $F = G \cdot f$ состоит из всех матриц f' с квадратом -1, т. е. из всех комплексных структур на $\mathbb{R}^{2n}$.

Мы утверждаем, что F обладает кэлеровой структурой, или, что то же самое, обладает G-инвариантной комплексной структурой, порождающей (вместе с ω_f) кэлерову структуру на $\mathfrak{g}/\mathfrak{g}(f)$. Чтобы убедиться в этом, используем вторую интерпретацию комплексной структуры — как подпространства в $(\mathbb{R}^{2n})_{\mathbb{C}} = \mathbb{C}^{2n}$. Тогда пространство F отождествляется с некоторым подмножеством (компактного комплексного) грассманова многообразия $\mathrm{Gr}(n, 2n)$ всех n-мерных подпространств в $\mathbb{C}^{2n}$. Из условия на подпространство, налагаемого соотношением (5.3b), следует, что F открыто в $\mathrm{Gr}(n, 2n)$. Это даёт инвариантную комплексную структуру; мы опускаем проверку того, что F кэлерово. Комплексная размерность F равна n^2, а сигнатура кэлеровой формы равна $((n^2 - n)/2, (n^2 + n)/2)$.

Однородные эрмитовы линейные расслоения на F параметризуются унитарными характерами группы изотропии $\mathrm{GL}(n, \mathbb{C})$. Они в свою очередь параметризуются $\mathbb{Z}$ в соответствии с формулой

$$\chi_m(g) = (\det(g)/|\det(g)|)^m.$$

Обозначим через $\mathscr{L}_m$ расслоение, отвечающее χ_m. Оказывается, пространство голоморфных сечений $\mathscr{L}_m$ всегда конечномерно; его размерность положительна в точности тогда, когда m не превосходит нуля. Возникающие представления G в голоморфных сечениях совпадают с картановскими степенями фундаментального представления, отвечающего среднему простому корню; они никогда не являются унитарными, за исключением тривиального представления (с $m = 0$).

Часть трудностей связана с тем, что у нашего комплексного многообразия F имеется большое компактное подмногообразие. Обозначим через B стандартное скалярное произведение в $\mathbb{R}^{2n}$, продолженное по комплексной линейности на $\mathbb{C}^{2n}$. Рассмотрим следующее подпространство в $\mathrm{Gr}(n, 2n)$:

(d) $F_K = \{V \mid B(V, V) = 0\}$.

Очевидно, это — алгебраическое подмногообразие. Из того факта, что B положительно-определённо на $\mathbb{R}^{2n}$, вытекает, что F_K содержится в F. В действительности это орбита f относительно ортогональной группы:

(e) $F_K = 0(2n) \cdot f \cong 0(2n)/\mathrm{U}(n)$.

Поэтому F_K является компактным комплексным подмногообразием в F комплексной размерности $(n^2 - n)/2$.

Имеется ряд соображений в пользу того, что здесь вместо голоморфных функций следует рассмотреть высшие когомологии Дольбо F с коэффициентами в $\mathscr{L}_m$:

(f) $V^p(m) = H^{0,p}(F, \mathscr{L}_m)$.

(Это было отмечено в чуть различных контекстах Костантом и Ленглендсом около двадцати лет назад.) Во всяком случае эти пространства несут на себе представления группы G.

Чтобы исследовать природу этих представлений, можно действовать так. Положим

(g) $(V^p(m))_K = H^{0,p}(F_K, \mathscr{L}_m)$.

Это конечномерное пространство несёт на себе представление, которое можно явно описать с помощью теоремы Ботта—Бореля—Вейля. Мы получим:

(h) $(V^p(m))_K \neq 0$, если и только если $p=0$ и $m \leqslant 0$ или $p=(n^2-n)/2$ и $m \geqslant n-1$.

Отображение ограничения для когомологий индуцирует отображение из $V^p(m)$ в $(V^p(m))_K$. Это показывает, что $V^p(m)$ особенно интересно, когда выполнено одно из двух условий, указанных в (h).

Мы уже отбросили случай неположительного m как неинтересный; ничего не остается, как рассмотреть другой случай. Определитель действия $G(f)$ на голоморфном кокасательном пространстве в нуле равен χ_{2n}, и можно предположить, что здесь где-то содержится сдвиг на n — подкрутка на «полуплотность».

После некоторого более серьёзного исследования, получается следующее. Пусть m больше или равно n и $p=(n^2-n)/2$. Тогда существует G-инвариантное плотное подпространство

(i) $\mathscr{H}(m-n) \subset V^p(m)$,

которое обладает G-инвариантным скалярным произведением, превращающим его в гильбертово пространство. Для m, больших n, полученное унитарное представление соответствует орбите $[(m-n)/2]F$.

Этот пример очень полезен в плане ответа на вопрос, как в общих чертах следует строить представления из эллиптических орбит. В деталях это значительно менее ясно; скажем, скалярное произведение в наших рассмотрениях возникло в конце примера совершенно ниоткуда. Бдительный читатель мог заметить некоторую трудность, с которой мы столкнулись ещё раньше. А именно, пространство когомологий Дольбо возникает (например) как пространство когомологий некоторого комплекса $(0, p)$-форм. Нет никаких очевидных причин для того, чтобы у нас дифференциал имел замкнутый образ, так что трудно даже определить топологию на $V^p(m)$.

На компактных многообразиях обе проблемы решаются теорией Ходжа: для всего можно найти гармонических представителей и работать с ними. В рассматриваемом случае для применения этой идеи есть три препятствия. Во-первых, кэлерова форма является неопределенной, так что оператор Лапласа не эллиптичен. Поэтому гармонические формы — не вполне хороший объект. Во-вторых, наше многообразие некомпактно, поэтому возникают проблемы сходимости, которые необходимо решить, прежде чем можно будет интегрировать формы для получения скалярного произведения в когомологиях. В-третьих, из неопределённости кэлеровой формы следует, что локальное скалярное произведение на $(0, p)$-формах также неопределённо; поэтому неясно, будет ли скалярное произведение в когомологиях положительным.

В данный момент эти проблемы кажутся неразрешимыми. (Шмид решил их для дискретной серии. Некоторые «набеги» на общий случай предприняты в [11].) Однако для целей теории представлений их можно считать полностью решёнными теорией когомологической параболической индукции Цуккермана (см. [12]). Эта теория строит представления, используя формальную имитацию комплексного анализа на подходящих однородных пространствах. Метод Цуккермана позволяет нам завершить эллиптический шаг метода орбит. Для упрощения изложения мы будем оставаться в рамках комплексного анализа.

Итак, пусть теперь группа G редуктивна и f_e — эллиптический элемент из $\mathfrak{g}^*$. Обозначим через X_e соответствующий элемент алгебры Ли. Оказывается, что $\operatorname{ad}(X_e)$ обладает чисто мнимыми собственными значениями, так что мы имеем разложение на собственные подпространства

$$\mathfrak{g}_{\mathbb{C}} = \sum_{r \in \mathbb{R}} (\mathfrak{g}_{\mathbb{C}})^r. \tag{5.7a}$$

Здесь

$$(\mathfrak{g}_{\mathbb{C}})^r = \{Y \in \mathfrak{g} \mid [iX_e, Y] = rY\}. \tag{5.7b}$$

Это разложение сохраняется действием группы $G(f_e)$, и весовое пространство веса нуль совпадает с $\mathfrak{g}(f_e)_{\mathbb{C}}$. Выполнены также соотношения, аналогичные (4.1d) и (4.1e).

Положим

$$\mathfrak{u}_e = \sum_{r>0} (\mathfrak{g}_{\mathbb{C}})^r \tag{5.8a}$$

и

$$\mathfrak{q}_e = \mathfrak{g}(f_e)_{\mathbb{C}} + \mathfrak{u}_e. \tag{5.8b}$$

Подпространство

$$\mathfrak{q}_e/\mathfrak{g}(f_e)_{\mathbb{C}} \subset [\mathfrak{g}/\mathfrak{g}(f_e)]_{\mathbb{C}} \tag{5.8c}$$

оказывается антиголоморфным касательным пространством к f_e для кэлеровой структуры на орбите

$$F_e = G \cdot f_e. \tag{5.9a}$$

Пусть

$$(s, r) \text{ — сигнатура кэлеровой формы на } F_e. \tag{5.9b}$$

Опишем теперь схематически эллиптический шаг метода орбит. (Когда он является частью программы, представленной формулами (3.11), G заменяется на $G(f_h)$.)

Предполагается, что унипотентный шаг должен дать нам некоторое подлинное унитарное представление π^{mp} группы $G(f_e)^{\mathrm{mp}}$. После подкручивания на корень квадратный из определителя действия $G(f_e)$ на голоморфном кокасательном пространстве в f_e мы получим в точности унитарное представление (π_e, V_e) группы $G(f_e)$. Оно индуцирует голоморфное гильбертово расслоение $\mathscr{V}_e$ на F_e. Интересующее нас представление реализуется в подходящем плотном подпространстве когомологий Дольбо

$$H^{0,s}(F_e, \mathscr{V}_e). \tag{5.10}$$

Как уже было отмечено раньше, для этих унитарных представлений имеется теорема существования.

«Теорема» 5.11 [12]. *Пусть G — редуктивная группа и f_e — эллиптический элемент из $\mathfrak{g}^*$. Ниже мы используем обозначения* (5.7)—(5.9). *По поводу метаплектической накрывающей $G(f_e)^{\mathrm{mp}}$ см. определение* 2.17.

(a) *Существует подлинный характер ρ_e группы $G(f_e)^{\mathrm{mp}}$, такой что $[(\rho_e)(g)]^2 = \det \mathrm{Ad}(g)$ на $\mathfrak{u}_e$ для g из $G(f_e)^{\mathrm{mp}}$.*

Зафиксируем неприводимое унитарное подлинное представление π^{mp} группы $G(f_e)^{\mathrm{mp}}$. Предположим, что: (1) *ограничение π^{mp} на коммутаторную подгруппу слабо унипотентно* [12, *определение* 8.16] *и* (2) *π^{mp} имеет дифференциал $(if_e)\mathrm{Id}$ на центре $\mathfrak{g}(f_e)$. Положим $\pi_e = \pi^{\mathrm{mp}} \otimes \rho_e$, и пусть $\mathscr{V}_e$ — индуцированное голоморфное гильбертово расслоение на F_e.*

(b) *Когомологии Дольбо пространства F_e с коэффициентами в $\mathscr{V}_e$ равны нулю во всех степенях, кроме s.*

(c) *Существует плотное G-инвариантное подпространство $V(f_e, \pi^{\mathrm{mp}}) \subset H^s(F_e, \mathscr{V}_e)$, в котором реализуется унитарное представление группы G.*

Кавычки означают следующее. Результат, по-видимому, верен, как он сформулирован, но имеющаяся техника позволяет доказать лишь некий его алгебраический аналог. (В действительности унитарное представление строится для алгебры Ли, чего здесь недостает — это геометрической реализации.) Первое из двух фигурирующих в формулировке теоремы предположений может рассматриваться как желаемое свойство метода орбит для нильпотентных орбит. Второе связано с первым предположением из определения 2.17. Один из способов удовлетворить обоим условиям — устроить так, чтобы π^{mp} было унитарным характером с дифференциалом f_e.

Представление $\bar{V}$, даваемое теоремой, может оказаться приводимым или нулевым; но если f_e — достаточно общего вида, ни одна из этих возможностей не реализуется.

6. НИЛЬПОТЕНТНЫЙ ШАГ

Осталось выяснить, какие представления нашей редуктивной группы G соответствуют нильпотентным коприсоединенным орбитам. Это пока еще в значительной степени окутано тайной. Нильпотентные орбиты, вообще говоря, не являются кокасательными расслоениями; они никогда не обладают инвариантной кэлеровой структурой, за исключением случая точки нуль,— и специалисты по теории представлений пока не знают других ответов на вопрос, поставленный в начале разд. 4.

Тем не менее, и здесь можно извлечь кое-что полезное из методов, используемых на гиперболическом и эллиптическом шагах. Во-первых, нулевой элемент в $\mathfrak{g}^*$ является одновременно гиперболическим и эллиптическим (равно как и нильпотентным); рассмотрения разд. 4 или 5 показывают, что соответствующие представления должны быть тривиальны на G_0. Вот более тонкая реализация той же идеи:

Лемма 6.1. *В обозначениях* (4.1) *и* (4.2) *существуют открытые орбиты $E_1, \ldots, E_k$ для G в $T^*(G/P_h)$. Каждая орбита E_i является конечным накрытием нильпотентной коприсоединенной орбиты F_i (как однородного симплектического пространства).*

Эта лемма подсказывает ввести следующее

Предположение 6.2. В условиях леммы 6.1 пусть $\mathscr{V}$ — однородное эрмитово векторное расслоение на G/P_h с инвариантной плоской связностью. Тогда унитарное представление G в L^2-сечениях $\mathscr{V}$ (подкрученных на полуплотности) должно быть одним из тех, которые ассоциированы с нильпотентными коприсоединёнными орбитами F_i.

Предположение в том же духе можно сформулировать на основе эллиптического шага:

Предположение 6.3. Пусть G — редуктивная группа и f_e — эллиптический элемент из $\mathfrak{g}^*$. Используем обозначения (5.7)—(5.9). Пусть π^{mp} — подлинное представление группы $G(f_e)^{\mathrm{mp}}$, тривиальное на компоненте единицы. Положим

$$\pi_e = \pi^{\mathrm{mp}} \otimes \rho_e$$

(ср. с теоремой 5.11), и пусть $\mathscr{V}_e$ — индуцированное голоморфное гильбертово расслоение на F_e. Тогда унитарное представление G, которое плотно в $H^s(F_e, \mathscr{V}_e)$, должно быть одним из тех, что связаны с нильпотентными коприсоединенными орбитами.

Нильпотентные орбиты, с которыми это представление должно быть связано, находятся в «ассоциированном конусе»

$$\lim_{t\to 0+} tF_e. \tag{6.4}$$

Можно показать, что представление является ненулевым, если и только если этот предельный конус имеет такую же размерность, как и F_e. (Орбиты из предположения 6.2 могут быть получены как ассоциированный конус для F_h. Они всегда имеют ту же размерность, что и F_h.)

Используя подобные соображения, можно собрать достаточное число фактов для обоснования предложенного описания метода орбит для нильпотентных орбит. Дополнительная помощь пришла из теории примитивных идеалов и теории автоморфных форм. Хотелось бы приложить к этим теориям наши знания о представлениях. Однако в каждой из этих областей есть своя, внутренняя интуиция, и можно предсказать, что может сказать о них теория представлений. Эти предсказания можно трактовать как гипотезы в теории представлений, — если повезёт, как описание представлений, связанных с нильпотентными орбитами. Ориентир в этом направлении — статья [1]. Отчёт о дальнейших результатах появится в [13].

ЛИТЕРАТУРА

1. J. Arthur, On some problems suggested by the trace formula, Lie Group Representations, II (College Park, Md., 1982/1983), Lecture Notes in Math., vol. 1041, Springer-Verlag, Berlin — Heidelberg — New York, 1983.

2. L. Auslander and B. Kostant, Polarization and unitary representations of solvable Lie groups, Invent. Math. **14** (1971), 255—354.

3. A. Borel and N. Wallach, Continuous cohomology, discrete subgroups, and representations of reductive groups, Princeton Univ. Press, Princeton, N. J., 1980.

4. M. Duflo, Théorie de Mackey pour les groupes de Lie algébriques, Acta Math. **149** (1982), 153—213.

5. V. Guillemin and S. Sternberg, Geometric asymptotics, Mathematical Surveys, No. 14, Amer. Math. Soc., Providence, R. I., 1978. [Имеется перевод: Гийемин Б., Стернберг С. Геометрические асимптотики. — М.: Мир, 1981.]

6. Кириллов А. А. Унитарные представления нильпотентных групп Ли.— УМН, 1962, т. 17, № 4, 57—110.

7. —, Элементы теории представлений. — 2-е изд. — М.: Наука, 1978.

8. A. Knapp, Representation theory of real semisimple groups: an overview based on examples, Princeton Univ. Press, Princeton, N. J., 1986.

9. B. Kostant, Quantization and unitary representations, Lectures in Modern Analysis and Applications (C. Taam, ed.), Lecture Notes in Math., vol. 170, Springer-Verlag, Berlin — Heidelberg — New York, 1970.

10. G. Mackey, Theory of unitary group representations, Univ. of Chicago Press, Chicago, Ill., 1976.

11. J. Rawnsley, W. Schmid, and J. Wolf, Singular unitary representations and indefinite harmonic theory, J. Funct. Anal. **51** (1983), 1—114.

12. D. Vogan, Unitarizability of certain series of representations, Ann. of Math. **120** (1984), 141—187.

13. —, Unitary representations of reductive Lie groups, Ann. of Math. Studies, 1987.

