

# МАТЕМАТИЧЕСКОЕ ПРОСВЕЩЕНИЕ

Третья серия

выпуск 5

МЦНМО  
Москва 2001

ББК 22.1  
М34

## Редакционная коллегия

Бугаенко В. О.	Винберг Э. Б.	Вялый М. Н.
Гальперин Г. А.	Глейзер Г. Д.	Гусейн-Заде С. М.
Егоров А. А.	Ильяшенко Ю. С.	Канель-Белов А. Я.
Константинов Н. Н.	Прасолов В. В.	Розов Н. Х.
Соловьев Ю. П.	Сосинский А. Б.	Тихомиров В. М.
Шарыгин И. Ф.	Яценко И. В.	

Главный редактор: В. М. Тихомиров

Отв. секретарь: М. Н. Вялый

Адрес редакции:

121002, Москва, Б. Власьевский пер., д. 11, к. 202

(с пометкой «Математическое просвещение»)

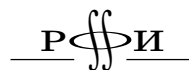
EMAIL: [matpros@mccme.ru](mailto:matpros@mccme.ru) WEB-PAGE: [www.mccme.ru/free-books](http://www.mccme.ru/free-books)

М34 **Математическое просвещение.** Третья серия, вып. 5. — М.: МЦНМО, 2001.  
— 240 с.

Темой очередного номера сборника «Математическое просвещение» является теория билиардных динамических систем. Публикуемые здесь материалы дают представление как о самых разных разделах этой теории. Раздел «Математический мир» содержит очерк, посвященный Л. А. Люстернику, продолжающий материалы предыдущих номеров; статьи о проблемах Гильберта, связанных с обыкновенными дифференциальными уравнениями, о теории хаоса, о квантовых вычислениях. Помимо этого, номер содержит материалы о теореме Понселе, избранные задачи Путнамовских олимпиад, и другие заметки об интересных математических фактах и сюжетах. Публикуются материалы конференции «Математика и общество. Математическое образование на рубеже веков».

ISBN 5-900916-77-4

©МЦНМО, 2001 г.



*Выпуск данного сборника поддержан грантом  
Российского Фонда Фундаментальных Исследований*



МЦНМО выражает благодарность компании Демос за  
предоставление высокоскоростного и качественного  
доступа в Интернет

## СОДЕРЖАНИЕ

### Математический мир

<i>Всероссийская Конференция «Математика и общество.</i> <i>Математическое образование на рубеже веков»</i> . . . . .	5
В. М. Тихомиров <i>Лазарь Аронович Люстерник и теория экстремума</i> . . . . .	12
А. А. Болибрух <i>Обыкновенные дифференциальные уравнения в проблемах Гильберта</i> . . . . .	20
Я. Г. Синай <i>Как математики изучают хаос</i> . . . . .	32
Д. Дойч, А. Эжерт, Р. Лупачини <i>Машины, логика и квантовая физика</i> . . . . .	47

### Тема номера: бильярды

Г. А. Гальперин <i>Бильярды и упругие столкновения частиц и шаров</i> . . . . .	65
Н. И. Чернов <i>Средняя длина пробега в бильярдных системах</i> . . . . .	100
Л. А. Бунимович <i>Рассеяние, дефокусировка и астигматизм</i> . . . . .	106
С. Л. Табачников <i>Внешний бильярд</i> . . . . .	125
Ф. С. Дужин <i>Сколько есть периодических траекторий у трехмерного бильярда?</i> . . . . .	136
Г. А. Гальперин <i>Бильярдная динамическая система для числа <math>\pi</math></i> . . . . .	137

### Две статьи о теореме Понселе

В. В. Прасолов <i>Доказательство теоремы Понселе по Дарбу</i> . . . . .	140
А. А. Панов, Д. А. Панов <i>Гомеодная плотность и теорема Понселе</i> . . . . .	145

### Наш семинар: математические сюжеты

Д. В. Аносов <i>О сумме логарифмически выпуклых функций</i> . . . . .	158
Ф. Бахарев, К. Кохась, Ф. Петров <i>Как складывать треугольники</i> . . . . .	164
Е. Горский, В. Гуровиц, И. Межиров, Е. Осьмова <i>Игра Ричмана</i> . . . . .	192

## **Олимпиады**

Е. М. Бронштейн

*Математические соревнования имени Уильяма Лоуэлла Путнама* . . . . . 192

С. А. Дориченко

*Венгерская задача о квадратах* . . . . . 205

А. Я. Канель

*История одной олимпиадной задачи* . . . . . 207

## **Нам пишут...**

В. В. Войтишек

*Программа курса лекций на двухгодичном потоке ФМШ*

*им. М. А. Лаврентьева, СУНЦ НГУ* . . . . . 209

*Уточнения* . . . . . 214

*Исправления к статье Д. Реповша и А. Скопенкова* . . . . . 215

## **Задачный раздел**

*Условия задач* . . . . . 216

*Решения задач из предыдущих выпусков* . . . . . 218

**Новые издания** . . . . . 230

# Математический мир

---

---

## Всероссийская Конференция «Математика и общество. Математическое образование на рубеже веков»

ОБРАЩЕНИЕ ВСЕРОССИЙСКОЙ КОНФЕРЕНЦИИ «МАТЕМАТИКА И ОБЩЕСТВО.  
МАТЕМАТИЧЕСКОЕ ОБРАЗОВАНИЕ НА РУБЕЖЕ ВЕКОВ»

ДУБНА, «РАТМИНО», 18–22 СЕНТЯБРЯ 2000 ГОДА

1. Мы, участники Всероссийской Конференции по математическому образованию, с удовлетворением отмечаем в качестве одного из важнейших достижений нашей Конференции сам факт ее проведения. Мы считаем, что подобные конференции должны стать регулярными. Тем самым мы не только возрождаем традицию российского математического образования начала XX века, но и содействуем более успешному его развитию в XXI веке. Очень важно, что в нашей Конференции приняли участие работники математического образования всех уровней — от начальной и до высшей школы. Мы все одна семья, у нас общие интересы и общие цели. Мы убеждены, что качество математического образования страны — один из важнейших факторов, определяющих уровень ее экономического и общественно-политического развития. Мы считаем, что математическое образование есть благо, на которое имеет право любой человек, и обязанность общества предоставить каждой личности возможность воспользоваться этим правом.

2. Мы считаем целесообразным создать общественный орган, постоянно действующий между конференциями и избираемый на очередной конференции. Этот орган призван не только исполнять функции Оргкомитета, но и следить за изменениями, происходящими в нашем общем и математическом образовании и помогать распространению лучших достижений. Следует добиваться, чтобы мнение этого органа учитывалось руководителями государства при принятии решений, затрагивающих интересы российского математического образования.

3. Многие недавние решения и проекты руководителей образования вызывают у нас серьезное беспокойство. Прежде всего, это переход на двенадцатилетнее обучение, замена конкурсного экзамена единым тестом. Мы опасаемся, что они не только не будут способствовать развитию образования в России, но, наоборот, приведут к снижению его уровня. Чрезмерное увлечение разного рода непродуманными инновациями, неразумное копирование иностранного опыта, завышенная оценка достижений в области образования западных стран, недооценка собственных, забвение национальных традиций, реформирование ради реформирования — таковы некоторые наблюдаемые сегодня тенденции, создающие внутреннее напряжение в российском математическом (и не только математическом) образовании. Развал сложившейся системы образования таит угрозу для национальной безопасности страны, что может привести к необратимым последствиям для судеб народов России.

4. Мы считаем необходимым, чтобы информация о нашей Конференции, о принятых ею решениях дошла до всех учителей и преподавателей России, до ученых и простых любителей математики и рассчитываем на более широкое и активное участие математической общественности в работе Оргкомитета и будущих конференций.

5. Мы обращаемся ко всем школьникам и студентам России, изучающим математику, независимо от их успехов и отношения к ней. Поверьте нам, мы заботимся о вашем будущем, о вашем интеллектуальном и даже психическом здоровье. Плохое математическое образование, низкая математическая культура в XXI веке могут стать серьезным препятствием не только на пути развития страны, но и в достижении успеха в жизни, значительно ограничить свободу личности. И наоборот, хорошее математическое образование, математическая культура могут защитить вас от многочисленных опасностей, таящихся на пути вашего развития, повысить ваши шансы на самореализацию в выбранной профессии.

6. Многие доклады Конференции были посвящены проблемам перехода от школы к ВУЗу. Мы прекрасно понимаем те социально-демографические проблемы, о которых заботятся руководители образования, знаем из первых рук все недостатки конкурсного экзамена в его традиционной форме, но все же опасаемся, что его замена на единый тест, да и просто на тестирование (при этом обычно имеются в виду самые примитивные формы тестирования) может иметь самые печальные последствия для нашего математического образования в целом, привести к обвальному снижению его уровня и даже к социальному напряжению и увеличению и без того значительного социального расслоения в образовании. Надо искать новые технологии, сочетающие достоинства и традиционного экзамена и тестовых форм оценки качества знания.

7. Мы считаем, что математики-профессионалы должны более активно и регулярно заниматься проблемами математического образования, практическими и научными. Примером в этом им могут служить многие выдающиеся математики России отдаленного и недавнего прошлого. Положительно оценивая деятельность Комиссии по школьному математическому образованию отделения математики РАН, мы все же считаем, что роль РАН в вопросах образования, учитывая интеллектуальные, политические и административные возможности РАН, могла бы быть значительнее. Важно, чтобы в вопросах образования про-

фессиональные ученые и преподаватели установили равноправное партнерство: учиться друг у друга и учить друг друга должны и ученые-математики, и учителя математики. Мы полагаем также, что вклад педагогической науки в школьное образование значительно меньше, чем вклад общества в развитие этой науки. Мы призываем математиков более активно осваивать новые предметные области приложений математики и активно обмениваться знаниями с коллегами. Мы просим руководителей образования всех уровней развивать дифференцированный подход к обучению математическим дисциплинам учащихся и студентов, в особенности будущих экономистов и гуманитариев.

8. Мы благодарим руководство Международной Комиссии по Математическому Образованию (ICMI) за моральную и материальную поддержку нашей конференции и надеемся на будущее более тесное сотрудничество с этой комиссией. В связи с этим считаем полезным создание в нашей стране бюро ICMI. Мы обращаемся к Исполкому ICMI с просьбой активнее привлекать к сотрудничеству ученых и преподавателей из России. Высокая квалификация наших специалистов в области математического образования будет содействовать развитию математической науки и образования в мире.

9. Сегодня в обществе складывается искаженное и даже негативное представление о математике и математическом образовании. К сожалению, математическое сообщество мало занимается общественной пропагандой математического знания, создания, как теперь принято говорить, положительного «имиджа» математики в общественном сознании. Это, безусловно, ошибка, которую надо исправлять. И делать это мы должны сообща, все вместе и каждый в отдельности, в меру своих возможностей. В этой связи мы хотели бы обратиться к российской общественности и представителям СМИ. Пожалуйста, не бойтесь предоставлять математикам, ученым и учителям, эфирное время и газетное пространство. Среди них много интересных людей, которым есть что сказать обществу и которые могут это сделать.

10. Необходимо вернуть традиционный для России высокий социальный статус профессии учителя, преподавателя ВУЗа, ученого. Необходимо стабильное эволюционное развитие системы образования, внимание к ней и поддержка со стороны общества и государственной власти, бизнеса, активная позиция научно-педагогической общественности, солидарность в отстаивании принципиальных позиций.

РЕШЕНИЕ ВСЕРОССИЙСКОЙ КОНФЕРЕНЦИИ «МАТЕМАТИКА И ОБЩЕСТВО.  
МАТЕМАТИЧЕСКОЕ ОБРАЗОВАНИЕ НА РУБЕЖЕ ВЕКОВ»

ДУБНА, 18-22 СЕНТЯБРЯ 2000 ГОДА

Мы, представители педагогической и научной общественности, собрались в год, объявленный ЮНЕСКО годом математики, чтобы обсудить тревожное состояние российского математического образования. Развал системы образования таит угрозу национальной безопасности страны, развитию гражданского общества, модернизации экономики, что может привести к катастрофическим последствиям для народов России.

Глобализация и усложнение экономических и социальных процессов требуют высокого уровня логической и математической культуры общества в целом. В этих условиях политика, направленная на уменьшение роли и веса математики в системе школьного и вузовского образования, представляется разрушительной.

В коренном улучшении нуждается система подготовки учителей. Незамедлительно следует предпринять шаги, способствующие возвращению должного уважения к профессиям учителя школы и преподавателя вуза, повышению их социального статуса, резкому повышению нынешнего уровня зарплаты учителей, преподавателей вузов, всех других работников образования. Необходимо также снизить наполняемость классов, нагрузку на учителя, улучшить снабжение учебно-методической литературой.

Конференция обсудила планируемые радикальные меры, существенно меняющие всю систему образования в стране. Мы высказываем опасение, что введение 12-летнего обучения и всеобщего тестирования как основного способа оценки знаний учащихся необратимо ухудшит уровень математического образования в России.

Предполагаемое введение системы тестирования является неэффективным способом решения проблемы унификации выпускных и вступительных экзаменов. По данным ЮНЕСКО, единственная из развитых стран мира, в максимальной степени применяющая эту систему, — США — находится на одном из последних мест по качеству математического образования. В России нет механизма, который в США и развитых странах компенсирует негативные последствия недостатков системы образования.

Мы убеждены, что сохранение качества образования на должной высоте является необходимым условием эффективного развития страны и обеспечения ее безопасности.

Конференция постановляет:

1. Обратиться с просьбой к Президенту Российской Федерации: рассмотреть вопрос о возможности значительного увеличения заработной платы и пенсии работникам государственной сферы образования.

2. Обратиться к Правительству Российской Федерации с просьбой

а) провести независимую экспертизу с привлечением широкого круга специалистов и общественности состояния образования в средней школе на предмет целесообразности перехода к 12-летней школе;

б) рассмотреть вопрос о недопустимости сокращения числа часов на математику как в школе, так и в вузах;



в) организовать издание массовыми тиражами нескольких вариантов дешевых базовых учебников для средней и высшей школы, поддержать издание методической литературы для учителей и математических журналов;

г) при внесении проектов законодательных и нормативных актов в области образования в обязательном порядке практиковать их обсуждение с привлечением Российской Академии Наук, Российской Академии Образования, научных обществ, школьной и вузовской общественности, средств массовой информации;

д) стабилизировать школьные и вузовские учебные планы и программы по математике на 5 – 10 лет;

е) довести до сведения органов государственной власти Российской Федерации настоящий документ.

3. Обратиться к Государственной Думе Федерального Собрания Российской Федерации с просьбами:

– рассмотреть возможность внесения поправок в федеральные законы «Об образовании», «О пенсионном обеспечении», «О бюджете» с целью существенно повышения качества образования. В частности, предусмотреть увеличение зарплат и пенсий учителей и преподавателей, а также включение в пенсионный стаж времени, посвященного повышению квалификации и уходу за детьми;

– при заслушивании проектов федеральных законов, касающихся проблем образования, практиковать привлечение Российской Академии Наук, Российской Академии Образования, школьной и вузовской общественности.

4. Обратиться к руководителям органов власти субъектов Федерации с просьбой обсудить возможности реализации предложений Конференции.

5. Просить руководство Российской Академии Наук и Российской Академии Образования принимать участие в регулярной экспертной оценке учебных программ, систем проверки качества знаний, реформ образования и т. д.

6. Обратиться ко всем работникам образования Российской Федерации с просьбой обсудить решения Конференции и принять участие в их реализации.

7. Создать на основе Оргкомитета конференции общественную Комиссию по математическому образованию. Поручить Комиссии подготовить предложения о создании Российской Ассоциации математики и математического образования.

8. Секретариату Конференции:

а) подготовить публикацию трудов Конференции,

б) опубликовать список участников Конференции,

в) предоставлять отечественным и иностранным СМИ возможно более полную информацию о работе Конференции.

КРУГЛЫЙ СТОЛ 18 СЕНТЯБРЯ 2000 ГОДА  
«АКТУАЛЬНЫЕ ПРОБЛЕМЫ МАТЕМАТИЧЕСКОГО ОБРАЗОВАНИЯ  
(КОНЦЕПЦИИ, СТАНДАРТЫ, 12-ЛЕТНЕЕ ОБРАЗОВАНИЕ)»

Представление о том, как было организовано обсуждение на этом круглом столе, можно получить из письма В.М. Тихомирова к участникам конференции (примерно так все и было проведено).

Дорогие друзья!

По давнему первоначальному плану мне поручалось проведение этого первого вступительного круглого стола. К сожалению, это оказалось невозможным по состоянию моего здоровья, и мне хочется кратко рассказать, как я планировал проведение этого круглого стола. Быть может, это поможет Вам успешнее его организовать.

Я предполагал выдвинуть для дискуссии два вопроса — очень важных и активно всюду обсуждаемых:

- 1) Разумен ли переход к двенадцатилетней школе, предлагаемый сейчас Министерством образования?
- 2) Разумен ли переход на единый общегосударственный тест (вместо вступительных экзаменов)?

Целью Конференции является, безусловно, обсуждение основных проблем содержания, стандартов образования и т. п. Но мне представляется разумным обсудить эти два вопроса, на которые в принципе можно получить ответ от всех участников Конференции (они — эти ответы — по сути бинарны).

Эти вопросы весьма сложны и не допускают однозначного ответа. Цель круглого стола не в том, чтобы добиться единодушия, но чтобы обозначить позиции участников нашего совещания и выслушать важнейшие аргументы за и против. И потому (несколько раз я проделывал это, и, как считаю, — с успехом) я предложил бы следующее: Пусть на специальном листе в левой его части каждый участник круглого стола до дискуссии выскажет свое мнение в одной из четырех форм: «да», «нет», «микс» («да» с оговорками, при осуществимости определенных условий) или «не знаю».

Анкетирование должно быть безусловно анонимным.

Основным принципом подобных дискуссий должен быть принцип свободы. Пусть каждый, кто хочет, свободно выразит свое мнение, опираясь на свой личный жизненный опыт, но осознавая при этом свою ответственность за будущее нашей профессии и математического образования.

Затем началась бы дискуссия. Я предполагал, что в качестве двух первых выступающих должны быть двое участников, придерживающихся противоположных точек зрения. После вопросов к ним естественно приступить к открытому диспуту.

Здесь важнейшим условием является уважение к человеческому достоинству.

Во всех мероприятиях, посвященных проблемам математического образования, в которых мне доводилось принимать участие, ощущалось чувство единства, но не убеждений и точек зрения (не нужно стремиться к возврату в прошлое, когда все поднимали руки одновременно — истина не решается методами голосования), но чувством общей кровной озабоченности за нашу науку, за наше

будущее, за судьбу наших детей и внуков. И пусть целью выступающих (регламент должен быть весьма жестким) будет не унижение своих оппонентов, не упражнение в сарказме и остроумии, но желание выдвинуть содержательные и глубокие аргументы в пользу или против рассматриваемого вопроса.

Дискуссии должны быть ограничены общим регламентом (скажем, 45 минут на вопрос), чтобы весь круглый стол в своей официальной части длился около двух часов. По завершении дискуссии я бы предложил заново ответить в правой части анкеты на те же самые вопросы и потом сдать анкеты для обработки. Эти данные, как мне представляется, окажутся адекватным выражением того, что и следует называть общественным мнением, с которым необходимо считаться людям и организациям, планирующим реформирование системы образования.

Разумеется, это только мое предложение, которым вы можете воспользоваться или пойти каким-то другим путем. В любом случае желаю вам успеха.

*В. М. Тихомиров*

#### ИТОГИ КРУГЛОГО СТОЛА

Итоги подсчета анкет, сданных участниками круглого стола, таковы:

	12-ЛЕТКА				Всего
	да	нет	mix	не знаю	
перед обсуждением	11	96	14	7	128
после обсуждения	9	107	8	4	128

	ЕДИНЫЙ ЭКЗАМЕН				Всего
	да	нет	mix	не знаю	
перед обсуждением	24	81	16	9	130
после обсуждения	22	88	14	7	131

## Лазарь Аронович Люстерник и теория экстремума

В. М. Тихомиров

В последний день позапрошлого года исполнилось сто лет со дня рождения выдающегося математика и замечательного человека — Лазаря Ароновича Люстерника.

Л. А. Люстерник родился в небольшом польском городке Здунска Воля, близ Лодзи 31 декабря 1899 года. В начале Первой мировой войны, желая избежать немецкой оккупации, он был вынужден перебраться в Смоленск. (Сёстры его остались в Польше и погибли во Вторую мировую войну.)

Люстерник поступил в Смоленскую гимназию (в которой ранее учился и которую в 1913 году закончил Павел Сергеевич Александров). В 1918 году, по окончании Смоленской средней школы (так стала именоваться гимназия), он поступает в Московский университет на физико-математический факультет.

Лазарь Аронович Люстерник принадлежал к блестящей плеяде учеников Николая Николаевича Лузина.

Его студенческие годы совпали с той порой нашей страны, о которой П. С. Александров писал так: «То были годы необычайного подъёма и увлечения внезапно открывшимися новыми творческими возможностями, годы подлинного цветения для многих молодых людей, впервые вкусивших радость творческого соприкосновения с наукой. Мало найдётся в истории математики периодов столь горячего энтузиазма, как начало двадцатых годов в Московском университете, когда в столь короткий срок, буквально в несколько лет, возникла большая научная школа, в значительной степени определившая развитие математики в нашей стране и сразу выдвинувшая целый ряд выдающихся ученых». Одним из наиболее ярких представителей в этом ряду был Лазарь Аронович Люстерник.

По окончании Университета (1921 г.) Лазарь Аронович был оставлен при нем в качестве «научного сотрудника 2-го разряда». В 1922 г. он становится аспирантом Института математики и механики МГУ.

В 1924 г., будучи аспирантом, Лазарь Аронович впервые выступает с докладом на Московском математическом обществе и сдает в печать первую научную работу. Лазарь Аронович обладал замечательным чувством юмора. После своего первого доклада на Московском математическом обществе, он был принят в члены этого общества, и тогда же ему был вручен ключ от специального туалета, куда допускались лишь члены Общества. «Это был первый закрытый распределитель в моей жизни», — так заключает Лазарь Аронович этот эпизод своей жизни.

Его «выпускная аспирантская работа» «Прямые методы вариационного исчисления» (1926 г.) по представлению Института математики и механики МГУ награждается премией Наркомпроса. В 1927 году Лазарь Аронович становится

приват-доцентом МГУ и начинает читать там свои первые специальные курсы. В 1928 году он участвует в работе Международного математического конгресса в Болонье, где выступает с докладом «Топологические методы в дифференциальной геометрии». В том же 1928 году он избирается на должность профессора Нижегородского Госуниверситета. Но там он проработал недолго.

В 1930 году Люстерника избирают на должность профессора МГУ (в звании профессора, а одновременно и в степени доктора физико-математических наук, он утверждается в 1935 году с образованием в нашей стране ВАКа).

Широта научных интересов Лазаря Ароновича была необыкновенной: дифференциальные уравнения, топология, вариационное исчисление, функциональный анализ, геометрия, вычислительная математика, специальные функции и многое другое. Люстерник был и геометром, и аналитиком. Для его математического стиля характерно движение от простого к сложному, в основе далёких обобщений у него всегда лежала простая модель.

О некоторых работах по теории экстремума, в частности, о его исследованиях в вариационном исчислении и по методу сеток мы далее поговорим более подробно.

Обо всём этом он впоследствии сказал в стихах:

«Я стал работать в направленьях  
Тогда в Москве совсем не модных —  
Вариационном исчисленьи,  
Задачах в частных производных...  
Я метод сеток развивал.»

В 20-е – 30-е годы Лазарь Аронович (совместно со Львом Генриховичем Шнирельманом) создаёт совершенно новую математическую область — топологические методы нелинейного анализа. Лазарь Аронович вводит новый гомотопический инвариант — категорию, и в работе, совместной с Л. Г. Шнирельманом, успешно применяет это понятие к оценке числа критических точек гладкой функции на гладком многообразии. Итогом их исследований явилось решение классической проблемы Пуанкаре о трех геодезических, и этот результат вошел в число высших мировых достижений нашего века в математике. (Об этом направлении рассказывалось в нашей совместной с В. В. Успенским статье о Л. Г. Шнирельмане, опубликованной в предыдущем номере «Математического просвещения».)

Предвоенные исследования Лазаря Ароновича были посвящены теории обыкновенных дифференциальных уравнений (им были получены замечательные результаты по качественному поведению собственных функций нелинейных задач типа Штурма-Лиувилля), функциональному анализу (доказанная им теорема о касательном пространстве лежит в самой основе современной теории экстремальных задач, об этом также пойдёт речь далее), геометрии (общеизвестным стало его обобщение классического неравенства Брунна – Минковского об объёме векторной суммы выпуклых тел на случай произвольных множеств).

С 1934 года Лазарь Аронович, не прекращая связи с мехматом, стал работать в Математическом Институте им. В. А. Стеклова АН СССР. В годы войны под его руководством там выполнялись спец-исследования оборонного значения. В частности, им были разработаны и внедрены таблицы, позволяющие

штурманам самолётов быстро определять по данным прибора положение самолёта. Для реализации этой работы требовались вычисления большого объёма. Это привело к необходимости создания и совершенствования вычислительных средств. Л. А. Люстерник проявил глубокое понимание перспектив развития прикладной математики. Некоторые его ученики и сотрудники уверяли меня, что сами термины «вычислительная математика» и «вычислительная техника» были впервые употреблены Лазарем Ароновичем!

В 1942 году Лазарем Ароновичем была в кратчайший срок (что было обусловлено органами НКВД) решена задача о составлении таблицы для определения курсового угла и расстояния — необходимые для пилота сведения. Они до войны вычислялись достаточно сложно. В разработанной таблице по координатам начального и конечного пунктов сразу же определялись путевые углы и длина при следовании по геодезической. Для этого две функции трёх переменных были представлены, как суперпозиции функций двух переменных — одна точно, а другая с достаточной степенью приближения.

Лазарь Аронович был избран член-корреспондентом АН СССР в 1946 г., в том же году он был удостоен звания лауреата Сталинской премии — высшего премиального отличия в те времена.

В послевоенные 40-е годы, когда развернулась работа по созданию, как тогда называли, «отечественных АЦМ» (автоматических цифровых машин), Л. А. Люстерник стал заведующим отделом Института точной механики и вычислительной техники АН СССР. Он был одним из инициаторов открытия в СССР такого института, также как и создания в МГУ и в некоторых других вузах нашей страны кафедр вычислительной математики. Лазарь Аронович активно занялся совершенно новым для того времени кругом проблем, связанных с программированием. Благодаря ему в СССР появилась первая научная группа по работе на вычислительных машинах, а затем и первая советская книга по программированию (Л. А. Люстерник, А. А. Абрамов, В. И. Шестаков, М. Р. Шура-Бура «Решение математических задач на автоматических цифровых машинах», М.: Изд-во АН СССР, 1952). Лазарь Аронович явился и первым в СССР лектором по курсу программирования.

В 50-е – 60-е годы у Лазаря Ароновича происходит новый творческий взлёт — и появляется блестящий цикл его работ (совместных с Марком Иосифовичем Вишиком) по асимптотическим разложениям решений уравнений с малым параметром, задачам с барьерами и с быстро меняющимися граничными функциями, по возмущению несимметрических матриц и операторов.

Огромную роль сыграл Л. А. Люстерник в самой истории механико-математического факультета МГУ. Он был первым, кто прочел на факультете курс лекций по функциональному анализу и открыл по нему (совместно с Абрамом Иезекииловичем Плеснером) научно-исследовательский семинар. Вместе с Михаилом Александровичем Лаврентьевым он коренным образом модернизировал университетский курс вариационного исчисления. Одним из первых он объявил факультетский семинар по вычислительной математике — еще до организации соответствующей кафедры, профессором которой он впоследствии работал (до перехода на кафедру ОПУ — общих проблем управления). Он же организовал затем семинар по математическим вопросам управления производством. Добавим, что он был первым заведующим кафедрой функционального анализа мех-

мата МГУ, был среди организаторов первых Московских школьных математических олимпиад, первым ответственным редактором журнала «Успехи математических наук». Среди учеников Лазаря Ароновича 20 докторов наук и около 50 кандидатов наук.

С приходом на кафедру ОПУ Лазарь Аронович сразу же стал читать на ней специальные курсы и вести научно-исследовательские семинары. Именно на этой кафедре он занялся разработкой нового научного направления — вероятностных методов в теории специальных функций. Незадолго до своей кончины им была опубликована основополагающая статья и в этой области.

С 1969 года под редакцией Л. А. Люстерника на мехмате МГУ стали выходить сборники трудов из серии «Математические вопросы управления производством». Серия возникла в связи с работой одноименного семинара под его руководством, но в дальнейшем включила в свой круг интересов и другие исследования, проводившиеся на кафедре ОПУ. Всего было издано девять сборников этой серии.

Последние годы своей жизни Лазарь Аронович провёл на кафедре общих проблем управления. До этого он работал на кафедре вычислительной математики, которая после организации нового университетского факультета — Вычислительной математики и кибернетики, перешла туда. Но Лазарь Аронович не желал расставаться с мехматом и из всех кафедр мехмата выбрал кафедру ОПУ. Я всегда воспринимал это, как подарок судьбы.

Лазарь Аронович, как и многие его великие сверстники, был легендарной фигурой. Про него можно было услышать много смешных рассказов. Часто люди боятся показаться смешными, и, как правило, это свидетельствует либо об ограниченности их личности, либо о том, что они преисполнены недоброжелательства. Светлая личность высвечивается лучами юмора. Вот два примера из жизни Лазаря Ароновича.

Однажды Евгений Михайлович Ландис, будучи студентом, должен был сдать спецкурс Люстернику. Они встретились в назначенное время, Люстерник задал вопрос и куда-то вышел. Вскоре он вернулся, но не начал экзамена, а стал разговаривать с Ландисом о математике (Ландис был активным участником многих семинаров). Это длилось около часа. Наконец, Люстерник нетерпеливо посмотрел на часы и сказал: «Ну куда же он пропал?». «Кто-то должен подойти?» — спросил Ландис. «Ну конечно, где же тот студент, которого мы с Вами должны были экзаменовать?».

В 1975 году Лазарь Аронович вышел на пенсию. Как-то незадолго до этого он определил тот момент, когда профессор МГУ должен выходить на пенсию. Многие знают, что для того, чтобы открыть парадную дверь главного входа МГУ требуются немалые усилия. Лазарь Аронович пошутил как-то, что профессор МГУ может работать лишь до того момента, когда он сам без посторонней помощи может открывать эту дверь.

Лазарь Аронович Люстерник скончался 23.07.1981.

В память о нем в сентябре 1999 года кафедра ОПУ, совместно с МИ РАН и Международным Банаховым центром, провели в Варшаве и Бендлево (близ Познани) мини-симпозиум, посвященный 100-летию со дня рождения Лазаря Ароновича Люстерника.

\* \* \* \* \*

А теперь коснёмся некоторых собственно математических тем в творчестве Лазаря Ароновича Люстерника.

Л. А. Люстерник стоял у истоков прямых методов и методов сеток в численном анализе. Еще в 1924 году он первым применил метод конечных разностей для доказательства существования решений уравнений с частными производными (уравнений Лапласа) в криволинейных областях. «Мне показалось естественным, — пишет он по этому поводу в своих воспоминаниях, — рассматривать вариационные задачи как пределы соответствующих задач для функций, заданных в конечном числе точек. Тогда, например, условие Якоби положительности второй вариации возникает, как предельное для условия Сильвестра. Поскольку в аспирантскую программу входило доказательство существования решения задачи Дирихле для плоского уравнения Лапласа, я попробовал доказать это предельным переходом от соответственной «сеточной задачи» и был даже удивлен, что такой «грубый» метод доказывал теорему в «тонких» случаях — даже когда граница не содержит континуальных компонент». От этой работы потянулся след, ощутимый и в наши дни. Тогда же состоялись плодотворные научные контакты Люстерника с Петровским, приведшие к первой публикации Ивана Георгиевича.

Заклучим наш рассказ сюжетом:

#### МЕТОД НЬЮТОНА, ТЕОРЕМА ОБ ОБРАТНОЙ ФУНКЦИИ И ТЕОРЕМА ЛЮСТЕРНИКА

Свои основные открытия в области математического анализа И. Ньютон совершил в начале шестидесятых годов семнадцатого столетия, когда ему было около двадцати лет. В частности, он научился решать уравнения  $f(x) = y$ , где  $f$  — функция одного переменного. Свой метод он проиллюстрировал на примере решения уравнения  $x^3 - 2x = 5$  (см. в книге И. Ньютон. Математические работы. М-Л, Изд-во технико-теоретической литературы, 1937, с. 9). В качестве начального приближения Ньютон выбирает число 2, далее полагает  $x = 2 + p$  и, подставляя в исходное уравнение, приходит к новому:  $p^3 + 6p^2 + 10p - 1 = 0$ , «у которого, — пишет Ньютон, — следует определить корень  $p$ , чтобы прибавить его к первому результату. Отсюда (пренебрегая  $p^3 + 6p$  по малости) имеем приблизительно  $10p - 1 = 0$  или  $p = 0.1$ . Поэтому я пишу в результате 0.1 и полагаю  $0.1 + q = p$ ; это выражение я подставляю, как и раньше, и при этом получается  $q^3 + 6.3q^2 + 11.23q + 0.061 = 0$ ». Совершив ещё одну итерацию, Ньютон получает такое приближение:  $x \approx 2.09455147$ . (На самом деле ответ такой: 2.094551481..., т. е. Ньютон вычислил корень с точностью до восьмого знака после запятой!)

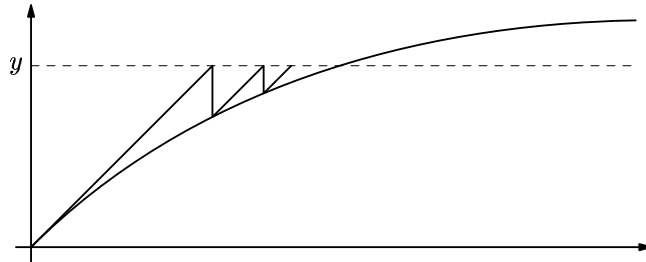
В нашем фрагменте по сути дела изложен метод Ньютона решения уравнения  $f(x) = y$ , состоящий в том, что после выбора начального приближения  $x_0$  далее применяется итеративная процедура:  $x_{k+1} = x_k + (f'(x_k))^{-1}(y - f(x_k))$ ,  $k = 0, 1, 2, \dots$

Роль, которую суждено было сыграть методу Ньютона в истории математики, совершенно исключительная. Одно из важнейших приложений его — доказательство теоремы об обратной (и неявной) функции.



Сформулируем теорему об обратной функции в самом простейшем (опять-таки — одномерном) случае.

**ТЕОРЕМА 1 (ОБ ОБРАТНОЙ ФУНКЦИИ В ОДНОМЕРНОМ СЛУЧАЕ).** Пусть  $f$  — функция одного переменного, определённая в окрестности нуля, равная нулю в нуле, непрерывно дифференцируемая в окрестности нуля причём  $f'(0) \neq 0$ . Тогда найдётся такое число  $\delta > 0$ , что для любого числа  $y$  такого, что  $|y| < \delta$ , существует единственное решение  $x(y)$  уравнения  $f(x) = y$ .



Геометрическая суть доказательства этой теоремы изображена нами на рисунке, а аналитическое содержание изображённого состоит в применении *модифицированного* метода Ньютона:

$$x_0 = 0, \quad x_{k+1} = x_k + (f'(0))^{-1}(y - f(x_k)), \quad k \geq 0.$$

Этот результат естественно приписать самому Ньютону. Двумерные обобщения этой теоремы появились лишь в XIX веке, в конце XIX века и в первом десятилетии XX столетия эта теорема получила многомерное развитие, а в 1934 году Лазарь Аронович Люстерник дал бесконечномерное обобщение этой теоремы. Оно было опубликовано в журнале «Математический сборник» (Л. А. Люстерник. Об условных экстремумах функционалов. Матем. сб. т. 41, №3, 1934, с. 390 – 401.) Сначала мы сформулируем прямое обобщение теоремы 1, а затем, в качестве простого следствия из него, выведем сам результат Люстерника.

**ТЕОРЕМА 2 (ОБ ОБРАТНОЙ ФУНКЦИИ В БЕСКОНЕЧНОМЕРНОМ СЛУЧАЕ).** Пусть  $X$  и  $Y$  — банаховы пространства,  $U$  — окрестность нуля в  $X$ ,  $F: U \rightarrow Y$  — отображение из  $X$  в  $Y$ ,  $F(0) = 0$ , непрерывно дифференцируемое в окрестности нуля, причём  $F'(0)$  отображает  $X$  на всё  $Y$ . Тогда найдётся такое  $\delta > 0$ , что для любого  $y \in Y$  такого, что  $\|y\|_Y < \delta$ , существует решение  $x(y)$  уравнения  $F(x) = y$  такое, что  $\|x(y)\|_X \leq K\|y\|_Y$  (для некоторой константы  $K > 0$ ).

Понятие банахова пространства обобщает понятие конечномерного евклидова пространства. Ныне студенты университета знакомятся с этим понятием на первых курсах, оно (и начала теории банаховых пространств) стали теперь абсолютно общепринятыми. А в том далёком 1934 году теории банаховых пространств исполнилось всего лишь два года. Автором теории был замечательный польский математик Стефан Банах, который в 1932 году опубликовал свой знаменитый мемуар «Théorie des opérations linéaires» (Теория линейных операций), в котором заложил основы теории, ставшей одной из существенных составных

частей функционального анализа. Знания по теории банаховых пространств в те годы в Москве черпались из двух экземпляров этой книги. Один принадлежал Плеснеру (и потому, вероятно, был доступен Люстернику), другой — Колмогорову. Поразительно, что Лазарь Аронович так быстро сумел извлечь из новой теории столь фундаментальный результат.

Школьник, не владеющий понятием банахова пространства, может считать, что  $X$  — это, например, трёхмерное, а  $Y$  — двумерное пространство с евклидовой нормой. При доказательстве используются только самые обычные свойства нормы (в основном, неравенство треугольника) и один важнейший принцип линейного анализа (т.е. теории банаховых пространств), а именно, теорема Банаха об обратном операторе, которую мы сформулируем в виде теоремы *о правом обратном*: пусть  $X$  и  $Y$  — банаховы пространства и  $\Lambda: X \rightarrow Y$  — линейный, непрерывный оператор, отображающий  $X$  на всё  $Y$ , тогда существуют оператор  $R: Y \rightarrow X$  и константа  $C > 0$  такие, что  $\Lambda Ry = y \forall y \in Y$  и  $\|R(y)\|_X \leq C\|y\|_Y$ . (Люстерник, естественно, использует этот факт, но ссылается почему-то не на Банаха, а на Хаусдорфа, не указывая даты его работы; историкам математики разумно разобраться, что это за работа и как она связана с теоремой Банаха.) А теперь приведём доказательство теоремы 2.

**ДОКАЗАТЕЛЬСТВО.** Доказательство теоремы основывается на том же модифицированном методе Ньютона:

$$x_0 = 0, \quad x_{k+1} = x_k + R(y - F(x_k)), \quad k \geq 0,$$

где  $R$  — оператор, правый обратный к  $F'(0)$ . Остаётся доказать, что для малых  $y$  метод итераций сходится к решению  $x(y)$  уравнения  $F(x) = y$ . При этом бесконечномерный случай несколько не сложнее одномерного.

Обозначим  $F'(0) = \Lambda$ , тогда  $\Lambda Ry = y$  и  $\|Ry\|_X \leq K\|y\|_Y \forall y \in Y$ . Выберем число  $\delta > 0$  столь малым, чтобы отображение  $F$  было непрерывно-дифференцируемым в  $\delta$ -окрестности нуля и из  $\|x'\| < \delta$  и  $\|x\| < \delta$  вытекало бы неравенство

$$\|F(x') - F(x) - \Lambda(x' - x)\| \leq \frac{1}{2K}\|x' - x\|.$$

Пусть  $\|y\|_Y < \frac{\delta}{2K}$ . Тогда  $\|x_1\|_X = \|Ry\|_Y \leq K\|y\|_Y < \delta/2 < \delta$ . По индукции покажем, что  $\|x_n\|_X < \delta$  для всех  $n \in \mathbb{N}$ . Пусть элементы  $\{x_k\}_{k=1}^n$  обладают этим свойством и  $1 \leq k \leq n$ . Тогда в силу того, что  $y - F(x_k) - \Lambda(x_{k+1} - x_k) = 0$ , получаем:

$$\begin{aligned} \|x_{k+1} - x_k\|_X &\leq \\ &\leq K\|y - F(x_k)\|_Y = K\|y - F(x_k) - y + F(x_{k-1}) + \Lambda(x_k - x_{k-1})\|_Y \leq \\ &\leq \frac{K}{2K}\|x_k - x_{k-1}\|_X = \frac{1}{2}\|x_k - x_{k-1}\|_X \leq \frac{1}{4}\|x_{k-1} - x_{k-2}\|_X \leq \dots \leq \frac{1}{2^k}\|x_1\|_X. \end{aligned}$$

Отсюда в силу неравенства треугольника будем иметь

$$\|x_n\|_X = \|x_n - x_{n-1} + x_{n-1} - \dots + x_2 - x_1 + x_1\|_X \leq \left(\frac{1}{2^{n-1}} + \dots + 1\right)\|x_1\|_X < \delta.$$

Итак,  $x_n$  определены для всех  $n$  и, как легко понять, эта последовательность является последовательностью Коши. Значит,  $x_n \rightarrow x(y)$  при  $n \rightarrow \infty$ . Ввиду

того, что  $\|y - F(x_n)\|_Y \leq \frac{\delta}{K2^n}$  получаем, что  $F(x(y)) = y$  и  $\|x(y)\|_X \leq 2\|x_1\|_X \leq 2K\|y\|_Y$ .  $\square$

**ЗАМЕЧАНИЕ.** Если бы мы начали свой итеративный процесс не с нуля, а с достаточно малого  $x$ , мы пришли бы к элементу  $\varphi(x, y)$  такому, что  $F(x + \varphi(x, y)) = y$  и при этом  $\|\varphi(x, y)\|_X \leq C\|y - F(x)\|_Y$ .

Теперь сформулируем и докажем теорему Люстерника о касательном пространстве. Пусть  $X$  — нормированное пространство и  $M$  — некоторое его подмножество. Элемент  $x$  называется *касательным* к  $M$  в точке  $\hat{x} \in M$ , если существует отображение  $r: [-1, 1] \rightarrow X$  такое, что  $\hat{x} + tx + r(t) \in M \forall t \in [-1, 1]$  и  $\|r(t)\| = o(t)$  при  $t \rightarrow 0$ . Множество касательных векторов к  $M$  в точке  $\hat{x}$  обозначается  $T_{\hat{x}}M$ . Если это множество является подпространством  $X$ , то оно называется *касательным пространством* ко множеству  $M$  в точке  $\hat{x}$ .

**ТЕОРЕМА 3 (ЛЮСТЕРНИКА О КАСАТЕЛЬНОМ ПРОСТРАНСТВЕ).** Пусть  $X$  и  $Y$  — банаховы пространства,  $U$  — окрестность точки  $\hat{x}$  и  $F: U \rightarrow Y$  — отображение, непрерывно-дифференцируемое в  $U$ , причём  $F'(\hat{x})$  отображает  $X$  на всё  $Y$ . Тогда если  $M = \{x \in X \mid F(x) = F(\hat{x})\}$ , то  $T_{\hat{x}}M = \text{Ker } F'(\hat{x})$ .

**ДОКАЗАТЕЛЬСТВО.** Не ограничив себя в общности, считаем, что  $\hat{x} = 0$  и  $F(0) = 0$ . Пусть  $x \in T_0M$ . Тогда  $0 = F(0) = F(tx + r(t)) = tF'(\hat{x})x + o(t)$ , откуда сразу следует, что  $F'(0)x = 0$ , т.е.  $T_0M \subset \text{Ker } F'(0)$ .

С другой стороны, если  $x \in T_0M$ , то, положив  $r(t) = \varphi(tx, 0)$ , получаем  $F(tx + r(t)) = F(tx + \varphi(tx, 0)) = 0$  (а значит,  $tx + r(t) \in M$ ), и при этом

$$\|r(t)\|_X = \|\varphi(tx, 0)\|_X \leq K\|F(tx)\|_Y = K\|tF'(0)x + o(t)\|_Y = o(t),$$

т.е.  $\text{Ker } F'(0) \subset T_0M$ . Теорема доказана.

Происхождение этой теоремы, по-видимому, связано с работой Л. А. Люстерника совместно с М. А. Лаврентьевым над учебником по вариационному исчислению. Он был издан в следующем году (М. А. Лаврентьев, Л. А. Люстерник. Основы вариационного исчисления. М.-Л.: ОНТИ, 1935, т. 1, 2.) Лазарь Аронович явно указывает в статье, что одной из целей статьи являлось включение вариационного исчисления в общую схему теории экстремума в функциональном анализе. Но почему-то в самом учебнике теория условий экстремума была построена традиционным путём, без применения методов функционального анализа.

Когда спустя сорок пять лет мы с Владимиром Михайловичем Алексеевым написали наш учебник по оптимальному управлению (В. М. Алексеев, В. М. Тихомиров. Оптимальное управление. М.: Наука, 1979), мы подарили его Лазарю Ароновичу. Спустя некоторое время (когда В. М. Алексеева уже не было в живых) Лазарь Аронович вдруг позвонил мне. Он поблагодарил за подаренную книгу и сказал с некоторым смущением, что не подозревал о том, что его теорема может лечь в основание общей теории экстремума.

Скорее всего, Лазарь Аронович лукавил немного, конечно, он нечто подобное «подозревал», но что-то, наверное, отвлекло его тогда от осуществления широкой программы модернизации теории экстремальных задач, а потом он позабыл о своих замыслах.

А теорема Люстерника ныне — одна из самых цитируемых его теорем.

## Обыкновенные дифференциальные уравнения в проблемах Гильберта

Академик РАН А. А. Болибрух

Среди знаменитых проблем Гильберта, сформулированных им на международном математическом конгрессе в Париже в 1900 году, две проблемы, 16-я (точнее, вторая её часть) и 21-я, относятся к аналитической теории обыкновенных дифференциальных уравнений. История их исследования стала одной из интересных страниц в развитии математики XX столетия. Обе некоторое время считались решенными, и в решении обеих были найдены ошибки. Лишь в 80–90-х годах было получено решение 21-й проблемы и частичное продвижение в 16-й, которая до сих пор всё ещё далека от своего окончательного решения. В этой заметке рассказывается об истории исследования указанных проблем и об их связи с некоторыми другими интересными задачами современной математики.

Вторая часть 16-й проблемы Гильберта формулируется следующим образом (см. [Hi]):

*Исследовать вопрос о максимальном числе и о расположении предельных циклов Пуанкаре для дифференциального уравнения первого порядка и первой степени вида*

$$\frac{dy}{dx} = \frac{Y}{X}, \quad (1)$$

где  $X, Y$  — целые рациональные функции  $n$ -й степени относительно  $x, y$ .

Что такое предельный цикл? Почему эта задача была включена Гильбертом в число его 23-х проблем? В чем её значение и важность? Ниже мы постараемся ответить на эти вопросы.

Целая рациональная функция  $n$ -й степени относительно  $x, y$  — это многочлен

$$P(x, y) = \sum_{i \geq 0, j \geq 0, i+j \leq n} a_{ij} x^i y^j$$

от двух переменных, у которого хотя бы один из коэффициентов  $a_{ij}$  при  $i + j = n$  отличен от нуля. При  $n$ , равном нулю, такая функция является постоянной, поэтому уравнение (1) в этом случае принимает вид

$$\frac{dy}{dx} = k, \quad k \in \mathbb{R}.$$

Общее решение последнего уравнения легко находится, это  $y = kx + c$ , где  $c$  — произвольная постоянная. Напомним, что решения дифференциального уравнения, изображенные на плоскости  $(x, y)$ , называются интегральными кривыми. В рассматриваемом случае интегральные кривые являются прямыми, которые заполняют всю плоскость и нигде не пересекаются (см. рис. 1).

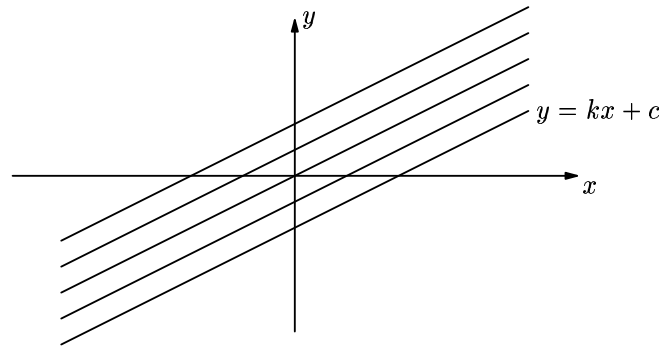


Рис. 1.

В таких случаях говорят, что на плоскости задано *слоение*, плоскость *расслоена* на эти прямые. Видно, что каждая интегральная кривая бесконечна в обе стороны. Никаких предельных циклов здесь нет, и при  $n = 0$  задачи, как таковой, просто не существует.

В случае  $n = 1$  рассмотрим простейший пример, когда уравнение имеет вид

$$\frac{dy}{dx} = -\frac{x}{y}.$$

Такое уравнение нетрудно проинтегрировать, домножив обе его части на  $ydx$  и выделив полный дифференциал:

$$ydy = -x dx, \quad d(y^2 + x^2) = 0.$$

Общее решение имеет вид  $y^2 + x^2 = c$ , и соответствующие интегральные кривые являются концентрическими окружностями с центром в точке 0 (см. рис. 2).

Каждая такая замкнутая интегральная кривая называется *циклом*. В этом случае вся плоскость расслоена на интегральные кривые, которые являются циклами. Но предельных циклов здесь еще нет.

Пропустим случай  $n = 2$  и перейдем сразу к случаю  $n = 3$ .

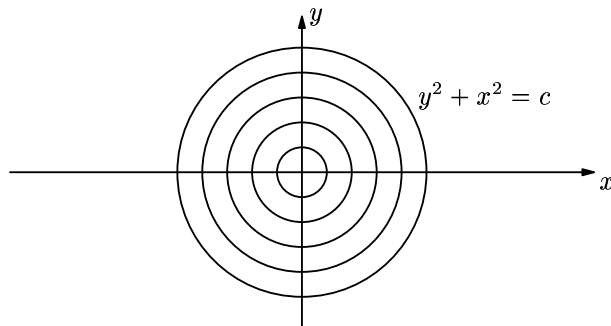


Рис. 2.

Следующее уравнение

$$\frac{dy}{dx} = \frac{-x + \mu(1 - x^2)y}{y}, \quad (2)$$

у которого в числителе правой части стоит многочлен степени три (указанную степень имеет моном  $-\mu x^2 y$ ), часто встречается в приложениях. Оно получается из уравнения Ван дер Поля

$$\frac{d^2x}{dt^2} + \mu(x^2 - 1)\frac{dx}{dt} + x = 0, \quad (3)$$

где  $t$  — время, заменой

$$\frac{dx}{dt} = y, \quad \frac{dy}{dt} = -x + \mu(1 - x^2)y,$$

и последующим делением второго равенства на первое. Поэтому интегральные кривые уравнения (2) можно интерпретировать как траектории движения материальной точки с координатами  $(x(t), y(t))$  в фазовом пространстве  $(x, y = \dot{x})$ , где  $x(t)$  — решение уравнения (3), а «точка» означает производную по времени.

Для уравнения (3) предъявить явное решение не удастся, но путем качественного его исследования можно показать, что среди решений этого уравнения имеется ровно одна замкнутая траектория — цикл. Это почти окружность радиуса 2; «почти», потому что она чуть-чуть продеформирована. Деформация зависит от значения параметра  $\mu$ , входящего в уравнение. Все же остальные решения могут быть описаны следующим образом (см. рис. 3): выберем некоторую точку  $z$  на нормали к этому циклу и рассмотрим интегральную кривую, проходящую через эту точку (на рисунке это кривая, изображенная черным). Эта интегральная кривая уже является незамкнутой: она прокручивается вокруг

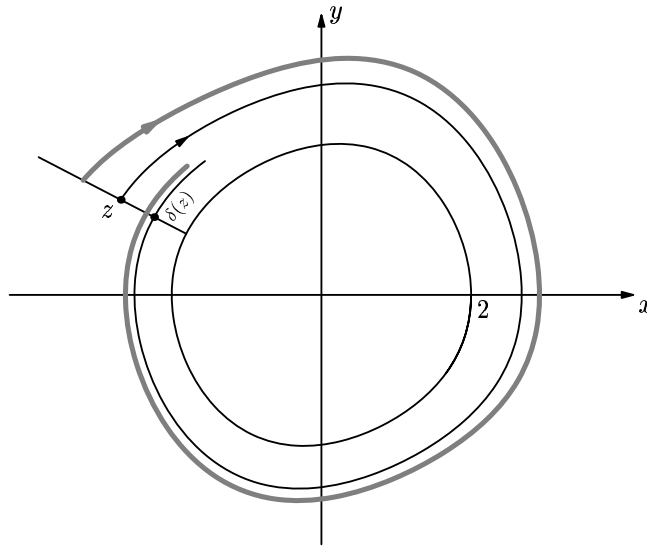


Рис. 3.

цикла и снова пересекает нашу нормаль в некоторой точке  $\delta(z)^1$ , расположенной ближе к циклу, чем точка  $z$ , и так далее. То есть, интегральная кривая с течением времени наматывается на цикл так же, как бесконечно длинная нить наматывается на катушку. И любая другая интегральная кривая в окрестности этого цикла, скажем, изображенная серым, ведет себя точно таким же образом: не пересекаясь ни с циклом, ни с предыдущей интегральной кривой, она тоже, как нить, но уже другого цвета, наматывается на наш цикл.

Другими словами, все интегральные кривые, рассмотренные в окрестности указанного цикла, с течением времени стремятся к этому циклу, неограниченно к нему приближаясь. Вот почему такой цикл называется *предельным*.

Почему же задача о предельных циклах представляет такой интерес? Почему она была выделена Гильбертом в отдельную проблему?

Кроме чисто внутриматематической мотивации (задачу о предельных циклах Гильберт связывает с задачей о числе овалов алгебраической кривой, которая составляет первую часть 16-й проблемы), чрезвычайно важным является то, что эта задача оказывается напрямую связанной с непосредственными приложениями математики. Рассмотрим эту связь на примере того же уравнения Ван дер Поля, которое часто возникает в различных задачах физики и техники, так, например, указанное уравнение описывает работу простейшего лампового генератора.

На практике требуется, чтобы этот прибор генерировал некоторый циклический режим, зависящий от времени  $t$ . Этот циклический режим изображается на плоскости  $(x, y)$  замкнутой траекторией, т. е., циклом. К сожалению, практически невозможно задать начальные условия работы прибора (то есть, выбрать начальную точку на интегральной кривой уравнения) так, чтобы попасть в точности на этот цикл. Возникает естественный вопрос: если мы чуть-чуть изменим начальные условия, т. е. рассмотрим точку, лежащую на близкой кривой, скажем, на черной траектории (см. рис. 3), то как будет работать прибор, что он будет генерировать?

Соответствующий режим работы в этом случае не будет периодическим, тем не менее с течением времени он будет неограниченно приближаться к периодическому, т. е. циклическому режиму. Такой режим работы, такой предельный цикл называется *устойчивым*. Ясно, что решение вопроса о существовании подобных циклов очень важно для практических приложений, для физики и техники.

В разобранным уравнении имеется единственный предельный цикл. Но число таких циклов может быть и большим, стало быть, большим может быть и число возможных устойчивых (и неустойчивых) периодических режимов. Поэтому вопрос о числе предельных циклов уравнения, об их устойчивости также является чрезвычайно важным для приложений.

В 1920-х годах известный французский математик Дюлак опубликовал работу, основной результат которой стал значительным продвижением в решении 16-й проблемы. А именно, он доказал, что *каждое конкретно взятое уравнение рассматриваемого типа имеет лишь конечное число предельных циклов*.

<sup>1)</sup> Отображение  $\delta$ , ставящее в соответствие точке  $z$ , из которой выходит траектория, первую после возвращения точку пересечения траектории с нормалью, называется *отображением последования Пуанкаре*; изучение свойств этого отображения играет важную роль в исследовании 16-й проблемы Гильберта.

Дюлак не получил оценки для числа таких циклов, тем более равномерной оценки, то есть оценки, верной для всех уравнений  $n$ -й степени, что, собственно, и составляет суть проблемы. Он доказал лишь конечность такого числа для каждого конкретного уравнения, но и это был очень важный, нетривиальный результат.

Следующей заметной вехой в исследовании 16-й проблемы стала замечательная работа Петровского и Ландиса, представленная в 1956 году на Всесоюзном съезде математиков в Москве, в которой была анонсирована требуемая оценка для всех значений  $n$ . Но, к сожалению, эта работа оказалась ошибочной, и поправить ее так и не удалось. Несмотря на это, работа Петровского и Ландиса оказала большое влияние на развитие всей аналитической теории дифференциальных уравнений благодаря тем замечательным нетривиальным методам, которые были в ней предложены для исследования 16-й проблемы Гильберта. В частности, в этой работе были заложены основы теории алгебраических комплексных слоений, которая затем интенсивно развивалась усилиями наших и зарубежных математиков.

Прошло всего 10–15 лет, и внимательный анализ показал, что и в классической работе Дюлака имеется ошибка. И так, где-то в 70-х годах выяснилось, что в 16-й проблеме по- существу не сделано практически ничего. Есть интересные подходы, интересные идеи, нетривиальная история, но нет решения.

И вот совсем недавно, в начале 80-х годов, в работах двух ученых — нашего математика Ю. Ильяшенко и французского математика Жана Экаля — было найдено доказательство результата Дюлака о конечности числа предельных циклов для каждого конкретного уравнения. Тем самым было показано, что работа Дюлака может быть поправлена. Доказательство в том и другом случае оказалось очень непростым (Ильяшенко при этом пользовался, в основном, геометрическими методами, а Экаль использовал, по преимуществу, алгебраический язык) и заняло целый том. Что же касается основной части проблемы, а именно той, которая, собственно, и была сформулирована Гильбертом, то до сих пор решения этой задачи нет. Эта проблема представляется чрезвычайно привлекательной для исследователей, благодаря простоте и естественности её формулировки и непосредственным связям с приложениями. Актуальность её — очевидна, тем не менее она все еще ждет своего решения.

\* \* \* \* \*

Следующая задача, являющаяся предметом нашего рассмотрения — 21-я проблема Гильберта. История исследования этой проблемы также оказалась довольно запутанной, в чем-то даже похожей на историю исследования 16-й проблемы. Формулируется 21-я проблема следующим образом:

|| Доказать, что всегда существует линейное дифференциальное уравнение фуксового типа с заданными особыми точками и данной группой монодромии.

Следует отметить довольно жесткую безальтернативную формулировку этой проблемы: Гильберт не ставит вопрос о возможности существования указанного уравнения (например, *существует или нет...*), а просто формулирует соответствующую теорему существования.



Вопрос о построении фуксового уравнения с заданной монодромией рассматривался ещё Б. Риманом в 1856 году, поэтому эта проблема (и её различные модификации) часто называется проблемой Римана – Гильберта.

Исторически рассматривались различные варианты 21-й проблемы Гильберта (см. комментарий на стр. 591 в [Hi]). Наиболее нетривиальным и важным для приложений оказался вариант проблемы, сформулированный для фуксовых систем линейных дифференциальных уравнений, который и станет предметом нашего дальнейшего рассмотрения.

Фуксовой системой линейных дифференциальных уравнений называется система вида

$$\frac{dy}{dz} = \left( \sum_{i=1}^n \frac{B_i}{z - a_i} \right) y, \quad (4)$$

где  $z$  — комплексный аргумент,  $y$  — неизвестная комплекснозначная вектор-функция  $p$  переменных, а  $B_1, \dots, B_n$  — матрицы размера  $(p \times p)$ .

Точки  $z = a_1, \dots, a_n$ , в которых знаменатели коэффициентов системы обращаются в нуль, называются *особыми точками* уравнения.

Неформально говоря, фуксова система уравнений — это система линейных обыкновенных дифференциальных уравнений на комплексной плоскости с простейшими особенностями (знаменатели коэффициентов системы имеют нули минимально возможного, первого порядка в особых точках).

Чтобы понять, что такое группа монодромии, рассмотрим следующий простой пример.

Уравнение

$$\frac{dy}{dz} = \frac{y}{2z},$$

где  $y$  — обычная скалярная функция, является уравнением фуксового типа, в котором особой точкой является точка  $z = 0$ , а матрица коэффициентов имеет размер  $(1 \times 1)$  и совпадает с числом  $1/2$ . Разумеется, это уравнение легко решается. Общее решение имеет вид  $y = c\sqrt{z}$ .

Как хорошо известно из курса комплексного анализа, функция  $\sqrt{z}$  является многозначной функцией. Так, например, в точке  $z = 1$  она может принимать два значения:  $+1$  и  $-1$ . Рассмотрим правую полуплоскость и выделим там то решение, которое в точке  $z = 1$  принимает значение  $+1$ . Обозначим это решение через  $(\sqrt{z})_+$ . Рассмотрим теперь окружность единичного радиуса с центром в особой точке  $z = 0$  и посмотрим, как меняется наше решение при продолжении вдоль этой окружности против часовой стрелки.

Напомним, что каждое комплексное число  $z$ , лежащее на единичной окружности, может быть записано в виде  $z = \cos \varphi + i \sin \varphi$ , где  $\varphi$  — угол, который составляет радиус-вектор числа  $z$  с осью  $x$ -ов; этот угол называется *аргументом* комплексного числа  $z$ . В этих обозначениях нетрудно выписать значение  $(\sqrt{z})_+$  для числа  $z$ , лежащего на верхней половине окружности рядом с числом 1. А именно,

$$(\sqrt{z})_+ = \cos \frac{\varphi}{2} + i \sin \frac{\varphi}{2}.$$

То есть для того, чтобы найти  $(\sqrt{z})_+$ , в данном случае надо просто рассмотреть радиус-вектор с половинным по отношению к заданному углом и отметить

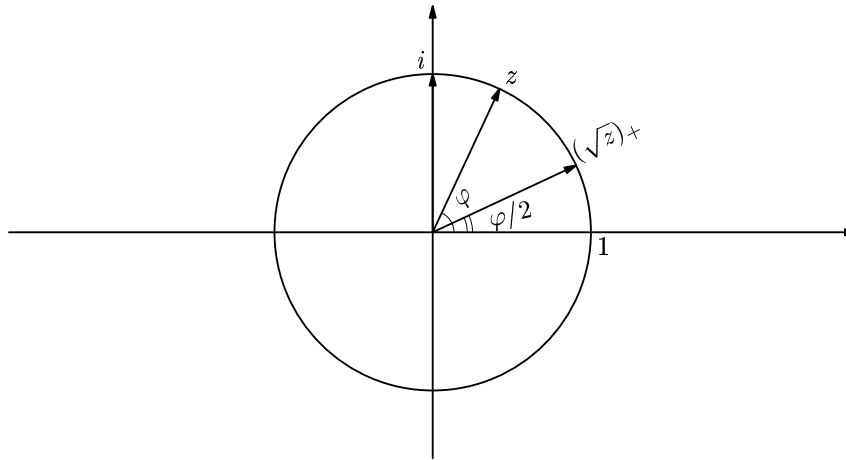


Рис. 4.

соответствующую точку на единичной окружности (см. рис. 4). Полученный ответ нетрудно проверить. Возведем выражение  $\cos \frac{\varphi}{2} + i \sin \frac{\varphi}{2}$  в квадрат. Воспользовавшись известными тригонометрическими формулами и тем, что  $i^2 = -1$ , получим в точности  $\cos \varphi + i \sin \varphi$ , что и требовалось доказать.

Давайте посмотрим, как меняется значение  $(\sqrt{z})_+$ , когда точка  $z$  перемещается вдоль нашей окружности против часовой стрелки. В точке  $i$  угол  $\varphi$  для радиус-вектора, изображающего число  $i$ , равен  $\pi/2$ . Поэтому для соответствующего значения  $(\sqrt{z})_+$  угол радиус-вектора точки будет равен  $\pi/4$ . Подставим в формулу для числа  $z$  вместо угла  $\pi/2$  угол  $\pi/4$ , получим  $(\sqrt{i})_+ = \sqrt{2}/2 + i\sqrt{2}/2$ .

Когда точка  $z$  в процессе движения добирается до точки  $-1$ , аргумент числа  $z$  становится равным  $\pi$ . Значит, согласно той же формуле,  $(\sqrt{-1})_+ = i$ , потому что число с аргументом  $\pi/2$  на единичной окружности соответствует точке  $i$ , и т. д. Когда точка  $z$  возвращается в точку  $z = 1$ , она проходит угол, равный  $2\pi$ . Поэтому, согласно нашему правилу мы должны взять угол  $2\pi$ , разделить его пополам, получить угол  $\pi$ , рассмотреть  $\cos \pi + i \sin \pi$  и получить значение, равное  $-1$ .

Тем самым наше решение, будучи продолженным в точку 1 вдоль окружности, принимает значение, равное не 1, а  $\cos \pi + i \sin \pi = -1$ . Другими словами, продолженное непрерывно вдоль окружности единичного радиуса наше решение  $(\sqrt{z})_+$  переходит в решение, равное  $-(\sqrt{z})_+$ . В таких случаях говорят, что точка 0 является *точкой ветвления* для этого решения, а про само решение — что оно *ветвится* в точке 0, при этом число  $-1$ , характеризующее это ветвление, называют *монодромией* уравнения в точке ноль.

В случае фуксовой системы уравнений (4) общего типа нужно продолжать не одно решение, а сразу всю систему независимых решений. Они также как-то преобразуются при таком продолжении вокруг особой точки: умножаются на константы, к ним прибавляются другие решения. То, как преобразуются решения уравнения при обходе особой точки, и характеризуется монодромией.

В общем случае это уже не одно число (как это было в разобранным примере), а некоторая числовая матрица.

Более точно, если  $(e_1(z), \dots, e_p(z))$  — базис в пространстве решений системы (4), рассмотренной в окрестности некоторой неособой точки  $z_0$ , то после аналитического продолжения вдоль петли  $g_i$ , обходящей особую точку  $a_i$ , этот базис вновь переходит в некоторый базис (вообще говоря, другой)  $(e'_1(z), \dots, e'_p(z))$  пространства решений той же самой системы. Постоянная матрица  $G_i$ , связывающая эти два базиса, и называется *матрицей монодромии* системы (4) в точке  $a_i$ .

Обозначим через  $G_g$  матрицу монодромии базиса  $(e_1(z), \dots, e_p(z))$ , соответствующую аналитическому продолжению вдоль произвольной петли  $g$  с началом в точке  $z_0$ , не проходящей через особые точки. В пространстве всех петель на комплексной плоскости с выколотыми особыми точками можно ввести операцию умножения (последовательный обход двух петель-сомножителей). Если петли рассматривать с точностью до непрерывной деформации, не задевающей особых точек, то эта операция превращает множество петель в группу. Тем самым превращается в группу и множество матриц  $G_g$  относительно операции обычного матричного произведения. Эта группа называется *группой монодромии* системы (4). Нетрудно показать, что её образующими являются матрицы монодромии  $G_1, \dots, G_n$ , удовлетворяющие априори единственному соотношению  $G_1 G_2 \dots G_n = I$  (где через  $I$  обозначена единичная матрица).

Итак, группа монодромии уравнения характеризует то, как преобразуются, ветвясь решения при обходе особых точек. Она задается конечным набором числовых матриц, каждая из которых соответствует своей особой точке.

Возникает естественный вопрос, который можно назвать обратной задачей монодромии. Рассмотрим набор точек на комплексной плоскости и соответствующий абстрактный набор матриц  $G_1, \dots, G_n$ , удовлетворяющий соотношению  $G_1 G_2 \dots G_n = I$ . Спрашивается, всегда ли существует система фуксового типа именно с данными особыми точками и с данной группой монодромии, которая определяется заданным набором матриц?

Именно эту обратную задачу и формулирует Гильберт в своей 21-й проблеме. Причем, формулирует в императивной форме, считая, что эта задача обязательно должна иметь положительное решение.

Какова мотивировка рассматриваемой задачи? Почему она интересна? С чем она связана? Сам Гильберт пишет об этом совсем немного. Он лишь отмечает, что решение указанной проблемы придало бы законченный вид аналитической теории линейных дифференциальных уравнений. То есть, он выдвигает чисто внутриматематическую мотивировку. Но, как это часто бывает с проблемами Гильберта, сформулированные даже в таком узкоспециальном смысле, они впоследствии оказываются важными и для всей математики в целом, и для ее приложений. Мы вернемся к этому вопросу чуть позже. А теперь поговорим немного об истории исследования этой проблемы, которая в чем-то даже более запутана, чем история, связанная с 16-й проблемой.

Уже в 1908 году замечательный югославский математик Племель, работавший во Франции, опубликовал работу, в которой предъявил положительное решение 21-й проблемы Гильберта (см. [P1]). Элегантная работа Племеля, опиравшаяся на использование теории интегральных уравнений Фредгольма,

была чистой теоремой существования и давала полное решение поставленной задачи.