14. D. Vogan and G. Zuckerman, Unitary representations with non-zero cohomology, Compositio Math. **53** (1984), 51—90.

МАССАЧУСЕТСКИЙ ТЕХНОЛОГИЧЕСКИЙ ИНСТИТУТ, КЕЙМБРИДЖ, МАССАЧУСЕТС 02139, США

ФИЗИКА И ГЕОМЕТРИЯ

Эдвард Виттен[1]

Уже не раз в прошлом задачи, возникавшие в теоретической физике, влияли на развитие математики, и наоборот, структуры, впервые появившиеся в математике, участвовали в развитии физики. В двадцатом веке самые яркие примеры тому — роль римановой геометрии в открытии общей теории относительности и влияние квантовой механики на развитие функционального анализа. Эти примеры, однако, связаны с событиями шестидесяти-семидесятилетней давности. В последние полвека математика и физика развивались в различных направлениях и взаимодействие этих дисциплин играло меньшую роль.

Отчасти это произошло потому, что математика осваивала абстрактные миры, внешне не связанные с банальным миром физика-теоретика, отчасти же из-за того, каким путём пошла физика. Две основные теории в физике двадцатого века — это общая теория относительности и квантовая теория поля. Их области приложения весьма различны. Общая теория относительности — эйнштейнова теория гравитации — имеет дело с крупномасштабными астрономическими явлениями, в то время как квантовая теория поля служит моделью, с помощью которой физикам удалось понять многие свойства элементарных частиц. Общая теория относительности была в своем окончательном виде сформулирована Эйнштейном в 1915 г., а квантовая теория остается открытой областью со времени своего возникновения в конце 20-х годов. В течение полувека основные продвижения в физике были связаны с развитием квантовой теории поля. Именно квантовая теория поля была в этот период главной ареной возможных взаимодействий между математикой и физикой.

---

[1] Edward Witten, Physics and geometry, ICM86, pp. 267—303.

В течение нескольких десятилетий после изобретения квантовой теории поля она оставалась довольно запутанной и неуклюжей теорией, с которой было трудно работать даже физикам. Было вовсе не очевидно, что она вообще допускает точную математическую формулировку. И что самое характерное, за первые несколько десятилетий существования квантовой теории поля в связи с ней не возникло новых интересных математических структур.

Это положение начало меняться в середине 70-х годов после появления неабелевых калибровочных теорий как основных объектов теоретической физики. В рамках этих теорий многие важные физические задачи оказались связанными с существенными понятиями современной математики. Например, изучение магнитных монополей и инстантонов использует классификацию векторных расслоений. Решение «U(1)-проблемы» в квантовой хромодинамике опирается на теорему Атьи—Зингера об индексе. Подлинное понимание локальных и глобальных «аномалий» требует привлечения довольно тонких свойств семейств эллиптических операторов. Можно привести ещё много других примеров.

Разумеется, приятно видеть «практическое» применение заумной на первый взгляд математики. В некоторых из перечисленных мною случаев решение физических задач потребовало открытия новых математических теорем. Но всё же взаимное влияние математики и физики оставалось бы, я полагаю, довольно ограниченным, если бы речь шла об одной только квантовой теории поля. Приложения современной математики к квантовой теории поля впечатляющи, но довольно специальны; то же можно сказать и о роли, которую квантовая теория поля играла до сих пор в развитии математики. Именно при попытках выйти за пределы квантовой теории поля физики ощутили реальную потребность в математике.

Ограниченность квантовой теории поля заключается в том, что, как мы отмечали выше, это лишь одна из двух фундаментальных теорий физики двадцатого века; другой является общая теория относительности. Обе эти теории описывают один и тот же реальный мир, поэтому более полное описание должно включать их обе. Однако уже с самых первых дней квантовой теории было ясно, что при попытке сочетать квантовую теорию поля с общей теорией относительности возникают серьёзные трудности. Формальное применение процедуры квантования к гравитации приводит к бессмысленным бесконечностям. Но в начале своего развития квантовая теория поля встретилась со многими трудностями, и указанная выше была лишь одной из них. По мере того как все прочие трудности были преодолены и квантовая теория поля превратилась во вполне адекватную схему, описывающую все силы природы, кроме гравитации, несогласованность между квантовой теорией и общей теорией относительности выступила как главное свидетельство ограниченности квантовой теории поля.

Эта проблема носит по преимуществу теоретический характер. Эксперименты здесь мало что дают сверх того, что обе теории участвуют в описании закона природы. К сожалению, гравитационные эффекты пренебрежимо малы во всех экспериментах, где реально играет роль квантовая теория поля, и обратно. Всё же несогласованность между двумя основными теориями в физике является, очевидно, существенной проблемой в логическом плане. В самом деле, история физики даёт много примеров, показывающих, как важны подобные проблемы. Так например, общая теория относительности возникла в результате попытки Эйнштейна разрешить противоречие между двумя ведущими теориями того времени: специальной теорией относительности и ньютоновской теорией гравитации. Точно так же квантовая теория поля родилась из попытки объединить нерелятивистскую квантовую механику со специальной теорией относительности.

При открытии общей теории относительности сначала появилась логическая схема. Эйнштейн вначале продумал физические принципы, которые должна воплощать новая теория, затем обнаружил в римановой геометрии подходящий математический аппарат и, наконец, сформулировал саму теорию. Развитие квантовой механики и квантовой теории поля шло совсем по-другому. Здесь не было никаких априорных принципов; основную роль играла феноменология. Как я уже говорил, при объединении общей теории относительности с квантовой теорией поля мало надежды на помощь со стороны экспериментальной физики. Поэтому можно прийти к мысли, что единственная надежда состоит в том, чтобы по образцу общей теории относительности изобрести чистым усилием мысли новую математическую схему, обобщающую риманову геометрию и способную охватить квантовую теорию поля. Многие честолюбивые теоретики пытались это сделать, но из этого пока ничего не вышло.

Прогресс был достигнут совсем другим путём. В попытке понять механизм сильных взаимодействий физики пришли в конце 60-х — начале 70-х годов к исследованию того, что стало потом известным под именем «струнной теории». На теорию струн первоначально «наткнулись» случайно или, во всяком случае, на весьма окольном пути при изучении так называемой «модели Венециано» [30]. Обзоры струнной теории можно найти в [11, 13, 17, 25]. По мере развития этой теории появлялись замечательно богатые математические структуры, однако всё менее и менее похожие на теорию сильных взаимодействий. Примерно в 1973—74 гг. теория сильных взаимодействий была успешно построена в рамках неабелевой калибровочной теории. Однако математическая сторона струнной теории сохранила свою привлекательность. Около 1974 г., как раз когда первоначальная мотивировка занятий струнной теорией пошла на убыль, возникло предположение, что её надо рассматривать не как теорию сильных взаимодействий, а как схему, позволяющую согласовать гравитацию с квантовой механикой [24]. У этой идеи немало странных следствий. Например, нужно примириться с тем, что пространство-время (если только привычные геометрические понятия вообще здесь применимы) имеет не четыре, а десять измерений. После нескольких лет забвения эта идея вновь появилась на авансцене в 80-х годах, и есть основания думать, что она близка к правильной теории.

За удачную находку на окольном пути приходится платить. Хотя мы уже знаем много о струнной теории, мы до сих пор не имеем логической схемы, в которой она «чувствовала» бы себя как дома. Дело обстоит примерно так, как если бы нам удалось сформулировать общую теорию относительности в каких-нибудь искусственных терминах, не ведая ничего о римановой геометрии; тогда естественно возникла бы задача построения римановой геометрии как математического аппарата теории гравитации. Сама мысль о формулировке общей теории относительности без римановой геометрии кажется странной, но именно такая ситуация в струнной теории. Никто не знает, какой окажется её естественная логическая схема. Мы можем сказать, что составными частями этой теории являются теория римановых поверхностей, теория модулярных форм и теория представлений бесконечномерных алгебр Ли. Это — необходимые предварительные сведения для занятий струнной теорией, так же как начала линейной алгебры образуют фундамент для римановой геометрии и общей теории относительности. Хотя мы и не знаем логических основ струнной теории, кажется правдоподобным, что они включают в себя некоторое фундаментальное обобщение обычной геометрии. Это обобщение должно иметь далеко идущие последствия для математики, равно как и для физики. Его открытие будет знаменовать начало нового золотого века в истории взаимодействия физики и математики.

Вполне вероятно, что в каком-то смысле число фундаментальных математических задач бесконечно. С другой стороны, лично я убежден, что в физике число по-настоящему фундаментальных проблем конечно. Если это так, в будущем нас ожидает лишь конечное число примеров существенного взаимодействия физики и математики. Похоже, что ближайшие десятилетия будут одним из таких периодов.

1. ФИЗИКА ЭЛЕМЕНТАРНЫХ ЧАСТИЦ В ВОСЬМИДЕСЯТЫХ ГОДАХ

Этот доклад состоит из четырех разделов. В первом из них я хочу дать обзор основных составляющих частей наших нынешних познаний в фундаментальной физике. В следующем разделе я попытаюсь объяснить, почему идея десятимерности нашего пространства-времени (что является одним из требований теории струн) не только согласуется с нашим повседневным опытом, но и является весьма привлекательной. Обзор тем, затрагиваемых в этих двух разделах, можно найти в [5, 13, 35]. В третьем разделе я дам очень краткое введение в квантовую теорию поля, выделяя те её аспекты, которые понадобятся для теории струн. Последний раздел посвящен теории струн.

Начнём наше изложение теоретической физики с общей теории относительности (ОТО). В этой теории роль пространства-времени играет псевдориманово многообразие M с метрикой сигнатуры $(-+\ldots+)$. Всюду в данном разделе M четырёхмерно. Я буду обозначать локальные координаты в пространстве-времени через x^i $i=1, \ldots, 4$. Основным уравнением ОТО является уравнение, выражающее вариационный принцип для действия

$$S_{\text{ОТО}} = \frac{1}{16\pi G} \int\limits_M R, \tag{1}$$

где R — скалярная кривизна на M, G — постоянная Ньютона. Соответствующее уравнение Эйлера—Лагранжа представляет собой условие обращения в нуль тензора Риччи R_{ij}:

$$R_{ij} = 0. \tag{2}$$

Из фундаментальных постоянных физики G, $\hbar$ (постоянная Планка) и c (скорость света) можно сконструировать выражение, имеющее размерность массы:

$$M_{\text{пл}} = \sqrt{\hbar c/G}. \tag{3}$$

Эта величина, называемая планковской массой, является естественной единицей массы в физике. В общепринятых единицах её приближённое значение таково:

$$M_{\text{пл}} \approx 1.02 \times 10^{-5} \text{ грамм}. \tag{4}$$

Точно так же мы можем образовать из фундаментальных постоянных фундаментальную длину

$$R_{\text{пл}} = \hbar/M_{\text{пл}}c \approx 10^{-33} \text{ сантиметров}, \tag{5}$$

называемую планковской длиной, и фундаментальное время

$$t_{\text{пл}} \approx 10^{-43} \text{ секунд}, \tag{6}$$

называемое планковским временем. Постоянные $\hbar$ и G настолько фундаментальны в физике, что естественнее всего работать в системе единиц, в которой $\hbar = c = G = M_{\text{пл}} = R_{\text{пл}} = t_{\text{пл}} = 1$. В этой системе любая физическая величина — любая длина, время, масса — будет просто числом.

Заметим, что значения у фундаментальных масс, длины и времени весьма странные. Планковская масса (4) — величина макроскопическая (равная примерно массе бактерии); тем самым это величина совсем другого порядка, нежели массы всех известных элементарных частиц. Масса электрона в 10^{22} раз меньше планковской, и самые тяжелые из получаемых нами в ускорителях частиц всё же в 10^{17} раз легче $M_{\text{пл}}$. Точно также (5) и (6) — величины, совершенно несравнимые по порядку с обычными в физике элементарных частиц. Все наши знания в квантовой теории поля получены из экспериментов на минимальных расстояниях порядка 10^{-16} см и на минимальных промежутках времени порядка 10^{-26} сек. Разумеется, эти величины очень малы по обычным меркам, но они оказываются очень большими, если измерять их в единицах, задаваемых фундаментальными физическими постоянными. Для непосредственного экспериментального изучения того, как в природе совмещаются квантовая механика и общая теория относительности, необходимы опыты, которые позволили бы нам регистрировать процессы, происходящие за время порядка $t_{\text{пл}}$ и на расстояниях порядка $R_{\text{пл}}$, или же опыты, в которых участвовали бы частицы, разогнанные до кинетической энергии порядка $M_{\text{пл}}$. К сожалению, это остаётся недостижимым для нас в обозримом будущем. Мы можем, конечно, надеяться получить из эксперимента какие-то косвенные указания, но настоящее продвижение в теории квантовой гравитации окажется возможным лишь при большой удаче и новом прозрении в теории.

На самом деле последствия того, что масса $M_{\text{пл}}$ велика, мы можем наблюдать и в повседневной жизни. Сказать, что $M_{\text{пл}}$ очень велика по сравнению с массой обычной частицы m, это то же самое, что сказать, что постоянная Ньютона G очень мала в масштабе, определяемом m, $\hbar$ и c:

$$G = \hbar c/M_{\text{пл}}^2 \ll \hbar c/m^2. \tag{7}$$

Малость постоянной Ньютона означает, что гравитационные силы, действующие между отдельными частицами массы m, очень малы. Тем самым гравитационное взаимодействие между обычными атомами становится заметным лишь при рассмотрении столь громадных образований, что эффект аккумуляции перевешивает чрезвычайную слабость гравитационных сил на атомном уровне. Это означает, что тела, образующиеся из отдельных атомов под действием сил притяжения, — такие как планеты и звёзды — должны быть очень большими. Точно также огромность астрономических расстояний по сравнению с атомными связана с тем, что атомные масштабы велики по сравнению с $R_{\text{пл}}$. Оба эти факта загадочны.

Обычные массы элементарных частиц настолько малы по отношению к естественным природным единицам, что должна существовать естественная идеализация, в которой они равны нулю. Одна из целей, преследуемых нами в этом разделе, — пояснить, что это за идеализация.

Теперь мне хотелось бы обсудить физическое содержание ОТО. В двумерном пространстве-времени интеграл (1) является топологическим инвариантом (эйлерова характеристика пространства-времени), и теория, определяемая одним этим интегралом, малосодержательна. В трёхмерном пространстве-времени из вариационного уравнения $R_{ij}=0$ следует, что пространство-время плоско (так как на трёхмерном многообразии тензор Римана может быть выражен через тензор Риччи). Специфические свойства ОТО впервые проявляются в размерности 4. В этом случае из уравнений Эйнштейна $R_{ij}=0$ уже не следует, что пространство-время плоско, — это уравнение имеет нетривиальные решения типа волн. Если η_{ij} — плоская лоренцева метрика ($\eta=\operatorname{diag}(-+\ldots+)$), то можно искать решение, близкое к плоскому:

$$g_{ij}=\eta_{ij}+h_{ij}, \tag{8}$$

где h мало. Отбрасывая члены высокого (по сравнению с h) порядка малости, получаем решение уравнения Эйнштейна в виде плоских волн:

$$h_{ij}=\varepsilon_{ij}e^{ikx}+\text{комплексно-сопряжённые члены}, \tag{9}$$

где k_i и ε_{ij} — коэффициенты, удовлетворяющие условиям

$$k_ik^i=k^i\varepsilon_{ij}=\varepsilon_i^i=0. \tag{10}$$

Решения (9) очень похожи на плоские волны — решения уравнения Максвелла, описывающие световые волны. Когда, сразу же после формулировки ОТО, было открыто решение (9) линеаризованных уравнений Эйнштейна, оно интерпретировалось как предсказание существования гравитационных волн, распространяющихся с той же скоростью, что и световые, так как и те и другие удовлетворяют условию $k_ik^i=0$. Тогда частицы («вещество») и волны рассматривались как совершенно разные вещи, и в предсказаниях общей теории относительности речь шла именно о новом виде волн, а не о новом виде частиц[1]. Через десяток лет, с развитием квантовой механики, стало ясно, что волны и частицы представляют собой просто две точки зрения на один и тот же объект — некую первичную субстанцию, выступающую как волна или как частица в зависимости от обстоятельств. Тем самым ОТО — это не просто теория гравитационных сил; она также описывает определённый тип «материи»; в концептуальном плане это блестящий триумф. Пытаясь просто придумать теорию гравитационных сил, основанную на римановой геометрии, Эйнштейн был вынужден изобрести объединённую теорию гравитации и вещества. Кое-что, однако, не получается. Похоже, что как квантовая теория ОТО не имеет смысла, и наблюдаемые в природе формы вещества намного разнообразнее, чем предсказывает ОТО.

Наша следующая цель — обсудить некоторые другие формы вещества (или, что эквивалентно, некоторые другие типы волн), наблюдаемые в природе. В первую очередь это неабелевы калибровочные силы. Тем самым пространственно-временное многообразие M помимо римановой метрики наделяется дополнительной структурой. Мы имеем главное расслоение над M

$$\begin{tikzcd} X \arrow[d, "G"] \\ M \end{tikzcd} \tag{11}$$

со структурной группой G, известной у физиков как «калибровочная группа». С каждым представлением R группы G можно связать некоторое векторное расслоение V_R: оно будет играть важную роль в наших последующих построениях. Что касается группы G, то мы знаем из эксперимента, что она содержит в качестве подгруппы произведение $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$, сомножители которого отвечают соответственно сильному, слабому и

---

[1] В то время, когда было сделано это предсказание, и в течение многих последующих десятилетий перспектива экспериментальной проверки этого предсказания казалась безнадёжно отдаленной — из-за чрезвычайной слабости гравитационных сил. Однако появление радиоастрономии и открытие в последние годы радиопульсаров сделали возможным косвенную, но неоспоримую экспериментальную проверку теории гравитационных волн.

электромагнитному взаимодействиям[1]. Пусть A — некоторая связность на V_R и F — соответствующая форма кривизны. Тогда действие Янга—Миллса (лагранжиан) задаётся формулой

$$S_{ЯМ}=\frac{1}{4e^2}\int\limits_M |F|^2, \tag{12}$$

где

$$|F|^2=g^{ii'}g^{jj'}\langle F_{ij}\,|\,F_{i'j'}\rangle. \tag{13}$$

Здесь g_{ij} — метрика пространства-времени, а $\langle\,|\,\rangle$ — форма Картана—Киллинга на алгебре Ли группы G. Постоянная e в (12) называется янг-миллсовской константой связи. В случае когда группа G не является простой, определение (12) можно обобщить, введя различные константы связи для каждого из простых сомножителей в калибровочной группе.

Разумеется, фигурирующая в (13) метрика g — та же самая, что и метрика в (1), поскольку всё происходит в одном и том же пространстве-времени. Мы можем сложить лагранжианы Эйнштейна и Янга—Миллса и изучить комбинированную теорию:

$$S=S_{ОТО}+S_{ЯМ}. \tag{14}$$

Записав соответствующие вариационные уравнения, мы получим объединенные уравнения для полей Янга—Миллса и гравитационных. Это подходящая основа для описания отклонения световых лучей гравитационным полем Солнца и многих других, более экзотических процессов, которые, к сожалению, слишком слабы, чтобы их можно было заметить.

Следующий принципиальный шаг — включить в рассмотрение то, что физики называют «фермионами», в противоположность «бозонам» — янг-миллсовским и гравитационным полям. Для этого введём алгебру Клиффорда, т. е. алгебру «дираковских матриц» Γ^i, $i=1, \ldots, n$, удовлетворяющих соотношениям

$$\Gamma^i\Gamma^j+\Gamma^j\Gamma^i=-2g^{ij}, \quad i,\ j=1, \ldots, n. \tag{15}$$

Для чётных n неприводимое представление S алгебры Клиффорда имеет размерность $2^{n/2}$, для нечётных — $2^{(n-1)/2}$. Клиффордов модуль S автоматически приобретает структуру представления группы Лоренца $SO(1, n-1)$. Оно называется спинорным

представлением, и генераторы группы Лоренца действуют в нём как

$$\Sigma^{ij}=-\Sigma^{ji}=\frac{1}{4}[\Gamma^i,\ \Gamma^j]. \tag{16}$$

Для n нечётных S будет неприводимым представлением группы Лоренца, а для n чётных — которые, собственно, нас и интересуют — S разлагается на неприводимые:

$$S=S_+\oplus S_-, \tag{17}$$

где S_+ и S_- — собственные подпространства для инволюции

$$\overline{\Gamma}=i^{(n+2)/4}\Gamma^1\Gamma^2\ldots\Gamma^n. \tag{18}$$

В размерности 4 пространства S_+ и S_- комплексно-сопряжены.

Если второй класс Штифеля—Уитни многообразия M обращается в нуль, мы можем определить «спинорный пучок» на M, который я буду обозначать $\hat{S}$. По сути дела это — векторное расслоение, слой которого над точкой $p\in M$ есть клиффордов модуль, определяемый соотношениями (15). Сечения ψ пучка $\hat{S}$ — это физические поля, называемые «фермионными полями со спином 1/2». Важными примерами служат лептоны (как например, электроны) и кварки (из которых состоят нейтроны и протоны). Выражение «спин 1/2» означает, что веса спинорного представления $SO(1, n-1)$ — полуцелые.

В чётных размерностях спинорный пучок разлагается в сумму

$$\hat{S}=\hat{S}_+\oplus\hat{S}_- \tag{19}$$

аналогично (17). Физическое поле, являющееся сечением $\hat{S}_+$ или $\hat{S}_-$, называется фермионным полем положительной или отрицательной киральности соответственно. Их также часто называют право- и левоспиральными: если мы пропустим пучок фермионов положительной киральности (т. е. частиц, математически описываемых как сечения $\hat{S}_+$) через вещество, он начнёт вращаться против часовой стрелки.

Различные принципы квантовой теории поля — как например, CPT-теорема — требуют вещественности фермионных полей. Поскольку в размерности 4 пространство $\hat{S}_+$ комплексно-сопряжено $\hat{S}_-$, то, вводя фермионное поле ψ — сечение $\hat{S}_+$, мы также должны ввести фермионное поле $\bar{\psi}$ — сечение $\hat{S}_-$. То же самое — и по тем же причинам — верно и в размерности $4k$ для любого k. По-другому обстоит дело в размерности $4k+2$, так как в этом случае $\hat{S}_+$ и $\hat{S}_-$ будут вещественными (или псевдовещественными), но этот случай мы обсудим в следующем разделе.

Если уж мы имеем в нашей теории фермионное поле со спином 1/2, то какому уравнению должно оно подчиняться?

---

[1] Эксперимент даёт нам сведения скорее об алгебре Ли группы G, чем о самой группе. Когда я говорю, что G содержит подгруппу $SU(3)\times SU(2)\times U(1)$, я на самом деле имею в виду, что алгебра Ли группы G содержит алгебру Ли группы $SU(3)\times SU(2)\times U(1)$; о глобальной структуре группы G никаких утверждений не делается. По той же причине в дальнейшем, как правило, не различаются группы Ли, имеющие одну и ту же алгебру Ли.

Существует естественный эллиптический оператор D первого порядка — оператор Дирака, отображающий сечения $\hat{S}$ в сечения $\hat{S}$. Он определяется формулой

$$D=\Gamma^i D_i, \tag{20}$$

где D_i — ковариантные дифференцирования. Пока мы рассматриваем уравнение Дирака для фермионов, взаимодействующих лишь с гравитационным полем; поэтому мы берём ковариантное дифференцирование, задаваемое связностью, согласованной с метрикой. В чётной размерности оператор D можно разложить:

$$D=D_+\oplus D_-, \tag{21}$$

где

$$D_+:\hat{S}_-\to\hat{S}_+,\quad D_-:\hat{S}_+\to\hat{S}_- \tag{22}$$

(другими словами, D_- отображает сечения $\hat{S}_+$ в сечения $\hat{S}_-$, и наоборот). Тем самым мы получаем следующие уравнения Дирака:

$$D_-\psi_+=0,\quad D_+\psi_-=0. \tag{23}$$

Это то, что обычно называют «безмассовым» уравнением Дирака. Мы можем ввести произвольное комплексное число λ и написать:

$$D_-\psi_+ +\lambda\psi_-=0,\quad D_+\psi_- +\lambda^*\psi_+=0 \tag{24}$$

(здесь λ^* — число, комплексно-сопряжённое к λ, так что второе уравнение — это просто комплексно-сопряженное к первому). В то время как уравнение (23) описывает «безмассовые» волны, распространяющиеся со скоростью света, подобно электромагнитным или гравитационным, уравнение (24) описывает фермионы ненулевой массы, а именно массы λ. Как я уже объяснял, во многих важных случаях (таких как лептоны и кварки) массовый член наших уравнений очень мал, меньше 10^{-17} в планковских единицах. Точнее, единственные известные нам случаи — это именно те, где массовый член обращается в нуль со столь высокой степенью точности, ибо наша технология не даёт нам возможности открыть фермионы с массой порядка 1, даже если таковые и имеются.

В рамках этой картины — фермионы, взаимодействующие только с гравитационным полем, — нет никакого естественного объяснения, почему член с λ должен отсутствовать. Чтобы получить такое объяснение — и описать одно из действительно центральных наблюдений в физике, — мы должны вернуться к рассмотрению неабелевых калибровочных полей, о которых временно забыли. Каждому представлению R калибровочной группы G отвечает ассоциированное с ним векторное расслоение V_R. Если $\bar{R}$ — дуальное, или комплексно-сопряжённое, представление, то $V_{\bar{R}}$ — расслоение, двойственное к V_R. Расслоение V_R канонически изоморфно $V_{\bar{R}}$ для вещественных или псевдовещественных R, но не для комплексных.

Теперь имеет смысл рассмотреть фермионные поля, являющиеся сечениями не $\hat{S}_+$, а

$$W_+=\hat{S}_+\otimes V_R. \tag{25}$$

Если мы вводим «правоспиральные» фермионы ψ_+, являющиеся сечениями W_+, то в размерности 4 CPT-теорема требует также введения «левоспиральных» фермионов — сечений комплексно-сопряжённого расслоения

$$\tilde{W}_-=\hat{S}_-\otimes V_{\bar{R}}. \tag{26}$$

Далее, поскольку расслоения V_R и $V_{\bar{R}}$ наделены связностями Янга—Миллса, можно написать операторы Дирака

$$D_-:W_+\to W_-,\quad D_+:\tilde{W}_-\to\tilde{W}_+. \tag{27}$$

Здесь

$$W_-=\hat{S}_-\otimes V_R,\quad \tilde{W}_+=\hat{S}_+\otimes V_{\bar{R}}. \tag{28}$$

Уравнения Дирака — формально — имеют тот же вид, что и раньше:

$$0=D_-\psi_+=D_+\psi_-, \tag{29}$$

но связность есть теперь сумма связностей Леви—Чивиты и Янга — Миллса. Однако между (29) и аналогичным уравнением (23), в которое поля Янга—Миллса не входят, есть большая разница. Она состоит в том, что если представление R комплексно, то мы не можем добавить к (29) массовый член. Уравнение

$$???\,0=D_-\psi_+ +\lambda\psi_- \tag{30}$$

имеет смысл, только если представление R изоморфно своему дуальному, поскольку $D_-\psi_+$ — сечение $\hat{S}_-\otimes V_R$, а ψ_- — сечение $\hat{S}_-\otimes V_{\bar{R}}$. Тем самым, *если* фермионы преобразуются по комплексному представлению калибровочной группы G, то мы можем естественно объяснить, почему они столь легки по сравнению с планковской массой.

В точности это и наблюдается в действительности, и именно в этом большинство физиков видит правильное объяснение того, почему все наблюдаемые фермионы такие легкие. В действительности это утверждение необходимо уточнить, поскольку наблюдаемое в природе представление R далеко не неприводимо. Разложение R на неприводимые представления калибровочной группы $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$, по всей видимости, содержит не менее 15 компонент:

$$R=\bigoplus_{i=1}^{15} R_i. \tag{31}$$

Я не буду давать здесь явного описания R_i, так как смогу сделать это более экономно после обсуждения «великого объединения». Разумеется, из (31) следует, что

$$\tilde{R}=\bigoplus_{j=1}^{15} \tilde{R}_j. \tag{32}$$

Заметим, что если бы существовали i и j с $R_i \approx \tilde{R}_j$, то соответствующие фермионы ψ_{i+} и ψ_{j-} могли бы иметь массы, поскольку в этом случае (30) имело бы смысл. Но для всех известных фермионов $R_i \not\approx \tilde{R}_j$, и это, вместе с другими обрисованными выше фактами, объясняет, почему они столь легки.

Давайте теперь посмотрим на это с немного иной точки зрения. Нет никаких оснований считать, что все существующие в природе фермионы исчерпываются экспериментально обнаруженными. Очень вероятно, что существует много, возможно даже бесконечно много фермионов с массами «порядка единицы», т. е. порядка $M_{\text{пл}}$. Такие фермионы, конечно, должны преобразовываться по вещественному представлению группы G. И обратно, кажется правдоподобным, что фермионы, отвечающие вещественным представлениям, будут иметь массы порядка единицы. Пусть U и $\tilde{U}$ — представления группы G, соответствующие базисным право- и левоспиральным фермионам в некоторой фундаментальной теории природы. Разумеется, они дуальны друг другу. Образуем — в духе K-теории — формальную разность представлений:

$$\Delta=U\ominus\tilde{U}. \tag{33}$$

Помимо наблюдаемых правых фермионов, преобразующихся по представлению R, U может содержать дополнительные правые фермионы, преобразующиеся по некоторому представлению U_0:

$$U=U_0\oplus R, \tag{34}$$

которое должно быть вещественным, так как частицы из U_0 имеют массу. Из (34) вытекает, что

$$\tilde{U}=U_0\oplus\tilde{R}. \tag{35}$$

Тем самым, в разности представлений слагаемое U_0 сократится:

$$\Delta=U\ominus\tilde{U}=R\ominus\tilde{R}. \tag{36}$$

Мы заключаем поэтому, что величина $R\ominus\tilde{R}$, которую мы можем определить в ускорителях, совпадает с разностью представлений (33), характеризующей фундаментальную теорию. Вот почему так важны квантовые числа фермионов низких энергий.

Здесь возникает несколько вопросов. Во-первых, может показаться, что мы слишком хорошо объяснили, почему лёгкие фермионы действительно лёгкие. Массы кварков и лептонов не равны нулю; они лежат в пределах от 10^{-22} до, возможно, 10^{-17} в естественных единицах, описанных выше. В нарисованной же только что картине получается, что массы кварков и лептонов должны быть строго ненулевыми, так как массовый член в (30) строго невозможен. Ответ связан с так называемым «нарушением симметрии». Если нам задано векторное расслоение со структурной группой G, то может случиться, что при некоторых условиях структурная группа расслоения может быть редуцирована до некоторой подгруппы G_0 группы G. Это явление имеет свой аналог в физике элементарных частиц: это так называемое калибровочное нарушение симметрии, играющее центральную роль в теории слабых взаимодействий Вайнберга — Салама—Глэшоу. Сейчас я просто буду считать, что математическая аудитория готова поверить в возможность существования «физического» аналога явления редукции структурной группы к её подгруппе.

Предположим, что на некотором низком уровне энергии m калибровочная группа G эффективно редуцируется до некоторой подгруппы G_0. Даже если представления R и $\tilde{R}$ не эквивалентны как представления G, они могут быть эквивалентными как представления G_0. В этом случае фермионы, бывшие безмассовыми из-за неэквивалентности R и $\tilde{R}$, могут приобрести массу порядка m. Похоже, что именно это и происходит в действительности. При массах частиц порядка $10^{-17}\,M_{\text{пл}}$ калибровочная группа $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$ редуцируется до $\mathrm{SU}(3)\times\mathrm{U}(1)$ [1]. При этом некоторые из калибровочных полей становятся массивными (таковы W- и Z-частицы, открытые в ЦЕРНе несколько лет назад, — самые тяжёлые из известных элементарных частиц). Помимо того представления R и $\tilde{R}$ изоморфны

---

[1] Эти группы отвечают сильным и электромагнитным взаимодействиям соответственно. Появляющаяся здесь группа $\mathrm{U}(1)$ нетривиально вкладывается в исходную группу $\mathrm{SU}(2)\times\mathrm{U}(1)$.

как представления $SU(3) \times U(1)$, поэтому лёгкие фермионы могут иметь и действительно приобретают массу. Во всём этом непонятно, почему уровень энергии, при котором происходит нарушение симметрии, так мал по сравнению с естественной единицей, $M_{\text{пл}}$. Как бы то ни было, из предыдущих рассуждений совершенно ясно, что идеализация, в которой все частицы имеют нулевую массу, получается в случае, когда калибровочная группа G не редуцируется до подгруппы.

В этом кратком обзоре современной физики идея, что представление R может быть комплексным, не подана как шокирующая. Однако исторически, когда это было открыто в 50-х годах (или, точнее, когда были открыты некоторые факты, которые сейчас интерпретируются описанным выше образом), это было ошеломляющей неожиданностью. Почему? В базисном разложении $S = S_+ \oplus S_-$ спинорного представления S на спиноры $S_\pm$ положительной и отрицательной киральности выбор S_+ или S_- — вопрос соглашения. При обращении ориентации пространства-времени (называемой физиками изменением чётности) S_+ и S_- меняются местами. Тем самым при изменении чётности R и $\tilde{R}$ меняются местами. Если предполагать, что законы природы инвариантны относительно изменения чётности, то R и $\tilde{R}$ должны быть изоморфными. Таким образом, наше объяснение лёгкости фермионов основывается на нарушении чётности. Открытие (в 50-х годах) того, что при слабых взаимодействиях чётность не сохраняется, произвело впечатление разорвавшейся бомбы. С другой стороны, при сильных и электромагнитных взаимодействиях чётность сохраняется; это значит, что R и $\tilde{R}$ изоморфны как представления группы $SU(3) \times U(1)$.