Однако неисследованным оставался вопрос о возможности алгоритмического построения фуксового уравнения по заданным особым точкам и матрицам монодромии (важная для приложений задача, являющаяся естественным развитием исходной постановки Гильберта). Среди результатов, полученных в этом направлении, следует прежде всего отметить выдающуюся работу нашего соотечественника Лаппо-Данилевского, который в 1928 году решил задачу для случая, когда все матрицы  $G_i$  близки к единичной матрице, то есть для случая, когда ветвления очень малы, и когда решения почти не ветвятся.

Следующее значительное продвижение в решении задачи об алгоритмическом построении фуксовой системы по данным монодромии было сделано в 1972 году голландским математиком Деккерсом, который предъявил алгоритм построения уравнения для матриц размера  $(2 \times 2)$ , из которого, в частности, следовало, что для случая системы двух уравнений 21-я проблема всегда имеет положительное решение (что подтверждало результат Племеля в размерности два).

В общем, после работы Племеля исследования в области 21-й проблемы Гильберта продвигались, но не слишком интенсивно. В каком-то смысле, эта задача оказалась на некоторое время на периферии развития математики.

Однако после открытия в начале 1970-х годов метода изомонодромных деформаций<sup>2)</sup> аналитическая теория дифференциальных уравнений получила новый мощный импульс к своему развитию.

Оказалось, что многие нелинейные уравнения математической физики могут быть проинтерпретированы как уравнения изомонодромных деформаций систем линейных дифференциальных уравнений. В качестве примера можно привести знаменитое уравнение, описывающее поведение плазмы, модифицированное уравнение Кортевега де Фриза, некоторые из решений которого выражаются через решения уравнений описанного типа (см. [AI]). При этом важную информацию о поведении решений этих нелинейных уравнений можно получить, исследуя соответствующие изомонодромные деформации линейных систем и, в частности, фуксовых систем дифференциальных уравнений. Но, чтобы построить изомонодромное семейство, надо вначале решить обратную задачу теории монодромии, задачу Римана – Гильберта. Так эта проблема вновь оказалась в центре внимания многих математиков.

Возникла новая мощная мотивировка для изучения этой задачи, появилось много новых интересных работ по этой проблеме. В конце семидесятых годов в Париже начал работать семинар знаменитых французских математиков Б. Мальгранжа, А. Дуади, Буте де Монвеля (последний очень хорошо известен и как математик, успешно работающий в области теоретической физики), на котором интенсивно изучалась указанная проблематика: задача Римана – Гильберта, изомонодромные деформации.

Одновременно у нас в стране работал семинар В. И. Арнольда и Ю. С. Ильяшенко, который занимался похожей тематикой. И почти одновременно в начале восьмидесятых годов и те, и другие обнаружили пробелы в доказательстве Пле-

<sup>2)</sup> В работах американских математиков Фляшки и Ньюэла, а также японских математиков Сато, Дзимбы и Мивы.

меля (см. [A1]). Оказалось, что его доказательство не является полным. Не то, чтобы они нашли ошибку... Ошибки как таковой вроде бы не было. Была обнаружена лакуна в доказательстве. Дело в том, что Племель вначале построил систему линейных дифференциальных уравнений с заданными особыми точками и монодромией в более широком классе, чем это требовалось в проблеме Гильберта, в классе систем с так называемыми *регулярными особыми точками* (т.е. в классе таких систем, все решения которых имеют не более, чем степенной рост в особых точках). Затем он применил к построенной системе некую процедуру, с помощью которой преобразовал её к системе, фуксовой во всех точках, кроме одной, и закончил своё доказательство следующей фразой: «А теперь, действуя точно так же, приведем нашу систему к фуксовой и в последней точке.» Однако, оказалось, что восстановить эту пропущенную часть доказательства не так-то просто. Во всяком случае, сделать это долго никому не удавалось, несмотря на то, что у всех без исключения математиков, работавших в этой области, было полное впечатление, что рассматриваемая проблема должна решаться положительно, как это и было сформулировано Гильбертом.

Тем более неожиданным оказалось появление в конце 1989 года контрпримера к 21-й проблеме Гильберта, полученного автором (см. [Bo1]). Оказалось, что сформулированное в ней утверждение в общем случае *неверно*, и что на самом деле *не всякие данные монодромии могут быть реализованы системой фуксовых уравнений*.

Построенный контрпример относится к случаю четырёх особых точек и матрицам монодромии размера  $(3 \times 3)$ . Это первая размерность и минимальное возможное количество особых точек, при которых такой пример в принципе возможен. Сам пример довольно сложен для изложения: искомая монодромия задаётся в нём неявно, в виде монодромии некоторой системы уравнений с регулярными особыми точками, а затем доказывается, что указанная система уравнений не может быть преобразована к фуксовой. К тому же оказывается, что этот контрпример нестабилен в следующем смысле. При почти любом малом изменении положения особых точек (с сохранением матриц монодромии, т.е. при изомодромной деформации построенной нефуксовой системы), новые данные монодромии (отличающиеся от исходных лишь положением особых точек) уже могут быть реализованы в качестве данных монодромии некоторой фуксовой системы.

Гораздо проще выглядит соответствующий контрпример в размерности четыре. Здесь можно обойтись тремя особыми точками и тремя матрицами монодромии. Рассмотрим специальный тип групп монодромии, который в дальнейшем будем называть *В-монодромией*.

Группа монодромии называется В-монодромией, если набор матриц монодромии приводим, и если жорданова нормальная форма каждой из матриц монодромии  $G_i$  состоит ровно из одной жордановой клетки. Напомним, что набор матриц называется приводимым, если у этих матриц есть общее инвариантное подпространство, отличное от нулевого пространства и от всего пространства  $\mathbb{C}^p$ .

Оказывается, что имеет место следующее утверждение (см. [Bo2]):

*Если В-монодромия может быть реализована как группа монодромии некоторой фуксовой системы, то произведение собственных значений матриц  $G_1, \dots, G_n$  монодромии должно равняться единице.*

Рассмотрим матрицы

$$G_1 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad G_2 = \begin{pmatrix} 3 & 1 & 1 & -1 \\ -4 & -1 & 1 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & -4 & -1 \end{pmatrix}, \quad G_3 = \begin{pmatrix} -1 & 0 & 2 & -1 \\ 4 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 4 & -1 \end{pmatrix}$$

и произвольный набор точек  $a_1, a_2, a_3$ .

Заметим, что  $G_1 \cdot G_2 \cdot G_3 = I$ , матрица  $G_2$  может быть преобразована к матрице  $G_1$ , а матрица  $G_3$  может быть преобразована к жордановой клетке с собственным значением  $-1$ . Действительно, для матрицы  $G_2$  имеем

$$S_2^{-1}G_2S_2 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad S_2 = \frac{1}{3} \begin{pmatrix} 3 & 0 & 0 & 0 \\ -6 & 3 & -3 & 4 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -2 & 3 \end{pmatrix},$$

а для матрицы  $G_3$  получаем

$$S_3^{-1}G_3S_3 = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad S_3 = \frac{1}{64} \begin{pmatrix} 0 & 16 & 4 & 3 \\ 64 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 \\ 0 & 0 & -16 & -12 \end{pmatrix}.$$

У матриц  $G_1, G_2, G_3$  имеется общее инвариантное подпространство размерности два. Стало быть, набор рассматриваемых матриц удовлетворяет определению Б-мондромии. Но произведение собственных значений этих матриц равно  $-1 \neq 1$ . Значит, набор этих матриц не может быть мондромией никакой фуксовой системы. То есть, *указанные данные мондромии доставляют контрпример к 21-й проблеме Гильберта*.

Метод, который удалось применить к решению 21-й проблемы Гильберта, пришел частично из алгебраической геометрии и был иницирован работами Х. Рерля, Т. Левельта, П. Делиня и других замечательных математиков, работавших в этой области. Его суть состоит в том, что удаётся связать между собой асимптотики решений фуксовой системы в особых точках и инварианты некоторого голоморфного векторного расслоения, построенного по исходной мондромии, а затем применить к задаче Римана-Гильберта хорошо разработанные алгебро-геометрические методы.

Полученное отрицательное решение проблемы не означало, что исследования в этой области полностью закончены. Скорее, этот результат поставил массу новых интересных и важных для приложений вопросов. Например, как описать тот класс групп мондромии, которые все же могут быть реализованы фуксовыми системами? На этот вопрос также удалось получить ответ (хотя и неполный)<sup>3)</sup>.

Значение проблем Гильберта состоит еще и в том, что при решении каждой из них обязательно возникают новые интересные методы и идеи, которые могут быть использованы при решении других задач.

<sup>3)</sup>Так, например, автору и независимо В. Костову удалось доказать, что любые данные мондромии с неприводимым набором матриц могут быть реализованы в качестве данных мондромии некоторой фуксовой системы, см. [Bo2].

Так и методы, развитые при решении 21-й проблемы, удалось применить для исследования особых точек уравнений Пенлеве. (Уравнения Пенлеве — это те 6 знаменитых уравнений математической физики, решения которых выступают в качестве новых специальных функций, через которые выражаются решения многих нелинейных дифференциальных уравнений.)

Удалось получить некоторые новые результаты, связанные с теорией векторных расслоений, исследовать расслоение с помощью ассоциированного с ним уравнения. Отметим также результаты, связанные с нормальной формой уравнения в окрестности особой точки, полученные с помощью упомянутых методов.

Другими словами, как это часто случается с проблемами Гильберта, 21-ю проблему, которая естественным образом помещается в раздел обыкновенных дифференциальных уравнений, можно смело поместить в разделы и математической физики, и алгебраической геометрии. Она оказалась тесно связана с этими областями математики.

В заключение этой заметки я бы хотел порекомендовать читателям — учителям, студентам и школьникам — прочитать доклад Гильберта «*Математические проблемы*» (см. [Hi]), особенно первую его часть, которая написана замечательным понятным языком, и в которой говорится о целях математики, о том, каким должно быть математическое доказательство. Прочитав этот доклад, проникнувшись его стимулирующим духом, можно понять, почему он оказал такое влияние на развитие математики XX столетия.

#### СПИСОК ЛИТЕРАТУРЫ

- [AI] *Арнольд В. И., Ильяшенко Ю. С.* Обыкновенные дифференциальные уравнения // Динамические системы – I / Под ред. Д. В. Аносова, В. И. Арнольда. М.: ВИНТИ, 1985. Т. 1. С. 7–149 (Современные проблемы математики. Фундаментальные направления).
- [Bo1] *Болибрух А. А.* Проблема Римана–Гильберта // Успехи Математических Наук. 1990. Т. 45, N 2. С. 3–47.
- [Bo2] *Болибрух А. А.* Фуксовы дифференциальные уравнения и голоморфные расслоения. М.: МЦНМО, 2000. 127 с.
- [Hi] *Гильберт Д.* Избранные труды. Т. 2. М.: Факториал, 1998. 607 с.
- [Pl] *Plemelj J.* Problems in the sense of Riemann and Klein, Interscience. New York, 1964.

# Как математики изучают хаос

Академик РАН Я. Г. Синай

В августе 2000 года я получил приглашение выступить с циклом Хедриковских лекций на заседании Математической ассоциации Америки. Аудитория состояла из математиков, преподающих в различных университетах США, профессиональные интересы которых разнородны и, как правило, далеки от тематики данных лекций. Этот текст представляет собой введение в математическую теорию хаоса и более или менее соответствует фактическому содержанию прочитанных лекций.

## Лекция 1. ХАОС И БЕСПОРЯДОК. ГИПЕРБОЛИЧНОСТЬ

В течение последних тридцати – сорока лет теория хаоса стала весьма популярной. В любом книжном магазине, где есть научный отдел, сегодня можно найти множество книг, в которых объясняется, что такое хаос. Это свидетельствует о том, что хаос — интересное и содержательное явление, причем не только для математика, но и для физика, инженера, экономиста, специалиста, работающего в разных областях приложений науки.

На самом деле то, что ныне зовётся хаосом, существовало в математике задолго до наших дней под именем эргодической теории. Однако физики не могли использовать эти слова, ибо термин «эргодичность» напоминал им эргодическую гипотезу и основания статистической механики. Согласно широко распространённой в теоретической физике точке зрения, жизнь человеческая слишком коротка, чтобы тратить ее на обоснования. Насколько я знаю, слова «детерминированный хаос» были впервые произнесены более тридцати лет назад российским физиком Чириковым и покойным профессором Фордом — физиком из Технологического института штата Джорджия. Этот термин понадобился потому, что в те годы появилось много результатов, относящихся к эргодической теории и ее приложениям.

Теория вероятностей также изучает хаос. Казино, азартные игры, карты — классические примеры хаоса, но это — недетерминированный хаос. Теория хаоса использует методы и основные результаты теории вероятностей, не являясь, однако, ее частью. В теории вероятностей всегда предполагают, что имеется случайный механизм, делающий исходы статистических экспериментов непредсказуемыми, и цель теории — изучать статистику будущего развития и делать разумные прогнозы.

В теории хаоса нет случайного механизма. Наоборот, предполагается наличие детерминированной динамики, детерминированных законов движения и т. п. В этом смысле теория хаоса может рассматриваться как часть теории динамических систем. Малые шумы, которые неизбежно присутствуют в конкретных



ситуациях, приводят лишь к эффектам более высокого порядка малости. Законы динамики определяют индивидуальные траектории движения; теория хаоса изучает статистические свойства этих траекторий.

Если задана траектория, можно поставить вопрос, имеется ли у нее некоторое усредненное по времени поведение. Вот типичные вопросы теории хаоса:

- а) существуют ли средняя частота затмений в нашей Солнечной системе?
- б) является ли Солнечная система устойчивой, иначе говоря, останутся ли расстояния между планетами ограниченными в будущем?
- с) можно ли предсказывать погоду?

После того, как у данной динамической системы установлено существование временных средних, для типичных в определенном смысле траекторий можно поставить вопрос об отклонениях траекторий от их среднего поведения. С точки зрения теории вероятностей это вопрос о справедливости центральной предельной теоремы в ситуациях, когда случайность (в обычном смысле слова) отсутствует.

Уже это простое обсуждение показывает, что в теории хаоса имеется большой параметр  $t$  — время, и нас интересуют различные вопросы, формулируемые в виде предельных теорем при  $t \rightarrow \infty$ . Удивительным открытием пятидесятых — шестидесятых годов было осознание того, что многие явления и идеи, связанные с хаосом, могут относиться к случаю, когда фазовое пространство маломерно и законы динамики выглядят весьма просто. Соответствующую теорию иногда называют малоразмерным хаосом. С другой стороны, имеется множество важных проблем, касающихся систем статистической механики с большим числом степеней свободы. В этих системах два параметра:  $t$  — время и  $N$  — число степеней свободы, и оба стремятся к бесконечности. Один из важнейших вопросов здесь — о стреле времени и необратимости. Если встать на точку зрения Демокрита, согласно которой окружающая среда состоит из слабо связанных между собой одинаковых частиц, то можно задаться вопросом, в какой мере основанные на этой картине модели отражают направленность времени. Конечно, это очень трудный вопрос, до сих пор остающийся открытым. Приведем пример. Если в некотором объеме имеется газ, содержащий 10 частиц, то в процессе эволюции можно обнаружить моменты времени, когда газ занимает только половину объема. Для 100 частиц вероятность этого события гораздо меньше, хотя и положительна. Необходимо наблюдать динамику существенно дольше, прежде чем подобное событие произойдет. Мы можем выбрать некоторый порог и считать, что событие, вероятность которого меньше этого порога, никогда не случается. На самом деле это — очень серьезный шаг; но теория описывает результаты эксперимента только после того, как этот порог выбран. Здесь следует вспомнить знаменитый численный эксперимент Ферми — Паста — Улама. Они обнаружили модель, в которой здравый смысл предсказывал потерю памяти о предыдущих состояниях и сходимости к равновесию, а реальное поведение оказалось совсем иным.

Если система со многими степенями свободы находится в термодинамическом равновесии, то это — состояние пространственного беспорядка, поскольку локальные конфигурации частиц не подчиняются никаким детерминистским законам, а имеют устойчивую предельную статистику при  $N \rightarrow \infty$ . Одним из

главных достижений теории хаоса стало открытие, что большой параметр  $t$  в динамике играет ту же роль, что и большой параметр  $N$  в статистической механике. В этом смысле детерминированный хаос есть динамический беспорядок. Соответствующая теория называется в динамике термодинамическим формализмом. Ниже мы приводим пример теоремы из этой теории, относящийся к предмету нашего обсуждения.

Пусть  $z = (z_1, \dots, z_d)$  — точка  $d$ -мерной решетки,  $z_i$  — целые числа. Для большинства наших задач достаточно рассматривать  $d = 1$ . Но важно, однако, иметь общую картину для произвольного  $d$ . Предположим, что  $z \in V$ , где  $V$  — ограниченная область, и в каждой точке  $z \in V$  есть переменная  $\sigma(z)$ , принимающая конечное число значений. Типичным примером является  $\sigma(z) = \pm 1$ , когда в каждой точке выбирается случайный знак. Рассмотрим функцию  $U(\sigma(z); \sigma(z'), |z - z'| \leq r)$ , которую назовём потенциалом взаимодействия переменной  $\sigma(z)$  с её соседями в радиусе  $r$ . Зафиксируем конфигурацию  $\bar{\sigma}(\partial V) = \{\sigma(z) \mid z \notin V, \text{dist}(z, V) \leq r\}$ . Сумма

$$\sum_{z \in V} U(\sigma(z); \sigma(z'), |z - z'| \leq r) = U(\sigma(z), z \in V \mid \bar{\sigma}(\partial V))$$

называется энергией конфигурации  $\sigma(z), z \in V$  при граничном условии  $\bar{\sigma}(\partial V)$ . Потом такие конфигурации будут использоваться для кодирования траекторий динамических систем. Типичный пример — так называемая ферромагнитная модель Изинга, где  $\sigma(z) = \pm 1$ ,  $r = 1$  и

$$U(\sigma(z); \sigma(z'), |z - z'| \leq 1) = \sum_{z': |z - z'| = 1} (\sigma(z) - \sigma(z'))^2 + h\sigma(z).$$

Последняя сумма при  $h = 0$  есть дискретный вариант известного интеграла Дирихле

$$\int_V (\nabla \sigma, \nabla \sigma) dz.$$

Однако во многих отношениях модель Изинга далека от её непрерывного аналога.

Выражение

$$\pi_\beta(\sigma(z), z \in V; \bar{\sigma}(\partial V)) = e^{-\beta U(\sigma(z), z \in V \mid \bar{\sigma}(\partial V))}$$

называется статистическим весом конфигурации  $\sigma(z), z \in V$ . Параметр  $\beta$  — обратная температура. Введем обозначение

$$\Xi(V \mid \beta, \bar{\sigma}(\partial V)) = \sum_{\{\sigma(z), z \in V\}} \pi_\beta(\sigma(z), z \in V; \bar{\sigma}(\partial V)).$$

Распределение Гиббса в объеме  $V$  при граничном условии  $\bar{\sigma}(\partial V)$  и обратной температуре  $\beta$  — это вероятностное распределение на конфигурациях  $\sigma(z), z \in V$ , где

$$p(\sigma(z), z \in V) = \frac{1}{\Xi(V \mid \beta, \bar{\sigma}(\partial V))} \pi_\beta(\sigma(z), z \in V; \bar{\sigma}(\partial V)).$$

Можно без труда обобщить все предыдущие понятия на случай, когда множество

значений  $\sigma(z)$  бесконечно или даже континуально, а потенциал  $U$  имеет бесконечный радиус взаимодействия  $r$ . Такие выражения часто возникают из динамики.

Первый вопрос — о пределе гиббсовского распределения при  $V \rightarrow \infty$  и зависимости этого предела от граничных условий. Пределы, получающиеся таким способом, могут быть определены с помощью так называемых уравнений Добрушина – Лэнфорда – Рюэля. О ситуациях, когда эти пределы зависят от граничных условий, говорят как о ситуациях, имеющих дальний порядок; он связан с фазовыми переходами первого рода. Сейчас есть разнообразные методы доказательства существования дальнего порядка. В частности, дальний порядок есть в упомянутой выше модели Изинга при  $d > 1$ ,  $h = 0$  и достаточно больших  $\beta$ . Однако задачи о фазовых переходах достаточно далеки от теории хаоса. Причина в том, что дальнего порядка заведомо нет при  $d = 1$ , потому что в этом случае распределения Гиббса — марковские цепи с конечной памятью. Впрочем, кое-что интересное может происходить, если  $r = \infty$ . Рассмотрим пример теоремы, имеющей приложения в динамике и теории хаоса.

Предположим, что задана функция  $p(\sigma(0) | \sigma(-1), \sigma(-2), \dots, \sigma(-n), \dots)$ , удовлетворяющая условиям:

- а)  $p(\sigma(0) | \sigma(-1), \sigma(-2), \sigma(-3), \dots, \sigma(-n), \dots) > \text{const} > 0$ ;
- б)  $\sum_{\sigma(0)} p(\sigma(0) | \sigma(-1), \sigma(-2), \sigma(-3), \dots, \sigma(-n), \dots) = 1$ .

Тогда  $p(\sigma(0) | \sigma(-1), \sigma(-2), \dots, \sigma(-n), \dots)$  можно рассматривать как условную вероятность  $\sigma(0)$  при заданной предыстории  $\sigma(-1), \sigma(-2), \dots, \sigma(-n), \dots$ . Зададимся вопросом: как много вероятностных распределений  $P$  совместимы с этой условной вероятностью? Последнее означает, что

$$\lim_{n \rightarrow \infty} \frac{P(\sigma(0), \sigma(-1), \dots, \sigma(-n))}{P(\sigma(-1), \dots, \sigma(-n))} = p(\sigma(0) | \sigma(-1), \dots, \sigma(-n), \dots).$$

Если  $p$  зависит только от конечного числа переменных, т.е.  $p = p(\sigma(0) | \sigma(-1), \sigma(-2), \sigma(-3), \dots, \sigma(-n))$ , то обычная теория цепей Маркова говорит, что  $P$  единственно. Для того, чтобы единственность имела место и при  $n = \infty$ , функция  $p$  должна достаточно хорошо приближаться функциями конечного числа переменных. Положим

$$\alpha_n = \sup_{\sigma, \sigma', \sigma''} \left| \frac{p(\sigma(0) | \sigma(-1), \dots, \sigma(-n), \sigma'(-n-1), \dots)}{p(\sigma(0) | \sigma(-1), \dots, \sigma(-n), \sigma''(-n-1), \dots)} - 1 \right|.$$

**ТЕОРЕМА.** Если  $\sum_n \alpha_n < \infty$ , то  $P$  единственно.

Если  $\sum_n \alpha_n = \infty$ , то, вероятно, может проявиться эффект дальнего порядка и  $P$  может стать неединственным.

В задачах теории хаоса  $P$  обычно единственно. Следующее, что нам нужно, — это так называемые временные корреляционные функции  $E_p(\sigma(0)\sigma(z))$  и скорость их сходимости к пределу при  $z \rightarrow \infty$ .

Теория хаоса состоит из двух основных частей. В первой из них главная задача — доказать, что некоторая динамическая система (или класс динамических систем) неустойчива или гиперболична. Вторая — это теория, показывающая,

что неустойчивые динамические системы имеют особые свойства, которые можно ассоциировать с хаосом.

Идея о связи неустойчивости и хаоса прослеживается уже в статьях и книгах Пуанкаре (особенно в его учебнике по теории вероятностей, где он обсуждает происхождение случайных событий), в работах Адамара, Биркгофа, Морса.

В более близкие к нам времена (тридцать – сорок лет назад) эта тема была центральной в работах физиков Н. С. Крылова, Б. Чирикова, Г. Заславского, Дж. Форда и др. В математике топологические аспекты этой идеи развивались С. Смейлом, Д. Аносовым, М. Шубом, Дж. Пейлисом и др. Математическая теория хаоса появилась в работах Д. Рюэля, Р. Боуэна, В. Оселедца, Я. Песина, моих и др.

Вывод основных свойств хаоса из неустойчивости является сейчас хорошо разработанной математической теорией. Мы обсудим ее в лекции 3. Во многих случаях основная трудность состоит в том, чтобы для данной динамической системы доказать достаточно сильную неустойчивость. В этой лекции мы обсудим общее понятие динамической неустойчивости.

Пусть мы имеем систему обыкновенных дифференциальных уравнений

$$\frac{dx_i}{dt} = f_i(x_1, \dots, x_d), \quad 1 \leq i \leq d, \quad (1)$$

здесь  $d$  — размерность фазового пространства. Функции  $f_i$  достаточно гладкие и регулярные, чтобы для любого начального условия  $x(0) = (x_1(0), \dots, x_d(0))$  система (1) имела единственное решение  $x(t) = (x_1(t), \dots, x_d(t))$ , определенное для всех  $t$ ,  $-\infty < t < \infty$ .

Рассмотрим малое возмущение начальных условий  $x'(0) = (x'_1(0), \dots, x'_d(0))$ ,  $x'_i(0) = x_i(0) + \delta x_i(0)$ . Вектор  $\delta x(0) = (\delta x_1(0), \dots, \delta x_d(0))$  характеризует размер возмущения. Возмущенное решение  $x'(t)$  отклоняется от начального решения  $x(t)$  и можно задать вопрос: насколько велико время  $t$ , при котором  $\|x'(t) - x(t)\| = O(1)$ , т. е.  $x'(t) - x(t)$  по порядку сравнимо с характерным размером фазового пространства? (В конкретных ситуациях обычно легко определить точное значение этой величины.) Чем меньше время  $t$ , тем более неустойчива динамика. Предположим, что  $\delta x(0) \rightarrow 0$  вдоль некоторого направления и  $t \sim \lambda \ln \|\delta x(0)\|$  для некоторого  $\lambda > 0$ . Конечно,  $\lambda$  зависит от выбранного направления, которое называется в этом случае неустойчивым. Последняя формула означает, что  $\|\delta x(t)\| \asymp e^{\lambda t} \|\delta x(0)\|$ . Однако правая часть значима только при таких  $t$  и  $\delta x(0)$ , что величина  $e^{\lambda t} \|\delta x(0)\|$  сравнима по порядку с характерным размером фазового пространства. Если  $t \rightarrow -\infty$ , то соответствующее направление называется устойчивым (терминология предложена Смейлом). Число  $\lambda$  в этом выражении называется показателем Ляпунова. Множество всех возможных  $\lambda$  называется ляпуновским спектром (данной траектории).

Точка  $x(0)$  называется гиперболической, если  $\mathbb{R}^d = R^1(x(0)) + E^{(u)}(x(0)) + E^{(s)}(x(0))$ , где  $R^1(x(0))$  — одномерное подпространство, порожденное вектором  $f(x(0)) = (f_1(x(0)), \dots, f_d(x(0)))$ ,  $E^{(u)}$  и  $E^{(s)}$  — подпространства, порожденные неустойчивыми и устойчивыми направлениями соответственно. Понятия гиперболической неподвижной точки и периодической орбиты хорошо известны в анализе, гидродинамике, прикладной математике. Теория хаоса имеет дело с

общими гиперболическими траекториями. Система (1) называется гиперболической, если каждая точка  $x(0)$  — гиперболическая. Гиперболичность может быть равномерной и неравномерной. Современное понятие гиперболичности введено Аносовым и Смейлом.

Пусть  $\dim E^{(u)}(x(0)) = k$ ,  $\dim E^{(s)}(x(0)) = l$  и эти размерности не зависят от  $x(0)$ .  $C^1$ -подмногообразие  $\Gamma^{(u)}(x)$  называется неустойчивым многообразием точки  $x = x(0)$ , если  $\dim \Gamma^{(u)}(x) = k$ , касательная плоскость к  $\Gamma^{(u)}(x)$  в точке  $x$  совпадает с  $E^{(u)}(x)$  и  $\Gamma^{(u)}(x) = \{y : \text{dist}(x(t), y(t)) \rightarrow 0 \text{ при } t \rightarrow -\infty\}$ . Аналогично,  $l$ -мерное  $C^1$ -подмногообразие  $\Gamma^{(s)}(x)$  называется устойчивым многообразием точки  $x = x(0)$ , если  $\dim \Gamma^{(s)}(x) = l$ , касательная плоскость к  $\Gamma^{(s)}(x)$  в точке  $x$  совпадает с  $E^{(s)}(x)$  и  $\Gamma^{(s)}(x) = \{y : \text{dist}(x(t), y(t)) \rightarrow 0 \text{ при } t \rightarrow \infty\}$ . Аналогичные определения можно дать в более общем случае гладких векторных полей на гладких многообразиях или гладких итераций на гладких многообразиях. Во всех этих случаях имеется однопараметрическая группа преобразований фазового пространства, образованная сдвигами вдоль траекторий. Такие группы и называются динамическими системами.

Утверждение, что при некоторых дополнительных условиях гиперболические точки имеют устойчивые и неустойчивые многообразия, называется теоремой Адамара – Перрона или теоремой об инвариантных многообразиях.

Ниже нам потребуется также понятие инвариантной меры. Нормированная мера  $\mu$  называется инвариантной для данной динамической системы, если для любой ограниченной измеримой функции  $\varphi$

$$\int \varphi(x(0)) d\mu(x(0)) = \int \varphi(x(t)) d\mu(x(0)).$$

Мера  $\mu$  абсолютно непрерывна (по отношению к мере Лебега), если  $d\mu(x) = p(x)dx_1 \cdot \dots \cdot dx_d$ . В этом случае функция  $p$  называется плотностью меры. Мы обсудим роль инвариантных мер в лекции 3.

## ЛЕКЦИЯ 2. ОСНОВНЫЕ ПРИМЕРЫ ГИПЕРБОЛИЧЕСКИХ ДИНАМИЧЕСКИХ СИСТЕМ

1. **Одномерные отображения.** Пусть  $f$  — определенная на  $[0, 1]$  кусочно  $C^2$ -гладкая функция, для которой  $f(x) \in [0, 1]$  при  $0 \leq x \leq 1$ . Такая функция порождает итерации  $T_f^n: T_f(x) = f(x)$ ,  $T_f^n(x) = f(T_f^{n-1}(x))$ . Отображение  $T_f$  — гиперболическое, если  $f'(x) > 1$  при всех  $x$ . Это неравенство может выполняться только для разрывных  $f$ , причем в точках разрыва имеются в виду левая и правая производные. Интересно отметить, что в этом и других случаях разрывность динамики не представляет серьезной трудности. Простейшая ситуация возникает, когда  $f$  обладает марковским свойством: отрезок  $[0, 1]$  можно разбить на конечное число подинтервалов  $[x_i, x_{i+1}]$  так, чтобы на каждом интервале  $f(x_i) = 0$ ,  $f(x_{i+1}) = 1$ , т.е.  $T_f[x_i, x_{i+1}] = [0, 1]$ . У гиперболического  $T_f$  всегда есть абсолютно непрерывная инвариантная мера. В марковском случае это — теорема Реньи, она может быть доказана при помощи теоремы из первой лекции. Общее утверждение доказано Ледрапье, Келлером, Хофбауэром и другими.

Самый известный случай в теории одномерных отображений — квадратичные отображения  $x \mapsto rx(1-x)$ ,  $0 \leq x \leq 1$ . Здесь  $r$  — параметр,  $0 \leq r \leq 4$ . Эти отображения изучались в голоморфной динамике, имеющей дело с итерациями голоморфных функций (Фату, Жюлиа, Карлесон, Милнор, Любич, Макмюлен, Дуади, Хаббард и многие другие). Они оказались важны для так называемой универсальности Фейгенбаума, описывающей возникновение хаоса.

Для хаоса важно, чтобы у отображения была абсолютно непрерывная инвариантная мера. Центральным результатом здесь является следующая теорема Якобсона.

**ТЕОРЕМА.** *Множество значений  $r$ , при которых у  $T_f$  есть абсолютно непрерывная инвариантная мера, имеет положительную меру Лебега.*

Недавно Любич доказал, что  $[0, 4] = C_1 \cup C_2 \cup C_3$ , где  $l(C_1) = 0$ ,  $l(C_2) > 0$  и для всех  $r \in C_2$  преобразование  $T_f$  имеет абсолютно непрерывную инвариантную меру, а для всех  $r \in C_3$  у  $T_f$  есть периодическая траектория, притягивающая почти все точки отрезка  $[0, 1]$ . Множество  $C_3$  открыто и всюду плотно. Множество значений  $r$ , при которых  $T_f$  гиперболично, имеет полную меру.

2. **ДВУМЕРНЫЕ ОТОБРАЖЕНИЯ.** Простейший пример двумерного гиперболического отображения — групповой автоморфизм двумерного тора,  $T(x_1, x_2) = (ax_1 + bx_2, cx_1 + dx_2)$ , где  $x = (x_1, x_2)$  — точка  $\mathbb{T}^2$ , координаты берутся по модулю 1, коэффициенты  $a, b, c, d$  — целые и  $ad - bc = 1$ . Отображение  $T$  гиперболично, если собственные числа  $\lambda_1, \lambda_2$  матрицы  $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$  таковы, что  $|\lambda_1| > 1 > |\lambda_2|$ . Неустойчивые и устойчивые многообразия  $\Gamma^{(u)}(x)$ ,  $\Gamma^{(s)}(x)$  — прямые, идущие вдоль направлений собственных векторов, соответствующих  $\lambda_1$  и  $\lambda_2$ .

3. **СТАНДАРТНЫЕ ОТОБРАЖЕНИЯ.** Рассмотрим периодическую гладкую функцию  $V(x)$  периода 1, имеющую две невырожденные критические точки. Отображения двумерного тора  $T(x_1, x_2) = (x'_1, x'_2)$ , где  $x'_1 = x_1 + kV(x) \pmod{1}$ ,  $x'_2 = x_2 + x'_1 \pmod{1}$ ,  $k$  — параметр, называются стандартными отображениями. Они были введены Б. В. Чириковым и британским физиком Тейлором и широко используются в физике плазмы, нелинейных колебаниях и т. п. Стандартные отображения при малых значениях  $k$  дают простейшие примеры применения теории КАМ (Колмогоров – Арнольд – Мозер). В некотором смысле стандартные отображения являются двумерными аналогами квадратичных отображений. Как только стандартные отображения стали популярными, появилась гипотеза, что множество гиперболических точек имеет положительную меру Лебега для всех  $k > 0$ . Эта очень трудная проблема все еще остается открытой.

4. **ГЕОДЕЗИЧЕСКИЕ ПОТОКИ НА ПОВЕРХНОСТЯХ ПОСТОЯННОЙ ОТРИЦАТЕЛЬНОЙ КРИВИЗНЫ И ДРУГИХ МНОГООБРАЗИЯХ.** Эти динамические системы являются естественными аналогами групповых автоморфизмов  $Tor^2$ . Зададим на верхней полуплоскости  $H = \{z : \text{Im } z > 0\}$  комплексной плоскости риманову метрику  $ds^2 = (dx^2 + dy^2)/y^2$ .

Хорошо известно, что  $H$  с такой метрикой имеет гауссову кривизну  $-1$ . Чтобы построить пример геодезического потока, возьмем группу  $\Gamma = \left\{ \begin{vmatrix} a & b \\ c & d \end{vmatrix} \right\}$ ,  $a, b, c, d$  — целые и  $ad - bc = 1$ . Можно выбрать фундаментальную область

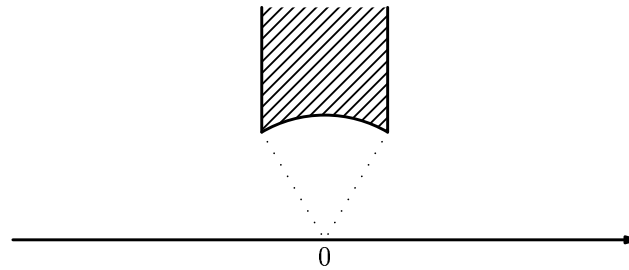


Рис. 1.

$Q = H/\Gamma$ , как изображено на рисунке 1. Легко подсчитать, что площадь  $Q$  конечна.

Для построения других примеров нужно брать различные дискретные подгруппы группы  $SL(2, \mathbb{R})$  действительных матриц  $\begin{pmatrix} x & y \\ z & w \end{pmatrix}$ ,  $xw - zy = 1$ .

Область  $Q$  играет важную роль в теории чисел, теории представлений, квантовом хаосе. Фазовое пространство  $M$  геодезического потока — это трехмерное многообразие единичных касательных векторов к  $Q$ . Геодезический поток — это динамическая система в  $M$ , перемещающая каждый вектор с единичной скоростью вдоль определяемой этим вектором геодезической.

Геодезические потоки можно ввести на любом римановом многообразии. В случае многообразий отрицательной кривизны они имеют особые свойства равномерно сильной гиперболичности. Устойчивые и неустойчивые многообразия — хорошо известные в гиперболической геометрии объекты, называемые орисферами. В двумерном случае устойчивые и неустойчивые многообразия — это одномерные семейства векторов, ортогональных к окружностям, касающимся прямой  $\text{Im } z = 0$ , или единичные векторы, ортогональные к горизонтальным прямым. Орисферы (в двумерном случае — орициклы) можно получать как пределы окружностей (см. рис. 3).

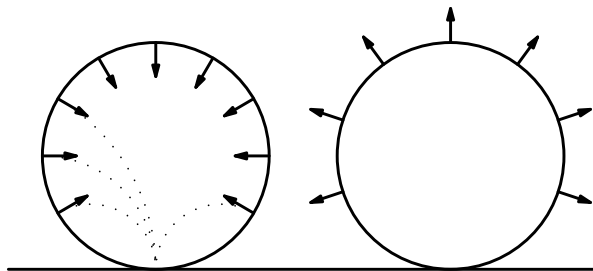


Рис. 2.

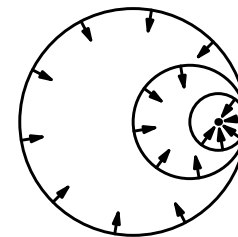


Рис. 3.

5. СИСТЕМЫ АНОСОВА. Динамические системы с однородной гиперболичностью всех траекторий называются системами Аносова (в честь русского математика Аносова, который их ввел). Они включают в себя групповые автоморфизмы торов и их нелинейные возмущения, геодезические потоки на компактных многообразиях отрицательной кривизны и некоторые другие системы.

6. РАССЕЙВАЮЩИЕ, или ГИПЕРБОЛИЧЕСКИЕ БИЛЛИАРДЫ. Биллиарды — это динамические системы, соответствующие равномерному движению точки положительной массы внутри некоторой области  $Q$ . Точка отражается от границы так, что угол падения равен углу отражения. Динамика такого движения разрывна. Свойства биллиарда зависят от свойств его границы  $\partial Q$ . Биллиард называется гиперболическим или рассеивающим, если его граница  $\partial Q$  — вогнута (см. рис. 4).

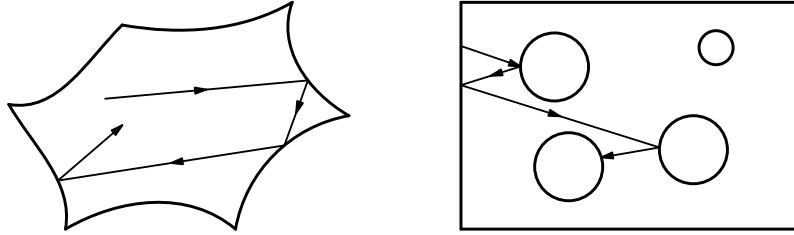


Рис. 4.

Как и в случае геодезических потоков на многообразиях отрицательной кривизны, типичная точка имеет устойчивое и неустойчивое многообразия. Примеры для двумерного случая приведены на рис. 5. Особые точки на этих многообразиях, имеющие вид клюва, соответствуют траекториям, которые в некоторый момент касаются границы. В этих точках динамика разрывна. Знаменитая система упруго сталкивающихся твердых шаров может также быть представлена как биллиардная система. Бунимович доказал гиперболичность биллиарда внутри стадиона (стадион Бунимовича). Газу из твердых сфер и другим биллиардам посвящена большая книга под редакцией Д. Саса, которая недавно вышла в издательстве Шпрингер.

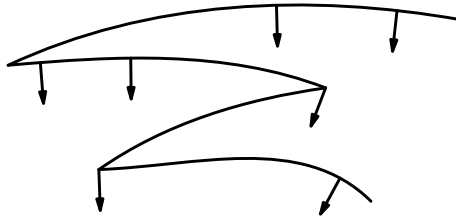


Рис. 5.

7. СИСТЕМЫ СМЕЙЛА С АКСИОМОЙ А. АТТРАКТОР ЛОРЕНЦА. С. Смейл предложил рассматривать динамические системы, у которых сильная однородная гиперболичность достигается на предельном множестве, притягивающем все траектории. Смейл назвал такие системы системами, удовлетворяющими аксиоме А. Они играют большую роль в топологической теории динамических систем.

В 1962 году Е. Лоренц привел пример системы из трех обыкновенных дифференциальных уравнений, которая обладает многими свойствами систем с ак-



сиемой А. Она имеет очень простой вид

$$\begin{aligned}\frac{dx_1}{dt} &= -\sigma x_1 + \sigma x_2, \\ \frac{dx_2}{dt} &= r x_1 - x_2 - x_1 x_3, \\ \frac{dx_3}{dt} &= -b x_3 + x_1 x_2.\end{aligned}\tag{2}$$

Здесь  $\sigma, r, b$  — числовые параметры. Е. Лоренц открыл эту систему при исследовании знаменитой проблемы Бенара в гидродинамике. Он обнаружил при численных расчетах, что система (2) при некоторых значениях  $\sigma, r, b$  проявляет сильные свойства гиперболичности. Десятью годами позже эта система стала весьма популярной, и идеи хаоса начали проникать в физику. Примерно в это же время Рюэль и Такенс (РТ) предложили понятие странного аттрактора применительно к проблеме турбулентности. Строго говоря, система (2) не удовлетворяет определению РТ, однако очень близка к нему. Объяснение гиперболичности (2) было дано Р. Вильямсом и (несколько иным способом) Афрамовичем, Быковым, Шильниковым. Недавно Такер при помощи компьютерных вычислений нашел доказательство гиперболичности системы Лоренца (2).

8. ОТОБРАЖЕНИЕ ЭНО И АТТРАКТОР ЭНО. Вдохновленный работой Е. Лоренца, французский астроном М. Эно предложил дискретную версию системы Лоренца (2), в численных экспериментах демонстрирующую аналогичные свойства. Она имеет вид  $(x_1, x_2) \rightarrow (1 + x_2 - a x_1^2, b x_1)$ , где  $a, b$  — параметры. М. Бенедик и Л. Карлесон доказали замечательные результаты о гиперболичности отображений Эно.

9. ПОДКОВА СМЕЙЛА, ГОМОКЛИНИЧЕСКИЕ И ГЕТЕРОКЛИНИЧЕСКИЕ ТРАЕКТОРИИ. Всё это — механизмы, порождающие инвариантные множества, состоящие из гиперболических траекторий. Обычно такие множества имеют меру нуль и в некотором смысле исключительны. Тем не менее, траектории на этих множествах проявляют хаотическое поведение. Преимущество подковы Смейла, гомоклинических и гетероклинических траекторий в том, что во многих случаях их гораздо легче обнаружить аналитическими или численными методами.

### ЛЕКЦИЯ 3. МАТЕМАТИКА УСТОЯВШЕГОСЯ ХАОСА

Из гиперболичности и неустойчивости траекторий легко следует, что в таких системах число возможных типов траекторий растёт экспоненциально. Многие из них — хаотические, потому что общее число нехаотических траекторий растёт медленнее, чем экспоненциально. Точная формулировка терминов и понятий, относящихся к этому вопросу, связана с теорией сложности динамических систем. Мы не будем здесь её обсуждать.

Детерминированная динамика хаотична, если она обладает свойствами случайных событий или случайных экспериментов из теории вероятностей. Первой основной теоремой теории вероятностей является закон больших чисел о существовании и устойчивости временных средних. Он имеет естественный аналог

в динамике. А именно, определим для любой ограниченной функции  $\varphi$  её временное среднее

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_u^{u+t} \varphi(x(s)) ds = \bar{\varphi}(x(0))$$

при условии, что этот предел существует. Он не зависит от  $u$ , но может зависеть от  $x(0)$ . Эргодическая теорема Биркгофа утверждает, что если  $\mu$  — инвариантная мера, то временное среднее существует для  $\mu$ -почти всех точек  $x(0)$ . Таким образом, первый шаг в анализе хаоса состоит в выборе инвариантной меры  $\mu$ . Обычно динамические системы имеют много типов траекторий: периодические траектории, асимптотические траектории, транзитные траектории и т. п. Если мы фиксируем  $\mu$ , то мы фиксируем некоторое множество траекторий с однородным режимом.

Есть общие теоремы о существовании по крайней мере одной инвариантной меры (один из примеров — теорема Боголюбова – Крылова для динамических систем с компактным фазовым пространством). Во многих случаях существование естественных инвариантных мер следует из законов динамики. Например, так называемые гамильтоновы системы, групповые автоморфизмы, геодезические потоки на компактных римановых многообразиях, бильярды имеют абсолютно непрерывные инвариантные меры.

Известен общий способ выбора естественной инвариантной меры, называемый иногда принципом Рюэля. Возьмём любую нормированную плотность  $\rho(x; 0)$ ,  $\int \rho(x; 0) dx_1 dx_2 \dots dx_d = 1$ . Она порождает меру  $\mu_0$ , для которой  $\frac{d\mu_0}{dx} = \rho(x; 0)$ . Для любого  $t > 0$  и ограниченной  $\varphi$

$$\int \varphi(x(t)) d\mu_0(x(0)) = \int \varphi(x(t)) \rho(x; 0) dx = \int \varphi(y) \rho(y; t) dy.$$

Плотность  $\rho(y; t)$  удовлетворяет уравнению Лиувилля

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho, f) = 0, \quad f(y) = (f_1(y), \dots, f_d(y))$$

или его аналогу для случая итераций. Если существует предел

$$\lim_{t \rightarrow \infty} \int \varphi(x(t)) d\mu_0(x(0)) = \int \varphi(y) d\mu_\infty(y),$$

где  $\mu_\infty$  — предельная мера и  $\mu_\infty$  не зависит от  $\mu_0$ , то  $\mu_\infty$  — естественная мера. Теорема Якобсона и результат Бенедикса – Карлесона, упомянутые в предыдущей лекции, гарантируют существование  $\mu_\infty$  для соответствующих систем. Для гамильтоновых систем и других систем, названных выше, трудной задачей является доказательство того, что  $\mu_\infty$  совпадает с естественной абсолютно непрерывной мерой. Эта задача сводится к доказательству перемешивающего свойства динамики и будет обсуждаться ниже.

Теперь предположим, что для данной динамической системы выбрана инвариантная мера  $\mu$ . Теорема Биркгофа позволяет временным средним  $\bar{\varphi}(x(0))$  зависеть от  $x(0)$ . Мера  $\mu$  называется эргодической, если  $\bar{\varphi}(x(0)) = \int \varphi(y) d\mu(y)$ . В этом случае эргодическая теорема принимает вид обычного закона больших чисел. Терминология происходит из статистической механики, а понятие эргодичности в немного ином виде было предложено Л. Больцманом.

Обычно очень трудно доказать, что данная инвариантная мера эргодична. Например, мы сейчас знаем благодаря КАМ теории, что многие гамильтоновы системы, описывающие динамику взаимодействующих частиц, неэргодичны. Однако для гиперболических динамических систем имеется общее рассуждение Э.Хопфа, которое устанавливает эргодичность. Мы опишем это рассуждение, но, конечно, это не будет полным строгим доказательством. Для простоты ограничимся случаем компактного фазового пространства. Достаточно рассмотреть непрерывные функции  $\varphi$ . Рассмотрим для точки  $x$  её локальное устойчивое многообразие  $\gamma^{(s)}(x)$ , т.е. малую окрестность  $x$ , принадлежащую устойчивому многообразию  $\Gamma^{(s)}(x)$ ,  $\gamma^{(s)}(x) \subset \Gamma^{(s)}(x)$ . Если существует  $\bar{\varphi}(x)$ , то и  $\bar{\varphi}(y)$ ,  $y \in \gamma^{(s)}$  также существует, причём  $\bar{\varphi}(y) = \bar{\varphi}(x)$ . Таким образом, пока  $x$  движется вдоль локальных устойчивых многообразий, временные средние в положительном направлении времени остаются теми же самыми. При движении  $x$  вдоль траектории временные средние сохраняются по очевидным причинам. Одно из следствий эргодической теоремы Биркгофа утверждает, что для почти всех  $y$

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_{-t}^0 \varphi(x(s)) ds = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \varphi(x(s)) ds$$

и оба предела существуют. Это показывает, что временные средние одни и те же для почти всех точек локальных неустойчивых многообразий. В гиперболических системах можно перейти от начальной точки к любой точке в малой окрестности начальной, двигаясь вдоль траекторий (1) и локальных устойчивых и неустойчивых многообразий. Поэтому временное среднее  $\bar{\varphi}(y)$  постоянно в малой окрестности  $x$ . Как правило, не слишком трудно перейти от локальной эргодичности к глобальной.

Э.Хопф использовал это рассуждение, чтобы доказать эргодичность геодезических потоков на компактных многообразиях отрицательной кривизны. Позднее оно применялось к системам Аносова, системам с аксиомой А и т.д. Однако оказалось, что применение рассуждения Хопфа к рассеивающим гиперболическим билиардам — трудная задача (Бунимович, Чернов, Синай, Шимани, Сас и др.).

Эргодическая теорема для марковских цепей утверждает, что при естественных условиях любое начальное распределение сходится к единственному стационарному распределению. Аналогом в динамике является упомянутый выше принцип Рюэля. В случае абсолютно непрерывной инвариантной меры с плотностью  $\rho$  это означает, что

$$\int \varphi(x) \rho(x, t) dx \rightarrow \int \varphi(x) \rho(x) dx, \quad t \rightarrow \infty. \quad (3)$$

В общей эргодической теории последнее свойство называется перемешиванием. Конечно, отсюда не следует поточечная сходимость  $\rho(x, t)$  к  $\rho(x)$ . Простые примеры показывают, что функция  $\rho(\cdot, t)$  при больших  $t$  становится очень нерегулярной и (3) выполняется вследствие того, что содержит некоторое усреднение. Во многих важных случаях предельная мера  $\mu_\infty$  сингулярна по отношению к мере Лебега.

Для гиперболических систем доказательство (3) и сходимости к пределу состоят из двух шагов, и оба шага можно естественно интерпретировать на языке термодинамического формализма, упомянутого в лекции 1. Первый шаг — доказательство того, что динамика порождает на каждом  $\Gamma^{(u)}(x(0))$  некоторую инвариантную, но не нормированную меру. Механизм опять-таки прост. Берём меру Лебега на  $\Gamma^{(u)}(x(-t))$  и её образ на  $\Gamma^{(u)}(x(0))$ , задаваемый динамикой. Не очень трудно доказать, что возникающие плотности сходятся. В гладких системах с сильной гиперболичностью сходимость экспоненциальна.

Второй шаг состоит в том, чтобы показать, что среднее на любом  $\Gamma^{(u)}(x)$  по построенной мере имеет предел. Этот шаг более трудный. Обычно для него нужны специальные разбиения фазового пространства, так называемые марковские разбиения. Мы обсудим их позднее. Возникающие таким путём меры иногда называются SRB-мерами (в честь Синая, Рюэля, Боуэна).

Следующим центральным результатом теории вероятностей является предельная теорема Гаусса, описывающая поведение флуктуаций

$$\frac{1}{t} \int_0^t \varphi(x(s)) ds - \bar{\varphi},$$

$\bar{\varphi}(y) = \int \varphi(x) d\mu(x)$ . Если эта предельная теорема выполняется, то указанная разность убывает как  $O(\frac{1}{\sqrt{t}})$ . У неё нет никакого асимптотического разложения в обычном смысле, но нормированная разность

$$\sqrt{t} \left( \frac{1}{t} \int_0^t \varphi(x(s)) ds - \bar{\varphi} \right)$$

имеет универсальное гауссово распределение при  $t \rightarrow \infty$ . Универсальность в данном контексте означает, что это распределение почти не зависит от  $\varphi$  и от меры  $\mu$ . Более точно, меру  $\mu$  можно заменить на произвольную неинвариантную меру, абсолютно непрерывную по отношению к  $\mu$ .

В теории вероятностей предельная теорема Гаусса доказывается при различных предположениях о независимости или слабой зависимости случайных величин.

Оказывается, это свойство также выполняется для гиперболических динамических систем. Дадим некоторые пояснения в случае дискретного времени.

Этот подход основывается на чисто геометрическом или, более точно, чисто динамическом понятии марковского разбиения. Замкнутое подмножество  $C$  фазового пространства называется параллелограммом, если оно состоит из локальных устойчивых и локальных неустойчивых многообразий и каждое локальное устойчивое многообразие пересекает каждое локальное неустойчивое многообразие в одной точке. Набор параллелограммов  $\{C_1, C_2, \dots, C_r\}$  называется разбиением, если  $\bigcup_{i=1}^r C_i$  — всё фазовое пространство и  $C_i \cap C_j = \partial C_i \cap \partial C_j$ . Разбиение  $\{C_1, C_2, \dots, C_r\}$  называется марковским, если для каждого локального неустойчивого  $\gamma^{(u)} \subset C_i$  любое пересечение  $T\gamma^{(u)} \cap C_j$  есть локальное неустойчивое многообразие в  $C_j$  и, аналогично, для каждого локального устойчивого

$\gamma^{(s)} \in C_i$  любое пересечение  $T^{-1}\gamma^{(s)} \cap C_j$  — локальное устойчивое многообразие в  $C_j$ .

Обычно нетрудно доказать, что гиперболическая система имеет марковское разбиение. Одно из главных достоинств марковских разбиений заключается в том, что порождаемые ими так называемые символические представления динамики оказываются довольно простыми. Для марковских разбиений бесконечное пересечение  $\bigcap_{-\infty}^{+\infty} T^k C_{i_k}$  непусто, если

$$C_{i_k} \cap T C_{i_{k+1}} \neq \emptyset \quad \text{при} \quad -\infty < k < +\infty$$

Марковские разбиения позволяют представить фазовое пространство как пространство реализаций стационарного случайного процесса из теории вероятностей и применить вероятностные методы. Во многих случаях SRB-меры можно представить как гиббсовские распределения в смысле лекции 1.

Вот пример результата, который можно доказать таким методом. Предположим, мы имеем периодическую конфигурацию окружностей на плоскости. Выберем случайное начальное условие  $(q, v)$ , где  $q = (q_1, q_2)$ ,  $v = (v_1, v_2)$ ,  $v_1^2 + v_2^2 = 1$ , и рассмотрим динамику бильярдной частицы с начальным условием  $(q, v)$ . Оказывается (Бунимович, Синай), что распределение нормированного смещения

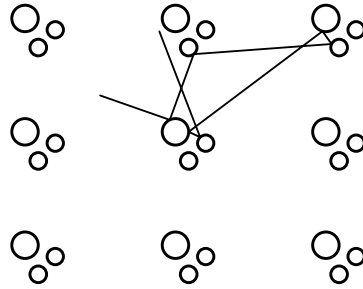


Рис. 6.

$q(t)/\sqrt{t}$  сходится к мере Винера, т.е. детерминированная динамика порождает диффузионный процесс. Обычная картина броуновского движения — это перемещение массивной частицы под воздействием многочисленных независимых столкновений с лёгкими частицами окружающего газа. В нашем случае лёгкая частица движется под воздействием столкновений с массивными неподвижными частицами и диффузия возникает из-за неустойчивости динамики. Это, возможно, перестает быть верным при замене окружностей на квадраты.

Многих интересует, в чём состоят основные достижения теории хаоса. Вот несколько ответов.

1. Результаты, показывающие, каким образом обратимая детерминированная динамика может порождать необратимость. Пример был приведен выше.
2. Анализ простых систем со странными аттракторами, наподобие системы Лоренца. Оказывается, сложная динамика может возникать при простых уравнениях движения.

3. Универсальность Фейгенбаума при возникновении хаоса и последовательности бифуркаций удвоения периода. Это замечательное открытие теории хаоса, которое наблюдалось во многих экспериментах. В математике благодаря ему возник так называемый метод ренорм-группы в теории динамических систем.

4. В приложениях теории динамических систем используются такие понятия теории хаоса, как топологическая и метрическая энтропия, ляпуновский спектр и т. п. Эти параметры можно оценить численно с довольно большой точностью, из чего часто удается заключить, что на протяжении больших, но конечных интервалов времени система ведёт себя как гиперболическая система с хаотическими свойствами.

5. Я не думаю, что теория хаоса дала что-либо интересное для гидродинамики, и, в частности, для турбулентности. Вероятно, мы до сих пор не имеем достаточно хорошего понимания динамики в этих случаях.

Этот текст следует более или менее близко Хедриковским лекциям, прочитанным в Лос-Анжелесе в августе 2000 года. Многие важные темы теории хаоса в нём пропущены. Прежде всего я имею в виду замечательные работы Д. Орнштейна и его коллег по проблеме изоморфизма динамических систем и сложности динамики, работы Дж. Мезера, Г. Ксия о неустойчивости и диффузии Арнольда в гамильтоновых системах и т. д. Все эти темы требуют отдельного изложения.

# Машины, логика и квантовая физика

Д. Дойч      А. Экерт      Р. Лупачини

Хотя истины логики и чистой математики объективны и не зависят от известных нам фактов или законов природы, наше *знание* этих истин существенно зависит от знания физических законов. Подтверждением тому служит недавний прогресс в квантовой теории вычислений, вынуждающий отказаться от классической точки зрения на вычисление (а, следовательно, и на математическое доказательство) как на чисто логическое понятие, не зависящее от физической природы вычисления. Отныне доказательство должно рассматриваться не как абстрактный объект или процесс, а как физический процесс (разновидность вычисления), результаты и достоверность которого зависят от нашего знания физики используемого вычислительного устройства.

## 1. МАТЕМАТИКА И ФИЗИЧЕСКИЙ МИР

Подлинно научное<sup>1)</sup> знание не имеет априорного обоснования. Оно должно формулироваться как гипотеза, проверяемая впоследствии экспериментом. Поэтому знания нужно выражать на языке, пригодном для высказывания точных, эмпирически проверяемых утверждений. Таким языком является математика.

Само признание возможности получения научных (в указанном выше смысле) знаний уже приводит к определённым представлениям о физическом мире. Как говорил Галилей: «Вселенная написана на языке математики» [5]. Введение Галилеем в физику математически формулируемых, эмпирически проверяемых теорий означало переход от аристотелевского понимания физики, основанного на априорных принципах, к современному статусу естественной науки. Вместо поисков непогрешимого универсального математического плана мироздания галилеева наука использует математику для выражения количественных предсказаний, относящихся к объективной физической реальности.

Итак, математика — язык описания наших знаний о физическом мире. Язык этот не только необычайно выразителен и точен, но и эффективен в практических приложениях. Юджин Вигнер обращал внимание на «непостижимую эффективность математики в физических науках» [12]. Но так ли уж непостижима или сверхъестественна эта эффективность?

Посмотрим, как мы изучаем математику. Имеем ли мы, точнее, наш мозг, прямой доступ к миру абстрактных понятий и отношений между ними? (В это верил Платон, а сейчас такую точку зрения защищает Роджер Пенроуз [8].) Или

<http://xxx.lanl.gov/math.H0/9911150> Перевод с незначительными сокращениями. Переводчик М. Н. Вялый.

<sup>1)</sup> По-русски обычно говорят — естественно-научное. *Прим. пер.*

мы изучаем математику опытным путём, взаимодействуя с физическими объектами? Мы склоняемся ко второму варианту. Это не означает, что математические сущности являются в каком-либо смысле частью физического мира или производны от него. Мы не отрицаем особую, не зависящую от предписаний законов природы, реальность чисел, множеств, групп и алгебр. Их свойства вполне объективны, как и полагал Платон. Но даны нам эти структуры только через физический мир. Лишь физические объекты, будь то компьютер или человеческий мозг, дают нам возможность заглянуть в абстрактный мир математики. Но как?

математики известно, что простые физические системы (пальцы, счётные палочки или счёты) можно использовать для представления некоторых математических сущностей и действий. Исторически именно арифметические действия удалось первыми поручить машинам. Как только стало ясно, что сложение и умножение можно разбить на последовательность базисных процедур, реализуемых физическими операциями, механические устройства, изобретённые Блезом Паскалем, Готфридом Вильгельмом Лейбницем и другими, начали освобождать людей от утомительных работ вроде перемножения двух больших чисел [6]. В XX веке следом за арифметикой удалось механизировать и логическое понятие вычислимости. Чтобы формализовать понятие «эффективности», подразумеваемое в интуитивном представлении о вычислимости, были изобретены машины Тьюринга. Алан Тьюринг предположил, что те абстрактные машины, с помощью которых он определил вычисление, способны выполнить любую *конечную эффективную* процедуру (алгоритм). Стоит отметить, что машины Тьюринга задумывались так, чтобы воспроизвести любое точно определённое действие, на которое способен *человек-вычислитель*, следующий предписанным инструкциям. Метод Тьюринга состоял в том, чтобы думать в терминах физических действий и представлять каждое действие, выполняемое вычислителем как «некоторое изменение физической системы, состоящей из вычислителя и его ленты» [11]. Поскольку результат работы не зависит от способа построения «машины для выполнения работы этого вычислителя», эффективность человека-вычислителя может быть имитирована логической машиной.

Машина Тьюринга была абстрактным понятием, но благодаря последующему прогрессу алгоритмы сейчас исполняются реальными вычислительными устройствами. Возникает естественный вопрос: какие логические процедуры может исполнить физическое устройство? Теория машин Тьюринга принципиально не может ответить на этот вопрос, равно как и любой подход, основанный на формализации традиционных представлений об эффективных процедурах. Вместо этого нужно обобщить идею Тьюринга о *механизации* процедур, в частности, процедур, связанных с понятием выводимости. Это позволило бы определить математическое доказательство как нечто механически воспроизводимое и за счёт этого эффективно проверяемое. Универсальность и достоверность логических процедур гарантируются при таком подходе наличием механических процедур, эффективно выполняющих логические действия. И не более того. Но что значит включить реальные физические машины в определение логического понятия? И что можно из этого заключить, в противоположность Вигнеру, об «эффективности физики в математических науках»?



Абстрактные модели машин, используемые в классической теории вычислений, являются чисто математическими понятиями, которым можно приписать любые не противоречащие друг другу свойства. Изучение реальных вычислительных устройств как физических объектов должно учитывать их физические свойства и потому опираться на законы природы. Машины Тьюринга (с произвольно длинными лентами) можно построить, но никто этого не делал, разве что шутки ради. Они были бы очень медленными и громоздкими. Имеющиеся сейчас компьютеры куда лучше. Но почему мы уверены, что компьютер порождает тот же выход, что и соответствующая абстрактная машина Тьюринга? Или даже, почему, покрутив ручку арифмометра, мы увидим в итоге правильный ответ? Ведь никто не проверяет машину, следуя за всеми возможными логическими шагами, или делая все вычисления, которые она может выполнить. Прежде всего, если бы у кого-то была возможность и желание делать такие проверки, не было бы нужды в изготовлении компьютеров. Причина, по которой мы доверяем машине, не может основываться целиком на логике, она также включает в себя наше знание природы этой машины. Мы опираемся на законы физики, управляющие вычислением, т. е. физическим процессом, который переводит машину из начального состояния (вход) в конечное состояние (выход). Более того, наш анализ опирается на физические теории, которые подтверждаются или опровергаются экспериментально. С этой точки зрения Тьюринг на самом деле предположил, что возможно сконструировать универсальный компьютер: машину, которую можно запрограммировать на любое вычисление, доступное какому-либо физическому объекту. Другими словами, возможно построение одного физического объекта, способного, при надлежащем обслуживании, источнике энергии и расширении памяти при необходимости, точно имитировать поведение и реакции любого другого физически возможного объекта или процесса. Гипотеза Тьюринга в этой форме (в аналогичном контексте Дойч назвал её принципом Чёрча – Тьюринга [3]) может рассматриваться как утверждение о физическом мире.

Есть очевидные и логические, и физические ограничения на возможные вычисления. Например, логика говорит нам, что никакая машина не найдёт более одного чётного простого числа, в то время как физика утверждает, что никакая машина не может нарушить законы термодинамики. Логические и физические ограничения по сути связаны между собой, как показывает «проблема остановки». Из логического анализа следует несуществование алгоритма, который определяет, остановится ли данная машина, начинающая работу из данного состояния. С физической точки зрения это означает, что некоторые машины нельзя создать. Сформулированная таким образом, проблема остановки становится утверждением о физической реальности.

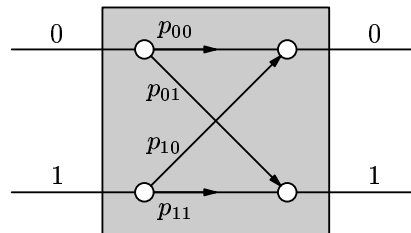
Так откуда же берётся эффективность математики? Это не просто чудо, «восхитительный дар, который мы не понимаем, и которого мы не заслуживаем» [12]. По крайней мере, это не большее чудо, чем способность к постижению эмпирического знания. Наше знание математики и логики неразрывно сплетено со знанием о физической реальности: приемлемость математического доказательства зависит от того, как мы представляем правила, которым подчиняются некоторые физические объекты, такие как компьютеры или человеческий мозг. Следовательно, прогресс в понимании физической реальности может дать

средства для улучшения понимания логики, математики и формальных понятий. Так что мы вынуждены признать зависимость математического *знания* от физики (но, подчеркнём ещё раз, не математических истин самих по себе). А раз так, пора отказываться от классической точки зрения на вычисление как на чисто логическое понятие, не зависящее от физической природы вычислителя. Ниже мы покажем как, в частности, квантовая механика меняет понимание природы вычисления.

## 2. КВАНТОВАЯ ИНТЕРФЕРЕНЦИЯ

Чтобы объяснить разницу между квантовыми компьютерами и их классическими аналогами, начнём с явления квантовой интерференции. Рассмотрим вычислительное устройство (машину), чьим входом может быть одно из двух состояний, помеченных 0 и 1.

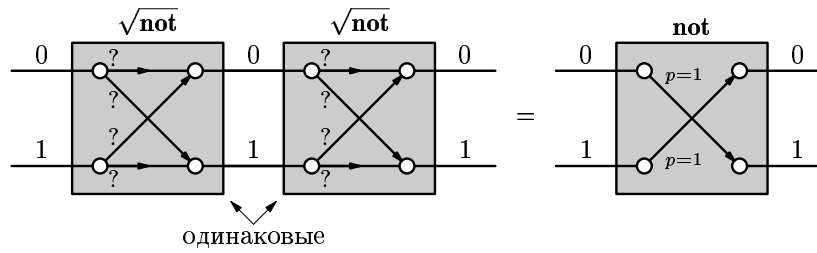
Эта машина преобразует вход  $a$  (0 или 1) в выход  $b$  (0 или 1) с вероятностью  $p_{ab}$ . Кажется очевидным, что рис. 1 (вероятности  $p_{ab}$  должны удовлетворять стандартному условию  $\sum_b p_{ab} = 1$ ) описывает все устройства, отображающие  $\{0, 1\}$  в себя. (Действие машины не зависит от иной информации, подаваемой на вход или хранящейся в памяти.) Есть два предельных случая, когда



**Рис. 1.** Схема наиболее общего устройства, отображающего  $\{0, 1\}$  в себя

поведение устройства детерминировано: при  $p_{01} = p_{10} = 0$ ,  $p_{00} = p_{11} = 1$  (тождественное преобразование) и при  $p_{01} = p_{10} = 1$ ,  $p_{00} = p_{11} = 0$  (отрицание **not**). В остальных случаях имеем устройство со случайным поведением. Пусть, к примеру,  $p_{01} = p_{10} = p_{00} = p_{11} = 0,5$ . Вновь кажется очевидным, что единственный вид такой машины, это случайный переключатель, равновероятно преобразующий каждый вход в один из двух возможных выходов. Однако это не так. Мы имеем в виду машину, удовлетворяющую тому же свойству, но такую, что при последовательном соединении двух одинаковых экземпляров её на выходе всегда получается отрицание входа (рис. 2).

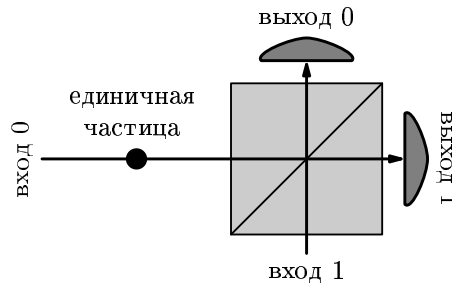
Это утверждение радикально противоречит интуиции. Каждая машина по отдельности выдаёт 0 или 1 равновероятно и независимо от входа, но две последовательно соединённые машины реализуют логическую операцию **not**. По этой причине назовём такую машину  $\sqrt{\text{not}}$ . В логике такой операции нет, так что разумно предположить, что машина  $\sqrt{\text{not}}$  не существует. Но на самом деле такая машина в природе есть! Для физика, изучающего одночастичную интерференцию, построение такой машины — обычное дело. Простейший вариант —



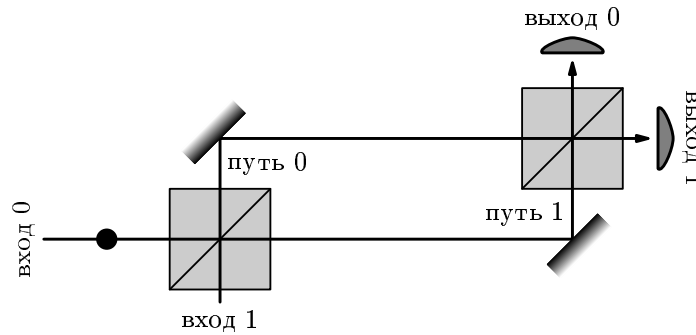
**Рис. 2.** Последовательное подключение двух одинаковых машин, отображающих  $\{0, 1\}$  в себя, каждая из которых работает как случайный переключатель. При последовательном подключении случайность исчезает: суммарный эффект есть логическая операция **not**. Это очевидным образом противоречит аксиоме аддитивности теории вероятностей!