После того как мы поняли, что нарушение симметрии играет важную роль в объяснении того, почему лёгкие фермионы не являются строго безмассовыми, естественно развить эту мысль дальше. Мы знаем, что калибровочная группа содержит по меньшей мере $G = SU(3) \times SU(2) \times U(1)$. Мы знаем также, что при очень низких энергиях она редуцируется до подгруппы $G_0 = SU(3) \times U(1)$. Может быть, если обратиться к более высоким энергиям, то наблюдаемая нами при реально достижимых значениях энергии калибровочная группа G сама окажется редукцией некоторой большей группы $\bar{G}$, действующей при больших энергиях? Куда более приемлемым было бы описание природы не с помощью громоздкого произведения $SU(3) \times SU(2) \times U(1)$, а с помощью какой-либо более простой структуры. В этом и состоит цель так называемых «теорий великого объединения». Наиболее очевидный пример группы, содержащей $SU(3) \times SU(2) \times U(1)$ — это $SU(5)$. Вложение $SU(3) \times SU(2)$ в множество 5×5-матриц, составляющих $SU(5)$, может быть

изображено так:

$$\begin{bmatrix} SU(3) & 0 \\ 0 & SU(2) \end{bmatrix}. \tag{37}$$

В качестве генератора группы $U(1)$ возьмём единственную 5×5-матрицу с нулевым следом, коммутирующую с указанным выше вложением $SU(3) \times SU(2)$, а именно

$$\operatorname{diag}(1,\ 1,\ 1,\ -3/2,\ -3/2). \tag{38}$$

Если наблюдаемая калибровочная группа $SU(3) \times SU(2) \times U(1)$ действительно должна быть вложена в $SU(5)$, то разность представлений $\Delta = R \ominus \tilde{R}$ должна допускать естественную интерпретацию как разность $SU(5)$-представлений. Так оно и оказывается, и здесь я в первый раз выпишу явную формулу для разности представлений, наблюдаемой в природе; на уровне $SU(3) \times SU(2) \times U(1)$ она выглядела бы очень громоздко. Пусть $\mathbf{5}$ обозначает фундаментальное пятимерное представление группы $SU(5)$, а $\bar{\mathbf{5}}$ — дуальное к нему представление. Пусть, далее $\mathbf{10}$ — десятимерная антисимметричная часть тензорного произведения $5 \otimes 5$, а $\overline{\mathbf{10}}$ — дуальное к нему представление. Тогда наблюдаемые в природе фермионы, описываются тремя экземплярами некой базисной структуры, обычно называемой «поколением». Левые фермионы заданного поколения преобразуются по представлению $\bar{\mathbf{5}} \oplus \mathbf{10}$, а правые — по дуальному представлению, т. е. по $\mathbf{5} \oplus \overline{\mathbf{10}}$. В природе мы наблюдаем сразу три экземпляра этой структуры; другими словами, наблюдаемая нами разность представлений есть

$$\Delta = 3\,[(\bar{\mathbf{5}} \oplus \mathbf{10}) \ni (\mathbf{5} \oplus \overline{\mathbf{10}})]. \tag{39}$$

Если мы собираемся расширить наблюдаемую калибровочную группу G до большей группы $\bar{G}$, то важно понимать, что это повлечёт за собой предсказание новых сил. Для $SU(5)$, как и для большинства других теорий великого объединения, эти новые силы приводят к драматическим следствиям: протон, ранее считавшийся абсолютно устойчивым, оказывается неустойчивым с характерным временем жизни порядка 10^{32}—10^{45} лет, в зависимости от того, какую именно теорию великого объединения мы рассматриваем. Оценивая грубо наибольшее возможное количество вещества в выполнимом эксперименте в 10^4 тонн воды, мы получаем, что там содержится примерно 10^{34} протонов; таким образом, считая, что мы в состоянии зарегистрировать случайный распад одного протона в год, становится возможной экспериментальная проверка нижней границы интересующего нас диапазона. До сих пор пронаблюдать распад протона ещё не удалось.

Выражение (39) выглядит намного лучше, чем аналогичная формула для $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$, и главным образом это дает нам основания полагать, что в SU(5)-модели есть зерно истины. Тем не менее естественно поинтересоваться, нет ли ещё групп, кроме SU(5), которые «работали» бы не хуже, а может, даже и лучше. Одной из таких групп является SO(10), с единственным нетривиальным вложением SU(5) в SO(10) (тавтологическое представление группы SO(10), которое мы обозначаем через **10**, как представление SU(5) разлагается в $\mathbf{5}\oplus\overline{\mathbf{5}}$). Группа SO(10) имеет два комплексно-сопряжённых спинорных представления, каждое размерности 16; обозначим их **16** и $\overline{\mathbf{16}}$. Относительно SU(5) представление **16** разлагается как $\mathbf{1}\oplus\overline{\mathbf{5}}\oplus\mathbf{10}$, где **1** — тривиальное одномерное представление SU(5) (разумеется, это — вещественное представление). Представление $\overline{\mathbf{16}}$, конечно, разлагается как комплексно-сопряжённое к **16**, т. е. как $\mathbf{1}\oplus\mathbf{5}\oplus\overline{\mathbf{10}}$. Тем самым на уровне SO(10) мы можем переписать (39) так:

$$\Delta=3(\mathbf{16}\ominus\overline{\mathbf{16}}). \tag{40}$$

Это почти простейший вид, который может иметь нетривиальная разность представлений, если не обращать внимания на странный множитель 3; некоторые попытки объяснить его мы обсудим в следующем разделе. Тем не менее естественно задаться вопросом, есть ли ещё группы, кроме SO(10), которые были бы столь же подходящими?

Оказывается, есть, а именно исключительная группа E_6. Простейшее нетривиальное представление E_6 — это комплексное представление, которое мы будем обозначать по его размерности 27; дуальное к нему будет обозначаться $\overline{\mathbf{27}}$. Группа E_6 содержит в качестве подгруппы SO(10), и представление **27** разлагается относительно действия SO(10) как $\mathbf{16}\oplus\mathbf{10}\oplus\mathbf{1}$, где **16** и **10** — это спинорное и тавтологическое представления группы SO(10). Тем самым на уровне E_6 мы можем написать

$$\Delta=3(\mathbf{27}\in\overline{\mathbf{27}}). \tag{41}$$

Это третий и последний пример группы, которая кажется подходящей для теории великого объединения в размерности 4. Естественно, заподозрив о возможной роли E_6, поинтересоваться, а нельзя ли пойти ещё дальше и построить теорию великого объединения на базе самой большой исключительной группы, E_8. Но тут мы сталкиваемся со следующей проблемой: все представления E_8 вещественны, и мы автоматически получили бы $\Delta=0$. Тем не менее E_8 можно использовать для великого объединения, но для этого необходимы дополнительные построения, которые мы обсудим в следующем разделе.

Если попытаться подытожить наши знания в области физики самым экономным образом, то есть три действительно фундаментальных наблюдения: (i) пространство-время — это псевдориманово многообразие M, снабжённое метрическим тензором и подчиняющееся геометрическим законам; (ii) над M задано векторное расслоение X с неабелевой калибровочной группой G; (iii) фермионы — это сечение расслоения $(\hat{S}_+\otimes V_R)\oplus(\hat{S}_-\otimes V_{\tilde{R}})$, где R и $\tilde{R}$ неизоморфны и именно то, что они не изоморфны, объясняет, почему лёгкие фермионы действительно лёгкие; возможно, что это различие между R и $\tilde{R}$ берёт начало в разности представлений Δ в некоторой фундаментальной теории. Всё это необходимо дополнить пониманием того, что геометрические законы, определяемые метрическим тензором, калибровочные поля и фермионы надлежит интерпретировать квантовомеханически.

2. ФИЗИКА В РАЗМЕРНОСТИ ДЕСЯТЬ

На счету рассмотренной нами стандартной физической модели много достижений, но многие вопросы остаются здесь открытыми. Разумеется хотелось бы найти объяснение множителю 3 в (40) и (41). Как-то рука не поднимается просто взять и предположить, что в основе природы лежат три экземпляра представления $\mathbf{16}\ominus\overline{\mathbf{16}}$ группы SO(10) или три экземпляра представления $\mathbf{27}\ominus\overline{\mathbf{27}}$ группы E_6. Должна существовать какая-то более экономная структура на более элементарном уровне. Так же странно было бы изучать цепочку, ведущую от SU(5) через SO(10) к E_6, и не найти никакого способа продолжить ее до E_6. Хотелось бы иметь какое-нибудь естественное объяснение последовательных нарушений симметрии вдоль этой цепочки. Наконец, тот факт, что всё, что мы наблюдаем в эксперименте, — это *разность* характеров ((33) или (36)), подсказывает, что истинная основополагающая структура может быть намного богаче, чем кажется. Фундаментальное представление U может быть даже бесконечномерным; разумеется, в этом случае определение разности характеров $\Delta=U\ominus\bar{U}$ потребует надлежащей регуляризации.

Я опишу подход к некоторым из этих вопросов. Давайте предположим, что пространство-время — это не *четырёх*мерное псевдориманово многообразие, а многообразие более высокой размерности; на самом деле для теории струн предпочтителен случай размерности 10. Таким образом, рассмотрим десятимерное многообразие $\bar{M}$, ориентированное, со спинорной структурой и метрикой сигнатуры $(-+\ldots+)$. Одной из составляющих десятимерной теории будет эйнштейновское действие

$$S_{\text{ОТО}} = \frac{1}{16\pi G}\int_{\bar{M}} R \qquad (42)$$

На первый взгляд кажется невозможным представить себе, что мир десятимерен. Настолько «очевидно», что мир четырёхмерен (если добавить время к трём очевидным пространственным измерениям). Есть, однако, следующее решающее соображение. Мы привыкли думать, что существует «пустое пространство» и все другие физические состояния получаются добавлением частиц к этому «пустому пространству». Однако в действительности то, что мы считаем пустым пространством, — это лишь некоторое решение фундаментальных физических уравнений, играющее особую роль в нашей теории. Причина существования такого играющего особую роль решения состоит в том, что уровень энергии наших экспериментов очень мал по сравнению с массой Планка $M_{\text{пл}}$. Имеющиеся в нашем распоряжении орудия столь слабы, что мы можем внести лишь минимальные возмущения в то решение фундаментальных уравнений, в котором нам довелось родиться, каким бы оно ни было. Чтобы разговор был более конкретным, предположим, что фундаментальными уравнениями физики являются уравнения Эйнштейна, выводимые из действия (42). Пусть M^4 — четырёхмерное пространство Минковского, а K — некоторое компактное шестимерное многообразие с нулевым тензором Риччи. Предположим, что «вакуумное состояние» десятимерного мира есть

$$\bar{M} = M^4 \times K. \qquad (43)$$

Предположим, наконец, что радиус K (под которым я понимаю любой характерный размер K) сравним с естественной единицей длины в физике, т. е. с $R_{\text{пл}}$. По человеческим масштабам эта величина настолько мала, что K будет неотличимо от точки; наличие дополнительных размерностей не будет заметно ни в обычной жизни, ни даже в опытах на ускорителях. Тем самым $\bar{M}$ будет неотличимо от четырёхмерного пространства Минковского, аналогичным образом можно сымитировать четырёхмерную космологическую модель.

На интуитивном уровне это кажется убедительным. Чтобы достичь большей ясности, обсудим один из важных аспектов десятимерной физики, а именно десятимерное уравнение Дирака Итак, мы вводим десятимерный оператор Дирака

$$D_{(10)} = \sum_{i=1}^{10} \Gamma^i D_i. \qquad (44)$$

Мы можем записать его в виде

$$D_{(10)} = D_{(4)} + D_K, \qquad (45)$$

где $D_{(4)}$ — четырёхмерный оператор Дирака:

$$D_{(4)} = \sum_{i=1}^{4} \Gamma^i D_i, \qquad (46)$$

а D_K — оператор Дирака на K:

$$D_K = \sum_{j=5}^{10} \Gamma^j D_j. \qquad (47)$$

Мы хотим сейчас проанализировать десятимерное уравнение Дирака

$$0 = D_{(10)}\Psi(x^i, y^j), \qquad (48)$$

где x^i и y^j, $i = 1, \ldots, 4$, $j = 5, \ldots, 10$, — координаты на M^4 и K соответственно, а Ψ — десятимерное спинорное поле.

В первую очередь давайте выясним, спинором какого типа должно быть Ψ. Неприводимое представление алгебры Клиффорда в размерности 10 разлагается в сумму двух неприводимых представлений группы Лоренца SO(1, 9). Они различаются собственными значениями оператора

$$\Gamma^{(10)} = \Gamma^1\Gamma^2 \ldots \Gamma^{10}. \qquad (49)$$

В размерности 10, в отличие от размерности 4, оба неприводимых спинорных представления SO(1,9), которые я обозначаю через $S_+^{(10)}$ и $S_-^{(10)}$, вещественны. Тем самым CPT-теорема позволяет нам рассмотреть теорию, в которой спинорное поле Ψ преобразуется по какому-то одному из этих представлений, скажем по $S_+^{(10)}$. Другими словами,

$$\Gamma^{(10)}\Psi = +\Psi. \qquad (50)$$

На самом деле соображения десятимерной суперсимметрии и теория струн заставляют нас рассмотреть именно этот вариант. Если нас интересует четырёхмерная физика, то мы должны разложить спинорное представление SO(1, 9) на неприводимые представления группы SO(1,3) × SO(6), где SO(1,3) — группа Лоренца для M^4, а SO(6) — структурная группа касательного расслоения к K. В это разложение будут входить только спинорные представления групп SO(1, 3) и SO(6), поскольку пространство представления для десятимерной алгебры Клиффорда будет, разумеется, и пространством представления для четырёхмерной и шестимерной подалгебр Клиффорда. Точное разложение легко получить, введя операторы, аналогичные (49):

$$\Gamma^{(4)} = i\Gamma^1\Gamma^2 \ldots \Gamma^4, \quad \Gamma^K = -i\Gamma^5\Gamma^6 \ldots \Gamma^{10}. \tag{51}$$

(Множители $\pm i$ вводятся по традиции, для того чтобы собственные значения операторов $\Gamma^{(4)}$ и Γ^K были равны ± 1.) Эти операторы подчиняются очевидному соотношению:

$$\Gamma^{(10)} = \Gamma^{(4)}\Gamma^K. \tag{52}$$

Следовательно, в неприводимом представлении группы SO(1, 9), в котором $\Gamma^{(10)} = +1$, операторы $\Gamma^{(4)}$ и Γ^K равны:

$$\Gamma^{(4)} = \Gamma^K. \tag{53}$$

Это значит, что разложение SO(1, 9)-спинорного представления $S_+^{(10)}$ положительной киральности как представления группы SO(1, 3)×SO(6) будет иметь вид

$$S_+^{(10)} = (S_+^{(4)} \otimes S_+^K) \oplus (S_-^{(4)} \otimes S_-^K). \tag{54}$$

Здесь $S_\pm^{(4)}$ и $S_\pm^K$ — спинорные представления положительной и отрицательной киральности групп SO(1, 3) и SO(6) соответственно.

Вернёмся к нашей проблеме. Мы хотим решить уравнение (48) при помощи разделения переменных, используя разложение (45). Поскольку $D_{(4)}$ и D_K не коммутируют, а антикоммутируют, стандартная процедура разделения переменных требует небольшого изменения. Удобно ввести оператор

$$D'_K = \Gamma^{(4)} D_K, \tag{55}$$

который коммутирует с $D_{(4)}$. Оператор D'_K унитарно эквивалентен D_K и потому имеет тот же спектр (ибо матрицы $\Gamma^{(4)}\Gamma^j$, $j = 5, \ldots, 10$, образуют ту же алгебру Клиффорда, что и матрицы Γ^j, а известно, что неприводимое представление этой алгебры единственно). Введём полное семейство собственных функций χ_m для D'_K:

$$D'_K \chi_m = \lambda_m \chi_m. \tag{56}$$

Далее, запишем

$$\Psi(x^i, y^j) = \sum_m \psi_m(x^i) \otimes \chi_m(y^j). \tag{57}$$

В результате уравнение (48) сводится к уравнению

$$0 = (D_{(4)} + \Gamma^{(4)}\lambda_m)\psi_m, \tag{58}$$

или, что эквивалентно,

$$0 = (D'_{(4)} + \lambda_m)\psi_m, \tag{59}$$

где мы ввели

$$D'_{(4)} = \Gamma^{(4)} D_{(4)}. \tag{60}$$

Оператор $D'_{(4)}$ унитарно эквивалентен $D_{(4)}$ (поскольку $\Gamma^{(4)}\Gamma^i$ порождают алгебру Клиффорда). Уравнение (59) равносильно уравнению Дирака (24) для тяжёлых фермионов.

Тем самым мы получили следующую важную информацию: собственные значения оператора Дирака D_K на компактном многообразии K соответствуют в четырёхмерных терминах массам фермионов ψ_m. Конечно, таких собственных значений бесконечно много. Но поскольку D_K — эллиптический оператор на компактном многообразии, он имеет лишь конечное число нулевых собственных значений — собственное значение 0 имеет конечную кратность. Ненулевые собственные значения D_K будут порядка $1/R$, где R — радиус K. Поскольку мы предполагаем, что радиус K — величина порядка планковской длины (5), ненулевые собственные значения D_K соответствуют фермионам с планковскими массами — фермионам, которые мы, безусловно, не сможем обнаружить экспериментально. Доступная эксперименту четырёхмерная физика будет определяться нулевыми собственными значениями D_K, соответствующими безмассовым частицам в размерности 4. Поскольку нулевых собственных значений конечное число, для всех практических применений десятимерная теория будет выглядеть как четырёхмерная с конечным числом фермионных полей. Структуры именно такого типа мы обсуждали в предыдущем разделе. Таким образом, мы достигли своей цели — показать, что теория, являющаяся на самом деле десятимерной, может выглядеть как четырёхмерная для наблюдателя, который в своих экспериментах ограничен низкими уровнями энергии.