полупрозрачное зеркало, т. е. зеркало, с вероятностью 50% отражающее падающий на него фотон, и с вероятностью 50% пропускающее его. Последовательное соединение двух этих машин реализуется двумя полупрозрачными зеркалами, при этом фотон означает 0, если он находится на одном из двух возможных путей, и 1 в противном случае.

Читатель может поинтересоваться, что делать с аксиомой аддитивности теории вероятностей, утверждающей про пару несовместных событий  $E_1$  и  $E_2$ , что вероятность события  $E_1$  **or**  $E_2$  есть сумма вероятностей событий  $E_1, E_2$ . Ведь переход  $0 \rightarrow 0$  в составной машине может происходить двумя взаимоисключающими способами: либо  $0 \rightarrow 0 \rightarrow 0$ , либо  $0 \rightarrow 1 \rightarrow 0$ . Их вероятности  $p_{00}p_{00}$  и  $p_{01}p_{10}$  соответственно, значит, сумма  $p_{00}p_{00} + p_{01}p_{10}$  равна вероятности перехода  $0 \rightarrow 0$  в новой машине. Она отлична от нуля, если  $p_{00}$  или  $p_{01}p_{10}$  отлична от нуля.



**Рис. 3.** Экспериментальная реализация элемента  $\sqrt{\text{not}}$ . Полупрозрачное зеркало отражает половину падающего на него света. Но единичный фотон не расщепляется: когда мы посылаем фотон сквозь такое зеркало, регистрируются выход 0 или выход 1 с равными вероятностями. Не нужно думать, однако, что фотон проходит зеркало по одному из двух случайно выбранных путей, он идёт по обоим путям! В этом можно убедиться, соединив последовательно два полупрозрачных зеркала, как показано на следующем рисунке



**Рис. 4.** Одночастичная интерференция — экспериментальная реализация двух соединённых последовательно элементов  $\sqrt{\text{not}}$ . Фотон, который входит в интерферометр через вход 0, всегда попадает в детектор у выхода 1 и никогда в детектор у выхода 0

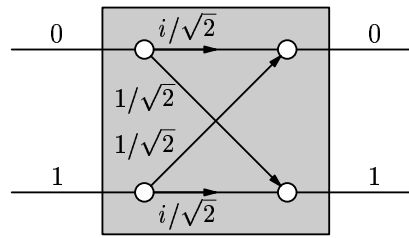
И тем не менее мы построили машину, в которой  $p_{00}$  и  $p_{01}p_{10}$  не равны нулю, а вероятность перехода  $0 \rightarrow 0$  в составной машине равна нулю. Что же неверно в приведённом рассуждении?

Неверно то, что процессы  $0 \rightarrow 0 \rightarrow 0$  и  $0 \rightarrow 1 \rightarrow 0$  взаимно исключают друг друга. Любое объяснение, которое предполагает, что фотон избирает один из двух путей по интерферометру, приводит к выводу, что каждый из детекторов должен срабатывать в среднем в половине случаев. Но эксперимент показывает обратное! Так что на самом деле происходят оба перехода одновременно. Это нельзя получить ни из теории вероятностей, ни из любой другой математической конструкции. Такое знание можно получить из наилучшей имеющейся в настоящий момент физической теории, а именно, квантовой механики. Квантовая теория объясняет поведение  $\sqrt{\text{not}}$  и правильно предсказывает вероятности для всех возможных выходов при любых способах соединения таких машин. Это знание возникло как результат гипотез, экспериментов и опровержений. Следовательно, на основании физических экспериментов, подтверждающих эту теорию, логики вправе ввести новую логическую операцию  $\sqrt{\text{not}}$ , поскольку физическая модель для неё существует в природе!

Рассмотрим теперь математический аппарат квантовой механики, с помощью которого описываются квантовые вычислительные устройства: от простейших, таких как  $\sqrt{\text{not}}$ , до самых сложных — квантового обобщения универсальной машины Тьюринга.

Квантовая механика вводит понятие *амплитуд вероятности* — комплексных чисел  $c$ , квадраты модуля которых  $|c|^2$  при некоторых обстоятельствах можно понимать как вероятности. Когда некоторый переход, наподобие рассмотренного выше, может происходить несколькими способами, его амплитуда вероятности есть сумма амплитуд вероятности по всем возможным способам, рассматриваемым порознь.

Вероятности переходов в машине  $\sqrt{\text{not}}$  изображены на рис. 5. Эта машина сохраняет значение бита с амплитудой вероятности  $c_{00} = c_{11} = i/\sqrt{2}$ , меняет



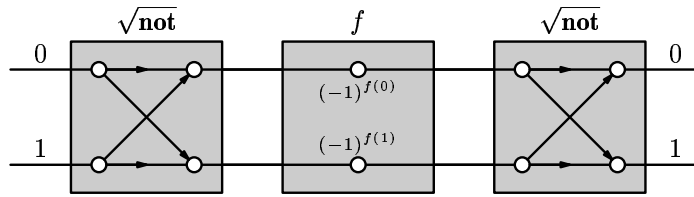
**Рис. 5.** Переходы в квантовых машинах описываются не вероятностями, а амплитудами вероятности

на противоположное — с амплитудой  $c_{01} = c_{10} = 1/\sqrt{2}$ . Соответствующие вероятности получаются возведением в квадрат модулей амплитуд вероятности, в каждом из случаев получаем  $1/2$ . Этим объясняется поведение  $\sqrt{\text{not}}$  на рис. 1. При последовательном соединении машин, как на рис. 2, для подсчёта вероятности выхода 0 при входе 0 нужно сложить амплитуды вероятности по всем путям, ведущим от входа 0 к выходу 0. Таких две:  $c_{00}c_{00} = -1/2$  и  $c_{01}c_{10} = 1/2$ . Их сумма равна 0, так что и вероятность выхода 0 равна 0. В отличие от вероятностей, амплитуды могут взаимно сокращаться!

### 3. КВАНТОВЫЕ АЛГОРИТМЫ

Сложение амплитуд вероятности вместо вероятностей является одним из основных правил квантовой механики и применимо ко всем физическим объектам, в том числе, и к квантовым вычислительным устройствам. Если вычисляющая машина начинает работать в заданной начальной конфигурации (вход), то вероятность того, что она, проходя через ряд промежуточных конфигураций, закончит работу в заданной конечной конфигурации (выход), равна квадрату модуля суммы амплитуд вероятности по всем вычислительным путям, соединяющим вход с выходом. Амплитуды — это комплексные числа. При сложении они могут взаимно сокращаться (*деструктивная интерференция*) или увеличиваться по абсолютной величине (*конструктивная интерференция*). Основная идея квантового вычисления состоит в том, чтобы усилить правильные ответы и подавить неправильные. Мы покажем это на примере одного из вариантов первого квантового алгоритма, предложенного Д. Дойчем в 1985 г.

Рассмотрим булевы функции  $f$ , отображающие  $\{0, 1\}$  в себя. Есть ровно четыре таких функции: две постоянных ( $f(0) = f(1) = 0$  и  $f(0) = f(1) = 1$ ) и две биекции ( $f(0) = 0, f(1) = 1$  и  $f(0) = 1, f(1) = 0$ ). Предположим, что можно вычислить значение функции *только один раз* (потому, например, что алгоритм вычисления функции очень длинный, или потому, что к таблице её значений позволено обратиться только один раз). Нужно определить, является ли  $f$  биекцией. Заметим, что конкретные значения  $f$  нас не интересуют, мы проверяем некоторое глобальное свойство функции  $f$ . Наша классическая интуиция подсказывает, а классическая теория вычислений подтверждает, что для проверки этого свойства нужно вычислить оба значения  $f(0)$  и  $f(1)$ , т.е. вычислять  $f$  дважды. Однако это неверно в реальном физическом мире, в котором можно



**Рис. 6.** Схема квантовой машины, решающей задачу Дойча за одно вычисление функции

осуществить квантовое вычисление, решающее задачу Дойча и вычисляющее значение  $f$  только раз. Машина, решающая эту задачу, использует квантовую интерференцию и состоит из двух  $\sqrt{\text{not}}$ , между которыми вставлено вычисление значения функции (рис. 6).

Нам нет нужды вдаваться в подробности реализации  $f(x)$ . Для нашего примера достаточно указать, что есть два пути, индексированные 0 и 1, а действие машины, реализующей  $f(x)$ , заключается в том, что амплитуда вероятности на пути  $x$  умножается на фазовый множитель  $\exp(\pi i f(x))$ , т. е. на  $(-1)^{f(0)}$  (на пути 0) и на  $(-1)^{f(1)}$  (на пути 1). Теперь подсчитаем амплитуду вероятности выхода 0 при входе 0. Амплитуды вероятности на двух различных вычислительных путях равны  $i/\sqrt{2} \times (-1)^{f(0)} \times i/\sqrt{2} = -1/2 \times (-1)^{f(0)}$  и  $1/\sqrt{2} \times (-1)^{f(1)} \times 1/\sqrt{2} = 1/2 \times (-1)^{f(1)}$ . Их сумма

$$\frac{1}{2} \left( (-1)^{f(1)} - (-1)^{f(0)} \right), \quad (1)$$

равна 0, если  $f$  — постоянная, и  $\pm 1$ , если  $f$  — биекция. Итак, вероятность (квадрат модуля амплитуды) выхода 0 на входе 0 равна 0 для постоянных функций и 1 для биекций.

С этого результата Дойча началась новая область исследований: квантовые вычисления. Последовательная работа по улучшению квантовых алгоритмов привела в 1994 г. к открытию Питером Шором квантового алгоритма, способного эффективно факторизовать числа [10]. А поскольку трудность факторизации лежит в основе надёжности многих известных криптографических систем, включая наиболее популярную криптосистему с открытым ключом RSA [9]<sup>2)</sup>, алгоритм Шора быстро стал популярен как первое “killer application” для квантового вычисления — нечто весьма полезное, что может делать только квантовый компьютер.

Мало кто из специалистов сомневается в том, что факторизация не принадлежит классу **ВРР** (**ВРР** означает «вероятностное вычисление за полиномиальное время с ограниченной вероятностью ошибки»). Интересно, однако, что это не доказано. В теории вычислительной сложности обычно трактуют принадлежность задачи классу **ВРР** как указание на её «несложность» или «практическую

<sup>2)</sup> В декабре 1997 г. британское правительство официально признало, что эта криптосистема с открытым ключом была изначально изобретена в Центре Правительственных Коммуникаций (GCHQ) в Челтенхэме. В 1975 году Джеймс Эллис, Клиффорд Кокс и Малькольм Вильямсон из GCHQ открыли то, что позднее было переполнено в академической науке и стало известно как RSA и обмен ключами Диффи – Хэллмана. — Прим. авт.

разрешимость», а задачи, не принадлежащие **ВРР**, напротив, понимаются как «трудные» или «нерешаемые на реальных компьютерах» (см., например, [7]<sup>3)</sup>). Алгоритм из класса **ВРР** — это такой эффективный (= быстрый) алгоритм, который на любом входе даёт правильный ответ с вероятностью, превышающей некоторый порог  $\delta > 1/2$ . Мы, вообще-то, не можем проверить правильность ответа. Но можно повторить все вычисление  $k$  раз и взять тот ответ, который встретился чаще всего. Увеличивая  $k$ , можно добиться вероятности правильного ответа, сколь угодно близкой к 1.

Результат Шора показывает, что в реальности, описываемой квантовой физикой, факторизация не является трудной задачей.

Еще Ричард Фейнман, выступая на первом конгрессе по физике вычислений в МИТ (1981 год), заметил, что эволюцию общей квантовой системы скорее всего нельзя эффективно моделировать на классическом вероятностном компьютере [4]. Количество классической информации, требуемой для описания квантового состояния эволюционирующей системы, растёт экспоненциально по времени, так что лобовое классическое моделирование эволюции происходит с экспоненциальным замедлением. Фейнман же предложил рассматривать этот факт не как препятствие, а как возможность. В самом деле, раз результат эксперимента по многочастичной интерференции описывается сложными и длинными вычислениями, то сам эксперимент и измерение выхода в нем равносильно выполнению некоторого сложного вычисления. Далее Фейнман предположил, что возможно эффективное моделирование квантовой эволюции, если само моделирующее устройство подчиняется законам квантовой механики. Более того, Фейнман предположил, что существует универсальное квантовое устройство, пригодное для моделирования квантовой эволюции произвольной системы. В 1985 году Дойч показал, что такое устройство (универсальный квантовый компьютер) действительно существует [3]. Тогда же было показано, что время и другие требуемые ресурсы не растут экспоненциально в зависимости от размера или степени точности описания моделируемой физической системы, так что такое моделирование является «эффективным» по стандартам теории сложности вычислений [1].

Это и есть иллюстрация нашей основной мысли: чем больше мы знаем о физике, тем больше можем узнать о вычислениях и математике. Только используя квантовую механику, удалось придумать, как эффективно решать задачу факторизации чисел; другого способа пока не найдено, а, возможно, его и вовсе нет.

#### 4. ДЕТЕРМИНИРОВАННЫЕ, ВЕРОЯТНОСТНЫЕ И КВАНТОВЫЕ КОМПЬЮТЕРЫ

Любой квантовый компьютер, включая и универсальный, можно описать так же, как мы описывали выше простые примеры машин, меняя вероятности на амплитуды вероятности. Начнем с классической машины Тьюринга. Она задается

---

<sup>3)</sup> Или предыдущий номер «Математического просвещения». Прим. пер.

конечным набором пятерок вида

$$(q, s, q', s', d), \quad (2)$$

в которых первые два символа описывают условие на применимость данного шага вычисления, а последние три — результат применения данной команды ( $q$  — текущее состояние,  $s$  — прочитанный в данный момент символ,  $q'$  — состояние, в которое переходит машина,  $s'$  — символ, которым заменяется  $s$ ,  $d$  указывает сдвиг головки относительно ленты (на один символ влево или вправо, или остаться на месте)). На этом языке вычисление описывается так. Машине на вход подается конечное слово в алфавите  $\Sigma$ , записанное в ячейках ленты, затем машина начинает работу из начального состояния  $q_0$  (читающая головка стоит над самым левым символом входа, выполняет последовательно элементарные действия, описанные выше, и останавливается при достижении финального состояния  $q_h$ . (При некоторых входах вычисление может продолжаться бесконечно.) Результат вычисления определяется как содержимое некоторой указанной части ленты после достижения финального состояния (если этого не случится, результат вычисления не определен)).

При вычислении возникает последовательность конфигураций машины (слово, записанное на ленте, состояние головки и ее положение на ленте). Например, начальная конфигурация описывается входным словом, состоянием  $q_0$  и положением головки над самым левым символом входа. Конфигураций бесконечно много, однако в успешном вычислении машина проходит лишь через конечное количество конфигураций. Переходы между конфигурациями полностью описываются пятерками (2).



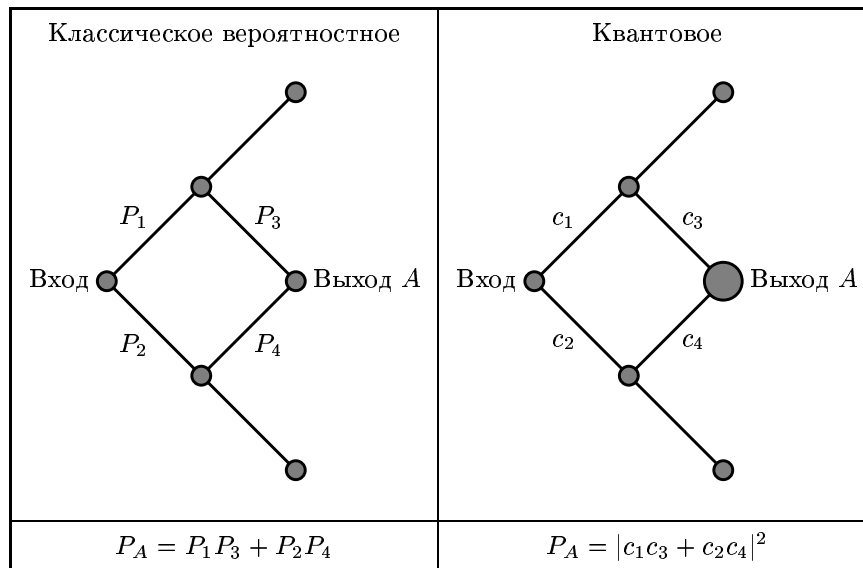
Рис. 7. Трехшаговое детерминированное вычисление

Вычисления, вообще говоря, не обязаны быть детерминированными. Мы можем расширить возможности машины Тьюринга, добавляя «подбрасывание монеты», что позволяет выбирать ей следующий шаг случайно. Такое вероятностное вычисление можно представлять ациклическим ориентированным графом с корнем, каждая вершина которого соответствует некоторой конфигурации машины, а ребро — одному шагу вычисления. Вычисление начинается с корня, представляющего начальную конфигурацию, каждая его ветвь ведет в вершины, отвечающие конфигурациям, достижимым из начальной с ненулевой вероятностью. Работа машины полностью задается конечным списком вида

$$\delta: Q \times \Sigma \times Q \times \Sigma \times \{\text{влево, вправо, на месте}\} \mapsto [0, 1], \quad (3)$$

где  $\delta(q, s, q', s', d)$  задает вероятность того, что будет выполнено действие, описываемое пятеркой  $(q, s, q', s', d)$ . Это описание должно согласовываться с законами теории вероятностей. Каждому ребру дерева вычислений можно сопоставить вероятность перехода по этому ребру, и нужно потребовать, чтобы сумма вероятностей на ребрах, выходящих из данной вершины, равнялась 1. Вероятность пути из корня в данную вершину есть произведение вероятностей ребер,





**Рис. 8.** В вероятностной машине (слева) вероятность выхода  $A$  есть сумма вероятностей вычислений, ведущих в  $A$ . В квантовой машине (справа) вероятность выхода  $A$  получается сложением амплитуд вероятности и взятием квадрата модуля полученной суммы. Так что в квантовом случае возможны как конструктивная (вероятности усиливаются), так и деструктивная (вероятности гасятся) интерференция

входящих в этот путь, а вероятность перехода в заданную конфигурацию после  $n$  шагов равна сумме вероятностей путей длины  $n$ , соединяющих начальную и заданную конфигурации.

Есть вероятностные алгоритмы, которые решают некоторые задачи (со сколь угодно близкой к 1 вероятностью успеха) гораздо быстрее, чем все известные детерминированные алгоритмы.

Описанная выше классическая модель подсказывает естественное квантовое обобщение. Квантовое вычисление можно представлять ориентированным ациклическим графом, как и классическое. Каждому ребру этого графа мы сопоставим амплитуду вероятности того, что вычисление пойдет по этому ребру. Как и выше, амплитуда вероятности пути есть произведение амплитуд вероятности по ребрам этого пути, а амплитуда вероятности заданной конфигурации — сумма амплитуд вероятности по всем путям, ведущим из начальной конфигурации в заданную. Если, в частности, финальная конфигурация достижима ровно по двум путям с амплитудами  $c$  и  $-c$ , вероятность попадания в эту конфигурацию  $|c - c|^2 = 0$ , несмотря на то, что вероятность того, что вычисление пойдет по любому из этих двух путей, равна  $|c|^2$ . Единственный квантовый компьютер может следовать по многим вычислительным путям одновременно и его выход определяется интерференцией всех этих путей. Напротив, классическая вероятностная машина Тьюринга следует по *единственному* (хоть и случайно

выбранному) пути вычисления. Работа квантовой машины полностью задается конечным списком вида

$$\delta: Q \times \Sigma \times Q \times \Sigma \times \{\text{влево, вправо, на месте}\} \mapsto \mathbb{C}, \quad (4)$$

где  $\delta(q, s, q', s', d)$  задает амплитуду вероятности выполнения действия, описываемого пятеркой  $(q, s, q', s', d)$ .

## 5. ДАЛЬНЕЙШИЕ СЛЕДСТВИЯ

Когда в начале 60-х годов начинали изучать физику вычислений, одним из основных стимулов было опасение, что квантовомеханические эффекты могут давать фундаментальные (=неустранимые) границы на точность, с которой физические объекты могут воспроизводить свойства абстрактных сущностей, таких как логические переменные и операции в теории вычислений. То есть боялись, что мощь и элегантность теории вычислений, такие ее понятия, как вычислительная универсальность, такие ее фундаментальные принципы, как тезис Чёрча – Тьюринга, равно как и более мощные результаты в современной теории сложности, окажутся не более чем плодами математического воображения, не имеющими отношения ни к чему в природе.

Эти опасения оказались беспочвенными. Квантовая механика не только не накладывает ограничений на возможности машин Тьюринга, но и дает новые возможности вычислений, обсуждавшиеся выше. Сохранилась элегантность теории, более того, оказалось, что квантовая теория вычислений весьма естественно сочетается с фундаментальными теориями в других областях, чего было трудно даже ожидать от классического приближения. Само слово «квант» означает то же самое, что и «бит» — элементарный кусочек — это замечательным образом согласуется с тем фактом, что классические физические системы, будучи подвержены неустойчивости общего вида, именуемой «хаос», не могут в принципе поддерживать цифровое вычисление (так что даже машины Тьюринга, теоретический образец всех классических компьютеров, всегда имели квантовомеханический характер, о чем умалчивалось). Тезис Чёрча – Тьюринга в классической теории (об эквивалентности всех «естественных» моделей вычисления) никогда не был доказан. Его аналог в квантовой теории вычислений (принцип Чёрча – Тьюринга, утверждающий, что универсальный квантовый компьютер способен моделировать поведение любой конечной физической системы) был доказан прямо из определения в работе Дойча 1985 года [3]. Более сильный результат (также предполагавшийся, но никогда не доказанный в классической теории), а именно, что такое моделирование возможно за время, зависящее полиномиальным образом от времени эволюции моделируемой системы, также был доказан в квантовом случае [1].

Наряду с прочими приложениями, квантовые вычисления существенно повлияли (по крайней мере, в принципе) на понятие математического доказательства. Выполнение вычисления, которое дает некоторый определенный результат, равносильно доказательству того, что наблюдаемый результат возможен при данном вычислении. Мы можем описать вычислительные операции математически, поэтому всегда можем перевести такое вычисление в доказательство некоторой теоремы. Так же обстоит дело и в классическом случае, но при отсутствии

интерференции возможно сделать запись всех шагов вычисления, порождая таким образом доказательство, удовлетворяющее классическому определению — «цепочка утверждений, каждое из которых является аксиомой или следует из предыдущих утверждений в цепочке по правилам вывода» (правильность такого доказательства также можно проверить пошагово). Теперь мы вынуждены отказаться от такого определения. Отныне доказательство должно рассматривать как процесс — вычисления самого по себе. Мы должны признать, что в будущем квантовые компьютеры будут доказывать теоремы методами, которые нельзя будет проверять пошагово, так как распечатка «цепочки утверждений», соответствующих этому доказательству, не поместится в наблюдаемой части Вселенной.

## 6. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

В этом коротком обсуждении мы прошли по самым верхам быстро развивающейся теории квантовых вычислений. Основное внимание было уделено принципиальным вопросам, оставляя в стороне физические подробности и нюансы технологической реализации. Стоит, однако, сказать, что квантовое вычисление — серьезная перспектива для будущих поколений вычислительных устройств. Это одна из причин, по которой квантовые вычисления привлекают все большее внимание как академических исследователей, так и промышленных кругов по всему миру. В настоящее время неясно когда, каким образом, да и будут ли вообще, реализованы полномасштабные квантовые компьютеры; но несмотря на это, уже сейчас квантовая теория вычислений играет гораздо более важную роль в общем понимании мира, чем ее классическая предшественница. Мы считаем, что всякий, кто добивается фундаментального понимания физики, вычислений или логики, должен учитывать достижения этой теории.

## СПИСОК ЛИТЕРАТУРЫ

- [1] *Bernstein E., Vazirani U.* Quantum complexity theory // Proc. of the 25th Ann. Symp. on the Theory of Comput. New York: ACM, 1993. P. 11–20.
- [2] *Cleve R., Ekert A., Macchiavello C., Mosca M.* Quantum Algorithms Revisited // Proc. of the Royal Soci., A, 1998. Vol. 454. P. 339–354.
- [3] *Deutsch D.* Quantum theory, the Church-Turing principle and the universal quantum computer // Proc. of the Royal Soc., A, 1985. Vol. 400. P. 97–117.
- [4] *Feynman R. P.* Simulating physics with computers // Int. J. of Theor. Physics, 1982. Vol. 21. P. 467–488.
- [5] *Galilei G.* Saggiatore, 1623. / Opere. (Favaro A. (ed.)) Vol. 6. Firenze: Edizione Nazionale, 1896.
- [6] *Goldstine H.H.* The Computer from Pascal to von Neumann. Princeton: Princeton University Press, 1972.
- [7] *Papadimitriou C. H.* Computational Complexity. Reading: Addison-Wesley, 1994.
- [8] *Penrose R.* Shadows of the mind. Oxford: Oxford University Press, 1994.

- 
- [9] *Rivest R., Shamir A., Adleman L.* On Digital Signatures and Public-Key Cryptosystems. Tech. Rep. MIT/LCS/TR-212. MIT Laboratory for Computer Science, January 1979.
- [10] *Shor P.* Algorithms for quantum computation: discrete log and factoring // Proc. of the 35th Ann. Symp. on the Foundations of Computer Science, S. Goldwasser (editor). Los Alamitos: IEEE Computer Society Press, 1994. P. 124–134.
- [11] *Turing A.* On computable numbers with an application to the Entscheidungsproblem // Proc. of the London Math. Soc., 2, 1936–37. Vol.42. P. 230–265.
- [12] *Wigner E. P.* The unreasonable effectiveness of mathematics in the natural sciences // Comm. on Pure and Appl. Math., 1960. Vol. 13. P. 1–14.

# Биллиарды и упругие столкновения частиц и шаров

Г. А. Гальперин

## I. ВВЕДЕНИЕ

Одной из важных и интересных задач физики, а точнее, кинетической теории идеальных газов, является исследование движения молекул газа в сосуде или пространстве. При этом молекулы можно представлять себе как упругие шарики одинаковой массы (если газ однородный) или различных масс (если сосуд заполнен разными видами газа); эти шарики летают с постоянными, но как правило, различными скоростями вдоль прямых, а при столкновениях мгновенно меняют свои скорости согласно законам упругого столкновения, после чего продолжают двигаться по новым прямым, пока не произойдут очередные столкновения. В результате многократных соударений молекулы «перемешаются», и если их много, то после продолжительного времени их движение будет казаться хаотическим, давление в сосуде выравняется («закон Паскаля»), в газе установится определенная «температура» (средняя кинетическая энергия молекулы), «плотность» газа будет постоянной и т. д. Все описанные (а также пропущенные здесь) характеристики идеального газа хорошо известны в физике, многие из них исследованы также чисто математически, как правило, с привлечением теории вероятностей и ее модной (с середины 50-х годов XX века и по настоящее время) ветви — эргодической теории. Грубо говоря, «математическое исследование» газа состоит в точном определении физических понятий («давление», «температура», «перемешивание» и т. д.), а затем в строгом решении математических задач, связанных с этими понятиями, в создании соответствующей математической теории, в математическом доказательстве известных формул и получении новых, и т. п. В этом направлении сейчас работают десятки, если не сотни, математиков из разных стран мира. Ключевыми словами в их исследованиях являются «хаос» (изучаемый эргодической теорией) и «биллиард» (геометрическая часть той же теории). Здесь следует заметить, что «хаос» и «биллиард» — взаимопроникающие и переплетающиеся понятия: «биллиард» изучает детерминированное поведение системы; это поведение с течением времени становится почти неотличимым от хаотического, и тогда, на последнем этапе, можно забыть про детерминированность и исследовать один только «хаос». «Хаос» очень сложная вещь, и здесь мы затронем только «детерминированную» часть поведения системы.

Цель настоящей статьи — объяснить, как задача о столкновении шариков сводится к биллиарду, т. е. к задаче о равномерном движении *одной* частицы и ее

бильярдному отражению от различных поверхностей (правда, в многомерном пространстве). После этого мы расскажем о замечательной задаче Я. Г. Синая — о числе соударений между молекулами газа и ее бильярдных модификациях, и опишем различные подходы к ее решению.

## II. НЕМНОГО ИСТОРИИ: ПОСТАНОВКА ЗАДАЧ И РЕЗУЛЬТАТЫ

### 1. ТРИ ПРОСТРАНСТВА: КОНФИГУРАЦИОННОЕ, «СКОРОСТНОЕ» И ФАЗОВОЕ

Идея заменять поведение сложной механической системы движением одной точки хотя и очень нетривиальна, но довольно стара — ее использовал в своих исследованиях великий физик Людвиг Больцман свыше ста лет назад (в 1872 г. Больцман написал свое знаменитое «уравнение Больцмана», связывающее давление газа и температуру с помощью константы Больцмана  $k_B$  (см., например, [6])).

Рассмотрим простой пример:  $n$  частиц движутся каким-то образом (пока что без всякого закона) в трехмерном пространстве  $\mathbb{R}^3$ . Каждая частица имеет три координаты  $(x_i, y_i, z_i)$ ,  $i = 1, \dots, n$ , меняющиеся во времени  $t$ . Заменяем эти  $n$  частиц на одну «многомерную» точку

$$\mathbf{x} = (x_1, y_1, z_1; x_2, y_2, z_2; \dots; x_n, y_n, z_n),$$

движущуюся в пространстве  $\mathbb{R}^{3n}$ . Движение точки  $\mathbf{x}$  состоит в том, что в каждый момент времени  $t$  известны все ее  $3n$  координат (они определяются координатами исходных частиц в момент времени  $t$ ), так что известна и вся траектория точки  $\mathbf{x} \in \mathbb{R}^{3n}$ . А если известны конкретные законы движения частиц, т. е. функции  $x_i(t)$ ,  $y_i(t)$ ,  $z_i(t)$ , то известен и закон движения «многомерной» частицы  $\mathbf{x}(t)$ . Во многих случаях следовать за точкой  $\mathbf{x}(t)$  легче, чем сразу за большим числом исходных частиц. Однако эта простота кажущаяся — уменьшая число частиц до одной, нам приходится увеличивать «высокомерие» пространства. Пространство  $\mathbb{R}^{3n}$ , в котором движется точка  $\mathbf{x}(t)$ , называется *конфигурационным*. Однако и его можно расширить, заменив конфигурационную точку  $\mathbf{x}(t)$  на *фазовую* точку  $\mathbf{z}(t)$ , координаты которой — не только положения частиц в пространстве  $\mathbb{R}^3$ , но и их скорости  $v_i = (v_{ix}, v_{iy}, v_{iz})$ ,  $i = 1, 2, \dots, n$ :

$$\mathbf{z}(t) = (\mathbf{x}(t), \mathbf{v}(t)) = (x_1, y_1, z_1; \dots; x_n, y_n, z_n; v_{1x}, v_{1y}, v_{1z}; \dots; v_{nx}, v_{ny}, v_{nz}).$$

Таким образом возникает фазовое пространство  $\mathbb{R}^{6n}$ , фазовая точка которого описывает движение (траекторию) исходной системы в  $\mathbb{R}^3$  целиком и полностью (в отличие от конфигурационной точки  $\mathbf{x}(t)$ , описывающей лишь пространственное положение частиц — без их скоростей). Фазовое пространство есть прямая сумма двух подпространств — конфигурационного  $\mathbb{R}_x^{3n}$  и пространства скоростей  $\mathbb{R}_v^{3n}$ :  $\mathbb{R}_z^{6n} = \mathbb{R}_x^{3n} \oplus \mathbb{R}_v^{3n}$ . Иногда бывает проще и нагляднее исследовать проекцию точки  $\mathbf{z}(t)$  на  $\mathbb{R}_x^{3n}$  и/или на  $\mathbb{R}_v^{3n}$  по отдельности, вместо того, чтобы описывать намного более сложное поведение фазовой точки  $\mathbf{z}(t)$  во всем  $\mathbb{R}^{6n}$ .

Следует отметить, что, как правило, движение фазовой (а также конфигурационной и «скоростной») точки осуществляется не во всем «пустом» пространстве  $\mathbb{R}_z^{6n}$  — ограничения, накладываемые на систему различными физическими законами, индуцируют геометрические препятствия (в виде поверхностей) в  $\mathbb{R}_z^{6n}$  (равно как и в  $\mathbb{R}_x^{3n}$ , и в  $\mathbb{R}_v^{3n}$ , соответственно).

Например, если исходные частицы в пространстве  $\mathbb{R}^3$  имеют одинаковую массу 1 и их движение подчиняется закону сохранения энергии:

$$v_1^2 + v_2^2 + \dots + v_n^2 = \text{const} \iff v_{1x}^2 + v_{1y}^2 + v_{1z}^2 + \dots + v_{nx}^2 + v_{ny}^2 + v_{nz}^2 = \text{const},$$

то проекция траектории точки  $z \in \mathbb{R}_z^{6n}$  на пространство  $\mathbb{R}_v^{3n}$  представляет собой кривую, лежащую на  $(3n - 1)$ -мерной сфере радиуса  $\sqrt{\text{const}}$ . Эта сфера и ограничивает движение точки  $z \in \mathbb{R}_z^{6n}$ .

Другой важный для нас пример: два одинаковых шара в пространстве  $\mathbb{R}^3$ . Каждый шар задается координатами своего центра и радиусом  $r$ . Если бы шары были прозрачными и проходили друг сквозь друга без препятствий, то их конфигурационным пространством оказалось бы все пространство  $\mathbb{R}_x^{3 \cdot 2} = \mathbb{R}_x^6$ . Однако шары непрозрачны и расстояние между их центрами всегда больше или равно  $2r$ :

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 \geq (2r)^2. \quad (1)$$

Неравенство (1) задает внешнюю часть цилиндра  $\Pi$  в 6-мерном пространстве, наклоненного под углом  $45^\circ$  к горизонтали. (Если ограничиться слева в формуле (1) только первым слагаемым и заменить знак неравенства знаком равенства, то мы получим полосу на плоскости  $Ox_1x_2$  между прямыми  $x_2 = x_1 - r$  и  $x_2 = x_1 + 2r$  — одномерный цилиндр, см. рис. 1).

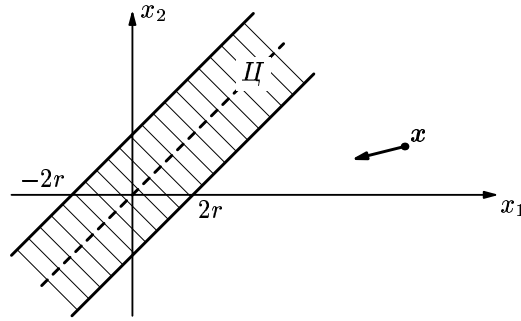


Рис. 1. Одномерный цилиндр  $\Pi$

Таким образом, конфигурационная точка  $x$  может двигаться только лишь в части пространства  $\mathbb{R}_x^6$  — вне цилиндра  $\Pi$ . Попадание точки  $x$  на цилиндр означает столкновение шаров.

В дальнейшем под конфигурационным, «скоростным» и фазовым пространствами мы будем иметь в виду только множество *возможных* положений соответствующей точки в  $\mathbb{R}_x^{3n}$ ,  $\mathbb{R}_v^{3n}$  или  $\mathbb{R}_z^{6n}$ . Описание этих пространств для специальных систем будет дано в следующем разделе.

## 2. Биллиард

Пусть точка  $x$  движется равномерно и прямолинейно в пространстве  $\mathbb{R}^n$ ; положим скорость ее движения равной 1:  $\|v\| = \left\| \frac{dx}{dt} \right\| = 1$ . Если в  $\mathbb{R}^n$  имеется плоскость  $\Pi$  размерности  $n - 1$ , то пусть точка  $x$  отражается от нее по закону

геометрической оптики: *угол отражения равен углу падения*. Или иначе: разложим скорость  $\mathbf{v}^-$  точки  $\mathbf{x}$  до отражения на две составляющие, одна из которых,  $\mathbf{v}_\perp^-$ , перпендикулярна плоскости  $\Pi$ , а вторая ей параллельна:  $\mathbf{v}_\parallel^- = \mathbf{v}^- - \mathbf{v}_\perp^-$ . При отражении от  $\Pi$  параллельная составляющая  $\mathbf{v}_\parallel^-$  не меняется, а перпендикулярная меняет знак:  $\mathbf{v}_\parallel^+ = \mathbf{v}_\parallel^-$ ,  $\mathbf{v}_\perp^+ = -\mathbf{v}_\perp^-$ . Новая скорость частицы (после отражения) равна

$$\mathbf{v}^+ = \mathbf{v}_\parallel^+ + \mathbf{v}_\perp^+ = \mathbf{v}_\parallel^- - \mathbf{v}_\perp^-.$$

Или, наконец, закон отражения может быть записан в виде векторной формулы

$$\mathbf{v}^+ = \mathbf{v}^- - 2 \cdot (\mathbf{v}^-, \mathbf{n}) \cdot \mathbf{n}, \quad (2)$$

где  $(\cdot, \cdot)$  означает скалярное произведение векторов в  $\mathbb{R}^n$ , а  $\mathbf{n}$  — вектор нормали к плоскости  $\Pi$ .

Возвращаясь к первоначальному закону геометрической оптики, опишем его чуть более подробно. Возьмем падающий луч света (точка  $\mathbf{x}$  приближается к  $\Pi$ ) и отраженный (точка  $\mathbf{x}$  удаляется от  $\Pi$ ). Через них проходит единственная двумерная плоскость  $\pi$  (если размерность пространства  $\mathbb{R}^n$  больше 2; в случае  $n = 2$  это сама плоскость  $\mathbb{R}^n$ ). В этой же плоскости расположен также и вектор  $\mathbf{n}$  нормали к отражающей  $(n - 1)$ -мерной плоскости  $\Pi$  (который берется в точке отражения). Тогда *угол падения*  $\alpha$  — это угол между падающим лучом света и вектором  $\mathbf{n}$  (или иначе, угол между векторами  $(-\mathbf{v}^-)$  и  $\mathbf{n}$ ), а *угол отражения*  $\beta$  — это угол между отраженным лучом и вектором  $\mathbf{n}$  (т.е. угол между векторами  $\mathbf{v}^+$  и  $\mathbf{n}$ ). Закон оптики гласит:  $\alpha = \beta$  (рис. 2). Он также называется *бильярдным законом*.

Легко видеть, что все три описания бильярдного отображения точки  $\mathbf{x}$  от плоскости  $\Pi$  эквивалентны.

Опишем теперь бильярдное отражение точки  $\mathbf{x}$  от гладкой поверхности  $S \subset \mathbb{R}^n$ . Проведем к  $S$  в точке отражения касательную  $(n - 1)$ -мерную плоскость  $\Pi$  и будем считать, что точка  $\mathbf{x}$  отражается от  $\Pi$  по описанному выше правилу. Это и есть бильярдный закон отражения от  $S$ . Если точка  $\mathbf{x}$  отражается по этому закону от нескольких гладких  $(n - 1)$ -мерных поверхностей (по отдельности) в пространстве  $\mathbb{R}^n$ , то мы будем говорить, что движение  $\mathbf{x}$  бил-

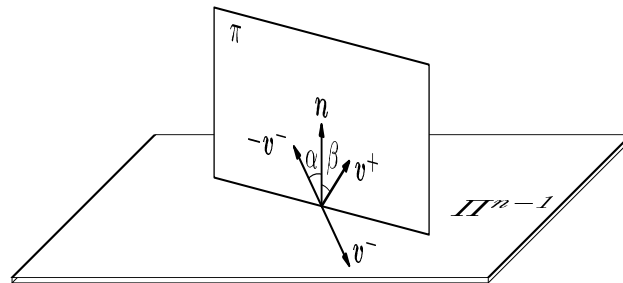


Рис. 2. Бильярдный закон:  $\alpha = \beta$



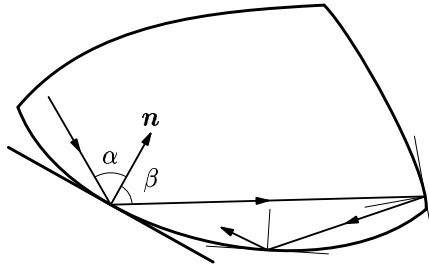


Рис. 3. Биллиардный стол и биллиардная траектория

лиардное, а сама точка  $x$  движется по биллиардной траектории в области пространства  $\mathbb{R}^n$ , ограниченной этими поверхностями. Поверхности отражения для краткости будем называть *стенками*, область, в которой движется точка  $x$ , — (многомерным) *биллиардным столом*, а все движение  $x$  — *биллиардом* (рис. 3). Саму точку  $x$  назовем *биллиардной частицей*.

Напомним, что по определению скорость биллиардной частицы всегда равна 1:  $\|v\| = 1$ . Отметим также, что если биллиардная частица попадает на пересечение двух или большего числа стенок, то ее движение после этого не определено (так как размерность пересечения стенок  $\leq n - 2$ , и касательная  $(n - 1)$ -мерная плоскость  $\Pi$  не определена однозначно). Мы будем считать в этом случае, что траектория частицы в такой точке обрывается. Мера множества начальных положений частицы, для которых траектория не определена (обрывается), равна 0, и такими траекториями мы будем пренебрегать. Таким образом, биллиардные траектории — только те, точки излома которых не принадлежат сразу нескольким стенкам биллиардного стола.

### 3. Задача Синая о столкновении одинаковых шаров

Примерно в середине 1960-х годов Я. Г. Синай, занимаясь доказательством эргодичности однородного газа в сосуде, поставил следующие вопросы. Пусть молекулы газа (шарики) летают в неограниченном пустом пространстве и упруго сталкиваются между собой. Шарики предполагаются одинаковыми (однородный газ), их число — конечным, а упругое соударение означает, что в момент удара скорости столкнувшихся шариков перераспределяются согласно закону сохранения импульса (количества движения) и энергии (точные формулы будут выписаны в следующем разделе).

- 1) Могут ли столкновения в такой системе продолжаться неограниченно долго?
- 2) Может ли число ударов в системе за *конечный* промежуток времени оказаться бесконечным?
- 3) Какое наибольшее число ударов может произойти между молекулами, если разрешается произвольно менять их начальные положения и скорости (при условии, что ответы на оба предыдущих вопроса отрицательны)?

Эти вопросы возникли у Я. Г. Синая естественным образом при исследовании стохастических свойств газа и создания им теории гиперболических биллиардов (которые называются также *рассеивающими* или *биллиардами Синая*,

см. фундаментальную работу Я.Г. Синая [8]). Дело в том, что, как мы покажем в следующей части, конфигурационным пространством системы шариков является многомерное евклидово пространство, из которого выкинута много цилиндров типа (1), где знак неравенства  $\geq$  заменен на противоположный знак  $\leq$  (т. е. из пространства выкинута цилиндры и их внутренности), а движение конфигурационной точки  $v(t)$  подчиняется бильярдному закону (2). Иными словами, газ в исходном трехмерном пространстве, состоящий из одинаковых сферически симметричных молекул, есть не что иное, как бильярд Синая в пространстве большой размерности: стенки бильярда (поверхности выкинутых цилиндров) «вдавлены» внутрь многомерного бильярдного стола — а это одно из требований на геометрию бильярда Синая.

Сделаем оговорку: 6-мерный цилиндр (1) задает во всем пространстве  $\mathbb{R}^{3n}$  другой цилиндр  $C = \Pi \times \mathbb{R}^{3n-6}$ , но граница у  $C$  уже в основном плоская, лишь по некоторым направлениям ее кривизна отлична от нуля. Так что возникающий бильярд — не строго рассеивающий, а полурассеивающий. Однако в работе [8] все трудности, связанные с «полурассеиванием» (т. е. с плоскими участками границы), также были с успехом преодолены.

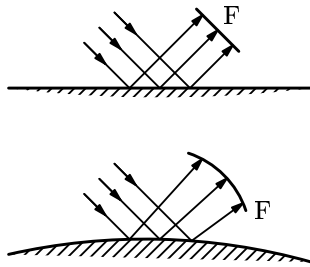


Рис. 4. Волновой фронт отраженных лучей

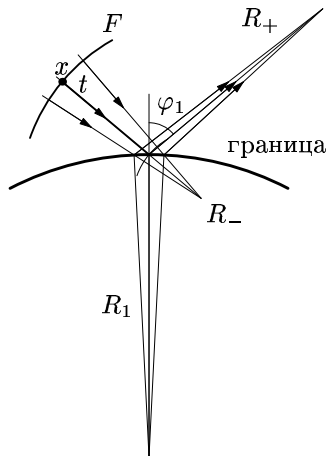


Рис. 5. К выводу уравнения зеркального отражения

Итак, газ сводится к бильярду Синая описанного выше типа. Зачем же нужно знать, конечно или бесконечно число соударений в системе шариков, или, что то же самое, в изоморфном этой системе бильярде?

Здесь мы можем дать ответ на этот вопрос лишь очень приближенно, оставляя основные идеи теории гиперболических бильярдных систем, кроме одной, за скобками. А эту одну идею продемонстрируем для двумерного случая.

Когда параллельный пучок «бильярдных» (или световых) лучей падает на плоский участок границы, он после отражения остается параллельным. Но если участок границы выпуклый (по отношению к параллельному пучку света), то этот пучок после отражения рассеется и волновой фронт  $F$  (поверхность, перпендикулярная отраженным лучам) будет криволинейным (рис. 4). (По этой причине бильярд и называется рассеивающим).

Как меняется кривизна фронта  $F$  при его движении в пространстве после каждого отражения от выпуклого участка границы?

Обратим время, т. е. запустим фронт  $F$  (вместе с пучком) в обратную сторону (см. рис. 5). Пусть  $x \in F$  и пусть фронт  $F$  движется до границы отражения время  $t_1$ . Тогда

$$R(x) = t_1 + R_-, \quad (3)$$

где  $R(x)$  — радиус кривизны фронта в точке  $x$ ,  $t_1$  — расстояние, которое проходит точка  $x$  до гра-

ницы,  $R_-$  — радиус кривизны фронта в точке отражения (до отражения пучка от границы),  $R_+$  — радиус кривизны фронта в точке отражения (после отражения пучка). Из геометрической оптики известна следующая элегантная связь радиусов  $R_-$ ,  $R_+$  и  $R_1$ :

$$\frac{1}{R_-} = \frac{1}{R_+} + \frac{1}{R_1} \cdot \frac{2}{\cos \varphi_1}, \quad (4)$$

где  $\varphi_1$  — угол отражения в точке отражения,  $R_1$  — радиус кривизны стенки; эта формула называется *уравнением зеркального отражения*. Обратные величины к радиусам — это кривизны, так что короче равенство (4) переписывается в терминах кривизн  $\chi_{\pm} = 1/R_{\pm}$ ,  $K_1 = 1/R_1$ :

$$\chi_- = \chi_+ + K_1 \cdot \frac{2}{\cos \varphi_1}. \quad (5)$$

Из (3) получаем  $\chi(x) = \frac{1}{R(x)} = \frac{1}{t_1 + R_-} = \frac{1}{t_1 + \frac{1}{\chi_-}}$ , откуда, используя (5),

находим

$$\chi(x) = \frac{1}{t_1 + \frac{1}{\frac{2K_1}{\cos \varphi_1} + \chi_+}}. \quad (6)$$

Итак, формула (6) дает нам кривизну фронта в точке  $x \in F$  после *одного* отражения от границы. Если же отражений было два, то мы должны принять кривизну  $\chi_+ = \chi_+^{(1)}$  за кривизну фронта в некоторой другой его точке  $y$  и заменить ее аналогичным выражением по формуле (6), только теперь все индексы сменятся с 1 на 2. В результате будем иметь формулу (7):

$$\chi(x) = \frac{1}{t_1 + \frac{1}{\frac{2K_1}{\cos \varphi_1} + \frac{1}{t_2 + \frac{1}{\frac{2K_2}{\cos \varphi_2} + \chi_+^{(2)}}}}}. \quad (7)$$

Продолжая итерировать, получим цепную дробь

$$\chi(x) = \frac{1}{t_1 + \frac{1}{\frac{2K_1}{\cos \varphi_1} + \frac{1}{t_2 + \frac{1}{\frac{2K_2}{\cos \varphi_2} + \frac{1}{t_3 + \dots}}}}}, \quad (8)$$

где многоточие следует заменить на аналогичные выражения столько раз, сколько имеется отражений точки при ее прямом движении во времени, прежде чем она попадет в  $x$ . Заметим, что мы начинали с параллельного пучка, которому отвечает  $\chi_+ = 0$ . Допустим теперь, что  $\chi_+^{(i)} \rightarrow 0$  при  $i \rightarrow \infty$ , т. е. что лучи становятся почти параллельными после бесконечного числа отражений. Тогда получаем бесконечную цепную дробь. В рассуждениях Я. Г. Синая важно, чтобы

эта дробь сходилась — тогда будет определена кривизна  $\chi(x)$  в точке  $x$ . В теории цепных дробей хорошо известен критерий сходимости бесконечной цепной дроби.

КРИТЕРИЙ СХОДИМОСТИ ЦЕПНОЙ ДРОБИ (SEIDEL – STERNE). *Бесконечная цепная дробь*

$$\frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}}, \quad a_i > 0, \quad i = 1, 2, \dots,$$

*сходится тогда и только тогда, когда*

$$\sum_{i=1}^{\infty} a_i = \infty.$$

(См., например, [11].)

Итак, для существования кривизны  $\chi(x)$  *необходимо*, чтобы

$$\sum_{i=1}^{\infty} \left( t_i + \frac{2K_i}{\cos \varphi_i} \right) = \infty.$$

А для этого *достаточно*, чтобы  $\sum_{i=1}^{\infty} t_i = \infty$  (поскольку  $2K_i/\cos \varphi_i > 0$  при всех  $i$ ). Аналогичные рассуждения, только намного более технические, работают и в многомерном случае.

Если число ударов в системе шаров на любом конечном интервале времени конечно, то этого достаточно для расходимости ряда  $\sum_{i=1}^{\infty} t_i$  и, тем самым, для сходимости цепной дроби (8). Вот почему достаточно ответить только на второй вопрос Я. Г. Синая; но первый и третий вопросы интересны сами по себе, и, кроме того, ответив на любой из них, мы одновременно смогли бы ответить также и на второй вопрос.

Я. Г. Синай поставил и решил задачу о столкновениях *одинаковых* шаров в своей (к сожалению, так и не опубликованной) рукописи [9]. А именно, он доказал, что если число шаров фиксировано, то существует такая константа  $C$ , что число соударений между этими шарами в пустом пространстве меньше  $C$ , *независимо* от начального состояния системы, т.е. от положений и скоростей шаров в нулевой момент времени.

Из этого результата вытекает отрицательный ответ на вопрос 2), играющий фундаментальную роль в сходимости цепной дроби (8). Этот ответ стал одним из ключевых моментов в доказательстве Я. Г. Синая эргодичности однородного газа в сосуде.

#### 4. СТОЛКНОВЕНИЯ ТРЕХ ОДИНАКОВЫХ ШАРОВ

Приведем любопытный и довольно тонкий результат о столкновениях *трех* одинаковых шаров в пространстве, известный как «гипотеза Уленбека». В работе американских физиков (Sandri, Sullivan, Norem [18]) 1964-го года была численно просчитана система из трех одинаковых шаров, и оказалось, что при рассмотренных ими начальных положениях и скоростях шаров, число столкновений

между шарами не превосходит 4: четыре столкновения могут произойти, а пять и больше — не могут!

Этот результат был также независимо анонсирован в 1964 г. в работе Thurston'a и Sandri [19]. Строгое его доказательство появилось не очень давно: в 1993 г. американские химик Т. J. Murphy и физик E. G. D. Cohen в совместной работе [16] доказали, что *возможны только* следующие столкновения в системе одинаковых шаров 1, 2 и 3:

- |    |     |                  |                  |
|----|-----|------------------|------------------|
| I. | a)  | (12)             | (одинарный удар) |
|    | b)  | (12)(23)         | (два удара)      |
|    | c1) | (12)(23)(31)     | (три удара)      |
|    | c2) | (12)(23)(12)     | (три удара)      |
|    | d1) | (12)(23)(12)(13) | (четыре удара)   |
|    | d2) | (12)(23)(13)(23) | (четыре удара)   |

и невозможны такие теоретически мыслимые столкновения:

- |     |     |                      |                |
|-----|-----|----------------------|----------------|
| II. | e1) | (12)(23)(12)(23)     | (четыре удара) |
|     | e2) | (12)(23)(13)(12)     | (четыре удара) |
|     | f1) | (12)(23)(12)(13)(12) | (пять ударов)  |
|     | f2) | (12)(23)(13)(23)(12) | (пять ударов). |

Отметим, что подобрать начальные условия в системе с четырьмя ударами d1) или d2) очень непросто; а доказательство невозможности четырех столкновений e1) и e2), а также пяти столкновений f1) и f2) легко влечет доказательство невозможности 6, 7, 8, ... ударов в системе из трех одинаковых шаров (поскольку любая последовательность из 6 ударов содержит одну из последовательностей e1), e2), f1), f2)). Т. Murphy и E. Cohen отмечают, что доказательство невозможности ударов типа II в очень кратком виде были ранее даны E. Cohen'ом в работе [14] 1966 года, и что утверждение про  $\leq 4$  столкновений верно также в пространстве любого числа измерений  $n > 3$ .

Задача про точную верхнюю оценку числа столкновений в системе из четырех или более одинаковых шаров никем не исследовалась.

##### 5. «Малая» и «Большая» бильярдные задачи Я. Г. Синая

Напрашивающееся обобщение задачи Я. Г. Синая об упругом столкновении *одинаковых* шаров состоит в рассмотрении шаров *произвольных* радиусов  $r_1, r_2, \dots, r_N$  и произвольных масс  $m_1, m_2, \dots, m_N$ , летающих и упруго сталкивающихся в пустом пространстве  $\mathbb{R}^n$ , где  $n = 1, 2, 3, \dots$ . Эта задача была поставлена Я. Г. Синаем в середине 1970-х годов на научном семинаре МГУ по динамическим системам под руководством В. М. Алексеева и Я. Г. Синая.

Имеется существенное отличие «одномерного газа» ( $n = 1$ ) от «многомерного газа» ( $n \geq 2$ ): в то время как в одномерной системе шар  $\#i$  может сталкиваться только с шарами  $\#(i - 1)$  и  $\#(i + 1)$ , в многомерном газе априори нет ограничений для столкновения шаров  $\#i$  и  $\#k$  с  $|i - k| \geq 2$ . (Такие ограничения могут появиться лишь в результате скрупулезного изучения системы, когда, как это было в случае трех одинаковых шаров, обнаруживается, что не всякие последовательности ударов осуществимы.) Одномерная система шаров (вернее, стержней на прямой) может быть редуцирована к системе точечных упругих *частиц* на

прямой (когда все радиусы  $r_i = 0$ ), в то время как в случае дисков на плоскости ( $n = 2$ ) или шаров в пространстве ( $n \geq 3$ ) такой редукции сделать нельзя.

Напомним, что мы исключаем из рассмотрения такие события, как более чем двойные столкновения (частиц, дисков или шаров), а также одновременные двойные столкновения более чем двух шаров: эти события происходят с вероятностью 0.

В следующем разделе будут описаны конфигурационные пространства (а) системы  $N$  частиц на прямой  $\mathbb{R}^1$  и (б) системы  $N$  дисков на плоскости  $\mathbb{R}^2$  или  $N$  шаров в пространстве  $\mathbb{R}^n$ ,  $n \geq 3$ . Они качественно различны: если в случае (а) конфигурационным пространством является многогранный угол специального вида в пространстве  $\mathbb{R}^N$ , то в случае (б) конфигурационное пространство — это пространство  $\mathbb{R}^{Nn}$ , из которого выкинута довольно замысловатая система из  $\binom{n}{2}$  цилиндров (и их внутренностей), причем все цилиндры имеют общее непустое пересечение. В общем случае, когда радиусы  $r_1, r_2, \dots, r_N$  и массы  $m_1, m_2, \dots, m_N$  различны, конфигурационная точка  $\mathbf{x}(t) \in \mathbb{R}^{Nn}$  движется между столкновениями со стенками — гранями угла или цилиндрами — вдоль прямых (как это и было раньше, когда все шары были одинаковыми), но ее отражения от стенок уже вовсе не бильiardные!

Тем не менее, сделав довольно простое — линейное! — преобразование пространства  $\mathbb{R}^{Nn}$ , Я. Г. Синай сумел свести задачу движения точки  $\mathbf{x}(t)$  к задаче о бильiardном движении новой точки  $\tilde{\mathbf{x}}(t)$  в преобразованном конфигурационном пространстве.

Таким образом, общая задача о шарах разных размеров и масс также сводится к бильiardу.

После такого сведения Я. Г. Синай сделал два естественных допущения: в задаче (а) про частицы он отказался от специального вида многогранного угла и рассмотрел бильiard в произвольном многогранном угле в пространстве  $\mathbb{R}^N$ , а в задаче (б) он заменил цилиндры в многомерном пространстве на конечное множество произвольных выпуклых тел в  $\mathbb{R}^K$  (где  $K$  не обязательно равно  $nN$ ) с таким свойством: границы тел гладкие и *все эти тела имеют непустое пересечение*, а конфигурационная точка  $\mathbf{z}(t)$  отражается от границ этих тел (стенок) по бильiardному закону. Отличие задачи (а) от (б) в том, что в задаче (а) такими выпуклыми телами служат полупространства, и бильiardное движение можно исследовать более или менее элементарными методами, а в задаче (б) у стенок имеются кривизны, которые создают рассеяние при отражении от них, и исследование становится намного более сложным. Задачу (б) тоже можно рассматривать как бильiard в многогранном угле, но со «вдавленными внутрь» стенками (рис. 6). Я назвал задачи (а) и (б) «малой» и «большой» бильiardными задачами потому, что это фактически одна и та же задача — о бильiardе в угле, только в «малой» задаче угол этот — обычный многогранный, с плоскими гранями, а в «большой» задаче грани-стенки многогранного угла — выпуклы внутрь него, и это обстоятельство (выпуклость стенок) создает «нелинейные» трудности при ее решении. Приведем точные формулировки бильiardных задач Я. Г. Синая.

«МАЛАЯ» ЗАДАЧА СИНАЯ. Доказать или опровергнуть, что в любом многогранном угле в  $\mathbb{R}^n$  с плоскими стенками число отражений бильiardной частицы от

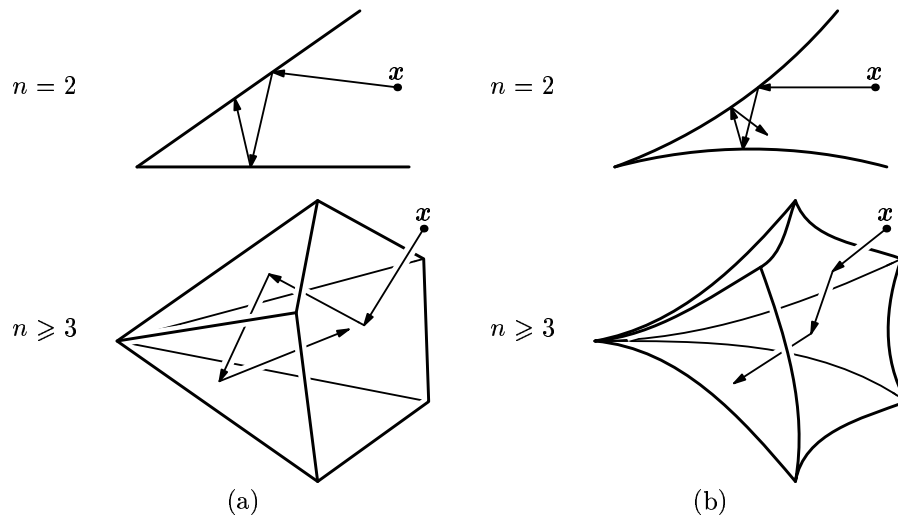


Рис. 6. «Малая» (a) и «большая» (b) задачи Синай

его стенок (i) конечно, и (ii) это число не превосходит некоторой константы  $C_1$ , зависящей только от геометрии угла, но не от начальных условий — положения и направления скорости частицы (рис. 6a).

«Большая» задача Синай. Доказать или опровергнуть, что в любом «криволинейном» угле с гладкими стенками, выпуклыми внутрь угла (стенки — границы выпуклых тел, имеющих в совокупности непустое пересечение) число отражений от его стенок (i) конечно, и (ii) при некоторых ограничениях на геометрию угла не превосходит некоторой константы  $C_2$ , зависящей только от геометрии угла, но не от начальных условий, накладываемых на бильiardную частицу (начального положения и скорости).

Под «некоторыми ограничениями на геометрию криволинейного угла» Я. Г. Синай имел в виду похожесть его на «плоский» угол, а именно: грани криволинейного угла должны быть трансверсальными, т. е. не иметь общей касательной.

## 6. Задача Синай о столкновениях произвольных шаров в пространстве

Одновременно с формулировками «малой» и «большой» бильiardных задач Я. Г. Синай сформулировал и их «динамические» аналоги.

Задача о частицах на прямой. Доказать или опровергнуть, что число столкновений в системе  $N$  частиц произвольных масс  $m_1, \dots, m_N$ , движущихся на бесконечной прямой, (i) конечно, и (ii) это число не превосходит некоторой константы  $C_1 = C_1(m_1, \dots, m_N)$ , зависящей только от масс, но не от начальных положений частиц и их скоростей. Выяснить характер функции  $C_1 = C_1(m_1, \dots, m_N)$ .

**ЗАДАЧА О ШАРАХ.** Доказать или опровергнуть, что число столкновений  $N$  шаров ( $N$  дисков в случае плоскости) произвольных ненулевых радиусов  $r_1, \dots, r_N$  и масс  $m_1, \dots, m_N$ , сталкивающихся в пространстве  $\mathbb{R}^n$  ( $n \geq 2$ ), (i) конечно и (ii) это число не превосходит некоторой константы  $C_2 = C_2(r_1, \dots, r_N, m_1, \dots, m_N)$ , зависящей только от радиусов и масс шаров, но не от начальных положений шаров и их скоростей. Выяснить характер функции  $C_2 = C_2(r_1, \dots, r_N, m_1, \dots, m_N)$ .

Бильярдные задачи Я. Г. Синая и его задачи о частицах и шарах привлекли внимание многих математиков; их решение заняло почти тридцать лет — с середины 1970-х годов и до конца XX века.

## 7. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

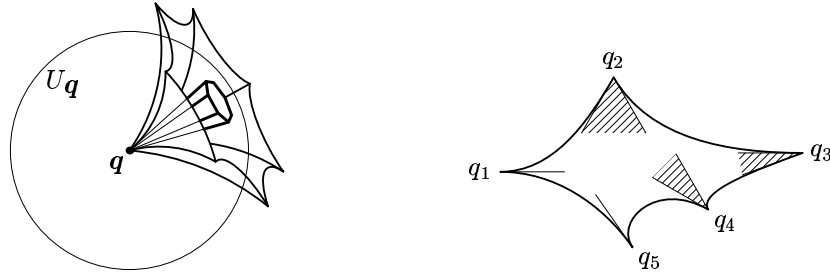
### 7.1. «МАЛАЯ» ЗАДАЧА СИНАЯ

«Малая» задача была решена положительно независимо и разными методами тремя математиками. Первым ее полностью решил в 1978 г. сам Я. Г. Синай в работе [10], доказав индукцией по размерности конфигурационного пространства существование равномерной оценки  $C_1$  для числа ударов. Фактически это доказательство неявно содержалось уже в его неопубликованной рукописи [9] о столкновениях *одинаковых* шаров. В этой рукописи, помимо большого числа тонких рассуждений, вводится одно важное понятие: «открытая точка края». Грубо говоря, если точка  $q$  принадлежит сразу нескольким стенкам многомерного бильярда, причем стенки *выступы внутрь* стола, как на рис. 6b), и в эту точку можно — в какой-нибудь ее малой окрестности  $U_q$  — вставить плоский ненулевой многогранный угол (как на рис. 6a) с вершиной  $q$  (точнее — окрестность вершины многогранного угла), то точка  $q$  называется открытой точкой края. Точнее, для открытой точки края  $q$  многогранный угол

$$Q_q = \{e \in \mathbb{R}^n \mid \forall i (e, n_i) > 0\} \quad (9)$$

не лежит в  $(n - 1)$ -мерной плоскости; здесь  $n_1, \dots, n_k$  — единичные векторы нормалей ко всем стенкам бильярда в их общей точке  $q$ , направленные внутрь бильярдного стола. На рис. 7a) указана открытая точка края  $q$  — вершина криволинейного угла, в который (в малой окрестности  $U_q$ ) вставлен ненулевой плоский многогранный угол. Открытыми точками края являются также точки  $q_2, q_3$  и  $q_4$  плоского «бильярда Синая» на рис. 7b), а вершины  $q_1$  и  $q_5$  того же бильярда открытыми точками не являются. Идея Я. Г. Синая состояла в том, что если максимальное число ударов бильярдной частицы в плоском многогранном угле с вершиной  $q$  не превосходит константы  $C_1$ , то в многогранном *криволинейном* угле с той же вершиной  $q$ , полученном из плоского угла «вдавливанием» его граней внутрь (при этом стенки угла становятся гладкими поверхностями в регулярных точках) число ударов не будет превосходить  $kC_1$ , где  $k \geq 1$ , поскольку вдавленные внутрь стенки могут лишь либо уменьшить общее число ударов, либо увеличить, но не более, чем в некоторое число  $k$  раз. Однако технические трудности, возникающие при обосновании этой идеи, настолько велики (они в основном связаны с почти касательными отражениями от криволинейных стенок), что решение «большой» задачи не было доведено в рукописи до конца. (О решении «большой» задачи см. ниже.)



а) Открытая точка края  $q$ б)  $q_2, q_3, q_4$  — открытые,  $q_1, q_5$  — неоткрытые точки края**Рис. 7.**

Другие два решения «малой» задачи появились после решения Я. Г. Синая с двумя разрывами примерно в 10 лет каждый.

Второе решение, полученное Г. А. Гальпериним в 1986 г., опубликовано в книге [3]. Как и решение Я. Г. Синая, оно проводится индукцией по размерности конфигурационного пространства, однако в рассуждения вовлекается биллиард на  $(n - 1)$ -мерной сфере и наглядные геометрические соображения.

Третье решение получено М. Б. Севрюком в [7]. Оно основано на следующей лемме: если какая-то точка в  $\mathbb{R}^n$  отражается относительно конечного числа плоскостей размерности  $(n - 1)$  и при этом некоторая линейная функция от этой точки строго возрастает при каждом отражении, то число таких отражений конечно и не превосходит универсальной константы, зависящей только от функции и от расположения плоскостей в пространстве, но не зависит от выбора отражаемой точки. Теорема Синая получается после этого с помощью рассмотрения довольно простой специальной линейной функции.

## 7.2. Задача о частицах на прямой

Положительное решение задачи о частицах есть прямое следствие положительного решения «малой» задачи Синая: движение  $n$  частиц моделируется биллиардным движением одной частицы в специальном многогранном угле (с плоскими стенками) в  $\mathbb{R}^n$ . Его можно рассматривать как *геометрическое* решение задачи о частицах.

В статье Г. А. Гальперина [1], появившейся в 1978 г. одновременно с работой [10], было предложено другое, *аналитическое* доказательство конечности числа столкновений упругих частиц на прямой — но без равномерной по начальным условиям оценки. Тем самым, было получено независимое решение задачи (i). Полное аналитическое доказательство части (ii) было впервые дано Г. А. Гальпериним в работе [2, Приложение 1], где приведено явное значение для константы  $C_1(m_1, \dots, m_N)$ :

$$C_1(m_1, \dots, m_N) = 2 \cdot \left( 8N^2(N - 1) \frac{m_{\max}}{m_{\min}} \right)^{N-2}, \quad (10)$$

здесь  $m_{\max}$  — максимальная, а  $m_{\min}$  — минимальная из масс  $m_1, \dots, m_N$ .

Другое, полуаналитическое-полугеометрическое решение дал в 1992 г. М. Б. Севрюк в упомянутой выше работе [7]. В нем он рассмотрел «скоростное» пространство системы вместо конфигурационного, к которому затем применил свою идею об отражении точки в гиперплоскостях этого пространства. В результате получилось новое доказательство существования равномерной оценки для числа ударов в системе частиц.

Интересно, что для трех частиц  $m_1, m_2, m_3$  на прямой можно явно выписать верхнюю оценку числа ударов между ними, причем эта оценка, в отличие от общей оценки (10), точна:

$$C_1(m_1, m_2, m_3) = \left\lceil \frac{\pi}{\arccos \sqrt{\frac{m_1 m_3}{(m_1 + m_2)(m_2 + m_3)}}} \right\rceil \quad (11)$$

(здесь  $\lceil x \rceil$  — наименьшее целое число, большее или равное  $x$ ). Конфигурационным пространством системы  $\{m_1, m_2, m_3\}$  является двугранный угол в  $\mathbb{R}^3$ , линейный угол которого равен  $\alpha = \arccos \sqrt{\left(1 + \frac{m_2}{m_1}\right)^{-1} \left(1 + \frac{m_2}{m_3}\right)^{-1}}$ , а число бильярдных ударов в плоском угле  $\alpha$  всегда  $\leq \lceil \pi/\alpha \rceil$  (см. теорему 2, раздел III).

### 7.3. «Большая» задача Синая и задача о шарах, часть (I)

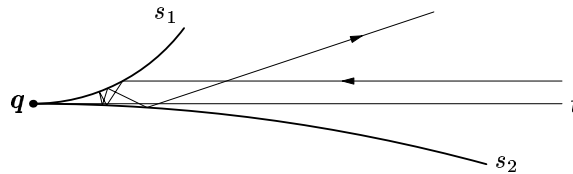
Ее решение, как и решение «малой» задачи, было также получено тремя математиками, однако, в отличие от «малой» задачи, не совсем независимо. Л. Н. Вассерштейн и Г. А. Гальперин в 1978–1979 гг. исследовали совместно два типа систем частиц в пространстве: (а) центрально отталкивающихся и (б) локально взаимодействующих. Для таких систем они построили аксиоматическую теорию взаимодействия и доказали существование так называемых «левых» и «правых» скоростей (некоторых специальных левых и правых пределов вида  $\lim \frac{\Delta s}{\Delta t}$ ), а также теорему о распадении системы (б) на не взаимодействующие с некоторого момента времени группы ограниченного размера (кластеры). Эти результаты составили первую часть статей [20] и [2].

Вторые части этих статей посвящены доказательству конечности числа ударов бильярдной частицы в криволинейном многогранном угле на любом конечном интервале времени, а затем доказательству конечности числа ударов в системе твердых шаров на всем бесконечном интервале времени  $[0, +\infty)$ . Доказательства этих утверждений в обеих статьях совершенно непохожи и основываются на разных геометрических (и физических!) идеях.

Еще одно альтернативное доказательство задачи о шарах дал в 1990 г. канадский математик R. Illner [15], упростивший доказательство Л. Н. Вассерштейна.

### 7.4. «Большая» задача Синая и задача о шарах, часть (II)

Отметим, что ответ на вопрос (ii) в «большой» задаче Синая будет заведомо отрицательным, если опустить слова «при некоторых ограничениях на геометрию угла» в ее формулировке. Действительно, рассмотрим криволинейный угол на плоскости, стороны которого  $s_1$  и  $s_2$  выпуклы навстречу друг другу, но имеют общую касательную  $t$  (рис. 8). Если запустить бильярдную частицу параллельно  $t$ , то она после *конечного* числа ударов со сторонами  $s_1$  и  $s_2$ , согласно



**Рис. 8.** Неограниченность количества ударов в «нулевом» угле для траектории, приближающейся к общей касательной к сторонам

теореме Вассерштейна – Гальперина – Иллнера, будет двигаться свободно вдоль прямой, удаляясь от вершины  $q$  и больше не соударяясь со сторонами угла. Однако, чем ближе к касательной  $t$  мы запустим частицу, тем больше соударений со сторонами угла она совершит, и это число будет стремиться к  $\infty$  при приближении траектории к  $t$ . Таким образом, чтобы надеяться на положительный ответ на вопрос (ii) в «большой» задаче Синая, совершенно необходимы «некоторые ограничения на геометрию угла». В двумерном случае криволинейный угол должен быть *ненулевым*, т. е. касательные  $t_1$  и  $t_2$  к сторонам  $s_1$  и  $s_2$  в вершине  $q$  должны быть различны; тогда  $\angle t_1 t_2 = \alpha \neq 0$ . Можно доказать, что как и в случае плоского угла, для криволинейного угла на плоскости максимальное число соударений частицы со сторонами угла  $\leq [\pi/\alpha]$  (см. раздел III, Утверждение 1). В многомерном случае «ограничением на геометрию угла» может служить открытость всех точек края (см. формулу (9) и рис. 7) или похожее условие. Мы уже отмечали выше, что использование понятия «открытой точки края» в прямом виде ведет к настолько большим техническим трудностям, что практически не позволяет аккуратно доказать верхнюю равномерную оценку на число ударов.

Намного более простым и существенно более эффективным оказалось другое ограничение на геометрию криволинейного угла, так называемая « $d/C$ -близость», введенная Д. Бураго, С. Ферлегером и А. Каноненко в работе [12].

Зафиксируем константу  $C > 0$  и для произвольной точки  $y$  внутри криволинейного угла найдем все расстояния от нее до стенок угла  $\text{dist}(y, W_k)$ , расстояние  $\text{dist}(y, \bigcap_{k=1}^N W_k) = \text{dist}(y, q)$  от  $y$  до вершины угла  $q$  (т. е. до точки пересечения всех стенок), затем найдем все отношения  $\rho_k = \text{dist}(y, W_k) / \text{dist}(y, q)$ ,  $k = 1, \dots, N$ . Криволинейный угол называется *невырожденным с константой  $C > 0$*  или  *$C$ -невырожденным*, если для каждого  $k = 1, \dots, N$  и для любой точки  $y$  выполнено  $\rho_k(y) \geq C$ . Проще говоря, это означает, что если точка  $y$   $d$ -близка ко всем стенкам  $W_1, \dots, W_N$  (т. е. находится на расстоянии  $\leq d$  от каждой стенки), то она  $\frac{d}{C}$ -близка к пересечению этих стенок (рис. 9). Криволинейный угол называется просто *невырожденным*, если найдутся такие константы  $\delta > 0$  и  $C > 0$ , что пересечение этого угла с произвольным  $\delta$ -шаром является  $C$ -невырожденным.

Основываясь на таком, «арифметическом» определении невырожденности угла, которое, отметим, эквивалентно невырожденности в смысле Синая, авторы статьи [12] доказали существование универсальной константы  $C_2$ , требуемой в «большой» задаче Синая, (ii).

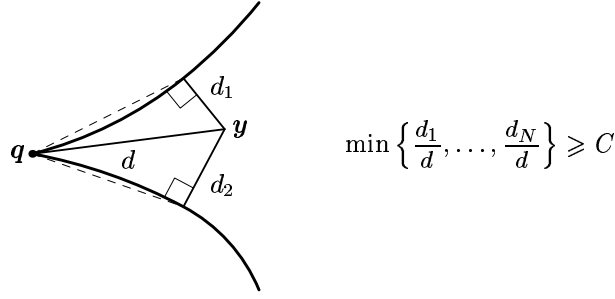


Рис. 9. Невырожденность угла с константой  $C$

Применяя этот результат к конфигурационному пространству системы шаров, они получили в [12] следующую равномерную оценку для числа ударов между шарами во всем пространстве:

$$C_2(r_1, \dots, r_N; m_1, \dots, m_N) = \left( 16 \sqrt{\frac{m_{\max}}{m_{\min}}} \cdot \frac{r_{\max}}{r_{\min}} \cdot N^{3/2} \right)^{N^2}. \quad (12)$$

Позднее, в работе [13], авторам удалось избавиться от соотношения между радиусами. Они доказали, что при любых радиусах шаров число столкновений в системе шаров во всем пространстве не превосходит универсальной константы  $C_2 = C_2(m_1, \dots, m_N)$ , которая не зависит от начальных положений и скоростей шаров, а зависит только от их масс:

$$C_2(m_1, \dots, m_N) = \left( 400 N^2 \cdot \frac{m_{\max}}{m_{\min}} \right)^{2N^4}. \quad (13)$$

Этими исследованиями была поставлена окончательная точка в решении задач Я. Г. Синая. В дальнейшем эти результаты были использованы для изучения системы шаров в сосудах с «вдавленными» внутрь и плоскими стенками.

### III. КРАТКИЕ ИДЕИ ДОКАЗАТЕЛЬСТВ

В этом разделе мы, прежде всего, сформулируем точное правило пересчета скоростей частиц и шаров после их столкновения. Затем дадим описание конфигурационного пространства для частиц на прямой и шаров в пространстве. Наконец, сделаем редукцию к бильярдам и сведем задачи про частицы и шары к бильiardным задачам Синая. После этого приступим к изложению идей доказательств теорем, о которых мы рассказывали в части II.

#### А. МОДЕЛЬ УДАРА

Если  $m_1, m_2$  — массы частиц на прямой,  $v_1, v_2$  — их скорости до удара, а  $u_1, u_2$  — скорости после удара, то  $u_1$  и  $u_2$  находятся из двух уравнений, первое из которых выражает закон сохранения импульса, а второе — закон сохранения

энергии:

$$\begin{cases} m_1 u_1 + m_2 u_2 = m_1 v_1 + m_2 v_2 = \text{const}_1, \\ \frac{m_1 u_1^2}{2} + \frac{m_2 u_2^2}{2} = \frac{m_1 v_1^2}{2} + \frac{m_2 v_2^2}{2} = \text{const}_2. \end{cases} \quad (14)$$

Для шаров, движущихся до столкновения со скоростями  $\mathbf{v}_1$  и  $\mathbf{v}_2$ , а после столкновения — со скоростями  $\mathbf{u}_1$  и  $\mathbf{u}_2$ , следует поступить так. Сначала каждую скорость  $\mathbf{v}_i$  надо разложить в векторную сумму двух: скорости  $\mathbf{v}_{i,\parallel}$  вдоль линии центров  $O_1 O_2$  и перпендикулярной ей скорости  $\mathbf{v}_{i,\perp}$  ( $i = 1, 2$ ), так что  $\mathbf{v}_{i,\parallel}$  — это ортогональная проекция  $\mathbf{v}$  на прямую  $O_1 O_2$ , а  $\mathbf{v}_{i,\perp} = \mathbf{v} - \mathbf{v}_{i,\parallel}$  (рис 10). После этого шары вдоль направления  $O_1 O_2$  следует рассмотреть как сталкивающиеся частицы, применить для их скоростей  $\mathbf{v}_{1,\parallel}$  и  $\mathbf{v}_{2,\parallel}$  формулы (14), вычислить величины новых скоростей этих частиц  $\mathbf{u}_{1,\parallel}$  и  $\mathbf{u}_{2,\parallel}$ , после чего найти скорости после удара:

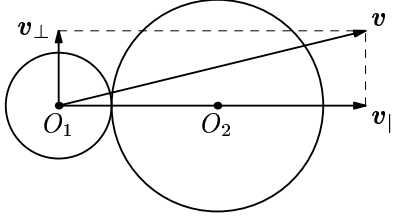


Рис. 10. Соударение шаров

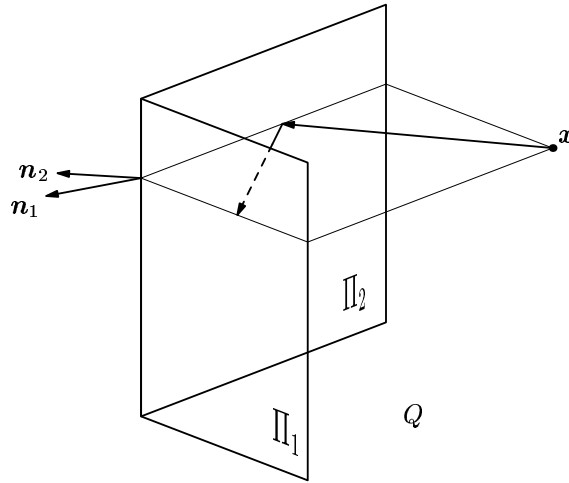
$$\mathbf{u}_1 = \mathbf{u}_{1,\parallel} + \mathbf{v}_{1,\perp}, \quad \mathbf{u}_2 = \mathbf{u}_{2,\parallel} + \mathbf{v}_{2,\perp}. \quad (15)$$

Итак, перераспределение скоростей шаров происходит лишь вдоль линии центров, а ортогональные составляющие скорости не меняются. Никакого закручивания шаров также не происходит. Иными словами, мы рассматриваем шары, как «распухшие» частицы. (Конечно, столкновению реальных шаров на бильярдном столе сопутствует вращение, но это происходит в основном из-за наличия трения. Без трения вращение отсутствует, и мы не должны принимать в расчет в нашей модели моменты инерции шаров.)

Интересно, что если шары имеют одинаковую массу, причем один шар покоится, то при центральном ударе (вдоль линии  $O_1 O_2$ ) налетевший шар полностью останавливается, а покоящийся начинает двигаться со скоростью налетевшего; дело происходит так, будто первый шар прошел сквозь второй и они просто обменялись номерами. При нецентральном ударе такая же картина наблюдается для скорости  $\mathbf{v}_{1,\parallel}$ , и поэтому одинаковые шары разлетаются (по теореме Пифагора) под углом  $90^\circ$ . Для шаров разных масс картина разлета намного сложнее.

### В. КОНФИГУРАЦИОННОЕ ПРОСТРАНСТВО ЧАСТИЦ НА ПРЯМОЙ

Пусть  $x_1 \leq x_2 \leq x_3 \leq \dots \leq x_N$  — координаты частиц с массами  $m_1, m_2, m_3, \dots, m_N$ , сталкивающихся на прямой согласно законам сохранения (14). Конфигурационная точка  $\mathbf{x} = (x_1, \dots, x_N) \in \mathbb{R}^N$  движется равномерно между соударениями частиц и меняет свою траекторию, если какие-то две соседние частицы  $m_i$  и  $m_{i+1}$  сталкиваются. В этот момент  $x_i = x_{i+1}$ , а все остальные координаты различны (более чем двойные столкновения мы для простоты отбросим). Уравнение  $x_i = x_{i+1}$  задает гиперплоскость  $\Pi_i$  в  $\mathbb{R}^N$ ; таким образом, точка  $\mathbf{x}$  как-то отражается от плоскости  $\Pi_i$ . Ясно, что если  $m_i = m_{i+1}$ , то это отражение бильярдное, но при  $m_i \neq m_{i+1}$  это уже не так. Плоскости  $\Pi_1, \dots, \Pi_{N-1}$  образуют стенки  $(N-1)$ -гранного угла в  $\mathbb{R}^N$  (например, в  $\mathbb{R}^3$  это двугранный угол, см. рис. 11), а часть  $Q$  пространства  $\mathbb{R}^N$ , задаваемая неравенствами



**Рис. 11.** Конфигурационное пространство  $Q$  системы частиц на прямой — многогранный угол  $x_1 \leq x_2 \leq \dots \leq x_N$  в  $\mathbb{R}^N$

$x_1 \leq x_2 \leq \dots \leq x_N$  — сам многогранный угол. Он и является конфигурационным пространством  $Q \subset \mathbb{R}^N$  системы  $N$  частиц.

### С. КОНФИГУРАЦИОННОЕ ПРОСТРАНСТВО ШАРОВ

Пусть шары радиусов  $r_1$  и  $r_2$  столкнулись в пространстве  $\mathbb{R}^n$ , тогда

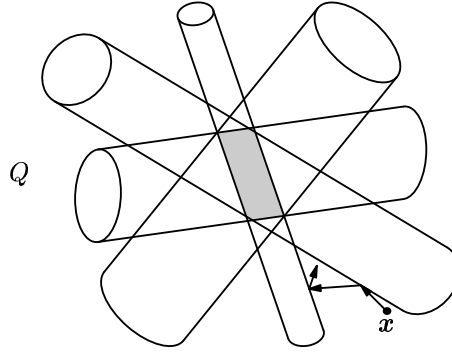
$$(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2 = (r_1 + r_2)^2, \quad (16)$$

где  $O_1 = (x_1, \dots, x_n)$  и  $O_2 = (y_1, \dots, y_n)$  — координаты центров шаров в момент столкновения. Равенство (16) задает  $(2n - 1)$ -мерный цилиндр  $\Pi$  в  $\mathbb{R}^{2n}$ . Если заменить равенство в (16) на неравенство  $<$ , то получится *невозможное* расположение шаров в  $\mathbb{R}^n$  и изображающая точка  $(x_1, \dots, x_n, y_1, \dots, y_n)$  поэтому не может оказаться внутри цилиндра  $\Pi$ . Однако она может попасть на поверхность цилиндра  $\Pi$ , что отвечает столкновению шаров, а также летать вне цилиндра — тогда равенство (16) заменяется на неравенство  $>$ , и шары находятся на некотором расстоянии друг от друга.

Рассмотрим теперь  $N$  шаров в пространстве  $\mathbb{R}^n$ , и пусть шары  $\#i$  и  $\#k$  столкнулись. Для них тогда выполняется равенство (16) (в котором ко всем координатам  $x$  надо добавить индекс  $i$ , ко всем  $y$  — индекс  $k$ , а  $r_1, r_2$  заменить на  $r_i, r_k$ ), а для всех остальных шаров — неравенство  $>$ . В результате мы получаем конфигурационную точку  $x$  в пространстве  $\mathbb{R}^{nN}$ , лежащую в множестве

$$C_{ij} = \Pi_{ij} \times \mathbb{R}^{nN-2n}. \quad (17)$$

$C_{ij}$  — это тоже цилиндр, но уже в пространстве  $\mathbb{R}^{nN}$ , и у него граница в направлениях, ортогональных подпространству  $\mathbb{R}^{2n}$ , в котором лежит цилиндр  $\Pi$ , плоская. Положительную кривизну цилиндр  $C_{ij}$  имеет только вдоль цилиндра  $\Pi_{ij}$ .



**Рис. 12.** Конфигурационное пространство шаров  $Q \subset \mathbb{R}^{nN}$

Точка  $x$  не может лежать внутри цилиндра  $C_{ij}$ , поскольку шары непрозрачны. Так что  $x$  лежит вне всех цилиндров, число которых равно  $\binom{N}{2} = \frac{N(N-1)}{2}$ , и поэтому конфигурационным пространством  $Q$  системы шаров является  $\mathbb{R}^{nN}$  с выкинутыми внутренностями цилиндров (17):

$$Q = \mathbb{R}^{nN} - \left( \bigcup_{i \neq j} (\text{int } C_{ij}) \right). \quad (18)$$

Важно отметить, что внутренности всех цилиндров  $C_{ij}$  — выпуклые тела, и что

$$\bigcap_{i \neq j} (\text{int } C_{ij}) \neq \emptyset. \quad (19)$$

Действительно, невозможное положение для конфигурационной точки  $x$  такое, когда один или несколько шаров проникли друг в друга. В частности, если все шары имеют непустое пересечение в  $\mathbb{R}^n$ , то им отвечает невозможная точка в  $\mathbb{R}^{nN}$ , лежащая в пересечении внутренностей всех цилиндров  $\text{int } C_{ij}$ . Поэтому пересечение внутренностей цилиндров непусто и является выпуклым множеством (как пересечение выпуклых множеств).

В результате получаем, что конфигурационное пространство  $Q$  — криволинейный многогранный угол (в силу (19)), у которого все стенки «вдавлены внутрь»  $Q$  (так как все цилиндры — выпуклы). Однако, как и в случае частиц на прямой, движение точки  $x$  в  $Q \subset \mathbb{R}^{nN}$  будет биллиардным только если  $m_i = m_j$ , и не будет таковым при  $m_i \neq m_j$  (рис. 12).

#### Д. Редукция к биллиарду

Рассмотрим конфигурационное пространство  $Q \subset \mathbb{R}^{nN}$  (частиц на прямой при  $n = 1$ , или шаров в пространстве  $\mathbb{R}^n$ ). Сделаем следующее линейное преобразование пространства  $\mathbb{R}^{nN}$ , которому также подвергнется как многогранный угол  $Q$ , так и траектория конфигурационной точки  $x$ :

$$\tilde{x}_i = \sqrt{m_i} \cdot x_i, \quad i = 1, \dots, nN. \quad (20)$$

Иными словами, матрица  $A$  линейного оператора — диагональная, с числами  $\sqrt{m_1}, \dots, \sqrt{m_N}$  по диагонали:  $A = \text{diag}(\sqrt{m_1}, \dots, \sqrt{m_N})$ . Заметим, что скорость движения конфигурационной точки подвергнется тому же преобразованию:

$$\tilde{v}_i = \frac{d\tilde{x}_i}{dt} = \sqrt{m_i} \frac{dx_i}{dt} = \sqrt{m_i} v_i. \quad (21)$$

Выбрав соответствующую единицу измерения, можно считать, что полная кинетическая энергия всех шаров (или частиц при  $n = 1$ ) равна  $E = 1/2$ . Тогда скорость  $\mathbf{v}(t) = (\sqrt{m_1}v_1(t), \dots, \sqrt{m_N}v_N(t)) = (\tilde{v}_1(t), \dots, \tilde{v}_N(t))$  конфигурационной точки  $\mathbf{x} = (x_1, \dots, x_{nN})$  имеет модуль 1:

$$|\mathbf{v}(t)|^2 = \sum_{i=1}^{Nn} |\tilde{v}_i(t)|^2 = 2E = 1. \quad (22)$$

Следовательно, длина вектора  $\mathbf{v}(t)$  всегда постоянна и равна 1. Закон сохранения импульса (14) может быть переписан в виде

$$m_1 v_1 + m_2 v_2 = \sqrt{m_1} \tilde{v}_1 + \sqrt{m_2} \tilde{v}_2 = \begin{pmatrix} \sqrt{m_1} \\ \sqrt{m_2} \end{pmatrix} \cdot \begin{pmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{pmatrix} = |\vec{m}| \cdot |\vec{v}| \cdot \cos \varphi = \text{const}, \quad (23)$$

где  $\vec{m} = \begin{pmatrix} \sqrt{m_1} \\ \sqrt{m_2} \end{pmatrix}$ ,  $\vec{v} = \begin{pmatrix} \tilde{v}_1 \\ \tilde{v}_2 \end{pmatrix}$ , а  $\varphi$  — угол между векторами  $\vec{m}$  и  $\vec{v}$ . Рассматривая столкновение только двух шаров, #1 и #2, и не меняя скоростей остальных шаров, получаем, что  $|\vec{v}| = 1$ ,  $|\vec{m}| = \text{const}_1$ , так что равенство (23) утверждает: *скалярное произведение двух постоянных по длине векторов не меняется после удара*. Значит,  $\cos \varphi = \text{const}_2$ , а так как угол  $\varphi$  должен измениться после отражения, то он просто меняет знак. Это и означает, что точка  $\tilde{\mathbf{x}}$  отражается от соответствующей стенки угла  $Q$  по бильiardному закону (подробное доказательство см. в [2, с. 164–165]). Редукция к бильiardу завершена.

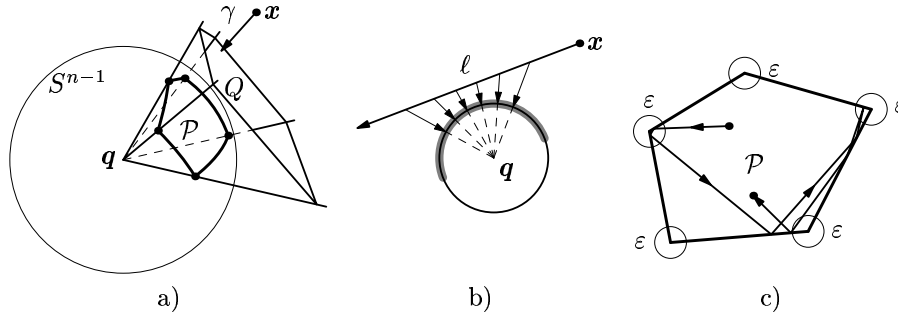
Теперь мы можем приступить к краткому изложению идей доказательств.

### 1. «Малая» задача Синая

**ТЕОРЕМА 1.** *Число отражений любой бильiardной траектории в многогранном угле  $Q$  с плоскими стенками в  $\mathbb{R}^n$  конечно; оно ограничено сверху некоторой константой  $C_1$ , зависящей только от геометрии угла.*

**ДОКАЗАТЕЛЬСТВО.** Приведем наглядное геометрическое доказательство, следуя [3]. Рассмотрим единичную сферу  $S^{n-1} \subset \mathbb{R}^n$  с центром в вершине  $\mathbf{q}$  многогранного угла и спроектируем этот угол на сферу из центра  $\mathbf{q}$  (см. рис. 13а). Получим выпуклый сферический многогранник  $\mathcal{P} \subset S^{n-1}$ . При центральной проекции бильiardная траектория  $\gamma$  частицы  $\mathbf{x}$  в угле перейдет в бильiardную траекторию  $\tilde{\gamma}$  проекции  $\mathbf{x}$  в многограннике  $\mathcal{P}$  (в сферической метрике). Выпрямим траекторию  $\gamma$  в  $\mathbb{R}^n$ , отражая симметрично угол  $Q$  вместе с траекторией  $\gamma$  относительно тех его граней, от которых отражается бильiardная частица  $\mathbf{x}$ , движущаяся вдоль  $\gamma$ . Получим прямую  $\ell$  в пространстве  $\mathbb{R}^n$ ; ее проекция на сферу — половина большой окружности и имеет длину  $\pi$  (см. рис. 13б). Следовательно, бильiardная траектория  $\tilde{\gamma}$  частицы  $\tilde{\mathbf{x}}$  в  $\mathcal{P}$  имеет конечную длину, равную  $\pi$ .





**Рис. 13.** а) Центральная проекция  $Q \rightarrow P$ ; б) Проекция выпрямленной траектории на  $S^{n-1}$ ; в) траектория  $\tilde{\gamma}$  в  $P$

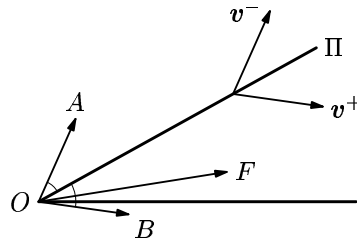
Доказательство конечности числа отражений в  $Q$  проведем индукцией по размерности. Рассмотрим (см. рис. 13с) объединение  $\varepsilon$ -окрестностей всех граней  $P$ , коразмерности которых не менее 2 (т.е. вершины, ребра и т.д., до граней размерности  $n-2$ ). Внутри каждой такой  $\varepsilon$ -окрестности бильярдная траектория  $\tilde{\gamma}$  отражается от угла меньшей, чем  $n$ , размерности. Поэтому число отражений частицы  $x$ , пока она находится в одной из  $\varepsilon$ -окрестностей, равномерно ограничено по предположению индукции. Уберем объединение всех этих  $\varepsilon$ -окрестностей; тогда объединение оставшихся частей граней  $P$  коразмерности 1 будет несвязно (оно состоит из объединения конечного числа кусков  $(n-1)$ -мерных граней), так что расстояние между любыми двумя такими частями отделено снизу от нуля некоторой константой  $C$ . Константа  $C$  зависит только от многогранника  $P$  и числа  $\varepsilon$ . Следовательно, частица  $\tilde{x}$  может побывать в каждой из  $\varepsilon$ -окрестностей не более  $\lceil \pi/C \rceil$  раз, поскольку общая длина траектории  $\tilde{\gamma}$  равна  $\pi$ .

Следовательно, частица  $x$  отразится от граней многогранника  $P$  конечное число раз, не превосходящее некоторой константы  $C_1$ , зависящей только от геометрии сферического многогранника  $P$ . Значит, и число отражений частицы  $\tilde{x}$  в многогранном угле  $Q$  также не превосходит  $C_1$ , причем  $C_1$  зависит только от геометрии угла  $Q$ . Теорема 1 доказана.

**ЗАМЕЧАНИЕ.** В работе Я. Г. Синая [10] доказательство ведется индукцией по размерности и от противного — явно выписывается некоторая константа, предполагается, что число ударов может ее превзойти, а затем получается противоречие.

В доказательстве М. Б. Севрюка в качестве функции, которая фигурирует в его лемме (см. II, 7.1), берется скалярное произведение скорости частицы с фиксированным вектором внутри угла  $Q$  (рис. 14).

Пусть  $O$  — вершина угла  $Q$ ,  $\vec{OA} = v^-$  и  $\vec{OB} = v^+$ ,  $F$  — фиксированная точка внутри угла. Так как  $\angle AOP = \angle BOP$  ( $v^+$  получается из  $v^-$  отражением относительно плоскости  $\Pi$ ), то  $\angle AOF > \angle BOF$ , поэтому  $f(B) = (\vec{OF}, \vec{OB}) = |\vec{OF}| \cdot |\vec{OB}| \cos \angle BOF > |\vec{OF}| \cdot |\vec{OA}| \cos \angle AOF = (\vec{OF}, \vec{OA}) = f(A)$ , и линейная функция  $f$  возрастает после каждого отражения точки относительно грани угла  $Q$ . В силу леммы Севрюка, такое возрастание может происходить лишь



**Рис. 14.** Функция  $f(x) = (\overrightarrow{OX}, \overrightarrow{OF})$  возрастает при отражении от грани:  $f(B) > f(A)$

конечное число раз, равномерно (по начальным условиям) ограниченное некоторой константой  $C_1$ , зависящей только от  $Q$ .

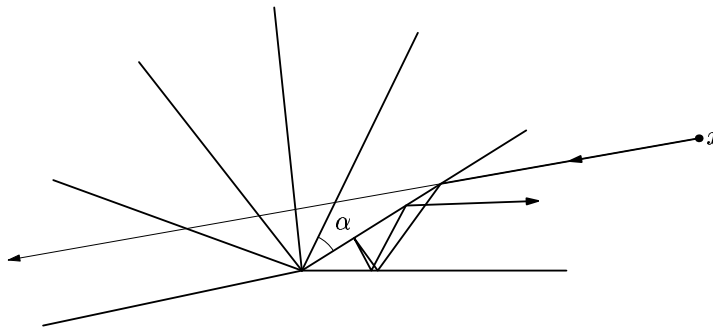
## 2. ЧАСТИЦЫ НА ПРЯМОЙ

**ТЕОРЕМА 2.** (a) Максимальное число ударов в системе трех частиц  $m_1, m_2, m_3$  задается формулой (11).

(b) Число ударов в системе частиц с произвольными массами конечно.

(c) Число ударов в системе частиц с произвольными массами равномерно ограничено константой  $C_1(m_1, \dots, m_n)$  по формуле (10).

**ИДЕИ ДОКАЗАТЕЛЬСТВА.** (a) Прежде всего, максимальное число  $N(\alpha)$  отражений бильярдной частицы в угле величины  $\alpha$  на плоскости равно  $\lceil \pi/\alpha \rceil$ , если  $\pi/\alpha \notin \mathbb{Z}$ , и равно  $\pi/\alpha$ , если  $\pi/\alpha \in \mathbb{Z}$  (рис. 15). Конфигурационное пространство системы трех частиц — это двугранный угол  $Q = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 \leq x_2 \leq x_3\}$ , или, после преобразования  $A = \text{diag}(\sqrt{m_1}, \sqrt{m_2}, \sqrt{m_3})$ , двугранный угол  $\tilde{Q} = \{(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) \in \mathbb{R}^3 \mid \frac{\tilde{x}_1}{\sqrt{m_1}} \leq \frac{\tilde{x}_2}{\sqrt{m_2}} \leq \frac{\tilde{x}_3}{\sqrt{m_3}}\}$ , в котором преобразованная траектория  $\tilde{\gamma}$  — бильярдная. Линейный угол  $\alpha$  этого двугранного угла равен углу между единичными нормальными  $\mathbf{n}_1 = (1/\sqrt{m_1}, -1/\sqrt{m_2}, 0)$  и



**Рис. 15.** Выпрямление бильярдной траектории в угле  $\alpha$  и неравенство  $N(\alpha) < \frac{\pi}{\alpha} + 1$

$\mathbf{n}_2 = (0, -1/\sqrt{m_2}, -1/\sqrt{m_3})$  к его граням (см. рис. 11). Поэтому

$$\alpha = \arccos \frac{(\mathbf{n}_1, \mathbf{n}_2)}{|\mathbf{n}_1| \cdot |\mathbf{n}_2|} = \arccos \frac{1/m_2}{\sqrt{\frac{1}{m_1} + \frac{1}{m_2}} \sqrt{\frac{1}{m_2} + \frac{1}{m_3}}}.$$

Ортогональная проекция билиардной траектории  $\bar{\gamma}$  на плоскость, ортогональную граням угла  $Q$ , также билиардная в плоском угле  $\alpha$ . Следовательно, максимальное число ее отражений равно  $\lceil \pi/\alpha \rceil = C_1(m_1, m_2, m_3)$ , где константа справа взята из формулы (11). Если  $m_1 = m_2 = m_3 = m$ , получаем  $C_1(m_1, m_2, m_3) = 3$ , как и должно быть для равных масс.

(b) Конечность числа ударов в системе частиц (без доказательства равномерной оценки) докажем от противного, индукцией по числу частиц. Для  $n = 3$  утверждение доказано в пункте (a).

Предположим, что число ударов в системе с  $n \geq 4$  частицами бесконечно. Допустим, что какие-то частицы  $i$  и  $i + 1$  сталкиваются лишь конечное число раз; тогда с какого-то момента  $t_0$  столкновения между ними прекращаются, и вся система разбивается на две невзаимодействующие подсистемы: I = частицы  $(1, 2, \dots, i)$  и II = частицы  $(i + 1, i + 2, \dots, n)$ . По предположению индукции число столкновений в каждой из подсистем I и II конечно, так что и во всей системе  $I \cup II$  из  $n$  частиц число столкновений конечно. Получили противоречие с предположением.

Следовательно, *любые две частицы  $i$  и  $i + 1$  в исходной системе обязаны сталкиваться бесконечно много раз между собой*. Однако учтем, что каждая частица получает от частицы слева положительный импульс, а от частицы справа — отрицательный. Самая правая частица системы  $\#n$  получает импульс только слева, поэтому ее импульс  $p_n(t) = m_n v_n(t)$  увеличивается после каждого удара. Однако скорость частицы  $v_n(t)$  ограничена сверху (это следует из  $m_n v_n^2 \leq 2E$ , где  $E$  — энергия всей системы). Поэтому функция  $p_n(t)$  возрастающая и ограниченная. Следовательно, по теореме Вейерштрасса существует предел  $p_n = \lim_{t \rightarrow \infty} p_n(t)$ .

Если рассмотреть систему частиц  $(\#(n-1), \#n)$ , то примерно тем же способом можно показать, что импульс ее центра масс — возрастающая и ограниченная по времени функция, а потому у нее также существует предел. Значит, существует предел и у функции  $p_{n-1}(t) = m_{n-1} v_{n-1}(t)$  — импульса частицы  $\#(n-1)$ :  $p_{n-1} = \lim_{t \rightarrow \infty} p_{n-1}(t)$ . Точно так же, рассматривая по очереди подсистемы частиц  $(\#k, \#(k+1), \#(k+2), \dots, \#(n-1), \#n)$ , где  $k = n-2, n-3, \dots, 1$ , можно доказать существование всех пределов  $p_k = \lim_{t \rightarrow \infty} p_k(t) = \lim_{t \rightarrow \infty} m_k v_k(t)$ .

В силу того, что между любыми двумя соседними частицами происходит бесконечно много ударов, нетрудно доказать, что предельные скорости всех частиц должны быть равными:  $\lim_{t \rightarrow \infty} v_1(t) = \dots = \lim_{t \rightarrow \infty} v_n(t) = v$ . Значит, скорость центра масс всей системы,  $(m_1 v_1(0) + \dots + m_n v_n(0)) / (m_1 + \dots + m_n)$  должна равняться  $v$ : *скорость каждой частицы стремится к скорости центра масс всей системы*.

Перейдем в инерциальную систему координат центра масс; в ней скорость центра масс равна нулю, так что скорость каждой частицы стремится к нулю в системе центра масс. Рассмотрим энергию всей системы — она также стремится к нулю! В силу закона сохранения энергии, начальная энергия также равна

нулю; но каждое ее слагаемое  $m_i v_i^2(0)/2$  неотрицательно; поэтому  $v_i(0) = 0$  для каждого  $i = 1, \dots, n$ . Мы получили, что в системе частиц вообще не было ударов!

Утверждение об отсутствии ударов мы получили, исходя из предположения о бесконечности числа ударов. Полученное противоречие доказывает, что число ударов конечно.

(с) Здесь мы опишем только шаги доказательства, опуская детали (подробности см. в [2, 3]).

Перейдем в систему центра масс, который поместим в начало координат, т. е. будем считать, что

$$m_1 x_1 + \dots + m_n x_n = 0, \quad m_1 v_1 + \dots + m_n v_n = 0,$$

Закон сохранения энергии запишется в виде  $\sum_{i=1}^n m_i v_i^2 = 2E$ , где  $E$  — полная энергия. Снова обозначим через  $p_i(t) = m_i v_i(t)$  импульс частицы  $i$ .

При  $k$ -м столкновении двух частиц импульс левой частицы уменьшается на величину  $\Delta p_k > 0$ , а правой — увеличивается на  $\Delta p_k$ . Число ударов в системе обозначим через  $K$  (оно конечно в силу (b)), так что  $1 \leq k \leq K$ .

ЭТАП 1.  $\Delta p_1 + \Delta p_1 + \dots + \Delta p_K \leq 2P_{\max}$ , где  $P_{\max} = \max_{0 \leq t < \infty} \mathcal{P}(t)$ , и

$$\mathcal{P}(t) = 1 \cdot p_1(t) + 2 \cdot p_2(t) + \dots + n \cdot p_n(t) = \sum_{i=1}^n i \cdot p_i(t).$$

Легко показывается, что  $\Delta \mathcal{P}_k \stackrel{\text{def}}{=} \mathcal{P}(t_k + 0) - \mathcal{P}(t_k - 0) = \Delta p_k$ , где  $t_k$  — момент  $k$ -го удара. Поэтому  $\sum \Delta p_k \leq P_{\max} - P_{\min}$ . Используя правило множителей Лагранжа, можно доказать, что

$$P_{\max} = \sqrt{2E \left( M_2 - \frac{M_1^2}{M} \right)}, \quad (24)$$

где  $M = \sum_{i=1}^n m_i$ ,  $M_1 = \sum_{i=1}^n i \cdot m_i$ ,  $M_2 = \sum_{i=1}^n i^2 \cdot m_i$ .

ЭТАП 2. На этом этапе вся ось времени разбивается на «быстрые промежутки», в каждом из которых какие-то два соседних шара ни разу не сталкиваются и имеют большую относительную скорость:

$$|v_i(t) - v_{i-1}(t)| \geq c_i$$

для всех  $t$  из соответствующего промежутка  $[t_1, t_2]$ .

Доказывается, что существует универсальный набор постоянных  $c_1, \dots, c_n$  (зависящих только от масс частиц и полной энергии системы), для которых вся ось времени покрывается быстрыми промежутками по всем парам  $(i-1, i)$ .

ЭТАП 3. Число быстрых промежутков оценивается сверху числом

$$K = 8 \left( \sum_{i=1}^n \frac{1}{m_i} \right) \cdot \sqrt{M_2 M - M_1^2}.$$

ЭТАП 4. На этом этапе число столкновений оценивается неравенством

$$C_1(m_1, \dots, m_n) \leq 2(2K)^{n-2},$$

где  $K$  — константа из этапа 3, после чего получается окончательная оценка (10).

## 3. «Большая» задача Синая и задача о шарах; часть (I)

ТЕОРЕМА 3. (а) Число отражений биллиардной траектории  $\gamma$  в многомерном многогранном угле  $Q$  с вдавленными внутрь стенками («криволинейном угле») конечно на всем интервале времени  $(0, +\infty)$ .

(б) Число ударов в системе  $N$  произвольных твердых шаров в пространстве  $\mathbb{R}^n$  конечно.

ИДЕИ ДОКАЗАТЕЛЬСТВА. (а1) Прежде всего, и это самый важный момент в доказательстве, мы покажем, что отражения не могут накапливаться на конечном интервале времени, или, иными словами, множество отражений от стенок угла дискретно. Мы следуем доказательству [2].

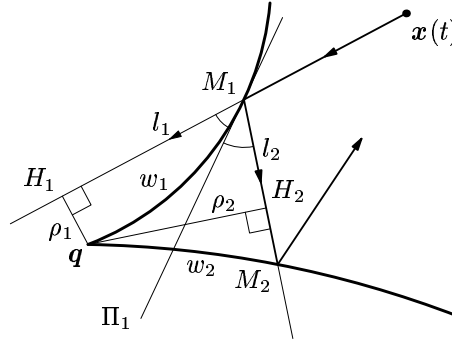


Рис. 16. Расстояние от вершины угла  $q$  до прямой  $l_1$  меньше расстояния от  $q$  до прямой  $l_2$ :  $\rho_1 < \rho_2$

Действительно, пусть моменты ударов накапливаются, тогда и точки, в которых происходят отражения в эти моменты, также накапливаются. Пусть  $q$  — точка накопления ударов; назовем ее вершиной угла (рис. 16). (Заметим, что многогранный угол может иметь много вершин — общих точек стенок; например, две трансверсальные стенки в  $\mathbb{R}^3$  имеют целую кривую вершин. На рисунке для наглядности изображена двумерная картинка криволинейного угла с одной вершиной.)

Поэтому для любого  $\varepsilon > 0$  найдется такой момент  $t$ , что  $\text{dist}(x(t), q) < \varepsilon$ . Мы сейчас получим противоречие с этим неравенством, доказав, что  $\text{dist}(x(t), q)$  всегда больше некоторой константы  $r$  (зависящей от начального движения точки).

Пусть частица  $x$  летела вдоль луча  $l_1$  до своего первого столкновения со стенкой  $w_1$  в точке  $M_1$ . Затем она отразилась от стенки  $w_1$  и полетела по лучу  $l_2$  к стенке  $w_2$ . Вектор скорости движения точки  $x$  отразился симметрично относительно касательной плоскости  $\Pi_1$  к стенке  $w_1$  в точке  $M_1$ , и часть луча  $l_1$  после точки  $M_1$  отразилась в «зеркале»  $\Pi_1$  и перешла в луч  $l_2$ . Заметим, что вся стенка  $w_1$  — граница выпуклого тела. Следовательно, перпендикуляр  $qH_1$ , опущенный из  $q$  на  $l_1$ , короче перпендикуляра  $qH_2$ , опущенного из  $q$  на  $l_2$  (ведь отрезок  $qH_2$  пересекает «зеркало»  $\Pi_1$ ):

$$|qH_1| < |qH_2|, \text{ или } \rho_1 < \rho_2. \quad (25)$$

Точно так же, после второго отражения — в точке  $M_2$  от стенки  $w_2$  — часть луча  $l_2$  перейдет в симметричный относительно плоскости  $\Pi_2$  луч  $l_3$ , и мы будем иметь по той же причине  $\rho_2 < \rho_3$ , и т. д.: расстояние от вершины  $q$  до лучей движения бильярдной частицы  $x(t)$  после каждого отражения строго возрастает:  $r = \rho_1 < \rho_2 < \rho_3 < \dots$ . Поскольку наклонная длиннее перпендикуляра,  $\text{dist}(x(t), q) > \rho_1$ . Следовательно, при  $\varepsilon < \rho_1$  мы получаем противоречие. Отметим, что при доказательстве мы использовали тот факт, что точка  $x$  приближается к  $q$ .

Итак, мы доказали, что на любом конечном интервале времени число отражений конечно. (Более того, можно показать (см. [2, с. 168–169]), что  $x(t)$  будет приближаться к вершине  $q$  время, меньшее  $\rho_1$ .) Как мы отмечали выше (раздел II.3), именно этот факт играет ключевую роль в доказательстве эргодичности системы молекул газа.

(a2) Интересно, что на плоскости даже для нулевого угла можно дать верхнюю оценку числа отражений (зависящую от начального положения точки  $x$ ), см. рис. 17.

**УТВЕРЖДЕНИЕ 1.** Пусть  $\varphi$  — угол первого отражения бильярдной точки  $x(t)$  от стороны криволинейного угла с вершиной  $q$  и  $\alpha$  — угол между касательными к сторонам угла в точках  $A$  и  $B$  пересечения окружности радиуса  $r = \rho_1$  с центром  $q$  (см. рис. 17). Тогда число отражений частицы от сторон угла

$$N(\varphi, \alpha) \leq \frac{2\varphi}{\alpha}. \quad (26)$$

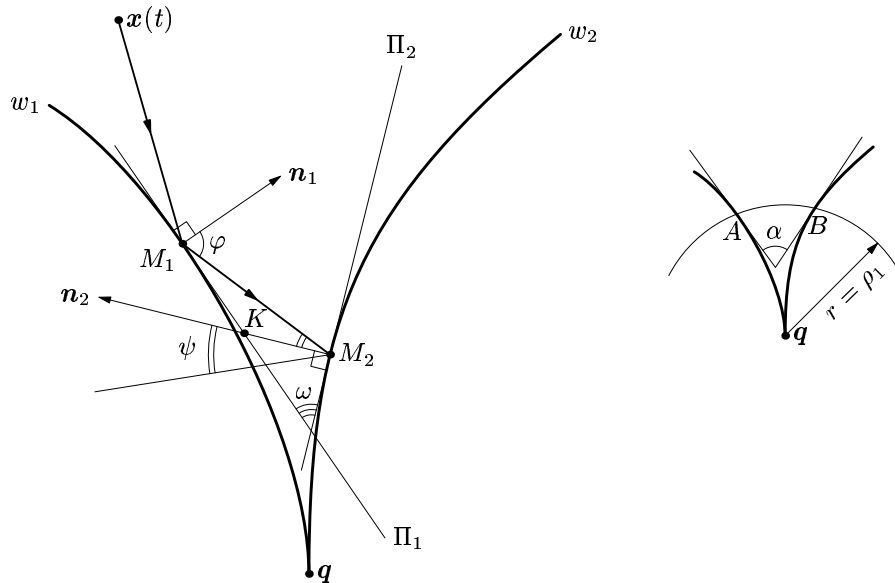


Рис. 17.  $\psi = \varphi - \omega$ ,  $\omega \geq \alpha$

СЛЕДСТВИЕ. Если криволинейный угол — ненулевой и вмещает в себя угол  $\alpha$ , то наибольшее число отражений от сторон этого угла  $\leq \frac{\pi}{\alpha} + 1$ .

Из рис. 17 видно, что два внутренних угла  $\triangle M_1 K M_2$  равны  $\frac{\pi}{2} - \varphi$  и  $\psi$ , а внешний угол в вершине  $K$  равен  $\frac{\pi}{2} - \omega$  ( $\omega$  — угол между касательными  $\Pi_1$  и  $\Pi_2$  в точках  $M_1$  и  $M_2$ ). Следовательно,  $\frac{\pi}{2} - \omega = \left(\frac{\pi}{2} - \varphi\right) + \psi$ , откуда  $\psi = \varphi - \omega$ . Но  $\omega \geq \alpha$ , поэтому  $\psi \leq \varphi - \alpha$ .

Значит, угол отражения каждый раз уменьшается по крайней мере на  $\alpha$ :

$$\psi_1 \leq \varphi - \alpha; \psi_2 \leq \psi_1 - \alpha \leq \varphi - 2\alpha; \dots; \psi_n \leq \varphi - n\alpha; \dots$$

Отсюда, поскольку  $0 \leq \psi_n \leq \psi_{n-1} \leq \dots \leq \psi_1$  при приближении точки  $x$  к вершине  $q$ , получаем, что  $n\alpha \leq \varphi$  и  $n \leq \varphi/\alpha$ . Итак, частица отражается не более  $\varphi/\alpha$  раз при ее приближении к вершине  $q$ .

Затем она удаляется от  $q$ . Можно считать, в силу обратимости движения во времени, что при удалении от  $q$  точка  $x$  отразится от сторон  $w_1$  и  $w_2$  не более того числа раз, которое она отразилась при приближении к  $q$ . Поэтому общее число отражений  $\leq 2\varphi/\alpha$ .

Так как  $\varphi \leq \frac{\pi}{2}$ , то  $N(\varphi, \alpha) \leq \frac{\pi}{\alpha}$ , и следствие вытекает немедленно.

(а3) Нам осталось доказать, что в последовательности отражений от сторон многомерного криволинейного угла наступит последнее отражение. Мы следуем одному из пунктов доказательства в [20].

Пусть  $t_0 < t_1 < t_2 < \dots < t_k < \dots$  — моменты отражений. Рассмотрим треугольники  $\Delta_1 = x(t_0)qx(t_1)$ ,  $\Delta_2 = x(t_1)qx(t_2)$ ,  $\dots$ ,  $\Delta_k = x(t_{k-1})qx(t_k)$ ,  $\dots$ , вырежем их ножницами из бумаги и расположим на (двумерной!) плоскости так, чтобы  $\Delta_k$  и  $\Delta_{k+1}$  примыкали по общей стороне  $qx(t_k)$  и не перекрывались (рис. 18). Иными словами, сделаем плоскую развертку «гармошки» треугольников  $\{\Delta_k\}_1^\infty$ , расположенных в пространстве  $\mathbb{R}^n$  и примыкающих по общим сторонам  $qx(t_k)$ .

Оказывается, для каждого  $k \geq 1$  угол  $\angle x(t_0)qx(t_k)$  строго меньше  $\pi$ :  $\gamma_1 + \gamma_2 + \dots \leq \pi$  (рис. 18). Это вытекает из выпуклости стенок угла.

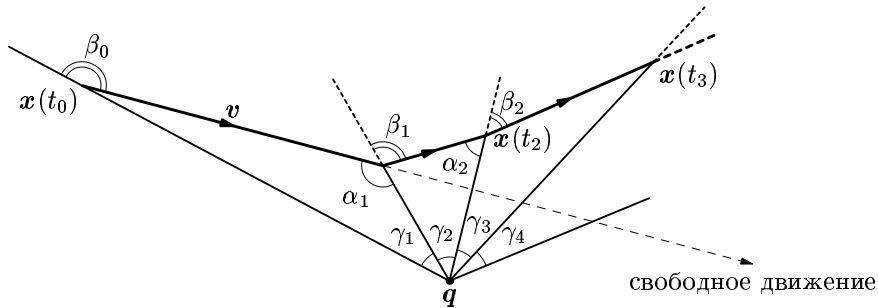
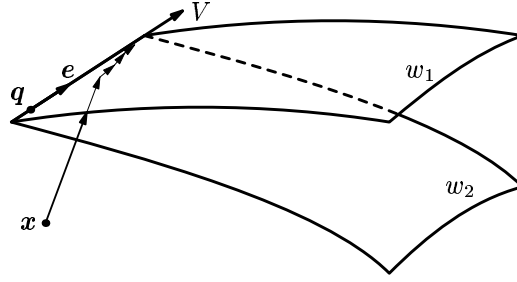


Рис. 18. Развертка цепочки треугольников на плоскость



**Рис. 19.** Луч  $V$  вдоль  $e$  лежит в пересечении всех стенок  $W = \bigcap W_i$  и частица  $x$  приближается к нему

Отсюда следует существование предельного единичного вектора

$$e = \lim_{t \rightarrow \infty} \frac{\mathbf{x}(t) - \mathbf{q}}{|\mathbf{x}(t) - \mathbf{q}|} = \lim_{t \rightarrow \infty} \frac{\overrightarrow{\mathbf{q}\mathbf{x}(t)}}{|\overrightarrow{\mathbf{q}\mathbf{x}(t)}|}, \quad (27)$$

если предположить, что число ударов бесконечно. Возьмем одну стенку угла  $W$ , частица отразится от нее бесконечно много раз, а так как  $|\overrightarrow{\mathbf{q}\mathbf{x}(t)}| \rightarrow \infty$ , мы получаем, что стенка  $W$  содержит целиком луч  $V$ , выходящий из  $\mathbf{q}$  в направлении  $e$ . Так как это верно для любой стенки угла  $Q$ , луч  $V$  лежит в пересечении  $\bigcap_{i=1}^N W_i$  всех стенок угла (рис. 19). Развертка треугольников позволяет также доказать тот факт, что функция  $d(t) = \text{dist}(\mathbf{x}(t), \bigcap W_i)$  выпукла по  $t$ . Мы хотим доказать, что существует такое  $T_0$ , что  $d(t)$  — линейная функция при всех  $t > T_0$ . Если число ударов бесконечно, то найдутся такие положительные константы  $\varepsilon > 0$  и  $c > 0$ , что

$$d(t) \geq \varepsilon t - c \text{ при всех } t.$$

Далее,  $\frac{\text{dist}(\mathbf{x}(t), V)}{|\overrightarrow{\mathbf{q}\mathbf{x}(t)}|} \rightarrow 0$  при  $t \rightarrow \infty$ . Однако

$$\begin{aligned} \text{dist}(\mathbf{x}(t), V) &\geq \text{dist}(\mathbf{x}(t), \bigcap W_i) = d(t) \geq \varepsilon t - c, \\ |\overrightarrow{\mathbf{q}\mathbf{x}(t)}| &\leq |\overrightarrow{\mathbf{q}\mathbf{x}(0)}| + |\overrightarrow{\mathbf{x}(0)\mathbf{x}(t)}| \leq |\overrightarrow{\mathbf{q}\mathbf{x}(0)}| + t \end{aligned}$$

(мы заменили  $|\overrightarrow{\mathbf{x}(0)\mathbf{x}(t)}|$  на  $t$ , так как частица движется вдоль траектории от  $\mathbf{x}(0)$  до  $\mathbf{x}(t)$  со скоростью 1). Поэтому

$$\frac{\text{dist}(\mathbf{x}(t), V)}{|\overrightarrow{\mathbf{q}\mathbf{x}(t)}|} \geq \frac{\varepsilon t - c}{|\overrightarrow{\mathbf{q}\mathbf{x}(0)}| + t} \rightarrow \varepsilon > 0 \text{ при } t \rightarrow \infty,$$

и левая часть неравенства не стремится к 0 — противоречие.

Переходим к доказательству пункта (b).

(b). Выше мы сделали редукцию системы шаров к бильярду в криволинейном угле. Поэтому утверждение о шарах есть следствие пункта (a).

Приведем еще одну идею, близкую к идее доказательства про частицы (см. подробности в [2]). Из (a1) следует, что в любом *конечном* интервале времени число ударов в системе шаров конечно. Как и при доказательстве теоремы 2(b), закончить доказательство про систему шаров можно так. Если на интервале  $(0, +\infty)$



число ударов в системе бесконечно, то вся система с какого-то момента  $T_0$  разбивается на группы (кластеры), в каждой из которых любой шар испытывает бесконечное число ударов от шаров того же кластера, а сами кластеры никак не взаимодействуют друг с другом: их центры масс движутся в пространстве равномерно и прямолинейно. Рассмотрим один кластер и перейдем в систему его центра масс.

Как и в случае частиц, можно доказать, что скорости всех шаров стремятся к скорости центра масс, т.е. к 0. А тогда энергия  $E(t)$  кластера также стремится к 0. Но она постоянна во времени при  $t > T_0$ , так что  $E(t) \equiv 0$ . Значит, шары в кластере вообще не двигались с момента  $T_0$ . Получили противоречие с бесконечностью числа ударов в кластере.

Следовательно, число ударов в системе может быть только конечным, что и требовалось доказать.

#### 4. «Большая» задача Синая и задача о шарах; часть (II)

**ТЕОРЕМА 4.** (а) Число отражений биллиардной траектории  $\gamma$  в многомерном криволинейном многогранном угле  $Q$  с вдавленными внутрь стенками, на который наложено условие невырожденности (см. II.7.4 и рис. 9), не превосходит некоторой константы  $C_2$ , зависящей только от геометрии угла  $Q$ .

(б) Число ударов в системе  $N$  шаров в  $\mathbb{R}^n$  с массами  $m_1, \dots, m_N$  не превосходит константы  $C_2(m_1, \dots, m_N)$ , определяемой равенством (13), какими бы ни были начальные положения и скорости шаров, а также их радиусы  $r_1, \dots, r_N$ .

**ИДЕИ ДОКАЗАТЕЛЬСТВ.** (а) Доказательство существования равномерной по начальным условиям константы  $C_2$  ведется индукцией по числу стенок в угле  $Q$ . Оно основано на леммах, сформулированных ниже.

Сначала определение. Обозначим конечный кусок биллиардной траектории, выходящей из точки  $X$  и заканчивающейся в точке  $Y$ , через  $\Gamma(X, Y)$ . Назовем кодом этого куска последовательность стенок  $w_{i_1}, \dots, w_{i_k}$ , от которых происходили отражения траектории  $\Gamma(X, Y)$ . Мы скажем, что другая ломаная  $\gamma(t)$ ,  $0 \leq t \leq 1$ , начинающаяся в точке  $X = \gamma(0)$  и кончающаяся в точке  $Y = \gamma(1)$ , имеет тот же самый комбинаторный тип, что и биллиардная ломаная  $\Gamma(X, Y)$ , если она отражается от тех же стенок в том же порядке, т.е. существуют моменты  $0 \leq t_1 \leq t_2 \leq \dots \leq t_k \leq 1$ , в которые  $\gamma(t_j) \in W_j$ .

**ВАЖНОЕ ЗАМЕЧАНИЕ:** указанные неравенства — нестрогие, так что вместо отражения от стенки  $W_i$ , а потом от  $W_j$ , ломаная может попасть сразу на пересечение  $W_i \cap W_j$ . В частности, если взять какую-нибудь вершину угла  $q \in \bigcap_{i=1}^N W_i$  (т.е. одну из точек на пересечении всех стенок угла), то двузвенная ломаная  $XqY$  имеет тот же комбинаторный тип, что и траектория  $\Gamma(X, Y)$ .

**ЛЕММА 1 (COMPARISON LEMMA).** Биллиардная траектория  $\Gamma(X, Y)$  короче любой ломаной  $\gamma(t)$  того же комбинаторного типа в угле  $Q$  с выпуклыми внутрь стенками.

Напомним, как обычно доказывается эта лемма для многогранного угла с плоскими стенками. См. рис. 15: мы выпрямляем траекторию, отражая угол  $Q$  относительно граней отражения траектории, и получаем отрезок прямой между

начальной и конечной точками. А он, конечно же, короче любой ломаной с теми же концами. Для криволинейного угла идея та же: *выпрямление траектории*. Только осуществить ее следует чуть иначе.

Возьмем столько копий криволинейного угла  $Q$ , сколько имеется звеньев у траектории  $\Gamma(X, Y)$  (т. е. на 2 больше числа точек излома траектории), и начнем их приклеивать одну к другой последовательно по стенкам  $W_{ij}$ , от которых происходили отражения, следуя коду ломаной  $\Gamma(X, Y)$ . Тогда на развертке или «гармошке» копий  $Q_1, Q_2, \dots$ , полученной в результате такого склеивания (в «высокой» математике «гармошка» называется обобщенным римановым, или александровским, пространством), траектория  $\Gamma(X, Y)$  «распрямляется» и будет *кратчайшей* (геодезической) кривой между двумя фиксированными точками —  $X$  (в первой копии гармошки) и  $Y$  (в последней копии). Отсюда и следует лемма 1: любая другая ломаная длиннее геодезической (см. подробности в [12]).

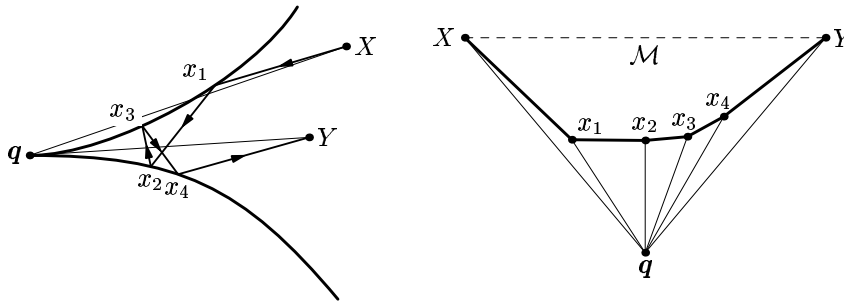
Приведем удивительное, на первый взгляд, утверждение.

**ЛЕММА 2.** Пусть  $q$  — вершина криволинейного угла  $Q$  ( $q \in \bigcap_{i=1}^N W_i$ ), а  $\Gamma(X, Y)$  — конечный кусок бильярдной траектории общей длины  $|\Gamma(X, Y)|$ . Тогда  $|qX| + |qY| > |\Gamma(X, Y)|$  (рис. 20).

Как и на последнем этапе доказательства «большой» теоремы Синяя, вырежем из бумаги треугольники  $\triangle qXx_1, \triangle qx_1x_2, \triangle qx_2x_3, \dots, \triangle qx_kY$  и склеим их по общим сторонам без перекрытий. Получим плоскую развертку «гармошки» тех же треугольников, расположенных в пространстве  $\mathbb{R}^n$  (рис. 18).

В силу выпуклости стенок  $W_i$  и закона «угол падения равен углу отражения» легко показать, что ломаная  $\gamma(XY) = Xx_1x_2x_3 \dots x_kY$  на плоскости *выпукла*: любая прямая  $Xx_1, x_1x_2, \dots, x_kY$  делит плоскость на две полуплоскости, в одной из которых находится вся ломаная  $\gamma(X, Y)$ , а в другой — точка  $q$ . При этом длина плоской ломаной  $\gamma(X, Y)$  равна длине пространственной бильярдной траектории:  $|\gamma(X, Y)| = |\Gamma(X, Y)|$ . (Для угла с плоскими стенками ломаная  $\gamma(X, Y)$  представляет собой отрезок прямой).

Соединим отрезком точки  $X$  и  $Y$  ломаной  $\gamma(X, Y)$  (фигура справа на рис. 20). Тогда многоугольник  $M = Xx_1 \dots x_kY$  — выпуклый, и он находится внутри треугольника  $\triangle XqY$ , который также выпуклая фигура на плоскости. По известной



**Рис. 20.** Доказательство леммы 2: периметр многоугольника  $M$  меньше периметра  $\triangle XqY$

теореме элементарной геометрии о вложенных выпуклых многоугольниках заключаем, что периметр  $\triangle XqY$  больше периметра  $M$ . Вычитая длину отрезка  $XY$ , получаем  $|qX| + |qY| > |\gamma(X, Y)|$ . Лемма 2 доказана.

ЛЕММА 3 (ТЕХНИЧЕСКАЯ ЛЕММА). *Для любых четырех точек  $A, B, C, D \in \mathbb{R}^n$ , если  $\frac{\text{dist}(C, AB)}{|CD|} \geq c > 0$ , то*

$$\frac{(|AD| + |DB|) - (|AC| + |CB|)}{|AC| + |CB| - |AB|} \leq K = K(c), \text{ где } K(c) = 16 \left(\frac{1}{c} + 1\right)^2.$$

Доказательство леммы 3 основывается на рассмотрении различных случаев расположения точек в  $\mathbb{R}^n$  (см. [12]) и мы его опускаем.

Доказательство теоремы 4(a) основывается на так называемой «процедуре укорачивания траектории  $\Gamma$ » с привлечением лемм 1, 2 и 3. Она состоит из переменной замены куска бильярдной траектории  $\Gamma(X_1, X_2)$ , где  $X_1$  и  $X_2$  — точки отражения от одной и той же стенки  $W_i$  угла  $Q$ , на

(1) двузвенную ломаную — объединение двух отрезков  $X_1D \cup DX_2$ , где  $D$  — точка на пересечении всех стенок  $\bigcap_{i=1}^N W_i$ , ближайшая к заранее выбранной точке  $Z$  траектории  $\Gamma$ ; или

(2) отрезок  $X_1X_2$ .

При этом точка  $Z$  траектории  $\Gamma$  предшествует точке  $X_2$  и следует за точкой  $X_1$  на  $\Gamma$ , и выполнено соотношение

$$\frac{\text{dist}(Z, W_i)}{\text{dist}(Z, \bigcap_{i=1}^N W_i)} \geq C, \quad (28)$$

где константа  $C$  взята из определения невырожденного угла  $Q$  (см. условие  $d/C$ -близости и рис. 9).

(1) Замену  $\Gamma(X_1, X_2)$  на  $X_1D \cup DX_2$  назовем «растяжением» траектории  $\Gamma$  (в силу леммы 2) и обозначим

$$(|X_1D| + |DX_2|) - (|X_1Z| + |ZX_1|) = s;$$

$s$  зависит от точек  $X_1, X_2$  и  $Z$ .

(2) Замену  $\Gamma(X_1, X_2)$  на  $X_1X_2$  назовем «сжатием» траектории и обозначим

$$|X_1Z| + |ZX_2| - |X_1X_2| = p.$$

Тогда по лемме 3

$$s \leq K(c) \cdot p, \quad K(c) = 16 \left(\frac{1}{c} + 1\right)^2. \quad (29)$$

ЛЕММА 4. *Число попарно различных конечных кусков траекторий  $\Gamma(A, B)$  бильярдной траектории  $\Gamma$ , отражающихся от всех  $N$  стенок угла  $Q$  и не имеющих общих точек (кроме, возможно, общих концов), не превосходит  $L = 2K(c) + 4$ , где  $K(c)$  — константа из леммы 3. Такие куски авторы [12] называют циклами.*

Доказательство леммы 4 авторы получают от противного — противоречием с леммами 1 и 2.

Дадим теперь краткое доказательство теоремы 4(а). Пусть при всех  $k < N$  любая траектория, отражающаяся от  $\leq k$  стенок  $W_i$ , имеет  $\leq M_k$  отражений в угле  $Q$ . По предположению индукции, каждый кусок траектории, содержащий  $\leq M_{N-1}$  отражений, содержит цикл. По лемме 4 число циклов  $\leq L$ ; следовательно всего отражений в угле  $Q$  может быть  $\leq M_N$ , где  $M_N = L(M_{N-1} + 1) \leq 2LM_{N-1}$ . Поскольку  $M_1 = 1$  (одно отражение от стенки), получаем

$$\begin{aligned} M_N &\leq (2L)^{N-1} = (2(2K(C) + 4))^{N-1} = (4K(C) + 8)^{N-1} = \\ &= \left(4 \cdot 16 \left(\frac{1}{C} + 1\right)^2 + 8\right)^{N-1} \leq \left[64 \left(\frac{1}{C} + 2\right)^2\right]^{N-1} = \left[8 \left(\frac{1}{C} + 2\right)\right]^{2(N-1)}. \end{aligned} \quad (30)$$

Таким образом, константу  $C_2$  в теореме 4 можно положить равной

$$C_2 = \left[8 \left(\frac{1}{C} + 2\right)\right]^{2(N-1)}, \quad (31)$$

где  $C$  — константа в условии невырожденности угла  $Q$  и  $N$  — число стенок угла.

(б) Для доказательства существования равномерной оценки для числа ударов  $N$  шаров надо прежде всего проверить условия невырожденности бильярда, образованного выкидыванием из  $\mathbb{R}^{nN}$   $\frac{N(N-1)}{2}$  цилиндров  $C_{ij}$  (см. формулы (17) и (18)). Положим  $m_{\min} = 1$ ,  $r_{\min} = 1$ .

Идея состоит в том, чтобы следить за движениями шаров в  $\mathbb{R}^n$ , если двигать конфигурационную точку  $\mathbf{x}$  к каждой из стенок  $W_{ij} = C_{ij}$  специальным образом, а также если двигать  $\mathbf{x}$  к пересечению всех стенок специальным образом. Оказывается, расстояние от  $\mathbf{x} \in \mathbb{R}^{nN}$  до  $C_{ij}$  в метрике  $\|\{x_{ij}\}\|^2 = \frac{1}{2E} \sum_{i,j} m_i x_{ij}^2 = \frac{1}{2E} \sum_{i,j} m_i |c_i|^2$ , где  $c_i = \{x_{ij}\} \in \mathbb{R}^n$  — центры шаров,  $E$  — энергия системы, не превосходит некоторой константы (зависящей от энергии  $E$  системы шаров), умноженной на расстояние между шарами  $i$  и  $j$  в  $\mathbb{R}^n$ . А именно, возьмем какую-нибудь группу шаров  $I$ , все пары шаров из группы  $I$ , и соответствующие им цилиндры  $C_{ij}$ .

Пусть  $d_{ij}$  — расстояние между центрами шаров  $i$  и  $j$ ,  $r_i, r_j$  — их радиусы и  $\delta_{ij} = d_{ij} - r_i - r_j \geq 0$ . Рассмотрим максимум всех  $\delta_{ij}$  и обозначим  $\delta = \delta_{i_1 j_1} = \max_{i,j \in I} \delta_{ij}$ . Можно показать, что  $\max_{C_{ij} \in I} \text{dist}(\mathbf{x}, C_{ij}) \geq \frac{\delta}{2\sqrt{2E}}$ , и что

$$\text{dist}(\mathbf{x}, \bigcap_{i,j \in I} C_{ij}) \leq \frac{N \cdot r_{\max} \delta \sqrt{m_{\max} N}}{\sqrt{2E}}.$$

Поэтому отношение левых частей дает значение константы невырожденности для группы  $I$ :

$$C_{I, \mathbb{R}^{nN}}(\mathbf{x}) \geq \frac{1}{2r_{\max} \cdot N \sqrt{m_{\max} N}}.$$

Так как правая часть неравенства не зависит ни от  $\mathbf{x}$ , ни от группы  $I$ , получаем универсальную константу невырожденности

$$C = \frac{1}{2r_{\max} N^{3/2} m_{\max}^{1/2}}. \quad (32)$$

Подставляя ее в формулу (31), находим, что число ударов в системе шаров не превосходит

$$\left(8 \cdot 2r_{\max} N^{3/2} \sqrt{m_{\max}}\right)^2 \cdot \left(\frac{N(N-1)}{2} - 1\right) < \left(16r_{\max} \cdot N^{3/2} \sqrt{m_{\max}}\right)^2 N^2.$$

Пока что получилась оценка (12), включающая радиусы шаров (напомним, что  $m_{\min} = 1$  и  $r_{\min} = 1$ , так что, если отбросить эти условия, мы должны писать везде отношения  $\frac{m_{\max}}{m_{\min}}$  и  $\frac{r_{\max}}{r_{\min}}$ ).