Что ещё важно отметить — очень интересным является то, что в рамках этой картины наблюдаемые четырёхмерные фермионы получаются как нулевые моды оператора Дирака D_K. Появление нулевых собственных значений у эллиптических операторов, таких как оператор Дирака, не может быть случайным; оно связано с соответствующими топологическими инвариантами. Таким образом, здесь (как и во многих других ситуациях), фундаментальные физические вопросы приводят нас к вопросу о топологии K.

Простейший топологический инвариант, с помощью которого можно предсказать существование нулевых собственных значений эллиптического оператора, — это его *индекс*. Обозначим через n_+ и n_- число собственных значений оператора Дирака D_K положительной и отрицательной киральности соответственно (т. е. с $\Gamma_K = \pm 1$). Разность $n_+ - n_-$ называется индексом оператора D_K и, как легко показать, является топологическим инвариантом. До сих пор неявно подразумевалось, что мы рассматриваем только уравнение Дирака (десятимерное) для взаимодействия фермионных полей с гравитационным, т. е. с геометрической структурой пространства-времени, без участия полей Янга — Миллса. В этом случае из теоремы Атьи — Зингера об индексе (или из различных других, более элементарных соображений) легко выводится, что в размерности 6 (и, более общо, в размерности $4k+2$) индекс оператора Дирака обращается в нуль. Это не должно вызывать удивления: в предыдущем разделе нам удалось найти объяснение существованию безмассовых фермионов лишь при наличии группы Янга — Миллса. Давайте поэтому обобщим наше обсуждение и вновь включим в него калибровочные поля Янга — Миллса.

Предыдущие рассмотрения носили весьма общий характер. С учётом достижений последних лет [12, 14] в теории струн, случай, в пользу которого больше всего свидетельств, — это десятимерная теория с калибровочной группой $E_8 \times E_8$. Для простоты я ограничусь рассмотрением просто E_8. Фермионы будут преобразовываться по присоединённому представлению E_8. В десятимерной теории, включающей в себя как поля Янга—Миллса, так и гравитационное поле, для описания вакуумного состояния недостаточно указать вакуумное многообразие. Необходимо также указать векторное расслоение X над пространством-временем со структурной группой E_8 и связность A в нём.

Каков естественный выбор X?[1] При изучении любого шестимерного многообразия K у нас всегда есть в распоряжении касательное расслоение T. Разумеется, оно наделено связностью Леви—Чивиты. Структурной группой T в общем случае будет SO(6). Всякое вложение SO(6) в E_8 даёт нам канонический способ построить из касательного расслоения T с его римановой связностью некоторое E_8-расслоение со связностью. Существует вложение SO(6) в E_8, являющееся в определённом смысле минимальным среди всех таких вложений. Оно получается из цепочки

$$\mathrm{SO}(6) \times \mathrm{SO}(10) \subset \mathrm{SO}(16) \subset E_8. \tag{61}$$

---

[1] Описываемая ниже конструкция рассматривалась в [4, 31].

Здесь SO(16) — максимальная подгруппа в E_8, а SO(6) $\times$ SO(10) — максимальная подгруппа в SO(16)[1]. Оказывается, что вложение (61) приводит к интересной картине четырёхмерной физики.

Здесь мы сталкиваемся с уже упоминавшимся в предыдущем разделе нарушением симметрии. Что будет интерпретироваться физиком низких энергий как калибровочная группа? Физик, имеющий дело с низкими энергиями, «замкнут» в мире с E_8-расслоением X и некоторой фиксированной связностью A в нём и не может сильно изменить его. Калибровочное преобразование из E_8, не сохраняющее A, не будет симметрией этого мира; оно будет переводить его в некоторый другой мир с некоторой другой связностью A'. Такое калибровочное преобразование затрагивало бы степени свободы с очень высокими уровнями энергии, недоступными наблюдателю, ограниченному низкими уровнями энергии; в этом причина того, что такому наблюдателю мир кажется четырёхмерным. Роль калибровочной группы для физика низких энергий будет играть подгруппа в E_8, которая действует на расслоении X, сохраняя связность A. При «типичном» выборе X и A эта подгруппа будет тривиальной. Но если расслоение X построено из касательного расслоения с помощью вложения (61) группы SO(6) в E_8, то подгруппой калибровочных преобразований, сохраняющих связность A, будет группа $G = \mathrm{SO}(10)$, коммутирующая с SO(6).

Это многообещающее продвижение, поскольку, хотя сама E_8 не подходит в качестве калибровочной группы для четырёхмерной теории (она имеет лишь вещественные представления и потому приводила бы к $\Delta = 0$), группа SO(10), как мы выяснили в предыдущем разделе, является одним из естественных кандидатов. Давайте вычислим разность представлений Δ, которая получается при таком подходе. Для этого необходимо разложить присоединённое представление E_8 на неприводимые компоненты относительно действия группы SO(6)$\times$SO(10). Присоединённое представление E_8, которое мы будем обозначать **248**, можно разложить относительно действия SO(6) $\times$ SO(10). Общий вид этого разложения такой:

$$\mathbf{248} = \sum_i L_i \otimes R_i, \tag{62}$$

где L_i и R_i — некоторые представления групп SO(6) и SO(10) соответственно. Для каждого L_i мы имеем соответствующий

---

[1] Напомним, что мы работаем на уровне алгебр Ли, не уточняя глобальную структуру различных групп; фигурирующая в (61) группа SO(16) — это в действительности $\mathrm{spin}(16)/Z_2$, и т. д.

оператор Дирака $D_K^{(i)} = D_K^{L_i}$, действующий на фермионах, преобразующихся под действием SO(6) как L_i. Безмассовые фермионы в размерности 4, преобразующиеся под действием SO(10) как R_i, получаются как нулевые моды $D_K^{(i)}$. Ввиду (53), нулевые собственные значения $D_K^{(i)}$ с $\Gamma^K = +1$ (соотв. -1) имеют $\Gamma^{(4)} = +1$ (соотв. -1). Обозначим через δ_i индекс $D_K^{(i)}$. Это разность между числом нулевых собственных значений $D_K^{(i)}$ с $\Gamma^K = 1$ и $\Gamma^K = -1$, или, что то же самое, разность между числом нулевых собственных значений $D_K^{(i)}$ с $\Gamma^{(4)} = +1$ и $\Gamma^{(4)} = -1$. Тем самым в разности представлений Δ для безмассовых фермионов представление R_i группы SO(10) входит с коэффициентом δ_i. В итоге получаем

$$\Delta = \sum_i \delta_i R_i. \tag{63}$$

Разберемся с (63) поподробнее. Введём некоторые обозначения. Если M — представление SO(6), а N — представление SO(10), то будем обозначать их тензорное произведение $M \otimes N$ через (M, N). Для представлений SO(6) и SO(10) в качестве обозначения будем использовать их размерность. Интересующие нас представления SO(6) — это присоединенное **15**, тавтологическое **6** и два спинорных **4** и $\bar{\mathbf{4}}$. Из представлений SO(10) нас интересуют присоединённое **45**, тавтологическое **10** и два спинорных **16** и $\overline{\mathbf{16}}$. Присоединённое представление E_8 под действием SO(6)×SO(10) разлагается следующим образом:

$$\mathbf{248} = (\mathbf{15}, \mathbf{1}) \oplus (\mathbf{1}, \mathbf{45}) \oplus (\mathbf{6}, \mathbf{10}) \oplus (\mathbf{4}, \mathbf{16}) \oplus (\bar{\mathbf{4}}, \overline{\mathbf{16}}). \tag{64}$$

Если L_i — вещественное представление SO(6), то (по теореме Атьи—Зингера об индексе или по многим другим, более элементарным соображениям) индекс δ_i равен нулю. Далее, если L_i и $L_{\bar{i}}$ комплексно сопряжены, то $\delta_i = -\delta_{\bar{i}}$. Единственные комплексные представления SO(6) в (64) — это **4** и $\bar{\mathbf{4}}$. Таким образом, (63) сводится к

$$\Delta = \delta_4 (\mathbf{16} \ominus \overline{\mathbf{16}}), \tag{65}$$

Сравнивая с (40), видим, что это согласуется с наблюдениями, если считать, что в природе $\delta_4 = \pm 3$[1].

Что касается реального значения δ_4, то это один из фундаментальных топологических инвариантов шестимерного многообразия. Комплекс де Рама дифференциальных форм можно построить из тензорного произведения двух спинорных представлений. Поскольку **4** есть одно из спинорных представлений SO(6), спиноры на K со значениями в представлении **4** группы SO(6) эквивалентны определённому набору дифференциальных форм. Поэтому δ_4 можно выразить через эйлерову характеристику χ и сигнатуру Хирцебруха σ — это именно те топологические инварианты, связанные с проблемой индекса, которые могут быть получены из комплекса де Рама. На самом деле в размерности 6 (и вообще в размерности $4k + 2$) $\sigma = 0$ и

$$\delta_4 = \chi/2. \tag{66}$$

Таким образом, мы видим, как связать наблюдаемую разность представлений Δ с чем-то более фундаментальным, а именно с топологией K. Хотя нам и не удалось полностью объяснить появление странного множителя 3 в (40), всё-таки мы отчасти сняли с него покров таинственности. Причина, по которой природа повторяет саму себя в нескольких «фермионных поколениях», состоит просто в том, что *нет* никаких причин, почему шестимерное многообразие обязано иметь эйлерову характеристику ±2. При подходящем выборе K мы можем, начав, как мы начали, с единственного спинорного поля Ψ в размерности 10, получить любое наперёд заданное число фермионных поколений в размерности 4.

Я попытался дать представление о том, как из геометрических и топологических свойств K можно вывести свойства четырёхмерной физики. Имеются многочисленные другие примеры, но для иллюстрации достаточно и этих. Хотелось бы тем не менее подчеркнуть, что, хотя мы и предположили, что *вакуумное состояние* имеет вид $M^4 \times K$, но общие физические возмущения не сохраняют этой структуры прямого произведения. Считается, что основные законы — это десятимерные законы, получаемые из (42) или (более вероятно) из некоторой более тонкой десятимерной теории; приближённая четырёхмерная картина становится возможной лишь потому, что вакуумное состояние допускает четырёхмерные, а не десятимерные симметрии из группы Пуанкаре. Почему это так и каким должно быть K — остаётся загадкой.

3. КВАНТОВАЯ ТЕОРИЯ ПОЛЯ НА РИМАНОВОЙ ПОВЕРХНОСТИ

В двух предыдущих разделах мы обрисовали некоторые ключевые составляющие физики в *классических* терминах, сделав мельком оговорку, что теорию Янга—Миллса, равно как

---

[1] Знак мы определить не можем, поскольку различие между представлениями **16** и $\overline{\mathbf{16}}$ — вопрос соглашения.

и всё остальное, надлежит интерпретировать квантовомеханически. В действительности это приводит к намного более богатой и обширной теории[1].

Квантовая теория поля пока ещё не стала важным орудием в современной математике, но многое указывает, что наступает период, когда это изменится. И в теории аффинных алгебр Ли, и в алгебраической геометрии в последнее время стали играть важную роль структуры, известные в квантовой теории поля. (Быть может мы услышим здесь об этом в докладе Г. Фальтингза «Новые результаты в арифметической алгебраической геометрии»[2].) Я попытаюсь сейчас дать очень краткое введение в квантовую теорию поля, выделяя те её аспекты, которые необходимы для понимания теории струн и которые, по-видимому, связаны с упомянутыми мной разделами математики.

Пусть Σ — риманова поверхность, возможно с краем, а φ — вещественнозначная функция на Σ, т. е. отображение из Σ в $\mathbb{R}$ (вещественные числа). Положим

$$I(\varphi)=\frac{1}{2}\int_{E} d\varphi \wedge * d\varphi. \tag{67}$$

Это выражение называется «функционалом действия свободной бозонной теории поля». Пусть f — вещественнозначная функция на $\partial\Sigma$ (край Σ), а $\Omega_f(\Sigma)$ — пространство непрерывных вещественнозначных функций

$$\varphi:\Sigma\to\mathbb{R}, \tag{68}$$

ограничение которых на $\partial\Sigma$ совпадает с f.

Если нам дана какая-либо настоящая метрика (а не класс конформной эквивалентности метрик) на Σ, то аффинное пространство $\Omega_f(\Sigma)$ снабжается естественной римановой метрикой. Для $\varphi, \varphi' \in \Omega_f(\Sigma)$ мы полагаем

$$|\varphi-\varphi'|^2=\int_{\Sigma}(\varphi-\varphi')^2. \tag{69}$$

В конечномерном случае риманова метрика всегда индуцирует некоторую меру («квадратный корень из детерминанта метрики»). В надежде, что это остаётся в силе и в бесконечномерном случае, можно попробовать ввести интеграл

$$Z_f(\Sigma)=\int_{\Omega_f(\Sigma)} e^{-I(\varphi)}. \tag{70}$$

---

[1] В качестве введения см., например, [10, 16].

[2] В докладе Фальтингза (с. 105—113 настоящего сборника) об этом ничего нет, зато в докладе Доналдсона (с. 88—104) кое-что есть. — *Прим. изд. ред.*

Давайте посмотрим, что для этого нужно. Поскольку $I(\varphi)$ — квадратичный функционал от φ, интеграл (70) похож на конечномерный гауссов интеграл

$$Z(A)=\int_{-\infty}^{\infty} d\varphi_1\, d\varphi_2 \ldots d\varphi_n e^{-(\varphi,\, A\varphi)/2}. \tag{71}$$

Здесь φ_i — координаты в n-мерном эвклидовом пространстве, а A — положительно-определённая квадратичная форма от n переменных. Как хорошо известно, значение интеграла (71) равно

$$Z(A)=(2\pi)^{n/2}\frac{1}{\sqrt{\det A}}=(2\pi)^{n/2}\prod_i \lambda_i^{-1/2}, \tag{72}$$

где λ_i — собственные значения A. При попытке обобщить формулу (72) на бесконечномерный случай ($n\to\infty$) не приходится ждать ничего хорошего от множителя $(2\pi)^{n/2}$. По этой причине, а также из-за трудностей с определением $\det A$ в бесконечномерном случае мы не в состоянии определить интеграл $Z_f(\Sigma)$ для заданных f и Σ, но отношениям вида $Z_{f_1}(\Sigma_1)/Z_{f_1}(\Sigma_2)$ смысл придать можно.

Что будет аналогом формы A в интеграле (70)? Очевидно, определённый равенством (67) интеграл $I(\varphi)$ равен

$$I(\varphi)=\frac{1}{2}\int_{\Sigma}\varphi\,\Delta\varphi, \tag{73}$$

где $\Delta={}^*d^*d$ — обычный оператор Лапласа. Рассмотрим сначала случай, когда Σ — многообразие без края. Тогда Δ — положительно-определённый оператор с дискретным спектром:

$$\Delta\varphi_i=\lambda_i\varphi_i, \quad i=1,\ 2,\ 3,\ \ldots. \tag{74}$$

Имеется одно нулевое собственное значение, отвечающее постоянной функции 1, а все остальные положительны. Интеграл в (70) должен быть точным бесконечномерным аналогом интеграла (71), поэтому нам нужно определить регуляризованный вариант произведения

$$\prod_i \lambda_i^{-1/2}. \tag{75}$$

Прежде чем пытаться определить это бесконечное произведение, необходимо исключить нулевое собственное значение. Итак, мы хотим придать смысл выражению

$$\prod_{\lambda_i\neq 0}\lambda_i^{-1/2}. \tag{76}$$

Это можно сделать регуляризацией с использованием дзета-функции. Положим

$$\zeta(s) = \sum_{\lambda_i \neq 0} \lambda_i^{-s}. \tag{77}$$

Этот ряд сходится, если вещественная часть s достаточно велика. Он определяет мероморфную функцию от s, и можно показать, что эта функция регулярна в нуле. Поэтому мы полагаем

$$\prod_{\lambda_i \neq 0} \lambda_i^{-1/2} = \exp(\zeta'(0)/2). \tag{78}$$

Таким образом, для случая $\partial\Sigma = 0$ мы вводим такое определение:

$$Z(\Sigma) = \exp(\zeta'(0)/2). \tag{79}$$

Теперь давайте выведем формулу, аналогичную (79), для более общего случая $\partial\Sigma \neq 0$. Разберём сперва случай $f = 0$. Тут едва ли есть что-либо новое: если $f = 0$, то (70) представляет собой интеграл по функциям φ, обращающимся в 0 на $\partial\Sigma$. При граничных условиях $\varphi|_{\partial\Sigma} = 0$ лапласиан Δ имеет дискретный спектр с положительными собственными значениями, и мы, как и раньше, полагаем

$$Z_{f=0}(\Sigma) = \exp(\zeta'(0)/2). \tag{80}$$

При $\partial\Sigma \neq 0$ и граничных условиях $\varphi = 0$ на $\partial\Sigma$ оператор Δ является строго положительно-определённым и потому при определении (80) не возникает необходимости исключать нулевое собственное значение. Остаётся обобщить определение (80) на случай $f \neq 0$. Для этого вспомним классическую теорему о том, что функционал $I(\varphi)$ с наложенными граничными условиями $\varphi|_{\partial\Sigma} = f$ имеет единственный экстремум. Обозначим экстремальную функцию φ_0 и запишем $\varphi = \varphi_0 + \varphi'$. Поскольку $I(\varphi)$ — квадратичный функционал от φ и $\varphi = \varphi_0$ — стационарная точка, имеем

$$I(\varphi) = I(\varphi_0) + I(\varphi'). \tag{81}$$

Если φ и φ_0 лежат в $\Omega_f(\Sigma)$, то $\varphi' = \varphi - \varphi_0$ лежит в $\Omega_0(\Sigma)$. Возвращаясь к интегралу (70), видим, что замена переменных $\varphi \to \varphi'$ переводит интеграл по $\Omega_f(\Sigma)$ в интеграл по $\Omega_0(\Sigma)$, и мы получаем

$$Z_f(\Sigma) = e^{-I(\varphi_0)} Z_0(\Sigma) = e^{-I(\varphi_0)} e^{\zeta'(0)/2}. \tag{82}$$

Разумеется, в нашем подходе (79) и (82) просто определения. Они становятся теоремами в рамках подходящей теории бесконечномерного интегрирования.