Однако шары можно двигать в  $\mathbb{R}^n$  несколько хитрее. Выберем опять некоторое множество стенок  $I$ .

Возьмем точку  $\mathbf{x}_0 \in Q = \mathbb{R}^{nN} \setminus \left(\bigcup_{(ij) \in I} C_{ij}\right)$  и найдем максимальное расстояние в  $\mathbb{R}^{nN}$  от  $\mathbf{x}_0$  до стенок (цилиндров) из множества  $I$ :  $\delta = \max_{(ij) \in I} \text{dist}(\mathbf{x}_0, C_{ij})$ . Оценим теперь расстояние  $\text{dist}(\mathbf{x}_0, \bigcap_{(ij) \in I} C_{ij})$  через  $\delta$ . Начнем двигать точку  $\mathbf{x}(0)$  в  $\mathbb{R}^{nN}$  так. Обозначим  $I_0 = \{i \mid C_{ij} \in I\}$ . Возьмем какое-то  $i_1 \in I_0$  и все шары с номерами  $k \in I_0 \setminus \{i_1\}$  запустим одновременно с постоянной скоростью вдоль прямых, соединяющих их центры с центром шара  $i_1$ , пока каждая пара шаров  $(k, i_1)$  не соприкоснется (как только соприкосновение произойдет, конфигурационная точка попадет на цилиндр  $C_{i_1 k}$ ). В момент, когда шар  $\#k$  коснется шара  $\#i_1$ , шар  $\#k$  прекращает движение, а остальные еще не соприкоснувшиеся шары продолжают равномерно двигаться к шару  $\#i_1$ . Таким образом, конфигурационная точка  $\mathbf{x}(t)$  сначала попадает на один из цилиндров вида  $C_{i_1 j_1}$ , потом движется по нему и попадает на другой цилиндр группы  $I$  вида  $C_{i_1 j_2}$  (теперь она на пересечении двух цилиндров), потом движется и попадает на третий цилиндр группы  $I$  вида  $C_{i_1 j_3}$ , и т. д., пока последний шар из группы  $I$  не остановится, соприкоснувшись с шаром  $\#i_1$ . В этот момент конфигурационная точка находится в некоторой точке  $\mathbf{x}_1 \in \mathbb{R}^{nN}$ . Так как при каждом движении  $\mathbf{x}_0(t)$  происходил сдвиг на  $\leq \delta$ , а движений было  $\leq N$  ( $N$  — число шаров в системе), то  $\text{dist}(\mathbf{x}_0, \mathbf{x}_1) \leq m_{\max} \cdot N\delta$ . Напомним, что «бильярдная» метрика в  $\mathbb{R}^{nN}$  задается равенством

$$\text{dist}((x_1, \dots, x_n), (y_1, \dots, y_n)) = \sqrt{\sum_{i=1}^N m_i \cdot \text{dist}(x_i, y_i)^2},$$

откуда и появляется множитель  $m_{\max}$ .

Так как все шары двигались *равномерно* вдоль прямых в  $\mathbb{R}^n$ , направленных в одну точку — центр шара  $\#i_1$ , расстояние между каждой парой шаров уменьшалось; следовательно, для новой точки  $\mathbf{x}_1$   $\max_{(ij) \in I} \text{dist}(\mathbf{x}_1, C_{ij}) \leq \delta$ .

Теперь возьмем другой номер  $i_2 \in I_0 \setminus \{i_1\}$  и, начав с  $\mathbf{x}_1$ , тем же способом найдем новую точку  $\mathbf{x}_2 \in \mathbb{R}^{nN}$ , такую, что  $\text{dist}(\mathbf{x}_1, \mathbf{x}_2) \leq m_{\max} \cdot N\delta$ ; затем третью точку  $\mathbf{x}_3$  с тем же свойством и т. д., пока не исчерпаем всех шаров из множества  $I_0$ . По построению, последняя конфигурационная точка  $\mathbf{x}_{|I_0|}$  принадлежит пересечению  $\bigcap_{(ij) \in I} C_{ij}$  всех цилиндров из множества  $I$ . Значит,

$$\text{dist}(\mathbf{x}_0, \mathbf{x}_{|I_0|}) \leq \sum_{i=1}^{|I_0|-1} \text{dist}(\mathbf{x}_i, \mathbf{x}_{i+1}) \leq m_{\max} \cdot N^2 \delta. \quad (33)$$

Значит, выполнено свойство невырожденности с константой  $C = \frac{1}{m_{\max} N^2}$  (для этого надо поделить  $\delta$  на правую часть неравенства (33) и поменять знак неравенства на « $\geq$ », как и должно быть в определении невырожденности: см. рис. 9).

Отметим также, что бильярдный стол  $Q = \mathbb{R}^{nN} \setminus \left( \bigcup_{i \neq j} C_{ij} \right)$  состоит из большого числа криволинейных многогранных углов  $\{Q_s\}$  (с двумя, тремя, ...,  $\frac{N(N-1)}{2}$  стенками), и мы должны оценку (31), справедливую в действительности лишь в шаре определенного радиуса, внутри которого отражения происходят только от стенок одного из углов  $\{Q_s\}$ , превратить в глобальную оценку на всем столе  $Q$ . Авторы [12] находят такую глобальную оценку: число ударов в  $Q$  не превосходит

$$\left( 200 \left( \frac{1}{C} + 2 \right) \right)^{2N^4}; \quad (34)$$

подставляя  $C = \frac{1}{m_{\max} N^2}$ , получаем заведомо завышенную оценку

$$\left( 200(m_{\max} N^2 + 2) \right)^{2N^4} \leq (400N^2 m_{\max})^{2N^4}, \quad (35)$$

которая совпадает с (13), если вспомнить, что у нас  $m_{\min} = 1$ . Теорема 4 доказана.

В заключение приведем теорему о газе в сосуде.

**ТЕОРЕМА 5.** *Для любого невырожденного полурассеивающего бильярда (с выпуклыми внутрь стенками) существует такая константа  $C > 0$ , зависящая только от геометрии стола, что бильярдная частица за время  $t$  отражится от его стенок  $\leq Ct$  раз. В частности, эта же оценка справедлива для разреженной системы системы молекул (шариков) в сосуде, стенки которого плоские или выпуклы внутрь сосуда.*

Доказательство этой теоремы следует из теоремы 4.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Гальперин Г. А. Упругие столкновения частиц на прямой // УМН, 1978. Т. 33. Вып. 1. С. 211–212.
- [2] Гальперин Г. А. О системах локально взаимодействующих и отталкивающихся частиц, движущихся в пространстве // Труды Московского Математического Общества. Т. 43. М.: МГУ, 1981. С. 142–196.
- [3] Гальперин Г. А., Земляков А. Н. Математические бильярды. М.: Наука, 1990.
- [4] Гальперин Г. А., Чернов Н. И. Бильярды и хаос // Серия «Математика, кибернетика», №5. М.: Знание, 1991.
- [5] Земляков А. Н. Геометрия и арифметика столкновений // Квант, 1978. №4. С. 14–21.

- [6] *Полак Л. С.* Людвиг Больцман (1844–1906). М.: Наука, 1987.
- [7] *Северюк М. Б.* К оценке числа столкновений  $n$  упругих частиц на прямой. Институт энергетических проблем хим. физики РАН, 1992.
- [8] *Синий Я. Г.* Динамические системы с упругими отражениями. Эргодические свойства рассеивающих билиардов // УМН, 1970. Т. 25. Вып. 2. С. 141–192.
- [9] *Синий Я. Г.* Равномерная верхняя оценка для числа упругих ударов в системе одинаковых твердых шаров в пространстве. Рукопись, ~1965.
- [10] *Синий Я. Г.* Билиардные траектории в многогранном угле // УМН, 1978. Т. 33. Вып. 1. С. 229–230.
- [11] *Хованский А. Н.* Приложения цепных дробей и их обобщений к вопросам приближенного анализа. М.: Мир, 1956.
- [12] *Burago D., Kanonenko A., Ferleger S.* Uniform estimates on the number of collisions in semi-dispersing billiards // Annals of Mathematics, 1998. V. 147. P. 695–708.
- [13] *Burago D., Kanonenko A., Ferleger S.* A geometric approach to semi-dispersing billiards // Ergodic Th.& Dyn. Syst., 1998. V. 18. P. 303–319.
- [14] *Cohen E. G. D.* // Lectures in Theoretical Physics, VIII A. Univ. of Colorado Press, 1966. P. 167.
- [15] *Illner R.* Finiteness of the number of collisions in a hard sphere particle systems in all space // Transport Theory Statist. Physics, 1990. V. 19. P. 573–579.
- [16] *Murphy T. J., Cohen E. G. D.* Maximum number of collisions among identical hard spheres // J. Stat. Physics, 1993. V. 71. P. 1063–1080.
- [17] *Rehacek J.* On the ergodicity of dispersing billiards // Random&Computational Dynamics, 1995. Vol. 3(1&2). P. 35–55.
- [18] *Sandri G., Sullivan R. D., Norem P.* Collisions of three hard spheres // Physical Review Letters, 1964. V. 13. No 25. P. 743–745.
- [19] *Turston W., Sandri G.* Classical hard sphere 3-body problem // Bull. Amer. Phys. Soc., 1964. V. 9. P. 386.
- [20] *Vaserstein L. N.* On systems of particles with finite-range and/or repulsive interactions // Comm. Math. Physics, 1979. V. 69. P. 31–56.

## Средняя длина пробега в билиардных системах

Н. И. Чернов

В статье выводится формула для средней длины свободного пробега в билиардных системах и на примерах показывается, как она работает. Применяя ее к газу твердых сфер, выводятся классические формулы Больцмана.

**ВВЕДЕНИЕ.** В этой заметке мы расскажем об одной замечательной, но не очень широко известной формуле в теории билиардов, дающей среднюю длину пробега между отражениями. Если скорость билиардной частицы равна единице (как это обычно принимают), то средняя длина пробега совпадает со средним временем пробега, и эта интерпретация также важна в физике.

Итак, пусть  $Q$  — ограниченная область в  $\mathbb{R}^d$  или на торе  $\mathbb{T}^d$  с кусочно-гладкой границей  $\partial Q$ . Обозначим  $\{\Phi^t\}$  поток движения на фазовом пространстве  $M = Q \times S^{d-1}$  (как обычно,  $S^{d-1}$  означает  $(d-1)$ -мерную единичную сферу векторов скоростей). Размерность пространства  $M$  равна  $d + (d-1) = 2d - 1$ . Поток  $\{\Phi^t\}$  сохраняет меру Лиувилля на фазовом пространстве  $M$ :

$$d\mu = c_\mu dq dv, \quad (1)$$

где  $dq$  — мера Лебега в области  $Q$ ,  $dv$  — мера Лебега на сфере  $S^{d-1}$ , а  $c_\mu$  — просто нормировочная константа, обеспечивающая условие  $\mu(M) = 1$ . Для нас важно знать эту константу точно:

$$c_\mu = \frac{1}{|Q| \cdot |S^{d-1}|}, \quad (2)$$

где  $|Q|$  —  $d$ -мерный объем области  $Q$ , а  $|S^{d-1}|$  —  $(d-1)$ -мерный объем (можно назвать его «площадью поверхности») сферы  $S^{d-1}$ . Последняя величина универсальна:

$$|S^{d-1}| = \frac{2\pi^{d/2}}{\Gamma(d/2)}; \quad (3)$$

здесь  $\Gamma(x)$  означает гамма-функцию. Нам достаточно знать, что  $\Gamma(n+1) = n!$  при целых  $n$ ,  $\Gamma(x+1) = x\Gamma(x)$  для всех  $x > 0$  и  $\Gamma(1/2) = \sqrt{\pi}$ .

Билиардное отображение задается на граничной поверхности  $M_1 \subset \partial M$ :

$$M_1 = \{x = (r, v) \in M : r \in \partial Q, \langle v, n(r) \rangle > 0\},$$

где  $n(r)$  означает единичный вектор нормали к  $\partial Q$  в точке  $r$ , направленный внутрь  $Q$ ,  $v$  — вектор скорости, а  $\langle \cdot, \cdot \rangle$  — скалярное произведение. Иными словами,  $M_1$  является семейством всех векторов с началом на  $\partial Q$ , направленных внутрь  $Q$ . Это — множество всех возможных векторов скорости частицы после отражений от  $\partial Q$ . Ясно, что  $M_1$  есть  $(2d-2)$ -мерное многообразие с краем ( $2d-2 = \dim M - 1$ ). Поток  $\{\Phi^t\}$  определяет отображение последования Пуанкаре на  $M_1$ , переводящее точку  $x = (r, v) \in M_1$  в точку  $T(x) = (r + \tau v, v_1) \in M_1$ , где



$\tau = \tau(x)$  — время свободного движения от точки  $x$  до нового отражения от  $\partial Q$ , а  $v_1$  — вектор скорости, приобретенный частицей после отражения. Заметим, что именно среднее значение функции  $\tau(x)$  интересует нас в этой статье.

Отображение  $T: M_1 \rightarrow M_1$  сохраняет меру

$$d\nu = c_\nu \langle v, n(r) \rangle dr dv, \quad (4)$$

где  $dr$  — мера Лебега на многообразии  $\partial Q$ ,  $dv$  — мера Лебега на  $S^{d-1}$  (как и ранее), а  $c_\nu$  — опять нормировочный множитель. Несложный интегральный подсчет показывает, что

$$c_\nu = \frac{1}{|\partial Q| \cdot |B^{d-1}|}, \quad (5)$$

где  $|\partial Q|$  означает  $(d-1)$ -мерный объем многообразия  $\partial Q$ , а  $|B^{d-1}|$  —  $(d-1)$ -мерный объем единичного шара  $B^{d-1} \subset \mathbb{R}^{d-1}$ . Последняя величина снова универсальна и равна:

$$|B^{d-1}| = \frac{|S^{d-2}|}{d-1} = \frac{2\pi^{(d-1)/2}}{(d-1)\Gamma((d-1)/2)}. \quad (6)$$

**ГЛАВНАЯ ФОРМУЛА.** Теперь можно непосредственно перейти к вычислению среднего значения длины (или времени) свободного пробега. Прежде всего, зададимся вопросом — что означает «среднее»? Интеграл по какой-либо мере? Тогда можно определить его двояко — интегрировать уже имеющуюся функцию  $\tau(x)$  по мере  $\nu$  на  $M_1$  или определить длину пробега между двумя соседними отражениями на всем пространстве  $M$  и интегрировать ее по мере  $\mu$ . Результаты будут разные, кстати. Мы примем «физический» смысл слова «среднее» как «временное среднее», т. е. предел средней длины пробега за  $n$  отражений, когда  $n \rightarrow \infty$ :

$$\bar{\tau}(x) = \lim_{n \rightarrow \infty} \frac{\tau(x) + \tau(Tx) + \dots + \tau(T^{n-1}x)}{n}.$$

По теореме Биркгофа,  $\bar{\tau}(x)$  существует почти всюду на  $M_1$ . Если бильярд эргодичен, то величина  $\bar{\tau}(x)$  постоянна почти всюду и равна

$$\bar{\tau} = \int_{M_1} \tau(x) d\nu(x). \quad (7)$$

Если бильярд не эргодичен, то функция  $\bar{\tau}(x)$  существенно зависит от точки  $x$ , но все же ее среднее равно (7) и имеет смысл его вычислить. Этим мы и займемся. Запишем

$$\bar{\tau} = c_\nu \int_{M_1} \tau(r, v) \langle v, n(r) \rangle dr dv = c_\nu \int_{M_1} \int_0^{\tau(r, v)} \langle v, n(r) \rangle dt dr dv, \quad (8)$$

где  $t$  — пока формальная переменная, меняющаяся от 0 до  $\tau(x)$ . Очевидно,  $t$  параметризует отрезок бильярдной траектории от точки  $r$  до  $r + tv$ , т. е. между двумя отражениями. Поскольку все фазовое пространство  $M$  состоит из таких отрезков траекторий, то координаты  $r, v, t$  описывают все  $M$ . Это очень интересная и необычная параметризация, хотя  $r, v, t$  могут быть легко рассчитаны по  $q, v$  и наоборот (см. рис. 1). В частности,  $v$  одно и то же в обеих системах координат,  $q = r + tv$ , а  $t$  теоретически можно найти из условия  $q - tv \in \partial Q$  (но нам это не понадобится).

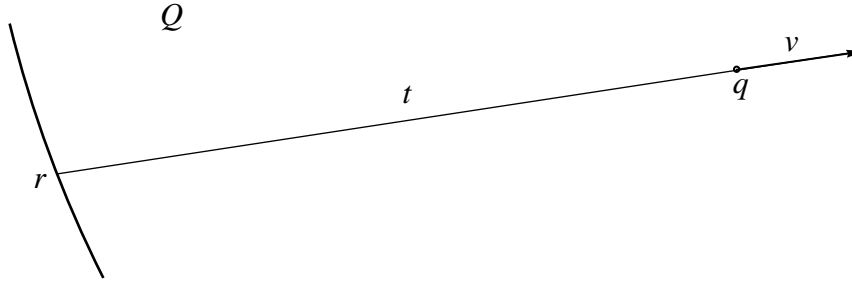


Рис. 1. Координаты  $r \in \partial Q$  и  $t$ , расстояние между  $r$  и  $q$

Главное — мера  $\mu$  может быть выражена в координатах  $r, v, t$  следующим образом:

$$d\mu = c_\mu \langle v, n(r) \rangle dr dv dt.$$

Это происходит потому, что элемент объема в  $M$  можно записать в виде

$$dq dv = \langle v, n(r) \rangle dr dv dt \quad (9)$$

и затем применить (1). Равенство (9) может показаться странным, но на самом деле оно геометрически просто. На рис. 2 изображено «доказательство» для случая плоскости ( $d = 2$ ). Мы представляем  $dq = dx dy$ , где  $x$  — координата, параллельная  $v$ , а  $y$  — ортогональная. Тогда  $dx = dt$ , а  $dy = dr \langle v, n(r) \rangle$  по определению косинуса. Аналогичный аргумент работает при  $d > 2$ , и мы оставляем его читателю.

Теперь мы легко закончим вычисление  $\bar{\tau}$  в (8):

$$\bar{\tau} = c_\nu \int_M dq dv = \frac{c_\nu}{c_\mu},$$

т. е., согласно (2) и (5),

$$\boxed{\bar{\tau} = \frac{|Q| \cdot |S^{d-1}|}{|\partial Q| \cdot |B^{d-1}|}}. \quad (10)$$

Это и есть замечательная формула для  $\bar{\tau}$ .

Отметим, что средняя величина свободного пробега  $\bar{\tau}$  не зависит от формы и характера границы  $\partial Q$ , область  $Q$  может быть даже несвязной.

Для классического бильярдного стола на плоскости имеем  $|S^1| = 2\pi$  и  $|B^1| = 2$ , и мы получаем

$$\bar{\tau} = \frac{\pi|Q|}{|\partial Q|}. \quad (11)$$

Аналогично, для пространства  $d = 3$  и

$$\bar{\tau} = \frac{4|Q|}{|\partial Q|}.$$

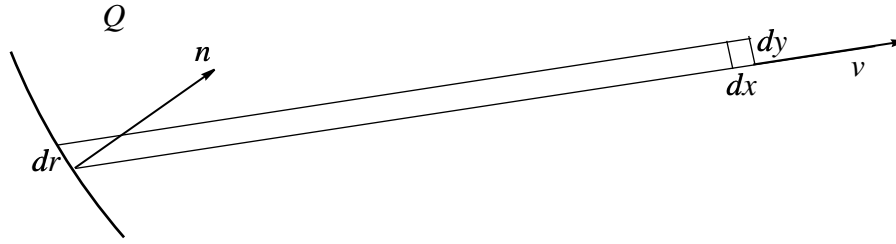


Рис. 2. Иллюстрация к соотношениям  $dx = dt$  и  $dy = dr\langle v, n(r) \rangle$

Пример. Рассмотрим билиард на плоском торе, из которого вырезан кружок радиуса  $\varepsilon > 0$ . При малых  $\varepsilon$  траектории долго «гуляют» по тору без отражений, а некоторые периодические траектории вообще не испытывают отражений. Наша формула дает среднее время свободного пробега:

$$\bar{\tau} = \frac{\pi(1 - \pi\varepsilon^2)}{2\pi\varepsilon},$$

откуда видно, что  $\bar{\tau}$  растет примерно как  $1/2\varepsilon$  при  $\varepsilon \rightarrow 0$ .

История. Формула (10), конечно, не нова. Она использовалась в некоторых работах по билиардной динамике. Но в основном она известна в интегральной геометрии и геометрической теории вероятностей. Доказательство для случая  $d = 2$  дано в книге Л. А. Сантало [4], поэтому (11) иногда называют формулой Сантало. Доказательство для всех  $d \geq 2$  можно получить из оценок в книге [3], хотя прямо эта формула там не приводится.

Автор настоящей заметки участвовал в работе научного семинара по динамическим системам под руководством В. М. Алексева и Я. Г. Синая с конца 70-х годов, и там формула (10) и ее доказательство были хорошо известны. Однако, похоже, она не была опубликована ни в одной работе по билиардам того времени. Позже некоторые западные физики «открывали» формулу (10) заново на основе численных расчетов и эвристических рассуждений. Автор данной статьи приводил формулу (10) с доказательством в двух своих работах [1, 2], из которых вторая была написана именно с целью сделать формулу (10) известной среди физиков.

ГАЗЫ ТВЕРДЫХ СФЕР. В заключение опишем одно интересное применение формулы (10) в статистической физике. Рассмотрим газ из  $N$  одинаковых  $k$ -мерных твердых шаров диаметра  $\sigma > 0$  на  $k$ -мерном торе  $\mathbb{T}_L^k$  со стороной  $L > 0$ . Величину  $n = N/L^k$  можно назвать средней плотностью газа, а

$$\rho = \frac{\sigma^k N |B^k|}{(2L)^k}$$

средней объемной плотностью (это часть объема, занимаемая шарами, так как объем одного шара равен  $(\sigma/2)^k |B^k|$ ). Мы считаем, что массы шаров равны 1, и

обозначим кинетическую энергию газа через  $EN = (v_1^2 + \dots + v_N^2)/2$ , где  $E > 0$  — средняя кинетическая энергия.

Известно, что динамика твердых шаров с упругими соударениями сводится к бильярдной задаче в области

$$Q = \mathbb{T}_L^{kN} \setminus \bigcup_{i \neq j} C_{ij},$$

где  $C_{ij} = \{\|q_i - q_j\| \leq \sigma\}$  — цилиндр, отвечающий перекрытиям  $i$ -го и  $j$ -го шара (перекрытия запрещены, поэтому цилиндры удаляются из  $\mathbb{T}_L^{kN}$ ). Запомним, что бильярдная частица в области  $Q$  движется со скоростью  $\sqrt{2EN}$ , а не с единичной, как принято.

Наша формула (10) позволяет рассчитать среднее время между последовательными отражениями от  $\partial Q$ , т. е. последовательными столкновениями шаров во всей системе. Для этого надо вычислить  $|Q|$  и  $|\partial Q|$ . Вычислить точно эти величины не представляется возможным, поскольку цилиндры  $C_{ij}$  пересекаются друг с другом весьма запутанным образом. Но можно получить приближенные значения  $|Q|$  и  $|\partial Q|$  при малых плотностях  $\rho$ , игнорируя пересечения цилиндров:

$$|Q| = L^{kN} (1 + o(1))$$

и

$$|\partial Q| = \frac{2^k k \rho (N-1)}{\sqrt{2} \sigma} L^{kN} (1 + o(1)),$$

где  $o(1)$  стремится к нулю при  $\rho \rightarrow 0$ . Детали вычислений мы оставляем читателю (см. [2]). С помощью (10) получаем

$$\bar{\tau} = \frac{\sigma(kN-1) \cdot |S^{kN-1}|}{2^k k \rho \sqrt{EN}(N-1) \cdot |S^{kN-2}|} \cdot (1 + o(1)), \quad (12)$$

где принято во внимание, что скорость движения частицы в  $Q$  равна  $\sqrt{2EN}$ .

Полученная формула правильная, но физически мало полезна. Дело в том, что газ из  $N$  шаров на торе сохраняет полный момент  $V = v_1 + \dots + v_N$ . Иначе говоря, система не эргодична — ее фазовое пространство расслаивается на инвариантные подмногообразия  $V = \text{const}$ . Ясно, что если  $\|V\|$  велико, то относительные скорости шаров малы (при фиксированной полной энергии  $2EN$ , разумеется), и столкновения происходят реже. Если  $\|V\|$  мало, то наоборот, относительные скорости велики и столкновения происходят чаще. Таким образом, формула (12) охватывает не одну, а целое семейство разных моделей с разной величиной  $\bar{\tau}$ . Наиболее интересная модель — это  $V = 0$ , где вся энергия тратится на взаимное движение шаров и нет общего «ветра», который несет весь газ в одном направлении. На физическом языке, модель  $V = 0$  находится в равновесии.

Условие  $V = 0$  определяет сечение области  $Q$  некоторой  $(kN-k)$ -мерной плоскостью, обозначим полученное сечение  $Q_0$ , в нем опять получается бильярдная задача (с отражениями от  $\partial Q_0$ ). Все такие сечения области  $Q$  параллельны и конгруэнтны, поэтому  $|Q_0|/|\partial Q_0| = |Q|/|\partial Q|$ . Но поскольку область  $Q_0$  будет  $(kN-k)$ -мерной, то формула (12) для модели  $V = 0$  принимает вид

$$\bar{\tau} = \frac{\sigma(kN-k-1) \cdot |S^{kN-k-1}|}{2^k k \rho \sqrt{EN}(N-1) \cdot |S^{kN-k-2}|} \cdot (1 + o(1)). \quad (13)$$

Эта формула дает среднее время между столкновениями во всей системе, которое, очевидно, стремится к нулю при  $N \rightarrow \infty$ . Попробуем рассчитать среднее время между последовательными столкновениями для одной типичной частицы. Поскольку в системе  $N$  частиц, и в каждом столкновении участвуют две частицы, то умножим  $\bar{\tau}$  на  $N/2$ :

$$\bar{\tau}_1 = \bar{\tau}N/2 = \frac{\sigma N(kN - k - 1) \cdot |S^{kN-k-1}|}{2^{k+1} k \rho \sqrt{EN}(N-1) \cdot |S^{kN-k-2}|} \cdot (1 + o(1)).$$

Оказывается, последняя величина имеет положительный предел при  $N \rightarrow \infty$ . Применяя (3) и простые свойства гамма-функции, нетрудно получить

$$\lim_{N \rightarrow \infty} \bar{\tau}_1 = \frac{\sqrt{\pi} \sigma}{2^{k+1} \rho \sqrt{Ek/2}} \cdot (1 + o(1)). \quad (14)$$

Интересно, что в физике средняя энергия  $E$  связана с температурой газа  $T$  стандартными соотношениями:  $E = k_B T$  в двумерном случае ( $k = 2$ ) и  $E = 3k_B T/2$  в трехмерном ( $k = 3$ ), здесь  $k_B$  — условный множитель, называемый постоянной Больцмана. Подставляя эти соотношения в (14), получим (отбрасывая член  $o(1)$ , т. е. считая плотность малой)

$$\bar{\tau}_1(k = 2, N = \infty) = \frac{1}{2\sigma n \sqrt{\pi k_B T}}$$

и

$$\bar{\tau}_1(k = 3, N = \infty) = \frac{1}{(2\sigma)^2 n \sqrt{\pi k_B T}}.$$

Последние две формулы — не что иное, как классические формулы Больцмана, выведенные эмпирически более 100 лет назад. Замечательно, что теория бильярдов позволяет дать математическое обоснование теории Больцмана.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Чернов Н. И. Новое доказательство формулы Синая для вычисления энтропии гиперболических бильярдov. Приложение к газу Лоренца и стадиону Бунимовича // Функ. анализ и его прил., 1991. Т. 25. С. 50–69.
- [2] Chernov N. I. Entropy, Lyapunov exponents and mean free path for billiards // J. Statist. Phys., 1997. V. 88. P. 1–29.
- [3] Matheron G. Random sets and integral geometry. J. New York: Wiley & Sons, 1975.
- [4] Santaló L. A. Integral geometry and geometric probability. Reading, Mass.: Addison-Wesley Publ. Co., 1976.

# Рассеяние, дефокусировка и астигматизм

Л. А. Бунимович

## ВВЕДЕНИЕ

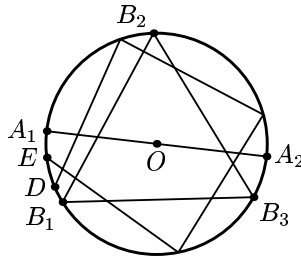
Рассмотрим движение луча света в ограниченной области  $Q$  евклидова пространства  $\mathbb{R}^d$  с кусочно гладкой границей. Предположим, что каждая гладкая компонента границы (стенка) зеркально отражает падающие на неё лучи. Это условие означает, что угол падения (угол между падающим на стенку лучом и направлением внутренней, по отношению к  $Q$ , нормали к стенке в этой точке) равен углу между этой нормалью и отражённым лучом, который называется углом отражения. (Мы предполагаем, что луч отражается во внутренней точке стенки.) Вместо луча света можно рассматривать, например, движение акустического луча или материальной точки. В последнем случае (тот же самый) закон отражения принято называть упругим (а не зеркальным). Получающаяся динамическая система называется бильярдом в  $Q$ .

Очевидно, что  $d = 1$  отвечает тривиальному случаю. Поэтому мы предположим, что размерность области  $Q$  («бильярдного стола»)  $d \geq 2$ . Уже из приведенного неформального описания модели ясно, что изучение её эволюции (динамики) имеет многочисленные приложения в оптике, акустике и классической механике. Оказывается, что изучение динамики бильярда позволяет получить важную информацию и о соответствующей квантовой задаче в области  $Q$  (см., например, [19]).

Пожалуй, наиболее естественным вопросом о поведении бильярдных орбит в области  $Q$  является следующий: будут ли орбиты бильярда всюду плотны (в области  $Q$ )? Обычно этот вопрос называется задачей об освещаемости области. (Задача об освещаемости может быть сформулирована несколькими способами в зависимости от того, какого типа источники света мы рассматриваем (см., например, [4]). В данной статье мы будем интересоваться только источниками типа фонаря, который испускает либо один луч, либо узкий пучок параллельных лучей.)

Орбиты бильярда представляют собой ломаные в области  $Q$ . Ясно, что такая ломаная не может пройти через все внутренние точки области  $Q$ . Поэтому задача об освещаемости формулируется именно как всюду плотность орбит соответствующего бильярда на бильярдном столе  $Q$ .

С другой стороны, ясно, что для многих областей легко построить такие орбиты, для которых соответствующие ломаные будут иметь только конечное число звеньев (рис. 1). Такие орбиты называются периодическими. Тем самым, задача об освещаемости должна ставиться для типичных, а не для всех орбит бильярда. (Далее мы объясним, как ввести естественным образом понятие типичности бильярдных орбит).



**Рис. 1.** Биллиард в круге.  $A_1A_2, B_1B_2B_3$  — периодические траектории

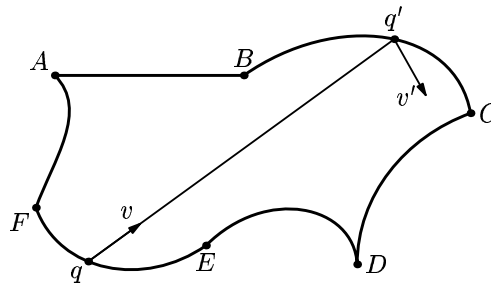
Более или менее сразу очевидно, что поведение биллиарда существенно определяется свойствами границы области  $Q$ .

Имеющиеся в настоящее время результаты относятся к случаю, когда регулярные компоненты границы области  $Q$  (стенки биллиардного стола) являются либо выпуклыми внутрь области  $Q$  (такие компоненты называются рассеивающими), либо вогнутыми внутрь области  $Q$  (такие стенки называются фокусирующими), либо  $(d - 1)$ -мерными гиперплоскостями (такие компоненты называются нейтральными), рис. 2. Случай, когда кривизна регулярной компоненты границы меняет знак, практически не исследован.

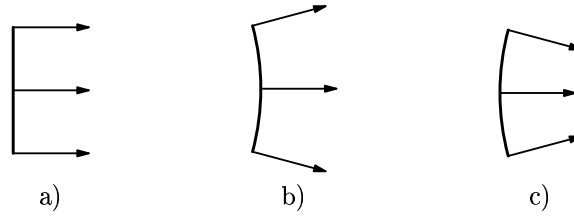
В данной статье мы будем рассматривать биллиарды в областях с хотя бы одной не нейтральной компонентой границы. (Тем самым, мы исключаем из рассмотрения многоугольники и многогранники).

Пусть пучок параллельных лучей отражается один раз от одной и той же компоненты границы  $\partial Q$ . Тогда, как легко видеть (ниже мы приведём формулы для кривизны фронтов таких отражённых пучков), возможны три ситуации: после отражения от нейтральной стенки пучок остаётся параллельным, рис. 3а), после отражения от рассеивающей стенки пучок становится расходящимся (рассеивающимся), рис. 3б), после отражения от фокусирующей стенки пучок становится сходящимся (фокусирующимся), рис. 3с).

Глядя на рис. 3, естественно предположить, что рассеивающие стенки способствуют увеличению освещённости области. Действительно, поскольку лучи в



**Рис. 2.**  $AB$  — нейтральная компонента,  $BC, FE$  — фокусирующие компоненты,  $CD, DE, FA$  — рассеивающие компоненты



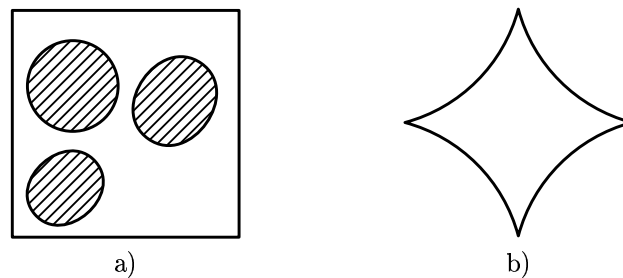
**Рис. 3.** а) параллельный; б) расходящийся (рассеивающийся) и в) сходящийся (фокусирующийся) пучки траекторий

пучке расходятся после отражения от рассеивающей стенки, этот пучок освещает со временем все более и более широкую подобласть. С другой стороны, после отражения от нейтральной стенки ширина пучка лучей не изменяется. Что же касается отражения от фокусирующей стенки, то в этом случае пучок сходится и освещает после отражения более узкую подобласть.

Поэтому наличие фокусирующей границы традиционно считалось препятствием к освещаемости. Действительно, легко видеть, например, что параллельный пучок лучей, не проходящий через центр круга, освещает только кольцо с внутренним радиусом, равным расстоянию от центра до ближайшего луча в этом пучке.

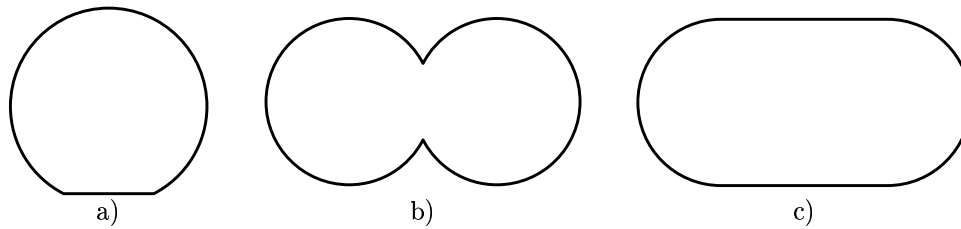
С другой стороны, в основополагающей работе Я. Г. Синая [6] было показано, что если все компоненты границы  $\partial Q$  являются рассеивающими и гладкими, то проблема освещаемости области  $Q$  решается положительно. Такие бильярдные столы, называемые бильярдами Синая, получаются, если, например, вырезать из тора конечное число выпуклых областей с гладкой границей (рис. 4). (На самом деле, в работе [6] были получены существенно более сильные результаты, а именно, доказаны эргодичность, перемешивание и  $K$ -свойство для бильярдов Синая.) Несколько позднее [3] было показано, что аналогичные результаты верны и для рассеивающих бильярдов с границей, имеющей сингулярные точки.

Отмеченные результаты, казалось бы, подтверждают интуитивное представление о том, что рассеивающие компоненты границы бильярдного стола способствуют освещаемости, в то время как фокусирующие компоненты препятствуют ей. Однако в работе [1] было показано, что существуют классы



**Рис. 4.** Бильярды Синая (рассеивающие бильярды)





**Рис. 5.** Простейшие формы фокусирующих бильярдov со стохастическим поведением  
 а) сегмент круга, вмещающий угол, меньший  $90^\circ$ ,  
 б) предыдущая область, отраженная относительно прямолинейного куска границы,  
 в) два полукруга, соединённые касательными прямолинейными отрезками («стадион»)

бильярдov, вообще не имеющие рассеивающих компонент, в которых, тем не менее, проблема освещаемости решается положительно. Более того, эти бильярдov обладают такими же сильными стохастическими свойствами, как и бильярдov Синая [1]. Простейшие примеры таких бильярдov приведены на рис. 5.

Работа [1] показала, что, вопреки общим представлениям, бытующим как среди математиков, так и среди физиков, имеется другой, отличный от механизма рассеяния, механизм, который может приводить к освещаемости области даже с чисто фокусирующей границей или, более общо, к стохастичности соответствующей динамической системы. Этот механизм называется механизмом *дефокусировки*. Более того, оказалось, что только эти два механизма, рассеяние и дефокусировка, могут генерировать стохастичность бильярдov.

Однако стохастичность (или более слабо, освещаемость) рассеивающих бильярдov была доказана для областей любой (конечной) размерности, в то время как для фокусирующих бильярдov все примеры ограничивались только двумерными бильярдами.

Только совсем недавно, через 25 лет после работы [1], было доказано, что механизм дефокусировки также может генерировать освещаемость (стохастичность) фокусирующих бильярдov в любой (конечной) размерности [15, 16, 17].

То, что этот вопрос был открыт в течение столь длительного времени, неудивительно. Дело в том, что в размерностях  $d > 2$  возникает очень серьёзное препятствие к действию механизма дефокусировки. Это препятствие есть, на самом деле, хорошо известное в геометрической оптике явление астигматизма [20].

Астигматизм означает, что при отражении пучка лучей от фокусирующего зеркала (фокусирующей компоненты границы) сила фокусировки оказывается существенно различной в разных плоскостях.

В данной статье мы объясним все эти результаты наглядно и с минимальным количеством формул, приводя ссылки на работы с полными доказательствами. Более того, мы рассмотрим и вопрос о том, какими свойствами должна обладать фокусирующая компонента границы бильярда (фокусирующее зеркало)

для того, чтобы его можно было использовать в качестве компоненты границы стохастического бильярда [10].

Известно, что любое рассеивающее зеркало можно использовать в качестве компоненты границы стохастического (освещаемого) бильярда.

Для фокусирующих зеркал это уже не так. Оказывается, что для того, чтобы фокусирующее зеркало можно было использовать для границы стохастического бильярда, оно должно быть абсолютно фокусирующим [10, 13].

Понятие абсолютно фокусирующего зеркала [10, 13] является новым понятием в геометрической оптике. Обычное понятие фокусирующего зеркала означает, что оно фокусирует, после первого отражения от него, все падающие на него параллельные пучки лучей. Однако параллельный пучок лучей может испытывать серию из любого конечного числа последовательных отражений от фокусирующего зеркала. (Отметим, что от рассеивающего зеркала у параллельного пучка лучей не может быть (в  $\mathbb{R}^d$ ) двух последовательных отражений.) Фокусирующее зеркало называется *абсолютно фокусирующим*, если оно фокусирует все падающие на него параллельные пучки лучей после последнего отражения в соответствующей серии последовательных отражений от этого зеркала (рис. 6).



Рис. 6. а) фокусирующее зеркало, б) абсолютно фокусирующее зеркало

## 1. БИЛЛИАРДЫ И ИХ ОСНОВНЫЕ СВОЙСТВА

Пусть  $Q \subset \mathbb{R}^d$  ( $d \geq 2$ ) —  $d$ -мерная область евклидова пространства с кусочно гладкой границей  $\partial Q$ . Мы предполагаем, что каждая компонента границы  $(\partial Q)_i$ ,  $i = 1, 2, \dots, k$ , принадлежит классу  $C^3$ , хотя некоторые формулируемые ниже результаты справедливы и в  $C^2$ . (Наше предположение позволяет избежать патологической ситуации, когда траектория бильярда может испытывать бесконечное число отражений в конечном интервале времени [23]).

Пусть граница  $\partial Q$  оснащена полем внутренних (по отношению к области  $Q$ ) нормальных к ней векторов  $n(q)$ ,  $q \in \partial Q$ . Ясно, что во всех регулярных точках  $q$  границы вектор  $n(q)$  определён однозначно. В сингулярных точках  $q \in \partial Q$  число нормальных векторов совпадает с числом регулярных компонент, пересекающихся в этой точке.

Зафиксировав единичный нормальный вектор в каждой регулярной точке  $q \in \partial Q$ , мы можем определить кривизну границы (или, если  $d > 2$ , вторую фундаментальную форму, иногда также называемую оператором кривизны  $K(q)$ )

в точке  $q$ . Мы предполагаем, что на каждой регулярной компоненте границы  $(\partial Q)_i$ ,  $i = 1, 2, \dots, k$ , оператор кривизны  $K(q)$  определён либо неотрицательно ( $K(q) \geq 0$ ,  $q \in (\partial Q)_i$ ), либо неположительно ( $K(q) \leq 0$ ,  $q \in (\partial Q)_i$ ), либо  $K(q) = 0$  во всех точках  $q \in (\partial Q)_i$ . (В двумерном случае,  $d = 2$ , это условие означает, что кривизна каждой компоненты границы положительна, отрицательна, либо тождественно равна нулю, см. рис. 2). Естественно называть регулярные компоненты границы таких областей рассеивающими, фокусирующими и нейтральными (или плоскими). Например, бильярд называется рассеивающим, если  $K(q) > 0$  во всех регулярных точках  $q$  границы  $\partial Q$ . Бильярд называется нигде не рассеивающим, если  $K(q) \leq 0$  во всех регулярных точках  $q \in \partial Q$ .

Мы уже не раз употребляли слово «бильярд», говоря пока лишь о форме рассматриваемых «бильярдных столов» (областей  $Q$ ).  $Q$  также называется конфигурационным пространством бильярда. Теперь же мы определим соответствующую динамическую систему, которая и называется бильярдом.

**ОПРЕДЕЛЕНИЕ 1.** Бильярд есть динамическая система, порождённая движением точечной частицы с единичной скоростью внутри  $Q$ . При достижении границы  $\partial Q$  точка отражается от неё по закону «угол падения равен углу отражения». Последнее означает, что при отражении от границы тангенциальная компонента скорости частицы сохраняется, а нормальная компонента скорости частицы меняет знак, т. е.  $v_+ = v_- - 2(n(q), v_-)n(q)$ , где  $v_+$  ( $v_-$ ) есть скорость частицы в момент непосредственно после (до) момента отражения.

Из определения бильярда сразу видно, что без ограничения общности можно считать скорость частицы единичной, т. е.  $\|v\| = 1$ .

Фазовое пространство  $\mathcal{M}$  бильярда представляет собой ограничение на область  $Q$  единичного касательного пучка к  $\mathbb{R}^d$ . Мы будем использовать стандартное обозначение для точек  $x = (q, v) \in \mathcal{M}$  фазового пространства, где  $q$  есть точка конфигурационного пространства  $Q$ , а  $v$  есть (единичный) вектор скорости частицы.

Бильярд представляет собой динамическую систему с непрерывным временем  $\{S^t\}$ ,  $-\infty < t < \infty$ . (Такие динамические системы называются потоками). Конечно, эта динамическая система определена только для таких точек  $x$ , траектории которых никогда не попадают в особые точки границы. Динамическая система  $\{S^t\}$  сохраняет меру Лиувилля  $dv = dq \cdot d\omega$ , где  $dq$  есть лебегова мера на  $Q$ , а  $d\omega$  — равномерная мера на  $(d-1)$ -мерной сфере. Нетрудно показать, что мера тех точек  $x$ , траектории которых  $\{S^t x\}$  хотя бы раз попадают в особые точки границы, равна нулю.

Бильярды обладают так называемым глобальным сечением Пуанкаре, которое позволяет свести изучение свойств бильярдного потока  $\{S^t\}$  к изучению динамической системы с дискретным временем, порождаемой бильярдным отображением  $T$ .

Обозначим  $M = \{x = (q, v), q \in \partial Q, (v, n(q)) \geq 0\}$ . Пусть  $\pi$  есть естественная проекция фазового пространства бильярда  $\mathcal{M}$  на его конфигурационное пространство  $Q$ , т. е.  $\pi(x) = q$ . Обозначим через  $\tau(x)$  первый положительный момент времени, когда траектория точки  $x = (q, v)$  отражается от границы  $\partial Q$ .

Определим теперь бильярдное отображение как  $Tx = (q', v') = S^\tau x$ . Таким образом,  $q'$  есть точка границы, в которой происходит следующее отражение

траектории точки  $x$  от  $\partial Q$ , а  $v'$  — вектор скорости частицы в момент сразу после этого отражения (рис. 2).

Пространство  $M$  называется фазовым пространством билиардного отображения  $T$ . Нетрудно убедиться (см., например, [5]), что  $T$  сохраняет проекцию  $d\mu$  меры Лиувилля  $dv$ , задаваемую формулой

$$d\mu(q, v) = \text{const}(v, n(q)) dq dv, \quad (1)$$

где  $dq$  есть  $(d-1)$ -мерная лебегова мера на границе  $\partial Q$ , порождаемая объёмом, а  $dv$  есть  $(d-1)$ -мерная лебегова (равномерная) мера на единичной сфере. Нормализующий множитель  $\text{const}$  обуславливает, что  $\mu(M) = 1$ .

## 2. РАССЕЯНИЕ, ФОКУСИРОВКА И ДЕФОКУСИРОВКА

Как уже отмечалось во введении, нас будет интересовать поведение узких пучков траекторий билиарда (лучей). На самом деле достаточно начать с параллельных пучков траекторий. При этом под узким пучком траекторий мы всегда будем иметь в виду инфинитезимально узкие пучки лучей, т.е. кривая, ортогональная всем лучам в пучке (фронт), является инфинитезимально малой и после любого конечного числа отражений пучок ни разу не попадает в особые точки границы.

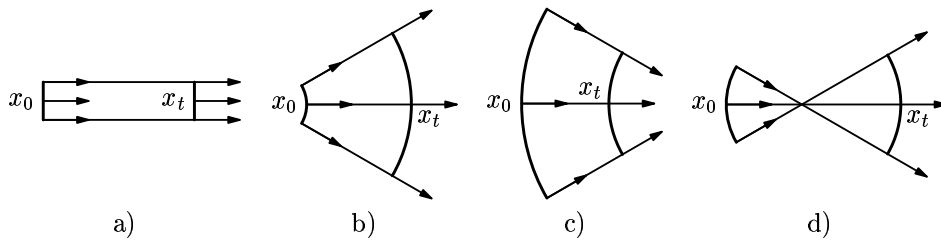
Как нам уже известно, узкие пучки лучей могут быть рассеивающимися, фокусирующимися или плоскими в зависимости от знака кривизны их фронта (рис. 3). Мы рассмотрим сейчас, что происходит с такими пучками в процессе динамики билиарда.

Очевидно, что эта динамика состоит из двух процессов:

1. Свободное движение частицы внутри области  $Q$  между двумя последовательными её отражениями от  $\partial Q$ .
2. Отражение частицы от границы.

Мы рассмотрим, что происходит с узкими пучками лучей (траекторий билиардов) при каждом из этих процессов.

Все возможные ситуации для свободного движения частицы между двумя последовательными отражениями изображены на рис. 7. Пусть начальная кривизна фронта такого пучка в его «центральной» точке  $x_0$  равна  $\varkappa(x_0)$ , а кривизна её



**Рис. 7.** Поведение пучков лучей между двумя последовательными отражениями

образа  $x_t = S^t x_0$  равна  $\varkappa(x_t)$ . Тогда имеем очевидное соотношение для соответствующих радиусов кривизны

$$\frac{1}{\varkappa(x_0)} + t = \frac{1}{\varkappa(x_t)}. \quad (2)$$

Откуда имеем

$$\varkappa(x_t) = \frac{\varkappa(x_0)}{1 + t\varkappa(x_0)}. \quad (3)$$

Из соотношения (3) сразу следует, что пучок лучей с плоским фронтом ( $\varkappa(x_0) = 0$ ) остаётся плоским (рис. 7а), рассеивающийся пучок с выпуклым фронтом ( $\varkappa(x_0) > 0$ ) остаётся выпуклым ( $\varkappa(x_t) > 0$ ), в то время как для фокусирующегося пучка с вогнутым фронтом ( $\varkappa(x_0) < 0$ ) возможны два случая. Действительно, такой пучок либо остаётся вогнутым (рис. 7с), если длина свободного пробега между двумя последовательными отражениями от границы  $\partial Q$  достаточно мала, либо этот пучок проходит через сопряжённую точку (в момент времени  $\hat{t} = -1/\varkappa(x_0)$ ) и становится расходящимся (рис. 7d). Ясно, что последняя ситуация имеет место только тогда, когда длина свободного пробега достаточно велика.

Рассмотрим теперь, что происходит с кривизной пучка при его отражении от границы. Обозначим через  $\varkappa_+(x)$  ( $\varkappa_-(x)$ ) кривизну фронта пучка в точке  $x$  в момент сразу после (до) его отражения от границы  $\partial Q$ . Тогда

$$\varkappa_+(x) = \varkappa_-(x) + \frac{2k(x)}{\cos \varphi(x)}, \quad (4)$$

где  $x$  — точка, в которой происходит отражение от границы, т. е.  $\pi(x) \in \partial Q$ ,  $k(x)$  — кривизна границы в точке  $\pi(x)$ , а  $\varphi(x)$  — угол падения (или отражения).

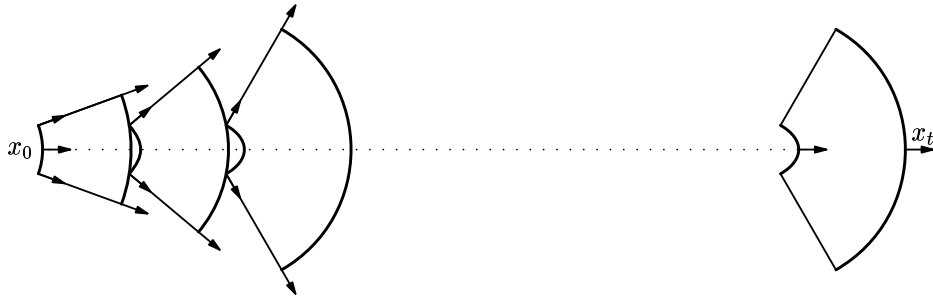
Формула (4) в геометрической оптике называется формулой зеркала. Читатель может либо сам вывести её, либо найти этот элементарный вывод в любой книге по геометрической оптике (или во многих статьях по теории бильярдов).

Непосредственно из формул (3) и (4) вытекает, как работают механизм рассеяния в рассеивающих бильярдах и механизм дефокусировки в бильярдах без рассеивающих компонент границы.

Мы начнём с более простого механизма рассеяния. Пусть расходящийся пучок лучей за время  $t$  испытал  $n$  отражений от границы  $\partial Q$  рассеивающего бильярда. Если рассматриваемый пучок — узкий, то по сделанному выше предположению все лучи в нём за рассматриваемое время отражались от одних и тех же регулярных компонент границы  $\partial Q$ , так что фронт такого пучка остаётся гладким. Его (локальная) эволюция изображена на рис. 8. Между каждыми двумя последовательными отражениями эволюция пучка такая же, как и на рис. 7b). Однако при каждом отражении кривизна пучка испытывает скачок, равный, согласно формуле зеркала (4),  $2k(x)/\cos \varphi(x)$ . Заметим, что  $0 \leq \cos \varphi \leq 1$  (напомним, что угол  $\varphi$  измеряется от внутренней нормали), а

$$k(x) \geq \min_{\substack{x \in \mathcal{M} \\ \pi(x) \in \partial Q}} k(x) > k_{\min} > 0.$$

Именно в этом месте мы используем, что все компоненты границы рассеивающего бильярда гладкие и строго выпуклые вовне области  $Q$ .

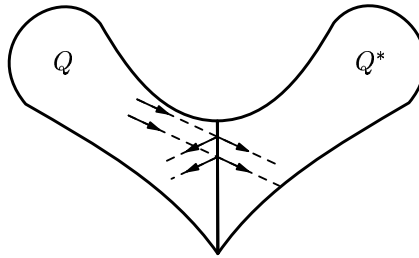


**Рис. 8.** Локальная эволюция расходящегося пучка лучей в рассеивающем бильярде

Отметим, что, если граница  $\partial Q$  имеет и плоские (нейтральные) компоненты, то отражение от них не меняет картины. Действительно, согласно формуле зеркала (4), при отражении от плоских компонент границы кривизна фронта пучка не меняется. Тем самым, можно считать, что пучок как бы «проходит сквозь такую компоненту границы» и продолжает расходиться.

Это соображение приводит к хорошо известному в геометрической оптике трюку (для построения изображения источника). Дело в том, что вместо того, чтобы отражать лучи пучка от плоской компоненты границы  $\partial Q$ , мы можем отразить область  $Q$  относительно этой компоненты, а луч в пучке продолжить прямолинейно (рис. 9). Тем самым, наличие у границы бильярда не только рассеивающих, но и нейтральных компонент, не изменяет характера эволюции расходящихся пучков траекторий. Действительно, отражение от нейтральной части границы фактически обуславливает только увеличение свободного пробега пучка между двумя последовательными отражениями от рассеивающей части границы, в то время как общая эволюция остаётся такой же, как на рис. 8. Единственно, что мы должны потребовать — это чтобы мера траекторий, которые отражаются не только от нейтральных компонент границы, была равна единице.

Удобно ввести следующее определение.



**Рис. 9.** Отражение области  $Q$  относительно плоской компоненты границы  $\partial Q$  и продолжение лучей

ОПРЕДЕЛЕНИЕ 2. Отражения от нейтральных компонент границы  $\partial Q$  называются *несущественными*, в то время как отражения от рассеивающих или фокусирующих компонент  $\partial Q$  называются *существенными*.

Тем самым, на рис. 8 изображены только существенные отражения. Из этого рисунка видно, что в рассеивающих бильярдах фронты расходящихся пучков лучей экспоненциально увеличиваются со временем (ниже мы приведём точную формулу). Это явление и приводит, в частности, к освещаемости соответствующих рассеивающих бильярдных столов (областей). Действительно, первоначально параллельный пучок лучей в таких областях уже после первого отражения от рассеивающей границы становится расходящимся.

Обратимся теперь к механизму дефокусировки. Этот механизм действует в некоторых (весьма широких) классах бильярдных столов, границы которых содержат фокусирующие компоненты.

Рассмотрим теперь эволюцию фокусирующегося (сходящегося) пучка лучей (рис. 3с). Как уже отмечалось, между двумя последовательными отражениями этого пучка от границы  $\partial Q$  возможны две ситуации: либо к моменту следующего отражения от границы пучок всё ещё остаётся сходящимся (рис. 7с), либо к этому моменту пучок уже стал расходящимся (рис. 7d). В последнем случае пучок лучей успевает пройти через точку пересечения (в линейном приближении, напомним, что пучок — узкий) лучей в пучке (сопряжённую точку).

ОПРЕДЕЛЕНИЕ 3. Мы будем говорить, что сходящийся (в момент сразу после отражения от границы  $\partial Q$ ) *пучок траекторий дефокусируется* во время его свободного пробега, если к моменту его следующего отражения от  $\partial Q$  он становится расходящимся.

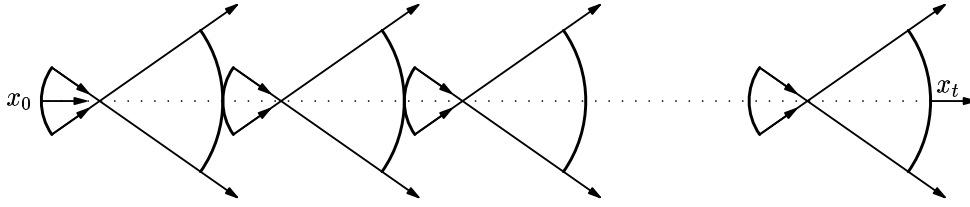
Ясно, что пучок дефокусируется, если время между последовательными его отражениями достаточно велико.

Как и при рассмотрении выше эволюции расходящихся пучков траекторий, в этом определении можно рассматривать только существенные отражения.

ОПРЕДЕЛЕНИЕ 4. Мы будем говорить, что в бильярде в области  $Q$  действует *механизм дефокусировки*, если для каждого сходящегося пучка лучей и серии его последовательных отражений от границы  $\partial Q$  после каждого отражения (за исключением, возможно, конечного, зависящего от этого пучка, числа начальных отражений в серии) от фокусирующей части границы  $\partial Q$  пучок успевает дефокусироваться.

Если все компоненты границы являются фокусирующими или нейтральными, то в бильярде, в котором действует механизм дефокусировки (возможно, после отбрасывания конечной и фиксированной для каждого сходящегося пучка серии) эволюция сходящегося пучка траекторий является такой, которая изображена на рис. 10.

Ясно, что без дефокусировки трудно рассчитывать, что проблема освещаемости соответствующей бильярдной области решается положительно. Действительно, в ситуации на рис. 7с) фронт пучка при его эволюции сужается. Однако и при действии механизма дефокусировки область  $Q$  может быть не освещаемой. Действительно, элементарный анализ [2] показывает, что механизм



**Рис. 10.** Возможная локальная эволюция сходящегося пучка лучей в фокусирующих бильярдах

дефокусировки действует в бильярде в круге, где проблема освещаемости, очевидно, решается отрицательно.

Действительно, механизм дефокусировки является гораздо более тонким, чем механизм рассеяния. При дефокусировке результат определяется конкуренцией двух процессов, фокусировки и рассеяния. В самом деле, между двумя последовательными отражениями пучок некоторое время фокусируется, а потом рассеивается (рис. 7d). Очевидно, что «победитель» определяется соответствующими временами.

Пусть  $\tau(x)$  — время свободного пробега «центрального» луча в сходящемся пучке. Тогда

$$\tau(x) = \tau_f(x) + \tau_d(x), \quad (5)$$

где  $\tau_f(x)$  ( $\tau_d(x)$ ) — время фокусировки (рассеяния) этого пучка.

Оказывается, что в круге для любого сходящегося пучка

$$\frac{\tau_f^{(n)}(x)}{\tau_d^{(n)}(x)} \xrightarrow{n \rightarrow \infty} 1.$$

В то время, как для положительного решения проблемы освещаемости  $Q$  нужно, чтобы длина фронта пучка достаточно быстро росла. Таким образом, в этом месте проблема уже становится не качественной, а количественной.

Для формулировки соответствующего условия рассмотрим любой (сходящийся, расходящийся или плоский) пучок траекторий (лучей). Пусть начальная кривизна фронта пучка в точке  $x_0$  была равна  $\varkappa_0(x)$ .

Найдём формулу для коэффициента растяжения  $\lambda_n(x)$  фронта этого пучка в точке  $T^n x$ , т. е. после  $n$ -го отражения этого пучка от границы  $\partial Q$ .

Из формулы (3) и рис. 7 легко следует [2], что

$$\lambda_n(x) = \prod_{i=0}^{n-1} (1 + |\varkappa_i(x)|\tau_i(x))^{-1}, \quad (6)$$

где  $\tau_i(x)$  — время между  $i$ -м и  $(i+1)$ -м отражениями от границы, а  $\varkappa_i(x)$  — кривизна фронта пучка в момент сразу после  $i$ -го отражения от границы.

Оказывается, что если величина  $\lambda_n(x)$  растёт экспоненциально, то задача освещаемости бильярдной области  $Q$  решается положительно.

Бильярды, для которых выражение (6) растёт экспоненциально, очевидно, обладают свойством локальной экспоненциальной неустойчивости. Это означа-



ет, что (локально!) их траектории экспоненциально расходятся в фазовом пространстве.

Динамические системы с таким свойством называются гиперболическими. Они образуют очень обширный и один из наиболее (если не просто наиболее) важный класс динамических систем.

Для многих классов бильярдов, например, для рассеивающих (бильярд Синая), бильярд на рис. 5 и их существенных обобщений (см. например [13, 9, 26, 24, 21]), доказаны гораздо более сильные свойства, чем положительное решение проблемы освещаемости.

А именно, для этих бильярд доказано, что они являются эргодическими и перемешивающими динамическими системами (см. точные определения, например, в [5]). Более того, эти системы метрически изоморфны бернуллиевским [22]. Тем самым, гиперболические бильярды являются динамическими системами с очень сильными стохастическими свойствами. Например, как следствие самого слабого из этих свойств, эргодичности, мы получаем сразу не только положительное решение задачи об освещаемости бильярдных областей  $Q$ , но и всюду плотность типичных траекторий в соответствующих фазовых пространствах  $M$ . Последнее означает, что типичная траектория бильярда не только проходит через каждую (малую) открытую подобласть бильярдного стола  $Q$ , но она пересекает эту подобласть по всюду плотному множеству направлений.

### 3. БИЛЛИАРДЫ В ВЫСШИХ (БОЛЬШЕ ДВУХ) РАЗМЕРНОСТЯХ И АСТИГМАТИЗМ

Мы обратимся теперь к многомерным бильярдам ( $d > 2$ ). Рассмотренный выше двумерный случай позволяет сразу указать, почему для бильярд Синая задача освещаемости (а также эргодичность, перемешивание и т. д.) была решена сразу для любой размерности, в то время как соответствующие вопросы для фокусирующих бильярд были открыты четверть века. (Отметим, что для рассеивающих бильярд с негладкой границей при  $d > 2$  возникают серьёзные, но чисто технические трудности при доказательстве их эргодичности [8, 7, 18]).

Действительно, вместо кривизны фронтов локальных пучков траекторий при  $d > 2$  приходится рассматривать соответствующие операторы кривизны (вторую квадратичную форму). Нетрудно видеть, что эти операторы для рассеивающих бильярд являются знакоопределёнными (положительно определёнными) вдоль всей (положительной по времени) траектории такого пучка, в то время как для бильярд в областях, где есть хотя бы одна фокусирующая компонента, это уже не так. Нет нужды специально подчёркивать, что положительно определённые операторы обычно много легче изучать, чем незнакоопределённые операторы. Эту проблему можно всё же, в какой-то мере, обойти, используя идею дефокусировки. Однако, существует препятствие, которое не обойдёшь, потому что это препятствие есть физическое явление — астигматизм.

Напомним, что механизм рассеяния является чисто количественным, в то время как механизм дефокусировки является качественным и количественным. Нам нужна не просто фокусировка, чтобы пучок успел дефокусироваться перед последующим отражением от границы, а достаточно сильная фокусировка

при (каждом!) отражении от границы. (Например, как мы уже знаем, круг не подходит.)

Астигматизм означает, что при отражении пучка лучей от фокусирующего зеркала сила фокусировки в разных плоскостях является различной. Более того, вдоль некоторых (на самом деле, непрерывного семейства) плоскостей фокусировка оказывается довольно слабой. Это обстоятельство несомненно препятствует действию механизма дефокусировки. Поэтому неудивительно, что были даже опубликованы некоторые работы, в которых ставилось под сомнение, что механизм дефокусировки может приводить к освещаемости областей с фокусирующей границей в высших ( $d > 2$ ) размерностях.

В этом параграфе мы объясним явление астигматизма, а также результаты недавних работ [15, 16, 17, 14], где было показано, что механизм дефокусировки работает и при  $d > 2$ .

Мы будем рассматривать только фокусирующие компоненты, являющиеся кусками сфер (сферические зеркала). Предположим, что узкий пучок лучей падает на такое зеркало. Рассмотрим его «центральный» луч  $x$ .

Динамика вблизи соответствующей бильярдной траектории может быть охарактеризована поведением второй фундаментальной формы локальной гиперповерхности  $W$  (фронта), ортогональной к потоку (рассматриваемому пучку траекторий). Матрицу этой квадратичной формы мы будем называть оператором кривизны и обозначать  $(Gx, x)$ .

Пусть  $n(q)$  — единичный вектор внутренней (т. е. направленной к центру сферы) нормали к сферическому зеркалу в точке  $q$ . Поскольку через луч  $x$  и центр сферы проходит единственная двумерная плоскость  $P$ , то из бильярдного закона отражения вытекает, что и отражённый луч будет лежать в плоскости  $P$ . (Это уже указывает на астигматизм — выделенность плоскости  $P$ .) Более того, легко видеть, что во всей серии последовательных отражений от сферического зеркала траектория луча  $x$  будет всё время находиться в плоскости  $P$  (рис. 11).

Рассмотрим  $(d - 1)$ -мерное касательное пространство  $U$  к фронту (гиперповерхности  $\tilde{U}$ ) в точке  $x$ . Плоскость  $P$  расщепляет  $U$  на два подпространства  $U = U_p \oplus U_t$ , где  $U_p = U \cap P$ , а  $U_t$  есть  $(d - 2)$ -мерное ортогональное дополнение к  $U_p$  в  $U$ .

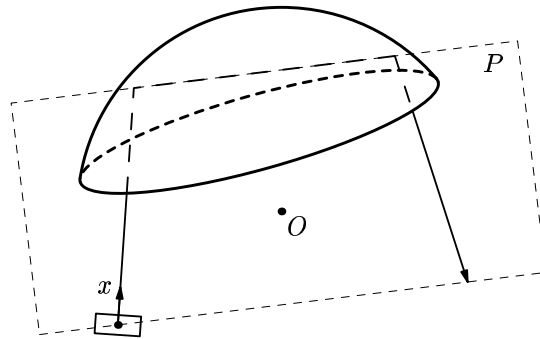


Рис. 11. Отражение луча от сферического зеркала

Мы приведём теперь формулы, дающие эволюцию оператора кривизны фронта. Поскольку в разные моменты времени операторы кривизны действуют в разных пространствах (касательных к соответствующим мгновенным фронтам), то нам придётся ввести и некоторые дополнительные операторы, переводящие эти пространства друг в друга [25].

Пусть  $V$  — изометричный оператор, отображающий плоскость  $T_-$ , касательную к фронту пучка в момент перед его отражением от границы  $\partial Q$ , в плоскость  $T_+$ , касательную к нему в момент после его отражения. Оператор  $V$  отображает  $T_-$  в  $T_+$  параллельно вектору нормали  $n(q)$  в точке отражения.  $V$  реализует необходимое вращение касательной гиперплоскости так, что  $T_+$  остаётся перпендикулярной вектору скорости. Аналогично,  $W$  — оператор, который отображает  $T_-$  на касательную к границе  $\partial Q$  в точке отражения гиперплоскость  $T_0$  вдоль направления, параллельного отражённому вектору скорости  $v_+$ , а  $W^*$  — сопряжённый оператор, отображающий  $T_0$  на  $T_+$ . Тем самым, оператор  $W$  преобразует оператор кривизны границы  $\partial Q$  в оператор, действующий в касательном пространстве к фронту.

Соответствующая формула имеет вид [25]:

$$G_+ = V^{-1}G_-V - 2(v_+, n(q))W^*KW, \quad (7)$$

где  $K$  — оператор кривизны границы  $\partial Q$  в точке отражения  $q$ . Соотношение (7) является многомерным аналогом формулы (4).

Формула для изменения кривизны фронта во время свободного (без отражений от границы) движения частицы выглядит проще и тоже полностью аналогична соответствующему соотношению (3) для двумерного случая. Она имеет вид:

$$G_t(y) = G_0(x)(1 + G_0(x))^{-1}, \quad (8)$$

где  $G_t(y)$  ( $G_0(x)$ ) — оператор кривизны, действующий на гиперплоскости  $U(y)$  ( $U(x)$ ), перпендикулярной направлению движения (вектору скорости), и  $y = S^t x$ .

Проиллюстрируем наглядно, что происходит с фронтом  $\tilde{U}$  пучка лучей при его отражении от границы. Фиксируем некоторое направление  $\ell$  в касательном пространстве  $U$  (рис. 12). Величина кривизны фронта  $\tilde{U}$  до отражения в

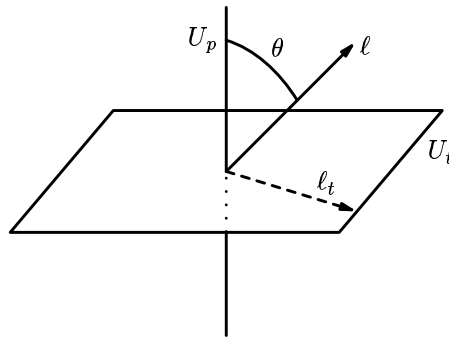


Рис. 12. Разложение касательного пространства

направлении  $\ell$  равна  $(G_- \ell, \ell)$ . Для того, чтобы получить аналогичное выражение для кривизны фронта после отражения в направлении  $\ell$ , т. е.  $(G_+ \ell, \ell)$ , напишем очевидное соотношение

$$\ell = \ell_p \cos \theta + \ell_t \sin \theta, \quad (9)$$

где  $\theta$  — угол между  $\ell$  и  $U_p$ , а  $\ell_p, \ell_t$  — (нормализованные) проекции  $\ell$  соответственно на  $U_p$  и  $U_t$ . Тогда изменение кривизны фронта (вдоль направления  $\ell$ ) после отражения равно

$$\kappa_+ - \kappa_- = \cos^2 \theta \left( \frac{2}{r \cos \theta} \right) + \sin^2 \theta \left( \frac{2 \cos \theta}{r} \right), \quad (10)$$

где  $r$  — радиус нашего сферического зеркала.

Формула (10) в качестве специального случая ( $\theta = 0$ ) даёт формулу (4). Это естественно, поскольку, как уже отмечалось, при отражении от сферического зеркала (изнутри!), падающий и отражённый лучи лежат в одной плоскости.

Однако для кривизны вдоль ортогонального к направлению распространения луча пространства вместо формулы (4) мы имеем

$$\kappa_+^\perp - \kappa_-^\perp = 2k \cos \varphi, \quad (11)$$

где  $k = 1/r$ .

Формула (11) показывает, что изменение кривизны фронта после отражения оказывается существенно различным вдоль разных направлений (астигматизм). При этом фокусировка оказывается (см. (11)) весьма слабой по некоторым направлениям. Это обстоятельство приводит к идее, что механизм дефокусировки не работает в размерностях  $d > 2$ . Действительно, этот механизм требует, чтобы между каждыми двумя последовательными отражениями в любой как угодно длинной серии последовательных (существенных!) отражений от границы  $\partial Q$  (здесь предполагается, что  $\partial Q$  не содержит рассеивающих компонент) имеет место ситуация, изображённая на рис. 7d) и рис. 10. (Как обычно, предполагается, что начальным, *фиксированным для каждой точки*  $x \in M$  куском траектории можно пренебречь.) Но из формул (10), (11) сразу следует, что для отражений от границы с как угодно большим номером пучок может стать сходящимся, но кривизна его фронта при этом может быть сколь угодно малой (рис. 13). Тогда, если эта кривизна меньше диаметра бильярдного стола  $Q$ , то пучок не успеет пройти через фокус (сопряжённую точку) до следующего отражения от границы. Тем самым, механизм дефокусировки разрушается.

Таково и было «общественное мнение» в течение многих лет. Однако, в работах [15, 16, 17, 14] было показано, что снова, как и за 25 лет до этого [1] «общественное мнение» оказалось неправильным. (Под «общественным мнением» мы

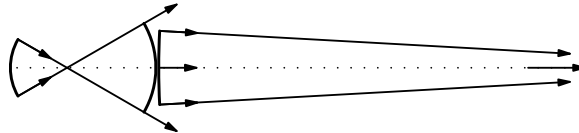


Рис. 13. Возможное нарушение механизма дефокусировки

здесь имеем в виду не только мнение математиков, но и физиков, механиков, инженеров и т. д. Следует отметить, тем не менее, что мы никогда не интересовались мнением по этому вопросу философов, филологов и представителей других гуманитарных профессий).

Мы приведём здесь (в описательной форме) соответствующие результаты и дадим один конкретный пример области с фокусирующими и плоскими компонентами границы, удовлетворяющей требуемым условиям.

Предположим, что бильiardная область  $Q$  обладает следующими свойствами:

0. Бильiard в  $Q$  является нигде не рассеивающим, т. е. граница  $\partial Q$  не содержит рассеивающих компонент.
1. Все фокусирующие компоненты границы являются сферическими зеркалами (т. е. получаются в результате «отсекания» от  $(d - 1)$ -мерной сферы сферической шапочки некоторой  $(d - 1)$ -мерной гиперплоскостью).
2. Соответствующие сферические шапочки не очень велики, а именно, соответствующий (максимальный) внутренний угол не превосходит  $60^\circ$  (угол  $\gamma$  на рис. 14).
3. Дополнение каждой фокусирующей компоненты  $\partial Q$  до соответствующей полной сферы целиком лежит внутри  $Q$ .
4. Граница каждой фокусирующей компоненты  $\partial Q$  (сферической шапочки) принадлежит какой-то (одной!) плоской компоненте границы  $\partial Q$ .
5. Если плоская компонента границы  $(\partial Q)_i$  содержит границу какой-то сферической шапочки (фокусирующей компоненты  $\partial Q$ ), то все плоские компоненты  $\partial Q$ , пересекающиеся с  $(\partial Q)_i$ , должны быть перпендикулярны гиперплоскости, содержащей  $(\partial Q)_i$ .
6. Мера множества траекторий, которые отражаются только от плоских компонент границы, равна нулю.

На рис. 14 изображена бильiardная область, которую естественно назвать многомерным стадионом из-за её аналогии стадиону (рис. 5с). Эта область состоит из куба, к противоположным граням которого приторочены сферические шапочки. (Если взять стадион (рис. 5с) и вращать его вокруг его большой оси, то бильiard в получившейся трёхмерной области не будет эргодическим. Нетрудно видеть, что в этом бильiarде есть первый интеграл (инвариантная функция,

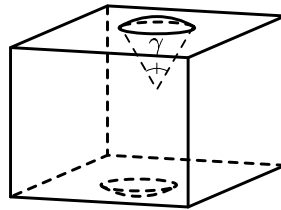


Рис. 14. Многомерный стадион

называемая в механике моментом). Эту (непрерывную!) симметрию надо разрушить. Взяв куб (вместо цилиндра), соединяющий сферические шапочки, мы и разрушаем её.)

В работах [15, 16, 17] рассматриваются существенно более общие области, бильярды в которых являются эргодическими, перемешивающими и являются бернуллиевскими системами. В статье [14] эти результаты обобщены для областей, границы которых могут включать не только фокусирующие и нейтральные, но и рассеивающие компоненты.

Мы не приводим здесь доказательств, отсылая к соответствующим статьям. Отметим только, что условие 2 (относительная «малость» фокусирующих компонент) — это цена, заплаченная астигматизму. Действительно, в двумерном случае фокусирующие компоненты эргодического бильярда могут быть сколь угодно близки к полным окружностям (рис. 5а).

#### 4. УСЛОВИЯ ЭРГОДИЧНОСТИ БИЛЛИАРДОВ И АБСОЛЮТНО ФОКУСИРУЮЩИЕ ЗЕРКАЛА

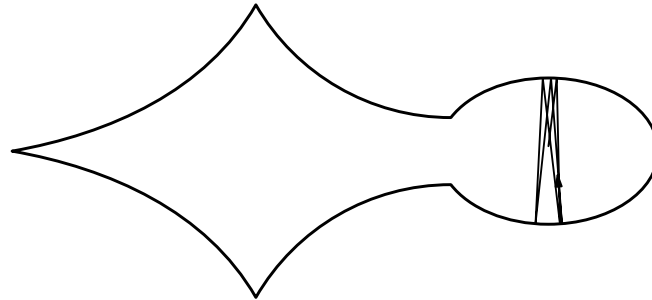
Естественно поставить общий вопрос: каким общим свойством должна обладать граница бильярдного стола  $Q$  для того, чтобы бильярд в  $Q$  был эргодичен?

Как мы уже отмечали, эргодичность влечёт не только положительное решение задачи освещаемости области  $Q$ , но и гораздо более сильные свойства.

В работах [11, 12] были сформулированы соответствующие условия и доказана теорема об эргодичности для двумерного случая.

Основным при этом оказывается вопрос о том, какие фокусирующие зеркала могут служить компонентами границы эргодического бильярда. Ясно, что не все фокусирующие зеркала могут быть использованы. Например, если фокусирующее зеркало представляет собой часть эллипса, содержит концы его малой полуоси и больше полуэллипса. Тогда, как хорошо известно, любой луч, стартующий вблизи этой полуоси, навечно остаётся в её окрестности (рис. 15).

Оказывается [12], что каждая фокусирующая компонента границы эргодического бильярда должна быть абсолютно фокусирующей. Как уже было объяснено во введении, фокусирующая компонента (кривая) является абсолютно фо-



**Рис. 15.** Островок устойчивости «внутри» фокусирующей компоненты

кусирующей, если любой падающий на неё луч фокусируется после последнего отражения в серии последовательных отражений от этой компоненты. Эквивалентно, любой плоский пучок траекторий, падающий на абсолютно фокусирующее зеркало, должен фокусироваться после последнего отражения в этой серии (рис. 6).

Примером фокусирующей, но не абсолютно фокусирующей кривой может служить полуэллипс (содержащий концы его малой полуоси), если отношение длин его осей больше  $\sqrt{2}$  (см. [21, 13]; работа [13] содержит элементарное доказательство этого факта).

В многомерном случае понятие абсолютно фокусирующего зеркала определяется аналогично. В работах [15, 16, 17, 14] был сделан только первый шаг к определению общих условий, при которых билиарды на многомерных билиардных столах ( $\dim Q > 2$ ), имеющие компоненты границы всех трёх типов (плоские, рассеивающие и фокусирующие), являются эргодическими. Наша гипотеза состоит в том, что и при  $d > 2$  каждая фокусирующая компонента границы эргодического билиарда должна быть абсолютно фокусирующей, как и в двумерном случае.

### СПИСОК ЛИТЕРАТУРЫ

- [1] *Бунимович Л. А.* Об эргодических свойствах некоторых билиардов // Функ. ан. и прил., 1974. Т. 8. С. 254–255.
- [2] *Бунимович Л. А.* О билиардах, близких к рассеивающим // Матем. Сб., 1974. Т. 94. С. 49–73.
- [3] *Бунимович Л. А., Синай Я. Г.* Об основной теореме теории рассеивающих билиардов // Матем. Сб., 1972. Т. 90. С. 415–431.
- [4] *Гальперин Г. А., Земляков А. Н.* Математические билиарды. М.: Наука, 1990.
- [5] *Корнфельд И., Синай Я. Г., Фомин С. В.* Эргодическая теория. М.: Наука, 1980.
- [6] *Синай Я. Г.* Динамические системы с упругими отражениями. Эргодические свойства рассеивающих билиардов // УМН, 1970. Т. 25. Вып. 2. С. 141–192.
- [7] *Синай Я. Г., Чернов Н. И.* Эргодические свойства некоторых систем двумерных дисков и трёхмерных шаров // УМН, 1987. Т. 42. С. 181–207.
- [8] *Чернов Н. И.* Построение трансверсальных слоёв в многомерных полурассеивающих билиардах // Функ. ан. и прил., 1982. Т. 16. С. 35–46.
- [9] *Bunimovich L. A.* On the ergodic properties of nowhere dispersing billiards // Comm. Math. Phys., 1979. Vol. 65. P. 295–312.
- [10] *Bunimovich L. A.* Many-dimensional nowhere dispersing billiards with chaotic behavior // Physica D, 1988. Vol. 33. P. 58–64.
- [11] *Bunimovich L. A.* A theorem on ergodicity of two-dimensional hyperbolic billiards // Comm. Math. Phys., 1990. Vol. 130. P. 599–621.

- [12] *Bunimovich L. A.* Conditions of stochasticity of two-dimensional billiards // *Chaos*, 1991. Vol. 1. P. 187–193.
- [13] *Bunimovich L. A.* On absolutely focusing mirrors // *Ergodic Theory and Related Topics* (eds. U. Krengel et al.). N.-Y.: Springer-Verlag, 1992. P. 62–82. (Lect. Notes Math., vol. 1514)
- [14] *Bunimovich L. A.* Hyperbolicity and astigmatism // *J. Stat. Phys.* To be published.
- [15] *Bunimovich L. A., Rehacek J.* Nowhere dispersing 3D billiards with non-vanishing Lyapunov exponents // *Comm. Math. Phys.*, 1997. Vol. 189. P. 729–757.
- [16] *Bunimovich L. A., Rehacek J.* How high-dimensional stadia look like // *Comm. Math. Phys.*, 1998. Vol. 197. P. 277–301.
- [17] *Bunimovich L. A., Rehacek J.* On the ergodicity of many-dimensional focusing billiards // *Annales Inst. H. Poincaré*, 1998. Vol. 68. P. 421–448.
- [18] *Burago D., Kononenko A., Ferleger S.* Uniform estimates on the number of collisions in semi-dispersing billiards // *Annals of Mathematics*, 1998. V. 147. P. 695–708.
- [19] *Chirikov B. V., Izrailev F. M., Shepelyansky D. I.* Dynamical stochasticity in classical and quantum mechanics // *Sov. Sci. Rev.*, 1981. Vol. 2. P. 209–267.
- [20] *Coddington H.* *Treatise on reflection and refraction of light.* London: Simpkin & Marshall, 1829.
- [21] *Donnay V. J.* Using integrability to produce chaos: billiards with positive entropy // *Comm. Math. Phys.*, 1991. Vol. 141. P. 225–257.
- [22] *Galavotti G., Ornstein D.* Billiards and Bernoulli schemes // *Comm. Math. Phys.*, 1974. Vol. 38. P. 83–101.
- [23] *Halpern B.* Strange billiard tables // *Trans. Amer. Math. Soc.*, 1977. Vol. 46. P. 297–305.
- [24] *Markarian R.* Billiards with Pesin region of measure one // *Comm. Math. Phys.*, 1988. Vol. 118. P. 87–97.
- [25] *Sinai Ya. G.* Development of Krylov's ideas // *Krylov N. S. Works on the foundations of statistical physics.* Princeton, NJ: Princeton Univ. Press, 1979. P. 239–281.
- [26] *Wojtkowski M. P.* Principles for the design of billiards with non-vanishing Lyapunov exponents // *Comm. Math. Phys.*, 1986. Vol. 105. P. 391–414.



# Внешний бильярд

С. Л. Табачников

## 1. ОПРЕДЕЛЕНИЕ И ПРИМЕРЫ

Для игры во внешний бильярд нужен бильярдный стол — ограниченная выпуклая область на плоскости. Игра состоит в следующем. Начальная позиция бильярдного шара — это точка  $A$  вне бильярдного стола. Из точки  $A$  можно провести две касательные (точнее, опорные) прямые к границе бильярдного стола  $\gamma$  — левую  $l$  и правую  $r$ , если смотреть из точки  $A$ . Продолжим правую касательную за точку касания  $X$  до точки  $B$  так, что  $BX = AX$  (см. рис. 1). Точка  $B$  и есть новое положение бильярдного шара. Обозначая внешнее бильярдное отображение через  $T$  (или, если нужно подчеркнуть зависимость от стола, через  $T_\gamma$ ), имеем:  $B = T(A)$ .

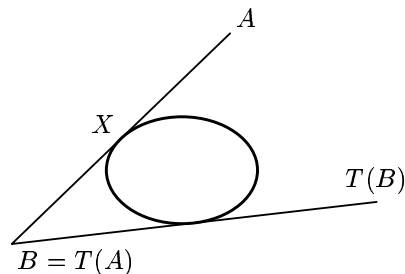


Рис. 1.

У данного определения есть один дефект: точка  $X$  может оказаться не единственной. Это произойдет, если граница бильярдного стола содержит отрезок прямой, например, если бильярдный стол — многоугольник. Этот недостаток определения неустраим и мы считаем, что отображение  $T$  не определено в тех точках, которые лежат на продолжениях отрезков, являющихся частью кривой  $\gamma$ . Это напоминает обычный (внутренний) бильярд: если бильярдный шар попадает в угол, его дальнейшее движение не определено.

Внешний бильярд изучен гораздо менее основательно, чем внутренний. Даже терминология не вполне устоялась: внешний бильярд известен также как «дуальный бильярд» и «антибильярд» (по-английски, «outer», «exterior», «dual» и «antibilliard»). Я не знаю, кто первым изобрел внешний бильярд. Знаменитый математик второй половины XX века Юрген Мозер описал внешний бильярд в оказавшей большое влияние книге [5]. Мозер узнал о внешнем бильярде из лекции австралийского математика Б. Ньюмана в конце 1950-х гг.

Рассмотрим несколько примеров.

ПРИМЕР 1. Если внешний бильярдный стол — круг, то каждая точка движется по концентрической окружности. Тем самым, каждая концентрическая окружность — инвариантная кривая внешнего бильярдного отображения  $T$ .

Отметим, что внешнее бильярдное отображение не чувствительно к аффинным преобразованиям плоскости. Точнее говоря, пусть  $f$  — аффинное преобразование, а  $\gamma$  — внешняя бильярдная кривая. Тогда

$$f \circ T_\gamma = T_{f(\gamma)} \circ f$$

(см. рис. 2).

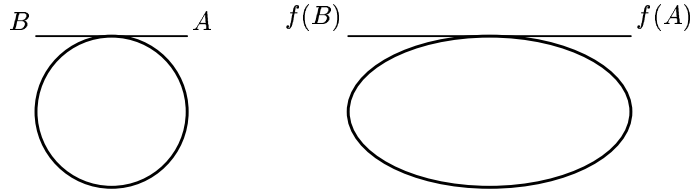


Рис. 2.

Поскольку эллипс — аффинный образ окружности, все, сказанное о круглом внешнем бильярдном столе, относится в равной мере и к эллиптическому.

ПРИМЕР 2. Если внешний бильярдный стол — квадрат, то движение каждой точки — периодическое. Структура орбит изображена на рис. 3: каждая точка квадрата, отмеченного номером  $n$ , посещает все остальные квадратики с этим номером (их  $4n$ ) по одному разу, прежде чем вернуться в исходное положение. Читатель легко установит аналогичные факты о треугольном внешнем бильярдном столе и о столе в форме правильного шестиугольника. В силу аффинной инвариантности, сказанное о квадрате относится и к произвольному параллелограмму.

ПРИМЕР 3. Если внешний бильярдный стол — правильный пятиугольник, то наряду с периодическими орбитами, есть и бесконечные. Одна такая орбита

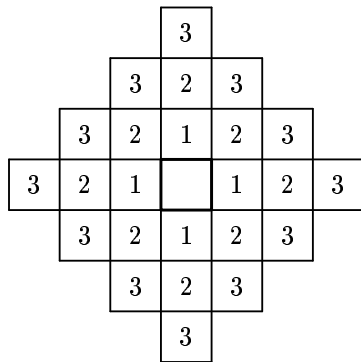


Рис. 3.

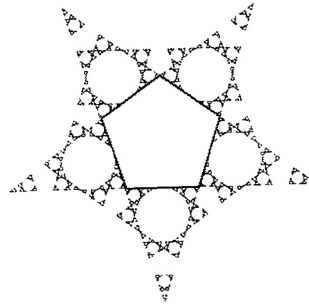


Рис. 4.

показана на рис. 4. Отметим бросающееся в глаза самоподобие изображенной орбиты (точнее, ее замыкания). Существование бесконечных орбит вовсе не очевидно, и доказательство (данное в [8] и [9]) опирается на отмеченное самоподобие. Оно же позволяет вычислить (дробную!) размерность множества бесконечных орбит, которая оказывается равной  $\lg 6 / \lg(\sqrt{5} + 2) = 1.24\dots$

Компьютерные эксперименты показывают, что нечто подобное происходит и с другими правильными  $n$ -угольниками (за исключением  $n = 3, 4, 6$  — см. предыдущий пример).

## 2. СОХРАНЕНИЕ ПЛОЩАДИ

Независимо от формы внешнего бильярдного стола, внешнее бильярдное отображение обладает свойством сохранять площадь. Докажем это.

Возьмем близкие точки  $X$  и  $X'$  на кривой  $\gamma$ . Выберем число  $r > 0$  и проведем касательные отрезки длины  $r$  к  $\gamma$  в каждой точке между  $X$  и  $X'$ . Концы этих отрезков пробегают кривые  $AA'$  и  $BB'$ , и при этом  $T(AA') = BB'$ . Прделаем то же построение, заменяя  $r$  на  $r - \varepsilon$  (предполагая конечно, что  $\varepsilon$  — бесконечно малое); мы получим кривую  $CC'$  и ее образ  $DD'$ . Обозначим через  $Y$  точку пересечения отрезков  $AB$  и  $A'B'$  и через  $\delta$  (еще одно бесконечно малое!) угол между этими отрезками (см. рис. 5).

Вычислим площадь криволинейного четырехугольника  $ACC'A'$  с точностью до малых второго порядка, то есть, пренебрегая  $\varepsilon^2$  и  $\delta^2$ . Имеем:

$$\text{Площадь } (AYA') = \frac{1}{2}\delta r^2, \quad \text{Площадь } (CYC') = \frac{1}{2}\delta(r - \varepsilon)^2 = \frac{1}{2}\delta r^2 - \delta\varepsilon r,$$

а значит

$$\text{Площадь } (ACC'A') = \delta\varepsilon r.$$

Аналогично,

$$\text{Площадь } (BDD'B') = \delta\varepsilon r,$$

то есть отображение  $T$  сохраняет площадь бесконечно малого четырехугольника  $ACC'A'$ . Это и доказывает свойство внешнего бильярдного отображения сохранять площадь.

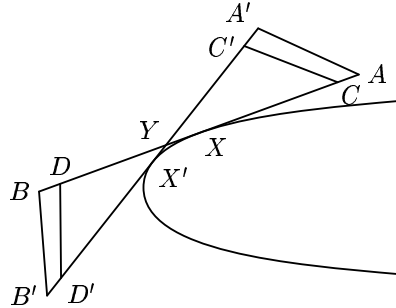


Рис. 5.

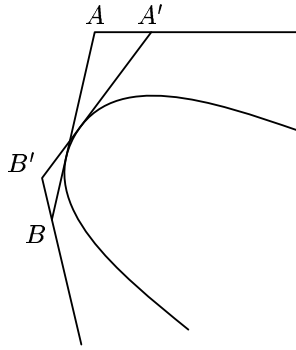


Рис. 6.

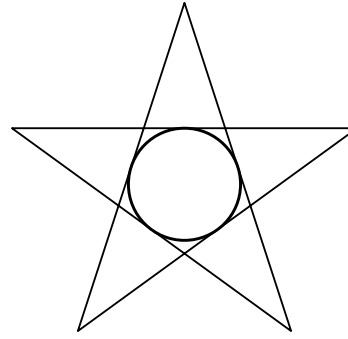


Рис. 7.

Сохранение площади имеет разнообразные следствия. Зададимся, например, таким вопросом. Предположим, что граница внешнего бильярдного стола  $\gamma$  — гладкая строго выпуклая кривая (т. е. ее кривизна всюду положительна). Пусть задано целое число  $n \geq 3$ . Существует ли у внешнего бильярдного отображения  $n$ -периодическая орбита?

Ответ на этот вопрос положительный, как показывает следующее рассуждение. Рассмотрим  $n$ -угольник наименьшей площади, описанный вокруг  $\gamma$ . Я утверждаю, что стороны этого многоугольника делятся пополам точками касания с кривой  $\gamma$ , то есть что это —  $n$ -периодическая орбита внешнего бильярдного отображения.

Действительно, если сторона  $AB$  не делится точкой касания пополам, то, слегка повернув отрезок  $AB$  в новое положение  $A'B'$ , можно уменьшить площадь многоугольника (см. рис. 6). Это рассуждение почти идентично предыдущему, доказывающему, что  $T$  сохраняет площадь.

Это же рассуждение доказывает, что у внешнего бильярдного отображения есть  $n$ -периодические орбиты, обходящие вокруг бильярдного стола  $k$  раз, где  $1 \leq k \leq n/2$  (см. рис. 7, на котором  $n = 5$ ,  $k = 2$ ).

В действительности для каждого  $n \geq 3$  и  $1 \leq k \leq n/2$  найдется по крайней мере две периодические орбиты внешнего бильярдного отображения, но доказать это значительно сложнее. Аналогичный факт для внутреннего бильярда был доказан Биркгофом в 1920-х гг.

Рассмотрим еще одно следствие сохранения площади. Предположим, что у внешнего бильярдного отображения есть выпуклая инвариантная кривая  $\Gamma$  (говоря об инвариантной кривой, я всегда предполагаю, что  $\Gamma$  обходит вокруг бильярдного стола ровно один раз). Можно ли восстановить бильярдный стол по  $\Gamma$ ? Ответ дается следующей конструкцией.

Зафиксируем число  $c > 0$  и рассмотрим семейство прямых, отсекающих от фигуры, ограниченной  $\Gamma$ , сегменты постоянной площади  $c$ . Полученное однопараметрическое семейство прямых имеет огибающую, т. е. такую кривую  $\gamma$ , которая касается прямых семейства в каждой своей точке (см. рис. 8).

Вообще говоря, кривая  $\gamma$  может не быть гладкой (этот вопрос подробно обсуждается в статье [10]), но мы предположим, что  $\gamma$  — гладкая кривая. Я утвер-

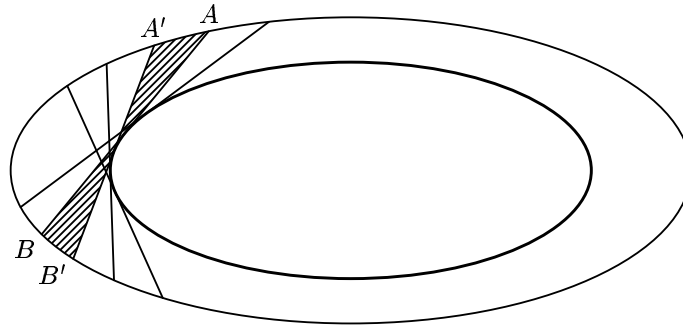


Рис. 8.

жду, что  $\Gamma$  — инвариантная кривая внешнего бильярдного отображения  $T_\gamma$ . Действительно, площади заштрихованных инфинитезимальных треугольников на рис. 8 равны, а это, как мы уже видели, означает, что отрезок  $AB$  делится точкой касания с кривой  $\gamma$  пополам.

Отметим, что описанная конструкция зависит от параметра  $s$ , поэтому кривая  $\Gamma$  задает целое однопараметрическое семейство бильярдных столов. Аналог этой конструкции для внутреннего бильярда хорошо известен (производится с помощью нерастяжимой нити) — см., например, [2].

Вернемся к примеру 1 из предыдущего параграфа. Если внешний бильярдный стол — круг, то каждая точка вне него лежит на инвариантной кривой; в этом случае говорят, что внешнее бильярдное отображение интегрируемо. Интегрируемость — очень редкое явление; она означает, что отображение весьма регулярно, т. е. поведение точки при итерациях отображения полностью предсказуемо. Существует гипотеза, что интегрируемость внешнего бильярдного отображения имеет место, только если внешний бильярдный стол — эллипс. По всей видимости, доказать это весьма трудно (я верю, что эта гипотеза верна); аналогичная гипотеза для внутреннего бильярда (приписываемая Биркгофу) тоже не доказана, несмотря на усилия многих математиков.

### 3. ДВОЙСТВЕННОСТЬ МЕЖДУ ВНЕШНИМ И ВНУТРЕННИМ БИЛЛИАРДАМИ

Читатель, вероятно, уже отметил своеобразную двойственность между внешним и внутренним бильярдами. Траектория внутреннего бильярда — вписанный многоугольник, а внешнего — описанный, причем первый имеет экстремальный периметр, а последний — экстремальную площадь (см. предыдущий параграф). Чем объяснить такую двойственность между длиной и площадью?

Ситуация прояснится, если рассмотреть бильярд не на плоскости, а на сфере. Роль прямых в сферической геометрии играют большие круги. Имеется замечательная двойственность между точками и ориентируемыми прямыми: полюсу отвечает экватор (см. рис. 9, на котором точки названы большими буквами, а соответствующие прямые — строчными). Отметим, что (сферическое) расстояние  $AB$  равно углу между прямыми  $a$  и  $b$ .

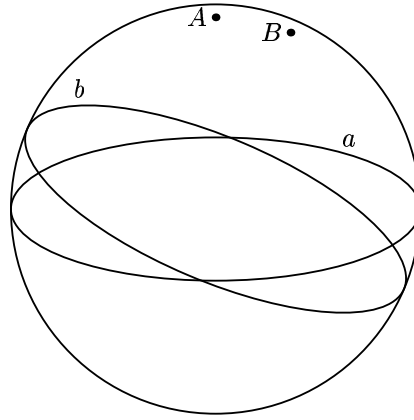


Рис. 9.

Важное свойство сферической двойственности состоит в следующем: если точка  $A$  лежит на прямой  $b$ , то двойственная точка  $B$  лежит на двойственной прямой  $a$  (докажите!).

Двойственность распространяется на сферические кривые. Гладкая ориентированная кривая  $\gamma$  определяет однопараметрическое семейство касательных прямых. Сопоставляя этим прямым двойственные точки, получаем однопараметрическое семейство точек, т. е. новую кривую  $\gamma^*$ . Это и есть кривая, двойственная кривой  $\gamma$ . Двойственную кривую  $\gamma^*$  можно построить так: каждую точку  $X$  исходной кривой  $\gamma$  нужно сдвинуть на расстояние  $\pi/2$  (то есть на четверть большего круга) вдоль ориентированной нормали к  $\gamma$  в точке  $X$ ; ориентация нормали определяется ориентацией кривой  $\gamma$  по правилу правой руки. Например, если  $\gamma$  — параллель северной широты  $\alpha$ , то, в зависимости от ориентации  $\gamma$ , кривая  $\gamma^*$  — параллель северной или южной широты  $\pi/2 - \alpha$ .

Отметим важное свойство двойственности: кривая, двойственная двойственной, центрально-симметрична исходной, т. е.

$$(\gamma^*)^* = -\gamma.$$

Читатель, не знакомый со сферической двойственностью, получит удовольствие от доказательства сформулированного свойства; подробное обсуждение геометрии сферических кривых содержится в статье [1].

Рассмотрим отражение бильярдного шара от бильярдной кривой  $\gamma$  в точке  $X$  (см. рис. 10, левая часть). Двойственная картина изображена на рис. 10 справа (как и раньше, двойственные объекты обозначены одинаковыми буквами).

Закон бильярдного отражения гласит: «угол падения равен углу отражения», т. е. углы, образованные прямыми  $a$  и  $b$  с касательной  $l$ , равны. На двойственной картине это означает, что  $AL = LB$ , а значит, двойственное бильярдное отображение относительно кривой  $\gamma^*$  переводит точку  $A$  в точку  $B$ . Итак, сферические внешний и внутренний бильярды двойственны друг другу.

А как объяснить двойственность между длиной и площадью? Траектория внутреннего бильярда — вписанный многоугольник экстремального периметра.

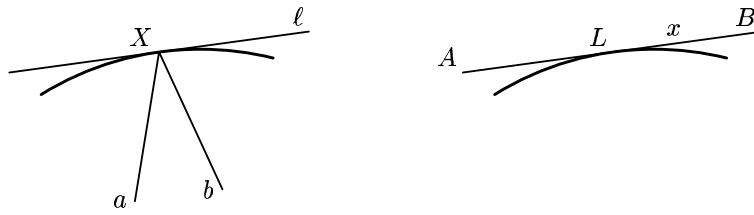


Рис. 10.

тра. Длины двойственны углам. Следовательно, траектория внешнего бильярда — описанный многоугольник, имеющий экстремальную сумму углов. Сумма углов сферического  $n$ -угольника связана с его площадью:

$$\sum \alpha_i = \pi(n - 2) + A$$

(конечно, полная площадь сферы равна  $4\pi$ ). Это — теорема Гаусса–Бонне для сферических многоугольников; см., например, [2].

Итак, орбита внешнего бильярдного отображения на сфере — это описанный многоугольник экстремальной площади. Плоскость можно представлять как сферу бесконечного радиуса. Сумма углов  $n$ -угольника перестает быть переменной (она равна  $\pi(n - 2)$ ), однако площадь сохраняет роль функции на описанных многоугольниках, чьи экстремумы отвечают орбитам внешнего бильярда.

#### 4. МОЖЕТ ЛИ ОРБИТА УЙТИ В БЕСКОНЕЧНОСТЬ?

Посмотрим на внешний бильярд «с высоты птичьего полета», т. е. зададимся вопросом: как устроены орбиты очень далеко от бильярдного стола? Если смотреть издали, то бильярдная кривая  $\gamma$  представляется точкой, а внешнее бильярдное отображение  $T$  — центральной симметрией. Эволюция точки под действием отображения  $T^2$  выглядит как непрерывное движение вдоль центрально-симметричной замкнутой кривой  $\Gamma$ , причем это движение удовлетворяет второму закону Кеплера: секториальная скорость постоянна (т. е. в равные времена радиус-вектор заметает равные площади; единица времени — это, конечно, одна итерация отображения  $T^2$ ). На рис. 11 изображены некоторые внешние бильярдные кривые  $\gamma$  (верхняя строка) и соответствующие инвариантные

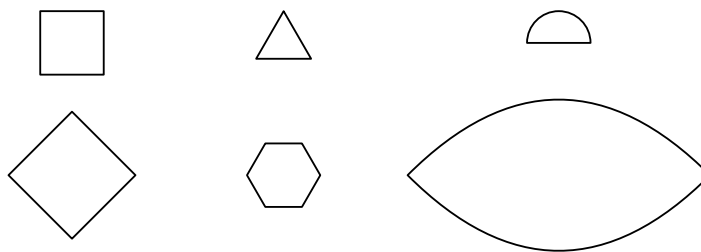


Рис. 11.

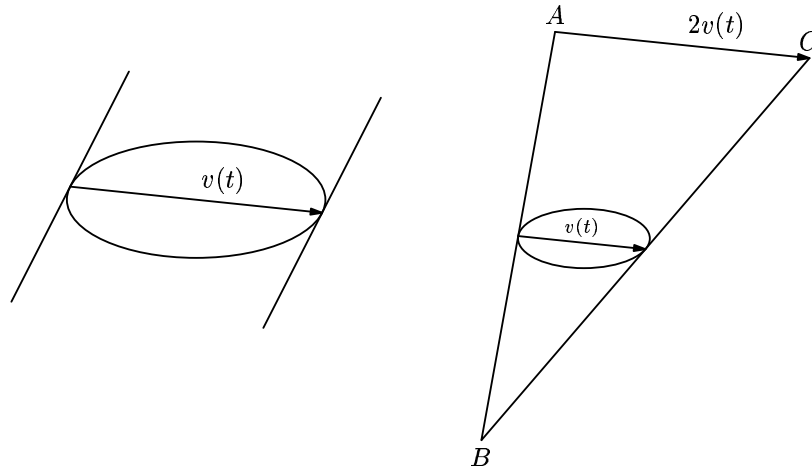


Рис. 12.

кривые в бесконечности  $\Gamma$  (последний столбец состоит из полукруга и из двух равных дуг парабол, пересекающихся под прямым углом).

Сейчас я объясню эти наблюдения. Для определенности, рассмотрим случай гладкой кривой  $\gamma$  (читателю предлагается самостоятельно изучить случай, когда  $\gamma$  — выпуклый многоугольник). Рассмотрим какую-нибудь параметризацию внешней бильярдной кривой  $\gamma(t)$  и обозначим через  $v(t)$  вектор, конец которого — точка  $\gamma(t)$ , а начало — точка касания опорной прямой к  $\gamma$ , параллельной вектору  $\gamma'(t)$  (см. левую часть рис. 12).

Для точек, очень удаленных от бильярдного стола, угол  $ABC$  на правой части рис. 12 пренебрежимо мал, поэтому искомая инвариантная кривая в бесконечности  $\Gamma(t)$  (определенная лишь с точностью до подобия) должна иметь касательный вектор, пропорциональный  $v(t)$ . Иными словами, мы хотим решить дифференциальное уравнение

$$\Gamma'(t) \sim v(t).$$

Ясно, что решение (если оно существует) единственно с точностью до гомотетии — ведь направление кривой  $\Gamma$  задано во всех ее точках вектором  $v$ .

Напомним, что векторное произведение  $a \times b$  — это ориентированная площадь параллелограмма, образованного векторами  $a$  и  $b$ . В частности,  $a \times b = 0$  тогда и только тогда, когда векторы  $a$  и  $b$  пропорциональны.

**УТВЕРЖДЕНИЕ.** *Решением дифференциального уравнения является следующая вектор-функция:*

$$\Gamma(t) = \frac{v'(t)}{v(t) \times v'(t)}.$$

**ДОКАЗАТЕЛЬСТВО.** Имеем:

$$\Gamma' = \frac{v''}{v \times v'} - \frac{v'(v \times v'')}{(v \times v')^2},$$



поэтому

$$v \times \Gamma' = \frac{v \times v''}{v \times v'} - \frac{(v \times v')(v \times v'')}{(v \times v')^2} = 0.$$

Следовательно,  $\Gamma'(t) \sim v(t)$ , что и требовалось.

Итак, у нас есть формула для кривой  $\Gamma$ . Остается объяснить закон Кеплера. Движение по кривой  $\Gamma(t)$  происходит со скоростью  $2v(t)$ , а секториальная скорость равна

$$2v(t) \times \Gamma(t) = 2 \frac{v \times v'}{v \times v'} = 2,$$

то есть не зависит от времени (снова подчеркнем, что значение константы не имеет особого смысла, так как все определено лишь с точностью до подобия).

Мы видим, что динамика внешнего бильярдного отображения (а точнее, его второй итерации) в бесконечности хорошо аппроксимируется непрерывным движением вдоль кривых, гомотетичных  $\Gamma$ . Важно понимать, однако, что это — всего лишь приближение к гораздо более сложной динамике. Тем не менее, описанная аппроксимация имеет важные следствия.

Первое следствие относится к случаю, когда кривая  $\gamma$  достаточно гладкая (насколько я знаю, достаточно шестикратной дифференцируемости). В этом случае, используя теорию КАМ (Колмогоров – Арнольд – Мозер), можно доказать, что отображение  $T_\gamma$  действительно имеет инвариантные кривые сколь угодно далеко от бильярдного стола (см. [5, 6]). Такая инвариантная кривая служит «оградой», которую не может пересечь орбита внешнего бильярдного отображения. Следовательно, ни одна орбита отображения  $T_\gamma$  не может уйти в бесконечность.

Второе следствие относится к случаю, когда бильярдный стол — выпуклый многоугольник. В этом случае  $\Gamma$  — центрально-симметричный выпуклый  $k$ -угольник (определенный с точностью до гомотетии). Каждой стороне  $\Gamma$  отвечает «время»  $t$  — отношение длины стороны к скорости  $v$ , с которой происходит движение вдоль этой стороны (эта скорость — одна из диагоналей многоугольника  $\gamma$ ; см. предыдущее обсуждение). Набор чисел  $\vec{t} = (t_1, \dots, t_k)$  определен с точностью до общего множителя. Внешний бильярдный стол называется *квазирациональным*, если все компоненты  $\vec{t}$  можно сделать рациональными (а тогда — и целыми).

Примером квазирационального многоугольника служит многоугольник, все вершины которого имеют целые координаты. Другой пример — правильный многоугольник, для которого  $\vec{t} = (1, \dots, 1)$ .

В статье [6] Ю. Мозер поставил следующий вопрос: может ли орбита внешнего бильярда уйти в бесконечность, если внешний бильярдный стол — выпуклый многоугольник? Частичный ответ дается следующей теоремой.

**ТЕОРЕМА.** *Если внешний бильярдный стол — квазирациональный многоугольник, то ни одна орбита внешнего бильярдного отображения не может уйти в бесконечность.*

Доказательство этой теоремы не использует «тяжелой артиллерии» современной математики, но тем не менее, совсем не просто. Читатель может найти

детали в статьях [4, 7, 3]. Полный ответ на вопрос Мозера до сих пор не известен (компьютерные эксперименты дают основания думать, что если внешний бильярдный стол — полукруг, то некоторые орбиты уходят в бесконечность).

Отметим следствие сформулированной теоремы: если внешний бильярдный стол — многоугольник, все вершины которого имеют целые координаты, то все орбиты внешнего бильярдного отображения периодические.

Действительно, целочисленность вершин означает, что каждая орбита внешнего бильярдного отображения дискретна, а теорема означает, что орбита ограничена. Ограниченная и дискретная орбита конечна, что и требуется.

Насколько я знаю, более простого доказательства сформулированного следствия, не использующего вышеуказанной теоремы, не известно; было бы интересно найти такое доказательство.

Еще один открытый вопрос, тоже относящийся к случаю, когда внешний бильярдный стол — выпуклый многоугольник: может ли внешнее бильярдное отображение вовсе не иметь периодических траекторий?

Отметим в заключение параграфа, что внешний бильярд имеет смысл и в геометрии Лобачевского, причем свойство сохранения площади остается в силе. Изучение таких бильярдных — интересная открытая задача.

## 5. МНОГОМЕРНЫЙ ВНЕШНИЙ БИЛЛИАРД

Внутренний бильярд определен в пространстве любой размерности, например, в обычном трехмерном пространстве. А как обстоит дело с внешним бильярдом?

Оказывается, внешний бильярд можно определить в любом четномерном пространстве (плоскость четномерна!). Вещественное пространство размерности  $2n$  можно рассматривать как комплексное пространство размерности  $n$ . Нам понадобится операция умножения на  $\sqrt{-1}$ , которую мы обозначим через  $J$  — это линейное отображение пространства  $\mathbb{C}^n$  в себя. На плоскости  $J$  — это просто поворот на  $\pi/2$  в положительном направлении, а в  $\mathbb{C}^n$  оператор  $J$  можно представлять себе как одновременный поворот каждого координатного пространства  $\mathbb{C}$  на  $\pi/2$ .

Внешний бильярдный стол — это так же, как и раньше, выпуклая ограниченная область. Обозначим ее границу — выпуклую замкнутую гиперповерхность в  $\mathbb{C}^n$ , через  $M$ . Внешнее бильярдное отображение можно было бы определить, если бы в каждой точке  $M$  была определена единственная касательная прямая. Проблема же в том, что таких прямых слишком много. Разрешается эта трудность так.

Обозначим через  $\nu$  внешний единичный нормальный вектор к  $M$  (предполагая, что  $M$  — гладкая строго выпуклая гиперповерхность). Вектор  $J(\nu)$  касается  $M$  и задает ориентированную касательную прямую в каждой точке  $M$ . Можно доказать, что через каждую точку  $A$  вне бильярдного стола проходит ровно две такие касательные прямые, причем одна из них,  $r$ , имеет направление к  $M$ , а другая,  $l$ , — от  $M$ . Теперь ясно, как определить внешнее бильярдное отображение: находим такую точку  $X \in M$ , что соответствующая прямая  $r$  проходит через  $A$ , и отражаем  $A$  в точке  $X$ , чтобы получить новую точку  $B = T(A)$  (сравните с началом параграфа 1).

Что данное определение «правильное», показывает следующий аналог свойства сохранения площади. Наряду с обычным скалярным произведением,  $n$ -мерное пространство обладает другой билинейной операцией — симплектической структурой. Симплектическая структура обобщает векторное произведение на плоскости, то есть ориентированную площадь параллелограмма. Обозначается симплектическая структура через  $\omega$  и задается следующей формулой:

$$\omega(a, b) = J(a) \cdot b,$$

где  $a$  и  $b$  — векторы, а точкой обозначено скалярное произведение. В отличие от последнего,  $\omega$  — кососимметричная операция:

$$\omega(a, b) = -\omega(b, a).$$

Так вот, для любого внешнего бильярдного стола, внешнее бильярдное отображение сохраняет симплектическую структуру. Эта теорема (доказанная в [8, 9]) имеет следствия, известные нам в случае плоскости. Например, у внешнего бильярдного отображения есть периодические траектории всех периодов, начиная с 3.

Тем не менее, о многомерном внешнем бильярде не известно почти ничего. Например, может ли орбита уходить в бесконечность? В многомерной ситуации теория КАМ больше не обеспечивает существования инвариантных гиперповерхностей. Другой интересный вопрос касается случая, когда внешний бильярдный стол — выпуклый многогранник (скажем, тетраэдр в четырехмерном пространстве). Этот вопрос тоже полностью открыт.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] *Арнольд В.* Геометрия сферических кривых и алгебра кватернионов // *Успехи Матем. Наук*, 1995. Т. 50, вып. 1. С. 3–68.
- [2] *Берже М.* Геометрия. М. Мир, 1984.
- [3] *Gutkin E., Simanyi N.* Dual polygonal billiards and necklace dynamics // *Comm. Math. Phys.*, 1991. V. 143. P. 431–450.
- [4] *Kolodziej R.* The antibilliard outside a polygon // *Bull. Pol. Acad. Sci.*, 1989. V. 37. P. 163–168.
- [5] *Moser J.* Stable and random motions in dynamical systems // *Ann. of Math. Stud.*, 1973. V. 77.
- [6] *Moser J.* Is the solar system stable? // *Math. Intell.*, 1978. V. 1. P. 65–71.
- [7] *Shaidenko A., Vivaldi F.* Global stability of a class of discontinuous dual billiards // *Comm. Math. Phys.*, 1987. V. 110. P. 625–640.
- [8] *Табачников С.* Дуальные бильярды // *Успехи Матем. Наук*, 1993. Т. 48, вып. 6. С. 75–102.
- [9] *Tabachnikov S.* Billiards // *SMF “Panoramas et Syntheses”*, 1995. No 1.
- [10] *Табачников С., Фукс Д.* Сегменты постоянной площади // *Квант*, 1990. №8. С. 26–31.

## Сколько есть периодических траекторий у трехмерного бильярда?

Ф. С. Дужин

Изучение замкнутых (периодических) траекторий бильярдных систем встречается в ряде математических задач. Например, замкнутые траектории бильярдных систем связаны с собственными значениями оператора Лапласа на соответствующей области, отвечающими за частоты звучания мембраны, имеющей форму рассматриваемого бильярда. Первым, кто занимался оценками количества периодических траекторий гладких бильярдных систем, был американский математик Джордж Биркгоф. Он рассматривал следующую задачу. Пусть на плоскости дана выпуклая область с гладкой границей. Внутри области движется бильярдный шар, отражаясь от ее границы по закону «угол падения равен углу отражения». Некоторые траектории движения шара могут быть периодическими: после  $p$  отражений он снова полетит по своему маршруту. Вопрос: какое может быть минимальное количество  $p$ -периодических траекторий?

Дж. Биркгоф доказал, что для простого  $p > 2$  существует по крайней мере  $p - 1$  геометрически различных  $p$ -периодических траекторий. Можно показать, что эта оценка точна: для любого простого  $p > 2$  существует такая выпуклая область, для которой количество  $p$ -периодических траекторий в точности равно  $p - 1$ .

Возникает вопрос: а что будет в трехмерном пространстве? Пусть в пространстве имеется выпуклая область с гладкой границей, внутри которой по описанному правилу движется шар. Предположим для простоты, что мы интересуемся только 3-периодическими траекториями. Можно показать, что таких геометрически различных траекторий имеется по крайней мере 4. (Строго говоря, это утверждение имеет место для области общего положения. Для областей специального вида некоторые из траекторий могут сливаться.) Доказательство основано на методах, связанных с теорией Морса.

Спрашивается: является ли эта оценка точной? Существует ли область в пространстве, для которой имеется ровно 4 периодические траектории периода 3? Если нет, то каково минимально возможное количество таких траекторий? Эти вопросы (и, конечно, аналогичные им для любого  $p$ ) остаются открытыми.

## Биллиардная динамическая система для числа $\pi$

Г. А. Гальперин

**ВВЕДЕНИЕ.** Имеется много разных методов вычисления числа  $\pi$ , известных как с древних времен, так и появившихся совсем недавно. Эти методы используют разнообразные изящные идеи — *геометрические* (вписывание и описывание многоугольников вокруг окружности), *теоретико-числовые* (теория цепных дробей дает приближение  $\pi \approx 355/113$  с точностью до одной миллионной, если ограничиваться дробями с трехзначными числителем и знаменателем), *аналитические* (с помощью рядов, интегралов и бесконечных произведений), *компьютерные* (которым в последнее время несть числа), и их многочисленные комбинации.

Кроме этих — *математических* — методов, с давних пор известен один *экспериментальный* способ определения числа  $\pi$  — так называемый «метод иглы Бюффона». В нем на разлинованную равноудаленными прямыми плоскость произвольно бросается игла, длина которой равна половине расстояния между соседними прямыми. (Так что игла либо не пересекает прямые, либо пересекает ровно одну при каждом бросании). Можно доказать, что отношение числа пересечений иглы с какой-нибудь линией к общему числу бросаний стремится к  $\pi$  при увеличении числа бросаний до бесконечности. Нужно сделать очень много испытаний, чтобы получить более-менее приличную точность приближения полученной дроби к  $\pi$ , а кроме того, при эксперименте надо внимательно следить, чтобы бросание иглы было «равновероятным»: метод иглы Бюффона существенным образом базируется на методах теории вероятностей.

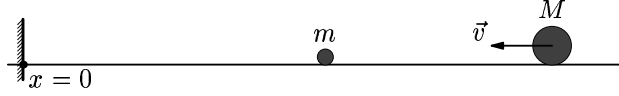
Здесь мы расскажем об одной новой идее определения числа  $\pi$  — *биллиардной*, придуманной автором при подготовке научного доклада в одном из американских университетов. Когда биллиардный метод был объявлен аудитории, никто ему поначалу не поверил, но затем последовало доказательство, простота которого сразу убедила всех сомневающихся. Позже автор выделил этот метод в отдельную тему и сделал несколько докладов в других университетах, с одной и той же реакцией аудитории: сначала полное недоверие, затем полная очевидность.

Биллиардный метод дает возможность определить  $\pi$  с *произвольной* точностью, т. е. узнать его любую наперед заданную цифру. Этот метод — детерминированный: все, что нужно сделать, это «запустить» биллиардную систему и подсчитать число ударов в ней.

**ПРОЦЕДУРА.** Положим на положительную числовую полуось  $x \geq 0$  два биллиардных шарика с массами  $m$  и  $M \geq m$ , и будем предполагать, что в начале координат  $x = 0$  расположена абсолютно упругая стенка, отражающая налетающий на нее шарик. При отражении от стенки скорость шарика меняется на строго противоположную. Размеры шариков несущественны, и для простоты мы будем считать их точечными частицами.

Фиксируем натуральное число  $N$ . Следующая процедура позволяет определить любое наперед заданное количество  $N$  последовательных цифр числа  $\pi$ :

- (1) Массы  $m$  и  $M$  подбираем так, что  $M/m = 100^N$ ;
- (2) Шар  $m$  располагаем между стенкой  $x = 0$  и шаром  $M$ ;
- (3) Запускаем шар  $M$  в сторону шара  $m$  с произвольной скоростью;
- (4) Подсчитываем общее количество ударов в системе (т. е. число столкновений между шарами плюс число отражений шара  $m$  от стенки);
- (5) Записываем полученное число в десятичной системе и обозначаем его через  $\pi(N)$ .



**ТЕОРЕМА.** А) Число ударов в описанной динамической системе всегда конечно и не зависит от начальных положений шариков и начальной скорости шара  $M$ .

Б) Число  $\pi(N)$  ударов в системе равно

$$\pi(N) = \lfloor \pi \cdot 10^N \rfloor = 3 \underbrace{14159265358979323 \dots}_{N \text{ десятичных знаков } \pi} \quad (1)$$

**Идея доказательства теоремы.** Она ясна из статьи «Бильярды и упругие столкновения частиц и шаров» (с. 65–99) и состоит в сведении описанной динамической системы к бильярду в угле величиной  $\alpha = \arctg(10^{-N})$ . Траектория этого бильярда оказывается параллельной одной из сторон угла  $\alpha$ , поэтому число ее отражений от сторон угла всегда конечно и равно  $\pi(N) = \lfloor \pi/\alpha \rfloor$  (если  $N \geq 1$ ). Этим доказана часть А. Воспользовавшись рядом Тейлора для  $\arctg x$  и «тонкой структурой» числа  $\pi$  — равенством  $\lfloor \pi/\arctg(10^{-N}) \rfloor = \lfloor \pi/10^{-N} \rfloor$ , получаем основной результат теоремы — часть Б:

$$\pi(N) = \left\lfloor \frac{\pi}{\arctg(10^{-N})} \right\rfloor = \left\lfloor \frac{\pi}{10^{-N}} \right\rfloor. \quad (2)$$

Полное доказательство будет приведено в следующем номере.

**Благодарности.** Прежде всего хочу поблагодарить профессора Eastern Illinois University Айру Розенхольца, в течение долгого времени настаивавшего на публикации результата этой статьи. Я признателен также руководителям и участникам всех математических и физических семинаров и коллоквиумов, на которых мне довелось выступить с этим результатом в 1996–2000 гг.

## Доказательство теоремы Понселе по Дарбу

В. В. Прасолов

Теорема Понселе состоит в следующем. Пусть заданы две коники  $\Gamma_1$  и  $\Gamma_2$ , не касающиеся друг друга. В комплексной ситуации из каждой точки  $A_1 \in \Gamma_1$ , отличной от точек пересечения  $\Gamma_1$  и  $\Gamma_2$ , можно провести ровно две касательных к  $\Gamma_2$ . Поэтому можно построить ломаную  $A_1A_2 \dots A_nA_{n+1}$  так, чтобы её вершины  $A_i$  лежали на  $\Gamma_1$ , а прямые  $A_1A_2, A_2A_3, \dots, A_nA_{n+1}$  касались  $\Gamma_2$  (подразумевается, что  $A_{i+2} \neq A_i$ ). Предположим, что для некоторой точки  $A_1$  точка  $A_{n+1}$  совпала с  $A_1$ , т. е. получилась замкнутая  $n$ -звенная ломаная. Тогда и при любом выборе точки  $A_1 \in \Gamma_1$  тоже получится замкнутая  $n$ -звенная ломаная.

Одно из наиболее понятных доказательств теоремы Понселе предложил Дарбу [Д]. Он заметил, что теорему Понселе удобно доказывать, используя систему координат, в которой положение точки  $A$  задаётся двумя касательными к некоторой фиксированной конике, проведёнными из точки  $A$ . В качестве фиксированной коники можно выбрать произвольную конику, поскольку проективным преобразованием любую (невырожденную) конику можно перевести в любую другую. Выберем конику, которая в однородных координатах  $(x : y : z)$  задаётся уравнением  $y^2 = xz$ . Точки этой коники имеют координаты  $(1 : t : t^2)$ ; при  $t = \infty$  получаем точку  $(0 : 0 : 1)$ . Касательная в точке  $(x_0 : y_0 : z_0)$  задаётся уравнением  $2y_0y = x_0z + z_0x$ , т. е.  $t^2x - 2ty + z = 0$ , где  $t$  — параметр, соответствующий точке  $(x_0 : y_0 : z_0)$ . Чтобы найти однородные координаты точки пересечения в точках с параметрами  $t_1$  и  $t_2$ , нужно решить систему линейных уравнений

$$t_i^2x - 2t_iy + z = 0, \quad i = 1, 2.$$

Решая её, получаем

$$2y = x(t_1 + t_2), \quad z = xt_1t_2.$$

**ПРИМЕР 1.** Исходная коника  $y^2 = xz$  в координатах  $(t_1, t_2)$  задаётся уравнением  $(t_1 + t_2)^2 = 4t_1t_2$ , т. е.  $(t_1 - t_2)^2 = 0$ .

**ПРИМЕР 2.** Касательная к исходной конике в точке  $(1 : \alpha : \alpha^2)$  задаётся уравнением  $(t_1 - \alpha)(t_2 - \alpha) = 0$ .

Нам потребуется описание семейства кривых минимальной степени, проходящих через все точки пересечения  $n$  прямых общего положения на плоскости. Каждая из этих прямых пересекает такую кривую в  $n - 1$  точках, поэтому степень кривой не может быть меньше  $n - 1$ .

ЛЕММА (ДАРБУ). Уравнение любой кривой степени  $n-1$ , проходящей через все точки пересечения прямых, заданных линейными уравнениями  $p_1 = 0, \dots, p_n = 0$ , имеет вид

$$p_1 \cdot \dots \cdot p_n \left( \frac{\lambda_1}{p_1} + \dots + \frac{\lambda_n}{p_n} \right) = 0,$$

где  $\lambda_1, \dots, \lambda_n$  — некоторые константы (предполагается, что все точки пересечения данных прямых попарно различны).

ДОКАЗАТЕЛЬСТВО. Пусть рассматриваемая кривая задаётся уравнением  $C = 0$ . Возьмём прямую  $l$ , пересекающую прямые  $p_1, \dots, p_n$  в  $n$  различных точках  $x_1, \dots, x_n$ , и выберем числа  $\lambda_i$  так, чтобы в точках  $x_i$  выполнялось равенство

$$C - p_1 \cdot \dots \cdot p_n \left( \frac{\lambda_1}{p_1} + \dots + \frac{\lambda_n}{p_n} \right) = 0. \quad (1)$$

Для этого нужно положить

$$\lambda_i = \frac{C(x_i)}{p_1(x_i) \cdot \dots \cdot \hat{p}_i(x_i) \cdot \dots \cdot p_n(x_i)}.$$

Тогда равенство (1) выполняется в  $n$  различных точках каждой прямой  $p_i$ . Если равенство (1) выполняется не для всех точек плоскости, то оно задаёт кривую степени не выше  $n-1$ . Но эта кривая должна содержать все прямые  $p_i$ , поэтому её степень не может быть меньше  $n$ . Приходим к противоречию, поэтому  $C = p_1 \cdot \dots \cdot p_n \left( \frac{\lambda_1}{p_1} + \dots + \frac{\lambda_n}{p_n} \right) = 0$ .

Выберем в качестве прямых  $p_1, \dots, p_n$  касательные к конике  $y^2 = xz$  в точках с параметрами  $\alpha_1, \dots, \alpha_n$ . Согласно лемме Дарбу кривая степени  $n-1$ , проходящая через точки пересечения этих прямых, задаётся уравнением

$$\sum_{i=1}^n \frac{\lambda_i}{(t_1 - \alpha_i)(t_2 - \alpha_i)} = 0.$$

Тождество

$$\frac{\lambda_i}{t_1 - \alpha_i} - \frac{\lambda_i}{t_2 - \alpha_i} = \frac{(t_2 - t_1)\lambda_i}{(t_1 - \alpha_i)(t_2 - \alpha_i)}$$

показывает, что после умножения на  $t_2 - t_1$  это уравнение можно записать в виде

$$\sum_{i=1}^n \frac{\lambda_i}{t_1 - \alpha_i} = \sum_{i=1}^n \frac{\lambda_i}{t_2 - \alpha_i}.$$

Напомним, что точки, для которых  $t_1 = t_2$ , лежат на исходной конике  $y^2 = xz$ . Пусть

$$\sum_{i=1}^n \frac{\lambda_i}{t - \alpha_i} = \frac{P_{n-1}(t)}{Q_n(t)} = R(t),$$

где  $Q_n(t) = (t - \alpha_1) \cdot \dots \cdot (t - \alpha_n)$  и  $P_{n-1}(t)$  — многочлен степени не выше  $n-1$ .



Равенство  $R(t_1) = R(t_2)$  эквивалентно равенству  $R_\mu(t_1) = R_\mu(t_2)$ , где

$$R_\mu(t) = \frac{P_{n-1}(t)}{Q_n(t) + \mu P_{n-1}(t)}.$$

Это означает, что для всех  $\mu$  точки пересечения касательных в точках с параметрами  $\alpha_1(\mu), \dots, \alpha_n(\mu)$ , где  $\alpha_1(\mu), \dots, \alpha_n(\mu)$  — корни многочлена  $Q_n(t) + \mu P_{n-1}(t)$ , лежат на одной и той же кривой степени  $n - 1$ . Действительно, рациональную функцию  $R_\mu(t)$  можно представить в виде

$$R_\mu(t) = \sum_{i=1}^n \frac{\lambda_i(\mu)}{t_1 - \alpha_i(\mu)};$$

здесь предполагается, что числа  $\alpha_1(\mu), \dots, \alpha_n(\mu)$  попарно различны.<sup>1)</sup>

Для  $n = 3$  и  $n = 4$  теорему Понселе теперь легко доказать. При  $n = 3$  берём треугольник, вписанный в конику  $\Gamma_1$  и описанный вокруг коники  $\Gamma_2$ . Существует однопараметрическое семейство троек прямых, которые касаются коники  $\Gamma_2$  и точки пересечения которых лежат на конике  $\Gamma_1$ . При  $n = 3$  все точки пересечения прямых являются вершинами рассматриваемой ломаной, а кроме того, никаких проблем с возможным слиянием точек  $\alpha_i(\mu)$  не возникает по геометрическим соображениям.<sup>2)</sup>

Для  $n = 4$  тоже нет проблем с возможным слиянием точек  $\alpha_i(\mu)$ , потому что в этом случае могут слиться только точки  $\alpha_i(\mu)$ , соответствующие соседним звеньям ломаной, а тогда получается касание коник  $\Gamma_1$  и  $\Gamma_2$ . Но для  $n = 4$  уже не все точки пересечения прямых являются вершинами рассматриваемой ломаной: для четырёхугольника помимо вершин есть ещё и точки пересечения продолжений сторон. Через эти 6 точек пересечения сторон четырёхугольника и их продолжений нужно провести кривую степени 3, причём кривая, проходящая через вершины четырёхугольника, должна быть данной коникой  $\Gamma$ . Такая кривая должна быть объединением коники  $\Gamma$  и прямой  $l$ , соединяющей точки пересечения продолжений сторон. Существует однопараметрическое семейство четвёрок прямых, которые касаются данной коники и точки пересечения которых лежат на  $\Gamma$  и  $l$ . Вообще говоря, вершина четырёхугольника могла бы при движении «переехать» с коники  $\Gamma$  на прямую  $l$ . Но  $\Gamma$  и  $l$  не имеют вещественных точек самопересечения.

При  $n \geq 5$  проблемы с возможным слиянием точек  $\alpha_i(\mu)$  и с переезжанием вершин ломаной с одной ветви кривой на другую становятся более сложными. Поэтому подойдём к задаче по-другому.

Рассмотрим ломаную  $A_1 A_2 A_3 \dots$ , вершины которой лежат на конике  $\Gamma_1$ , а звенья касаются коники  $\Gamma_2$ . Мы снова будем задавать положение точки координатами, связанными с коникой  $\Gamma_2$ . Будем предполагать, что точка  $A_i$  имеет

<sup>1)</sup>Совпадение чисел  $\alpha_i(\mu)$  и  $\alpha_j(\mu)$  при  $i \neq j$  соответствует тому, что при данном значении  $\mu$  вместо замкнутой  $n$ -звенной ломаной появляется  $(n/k)$ -звенная ломаная, которая обходит  $k$  раз. В действительности такого не бывает, но это требует отдельного доказательства.

<sup>2)</sup>В условии теоремы Понселе предполагается, что коники  $\Gamma_1$  и  $\Gamma_2$  не касаются. Из слияния двух точек  $\alpha_i(\mu)$ , соответствующих соседним звеньям ломаной, вытекает касание коник. А в треугольнике любая пара сторон соседняя.

координаты  $(\alpha_{i-1}, \alpha_i)$ . Коника  $\Gamma_1$  задаётся уравнением  $f(t_1, t_2) = 0$ , где

$$f(t_1, t_2) = at_1^2t_2^2 + bt_1t_2(t_1 + t_2) + c(t_1^2 + t_2^2) + dt_1t_2 + e(t_1 + t_2) + f.$$

Выясним, каким уравнением задаётся кривая, на которой лежат точки пересечения прямых  $A_{i-1}A_i$  и  $A_{i+1}A_{i+2}$ . Для этого нужно исключить  $\alpha_i$  из соотношений  $f(\alpha_{i-1}, \alpha_i) = 0$  и  $f(\alpha_i, \alpha_{i+1}) = 0$ .

Функцию  $f$  можно представить в виде

$$\begin{aligned} f(t_1, t_2) &= \alpha(t_1)t_2^2 + \beta(t_1)t_2 + \gamma(t_1) = \\ &= \alpha(t_2)t_1^2 + \beta(t_2)t_1 + \gamma(t_2). \end{aligned}$$

Чтобы исключить  $\alpha_i$  из уравнений

$$\begin{aligned} \alpha(\alpha_{i-1})\alpha_i^2 + \beta(\alpha_{i-1})\alpha_i + \gamma(\alpha_{i-1}) &= 0, \\ \alpha(\alpha_{i+1})\alpha_i^2 + \beta(\alpha_{i+1})\alpha_i + \gamma(\alpha_{i+1}) &= 0, \end{aligned}$$

нужно составить результат

$$\begin{vmatrix} \alpha(\alpha_{i-1}) & \beta(\alpha_{i-1}) & \gamma(\alpha_{i-1}) & 0 \\ 0 & \alpha(\alpha_{i-1}) & \beta(\alpha_{i-1}) & \gamma(\alpha_{i-1}) \\ \alpha(\alpha_{i+1}) & \beta(\alpha_{i+1}) & \gamma(\alpha_{i+1}) & 0 \\ 0 & \alpha(\alpha_{i+1}) & \beta(\alpha_{i+1}) & \gamma(\alpha_{i+1}) \end{vmatrix} = \Phi(\alpha_{i-1}, \alpha_{i+1}).$$

Точки пересечения прямых  $A_{i-1}A_i$  и  $A_{i+1}A_{i+2}$  лежат на кривой, заданной уравнением  $\Phi(t_1, t_2) = 0$ . Непосредственно из определения видно, что  $\Phi(t_1, t_2) = \Phi(t_2, t_1)$  и  $\Phi$  делится на  $t_1 - t_2$ , поэтому  $\Phi$  делится на  $(t_1 - t_2)^2$ . Рассматриваемый результат является многочленом степени 4 (по каждой переменной), поэтому получаем симметричное квадратичное соотношение  $f_1(\alpha_{i-1}, \alpha_{i+1}) = 0$ , которое задаёт конику.

Покажем, что и при любом  $k$  точки пересечения прямых  $A_{i-1}A_i$  и  $A_{i+k-1}A_{i+k}$  будут лежать на одной конике. Предположим, что уже известны симметричные квадратичные (по каждой переменной) соотношения

$$f(\alpha_{i-1}, \alpha_i) = 0, \quad f_1(\alpha_{i-1}, \alpha_{i+1}) = 0, \quad \dots, \quad f_{k-1}(\alpha_{i-1}, \alpha_{i+k-1}) = 0.$$

Мы хотим получить соотношение  $f_k(\alpha_{i-1}, \alpha_{i+k}) = 0$ . Для этого нужно исключить  $\alpha_{i+k-1}$  из системы уравнений  $f_{k-1}(\alpha_{i-1}, \alpha_{i+k-1}) = 0$ ,  $f(\alpha_{i+k-1}, \alpha_{i+k}) = 0$ . В результате получим соотношение 4-й степени  $\Phi(\alpha_{i-1}, \alpha_{i+k}) = 0$ . Функция  $f(\alpha_{i+k-1}, t)$  обращается в нуль не только при  $t = \alpha_{i+k}$ , но и при  $t = \alpha_{i+k-2}$ . Поэтому  $\Phi(\alpha_{i-1}, \alpha_{i+k-2}) = 0$ . Это означает, что многочлен  $\Phi(x, y)$  делится на  $f_{k-2}(x, y)$ , т.е. делится на соотношение, связывающее  $\alpha_{i-1}$  и  $\alpha_{i+k-2}$ . Остаётся проверить, что многочлен  $f_k = \Phi/f_{k-2}$  симметричен.

Чтобы симметричность соотношения, связывающего  $\alpha_{i-1}$  и  $\alpha_{i+k}$ , была очевидна, можно это соотношение получить по-другому. Будем последовательно исключать  $\alpha_i, \alpha_{i+1}, \dots, \alpha_{i+k-1}$  из соотношений  $f(\alpha_{i-1}, \alpha_i) = 0, f(\alpha_i, \alpha_{i+1}) = 0, \dots, f(\alpha_{i+k-1}, \alpha_{i+k}) = 0$ , причём каждый раз полученный многочлен будем сокращать на  $f_j$ . Если воспользоваться симметричностью многочленов  $f, f_1, \dots, f_{k-1}$  и обратить последовательность вычислений, т.е. исключать  $\alpha_{i+k-1}, \dots, \alpha_{i+1}, \alpha_i$ , из  $f(\alpha_{i+k-1}, \alpha_{i+k}) = 0, \dots, f(\alpha_i, \alpha_{i+1}) = 0, f(\alpha_{i-1}, \alpha_i) = 0$ , то в результате получим тот же самый многочлен  $f_k$ . Поэтому многочлен  $f_k$  симметричен.

Теперь при  $n \geq 5$  теорема Понселе легко доказывается, поскольку точки  $A_1, \dots, A_5$  однозначно задают конику, а значит, точки пересечения прямых  $A_i A_{i+1}$  и  $A_{i+n-1} A_{i+n} = A_{i+n-1} A_i$  лежат на той же самой конике.

Если  $n = 2m$ , то кривая, по которой движутся точки пересечения продолжений звеньев вписанно-описанной ломаной, должна состоять не из  $m$  коник, а из  $m - 1$  коник и одной прямой, поскольку степень этой кривой равна  $n - 1$ . Поясним, почему одна из коник (и какая именно) вырождается в прямую.

Уравнение  $f_k(\alpha_{i-1}, x) = 0$  имеет корни  $\alpha_{i \pm k}$ . Если  $k = m$ , то эти корни совпадают, поэтому многочлен  $f_m$  должен быть квадратом симметричной линейной функции.

#### СПИСОК ЛИТЕРАТУРЫ

[Д] Дарбу Г. Принципы аналитической геометрии. Л.-М.: ГОНТИ, 1938.