Здесь необходимо сделать несколько замечаний. В первую очередь, выше мы обсуждали лишь *вещественнозначные* функции на Σ, т. е. отображения $\Sigma \to \mathbb{R}$. Более общо, можно взять риманово многообразие X с метрическим тензором γ и рассмотреть отображения из Σ в X. Зафиксировав отображение f: $\partial\Sigma \to X$, определим $\Omega_f(\Sigma, X)$ как пространство непрерывных отображений из Σ в X, ограничение которых на $\partial\Sigma$ совпадает с f. Выбрав локальные координаты φ_i на X, мы можем локально записать произвольное отображение Φ: $\Sigma \to X$ с помощью вещественнозначных функций φ^i на Σ. Обобщим (67) так:

$$I(\Phi) = \int_{\Sigma} \gamma_{ij}(\Phi)\, d\varphi^i \wedge * d\varphi^j. \tag{83}$$

Вместо (70) мы хотим рассмотреть интеграл

$$Z_f(\Sigma, X) = \int_{\Omega_f(\Sigma, X)} e^{-I(\Phi)}. \tag{84}$$

Заметим, что интеграл (70) был интегралом по аффинному пространству и его можно было явно выразить через дзета-функцию по формулам (79) и (82). Формула (70) — это один из путей определить то, что физики называют теорией свободного (бозонного) поля. Если X не плоско, то (84) — это интеграл по нелинейному пространству и иметь с ним дело довольно трудно. Интеграл (84) — это пример того, что физики называют нелинейной теорией поля или теорией поля с взаимодействиями. Квантовая теория поля, определяемая интегралом (84), известна как нелинейная сигма-модель и особенно важна в теории струн. Определение этого интеграла требует намного большего, чем просто регуляризации с использованием дзета-функции, которой мы обошлись в случае теории свободного поля. На самом деле не существует удовлетворительного определения интеграла (84) для произвольного X. Вообще говоря, этот интеграл можно определить для многообразий X с положительным или нулевым тензором Риччи, но не для многообразий с отрицательным тензором Риччи. Про нелинейную сигма-модель известно многое, хотя доказана от этого многого лишь небольшая часть. К сожалению, для того чтобы объяснить здесь, что же известно про интеграл (84), потребовалось бы долго объяснять такие понятия, как ренормгруппа, $1/N$-разложение, факторизуемые S-матрицы и т. д.

Следующее замечание состоит в том, что функционал действия и подынтегральное выражение в (70) зависят только от конформного класса метрического тензора на Σ — другими словами, они зависят только от комплексной структуры Σ. Однако чтобы определить — хотя бы формально — меру, по которой проводится интегрирование в (70), нам необходима настоящая риманова метрика на Σ, а не конформный класс римановой структуры. Таким образом, наш окончательный ответ (74) не будет функцией на пространстве модулей римановых поверхностей Σ — он не инвариантен относительно конформных замен метрики на Σ. Тем не менее можно показать, что при конформных заменах метрики на Σ выражение (79) преобразуется довольно простым образом. Это было проинтерпретировано Куилленом как определение метрики на некотором голоморфном линейном расслоении на пространстве модулей [23]; эта структура играет важную роль в исследованиях, о которых пойдёт речь в докладе Фальтингза.

Следующий существенный шаг в объяснении квантовой теории поля, как её понимают физики, состоит в том, чтобы объяснить связь между только что описанными нами фейнмановскими интегралами по траекториям и более старой интерпретацией квантовой теории поля с точки зрения гильбертовых пространств. Если граница поверхности Σ не пуста, то она состоит из нескольких попарно непересекающихся окружностей S_i. Рассмотрим случай одной-единственной окружности S. Пусть $\Omega(S)$ — пространство непрерывных вещественнозначных функций на S. Как и при нашем обсуждении Σ, риманова структура на S индуцирует риманову структуру и, следовательно, меру на $\Omega(S)$. Имея такую меру, мы можем говорить о квадратично-интегрируемых комплекснозначных функциях на $\Omega(S)$ т. е. о функциях ψ с

$$\int\limits_{\Omega(S)} |\psi|^2 < \infty.$$

Конечно, при этом необходимо строго определить меру на $\Omega(S)$, точно так же как нам потребовалась регуляризация с использованием дзета-функции (или какая-нибудь другая) при обсуждении интегралов по $\Omega(S)$. Хотя это нетрудно сделать вполне аккуратно (по крайней мере в случае теории свободного поля), в дальнейшем я буду действовать чисто формально. Обозначим гильбертово пространство квадратично-интегрируемых функций на $\Omega(S)$ через H_S. Оно известно как гильбертово пространство обсуждаемой нами квантовой теории поля.

Теперь я хотел бы определить некоторый линейный оператор на H_S, известный как «гамильтониан» этой теории. Рассмотрим фейнмановский интеграл (70) на римановой поверхности с краем Σ весьма специального вида — на цилиндре (см. рис. 1) со стандартной плоской метрикой. Его граница состоит из двух компонент, S_1 и S_2. Будем представлять себе S_1 и S_2 как два экземпляра «одной и той же» окружности S, разнесённые по вертикали на расстояние («собственное время») β. Соответственно при необходимости будем свободно отождествлять пространства функций $\Omega(S_1)$ и $\Omega(S_2)$, равно как и гильбертовы пространства H_{S_1} и H_{S_2}. Таким образом, для $\psi_1 \in H_{S_1}$ и $\psi_2 \in \in H_{S_2}$ мы имеем скалярное произведение

$$\langle\psi_1 | \psi_2\rangle = \int\limits_{\Omega(S)} \psi_2^* \psi_1. \tag{86}$$

Вещественнозначная функция f на границе Σ — это то же самое, что пара функций f_1 и f_2 на компонентах границы; будем писать $f = (f_1, f_2)$. Мы уже обсуждали фейнмановский интеграл на Σ с заданными граничными условиями (см. (70)):

$$Z_{(f_2, f_1)}(\Sigma) = \int\limits_{\Omega(f_2, f_1)} e^{-I}. \tag{87}$$

Теперь сделаем следующий шаг и определим некоторый билинейный функционал на гильбертовых пространствах H_{S_1} и H_{S_2}. Для $\psi_1 \in H_{S_1}$ и $\psi_2 \in H_{S_2}$ положим

$$R(\psi_2, \psi_1) = \int\limits_{\Omega(S_2)} \psi_2^*(f_2) \int\limits_{\Omega(S_1)} \psi_1(f_1)\, Z_{(f_2, f_1)}(\Sigma). \tag{88}$$

Функционал $R(\psi_2, \psi_1)$, очевидно, линеен по ψ_1 и антилинеен по ψ_2, и потому его можно записать в виде

$$R(\psi_2, \psi_1) = \langle\psi_2 | T_\beta | \psi_1\rangle \tag{89}$$

для некоторого линейного оператора T_β (напомним, что β — это высота цилиндра Σ, см. рис. 1), это просто тот линейный оператор на $\Omega(S)$, ядром которого служит

$$T_\beta(f_2, f_1) = Z_{(f_2, f_1)}(\Sigma). \tag{90}$$

Теперь предположим, что мы склеиваем, как на рис. 2, два цилиндра Σ_1 и Σ_2 высоты β_1 и β_2 по общей компоненте границы S, образуя цилиндр Σ высоты $\beta = \beta_1 + \beta_2$. Граница цилиндра Σ_1 состоит из двух компонент, S_1 и S, и граница Σ_2 состоит из двух компонент, S и S_2. Предположим, что мы фиксировали вещественнозначные функции f_1, f_2 и f на S_1, S_2 и S соответственно; рассмотрим произведение

$$T_{\beta_2}(f_2, f)\, T_{\beta_1}(f, f_1) = Z_{(f_2, f)}(\Sigma_2)\, Z_{(f, f_1)}(\Sigma_1). \tag{91}$$

Это произведение есть интеграл по вещественнозначным функциям на Σ, принимающим заданные значения на трёх окружностях S_1, S_2 и S (другими словами, по функциям, ограничение которых на S_1, S_2 и S совпадает с f_1, f_2 и f). Если мы *не* хотим фиксировать значения наших функций на S, мы можем избежать этого, проинтегрировав (91) по $\Omega(S)$. Итак, рассмотрим

$$\int_{\Omega(S)} T_{\beta_2}(f_2, f) T_{\beta_1}(f, f_1) = \int_{\Omega(S)} Z_{(f_2, f)}(\Sigma_2) Z_{(f, f_1)}(\Sigma_1). \qquad (92)$$

Здесь мы интегрируем по вещественнозначным функциям, значения которых заданы лишь на $\partial\Sigma$, — другими словами, мы рассматриваем наш основной интеграл (70). Поэтому значение (92) совпадает с ядром $T_\beta(f_2, f_1)$, отвечающим, согласно (90), интегрированию по вещественнозначным функциям на цилиндре Σ высоты $\beta = \beta_1 + \beta_2$:

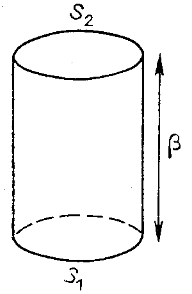


Рис. 1. Цилиндр со стандартной плоской метрикой. Его граница состоит из двух компонент — окружностей S_1 и S_2 длины 2π. «Высота» цилиндра обозначена через β.

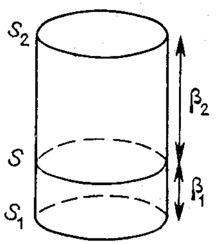


Рис. 2. Цилиндр высоты $\beta_1 + \beta_2$, полученный склеиванием цилиндров Σ_1 и Σ_2 высоты β_1 и β_2 соответственно. Граница цилиндра Σ_1 состоит из окружностей S_1 и S, а граница Σ_2 — из окружностей S и S_2.

$$T_{\beta_1+\beta_2}(f_2, f_1) = \int_{\Omega(S)} T_{\beta_2}(f_2, f) T_{\beta_1}(f, f_1). \qquad (93)$$

Это уравнение есть просто закон умножения в полугруппе:

$$T_{\beta_1+\beta_2} = T_{\beta_2} T_{\beta_1}. \qquad (94)$$

Из равенства (94) вытекает, что

$$T_\beta = e^{-\beta H} \qquad (95)$$

для некоторого линейного оператора H. Этот оператор известен как «гамильтониан» квантовой теории поля. В обсуждаемой нами теории свободного поля легко написать явное выражение

для H. Однако рассуждения, с помощью которых мы дали фейнмановскому интегралу по траекториям интерпретацию на языке гильбертовых пространств и получили гамильтониан, являются намного более общими.

Нам бы хотелось вычислить след оператора $e^{-\beta H} = T_\beta$. Для этого, разумеется, надо проинтегрировать ядро $T_\beta(f_2, f_1)$ вдоль диагонали:

$$\operatorname{Tr} e^{-\beta H} = \int_{\Omega(S)} T_\beta(f, f). \qquad (96)$$

Фактически (96) — это не что иное, как интеграл по вещественнозначным функциям на цилиндре Σ (см. рис. 1), удовлетворяющим тому условию, что у каждой из них ограничения на две компоненты границы совпадают между собой, а в остальном произвольным. Отождествив компоненты границы Σ, мы получим риманову поверхность (без края) рода 1, которую обозначим $\bar{\Sigma}$. Интеграл (96) представляет собой просто интеграл по траекториям (70) на $\bar{\Sigma}$:

$$\operatorname{Tr} e^{-\beta H} = \iint_{\Omega(\bar{\Sigma})} e^{-I}. \qquad (97)$$

Прежде чем обсуждать роль интеграла (97), произведём необходимое обобщение. Для этого вернёмся к гильбертову пространству H_S, связанному с окружностью S. Операция поворота окружности на угол θ очевидно, естественным образом действует на вещественнозначных функциях $\Omega(S)$ и, следовательно, на гильбертовом пространстве H_S. Тем самым мы имеем линейный оператор R_θ, представляющий действие поворота на H_S. (Мы будем использовать тот же символ R_θ и для обозначения действия этого поворота на $\Omega(S)$.) Снова перед нами очевидный закон умножения, определяющий полугруппу: $R_{\theta_1+\theta_2} = R_{\theta_2} R_{\theta_1}$, поэтому

$$R_\theta = e^{i\theta P}, \qquad (98)$$

где P — некоторый линейный оператор. Он известен как оператор импульса квантовой теории поля. Опять же не составляет труда выписать явную формулу для P. Легко видеть, что P коммутирует с H. Теперь мы хотим обобщить (97) и вычислить след оператора $T_{\beta,\theta} = e^{-\beta H + i\theta P}$. Ядро этого оператора есть просто

$$T_{\beta,\theta}(f_2, f_1) = T_\beta(f_2, R_\theta f_1). \qquad (99)$$

Здесь $T_\beta(f_2, f_1)$ — ядро оператора $e^{-\beta H}$, подробно обсуждавшееся выше. Формула (99) справедлива потому, что действие оператора $e^{i\theta P}$ в правой части равенства $e^{-\beta H+i\theta P}=e^{-\beta H}e^{i\theta P}$ сводится к замене f_1 на $R_\theta f_1$. Вычисляя след, как и раньше, интегрированием вдоль диагонали, получаем

$$\operatorname{Tr} e^{-\beta H+i\theta P}=\int\limits_{\Omega(S)} T_\beta(f,\ R_\theta f). \tag{100}$$

Как и в предшествующем обсуждении случая $\theta=0$ равенство (100) допускает простую интерпретацию. С использованием обозначения (70), правая часть (100) есть

$$\int\limits_{\Omega(S)} Z_{(f,\,R_\theta f)}(\Sigma)=\int\limits_{\Omega(S)}\ \int\limits_{\Omega_{(f,\,R_\theta f)}(\Sigma)} e^{-I}. \tag{101}$$

Это легко описать словами. Правая часть (101) — это интеграл по траекториям на цилиндре, изображённом на рис. 1. Интеграл берётся по всем вещественнозначным функциям, значения которых на компонентах границы S_1 и S_2 совпадают после поворота S_1 на угол θ. Если мы возьмём наш цилиндр Σ высоты β и склеим две компоненты границы, повернув одну из них на угол θ, то получится риманова поверхность рода 1 без края. Назовём её $\Sigma(\beta,\theta)$. Итак, мы заключаем, что

$$\operatorname{Tr} e^{-\beta H+i\theta P}=\int\limits_{\Omega(\Sigma(\beta,\,\theta))} e^{-I}. \tag{102}$$

Формула (102) — весьма общая формула квантовой теории поля. Она особенно интересна в случае, когда функционал действия I конформно-инвариантен — не меняется при конформных заменах метрики. Такие теории называются конформными теориями поля. Например, конформной теорией поля является теория свободного бозонного поля. Может показаться, что в случае конформной теории поля интеграл, стоящий в правой части (102), определяет конформно-инвариантную величину, зависящую только от конформной структуры на поверхности $\Sigma(\beta,\theta)$. Это не совсем так, поскольку не существует конформно-инвариантного способа определить меру в интеграле по $\Omega(S)$. Тем не менее существуют простые формулы, измеряющие «отклонение» правой части (102) от конформной инвариантности. Для математиков эти формулы связаны с теорией детерминантных пучков, для физиков — с теорией аномалий.