## Гомеоидная плотность и теорема Понселе

А. А. Панов

Д. А. Панов

Теорема Понселе — одна из самых привлекательных в проективной геометрии. Известно несколько вариантов ее доказательства, но, как сказано в [1] и повторено в [2], «все они достаточно длинны и неочевидны». Возможно, хотя бы одно из предложенных здесь доказательств будет свободно по крайней мере от одного из этих недостатков. Напомним формулировку теоремы Понселе.

**ТЕОРЕМА ПОНСЕЛЕ.** Пусть имеются две коники  $C_0$  и  $C_1$  и пусть  $A_1A_2\dots A_{n+1}$  — ломаная, вершины которой лежат на  $C_0$ , а ребра касаются  $C_1$ . Тогда, если для некоторой точки  $A_1$  эта  $n$ -звенная ломаная замкнута, то она будет замкнутой и при любом другом выборе точки  $A_1$  на конике  $C_0$ .

Заметим, что конструкция, рассматриваемая в теореме Понселе, имеет естественный характер и возникает в определенных механических и оптических задачах. Например, при движении светового луча в плоском эллиптическом резонаторе прямолинейные отрезки луча имеют огибающую (на оптическом языке — каустику), представляющую собой конику, софокусную с ограничивающим эллипсом [3, 4].

Мы определяем гомеоидную плотность на эллипсе как аффинный образ равномерной плотности на окружности. При этом эллипс наследует многие свойства окружности. Например, эллипс, снабженный гомеоидной плотностью, не притягивает на плоскости и касательные к эллипсу, проведенные из одной точки, становятся (в некотором смысле) равными. Гомеоидная плотность также допускает и естественную физическую интерпретацию. Заряд (здесь имеется в виду «плоский» заряд, для которого кулоновская сила обратно пропорциональна расстоянию), помещенный на проводящий эллипс, распределится на нем с гомеоидной плотностью. Введение такой плотности позволяет работать с парой коник

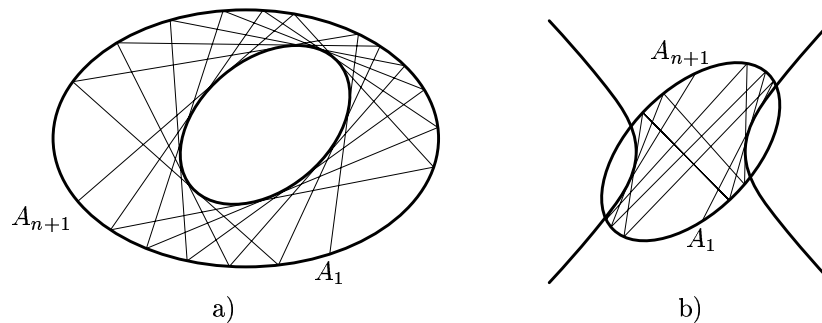


Рис. 1.

$C_0$  и  $C_1$  как с парой окружностей и апеллировать к известному доказательству теоремы Понселе для пары окружностей [2].

Гомеоидную плотность можно ввести и на других кониках. Сначала мы это сделаем для пары прямых, а потом получим известное аналитическое представление плотности [5] и для остальных коник. В [5] дано простое геометрическое определение гомеоидной плотности на кониках. Мы покажем, что гомеоидная плотность на кониках имеет и другое естественное геометрическое представление, а именно, представление двояко-линейчатой поверхностью, которое позволяет сформулировать следующий пространственный вариант теоремы Понселе [6], [7].

**ТЕОРЕМА.** Пусть имеются две квадрики общего положения  $Q_0$  и  $Q_1$  в  $CP^3$ . Через  $L'Q_1$  и  $L''Q_1$  обозначим две системы линейных образующих квадрики  $Q_1$ . Пусть  $B_1B_2 \dots B_{2n+1}$  — ломаная, вершины которой лежат на пересечении  $T = Q_0 \cap Q_1$ , а ребра принадлежат (чередуясь) системам линейных образующих  $L'Q_1, L''Q_1$ . Тогда, если для некоторой точки  $B_1$  эта  $2n$ -звенная ломаная замкнута, то она будет замкнутой и при любом другом выборе точки  $B_1$  на  $T$ .

Теорема специально сформулирована таким образом, чтобы непосредственно была видна ее связь с теоремой Понселе. После ее доказательства мы приведем более общий и симметричный вариант для пучка квадрик в  $CP^3$ . В конце сделано несколько добавлений о связи между гомеоидностью и интегрируемостью, о компьютерных экспериментах с эллиптическим бильярдом и о других вопросах.

## ГОМЕОИДНАЯ ПЛОТНОСТЬ — ПЕРВОЕ ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ ПОНСЕЛЕ

Для доказательства теоремы Понселе нам понадобятся следующие факты:

- ▷ окружность на плоскости не притягивает,
- ▷ эллипс на плоскости не притягивает,
- ▷ касательные к эллипсу, проведенные из одной точки, равны.

Поскольку в традиционном понимании все эти утверждения одинаково неверны, они нуждаются в некоторой интерпретации.

### *Окружность на плоскости не притягивает*

Этот факт верен по крайней мере наполовину. Известно, что в двумерном мире сила притяжения обратно пропорциональна расстоянию. В этой ситуации масса, равномерно распределенная вдоль окружности, создает внутри нее нулевое поле силы тяжести. Выполняется локальный баланс сил — две противоположные дуги, отсекаемые хордами, проходящими через данную точку, действуют на эту точку в точности с противоположными силами. Это утверждение следует из подобия инфинитезимальных треугольников на рисунке 2, в частности, из соотношения

$$\frac{d\varphi_1}{l_1} = \frac{d\varphi_2}{l_2}. \quad (1)$$

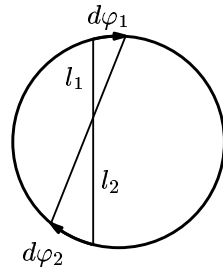


Рис. 2.

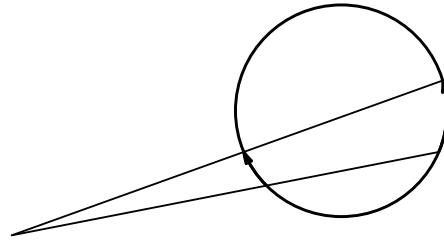


Рис. 3.

Посмотрим, как нужно изменить закон тяготения, чтобы и вне окружности поле было нулевым. Сначала заметим, что инфинитезимальные треугольники на рисунке 3 тоже подобны и для них после введения соответствующих обозначений выполняется то же самое условие (1). Сохраним для сил закон обратной пропорциональности и будем считать, что дуги, расположенные по разные стороны от точки (рис. 2), притягивают ее. Но, если дуги расположены по одну сторону от точки (рис. 3), то считаем, например, что дальняя притягивает, а ближняя отталкивает. Соотношение (1) гарантирует, что при таком понимании окружность на плоскости на самом деле не притягивает.

**ЗАМЕЧАНИЕ ОБ ОРИЕНТАЦИИ.** В дальнейшем нам придется следить за направлением инфинитезимального сдвига  $d\varphi$ . Поэтому вернемся к рисунку 2. Представим, что одна хорда на рисунке получается из другой малым поворотом вокруг их общей точки. Тогда этот поворот индуцирует согласованную ориентацию на дугах  $d\varphi_1$  и  $d\varphi_2$ . Если же, как на рисунке 3, секущая получается малым поворотом из секущей, то ориентация высекаемых ими дуг противоположна.

*Гомеоидная плотность — и эллипс тоже не притягивает*

Эллипс является аффинным образом единичной окружности. При аффинном отображении равномерная плотность на окружности по правилу «масса образа равна массе прообраза» порождает *гомеоидную* плотность на эллипсе [5]. Последнюю тоже будем обозначать  $d\varphi$ . Поскольку при аффинном преобразовании отношение длин отрезков, лежащих на одной прямой, сохраняется, для гомеоидной плотности выполняется то же самое соотношение (1). Отсюда, в частности, следует, что эллипс, снабженный гомеоидной плотностью, на плоскости тоже не притягивает.

*Гомеоидная плотность — и касательные к эллипсу равны*

Обозначим гомеоидную плотность на эллипсе на этот раз  $d\psi$  и запишем ее в виде  $d\psi = k(\psi) ds$ , где  $ds$  длина дуги эллипса. Из определения гомеоидной плотности следует, что величина  $1/k$  — это коэффициент растяжения касательного направления в соответствующей точке окружности при аффинном отображении этой окружности на эллипс. Пусть  $l(\psi_1)$  и  $l(\psi_2)$  длины касательных, проведенных к эллипсу из одной точки. Тогда величины  $l(\psi_1)k(\psi_1)$  и  $l(\psi_2)k(\psi_2)$  равны

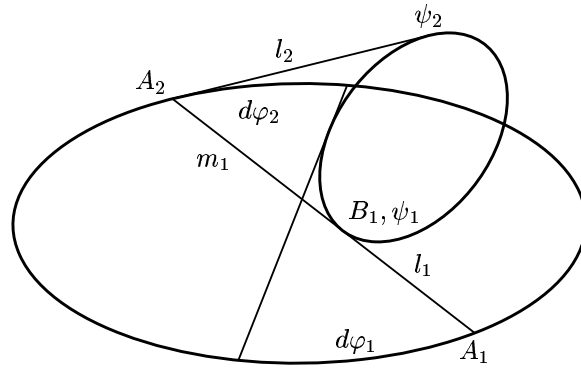


Рис. 4.

между собой, как длины соответствующих касательных к окружности —

$$l(\psi_1)k(\psi_1) = l(\psi_2)k(\psi_2). \quad (2)$$

Если теперь назвать величину  $l(\psi)k(\psi)$  длиной касательной к эллипсу, то, на самом деле, касательные к эллипсу, проведенные из одной точки, будут равны.

Все вспомогательные предложения рассмотрены, можно приступать к доказательству.

#### Доказательство теоремы Понселе

Будем считать, что обе коники являются эллипсами. На конике  $C_0$  введем гомеоидную плотность  $d\varphi$ , которая создает нулевое силовое поле. То есть такую плотность, для которой выполняется соотношение (1). На конике  $C_1$  введем гомеоидную плотность  $d\psi = k(\psi) ds$ , которая гарантирует равенство касательных в виде (2).

Пусть  $A_1A_2$  — хорда коники  $C_0$ , касающаяся коники  $C_1$ , причем  $A_1 = A(\varphi_1)$  и  $A_2 = A(\varphi_2)$  (рис. 4). Через  $l_1$  обозначим отрезок от точки  $A_1$  до точки касания  $B_1 = B(\psi_1)$ . Через  $l_2$  обозначим аналогичный отрезок, выходящий из точки  $A_2$ . Наконец, через  $m_1$  обозначим отрезок от точки касания  $B_1$  до точки  $A_2$  и положим  $k_1 = k(\psi_1)$ ,  $k_2 = k(\psi_2)$ .

Сдвинув начальную точку  $A_1$  на  $d\varphi = d\varphi_1$ , из соотношения (1) получаем

$$\frac{d\varphi_1}{l_1} = \frac{d\varphi_2}{m_1}.$$

В свою очередь из соотношения (2) получаем

$$m_1k_1 = l_2k_2.$$

Откуда следует, что

$$\frac{d\varphi_1}{l_1k_1} = \frac{d\varphi_2}{l_2k_2}.$$

Сначала развернем это равенство до

$$\frac{d\varphi_1}{l_1k_1} = \frac{d\varphi_2}{l_2k_2} = \dots = \frac{d\varphi_{n+1}}{l_{n+1}k_{n+1}}.$$

Потом снова сократим до

$$\frac{d\varphi_1}{l_1 k_1} = \frac{d\varphi_{n+1}}{l_{n+1} k_{n+1}}.$$

Если ломаная  $A_1 A_2 \dots A_{n+1}$  замкнута, то  $l_{n+1} = l_1$ , и  $k_{n+1} = k_1$ . Откуда следует, что

$$d\varphi_{n+1} = d\varphi_1. \quad (3)$$

Это соотношение гарантирует лишь равенство абсолютных величин сдвигов, но пока не совпадение их направлений.

Пусть  $\sigma$  — число точек касания ломаной с коникой  $C_1$ , лежащих вне эллипса  $C_0$  (число *внешних* касаний). Сделанное ранее замечание об ориентации показывает, что для ориентированных дуг соотношение (3) может быть записано в виде

$$d\varphi_{n+1} = (-1)^\sigma d\varphi_1.$$

Легко понять, что при любом расположении эллипсов число  $\sigma$  внешних касаний для замкнутой ломаной обязательно четно. Отсюда следует, что равенство (3) верно и с учетом ориентации. То есть куда сдвинется точка  $A_1$ , ровно туда же сдвинется точка  $A_{n+1}$  и ломаная останется замкнутой. Теорема Понселе почти доказана.

#### *Невырожденные коники приводятся к паре эллипсов*

Осталось показать, что, действительно, доказательство достаточно провести только для пары эллипсов. Конструкция, описанная в теореме Понселе, инвариантна относительно проективных преобразований, так как при проективном преобразовании коники переходят в коники и касательные в касательные. Если на проективной плоскости существует прямая, не пересекающая ни одну из коник  $C_0, C_1$ , то отправив эту прямую на бесконечность, сразу получим пару эллипсов.

Рассмотрим случай, когда любая прямая пересекает хотя бы одну из коник (речь идет о расположении, подобном рис. 1 b). Воспользуемся тем, что теорема Понселе инвариантна относительно еще одного типа преобразований — относительно перехода к двойственной конфигурации [1, 2]. При таком преобразовании коники лишь обмениваются ролями — образ ломаной будет вписан в образ  $C_1$  и будет касаться образа  $C_0$ . Но главное, что появляются прямые, не пересекающие образы коник, и значит опять дело можно свести к паре эллипсов.

### ГОМЕОИДНАЯ ПЛОТНОСТЬ — АНАЛИТИЧЕСКОЕ И ГЕОМЕТРИЧЕСКОЕ ПРЕДСТАВЛЕНИЯ

Гомеοидную плотность, как сейчас будет показано, можно ввести не только на эллипсе, но и на других кониках. Так что изложенное выше доказательство на самом деле проходит для любой пары коник и можно обойтись без каких-либо дополнительных преобразований.

#### *Две прямые не притягивают*

Пусть коника представлена двумя прямыми  $L_1$  и  $L_2$ , пересекающимися в точке  $O$  (рис. 5). Возьмем произвольную точку  $P$  на плоскости и выпустим из нее



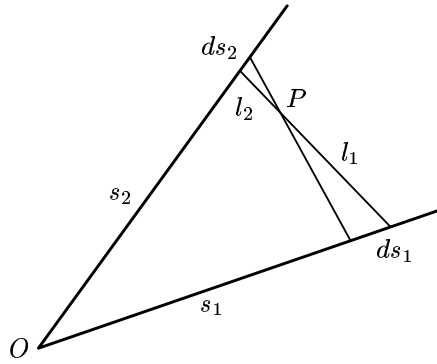


Рис. 5.

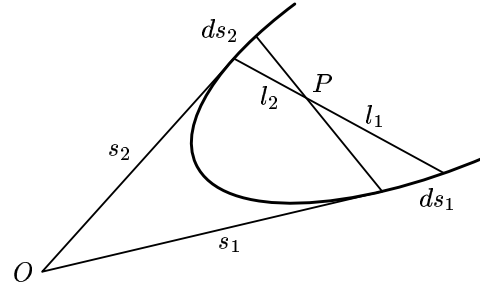


Рис. 6.

две близкие прямые. Они пересекут  $L_1$  и  $L_2$  по отрезкам  $ds_1$  и  $ds_2$ . Расстояния от точки  $P$  до этих отрезков обозначим через  $l_1$  и  $l_2$ , а расстояния от точки  $O$  до них обозначим  $s_1$  и  $s_2$ . Легко убедиться, что для данной конфигурации выполняется соотношение

$$\frac{ds_1/s_1}{l_1} = \frac{ds_2/s_2}{l_2}, \quad (4)$$

которое показывает, что после введения плотности  $d\varphi = ds/s$  прямые на плоскости не притягивают.

*Равенство касательных и отсутствие притяжения — две стороны одной медали*

При обсуждении ситуации с эллипсом, мы показали, что гомеоидная плотность обеспечивает отсутствие притяжения (из-за локального баланса сил со стороны высекаемых дуг) и одновременно равенство касательных. Сейчас мы покажем, что эти свойства равносильны.

Пусть, например, на кривой введена плотность  $d\varphi = k(s) ds$ , обеспечивающая равенство касательных, (рис. 6) — касательные длины  $s_1$  и  $s_2$  пересекаются и имеет место равенство

$$k_1 s_1 = k_2 s_2. \quad (5)$$

Пусть из произвольной точки  $P$  проведены две близкие прямые, высекающие на кривой линейные элементы длины  $ds_1$  и  $ds_2$ , находящиеся от  $P$  на расстояниях  $l_1$  и  $l_2$  (рис. 6). Пусть эти элементы пересекаются в точке  $O$ , находящейся от них на расстояниях  $s_1$  и  $s_2$ . Тогда одновременно выполняются равенства (4) и (5), из которых следует  $k_1 ds_1/l_1 = k_2 ds_2/l_2$ , т. е.

$$\frac{d\varphi_1}{l_1} = \frac{d\varphi_2}{l_2}.$$

Доказано, что кривая не притягивает, так как выполняется локальный баланс сил. Это доказательство обратимо, поэтому из локального баланса сил также вытекает равенство касательных.

*Аналитическое представление гомеоидной плотности*

Пусть коника на плоскости задана уравнением  $F(x, y) = 0$ , где

$$F(x, y) = [x \ y \ 1] A \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

Пусть в точках  $(x_1, y_1)$  и  $(x_2, y_2)$  на конике проведены касательные, пересекающиеся в точке  $O$ . Тогда для расстояния  $s_1$  от  $O$  до первой из точек имеется формула

$$s_1 = 2 \left( \left( [x_1 \ y_1 \ 1] A \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \right) / |\nabla F(x_1, y_1) \wedge \nabla F(x_2, y_2)| \right) \cdot (|\nabla F(x_1, y_1)|).$$

Мы видим, что один из сомножителей здесь симметричен относительно точек касания, а другой зависит только от первой из них. Поэтому после введения на конике плотности

$$d\varphi = \frac{ds}{|\nabla F|}$$

касательные становятся равными и коника перестает притягивать из-за локального баланса сил — эта плотность гомеоидна.

*Двояко-линейчатая поверхность —  
геометрическое представление гомеоидной плотности*

Пусть на кривой  $\mathcal{C}$ , лежащей в плоскости  $\pi$ , задана плотность  $d\varphi = k ds$ , гарантирующая равенство касательных. На кривой введем ориентацию, а плоскость вложим в  $R^3$ . В  $R^3$  построим семейство прямых, параметризованное точками кривой  $\mathcal{C}$ . Именно, в каждой точке  $P \in \mathcal{C}$  проведем прямую, которая проектируется на  $\pi$  в касательную к  $\mathcal{C}$  и имеет относительно  $\pi$  угловой коэффициент  $k(P)$ . Полученная линейчатая поверхность  $\mathcal{Q}$  на самом деле будет двояко-линейчатой, так как условие равенства касательных гарантирует, что на  $\mathcal{Q}$  лежат и все прямые с угловым коэффициентом  $-k(P)$ .

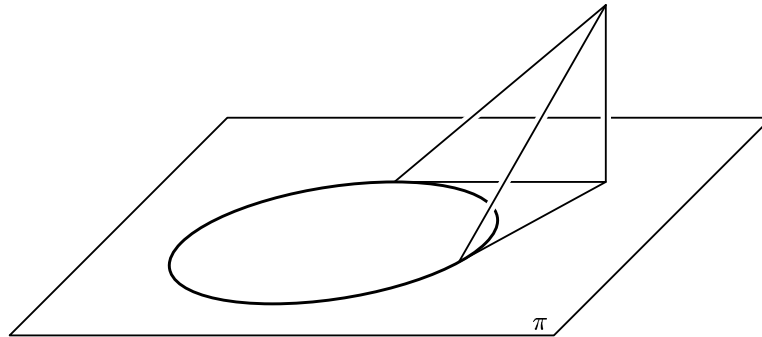


Рис. 7.

Хорошо известно [2], что всякая двояко-линейчатая поверхность это либо однополостный гиперболоид, либо гиперболический параболоид. Отсюда, кстати, следует, что исходная кривая  $\mathcal{C}$  — обязательно коника. И, значит, гомеодную плотность с равенством касательных и отсутствием притяжения из-за локального баланса сил можно ввести только на кониках.

### ОТ КОНИК К КВАДРИКАМ — ВТОРОЕ ДОКАЗАТЕЛЬСТВО

Итак, в предыдущем разделе мы показали, что гомеодная плотность на конике допускает естественное представление в виде квадрики в трехмерном пространстве.

#### *От коник к квадрикам*

Вернемся к теореме Понселе. Будем считать, например, что мы имеем дело с двумя эллипсами  $\mathcal{C}_0$  и  $\mathcal{C}_1$ , лежащими в координатной плоскости  $xy$ . Над эллипсом  $\mathcal{C}_0$  построим цилиндр  $\mathcal{Q}_0$  с образующей, параллельной оси  $z$ . Используя гомеодную плотность на эллипсе  $\mathcal{C}_1$ , построим над ним однополостный гиперболоид  $\mathcal{Q}_1$ . В описанной ситуации связь между теоремой Понселе для коник  $\mathcal{C}_0$ ,  $\mathcal{C}_1$  и теоремой для квадрик  $\mathcal{Q}_0$ ,  $\mathcal{Q}_1$ , сформулированной во введении, становится очевидной — линейные образующие  $\mathcal{Q}_1$  проектируются на плоскость  $xy$  в касательные к эллипсу  $\mathcal{C}_1$ , а пересечение  $\mathcal{Q}_0 \cap \mathcal{Q}_1$  в конику  $\mathcal{C}_0$ .

#### *Квадрики и их пересечения*

В пространстве  $CP^3$  все невырожденные квадрики проективно эквивалентны. Приведем невырожденную квадрику к виду  $x_1x_2 = x_3x_4$ . Такая квадрика содержит два семейства прямых —  $\lambda x_1 = \mu x_3$ ,  $\mu x_2 = \lambda x_4$  и  $\lambda x_1 = \mu x_4$ ,  $\mu x_2 = \lambda x_3$ . Каждое из семейств представляет собой  $CP^1$  (с однородными координатами  $\lambda$ ,  $\mu$ ) и через каждую точку квадрики проходит ровно по одной прямой из каждого семейства. Таким образом, невырожденная квадрика в  $CP^3$  является прямым произведением  $CP^1 \times CP^1$ .

Рассмотрим две квадрики общего положения в  $CP^3$ . Пусть  $\mathcal{T} = \mathcal{Q}_0 \cap \mathcal{Q}_1$  — их пересечение. Через  $L\mathcal{Q}_1$  обозначим одну из двух систем линейных образующих квадрики  $\mathcal{Q}_1$ . Эта система представляет собой  $CP^1$  или, что то же самое,  $S^2$ . Каждая из прямых, входящих в  $L\mathcal{Q}_1$ , пересекает  $\mathcal{Q}_0$  а, значит, и  $\mathcal{T}$  ровно в двух точках (с учетом кратности). Таким образом,  $\mathcal{T} = \mathcal{Q}_0 \cap \mathcal{Q}_1$  является двулистным накрытием  $S^2$ . У этого накрытия имеются четыре точки ветвления (как раз те точки, где прямые из  $L\mathcal{Q}_1$  касаются  $\mathcal{T}$ ). Отсюда следует, что  $\mathcal{T}$  — тор (подробности в [8]).

#### *Второе доказательство*

Докажем теорему для двух квадрик, сформулированную во введении. Рассмотрим две квадрики общего положения  $\mathcal{Q}_0$  и  $\mathcal{Q}_1$  в  $CP^3$ , пересекающиеся по тору  $\mathcal{T} = \mathcal{Q}_0 \cap \mathcal{Q}_1$ . Пусть опять  $L\mathcal{Q}_1$  одна из систем линейных образующих квадрики  $\mathcal{Q}_1$ . Определим инволюцию  $\sigma$  тора по правилу: точки пересечения  $p'$  и  $p''$

тора с линейной образующей меняются местами —  $p' \leftrightarrow p''$ . Присутствие кратных пересечений гарантирует наличие у этой инволюции неподвижных точек. У тора имеются два типа аналитических инволюций. Это сдвиг на полупериод и симметрия относительно точки. Существование неподвижных точек показывает, что  $\sigma$  представляет собой симметрию относительно точки.

На квадрике  $Q_1$  есть две системы линейных образующих  $L'Q_1, L''Q_1$ . Им соответствуют две инволюции  $\sigma'$  и  $\sigma''$ . Произведение четного числа этих инволюций  $\sigma''\sigma' \dots \sigma''\sigma'$  является сдвигом тора и наличие неподвижной точки гарантирует тождественность этого отображения. Теорема доказана.

#### Обратно к коникам

Покажем, как предыдущие рассуждения приводят к теореме Понселе для коник. Воспроизведем конструкцию из пункта *От коник к квадрикам*. Пусть расположенная в плоскости  $xy$  замкнутая ломаная  $A_1A_2 \dots A_{n+1}$ ,  $A_1 = A_{n+1}$  вписана в  $C_0$  и описана около  $C_1$ . Поднимем ее на квадрику  $Q_1$  так, чтобы получилась ломаная  $B_1B_2 \dots B_{n+1}$ , у которой отрезки  $B_{2i-1}B_{2i}$  лежат на образующих из  $L'Q_1$ , а отрезки  $B_{2i}B_{2i+1}$  — на образующих из  $L''Q_1$ . Легко убедиться: если отрезок  $A_iA_{i+1}$  касается коники  $C_1$  внешним образом (точка касания лежит вне  $C_0$ ), то точки  $B_i, B_{i+1}$  расположены по одну сторону от плоскости  $xy$ , если же касание внутреннее, то  $B_i, B_{i+1}$  лежат по разные стороны.

Вспомним еще, что для замкнутой ломаной  $A_1A_2 \dots A_{n+1}$  число внешних касаний  $\sigma$  четно. Отсюда сразу следует, что в случае четного  $n$  ломаная на квадрике тоже будет замкнутой, то есть  $B_1 = B_{n+1}$ . Но тогда ломаная  $B_1B_2 \dots B_{n+1}$  будет замкнута при любом выборе точки  $B_1$ , а вместе с ней будет замкнута ломаная  $A_1A_2 \dots A_{n+1}$  при любом выборе точки  $A_1$ .

Если же  $n$  нечетно, то будет выполняться соотношение  $B_1 = \sigma_z B_{n+1}$ , где  $\sigma_z$  симметрия относительно плоскости  $xy$ . Отсюда следует, что  $B_1$  неподвижная точка произведения четного числа инволюций  $\sigma_z\sigma' \dots \sigma''\sigma'$  и, следовательно, это произведение есть тождественное преобразование. Таким образом, при любом выборе точки  $B_1$  будет выполняться соотношение  $B_1 = \sigma_z B_{n+1}$  и, значит, при любом выборе точки  $A_1$  опять же будет  $A_1 = A_{n+1}$ .

#### Еще одна теорема Понселе

Как было обещано во введении, приведем еще один вариант пространственной теоремы Понселе.

**ТЕОРЕМА.** Пусть имеется пучок квадрик  $Q_\lambda$  в  $CP^3$ . Пусть пересечение  $T = \bigcap_\lambda Q_\lambda$  является тором. Рассмотрим  $2n$  квадрик  $Q_1, \dots, Q_{2n}$ , принадлежащих пучку. На каждой из них выберем по системе линейных образующих  $LQ_i$ . Пусть  $\sigma_i$  — соответствующие им инволюции тора  $T$ . Тогда, если произведение  $\sigma_1 \dots \sigma_{2n}$  имеет неподвижную точку, то это тождественное преобразование тора.

Доказательство этой теоремы ничем не отличается от предыдущего доказательства теоремы для двух квадрик.

## ДОБАВЛЕНИЯ

*Интегрируемость*

Как уже было сказано во введении, конфигурация, описанная в теореме Понселе, естественным образом возникает при движении частицы в эллиптическом бильярде. Траектория частицы, движущейся вдоль границы бильярда, все время касается эллипса, софокусного с границей бильярда. А все отрезки траектории, проходящей между фокусами, лежат на касательных к софокусной гиперболы. Такое упорядоченное поведение траекторий обязано интегрируемости эллиптического бильярда.

Пусть  $(x_n, y_n)$  — последовательные точки соударения частицы с границей бильярда

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

а  $V_n = (u_n, v_n)$ ,  $|V_n| = 1$  — скорость частицы перед соответствующим соударением. Тогда

$$\frac{(x_{n+1} + x_n)(x_{n+1} - x_n)}{a^2} + \frac{(y_{n+1} + y_n)(y_{n+1} - y_n)}{b^2} = 0$$

и это равенство означает, что естественное поле нормалей

$$\nabla \left( \frac{x^2}{a^2} + \frac{y^2}{b^2} \right) = \left( \frac{2x}{a^2}, \frac{2y}{b^2} \right)$$

обладает следующим свойством: сумма нормалей, взятых на концах любой хорды эллипса, перпендикулярна этой хорде. Это приводит к тому, что проекция поля нормалей на траекторию движения частицы есть величина постоянная, то есть имеется интеграл вида

$$F(x_n, y_n, u_n, v_n) = \frac{x_n u_n}{a^2} + \frac{y_n v_n}{b^2}. \quad (6)$$

*Интегрируемость и гомеодность*

Рассмотрим произвольную кривую с тем же самым механизмом интегрируемости, т. е. кривую с таким векторным полем  $N$ , что сумма его векторов на концах любой хорды перпендикулярна этой хорде. Беря малую хорду, видим, что это обязательно поле нормалей. Пусть на концах некоторой хорды векторы этого поля имеют длины  $N_1$  и  $N_2$ , а длины касательных от концов хорды до их пересечения имеют длины  $l_1$  и  $l_2$ . Тогда, как показывает рисунок 8,

$$\frac{l_1}{N_1} = \frac{l_2}{N_2},$$

то есть длины касательных с весом  $1/|N|$  равны, то есть плотность  $1/|N|$  гомеодная, а значит, сама кривая — обязательно коника.

*Фазовый портрет*

Вернемся к интегралу (6) и положим в нем

$$\begin{aligned} x &= a \cos t, & y &= b \sin t \\ u &= \cos \varphi, & v &= \sin \varphi. \end{aligned}$$

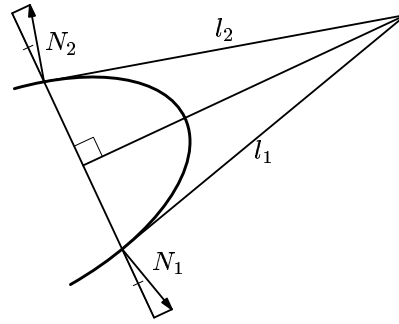


Рис. 8.

Тогда этот интеграл примет вид

$$F(t, \varphi) = (a + b) \cos(t - \varphi) - (a - b) \cos(t + \varphi). \quad (7)$$

■значит, в фазовом пространстве с координатами  $(t, \varphi)$  точка  $(t_n, \varphi_n)$ , соответствующая точке отражения  $(x_n, y_n)$  и вектору скорости перед соударением  $(u_n, v_n)$ , движется по линиям уровня функции  $F(t, \varphi)$ . Фазовый портрет эллиптического бильярда, изображенный на рисунке 9, получен в результате следующего компьютерного эксперимента. Из нижней точки эллипса в нескольких направлениях запускались частицы и для каждой частицы в течение продолжительного времени на фазовой плоскости отмечались точки  $(t_n, \varphi_n)$ .

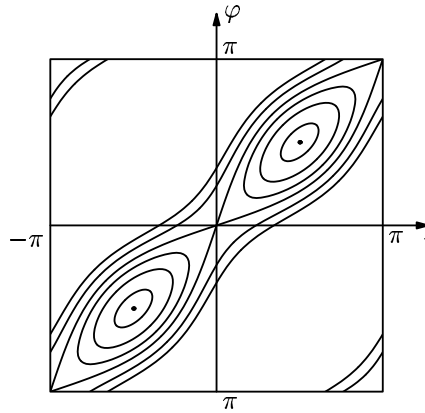


Рис. 9.

#### Геометрический и компьютерный бильярды

На первый взгляд существенных расхождений между этими бильярдами не должно быть. Сравнение рисунка 9 и картины линий уровня функции (7) подтверждает это. Однако, имеется ситуация, в которой различие существует и бросается в глаза.

Основное оптическое свойство эллипса гарантирует, что частица, один раз прошедшая через фокус, будет проходить через фокусы и после каждого следующего соударения. ■апустим частицу из нижней вершины эллипса через правый фокус. Очевидно, что, отражаясь и последовательно проходя через фокусы, частица будет приближаться к большой оси эллипса. Легко оценить скорость этого приближения. Пусть  $c = \sqrt{a^2 - b^2}$  — это расстояние от центра эллипса до его фокуса. Тогда при каждом прохождении бильярда отклонение частицы от большой оси уменьшается примерно в  $(a + c)/(a - c)$  раз.

В компьютерном эксперименте начальное поведение частицы, проходящей через фокус, полностью соответствует этому сценарию. Однако, после прижатия к оси и нескольких прохождений вдоль нее частица отбрасывается от большой оси эллипса. После этого ее снова прижимает к оси и снова отбрасывает и так далее — возникают биения. Дадим качественное объяснение этому явлению.

■апустим в «геометрическом» бильярде частицу через фокус, подождем пока она достаточно прижмется к оси и обратим ее направление. Обратная траектория — это тоже траектория некоторой частицы, и первоначально эта траектория будет проходить вдоль оси, но уже будет отталкиваться от нее. Таким образом, вблизи большой оси бильярда существуют две траектории: одна притягивается к этой оси, другая отталкивается от нее. В «компьютерном» бильярде за счет ошибок округления происходит «пересадка» с притягивающейся траектории на отталкивающуюся и частица начинает удаляться от оси бильярда. Этим и объясняются биения. Влияние ошибок округления можно подтвердить, проводя вычисления с удвоенной точностью, — в этом случае отбрасывание частицы от оси бильярда происходит после большего числа отражений.

#### *Пара коник приводится к виду двух окружностей*

Этот раздел предназначен для тех, кто имеет привычку просматривать тексты задом наперед — от конца к началу. ■десь объясняется, почему можно не читать эту статью, а значит, почему ее можно было и не писать. В [2] имеется простое доказательство теоремы Понселе для двух вложенных окружностей. Собственно, наше первое доказательство скроено по этому образцу. Покажем, что двумя окружностями можно было бы и ограничиться. Рассмотрим случай двух вложенных эллипсов. С помощью аффинного преобразования превратим внешний из них в окружность. Ограниченный ею круг будем считать проективной моделью плоскости Лобачевского. В этой модели движения — это проективные преобразования, сохраняющие круг [2]. Рассмотрим однопараметрическую группу  $H_1$  сдвигов вдоль горизонтального диаметра круга. Движения из  $H_1$  переставляют хорды, перпендикулярные диаметру, при этом точки круга движутся по эквидистантам, представленным полуэллипсами, касающимися внешней полуокружности в граничных точках диаметра. Аналогичным образом действует однопараметрическая группа  $H_2$  сдвигов вдоль вертикального диаметра. Пусть  $h_1 \in H_1$  и  $h_2 \in H_2$ . Тогда под действием преобразования  $h_1 h_2$  точка сначала сдвигается вдоль меридиана юг-север, а потом вдоль меридиана запад-восток. При этом для любой пары точек  $x, y$  однозначно определена пара преобразований  $h_1, h_2$ , такая что  $y = h_1 h_2(x)$ .

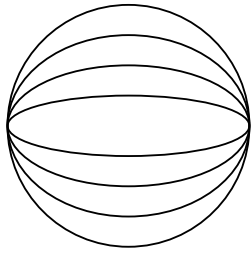


Рис. 10.

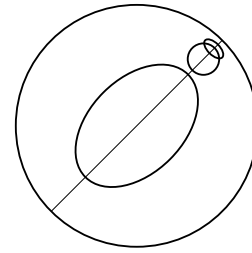
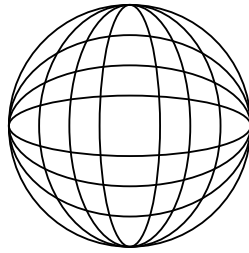


Рис. 11.

При движении плоскости Лобачевского эллипс переходит в эллипс, но при этом его центр не обязательно переходит в центр образа. Тем не менее, очевидно, что для любого эллипса существует преобразование вида  $h_1 h_2$ , переводящее его в эллипс, центр которого совпадает с центром круга. Проведем это преобразование и после этого рассмотрим однопараметрическую группу сдвигов вдоль диаметра, содержащего большую ось центрированного эллипса. При движении под действием этой группы и большая, и малая оси эллипса уменьшают свои размеры. Но большая ось делает это более интенсивно и в некоторый момент эллипс превращается в окружность. Таким образом, пара коник может быть приведена к виду двух окружностей, и теперь можно сослаться на соответствующее доказательство теоремы Понселе из [2].

## СПИСОК ЛИТЕРАТУРЫ

- [1] Берже М. Геометрия. М.: Мир, 1984.
- [2] Прасолов В. В., Тихомиров В. М. Геометрия. М.: МЦНМО, 1997.
- [3] Козлов В. В., Трещев Д. В. Биллиарды. М.: МГУ, 1991.
- [4] Синай Я. Г. Введение в эргодическую теорию. М.: ФАИИС, 1996.
- [5] Арнольд В. И. Математические методы классической механики. М.: Наука, 1989.
- [6] Griffiths P., Harris J. A Poncelet Theorem in Space // Comment. Math. Helvetici, 1977. Vol 52. P. 145–160.
- [7] Barth W., Bauer Th. Poncelet Theorems // Expo. Math. 1996. Vol 14. P. 125–144.
- [8] Прасолов В. В., Шварцман О. В. Алгебра римановых поверхностей. М.: ФАИИС, 1999.



# Наш семинар: математические сюжеты

---

---

## О сумме логарифмически выпуклых функций

Академик РАН Д. В. Аносов

Предлагается простое доказательство логарифмической выпуклости суммы логарифмически выпуклых функций.

1. В этой статье все функции принимают вещественные значения и определены на замкнутом, полуоткрытом или открытом интервале  $I$ , который может быть конечным или бесконечным (включая и случай всей числовой прямой  $\mathbb{R}$ ); иными словами,  $I$  — связное подмножество  $\mathbb{R}$ .

Как известно, функция  $f$  называется выпуклой, если для любого отрезка  $[a, b] \subset I$  график ограничения  $f|_{[a, b]}$  функции  $f$  на этот отрезок лежит ниже отрезка, соединяющего концы  $(a; f(a))$  и  $(b; f(b))$  этого графика. Подробнее и точнее это сокращённое выражение, которым я буду пользоваться и далее, означает, что для любого  $x \in [a, b]$  точка  $(x; f(x))$  графика лежит не выше точки указанного отрезка с той же абсциссой. Аналитически данное условие записывается так:

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b) \quad \text{при всех } t \in [0, 1]. \quad (1)$$

Функция  $f$  называется логарифмически выпуклой, если она принимает только положительные значения и её логарифм  $\log f(x)$  является выпуклой функцией. (Выбор основания логарифмов здесь не играет роли, ибо  $\log_{c_1} f = \log_{c_1} c_2 \log_{c_2} f$ . Но, конечно, подразумевается, что основание  $> 1$ .) Это же можно выразить так: для любого  $[a, b] \subset I$

$$f((1-t)a + tb) \leq f(a)^{1-t} f(b)^t \quad \text{при всех } t \in [0, 1]. \quad (2)$$

На геометрическом же языке можно сказать, что график  $f|_{[a, b]}$  лежит ниже графика экспоненциальной функции, принимающей те же самые значения, что и  $f$ , в концах отрезка. (Под экспоненциальной функцией сейчас понимается не только «чистая экспонента»  $e^{\beta x}$ , но и та же экспонента с положительным постоянным

множителем, т.е. функция вида  $\alpha e^{\beta x}$ , где  $\alpha > 0$ .) Графики ограничений всевозможных экспонент на всевозможные  $[a, b]$  образуют некоторое семейство дуг  $\mathcal{E}$  в полуплоскости  $\mathbb{R} \times (0, \infty)$ . Для любых двух точек этой полуплоскости с различными абсциссами имеется ровно одна дуга из  $\mathcal{E}$  с концами в этих точках. Логарифмически выпуклые функции суть те, для которых графики их ограничений на всевозможные  $[a, b] \subset I$  лежат ниже дуг из  $\mathcal{E}$  с теми же концами. В определении выпуклых функций аналогичную роль играет семейство  $\mathcal{L}$  всех невертикальных отрезков (графиков линейных функций) в  $\mathbb{R}^2$  (при этом семейство  $\mathcal{L}$  обладает в  $\mathbb{R}^2$  свойством, аналогичным указанному выше свойству семейства  $\mathcal{E}$  в  $\mathbb{R} \times (0, \infty)$ : любые две точки плоскости с различными абсциссами являются концами ровно одного отрезка из  $\mathcal{L}$ ).

Каждая линейная функция выпукла, а каждая экспонента — логарифмически выпукла. Кроме того, каждая экспонента выпукла (это простое упражнение по анализу). Отсюда сразу следует, что каждая логарифмически выпуклая функция  $f$  выпукла (график  $f|_{[a,b]}$  лежит под графиком соответствующей экспоненты, а тот — под соответствующим отрезком). Для экспоненты  $f = \alpha e^{\beta x}$  с  $\beta \neq 0$  любая точка графика  $f|_{[a,b]}$ , кроме его концов, лежит строго ниже точки соответствующего отрезка с той же абсциссой. Линейные функции, кроме констант, не являются логарифмически выпуклыми. Пересечение  $\mathcal{L} \cap \mathcal{E}$  состоит из горизонтальных отрезков.

Совершенно очевидно, что сумма выпуклых функций выпукла; отсюда следует, что произведение логарифмически выпуклых функций логарифмически выпукло. Отнюдь не столь очевидно

**ТЕОРЕМА 1.** *Если  $f, g$  логарифмически выпуклы, то  $f + g$  тоже логарифмически выпукла.*

Эта теорема давно известна, но излагаемое ниже её доказательство вполне может быть новым; довольно уверенно можно сказать, что в литературе на русском языке оно не публиковалось. Докажем сперва другую теорему.

**ТЕОРЕМА 2.** *Функция  $f$  логарифмически выпукла тогда и только тогда, когда при всех  $\alpha \in \mathbb{R}$  функции  $e^{\alpha x} f(x)$  выпуклы.*

**ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 2.** В одну сторону утверждение очевидно:  $\log(e^{\alpha x} f(x)) = \alpha x + \log f(x)$  отличается от  $\log f(x)$  на линейную функцию, так что  $\log(e^{\alpha x} f(x))$  и  $\log f(x)$  выпуклы или не выпуклы одновременно. А если функция  $e^{\alpha x} f(x)$  логарифмически выпукла, то она тем более выпукла.

Обратно, пусть дано, что при всех  $\alpha \in \mathbb{R}$  функции  $e^{\alpha x} f(x)$  выпуклы. Пусть  $[a, b] \subset I$ . Условие (2) выполняется тогда и только тогда, когда аналогичное условие выполняется для функции  $e^{\alpha x} f(x)$  (ведь  $e^{\alpha((1-t)a+tb)} = (e^{\alpha a})^{1-t}(e^{\alpha b})^t$ ; можно также воспользоваться логарифмированием). Существует такое  $\alpha$ , что  $e^{\alpha a} f(a) = e^{\alpha b} f(b)$  (именно,

$$\alpha = \frac{1}{b-a} \ln \frac{f(a)}{f(b)}; \quad (3)$$

можно также сослаться на указанное выше свойство семейства  $\mathcal{E}$ , взять дугу, соединяющую концы графика функции  $f|_{[a,b]}$ , и разделить  $f$  на соответствующую экспоненту). Так как концы дуги  $y = e^{\alpha x} f(x)$ ,  $x \in [a, b]$  лежат на одинаковой

высоте, то соединяющая их дуга из  $\mathcal{E}$  — это горизонтальный отрезок. А раз функция  $e^{\alpha x} f(x)$  выпукла, то её график лежит под этим отрезком. Здесь последний выступает как дуга из  $\mathcal{L}$ , но он же является и дугой из  $\mathcal{E}$ , соединяющей концы этого графика. Вот и выходит, что для  $e^{\alpha x} f(x)$  выполняется аналог условия (2). Аналитически то же выражается формулами

$$\begin{aligned} e^{\alpha((1-t)a+tb)} f((1-t)a+tb) &\leq (1-t)e^{\alpha a} f(a) + te^{\alpha b} f(b) = e^{\alpha a} f(a) = \\ &= (e^{\alpha a} f(a))^{1-t} (e^{\alpha b} f(b))^t, \end{aligned}$$

где сперва использована выпуклость  $e^{\alpha x} f(x)$ , а затем — что  $e^{\alpha a} f(a) = e^{\alpha b} f(b)$ .

Теорема 2 даёт своего рода характеризацию логарифмической выпуклости с помощью выпуклости. Для логарифмической выпуклости функции  $f$  необходима её выпуклость, но этого далеко не достаточно (линейные функции). Для доказательства логарифмической выпуклости надо как бы подвергнуть  $f$  бесконечному числу «тестов» — она должна удовлетворять (1) (со всевозможными  $a, b$ ) не только сама, но и после умножения на экспоненту. Из доказательства видно, почему это так — при одном из этих «тестов» (отвечающем (3)) условие (1) для  $e^{\alpha x} f(x)$  оказывается совпадающим с условием (2).

**ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ 1.** Мы хотим доказать, что в предположениях этой теоремы функция  $e^{\alpha x}(f(x) + g(x))$  выпукла. Но она является суммой двух логарифмически выпуклых и тем более выпуклых функций  $e^{\alpha x} f(x)$ ,  $e^{\alpha x} g(x)$ .

2. Здесь я остановлюсь на имеющемся в литературе доказательстве теоремы 1.

Когда  $f$  дважды дифференцируема, то условие её выпуклости состоит в том, что всюду  $f'' \geq 0$ , а условие логарифмической выпуклости — в том, что всюду  $f''f - f'^2 \geq 0$ . Для дважды дифференцируемых логарифмически выпуклых  $f, g$  утверждение теоремы 1 состоит в том, что всюду  $(f+g)''(f+g) - (f'+g')^2 \geq 0$ . Оно сводится к такому чисто алгебраическому утверждению: если  $c, \gamma > 0$ ,  $ac \geq b^2$  и  $\alpha\gamma \geq \beta^2$ , то  $(a+\alpha)(c+\gamma) \geq (b+\beta)^2$ . (Мы принимаем  $c = f$  и  $\gamma = g$ , поэтому  $c, \gamma$  положительны.) Для доказательства последнего достаточно заметить, что неравенство  $ac \geq b^2$  является при  $c > 0$  необходимым и достаточным условием неотрицательности квадратного трёхчлена  $ax^2 + 2bx + c$  при всех  $x$ . После того как мы перефразируем аналогичным образом неравенства  $\alpha\gamma \geq \beta^2$  и  $(a+\alpha)(c+\gamma) \geq (b+\beta)^2$ , остаётся заметить, что сумма неотрицательных трёхчленов неотрицательна. Можно и не обращаться к квадратным трёхчленам, а рассуждать так. Если  $a = 0$  или  $\alpha = 0$ , то утверждение тривиально (почему?). Если же  $a \neq 0$  и  $\alpha \neq 0$ , то

$$\begin{aligned} (a+\alpha)(c+\gamma) - (b+\beta)^2 &= \underline{ac} + ac + a\gamma + \underline{\alpha\gamma} - \underline{b^2} - 2b\beta - \underline{\beta^2} \geq \\ &\geq ac + a\gamma - 2b\beta = a\alpha \left( \frac{c}{a} + \frac{\gamma}{\alpha} \right) - 2b\beta \geq a\alpha \left( \left( \frac{b}{a} \right)^2 + \left( \frac{\beta}{\alpha} \right)^2 \right) - 2b\beta = \\ &= \frac{\alpha}{a} b^2 + \frac{a}{\alpha} \beta^2 - 2b\beta = \left( \sqrt{\frac{\alpha}{a}} b \right)^2 + \left( \sqrt{\frac{a}{\alpha}} \beta \right)^2 - 2 \left( \sqrt{\frac{\alpha}{a}} b \right) \left( \sqrt{\frac{a}{\alpha}} \beta \right) = \\ &= \left( \sqrt{\frac{\alpha}{a}} b - \sqrt{\frac{a}{\alpha}} \beta \right)^2 \geq 0. \end{aligned}$$

Здесь сперва использовано, что подчёркнутые члены с плюсом не меньше подчёркнутых членов с минусом, затем — что неравенство  $ac \geq b^2$  влечёт положительность  $a$  (ввиду  $a \neq 0, c > 0$ ) и эквивалентно неравенству  $\frac{c}{a} \geq \frac{b^2}{a^2}$ , а также аналогичное соображение для  $\alpha, \beta, \gamma$ .

В общем случае можно вывести теорему 1 из того же самого алгебраического утверждения. Непосредственно с его помощью доказывается, что если  $f, g > 0$ ,

$$f\left(\frac{a+b}{2}\right)^2 \leq f(a)f(b) \quad (4)$$

и аналогично для  $g$ , то и

$$\left(f\left(\frac{a+b}{2}\right) + g\left(\frac{a+b}{2}\right)\right)^2 \leq (f(a) + g(a))(f(b) + g(b)). \quad (5)$$

Это (меняя обозначения) снова сводится к тому, что

$$\text{из } c, \gamma > 0, ac \geq b^2, \alpha\gamma \geq \beta^2 \text{ следует } (a + \alpha)(c + \gamma) \geq (b + \beta)^2$$

(только в данном случае можно с самого начала считать, что все числа  $a, b, c, \alpha, \beta, \gamma$  больше 0). Надо, конечно, пояснить, почему частного случая (2) (с  $f + g$  вместо  $f$ ), отвечающего  $t = \frac{1}{2}$ , достаточно для заключения о логарифмической выпуклости  $f + g$ . Опять меняя обозначения, мы приходим к вопросу: можно ли из частного случая (1), отвечающего  $t = \frac{1}{2}$ , т.е. из того, что

$$2f\left(\frac{a+b}{2}\right) \leq f(a) + f(b) \quad \text{при } [a, b] \subset I, \quad (6)$$

сделать вывод о выпуклости  $f$ ?

Из (6) легко следует, что

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b) \quad \text{при двоично-рациональных } t \in [a, b].$$

Перейти к любым  $t \in [a, b]$  можно, если известно, что  $f$  непрерывна. Точнее, достаточно, чтобы  $f$  была непрерывна во внутренних точках  $I$ , а в конце  $A$  интервала  $I$ , если  $A \in I$ , достаточно выполнения условия

$$\overline{\lim}_{x \rightarrow A} f(x) \leq f(A). \quad (7)$$

Но сравнительно легко доказать, что выпуклая функция непрерывна внутри  $I$ , а если  $I$  содержит какой-нибудь свой конец  $A$ , то в нём выполняется (7). Значит, то же верно и для логарифмически выпуклой функции. Получается, что в условиях теоремы 1  $f$  и  $g$  можно считать непрерывными внутри  $I$  и удовлетворяющими условию (7) в конце интервала  $I$ , если этот конец содержится в  $I$ . Стало быть, то же самое справедливо и для  $f + g$ . А тогда нам достаточно (5).

Ради полноты я приведу доказательство непрерывности  $f$  внутри  $I$  и свойства (7) для выпуклой  $f$ . Достаточно доказать два утверждения:

- а). Если  $[a, b] \subset I$ , то  $f(a) \geq \overline{\lim}_{x \rightarrow a, x \in [a, b]} f(x)$  и аналогично для  $f(b)$ .
- б). Если  $a$  — внутренняя точка  $I$ , то  $f(a) \leq \underline{\lim}_{x \rightarrow a} f(x)$ .

К а). Для  $x \in [a, b]$  имеем  $x = (1-t(x))a + t(x)b$ , где  $t = \frac{x-a}{b-a} \rightarrow 0$  при  $x \rightarrow a$ .

А  $f(x) \leq (1 - t(x))f(a) + t(x)f(b)$ . Правая часть при  $x \rightarrow a$  стремится к  $f(a)$ , а верхний предел левой есть  $\lim_{x \rightarrow a, x \in [a, b]} f(x)$ .

К б). При достаточно малых  $h$  точки  $a \pm h \in I$ . При этом  $2f(a) \leq f(a - h) + f(a + h)$ . Следовательно,

$$2f(a) \leq \overline{\lim}_{h \rightarrow 0} f(a - h) + \underline{\lim}_{h \rightarrow 0} f(a + h) \leq f(a) + \underline{\lim}_{x \rightarrow a} f(x)$$

(здесь использовано, что ввиду а)  $\overline{\lim}_{h \rightarrow 0} f(a - h) \leq f(a)$ ). Отсюда получается б).

Вместе взятые, рассуждения п. 2 заметно длиннее, чем п. 1. Однако в любом случае, говоря о выпуклых функциях, надо остановиться на их непрерывности, так что эту часть п. 2 можно не учитывать при оценке длины доказательства теоремы 1. А тогда данное в п. 2 доказательство этой теоремы оказывается не длиннее, чем в п. 1. Но оно кажется формальнее: какое отношение вспомогательные квадратные трёхчлены имеют к нашей задаче? А если обходиться без них, то вообще получается какой-то алгебраический трюк. Мы не видим «движущих пружин» доказательства. В п. 1 же наши рассуждения были непосредственно связаны с геометрией задачи. Всё это отдаёт субъективизмом, но думаю, что многие (как и я) найдут рассуждения п. 1 более прозрачными.

Доказательство, приведённое в п. 2, имеется в книжке Э. Артина о  $\Gamma$ -функции [1], а также у И. И. Привалова [2] и Н. Бурбаки [3]. В последней оно приводится (только для дважды дифференцируемых функций, что по существу не проще) при изложении свойств  $\Gamma$ -функции, воспроизводящем текст Артина. В теореме 1 легко перейти от суммы к интегралу, что доставляет логарифмическую выпуклость  $\Gamma$ -функции. Это свойство несколько облегчает изучение последней, на что впервые обратили внимание Г. Бор и Й. Моллеруп в своём учебнике анализа [4] (пп. 39, 41, а также пример 6 на с. 161. Впоследствии этот учебник переиздавался, но в новых изданиях — по крайней мере тех, которые я видел, — о логарифмической выпуклости  $\Gamma$ -функции не говорилось. Возможно, авторы сочли, что после выхода книжки Артина писать об этом незачем). В [4] не было утверждения о логарифмической выпуклости суммы, а сразу доказывалась логарифмическая выпуклость интеграла, что делалось так же, как позднее в [1] доказывалось первое утверждение. Выделение первого утверждения как естественно предшествующего второму — это, по-видимому, усовершенствование Артина.

Сомнительно, чтобы вопрос о логарифмической выпуклости (как в общем случае суммы логарифмически выпуклых функций или интеграла от таковых, так и в специальном случае  $\Gamma$ -функции) мог возникнуть в XIX веке, так что рассуждения, имеющиеся у Бора – Моллерупа – Артина, скорее всего, им и принадлежат.

И. И. Привалов сообщает, что он узнал доказательство теоремы 1 от А. И. Плеснера. Последний незадолго до того эмигрировал в СССР из Германии, где он вполне мог узнать это доказательство от самого Артина или из его книги.

Теорема 2 принадлежит П. Монтелю [5]. Его доказательство отлично от нашего. Та же теорема имеется в [2]. Привалов ссылается на [5], но приводит другое доказательство. В конечном счёте его основная идея — та же, что и в п. 1

(подбор подходящей экспоненты — такой, при которой неравенства (1) и (2) для  $e^{\alpha x} f(x)$  становятся эквивалентными), но оформлена она иначе и, по-моему, менее прозрачно.

Психологическая загадка: почему, имея теорему 2, Привалов не вывел из неё теорему 1? Мне кажется, дело в том, что в основной части [2] большую роль играет оператор, который как бы заменяет оператор Лапласа (для более общих функций) и который теперь называют «оператором Привалова». А в главе о выпуклых функциях Привалов в порядке подготовки делал упор на использование аналогичного (но более простого) разностного оператора и связанные с ним обстоятельства типа принципа максимума. По-видимому, при написании этой главы он менее заботился о различных взаимосвязях, не имеющих отношения к этому оператору (хотя всё-таки счёл обязательным привести теорему 1, доказываемую без его помощи).

#### СПИСОК ЛИТЕРАТУРЫ

- [1] *Артин Э.* Введение в теорию гамма-функции. М.-Л.: ГТТИ, 1934.
- [2] *Привалов И. И.* Субгармонические функции. М.-Л.: ОНТИ НКТП, 1937.
- [3] *Бурбаки Н.* Функции действительного переменного. М.: Наука, 1965.
- [4] *Bohr H., Møllerup J.* Laerebog i Matematisk Analyse, t. III. København: Jul. Gjellerups Forlag, 1922.
- [5] *Montel P.* Sur les fonctions convexes et les fonctions susharmoniques. J. de math. pures et appl., 1928, t. 7, v. 1, 29-60.

# Как складывать треугольники

Ф. Бахарев      К. Кохась      Ф. Петров

Мы доказываем конфигурационную теорему, похожую на теорему Дезарга, которая позволяет ввести структуру абелевой группы на множестве треугольников, перспективных относительно некоторой точки.

## 1. КОНФИГУРАЦИОННЫЕ ТЕОРЕМЫ

### 1.1. ТЕОРЕМА ДЕЗАРГА И ЕЕ ОБОБЩЕНИЕ

**ОПРЕДЕЛЕНИЕ 1.** Если две плоские конфигурации, составленные из точек и прямых, могут быть приведены в соответствие так, что все прямые, проходящие через соответственные точки, пересекаются в одной точке  $O$ , то мы говорим, что эти две конфигурации *перспективны относительно точки  $O$* . Точка  $O$  в этом случае называется *центром перспективы*. Если же соответствие таково, что точки пересечения соответственных прямых лежат на одной прямой  $\ell$ , то мы говорим, что эти две конфигурации *перспективны относительно прямой  $\ell$* . Прямая  $\ell$  в этом случае называется *осью перспективы*.

**ТЕОРЕМА (ДЕЗАРГ).** *Если два треугольника перспективны относительно некоторой точки, то они также перспективны и относительно некоторой прямой.*

Верна и обратная (или двойственная) теорема Дезарга: если два треугольника перспективны относительно некоторой прямой, то они перспективны относительно некоторой точки.

Доказательство теоремы Дезарга любопытный читатель может найти в книгах [1], [2], [4] или получить самостоятельно.

Итак, теорема Дезарга утверждает, что у перспективных треугольников точки пересечения соответственных сторон лежат на одной прямой. А можно ли что-нибудь утверждать про точки пересечения *несоответственных* сторон? Оказывается, можно.

**ОПРЕДЕЛЕНИЕ 2 (ОСНОВНАЯ КОНСТРУКЦИЯ).** Пусть треугольники  $A_1A_2A_3$  и  $B_1B_2B_3$  перспективны относительно точки  $S$ . Пусть  $(i, j, k)$  — перестановка чисел 1, 2, 3. Обозначим через  $P_{ij}$  точку пересечения прямых  $A_iA_k$  и  $B_jB_k$ . Точку пересечения прямых  $A_iA_j$  и  $B_iB_j$  обозначим через  $S_k$ . Наконец, положим  $C_k = P_{ik}P_{ki} \cap P_{jk}P_{kj}$  (см. рис. 1).

**ТЕОРЕМА 1 (ОБОБЩЕННАЯ ТЕОРЕМА ДЕЗАРГА).** *Треугольник  $C_1C_2C_3$  перспективен треугольникам  $A_1A_2A_3$  и  $B_1B_2B_3$  относительно точки  $S$ .*

Мы дадим два доказательства этой теоремы.

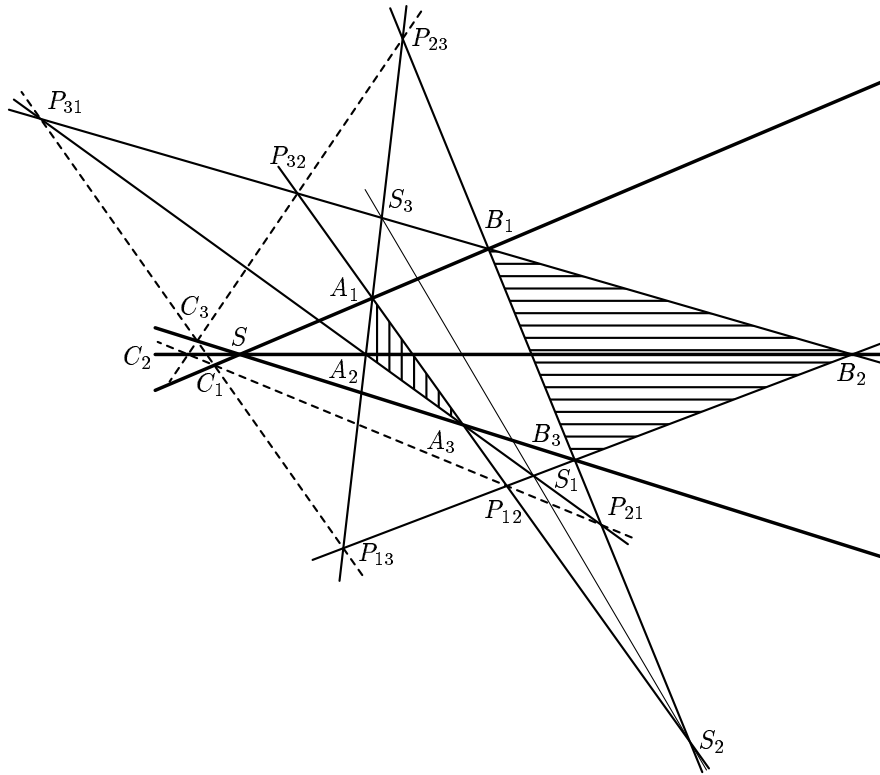


Рис. 1. Теорема Дезарга и ее обобщение

ДОКАЗАТЕЛЬСТВО 1. Проверим, что треугольники  $\Delta_1 = S_1P_{21}P_{12}$  и  $\Delta_2 = S_3P_{23}P_{32}$  перспективны относительно точки  $S_2$ . Для этого проверим, что прямые  $P_{21}P_{23}$ ,  $P_{12}P_{32}$  и  $S_1S_3$  проходят через точку  $S_2$ . Заметим, что прямая  $P_{21}P_{23}$  — это на самом деле прямая  $B_3B_1$ , а прямая  $P_{12}P_{32}$  — это прямая  $A_1A_3$ , поэтому эти прямые пересекаются в точке  $S_2$ . Осталось заметить, что точки  $S_1$ ,  $S_2$ ,  $S_3$  лежат на одной прямой в силу теоремы Дезарга для исходных треугольников  $A_1A_2A_3$  и  $B_1B_2B_3$ .

Итак, треугольники  $\Delta_1$  и  $\Delta_2$  перспективны относительно точки, следовательно, по теореме Дезарга, точки пересечения соответствующих сторон

$$\begin{aligned} A_2 &= S_1P_{21} \cap S_3P_{23}, \\ B_2 &= S_1P_{12} \cap S_3P_{32}, \\ C_2 &= P_{21}P_{12} \cap P_{23}P_{32} \end{aligned}$$

лежат на одной прямой. Аналогично  $C_1 \in A_1B_1$ ,  $C_3 \in A_3B_3$ , что и требовалось доказать.

ДОКАЗАТЕЛЬСТВО 2. Нам понадобится следующий факт об алгебраических кривых (см. [3, гл. 1, § 1, теорема 4 для  $n = 4$ ]):



Пусть  $A_{ij}$  — (попарно различные) точки пересечения прямых  $p_i, q_j$ ,  $1 \leq i, j \leq 4$ . Если все точки

$$\begin{array}{cccc} A_{11}, & A_{12}, & A_{13}, & A_{14}, \\ A_{21}, & A_{22}, & A_{23}, & A_{24}, \\ A_{31}, & A_{32}, & A_{33}, & \\ A_{41}, & A_{42} & & \end{array}$$

лежат на некоторой кривой степени 4, то остальные точки —  $A_{34}, A_{43}, A_{44}$  — тоже лежат на этой кривой.

Рассмотрим треугольники  $A_1A_2A_3$  и  $B_1B_2B_3$ , перспективные относительно точки  $S$ , и построим точки  $P_{ij}$ . Нам удобно будет определить точки  $C_i$  не так, как это сделано в основной конструкции, поэтому положим  $C'_1 = P_{12}P_{21} \cap A_1B_1$ ,  $C'_2 = P_{12}P_{21} \cap P_{23}P_{32}$ ,  $C'_3 = P_{23}P_{32} \cap A_3B_3$ . Будем обозначать прямые избыточно, перечисляя лежащие на них точки получившейся конфигурации. Рассмотрим прямые

$$\begin{array}{lll} p_1 = P_{23}A_1A_2P_{13}, & q_1 = P_{31}A_2A_3P_{21}, & r_1 = P_{31}P_{13}, \\ p_2 = P_{31}P_{32}B_1B_2, & q_2 = P_{13}P_{12}B_3B_2, & r_2 = P_{32}A_1A_3P_{12}, \\ p_3 = C'_3SA_3B_3, & q_3 = C'_1SA_1B_1, & r_3 = P_{23}B_1B_3P_{21}, \\ p_4 = C'_2C'_1P_{12}P_{21}, & q_4 = C'_2C'_3P_{32}P_{23}, & r_4 = SA_2B_2. \end{array}$$

Мы видим, что точки пересечения прямых  $p_i$  и  $q_i$  — это 16 точек нашей конфигурации, и все они, за исключением, быть может,  $C'_1, C'_2, C'_3$ , лежат на кривой степени 4, которая равна объединению четырех прямых  $r_1, r_2, r_3, r_4$ . В обозначениях упомянутой теоремы,  $C'_1 = A_{43}, C'_2 = A_{44}, C'_3 = A_{34}$ , поэтому эти точки также должны принадлежать объединению прямых  $r_i$ . Нетрудно проверить, что ни на одной из прямых  $r_i$  не может оказаться 5 точек нашей конфигурации, поэтому точки  $C'_3, C'_1$  должны лежать на прямой  $r_1$ , а точка  $C'_2$  — на прямой  $r_4$ .

В книге [2, § 22] описана конфигурация Рейе с параметрами (12, 4, 16, 3). Вершины конфигурации Рейе — это вершины куба, его центр и три бесконечно удаленные точки, соответствующие направлениям ребер; прямые — это ребра куба и его четыре главные диагонали. Как нетрудно проверить, конфигурация, описанная в теореме 1, — двойственная к конфигурации Рейе.

Теорема Паскаля, которую доказывают похожим рассуждением, оказывается в результате теоремой об ассоциативности операции сложения точек эллиптической кривой. Кажется, за нашей теоремой не скрывается никакой операции для объектов, связанных с кривой четвертого порядка. Не видно даже конфигурационной теоремы для общей кривой четвертого порядка.

## 1.2. ТЕОРЕМЫ ПАППА И ПАСКАЛЯ

В этом параграфе мы докажем обобщения теорем Паскаля и Паппа, которые, впрочем, далее нам не понадобятся. Любителям полярных преобразований и двойственных утверждений не составит труда обобщить в том же духе и теорему Брианшона.

**ТЕОРЕМА (ПАСКАЛЬ).** *Точки пересечения противоположных сторон вписанного в коническое сечение шестиугольника лежат на одной прямой.*

ТЕОРЕМА (ПАПП). Пусть точки  $A_1, A_2, A_3$  лежат на одной прямой, а точки  $B_1, B_2, B_3$  — на другой. Тогда точки пересечения пар прямых  $A_1B_2$  и  $A_2B_1$ ,  $A_1B_3$  и  $A_3B_1$ ,  $A_2B_3$  и  $A_3B_2$  также расположены на одной прямой.

Доказательства этих теорем можно найти в [3], [1].

Обобщение теоремы Паскаля получается, если мы изучаем точки пересечения непровотивоположных сторон, а в формулировке теоремы Паппа — если мы рассматриваем точки пересечения «несимметричных» прямых. Введем обозначения. Пусть  $A_1, A_2, A_3, B_1, B_2, B_3$  — исходные точки (в теореме Паппа) или вершины шестиугольника (в теореме Паскаля; обратите внимание на порядок вершин — см. рис. 2). Для каждой перестановки  $(i, j, k)$  чисел 1, 2, 3 обозначим через  $Q_{ij}$  точку пересечения прямых  $A_iB_k$  и  $B_jA_k$ .

ТЕОРЕМА 2 (ОБОБЩЕННАЯ ТЕОРЕМА ПАСКАЛЯ). Пусть  $A_1B_3A_2B_1A_3B_2$  — вписанный шестиугольник. Тогда прямые  $Q_{12}Q_{21}$ ,  $Q_{13}Q_{31}$  и  $Q_{23}Q_{32}$  пересекаются в одной точке.

ТЕОРЕМА 3 (ОБОБЩЕННАЯ ТЕОРЕМА ПАППА). В условиях теоремы Паппа прямые  $Q_{12}Q_{21}$ ,  $Q_{13}Q_{31}$  и  $Q_{23}Q_{32}$  пересекаются в одной точке.