Хорошо известно, что каждая риманова поверхность рода 1 изоморфна поверхности вида $\Sigma(\beta,\theta)$ для некоторых значений β, θ из промежутков

$$0\leqslant\beta<\infty,\quad 0\leqslant\theta<2\pi. \tag{103}$$

Это соответствие не является взаимно-однозначным. Если ввести комплексную переменную $\tau=(\theta+i\beta)/2\pi$, то на значениях τ действует группа $SL(2,\mathbb{Z})$ по формуле

$$\tau\to\frac{a\tau+b}{c\tau+d}, \tag{104}$$

где a, b, c, d — целые и $ab-bc=1$. Римановы поверхности со значениями τ, связанными соотношением (104), изоморфны. В теории модулярных форм обычно вводится переменная $q=e^{2\pi i\tau}$. Если положить

$$H_\pm=(H\pm P)/2, \tag{105}$$

то (102) принимает вид

$$\operatorname{Tr} q^{H_+}\bar{q}^{H_-}=\int\limits_{\Omega(\Sigma(\tau))} e^{-I}. \tag{106}$$

Конформные теории поля, в которых $H_-=0$, называются киральными теориями. Теория свободного бозонного поля не киральна, простейший пример киральной теории — это теория киральных фермионов, но я не буду здесь её объяснять. В киральных теориях правая часть формулы (106) определяет некоторую голоморфную функцию от q или, другими словами, определяет некоторую «модулярную (возможно, лишь слабо модулярную) функцию». В любом случае в конформной теории поля, независимо от того, киральна она или нет, можно получить простые формулы для преобразования выражения (106) при конформных преобразованиях, используя теорию детерминантных пучков или теорию аномалий [32, 9].

Например, в теории аффинных алгебр Ли известно, что характеры интегрируемых модулей старшего веса простым образом преобразуются под действием $SL(2,\mathbb{Z})$ (см. обзор [18]). На первый взгляд это кажется удивительным; группа $SL(2,\mathbb{Z})$ не входит в структуру этих алгебр никаким очевидным образом; она безусловно не является группой их автоморфизмов. Тем не менее известно, что модули старшего веса над аффинными алгебрами Ли допускают естественное описание в терминах квантовой теории поля. Для большинства аффинных алгебр Ли описание модулей уровня 1 можно получить, используя свободные бозоны или свободные фермионы [6, 20, 26, 7]. Вообще любые интегрируемые модули старшего веса над аффинной алгеброй Ли имеют реализации в квантовой теории поля, хотя и более сложные [33]. Соответствующие квантовые теории поля будут конформными, поэтому появление $SL(2,\mathbb{Z})$ в теории аффинных алгебр Ли является частным случаем нашего утверждения о том, что $\mathrm{SL}(2, \mathbb{Z})$ простым образом действует на выражение (106).

Большинство исследований по аффинным алгебрам Ли было проведено в рамках гамильтонова подхода к квантовой теории поля. Как мы только что видели, роль группы $\mathrm{SL}(2, \mathbb{Z})$ становится ясной только при подходе, основанном на интегралах по траекториям. Наше обсуждение прояснило, что перспективы возможного применения гамильтонова подхода связаны с интегралами по траекториям на цилиндре. Перспективы гамильтоновой теории самой по себе выглядят довольно ограниченными. Подход, основанный на интегрировании по траекториям, доставляет очевидное обобщение теории на другие римановы поверхности, которое едва ли пришло бы нам в голову, если бы мы рассуждали только в гамильтоновых терминах. Если в данной теории поля интегралы по траекториям на цилиндре дают нам модуль старшего веса над аффинными алгебрами Ли, то каков математический смысл тех же интегралов по траекториям, скажем на римановой поверхности рода g? Я не могу предложить ответа на этот вопрос. Одна из причин, позволяющих мне надеяться, что этот вопрос привлечёт интерес, состоит в том, что правильный ответ мог бы пролить некоторый свет на теорию струн.

Обсуждавшиеся выше до сих пор интегралы по траекториям никоим образом не являются самыми общими из используемых в квантовой теории поля. На самом деле мы рассмотрели лишь весьма частные случаи употребляющихся конструкций. Обычно вводят так называемые локальные операторы. Рассмотрим, например, теорию свободного бозонного поля, которую мы приводили в качестве примера простейшей квантовой теории поля. Пусть Σ — риманова поверхность, и пусть P — точка на Σ, а x^i — локальные координаты в её окрестности. При изучении теории свободного бозонного поля мы рассматривали интеграл по вещественнозначным функциям

$$\varphi : \Sigma \to \mathbb{R}$$

Под локальным оператором в точке P мы понимаем всякий функционал O от φ, который зависит лишь от $\varphi(P)$ и от значений производных φ в точке P, кроме того, требуется, чтобы зависимость O от производных φ была полиномиальной. Таким образом, примерами локальных операторов будут

$$F(\varphi(P)), \quad F^i(\varphi(P))\frac{\partial\varphi}{\partial x^i}, \quad F^{ij}(\varphi(P))\frac{\partial^2\varphi}{\partial x^i\,\partial x^j}, \tag{107}$$

где F, F^i, F^j — произвольные вещественнозначные функции от $\varphi(P)$. (В теории струн особенно важен случай $F = e^{i\lambda\varphi}$, $\lambda \in \mathbb{R}$.)

Как используются такие локальные операторы? Пусть P_i — некоторый набор точек на Σ, а O_i — локальный оператор в точке P_i. Мы обобщаем интеграл (70), полагая

$$Z_f(O_i;\,\Sigma) = \int\limits_{\Omega_f(S)} e^{-I} \prod_i O_i(P_i). \tag{108}$$

Это выражение можно рассматривать как интеграл по траекториям на поверхности Σ со «вставленными» операторами $O_i(P_i)$. Обычно вводится «корреляционная функция»

$$\langle O_1(P_1)\,O_2(P_2)\,\ldots\,O_n(P_n)\rangle\,\frac{Z_f(O_i;\,\Sigma)}{Z_f(\Sigma)}. \tag{109}$$

Для многих важных операторов O_i (например, для операторов вида (107) с $F = e^{i\lambda\varphi}$) можно вывести для таких корреляционных функций явные формулы, в которые входит лишь функция Грина для оператора Лапласа Δ. Здесь я не буду в это вдаваться.

Остаётся объяснить, почему локальные функционалы $O(P)$ называются локальными операторами. Они в самом деле соответствуют операторам в гамильтоновом подходе к квантовой теории поля. Рассмотрим снова интеграл по траекториям на знакомом нам цилиндре Σ, но теперь с добавлением оператора $O(P)$ в точке P (см. рис. 3). Пусть β_1 и β_2 — расстояния от точки P до S_1 и S_2 соответственно. Обобщая определение (88), мы вводим интеграл

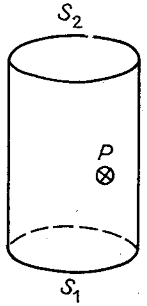


Рис. 3. Интеграл по траекториям на цилиндре со вставленным локальным оператором $O(P)$. Граница Σ снова состоит из двух окружностей S_1 и S_2.

$$R(O(P);\,\psi_2,\,\psi_1) = \int\limits_{\Omega(S_2)} \psi_2^*(f_2) \int\limits_{\Omega(S_1)} \psi_1(f_1)\,Z_{(f_2,\,f_1)}(O;\,\Sigma). \tag{110}$$

Как и в предыдущем случае, линейность по ψ_i обеспечивает нам, что этот интеграл записывается в виде $\langle\psi_2|U|\psi_1\rangle$ для некоторого оператора U в гильбертовом пространстве H_S квантовой теории поля. Определим теперь линейный оператор $\hat{O}(P)$ равенством

$$R(O(P);\,\psi_2,\,\psi_1) = \langle\psi_2|\,e^{-\beta_2 H}\hat{O}(P)\,e^{-\beta_1 H}\,|\,\psi_1\rangle. \tag{111}$$

Это и есть каноническое соответствие между локальными функционалами $O(P)$, которые можно «вставить» в интеграл по траекториям, и локальными операторами $O(P)$ в гильбертовом пространстве (не буду уточнять, в каком именно точном смысле эти операторы являются «локальными»). Локальные операторы $O(P)$ играют решающую роль в физических приложениях квантовой теории поля, подходящие локальные операторы служат естественным средством для описания модулей старшего веса над аффинными алгебрами Ли.

Кроме этого, существует также важное соответствие между операторами и векторами в гильбертовом пространстве. Рассмотрим простейшую риманову поверхность с краем, а именно диск D. Выберем в D внутреннюю точку P. Мы имеем гильбертово пространство H_S, ассоциированное с вещественнозначными функциями на границе S диска D. Мы также имеем локальные операторы в точке P, они образуют некоторое линейное пространство H_P. Сейчас я опишу естественное отображение из H_P в H_S. Пусть $O(P)$ — локальный оператор в P. Мы хотим построить по нему элемент из H_S. Пусть f — вещественнозначная функция на S, положим

$$\psi_O(f) = Z_f(O;\ D). \qquad (112)$$

Соответствие $O \to \psi_O(f)$ и даёт нам требуемое отображение $H_P \to H_S$. В конформной теории поля это соответствие является изоморфизмом между H_P и H_S и имеет особое значение.

Один из особенно важных локальных операторов строится следующим образом. При инфинитезимальной замене метрики g_{ij} на Σ изменение действия I квантовой теории поля задаётся формулой

$$\delta I = \int_{\Sigma} \delta g^{ij} T_{ij}, \qquad (113)$$

где T_{ij} — симметричный тензор, известный как тензор энергии-импульса. Конформные теории поля на римановой поверхности — это в точности те теории, для которых $g^{ij}T_{ij} = 0$. В конформной теории поля T_{ij} имеет только две ненулевые компоненты, преобразующиеся как дифференциалы типов $(2, 0)$ и $(0, 2)$ соответственно. Назовём их T и $\tilde{T}$. Обозначив угловую переменную на нашей канонической окружности S через σ, положим

$$L_n = \int_S e^{in\sigma} T(\sigma), \quad \tilde{L}_n = \int_S e^{in\sigma} T(\sigma). \qquad (114)$$

Можно показать, что эти величины удовлетворяют соотношениям

$$\begin{aligned} &[L_n,\ L_m] = (m - n) L_{n+m} + c\delta_{m+n}(m^3 - m), \\ &[\tilde{L}_n,\ \tilde{L}_m] = (m - n) \tilde{L}_{n+m} + \tilde{c}\delta_{m+n}(m^3 - m), \qquad (115) \\ &[L_n,\ \tilde{L}_m] = 0, \end{aligned}$$

где c и $\tilde{c}$ — константы, а $\delta_n = 1$ при $n = 0$ и 0 при всех остальных n. Если ограничиться рассмотрением L_n, то формулы (115) определяют так называемую алгебру Вирасоро — центральное расширение алгебры Ли группы диффеоморфизмов S. Последняя алгебра порождена векторными полями

$$D_n = -ie^{in\sigma} d/d\sigma, \qquad (116)$$

которые, как легко видеть, подчиняются соотношениям $[D_n, D_m] = (m - n) D_{m+n}$; первая из формул (115) задаёт центральное расширение алгебры (116). Представление алгебры Вирасоро называется представлением старшего веса, если L_0 ограничено сверху; вектор старшего веса — это вектор, аннулируемый L_n с $n > 0$. Конформные теории поля приводят к представлениям старшего веса алгебры Вирасоро, частным случаем этого утверждения является тот факт, что представления старшего веса аффинных алгебр Ли могут быть продолжены на полупрямое произведение с алгеброй Вирасоро.

Данное мной введение в квантовую теорию поля было очень схематичным, чтобы не сказать больше. Квантовая теория поля — богатая и сложная наука, я же объяснил только несколько общих положений. Описание квантовой теории поля как математической теории станет намного более лёгким и естественным, когда начнут проявляться специфически математические приложения. Есть основания полагать, что сейчас — через шестьдесят лет после возникновения квантовой теории поля — это время уже почти подошло.

4. ТЕОРИЯ СТРУН

Наконец, я хотел бы кратко описать, в чём состоит теория струн. Теория струн — ещё более обширная область, чем обсуждавшиеся нами выше, и моё изложение будет ещё более схематичным. Вернёмся к общей теории относительности, описываемой функционалом действия

$$S = \frac{-1}{16\pi G} \int_M R, \qquad (117)$$

где R — скалярная кривизна римановой метрики g на пространственно-временном многообразии M. Построить квантовую теорию, связанную с действием (117), означало бы следующее. Зафиксируем топологию на M и обозначим через Λ пространство метрик на M с точностью до изоморфизма. При квантовании ОТО должны были бы появиться интегралы наподобие

$$Z = \int\limits_{\Lambda} e^{-S}, \tag{118}$$

а также их обобщения: со вставленными операторами, с различными граничными условиями для случая $\partial M \neq 0$ и т. д. Хотя никаких строгих теорем на этот счёт нет, но всё указывает, что никакого удовлетворительного способа придать смысл мере, по которой производится интегрирование в (117), нет. Это одно из проявлений несогласованности ОТО и квантовой механики, характеризованной во введении как центральная проблема теоретической физики.

Прежде чем обсуждать обобщение интеграла (117), предлагаемое в теории струн, обсудим разложение в ряд по возмущениям в ОТО. Пусть η_{ij} — метрический тензор плоского пространства Минковского. Представим метрический тензор g_{ij} в виде

$$g_{ij} = \eta_{ij} + h_{ij}, \tag{119}$$

где h_{ij} — возмущение метрики. Если мы рассматриваем тензор g, близкий к метрике плоского пространства, то можно считать, что h_{ij} принимает значения в некотором линейном пространстве Λ_0. Если выразить функционал действия I через h, то линейный член будет равен нулю, так как плоское пространство Минковского есть решение уравнений Эйнштейна. Общий вид квадратичного члена даётся формулой

$$S_2 = \frac{-1}{16\pi G} \int\limits_{M} h \Delta_L h, \tag{120}$$

где Δ_L — некоторый линейный оператор, известный как лапласиан Лихнеровича. Решения уравнения $\Delta_L h = 0$ — это линеаризованные гравитационные волны, обсуждавшиеся в разд. 1. Кубический член имеет вид

$$S_3 = \frac{-1}{16\pi G} \Phi_3(h), \tag{121}$$

где Φ_3 — довольно сложное кубическое выражение, точный вид которого нас сейчас не интересует. Подставив в интеграл (118) разложение $S = S_2 + S_3 + \ldots$ функционала действия в ряд по степеням h, можно попытаться вычислить влияние возмущения, рассматривая h как малый параметр. Начальное приближение, пренебрегающее всеми членами, кроме S_2, даёт гауссовский интеграл, аналогичный обсуждавшемуся в предыдущем разделе. Но на вычислении поправочных членов мы спотыкаемся: для ОТО получаются не имеющие смысла бесконечности. Это одна из главных причин, заставляющих нас думать, что ОТО не имеет смысла как квантовая теория.

В теориях, имеющих осмысленные квантовые интерпретации, наше понимание основывается в основном на рассмотрении разложения в ряд по возмущениям, подобного описанному выше. Никакой красоты в этом нет. Если в квантовой теории поля и есть красота, то её нужно искать в основных формулах, вроде формулы (117), а не в трюках теории возмущений. Тем не менее в теории струн нам известно именно разложение в ряд по возмущениям! Мы не знаем ни основных формул типа формулы для действия (117), ни того, каковы фундаментальные физические понятия, играющие в теории струн ту роль, которую метрики, связности и кривизны играют в ОТО.

Теперь я расскажу по порядку о том, что́ в теории струн играет роль h, Λ_0, I_2 и I_3. Во-первых, линейное пространство Λ_0 заменяется гильбертовым пространством некоторой квантовой теории поля. Наше обсуждение квантовой теории поля в предыдущем разделе было довольно формальным, и мы не обсуждали физического смысла римановой поверхности Σ или квантового поля φ. В традиционных приложениях квантовой теории поля Σ играет роль пространства-времени, а φ — поле (аналогичное электромагнитному), распространяющееся в пространстве-времени. В теории струн интерпретация обратная: риманова поверхность Σ — это некоторый вспомогательный объект, а φ заменяется отображением Φ: $\Sigma \to M$, где M интерпретируется как пространство-время. Итак, пусть M — плоское многообразие размерности D со стандартными координатами X^i, $i = 1, \ldots, D$. Отображение Φ: $\Sigma \to M$ определяется заданием D вещественнозначных функций X^i на M. Тем самым функционал действия записывается в виде

$$I = \frac{1}{2\pi} \int\limits_{\Sigma} \eta_{ij}\, dX^i \wedge {}^{*} dX^j. \tag{122}$$

Очевидно, что этот функционал инвариантен относительно действия D-мерной группы Пуанкаре на X^i.

Основные составляющие линеаризованной ОТО — это линейное пространство Λ_0 возмущений метрики и квадратичный функционал S_2 действия на этом пространстве. Каковы их аналоги в теории струн? Аналогом Λ_0 служит гильбертово пространство H_S квантовой теории поля, определяемой действием (122). Замена Λ_0 на H_S — действительно принципиальный шаг, поскольку пространство H_S «намного бесконечномернее», чем Λ_0. В самом деле, элемент пространства Λ_0 — это конечный набор функций

$$h_{ij}(x^k), \tag{123}$$

где x^k, $k=1, \ldots, D$ — координаты на пространственно-временном многообразии M. Что же представляет собой элемент из H_S? В предыдущем разделе мы определили пространство Ω_S непрерывных отображений окружности S в M. Обозначив угловой параметр на окружности через σ, такое отображение можно явно задать набором вещественнозначных функций $X^k(\sigma)$. Разложим каждую из них в ряд Фурье:

$$X^k(\sigma) = x^k + \sum_{n \neq 0} e^{in\sigma} x_n^k. \tag{124}$$

Мы выделили здесь нулевой член ряда Фурье, известный как «центр масс струны»; хочется думать, что x^k в разложении (124) соответствует переменной x^k в (123). Конечно, коэффициент x_n^k комплексно-сопряжён с x_{-n}^k. Элемент пространства H_S — это вещественнозначная функция на Ω_S, а конкретнее функция

$$\Phi\left(x^k;\ x_{\pm 1}^k, \ldots\right). \tag{125}$$

Тем самым Φ зависит не только от координат центра масс x^k, фигурирующих уже в (123), но ещё от бесконечного числа переменных x_n^k. Квантовую теорию поля можно рассматривать как приближение к теории струн в том смысле, что если x_n^k в разложении (124) считать малыми, то Φ в первом приближении сводится к функции от x^k — как и гравитационное поле или любое другое поле, обычно рассматриваемое в физике. Точнее говоря, предположим, что x_n^k при $n \neq 0$ достаточно малы, чтобы можно было воспользоваться разложением в ряд Тэйлора вблизи $x_n^k = 0$. Это разложение будет иметь вид[1]

$$\Phi\left(x^k;\ x_n^k\right) = \varphi(x^k) + x_1^j B_j(x^k) + x_1^i x_{-1}^j h'_{ij}(x^k) + \ldots . \tag{126}$$

В этом разложении $\varphi(x^k)$, $B_i(x^k)$, $h'_{ij}(x^k)$ и т. д. — обычные функции от координат центра масс, т. е. обычные функции на M. Отсюда первый вывод: мы можем рассматривать Φ как бесконечный набор функций на пространстве-времени. Общая теория относительности основана на некоторых уравнениях в частных производных для h_{ij}. Аналогично в теории струн мы можем написать нелинейные уравнения в частных производных для Φ (или для некоторой большей системы переменных). Более конкретно нелинейные уравнения для Φ можно представлять себе как систему нелинейных уравнений для бесконечного множества переменных на пространстве-времени. Второй важный вывод из (126) состоит в том, что можно точно определить, в каком именно смысле следует рассматривать Φ как обобщение возмущения метрики (123) в ОТО. В формуле (126) для Φ фигурирует тензорное поле h'_{ij}; его можно трактовать как аналог поля h_{ij}. Нелинейные уравнения, рассматриваемые в теории струн, обладают следующим свойством: при ограничении на h'_{ij} они приближённо сводятся к уравнениям Эйнштейна, которым удовлетворяют h_{ij}, при длинах, больших по сравнению с планковской длиной.

Итак, мы объяснили, каков теоретико-струнный аналог пространства Λ_0. Следующий шаг состоит в том, чтобы объяснить, каков в теории струн аналог квадратичного функционала действия S_2 линеаризованной ОТО. Грубо говоря, аналогом S_2 будет просто

$$S_2(\Phi) = \langle \Phi \,|\, H \,|\, \Phi \rangle, \tag{127}$$

где H — гамильтониан квантовой теории поля. Это не совсем точное определение; необходимо дополнить его некоторым ограничением на векторы старшего веса над алгеброй Вирасоро. Тем не менее сейчас оно нас устроит.

Что будет в теории струн аналогом нелинейных членов $S_3, \ldots$ в ОТО? Правильный подход основан на естественном развитии идей предыдущего раздела. Рассмотрим риманову поверхность Σ, граница которой состоит из трёх окружностей S_1, S_2 и S_3, как на рис. 4. Чуть обобщая конструкцию из предыдущего раздела, можно рассмотреть интегралы по траекториям на Σ с граничными условиями на каждой компоненте границы, определяемыми Φ:

$$S_3(\Phi) = \int\limits_{\Omega(S_1)} \Phi(f_1) \int\limits_{\Omega(S_2)} \Phi(f_2) \int\limits_{\Omega(S_3)} \Phi(f_3)\, Z_{(f_1,\, f_2,\, f_3)}(\Sigma). \tag{128}$$

Это и есть, грубо говоря, аналог кубического члена $S_3(h)$ в эйнштейновском действии. Подход очевидным образом обобщается на случай границы, состоящей из n компонент.

Но имеется и другой язык, на котором можно изложить всё это. В конце предыдущего раздела мы заметили, что в кон-

---

[1] Приводимое ниже разложение — не совсем то, что нужно, но для наших целей оно сгодится.

формной теории поля существует естественный изоморфизм $H_P \approx H_S$ между локальными операторами $O(P)$, которые можно вставить в точке P, и векторами ψ_O гильбертова пространства H_S. Состояние Φ в формуле (128) соответствует при этом изоморфизме некоторому оператору V_Φ (называемому **вертексным оператором**). В дальнейшем мы будем обозначать его просто V. Согласно определению (112), Φ и V связаны между собой следующим образом: если ввести для каждого $i = 1, 2, 3$ диск D_i с границей S_i, то

$$\Phi(f_i) = Z_{t_i}(V;\ D_i). \tag{129}$$

Предположим, что мы компактифицировали риманову поверхность Σ, заклеив граничные окружности S_i дисками D_i, $i=1, 2, 3$. Назовём полученную замкнутую поверхность $\bar{\Sigma}$.

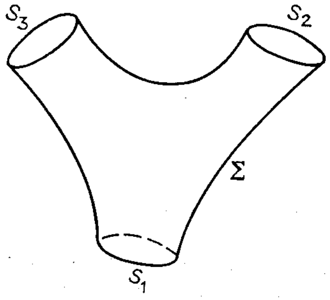


Рис. 4. Риманова поверхность, граница которой состоит из трёх компонент.

Мы хотим записать интеграл (128) как интеграл по $\bar{\Sigma}$. В (128) функция Φ использовалась для определения граничных условий на каждой граничной окружности S_i. Согласно (129), определить граничные условия на S_i с помощью Φ — это то же самое, что заклеить S_i диском D_i со вставленным оператором V в точке $P_i \in D_i$. Тем самым интеграл (128) на самом деле равен некоторому интегралу по траекториям на $\bar{\Sigma}$, со вставлением в трёх точках P_1, P_2 и P_3 вертексного оператора V:

$$S_3(\Phi) = Z(V(P_1),\ V(P_2),\ V(P_3);\ \bar{\Sigma}). \tag{130}$$

Мы не очень озаботились тем, чтобы объяснить, какие именно три точки P_1, P_2 и P_3 выбираются в формуле (130). В определённом смысле это все равно. Риманова поверхность $\bar{\Sigma}$ изоморфна римановой сфере. Хорошо известно, что на множестве троек различных точек на римановой сфере транзитивно действует группа $SL(2, \mathbb{C})$, поэтому с выбором точек P_i не связано никакого конформного инварианта. Но если мы хотим определить не $S_3(\Phi)$, а аналогичную величину $S_n(\Phi)$ для $n>3$, то перед нами уже встаёт вопрос о надлежащем выборе точек P_i. Как показал Поляков [22], правильный рецепт состоит в том, чтобы проинтегрировать по пространству модулей всех расположений n точек на $\bar{\Sigma}$:

$$S_n(\Phi) = \int\limits_{\mathcal{M}_n} Z\left(\prod V(P_i);\ \bar{\Sigma}\right). \tag{131}$$

Здесь $\mathcal{M}_n$ — пространство модулей n-точечных конфигураций на римановой сфере. На рис. 5 показан случай пяти точек.

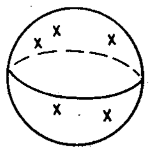


Рис. 5. Интеграл по пространству модулей расположений нескольких точек на римановой сфере.

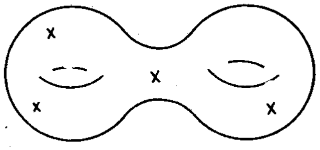


Рис. 6. Интеграл по траекториям из римановой поверхности рода два с несколькими вставленными вертексными операторами.

На самом деле формула (131) — не просто теоретико-струнное обобщение эйнштейновского действия; она содержит дополнительную информацию. Формула (131) является аналогом в теории струн рядов по возмущениям, получаемых из основной формулы (117) при определении «древесного приближения» к ОТО. Физическая интерпретация формулы (131) состоит в том, что она даёт вероятность рассеяния n частиц типа Φ. К сожалению, чтобы объяснить это замечание, потребовалось бы заметно расширить краткое введение в квантовую теорию поля, данное в разд. 3.

Далее, $\bar{\Sigma}$ в формуле (131) — риманова поверхность рода нуль, поскольку именно к роду нуль привели нас попытки найти теоретико-струнный аналог формулы для действия (117). Тем не менее в чисто математическом смысле формула (131) допускает очень естественное обобщение (рис. 6) с заменой $\bar{\Sigma}$ на риманову поверхность рода больше нуля. Это обобщение играет ключевую роль в теории струн. Я уже отмечал, что при попытке интерпретировать (117) как лагранжиан квантовой теории поля мы сталкиваемся с серьёзными трудностями. Попытки вычислить квантовые поправки к классической ОТО приводят к бесконечностям. С другой стороны, в теории струн существует осмысленный и корректно определённый рецепт вычисления квантовых поправок к классическим ответам. Мы просто заменяем $\tilde{\Sigma}$ римановой поверхностью Σ_k рода $k > 0$. Тем самым, если мы хотим вычислить вероятность рассеяния n частиц типа Φ, то в теории струн классический ответ, применимый при $\hbar = 0$, даётся интегралом (131). Если же мы хотим вычислить квантовую поправку к (131) порядка $\hbar^k$, то мы вычисляем не (131), а

$$\int_{\mathscr{M}_{k,n}} Z\left(\prod V(P_i);\ \tilde{\Sigma}\right). \tag{132}$$

Здесь $\tilde{\Sigma}$ — поверхность рода k, а $\mathscr{M}_{k,n}$ — пространство модулей римановых поверхностей рода k с n отмеченными точками. Именно таким образом физики заинтересовались пространствами модулей римановых поверхностей и интегралами по траекториям на этих пространствах. Интегралы вида (132) в действительности обладают замечательно красивыми свойствами; некоторые из них описаны Ю. Маниным в его докладе «Квантовые струны и алгебраические кривые» на этом конгрессе[1]. Но даже если бы интегралы в формуле (132) и не были столь замечательно красивы, уже одного того, что эта формула свободна от ультрафиолетовых расходимостей, отравляющих аналогичные формулы в квантовой ОТО, было бы достаточно, чтобы оправдать значение, придаваемое ей в физике.

В нашем изложении много серьёзных пробелов. Намного лучшее описание получается, если использовать не H_S, а некоторые когомологии алгебры Вирасоро с коэффициентами в H_S. Соответствующая теория когомологий [3, 29, 19, 8] докладывалась на этом конгрессе И. Френкелем[2], и я не буду останавливаться на ней. Я также обошёл вопрос о том, какие именно вертексные операторы V мы используем в интегралах (131) и (132). На самом деле в этих формулах приходится ограничиться вертексными операторами на римановой поверхности, преобразующимися как дифференциальные формы типа (1, 1) (при нашем каноническом соответствии между операторами и векторами из H_S они отвечают векторам старшего веса над алгеброй Вирасоро). Теория когомологий модулей старшего веса, которую я только что упоминал, является подходящей схемой для описания квадратичного действия $S_2(\Phi)$ [27, 1, 28]; она также очень полезна для описания того немногого, что мы знаем про нелинейную теорию с обсуждавшимся нами разложением в ряд по возмущениям [34, 15, 21].

Я попытался убедить вас в том, что, используя интегралы по траекториям на римановых поверхностях, можно сформулировать обобщение ОТО. И самое важное, получающееся обобщение (особенно в своей суперсимметричной форме) свободно от тех недугов, которыми поражена квантовая ОТО. Если логика рассуждений и была временами немного шаткой, то это, по крайней мере частично, объясняется тем, что почти всё, что мы знаем в теории струн, сводится к построенному методом проб и ошибок разложению в ряд по возмущениям. Интегралы (131) и (132) — вероятно, самые красивые формулы, известные в теории струн, тем не менее и они суть просто члены разложения в ряд теории возмущений (по степеням Φ и $\hbar$) некоторой фундаментальной структуры. Если когда и стояла перед наукой жизненно важная проблема, так это открытие этой структуры.

ЛИТЕРАТУРА

1. T. Banks and M. Peskin, Gauge invariance of string fields, Nuclear Phys. B **264** (1986), 513.
2. J.-P. Bismut and D. Freed, Analysis of elliptic families, Preprint, 1985.
3. C. Becchi, A. Rouet, and R. Stora, Ann. Physics **98** (1974), 287.
4. P. Candelas, G. Horowitz, A. Strominger, and E. Witten, Nuclear Phys. B **258** (1985), 46.
5. T. Appelquist, A. Chodos, P. G. O. Freund (editors), Modern Kaluza-Klein theory and its applications, Benjamin-Cummings (to appear).
6. A. Feingold and J. Lepowsky, Adv. in Math. **29** (1978), 271.
7. I. Frenkel and V. Kac, Invent Math. **62** (1980), 23.
8. I. Frenkel, H. Garland, and G. Zuckerman, Preprint, Yale Univ., New Haven, Conn., 1986.
9. D. Freed, MIT Preprint, 1986.
10. J. Glimm and A. Jaffe, Quantum physics: A functional integral point of view, Springer-Verlag, 1981. [Имеется перевод: Глимм Дж., Джаффе А. Математические методы квантовой физики. Подход с использованием функциональных интегралов. — М.: Мир, 1984.]
11. M. B. Green and D. J. Gross (editors), Unified string theories, World Scientific, Singapore, 1986.
12. M. B. Green and J. H. Schwarz, Phys. Lett. **54B** (1985), 502.
13. M. B. Green, J. H. Schwarz, and E. Witten, Superstring theory, Cambridge Univ. Press, New York, 1987.
14. D. J. Gross, J. A. Harvey, E. Martinec, and R. Rohm, Phys. Rev. Lett. **54** (1984), 502.
15. H. Hata, K. Itoh, T. Kugo, H. Kunitomo, and K. Ogawa, Covariant string field theory, Preprint, Kyoto Univ., 1986.
16. C. Itzykson and J.-B. Zuber, Quantum field theory, McGraw-Hill, New York, 1980. [Имеется перевод: Ициксон К., Зюбер Ж.-Б. Квантовая теория поля: В 2-х т. — М.: Мир, 1983.]
17. M. Jacob (editor), Dual models, North-Holland, Amsterdam, 1974.

---

[1] Yu. I. Manin, Quantum strings and algebraic curves, ICM86, 1286—1295 (текст доклада был зачитан И. Зингером на секции «Математическая физика»). — *Прим. изд. ред.*

[2] I. B. Frenkel, Beyond affine Lie algebras, ICM86, pp. 821—839 (сообщение на секции «Группы Ли и теория представлений»). — *Прим. изд. ред.*

18. V. Kac, Infinite dimensional Lie algebras, Birkhäuser, 1983. [В изд-ве «Мир» готовится перевод 3-го (1990 г.) издания этой книги.]

19. M. Kato and K. Ogawa, Nuclear Phys. B **212** (1983), 443.

20. J. Lepowsky and R. L. Wilson, Comm. Math. Phys. **62** (1978), 43.

21. A. Neveu and P. C. West, Phys. Lett. B **168** (1986), 192.

22. A. M. Polyakov, Phys. Lett. B **203** (1981), 207.

23. D. Quillen, Determinants of Cauchy-Riemann operators on a Riemann surface, Preprint, Inst. Hautes Études Sci.

24. J. Scherk and J. H. Schwarz, Phys. Lett. B **52** (1974), 374.

25. J. H. Schwarz (editor), Superstrings, World Scientific, Singapore, 1986.

26. G. Segal, Comm. Math. Phys. **80** (1981), 301.

27. W. Siegel, Phys. Lett. B **151** (1985), 391, 396.

28. W. Siegel and B. Zweibach, Nuclear Phys. B **263** (1986), 105.

29. Тюпин И. В. Препринт ФИАН № 39, 1975.

30. G. Veneziano, Nuovo Cimento **A 57** (1986), 190.

31. E. Witten, Fermion quantum numbers in Kaluza-Klein theory, Shelter Island II: Proceedings of the 1983 Shelter Island Conference on Quantum Field Theory and the Fundamental Problems of Physics, N. Khuri et al., editors, M. I. T. Press, Cambridge, Mass., 1985.

32. —, Global anomalies in string theory, Anomalies, Geometry, and Topology, A. White and W. Bardeen, editors, World Scientific, Singapore, 1985.

33. —, Comm. Math. Phys. **92** (1984), 495.

34. —, Nuclear Phys. B **268** (1986), 253.

35. A. Zee (editor), Unity of forces in the universe, World Scientific, Singapore, 1982.

ПРИНСТОНСКИЙ УНИВЕРСИТЕТ, ПРИНСТОН, НЬЮ ДЖЕРСИ 08544, США

ИМЕННОЙ УКАЗАТЕЛЬ

Авторы, фигурирующие в тексте лишь в латинской транскрипции

СОДЕРЖАНИЕ

Уважаемый читатель!

Ваши замечания о содержании книги, её оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., д. 2, издательство «Мир».