ДОКАЗАТЕЛЬСТВО. Рассмотрим треугольники  $Q_{12}Q_{23}Q_{31}$  и  $Q_{21}Q_{32}Q_{13}$ . Точки пересечения противоположных сторон этих треугольников — это точки, фигурирующие в теореме Паскаля (для шестиугольника  $A_1B_3A_2B_1A_3B_2$ ). Стало быть, треугольники перспективны относительно прямой, а потому и относительно точки, что и требовалось.

Отметим еще один известный способ обобщить теорему Паскаля.

ТЕОРЕМА (ЕЩЕ ОДНО ОБОБЩЕНИЕ ТЕОРЕМЫ ПАСКАЛЯ). Обозначим  $S_{ik} = A_iA_j \cap B_jB_k$ . Тогда три прямые  $l_i = S_{jk}S_{kj}$  пересекаются в одной точке.

ДОКАЗАТЕЛЬСТВО. Обозначим  $S = A_3B_3 \cap A_2B_2$ . По теореме Паскаля (для перенумерованных точек) точки  $S, S_{32}, S_{23}$  лежат на одной прямой  $l$ . По теореме Паппа,

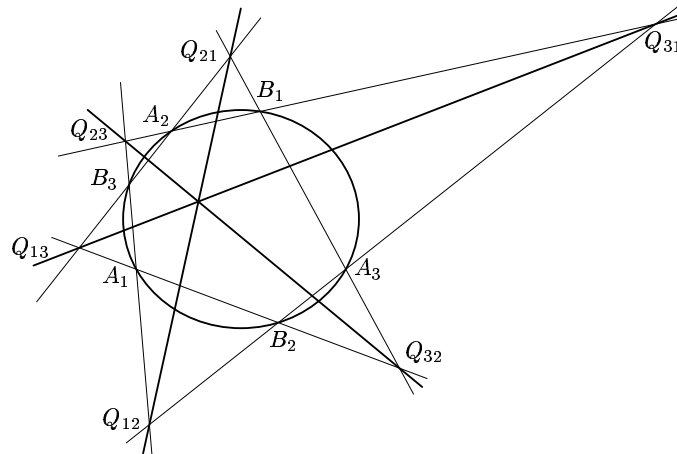


Рис. 2. Одно из обобщений теоремы Паскаля

примененной к тройкам точек  $B_3, S_{13}, B_2$  и  $A_2, S_{21}, B_3$ , получаем, что на прямой  $l$  лежит точка  $U = B_2S_{21} \cap A_3S_{13}$ . Еще раз применим теорему Паша, на этот раз к тройкам точек  $B_2, S_{12}, S_{13}$  и  $A_3, S_{31}, S_{21}$ . Получим, что на прямой  $l$  лежит точка  $Z = S_{12}S_{21} \cap S_{13}S_{31}$ . Что и требовалось.

### 1.3. ДВОЙСТВЕННАЯ КОНФИГУРАЦИЯ

Будем считать, что основная конструкция задана на проективной плоскости. Заметим, что все ее элементы сохраняются при проективных преобразованиях. При этом выбор точки  $S$  и прямых  $l_1, l_2, l_3$  с проективной точки зрения не нарушает никакой общности: при помощи подходящего проективного преобразования можно отобразить эту тройку конкурентных прямых в любую другую тройку конкурентных прямых. В частности, структура «аддитивной группы треугольников», описанной в разделе 2, не зависит от этого выбора.

Фиксируем полярное преобразование на проективной плоскости. Применяя его ко всем элементам основной конфигурации, мы получаем теорему, двойственную к теореме 1:

**ТЕОРЕМА 4.** Пусть треугольники  $A = A_1A_2A_3$  и  $B = B_1B_2B_3$  перспективны относительно прямой  $s$ . Для  $i = 1, 2, 3$  обозначим  $L_i$  — точки пересечения соответственных сторон:  $L_i = A_kA_j \cap B_kB_j$ . Положим  $C_k = A_iB_j \cap A_jB_i$ . Тогда  $C = C_1C_2C_3$  — треугольник, перспективный треугольникам  $A_1A_2A_3$  и  $B_1B_2B_3$  относительно прямой  $s$  (см. рис. 3).

Доказательство теоремы 1 является одновременно и доказательством теоремы 4 (так как полярное преобразование инволютивно).

**ОПРЕДЕЛЕНИЕ 3.** Конструкцию теоремы 1 (рис. 1) будем называть *центральной моделью* основной конструкции, а конструкцию теоремы 4 (рис. 3) — *осевой моделью*.

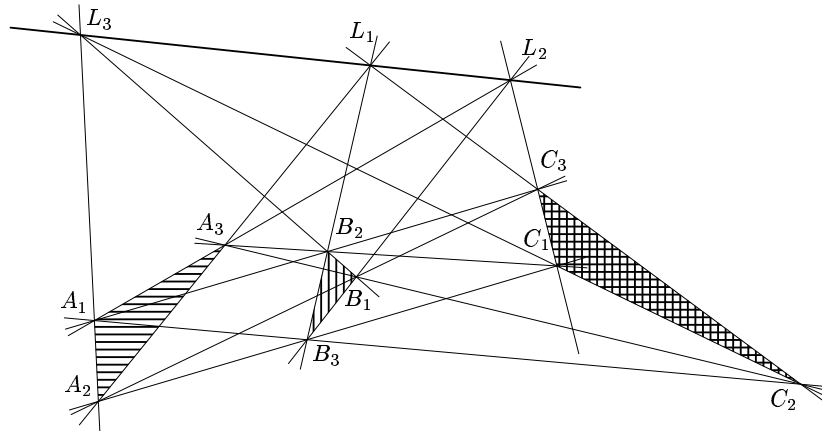


Рис. 3. Конструкция, двойственная к основной

В центральной модели мы считаем фиксированными точку  $S$  и три прямые  $\ell_1, \ell_2, \ell_3$ , проходящие через эту точку, в осевой модели зафиксируем прямую  $s$  и три точки  $L_1, L_2, L_3$ , лежащие на ней. Треугольник  $A_1A_2A_3$  (где  $A_i \in \ell_i$  в центральной модели или  $A_iA_j$  содержит точку  $L_k$  в осевой модели) будем для краткости обозначать  $A$ .

## 2. АДДИТИВНАЯ ГРУППА ТРЕУГОЛЬНИКОВ

Напомним читателю, как определяют сложение точек эллиптической кривой. Сначала вводят некоторую «геометрическую» конструкцию, которая по двум точкам  $A$  и  $B$  позволяет построить еще одну точку кривой, а именно:  $C$  — это третья точка пересечения прямой  $AB$  с данной эллиптической кривой, обозначим  $C = A \cdot B$ . После этого сама операция сложения задается так: мы фиксируем произвольную точку  $E$  на кривой и полагаем  $A + B = E \cdot (A \cdot B)$ . Точка  $E$  при этом будет играть роль нейтрального элемента («нуля»).

Этот способ, однако, требует обсуждения некоторых технических деталей, и по ходу дела оказывается, что кривую естественно рассматривать не на плоскости, а на комплексной проективной плоскости, определение точки  $C$  требует специальных оговорок, если  $A = B$  или в случае, когда прямая  $AB$  касается кривой, наконец, очень часто в качестве точки  $E$  берут бесконечно удаленную точку, что приводит к несколько более «загадочному» определению суммы — построив точку  $A \cdot B$ , ее симметрично отражают относительно оси  $x$ . Вершиной теории является построение эллиптических функций, которые являются гомоморфизмами между аддитивной группой точек эллиптической кривой и комплексным тором.

В этом разделе мы определяем операцию сложения на множестве перспективных треугольников. При помощи описанной в предыдущем параграфе конфигурационной теоремы мы можем по двум треугольникам построить третий. Здесь, правда, нужны некоторые технические комментарии, скажем, для случая совпадающих треугольников. Далее, чтобы определить сложение, нам нужно всего лишь повторить эту операцию с каким-нибудь фиксированным заранее треугольником. Однако введение такой операции требует некоторого пополнения множества треугольников, при этом в качестве фиксированного элемента удобно выбрать как раз «бесконечно удаленный треугольник». Роль эллиптической функции будут исполнять при этом барицентрические координаты.

### 2.1. СУММА ПЕРСПЕКТИВНЫХ ТРЕУГОЛЬНИКОВ

**ОПРЕДЕЛЕНИЕ 4.** Пусть даны два перспективных (относительно точки  $S$  или прямой  $s$ ) треугольника  $A$  и  $B$ . В силу обобщенной теоремы Дезарга эти треугольники определяют третий треугольник —  $C$ , который мы будем называть *предсуммой* треугольников  $A$  и  $B$  и записывать  $C = A \boxplus B$ .

Отметим очевидные свойства этой операции.

**ЛЕММА 1.** *Операция «предсумма» обладает следующими свойствами:*

- 1)  $A \boxplus B = B \boxplus A$ .
- 2) Если  $A \boxplus B = C$ , то  $A \boxplus C = B$  и  $B \boxplus C = A$ .

ЗАМЕЧАНИЕ 1. Множество треугольников не замкнуто относительно операции «предсумма». Например, разглядывая рис. 1, читатель легко поймет, что при неудачном выборе треугольников  $A$  и  $B$  все три вершины треугольника  $C$  могут совпасть с точкой  $S$ . Кроме того, не для любых двух перспективных треугольников наши построения корректны.

ОПРЕДЕЛЕНИЕ 5. (Общее определение суммы треугольников) Определим *сумму* треугольников  $A$  и  $B$  формулой  $A + B = F \boxplus (A \boxplus B)$ , в том случае, когда эта формула имеет смысл, где  $F$  — некоторый фиксированный треугольник.

ЗАМЕЧАНИЕ 2. Множество с операцией, обладающей свойствами из леммы 1, называется квазигруппой. Из аксиом квазигруппы никоим образом не следует, что операция  $A + B = F \boxplus (A \boxplus B)$  ассоциативна. Так, теорема Паппа позволяет по двум тройкам коллинеарных точек  $A$  и  $B$  построить третью тройку  $C = A \boxplus B$ . Нетрудно видеть, что такая операция удовлетворяет лемме 1. Но возникающая по аналогии с определением 5 «сумма» не ассоциативна.

Далее мы будем работать в осевой модели. Будем считать, что ось перспективы  $s$  бесконечно удалена (для этого достаточно сделать подходящее проективное преобразование). Кроме того, в технических целях нам удобно будет в качестве  $F$  взять, в некотором смысле (в каком — будет ясно позднее), «бесконечно удаленный треугольник». Поэтому в оставшейся части этой статьи мы будем пользоваться следующим определением суммы.

ОПРЕДЕЛЕНИЕ 6. Пусть даны два перспективных относительно бесконечно удаленной прямой  $s$  треугольника  $A$  и  $B$ . Пусть  $C = A \boxplus B$ , обозначим  $C_0$  — центр тяжести треугольника  $C$ . Треугольник  $D$ , центрально симметричный треугольнику  $C$  относительно точки  $C_0$ , будем называть *суммой* треугольников  $A$  и  $B$  и обозначать  $D = A + B$ .

Ниже мы докажем, что эта операция коммутативна и ассоциативна (что оправдывает ее название), а симметрия относительно центра тяжести — это на самом деле вычисление предсуммы данного треугольника и «бесконечно удаленного треугольника» (см. пример 4), который и будет играть роль нейтрального элемента (нуля) для операции сложения. Операция, заданная общим определением 5, также ассоциативна и коммутативна, и порождает ту же самую группу, но мы не будем этого здесь доказывать, поскольку доказательства совершенно аналогичны.

## 2.2. БАРИЦЕНТРИЧЕСКИЕ КООРДИНАТЫ И ПРОСТРАНСТВО ТРЕУГОЛЬНИКОВ

Итак, сделав подходящее проективное преобразование, мы можем считать, что прямая  $s$  в осевой модели — бесконечно удаленная. В этом случае соответственные стороны треугольников  $A$ ,  $B$ ,  $C = A \boxplus B$  параллельны (см. рис. 4). Ниже под *треугольниками* понимаются только треугольники с теми же направлениями сторон.

Зафиксируем какой-нибудь *треугольник*  $E = E_1E_2E_3$ . Ниже под координатами точки аффинной плоскости мы всегда будем понимать барицентрические координаты этой точки относительно *треугольника*  $E$  (см. [5]). С помощью этих координат введем координаты на множестве *треугольников*.

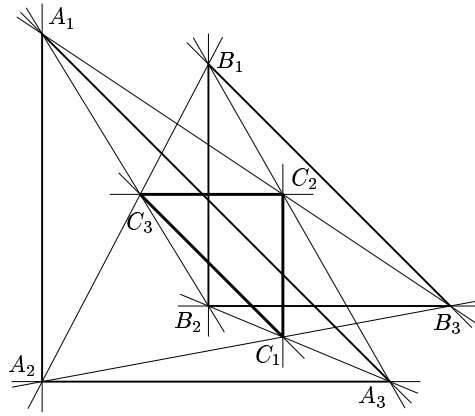


Рис. 4. Двойственная конструкция с бесконечно удаленной осью

ОПРЕДЕЛЕНИЕ 7. (Барицентрические координаты на множестве треугольников.) Каждый треугольник  $D$  однозначно задается координатами  $(d_1, d_2, d_3)$  ( $d_1 + d_2 + d_3 = 1$ ) его центра тяжести и коэффициентом  $d$  единственной гомотетии, переводящей  $D$  в  $E$  (параллельный перенос по определению считаем гомотетией с коэффициентом 1). Значит, этот треугольник однозначно определяется тройкой параметров  $(\delta_1, \delta_2, \delta_3)$ , таких что

$$\delta_1 : \delta_2 : \delta_3 = d_1 : d_2 : d_3, \quad \delta_1 + \delta_2 + \delta_3 = d.$$

Параметры  $(\delta_1, \delta_2, \delta_3)$  мы будем называть *барицентрическими координатами* треугольника  $D$ .

Переход от координат  $(d_1, d_2, d_3; d)$  к барицентрическим координатам  $(\delta_1, \delta_2, \delta_3)$  и обратно осуществляется по формулам:

$$\delta_i = d \cdot d_i, \quad d = \delta_1 + \delta_2 + \delta_3; \quad d_i = \frac{\delta_i}{d} = \frac{\delta_i}{\sum \delta_i} \quad (i = 1, 2, 3).$$

ПРИМЕР 1. Выразим координаты вершин треугольника  $D = D_1D_2D_3$  с центром  $D_0$  через его барицентрические координаты  $\delta_1, \delta_2, \delta_3$ . Для этого заметим, что вектор  $\overrightarrow{D_0D_3}$  гомотетичен вектору  $\overrightarrow{E_0E_3}$  с коэффициентом  $\frac{1}{d}$ :

$$\begin{aligned} D_3 &= D_0 + \frac{1}{d} (E_3 - E_0) = \frac{1}{d} (\delta_1, \delta_2, \delta_3) + \frac{1}{d} \left( -\frac{1}{3}, -\frac{1}{3}, \frac{2}{3} \right) = \\ &= \left( \frac{\delta_1 - 1/3}{d}, \frac{\delta_2 - 1/3}{d}, \frac{\delta_3 + 2/3}{d} \right). \end{aligned}$$

Аналогично вычисляются координаты остальных вершин.

Отметим, что если сумма координат равна нулю, никакого (геометрического) треугольника не получается.

ОПРЕДЕЛЕНИЕ 8. Добавим ко множеству геометрических треугольников формальные элементы, задаваемые в барицентрических координатах

параметрами  $(\delta_1, \delta_2, \delta_3)$ , где  $\delta_1 + \delta_2 + \delta_3 = 0$ . Полученное пространство, формально изоморфное  $\mathbb{R}^3$ , будем называть *пространством треугольников*.

Поскольку *треугольники* с малым значением суммы координат имеют «большой размер», естественно ожидать, что эти формальные элементы можно интерпретировать как в каком-то смысле «бесконечно удаленные» *треугольники* (см. раздел 3.1).

### 2.3. ОСНОВНАЯ ТЕОРЕМА

**ТЕОРЕМА 5.** *Описанная в определении 6 операция «+» в барицентрических координатах совпадает с операцией сложения в аддитивной группе  $\mathbb{R}^3$ .*

**ДОКАЗАТЕЛЬСТВО.** Пусть  $A$  и  $B$  — два *треугольника* с координатами  $(\alpha_1, \alpha_2, \alpha_3)$ ,  $\sum \alpha_i = a \neq 0$ , и  $(\beta_1, \beta_2, \beta_3)$ ,  $\sum \beta_i = b \neq 0$  соответственно, и пусть их предсумма  $C = A \boxplus B$  — тоже *треугольник*, т. е.  $\sum \gamma_i = c \neq 0$ . Вычислим координаты *треугольника*  $C$ . *Треугольники*  $A_1C_3A_2$  и  $B_2C_3B_1$  (см. рис. 4) подобны с коэффициентом  $A_2A_1/B_2B_1 = \frac{1/a}{1/b}$ . Значит,  $a \cdot A_1C_3 = b \cdot C_3B_2$ , и по правилу рычага точка  $C_3$  — центр масс материальных точек  $(a, A_1)$  и  $(b, B_2)$ . Отсюда

$$\begin{aligned} C_3 &= \frac{1}{a+b} (a \cdot A_1 + b \cdot B_2) = \\ &= \frac{1}{a+b} ((\alpha_1 + 2/3, \alpha_2 - 1/3, \alpha_3 - 1/3) + (\beta_1 - 1/3, \beta_2 + 2/3, \beta_3 - 1/3)) = \\ &= \left( \frac{-(\alpha_1 + \beta_1) - 1/3}{-(a+b)}, \frac{-(\alpha_2 + \beta_2) - 1/3}{-(a+b)}, \frac{-(\alpha_3 + \beta_3) + 2/3}{-(a+b)} \right). \end{aligned} \quad (1)$$

Аналогично вычисляются координаты точек  $C_2$  и  $C_1$ . Заметим, что

$$C_2 - C_1 = \frac{1}{-(a+b)} (-1, 1, 0) = \frac{1}{-(a+b)} (E_2 - E_1).$$

Следовательно, сумма координат *треугольника*  $C$  (то есть коэффициент гометии  $C \rightarrow E$ ) равна  $-(a+b)$ . Теперь для того чтобы найти координаты *треугольника*  $C$ , мы должны умножить на эту величину координаты его центра тяжести  $C_0$ . Координаты центра точки  $C_0$  равны

$$\frac{C_1 + C_2 + C_3}{3} = \left( \frac{-(\alpha_1 + \beta_1)}{-(a+b)}, \frac{-(\alpha_2 + \beta_2)}{-(a+b)}, \frac{-(\alpha_3 + \beta_3)}{-(a+b)} \right).$$

Значит, искомые координаты *треугольника*  $C$  равны

$$(-(\alpha_1 + \beta_1), -(\alpha_2 + \beta_2), -(\alpha_3 + \beta_3)). \quad (2)$$

Очевидно, что симметрия относительно центра тяжести меняет координаты *треугольника* на противоположные. Таким образом, координаты суммы  $A + B$  равны

$$(\alpha_1 + \beta_1, \alpha_2 + \beta_2, \alpha_3 + \beta_3). \quad (3)$$

**СЛЕДСТВИЕ 1.** *Описанная в определении 6 операция «+» продолжается до коммутативной, ассоциативной операции в пространстве *треугольников*. Построенная таким способом «аддитивная группа *треугольников*» изоморфна аддитивной группе  $\mathbb{R}^3$ .*

## 3. ТЕХНИЧЕСКИЕ ПОДРОБНОСТИ

## 3.1. ГЕОМЕТРИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ ФОРМАЛЬНЫХ ТРЕУГОЛЬНИКОВ

Для почти всех формальных треугольников мы можем дать геометрическую интерпретацию.

**ОПРЕДЕЛЕНИЕ 9.** (Барицентрические координаты на бесконечно удаленной прямой.) Для точки  $X$  с барицентрическими координатами  $(x_1, x_2, x_3)$ ,  $\sum x_i = 1$ , имеем  $\overrightarrow{OX} = \sum x_i \overrightarrow{OE_i}$  для любой точки  $O$ . Аналогично, для тройки чисел  $(y_1, y_2, y_3)$  с суммой 0 рассмотрим вектор  $\sum y_i \overrightarrow{OE_i}$ . Ясно, что этот вектор не зависит от точки  $O$ , и потому тройка  $(y_1, y_2, y_3)$  определяет направление (и даже вектор; направление будет, если факторизовать по однородности). Будем называть однородную тройку  $(y_1, y_2, y_3)$  *барицентрическими координатами* этого направления (= точки бесконечно удаленной прямой).

**ОПРЕДЕЛЕНИЕ 10.** *Псевдо-треугольник* с координатами  $(p_1, p_2, p_3)$ ,  $\sum p_i = 0$  — это упорядоченная тройка направлений с координатами

$$(p_1 + \frac{2}{3}, p_2 - \frac{1}{3}, p_3 - \frac{1}{3}), \quad (p_1 - \frac{1}{3}, p_2 + \frac{2}{3}, p_3 - \frac{1}{3}), \quad (p_1 - \frac{1}{3}, p_2 - \frac{1}{3}, p_3 + \frac{2}{3}).$$

Соответствующие этим направлениям точки бесконечно удаленной прямой (или сами направления) мы будем называть *вершинами* псевдо-треугольника. Формальный элемент пространства треугольников с координатами  $(p_1, p_2, p_3)$ ,  $\sum p_i = 0$ , отождествим с псевдо-треугольником  $(p_1, p_2, p_3)$ .

**ПРИМЕР 2.** Пусть  $A$  и  $C$  — два треугольника, центрально симметричные относительно точки, отличной от середин сторон этих треугольников. Найдем их предсумму. Выполняя построения основной конструкции, мы видим, что для всех пар  $i, j$  прямые  $A_i C_j$  и  $A_j C_i$  параллельны (см. рис. 5) и, таким образом, «треугольник»  $A \# C$  — это тройка бесконечно удаленных точек, т. е. в точности псевдо-треугольник.

**ПРИМЕР 3.** Выясним, что представляет собой псевдо-треугольник с координатами  $(0, 0, 0)$ . Его вершины — это направления  $(\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3})$ ,  $(-\frac{1}{3}, \frac{2}{3}, -\frac{1}{3})$ ,  $(-\frac{1}{3}, -\frac{1}{3}, \frac{2}{3})$ . Выбирая в качестве точки  $O$  из определения 9 центр тяжести треугольника  $E$ , мы видим, что эти направления суть в точности направления медиан треугольника  $E$  (или любого другого треугольника).

На проективной плоскости основная конструкция без специальных оговорок позволяет находить предсумму треугольника и псевдо-треугольника.

**ПРИМЕР 4.** Найдем предсумму треугольника  $A$  и псевдо-треугольника  $B$  с координатами  $(0, 0, 0)$ . Вершины псевдо-треугольника  $B(0, 0, 0)$  — это направления медиан треугольника  $A$  (см. пример 3). Выполняя построения основной конструкции, находим, что искомая предсумма — это просто отражение треугольника  $A$  относительно его центра тяжести (см. рис. 5).

Нетрудно видеть, что, хотя барицентрические координаты направления определены с точностью до умножения на константу, псевдо-треугольники с разными координатами определяют разные тройки направлений.



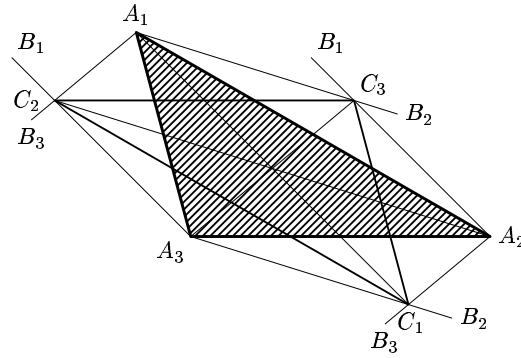


Рис. 5. Сложение треугольника с нулевым

Определение псевдо-треугольника не имеет смысла лишь для трех наборов параметров:  $(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3})$ ,  $(\frac{1}{3}, -\frac{2}{3}, \frac{1}{3})$  и  $(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3})$ , поскольку точки бесконечно удаленной прямой в силу определения 9 не могут иметь три нулевые координаты.

ОПРЕДЕЛЕНИЕ 11. Три формальных элемента пространства треугольников с координатами  $(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3})$ ,  $(\frac{1}{3}, -\frac{2}{3}, \frac{1}{3})$  и  $(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3})$  будем называть *совсем-псевдо-треугольниками*.

Таким образом, формальные элементы из определения 8 исчерпываются псевдо-треугольниками и совсем-псевдо-треугольниками.

### 3.2. Поправки к основной конструкции

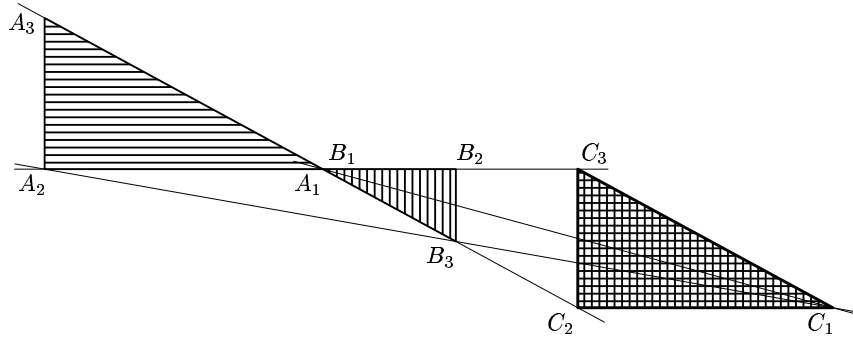
Мы по-прежнему работаем в осевой модели. Нетрудно видеть, что неопределенности в основной конструкции для *треугольников*  $A$  и  $B$  возникают в трех случаях:

- 1) если две стороны *треугольников*  $A$  и  $B$  оказались на одной прямой (например, если точки  $A_1, A_2, B_1, B_2$  лежат на одной прямой, то прямые  $A_1B_2$  и  $A_2B_1$  совпадают и вершина  $C_3$  не определена);
- 2) если  $A = B$  (более глубокий случай предыдущего вырождения; непосредственно не построить ни одной вершины *треугольника*  $C$ );
- 3) если *треугольники*  $A$  и  $B$  центрально симметричны друг другу относительно середины их (общей) стороны (тогда если, скажем,  $A_1A_2 = B_2B_1$ , то прямые  $A_1B_2$  и  $A_2B_1$  не определены).

В первом и втором случае результат вычисления предсуммы — *треугольник*, в третьем — *совсем-псевдо-треугольник*. Покажем, каким образом нужно изменить основную конструкцию, чтобы найти в этих случаях предсумму геометрически.

#### 3.2.1. Совпадение двух сторон

Пусть прямые, содержащие по стороне наших *треугольников*, совпадают, скажем,  $A_1 = B_1$  (но  $A_2 \neq B_2$  и *треугольники*  $A$  и  $B$  — не центрально симметричны, см. рис. 6). Тогда  $C_1$  определим по конструкции, а про  $C_2$  и  $C_3$  мы

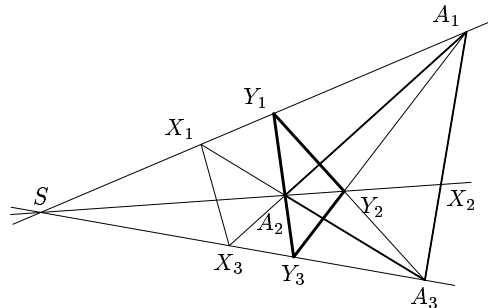


**Рис. 6.** Простейший вырожденный случай предложения

знаем лишь, что  $C_2 \in A_1B_3 = B_1A_3$ , а  $C_3 \in A_1B_2 = A_2B_1$ . Это позволяет найти  $C_2$  и  $C_3$  простыми построениями (скажем,  $C_2 = A_1B_3 \cap C_1S_3$ , где  $S_3$  — бесконечно удаленная точка прямой  $A_1A_2$ ). Случай, когда, скажем,  $A_2 = B_1$  или просто точки  $A_1, B_1, A_2, B_2$  лежат вчетвером на одной прямой, разбираются аналогично. То, что в этом случае координаты предсуммы по-прежнему определяются формулой (2) — понятно (например, сделаем предельный переход).

### 3.2.2. ГЕОМЕТРИЧЕСКОЕ ОПРЕДЕЛЕНИЕ $A \boxplus A$

Будем исходить из того, что должно выполняться свойство  $A \boxplus (A \boxplus A) = A$ . В центральной модели это значит, что стороны треугольника  $A \boxplus A$  должны проходить через вершины треугольника  $A$ . Построим такой треугольник. Обозначим  $X_i = A_jA_k \cap SA_i$  (см. рис. 7). Пусть  $Y_j$  — гармоническое дополнение точки  $S$  к паре  $\{A_j, X_j\}$ . Тогда три четверки вида  $\Gamma_j = \{S, Y_j\}\{A_j, X_j\}$  — гармонические. При проецировании из точки  $A_i$  четверки  $\Gamma_j$  на прямую  $SA_k$  получается (снова гармоническая) четверка  $\{S, Y'_k\}\{X_k, A_k\}$ , где  $Y'_k = A_iY_j \cap SA_k$ . Но точка  $Y_k$  в этом случае должна совпадать с точкой  $Y'_k$ , так как обе они являются гармоническими дополнениями точки  $S$  к паре  $\{A_k, X_k\}$ . Из этого следует, что точки  $A_i, Y_j, Y_k$  — коллинеарны и потому треугольник  $Y_1Y_2Y_3$  — искомый.



**Рис. 7.** Сложение треугольника с собой

Кстати,  $X_1X_2X_3 = A/2$ , то есть  $X_1X_2X_3 \boxplus X_1X_2X_3 = A$ .

В осевой модели условие «стороны  $A \boxplus A$  проходят через вершины  $A$ » превратится в условие «вершины  $A \boxplus A$  лежат на сторонах  $A$ ». Поэтому  $A \boxplus A$  — это в точности *серединный треугольник*. Легко видеть, что это синтетическое определение согласовано с определением 7 и формулой (2).

### 3.2.3. СУММА ТРЕУГОЛЬНИКОВ, СИММЕТРИЧНЫХ ОТНОСИТЕЛЬНО СЕРЕДИНЫ СТОРОНЫ

В этом случае мы можем предложить лишь формальное правило.

Пусть  $A_1 = B_2$ ,  $A_2 = B_1$ . Конструкция позволяет построить две (бесконечно-удаленные) вершины, третья вершина  $(A_1B_2 \cap A_2B_1)$  — не определена. В рассматриваемом случае  $\alpha_1 + \alpha_2 + \alpha_3 = -(\beta_1 + \beta_2 + \beta_3)$ . Аналогично примеру 1 убеждаемся, что координаты треугольников удовлетворяют соотношениям

$$\alpha_1 = -\beta_1 - \frac{1}{3}, \quad \alpha_2 = -\beta_2 - \frac{1}{3}, \quad \alpha_3 = -\beta_3 + \frac{2}{3}.$$

Будем считать, что в этом случае результат предложения — совсем-псевдо-треугольник с координатами  $(-\frac{1}{3}, -\frac{1}{3}, \frac{2}{3})$ .

Обратно: результат предложения совсем-псевдо-треугольника  $(-\frac{1}{3}, -\frac{1}{3}, \frac{2}{3})$  с *треугольником*  $A$  — *треугольник*, симметричный  $A$  относительно стороны  $A_1A_2$ .

Эти определения согласованы со свойствами из леммы 1 и отлично вписываются в формулу (2).

### 3.3. ПРЕДСУММА ДВУХ ПСЕВДО-ТРЕУГОЛЬНИКОВ

Нетрудно проверить, что при построении на проективной плоскости предсуммы двух центрально симметричных *треугольников* с координатами  $(\alpha_1, \alpha_2, \alpha_3)$  и  $(\beta_1, \beta_2, \beta_3)$  (см. пример 2) получается *псевдо-треугольник* с координатами  $p_i = -(\alpha_i + \beta_i)$ , то есть по-прежнему выполнена формула (2). Так же вычисляются координаты предсуммы *треугольника* и *псевдо-треугольника*.

Для геометрического сложения псевдо-треугольников рассмотрим следующую параметризацию множества псевдо- (в том числе, совсем-псевдо-) *треугольников*. Будем рассматривать псевдо- и совсем-псевдо-треугольники как предсуммы фиксированного *треугольника* (например, *треугольника*  $E$ ) с ему симметричными. Эти симметричные *треугольники* и будем рассматривать как параметры псевдо-треугольников.

Ясно (напрямую, с оглядкой на барицентрические координаты), что для каждого псевдо-треугольника найдется единственный параметризующий его *треугольник*. Предложение двух псевдо-треугольников описывается следующей леммой:

**ЛЕММА 2.** *Если треугольники  $B$  и  $C$  симметричны треугольнику  $E$ , а псевдо-треугольники  $X$  и  $Y$  определены как предсуммы  $E \boxplus B$  и  $E \boxplus C$ , то  $X \boxplus Y = E \boxplus D$ , где  $D_i$  — середина отрезка  $B_iC_i$ .*

Доказательство леммы оставляем читателю.

## СПИСОК ЛИТЕРАТУРЫ

- [1] *Коксетер Г. С. М., Грейтцер С. Л.* Новые встречи с геометрией. М., 1978.
- [2] *Гильберт Д., Кон-Фоссен С.* Наглядная геометрия. М., 1981.
- [3] *Прасолов В. В., Соловьев Ю. П.* Эллиптические функции и алгебраические уравнения. М., 1997.
- [4] *Ефимов Н. В.* Высшая геометрия. М.-Л., 1949.
- [5] *Балж М. Б., Болтянский В. Г.* Геометрия масс. (Библиотечка «Квант», вып. 61) М., 1987.

## Игра Ричмана

Е. Горский      В. Гуровиц      И. Межиров      Е. Осьмова

Эта статья написана по материалам летней конференции Турнира Городов 2000 года. На этих конференциях школьникам предлагается для решения несколько циклов задач исследовательского характера. На последней конференции самым популярным оказался цикл задач, посвященный различным вариантам игры Ричмана, о которой пойдет речь ниже. Задач в этом цикле было предложено очень много. Здесь обсуждается только часть затронутых на конференции вопросов. Заинтересованный читатель может найти дополнительную информацию в электронном архиве конференции [1].

Среди многообразия математических игр можно выделить два широких класса. К одному из них (к так называемым *комбинаторным* играм) относятся наиболее распространенные игры: шашки, шахматы, ним... В этих играх возможно лишь конечное число позиций, в каждой из которых игрок имеет конечное число возможных ходов. Исход такой игры не зависит от случая, вся информация, известная одному игроку, известна и другому.

Игры из второго класса (так называемые *матричные* или *вероятностные* игры) в основном имеют отношение к реальным (например, стратегическим) приложениям теории игр, оперирующим не с дискретными, а с непрерывными объектами. В них ход игрока не предопределен позицией, а в большой мере произволен (например, «камень-ножницы-бумага»).

Однако игру, которую мы рассматриваем в данной статье, сложно отнести к одному из этих двух классов. Она носит название «игра Ричмана» в честь математика, который ее изобрел.<sup>1)</sup>

### 1. ПРАВИЛА ИГРЫ

В классической версии этой игры игроков двое: они условно называются Синим и Красным. В качестве игрового поля берется произвольный ориентированный граф, одна из вершин которого отмечается синим цветом (далее она будет обозначаться  $b$ ), а другая — красным (соответственно,  $r$ ), причем из любой вершины ведет ориентированный путь хотя бы в одну из выделенных вершин. Остальные вершины графа мы будем называть черными.

В начале игры в некоторой вершине графа находится фишка, а игроки имеют некоторые начальные капиталы, равные, соответственно,  $\mathbf{B}$  и  $\mathbf{R}$  условных единиц (можно считать, что  $\mathbf{B} + \mathbf{R} = 1$ ). Каждый игрок делает неотрицательную ставку, не превосходящую его капитала (ставки делаются одновременно и

<sup>1)</sup> Дэвид Росс Ричман (1956–1991) — известный американский математик, работавший в основном в комбинаторных областях: теории графов, теории игр. Трагически погиб в авиакатастрофе.

в тайне от соперника: «закрытый аукцион»). Сделавший бóльшую ставку получает право сдвинуть фишку по ребру графа, соблюдая ориентацию. Взамен он должен отдать противнику сумму, равную своей ставке. Если ставки совпали по величине, то победитель определяется жребием, но деньги все равно передаются.

Если игрок смог привести фишку в вершину своего цвета, то он выигрывает. Иначе игра продолжается бесконечно, и считается, что она закончилась вничью. Наша цель состоит в том, чтобы определить, при каких начальных условиях игроки имеют беспроигрышную или выигрышную стратегию.

## 2. ПРИМЕРЫ

Априори не ясно, почему при каких-то начальных условиях у кого-то из игроков вообще есть выигрышная стратегия. Рассмотрим несколько конкретных примеров.

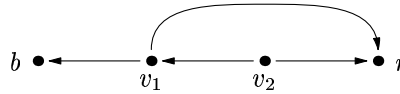


Рис. 1.

Если начальная вершина —  $v_1$ , а денег у игроков поровну, то оба ставят «ва-банк» (ставят весь свой капитал), и исход игры решает монетка, т. е. беспроигрышной стратегии нет ни у одного из игроков. В любом другом случае, обладающий большим капиталом, ставит весь свой капитал и выигрывает игру. Итак, при  $\mathbf{B} > 0,5$  Синий имеет выигрышную стратегию, при  $\mathbf{B} < 0,5$  выигрышную стратегию имеет Красный (из дальнейшего будет ясно, почему удобно записывать условие именно в такой форме).

Пусть фишка стоит в вершине  $v_2$ , а начальный капитал Синего равном  $0,7$  у. е. Тогда выигрышную стратегию имеет Красный игрок: первым ходом он ставит  $0,3$ . Если Синий ставит меньше, то он сразу проигрывает игру. Если же он ставит не меньше  $0,3$ , то он выигрывает право хода, двигает фишку в соседнюю черную вершину, но у него остается не больше  $0,4$  у. е., а у противника, соответственно, не меньше  $0,6$ , и Красный, ставя ва-банк, выигрывает аукцион и всю игру в целом.

При другом капитале (и начальной вершине  $v_2$ ) Синий игрок, чтобы выиграть первый аукцион, должен поставить не меньше  $\mathbf{R} = 1 - \mathbf{B}$ , что уменьшит его капитал по меньшей мере до  $2\mathbf{B} - 1$ . Тогда  $2\mathbf{B} - 1 > 0,5$  (так как он должен выиграть, находясь в вершине  $v_1$ ), откуда  $\mathbf{B} > 0,75$ . Итак, при  $\mathbf{B} > 0,75$  выигрышную стратегию имеет Синий, при  $\mathbf{B} < 0,75$  — Красный.

Рассмотрим более сложный граф (рис. 2).



Рис. 2.

Исследуем, кто выигрывает при начальной вершине  $v_1$  в зависимости от начального капитала Синего игрока  $\mathbf{B}$ . Наилучшая первая ставка Синего —  $\mathbf{B}$  (так как в этом случае он либо сразу выигрывает, либо гарантированно получает от противника наибольшую возможную сумму). Если Красный поставит меньше, то Синий сразу выиграет всю игру. Иначе Синий получит от него по крайней мере  $\mathbf{B}$ , т. е. капитал у него станет больше чем  $2\mathbf{B}$ , и фишка сдвинется в вершину  $v_2$ . Во избежание проигрыша Синий обязан поставить больше текущего капитала Красного, т. е.  $1 - 2\mathbf{B} + \varepsilon$ , где  $\varepsilon > 0$ . Фишка вернется на исходную позицию, а у Синего останется больше  $2\mathbf{B} - 1 + 2\mathbf{B} - \varepsilon = 4\mathbf{B} - 1 - \varepsilon$ . Итак, за один цикл  $\mathbf{B}$  превращается по крайней мере в  $4\mathbf{B} - 1 - \varepsilon$ . Если  $\mathbf{B} < \frac{1}{3}$ , то изменение капитала, равное  $3\mathbf{B} - 1 - \varepsilon$ , отрицательно и тем меньше, чем меньше  $\mathbf{B}$ . Сделав много таких циклов, Красный может уменьшить капитал Синего сколь угодно сильно, в частности, настолько, чтобы в некоторый момент в точке  $v_2$  тот имел капитал, меньший 0,5, т. е. Красный смог бы выиграть. Аналогично, если  $\mathbf{B} > \frac{1}{3}$ , то при подходящем  $\varepsilon$  капитал Синего будет увеличиваться, что позволит ему в какой-то момент поставить больше противника, когда фишка окажется в вершине  $v_1$ , т. е. выиграть партию. Итак, при правильной игре при  $\mathbf{B} > \frac{1}{3}$  выигрывает Синий, при  $\mathbf{B} < \frac{1}{3}$  — Красный.

Из приведенных примеров видно, что для заданной начальной вершины  $v$  графа удастся подобрать такое число (назовем его  $R(v)$ ), что при начальном капитале Синего  $\mathbf{B} > R(v)$  он имеет выигрышную стратегию, а при  $\mathbf{B} < R(v)$  выигрышную стратегию имеет Красный. Это утверждение оказывается верным и в общем случае. Ниже мы докажем его.

### 3. ОСНОВНЫЕ ТЕОРЕМЫ

Пусть  $U(v)$  — множество последователей данной вершины  $v$  (т. е. тех вершин, куда ведет ребро из  $v$ ). Назовем числовую функцию  $R(v)$  на графе *функцией Ричмана*, если она удовлетворяет следующим соотношениям:

$$R(r) = 1, \quad R(b) = 0, \quad R(v) = \frac{\min_{w \in U(v)} R(w) + \max_{w \in U(v)} R(w)}{2} \quad \text{для } v \neq b, r.$$

Мы не случайно использовали вновь обозначение  $R(v)$ . Основным результатом классической теории игры Ричмана как раз и является доказательство эквивалентности этих определений для  $R(v)$ , к которому мы сейчас переходим.

**ЛЕММА 1.** Пусть  $R(v)$  — функция Ричмана данного графа,  $\mathbf{B} > R(v_0)$ . Тогда Синий, начиная игру с вершины  $v_0$ , имеет беспроигрышную стратегию.

**ДОКАЗАТЕЛЬСТВО.** Докажем, что Синий может играть так, что в любой момент его текущий капитал будет превосходить значение функции Ричмана в текущей вершине. Пусть в какой-то момент фишка находится в вершине  $v$ , а Синий имеет капитал  $\mathbf{B} > R(v)$ . Тогда он должен сделать ставку

$$s = \frac{\max_{w \in U(v)} R(w) - \min_{w \in U(v)} R(w)}{2}.$$

Исходя из определения функции Ричмана, получим, что

$$R(v) + s = \max_{w \in U(v)} R(w)$$

и

$$R(v) - s = \min_{w \in U(v)} R(w).$$

Если аукцион выиграет Синий, то, передвинув фишку в вершину с наименьшим значением функции Ричмана среди последователей вершины  $v$ , он обеспечит выполнение условия  $\mathbf{B} > R(v)$ , так как  $\mathbf{B} - s > \min_{w \in U(v)} R(w)$ . Если же аукцион выиграет Красный, то, куда бы он ни передвинул фишку, значение функции Ричмана в новой вершине не превзойдет  $\max_{w \in U(v)} R(w)$  (так как  $\mathbf{B} + s > \max_{w \in U(v)} R(w)$ ), поэтому условие  $\mathbf{B} > R(v)$  сохранится и в этом случае.

Итак, у Синего есть стратегия, которая гарантирует выполнение условия  $\mathbf{B} > R(v)$ . Но так как  $R(r) = 1$ , а  $\mathbf{B} \leq 1$ , фишка не сможет попасть в красную вершину, а это и означает, что у Синего есть беспроигрышная стратегия.

**ЗАМЕЧАНИЕ 1.** Если наименьшее значение функции Ричмана достигается в нескольких последователях, то описанная выше стратегия неоднозначна. Лемма 1 верна при произвольном выборе последователя с минимальным значением функции Ричмана.

Для дальнейшего удобно считать, что Синий выбирает такого последователя, который *ближе* к вершине  $b$  (длина кратчайшего пути меньше).

Эта лемма дала бы нам возможность проанализировать игру, если бы мы могли в явном виде построить хотя бы одну функцию Ричмана для данного графа.

Пусть  $M_b(v)$  и  $M_r(v)$  — множества таких  $\mathbf{B}$ , что беспроигрышную стратегию при начальной вершине  $v$  и начальном капитале Синего  $\mathbf{B}$  имеют соответственно Синий и Красный. Пусть  $I(v) = \inf M_b(v)$ ,  $S(v) = \sup M_r(v)$ .

**ЛЕММА 2.**  $I(v)$  и  $S(v)$  — функции Ричмана данного графа.

**ДОКАЗАТЕЛЬСТВО.** Пусть Синий имеет беспроигрышную стратегию. Пусть в какой-то момент фишка находилась в вершине  $v$ , а капитал Синего был равен  $\mathbf{B}$ . Пусть ставка Синего равна  $s$ . Если Синий выиграет аукцион, то его новый капитал должен быть не меньше минимально возможного нового значения  $I$ :

$$\begin{aligned} \mathbf{B} - s &\geq \min_{w \in U(v)} I(w), \\ s &\leq \mathbf{B} - \min_{w \in U(v)} I(w). \end{aligned} \tag{1}$$

С другой стороны, если у Красного недостаточно денег, чтобы перекупить ставку, то  $\mathbf{R} < s$ , откуда

$$s > 1 - \mathbf{B}, \tag{2}$$

а иначе он сможет сдвинуть фишку в худшем случае в вершину с наибольшей  $I(v)$ :

$$\mathbf{B} + s \geq \max_{w \in U(v)} I(w),$$



т. е.

$$s \geq \max_{w \in U(v)} I(w) - \mathbf{B}. \quad (3)$$

Итак, у Синего будет беспроегрешная стратегия тогда и только тогда, когда найдется такое  $s$ , что выполняется (1) и либо (2), либо (3). Так как (3) следует из (2) (поскольку  $I(v) \leq 1$ ), необходимо и достаточно, чтобы

$$\mathbf{B} - \min_{w \in U(v)} I(w) \geq \max_{w \in U(v)} I(w) - \mathbf{B},$$

откуда

$$\mathbf{B} \geq \frac{\min_{w \in U(v)} I(w) + \max_{w \in U(v)} I(w)}{2}.$$

Полученное неравенство вместе с определением  $I(v) = \inf M_b(v)$  приводит к равенству

$$I(v) = \frac{\min_{w \in U(v)} I(w) + \max_{w \in U(v)} I(w)}{2}.$$

Ясно, что  $R(b) = 0$  (уже находясь в  $b$ , Синий выигрывает),  $R(r) = 1$  (находясь в  $r$ , он проигрывает), т. е.  $I(v)$  — действительно функция Ричмана данного графа.

Аналогично и  $S(v)$  — функция Ричмана, т. е. согласно лемме 1, если  $\mathbf{B} > S(v)$ , то Синий имеет беспроегрешную стратегию, а значит, по определению  $S(v) \geq I(v)$ .

Лемма 1 не дает гарантий того, что игра не будет продолжаться бесконечно: она гарантирует одному из игроков не выигрешную, а лишь беспроегрешную стратегию. Вопрос о выигрешной стратегии решает следующая лемма.

**ЛЕММА 3.** Пусть  $R(v)$  — функция Ричмана данного графа, начальная вершина —  $v_0$ ,  $\mathbf{B} > R(v_0)$ . Тогда Синий имеет выигрешную стратегию.

**ДОКАЗАТЕЛЬСТВО.** Пусть  $d$  — наибольшая длина ориентированного пути из произвольной вершины графа в  $b$ ,  $\varepsilon = \frac{\mathbf{B} - R(v_0)}{2^{d+1}}$ . Синий будет действовать согласно беспроегрешной стратегии из леммы 1, но прибавляя к своим ставкам с каждым разом все больший добавок:  $\varepsilon$ ,  $2\varepsilon$ ,  $4\varepsilon$ , ..., до тех пор, пока его не перекупит Красный. Пока аукцион выигрывает Синий, после каждого его хода уменьшается значение функции Ричмана или расстояние до вершины  $b$ . Так что Синий будет двигаться по некоторому ориентированному пути в  $b$ , поэтому Красный обязан будет в какой-то момент перекупить его ставку. Пусть в этот момент прибавок был равен  $2^k \varepsilon$ . Тогда всего Красный получил от Синего  $\varepsilon \times (1 + 2 + 4 + 8 + \dots + 2^{k-1}) = (2^k - 1)\varepsilon$ , т. е. в итоге Синий получит прибыль  $\varepsilon$ . Итак, действуя таким образом, Синий на каждой перекупке Красного будет получать одну и ту же прибыль, т. е. капитал Красного будет неограниченно падать. Так как это не может продолжаться до бесконечности, то в какой-то момент Красный уже не сможет перекупить ставку, т. е. Синий сможет довести фишку до  $b$  и выиграть. Итак, у Синего есть выигрешная стратегия.

Так как  $I(v)$  — функция Ричмана, то при начальном капитале Синего, большем  $I(v)$ , Красный игрок не может иметь беспроегрешной стратегии, т. е.  $S(v) \leq I(v)$ , и, значит,  $S(v) = I(v)$ . Кроме того, если задана какая-то функция

Ричмана  $R(v)$  на графе, то  $1 - R(v)$  — «функция Ричмана для Красного игрока», т. е. аналогично лемме 3, при  $\mathbf{R} > 1 - R(v)$  (т. е.  $\mathbf{B} < R(v)$ ) Красный имеет выигрышную стратегию, а при  $\mathbf{B} > R(v)$  Синий имеет выигрышную стратегию, т. е.  $R(v) = I(v) = S(v)$ .

Остался неисследованным случай  $\mathbf{B} = R(v)$ . Но в нем динамика игры понятна, исходя из сказанного выше: игроки должны ставить ровно ту сумму, которая им предписывается беспроигрышной стратегией. Если один из игроков поставит больше или меньше, то его противник превысит свою функцию Ричмана, т. е. сможет выиграть. Поэтому исход каждого аукциона будет определяться броском монеты, т. е. ни у одного из игроков нет выигрышной стратегии.

Итак, получен основной результат классической теории игры Ричмана: *функция Ричмана  $R(v)$  существует и единственна; при  $\mathbf{B} > R(v)$  Синий имеет выигрышную стратегию, при  $\mathbf{B} < R(v)$  ее имеет Красный, при  $\mathbf{B} = R(v)$  ни один из игроков не имеет даже беспроигрышной стратегии.*

Классическая игра Ричмана допускает различные обобщения: можно сделать информацию о противнике неполной (например, игрок не будет знать начального капитала противника), добавить третьего игрока и так далее.

#### 4. ИГРА РИЧМАНА С НЕПОЛНОЙ ИНФОРМАЦИЕЙ

Предположим, что Синий не знает начального капитала противника, но знает, что его относительный капитал достаточно велик:  $\frac{\mathbf{B}}{\mathbf{B} + \mathbf{R}} > R(v)$  (т. е. он бы смог гарантированно выиграть, зная ставку противника). Докажем, что и в этом случае у него есть беспроигрышная стратегия. Синий игрок должен считать, что капитал Красного равен  $R_0 = \frac{\mathbf{B}}{R(v)} - \mathbf{B}$  (т. е. больше, чем на самом деле). Исходя из этого, Синий может строить свою беспроигрышную стратегию, аналогично стратегии из леммы 1. Так как капитал Красного на самом деле еще меньше, Красный не сможет помешать Синему, т. е. такая стратегия останется беспроигрышной и в этом случае.

Если игровое поле не имеет ориентированных циклов, то игра не сможет продолжаться бесконечно, и, так как Синий не сможет проиграть, он выигрывает.

В общем случае Синий также имеет стратегию, дающую вероятность выигрыша Синего, равную 1. Пусть он будет действовать согласно своей беспроигрышной стратегии, следя за тем, какие именно ставки делает Красный. Если ставка Красного в какой-то момент перестанет быть равной ставке Синего и будет отличаться от нее на  $\varepsilon$ , то Синий после этого хода будет твердо знать, что его относительный капитал превосходит  $R(v)$  по меньшей мере на  $\frac{\varepsilon}{\mathbf{B} + R_0} = \alpha$ , что позволяет ему использовать выигрышную стратегию из леммы 3. Если же Красный всегда будет повторять ставку Синего, то каждый аукцион будет решаться бросанием монетки. Тогда очередность хода будет определяться случайно. Пусть  $N$  — наибольшая длина пути из произвольной вершины графа в  $b$ . Синему достаточно  $N$  раз подряд выиграть монетку, чтобы выиграть всю игру (так как он не может проиграть). Вероятность выигрыша на первой серии из  $N$  бросков равна  $\frac{1}{2^N}$ , т. е. вероятность проигрыша на одной серии равна

$1 - \frac{1}{2^N}$ , а на  $k$  сериях —  $(1 - \frac{1}{2^N})^k$ , то есть вероятность проигрыша вообще равна  $\lim_{k \rightarrow \infty} (1 - \frac{1}{2^N})^k = 0$ , т.е. Синий достоверно выигрывает. И так, игра с неполной информацией устроена почти так же, как и классическая игра Ричмана.

## 5. О ДОКАЗАТЕЛЬСТВЕ СУЩЕСТВОВАНИЯ И ЕДИНСТВЕННОСТИ ФУНКЦИИ РИЧМАНА БЕЗ ИСПОЛЬЗОВАНИЯ ИГРЫ РИЧМАНА

Определим оператор Ричмана  $\rho$  как оператор, сопоставляющий каждому набору чисел  $M$  на вершинах графа другой набор чисел  $\rho(M)$  такой, что для любой черной вершины  $v$  выполнено  $\rho(M)(v) = (M_+(v) + M_-(v))/2$ , где  $M_+(v)$  и  $M_-(v)$  — максимальное и минимальное среди значений  $M$  в потомках  $v$  (здесь и далее подразумевается, что на синей вершине всегда стоит 0, а на красной — 1). По определению, функция Ричмана — это неподвижная точка оператора Ричмана, т.е. такой набор чисел  $R$ , что  $\rho(R) = R$ .

Существование  $R$  сразу следует из топологических соображений. В самом деле,  $\rho$  — непрерывный оператор. Наборы чисел на вершинах графа можно рассматривать как точки многомерного куба. Тогда оператор  $\rho$  — непрерывное отображение многомерного куба в себя, а такое отображение по теореме Брауэра имеет неподвижную точку — функцию Ричмана.

Докажем единственность функции Ричмана. Во-первых, заметим, что из любой вершины можно прийти в синюю, переходя из некоторой вершины  $v$  только в такие вершины  $u$ , что  $R(u) = R_-(v)$ . Действительно, при таком переходе значение  $R$  в текущей вершине не возрастает, а из любой вершины  $v$  достаточно идти по некоторому пути в синюю, чтобы встретить вершину, из которой можно будет выйти в вершину со строго меньшим значением  $R$ . Далее, предположим наличие двух функций Ричмана —  $R^1$  и  $R^2$ . Докажем, что  $D = R^1 - R^2$  неположительно. Пусть  $D$  принимает максимальное значение в вершине  $v$ . Тогда несложно убедиться, что если  $R^1(u) = R^2_-(v)$  и  $u$  — потомок  $v$ , то  $D(u) = D(v)$ . Таким образом можно дойти до синей вершины и доказать, что  $D(b) = D(v)$ , откуда  $D(v) = 0$  и во всех вершинах  $D$  неположительно. Аналогично доказывается, что  $D$  неотрицательно, откуда  $D = 0$  и  $R^1 = R^2$ .

## 6. ВЕРОЯТНОСТНАЯ ИНТЕРПРЕТАЦИЯ ИГРЫ РИЧМАНА

Наряду с классической, «денежной» игрой Ричмана существует также другая игра, приводящая к той же функции.

Как и в классической игре, поле игры — это ориентированный граф с двумя выделенными вершинами, цель Синего и Красного — привести фишку в вершину своего цвета. Однако очередность хода определяется подбрасыванием монеты.

Оказывается, что функция Ричмана  $R(v)$  — это вероятность выигрыша Красного «при правильной игре», т.е. у Красного существует такая стратегия игры, что при любой игре Синего вероятность выигрыша Красного не меньше  $R(v)$ . С другой стороны, у Синего существует стратегия, гарантирующая ему вероятность выигрыша не меньше  $1 - R(v)$ .

Поскольку исход игры от предыстории не зависит, стратегия должна указывать, куда делается ход из данной вершины, т.е. это функция  $\mathcal{R}: V \rightarrow V$ , сопоставляющая вершине одного из ее потомков<sup>2)</sup>.

Пусть  $R(v)$  — функция Ричмана для данного графа с отмеченными вершинами  $b$  и  $r$ . Тогда оптимальная стратегия Красного  $\mathcal{R}^*$  состоит в том, чтобы из вершины  $v$  делать ход в вершину  $w$ , на которой достигается максимум  $R(\cdot)$  среди всех потомков  $v$  (аналогично, оптимальная стратегия Синего состоит в выборе вершины с минимальным значением  $R(v)$ ). Докажем это для оптимальной стратегии Синего, для оптимальной стратегии Красного — аналогично.

Прежде всего заметим, что вероятность того, что игра не закончится в  $k$  ходов, стремится к 0 при  $k \rightarrow \infty$ . В самом деле, для данного графа существует некоторое число  $d$  такое, что Красный, сделав  $d$  ходов подряд, выигрывает. Но вероятность выпадения серии из  $d$  одинаковых результатов после  $k$  подбрасываний монеты стремится с ростом  $k$  к 1.

Пусть Синий придерживается некоторой стратегии  $\mathcal{B}$ . Тогда для каждой вершины  $v$  определена вероятность  $p^{\mathcal{R}^*, \mathcal{B}}(v)$  выигрыша Красного. Докажем, что  $\Delta(v) = p^{\mathcal{R}^*, \mathcal{B}}(v) - R(v) \geq 0$ . Действительно, из определения функции Ричмана следует, что

$$\Delta(v) \geq \frac{\Delta(\mathcal{R}^*(v)) + \Delta(\mathcal{B}(v))}{2},$$

а  $\Delta(b) = \Delta(r) = 0$ . Отсюда следует, что  $\Delta(v) \geq 0$ . Доказательство аналогично приведенному выше доказательству единственности функции Ричмана: возьмем минимум  $\Delta(\cdot)$ , для него неравенство, написанное выше, превращается в равенство, откуда следует, что на  $r$  функция  $\Delta$  принимает минимальное значение.

**ЗАМЕЧАНИЕ 2.** Мы ограничились анализом случая, когда Синий придерживается некоторой стратегии. Нетрудно видеть, что использованное выше неравенство не изменится, если предположить, что ход Синего зависит не только от вершины, но и от номера хода.

Рассмотрение вероятностной игры Ричмана позволяет сделать понятие стратегии игрока более простым, поскольку вместо двух решений — куда пойти и сколько поставить — игрок принимает лишь одно.

## 7. АЛГОРИТМ НАХОЖДЕНИЯ ФУНКЦИИ РИЧМАНА ДЛЯ НЕОРИЕНТИРОВАННОГО ГРАФА

Существует интересный алгоритм, позволяющий быстро (за время, полиномиально ограниченное размером графа) находить функцию Ричмана для неориентированного графа (т.е. графа, у которого на всех ребрах стрелки двусторонние).

Дадим вначале неформальное описание этого алгоритма. Предположим, что имеется модель графа, сделанная из тяжелых шариков и ниток одинаковой длины. Разместим эту модель в вертикальной плоскости. Каждый шарик будет стремиться повиснуть примерно посередине между самым левым и самым правым

<sup>2)</sup>В теории игр рассматриваются и *смешанные* стратегии, когда игрок подбрасывает монетку для принятия решения. В данном случае смешанные стратегии не нужны — оптимальная стратегия оказывается *чистой*.

своим соседом. Таким образом, горизонтальные координаты шариков примерно дают функцию Ричмана. Физически очевидно, что если самый левый и самый правый шарик, соответствующие синей и красной вершинам, оттянуть в разные стороны, то натянутся нитки по самому короткому пути между ними. Значит, можно ожидать, что на самом коротком пути между синей и красной вершиной функция Ричмана образует арифметическую прогрессию. Если постепенно удлинять нитки, образующие кратчайший путь (оставляя их натянутыми), то еще один путь в графе натянется. Продолжая удлинение ниток, можно узнать функцию Ричмана на всем графе.

Теперь опишем детали алгоритма более формально.

Алгоритм работает тактами, перед каждым тактом в части вершин значение функции Ричмана уже определено (такие вершины будем называть *пройденными*). Перед первым тактом пройденными вершинами будут только синяя и красная. За каждый такт работы алгоритм присоединяет к пройденным одну или несколько вершин. Таким образом, в конце работы алгоритма все вершины будут пройдены.

Опишем теперь такт работы алгоритма. Назовем *наклоном* некоторого пути от пройденной вершины к другой пройденной вершине, не содержащего промежуточных пройденных вершин, отношение разности известных алгоритму значений функции Ричмана на концах пути к количеству ребер в пути. Другими словами, наклон пути — это разность арифметической прогрессии, «натянутой» на его концы с заданными значениями. Алгоритм находит среди всех путей по непройденным вершинам, соединяющих две пройденные вершины, путь с максимальным наклоном. Для нахождения такого пути достаточно рассмотреть все кратчайшие пути между пройденными вершинами, поскольку от сокращения пути наклон только возрастает. Найдя путь с максимальным наклоном, алгоритм делает его промежуточные вершины пройденными, значения функции Ричмана в этих вершинах образуют арифметическую прогрессию вместе с уже известными значениями на концах пути.

Докажем по индукции корректность работы алгоритма. На 0-м шаге известные числа, 0 и 1, найдены, разумеется, правильно. Пусть перед каким-то тактом все известные алгоритму значения функции Ричмана найдены правильно; докажем, что на этом такте построенные алгоритмом значения также правильны.

Пусть  $v_1 - v_2$  — ребро с максимальной разностью значений функции Ричмана в  $v_1$  и  $v_2$  среди всевозможных ребер с концами в непройденных вершинах  $v_1$  и  $v_2$ . Предположим для определенности, что значение функции Ричмана в  $v_2$  больше. Тогда  $v_2$  — потомок  $v_1$  с максимальным значением функции Ричмана среди всех потомков. Пусть  $v_0$  — потомок  $v_1$  с минимальным значением функции Ричмана.  $R(v_1) = \frac{1}{2}(R(v_0) + R(v_2))$ , откуда  $R(v_1) - R(v_0) = R(v_2) - R(v_1)$ . Аналогично, пусть  $v_{k+1}$  — сосед  $v_k$  с максимальным значением  $R$ ,  $v_{k-1}$  — сосед  $v_k$  с минимальным значением. Последовательность обязательно придет в пройденную вершину и справа, и слева, потому что иначе бы последовательность  $\dots, R(v_{-1}), R(v_0), R(v_1), \dots$  была бы бесконечной (может быть, только в одну сторону) арифметической прогрессией, что противоречит ограниченности  $R$ .

Будем считать, что построение последовательности в обе стороны обрывается на пройденных вершинах. Заметим, что  $R(v_k) = \frac{1}{2}(R(v_{k-1}) + R(v_{k+1}))$ , откуда  $R(v_k) - R(v_{k-1}) = R(v_{k+1}) - R(v_k)$ . Значит, последовательность  $R(v_i)$  является конечной арифметической прогрессией с известными алгоритму концами. Осталось только заметить, что последовательность  $v_i$  — путь с наибольшим наклоном, поскольку его наклон равен длине ребра с самой большой разностью значений  $R$  на концах, а ни один путь не может иметь больший наклон. Тогда алгоритм выберет путь с точно таким же наклоном; но это возможно только в том случае, если значения  $R$  на этом пути образуют арифметическую прогрессию. Доказательство закончено.

Интересно, что этот алгоритм в некоторых случаях позволяет находить функцию Ричмана на бесконечном графе, заодно доказывая ее единственность (а для бесконечного графа функция Ричмана может быть не единственна!). Для этого необходимо найти способ отыскания кратчайших путей на заданном бесконечном графе, из которого выброшено конечное число вершин (конечно, такой способ не всегда существует). Например, для бесконечной квадратной сетки несложно описать требуемый способ. После этого алгоритм будет работать как обычно, хотя, конечно, в общем случае невозможно гарантировать, что для какой-то вершины алгоритм рано или поздно пройдет через нее.

Чтобы доказать правильность работы алгоритма в случае бесконечного графа, приведенное доказательство потребует слегка изменить.

Предположим, что алгоритм на некотором шаге ошибся и значения  $R$  на выбранном им пути не образуют арифметической прогрессии. Тогда выбранный алгоритмом путь содержит ребро  $v_1 - v_2$ , для которого  $R(v_2) - R(v_1)$  больше наклона пути, выбранного алгоритмом. Можно построить последовательность  $v_k$  аналогично приведенному выше построению. Заметим, что  $R(v_k)$  при  $k > 1$  — среднее арифметическое  $R(v_{k+1})$  и чего-то еще, меньшего или равного  $R(v_{k-1})$ , откуда  $R(v_{k+1}) - R(v_k)$  не меньше  $R(v_k) - R(v_{k-1})$ . Аналогичное рассуждение при  $k < 2$  показывает, что  $R(v_k) - R(v_{k-1})$  не меньше  $R(v_{k+1}) - R(v_k)$ . Значит, при всех допустимых  $k$   $R(v_{k+1}) - R(v_k)$  не меньше  $R(v_2) - R(v_1)$ . Поэтому наклон пути  $v_i$  больше наклона пути, построенного алгоритмом, что противоречит описанию алгоритма. Доказательство закончено.

## 8. ОБ АЛГОРИТМАХ НАХОЖДЕНИЯ ФУНКЦИИ РИЧМАНА

В настоящее время не известно алгоритма для нахождения функции Ричмана, про который было бы доказано, что он заканчивает работу за полиномиальное от числа вершин графа время.

Простейший алгоритм поиска функции Ричмана — это перебор всевозможных перестановок  $n$  чисел (здесь и далее  $n$  — количество вершин графа, исключая синюю и красную). Каждая перестановка  $(a_i)$  соответствует предположению  $R(v_{a_1}) \leq R(v_{a_2}) \leq \dots \leq R(v_{a_n})$ , где  $v_i$  — вершины графа. Чтобы в этом предположении найти  $R$ , достаточно заметить, что для каждой вершины ее потомок, в котором достигается максимальное (минимальное) значение, уже определился. Тогда определение функции Ричмана превращается в систему

линейных уравнений вида

переменная = среднее арифметическое двух других переменных.

Решив ее, получим значения предполагаемой функции Ричмана. Нетрудно проверить, являются ли они в действительности нужными значениями. Если нет, предположение отбраковывается.

Этот алгоритм, как нетрудно видеть, работает за время  $O(n!)$ . Однако, известны алгоритмы, работающие за экспоненциальное время, т.е. значительно быстрее. Опишем идею одного из таких алгоритмов.

Рассмотрим на заданном графе некоторый набор чисел,  $M$ , например, нули (естественно, в красной вершине по-прежнему 1). Примем  $M$  за нулевое приближение к функции Ричмана. На каждом шаге будем переходить от полученного на данном шаге набора чисел  $N$  к следующему приближению  $\rho(N)$ .

Предположим, что текущее приближение задает такой же порядок на вершинах, что и значения функции Ричмана. Описанным выше способом можно это проверить, решив систему уравнений. Если найденное решение системы является функцией Ричмана, то работа закончена, иначе переходим к следующему шагу.

Нельзя не отметить одну тонкость, связанную с работой такого алгоритма. По набору чисел  $N$ , полученному на очередной итерации, для построения системы уравнений необходимо определить порядок вершин. Разумеется, если число в одной вершине меньше числа в другой, то и эта вершина будет стоять раньше в искомом порядке. Однако, может так случиться, что в нескольких вершинах значения будут одинаковыми. Как в таком случае определить их порядок?

Оказывается, что алгоритм работает корректно не при любом правиле упорядочения вершин с одинаковым значением  $N$ . Приведем пример правила, которое гарантирует корректность алгоритма. Для группы вершин с одинаковым значением  $N$  положим, что та из них «меньше», из которой кратчайший путь к вершинам с меньшим значением  $N$  короче, если же длина этих путей одинакова, то кто меньше — неважно.

Заметим, что не любой набор чисел годится в начальное приближение: необходимо, чтобы для любой вершины существовал путь в синюю вершину, на котором числа начального приближения убывают.

У описанной идеи есть огромное количество вариантов вычисления следующего приближения. К примеру, можно в качестве следующего приближения использовать полученное решение системы уравнений, а можно придать каждому числу скорость в направлении среднего арифметического минимума и максимума в потомках, а затем подождать до первого обгона одного числа другим. Можно также задействовать нестандартные игровые элементы, изменяющие саму функцию Ричмана, вроде разрешения Синему до подбрасывания монетки выбрать из нескольких возможных вариантов набор ребер, ведущих из текущей вершины, или устройства, с некоторой вероятностью перебрасывающее фишку из одной вершины в другую (главное — следить, чтобы вероятность ничьей была 0). Затем в ходе работы алгоритма нестандартные элементы постепенно исчезают, и игра превращается в обычную. К сожалению, про все подобные алгоритмы известно мало, иногда даже не известно доказательство окончания их работы.

## 9. ИГРА РИЧМАНА ДЛЯ ТРЕХ ИГРОКОВ

Добавление третьего игрока вносит качественные изменения в игру Ричмана, делая ее несводимой к классической игре.

Это направление исследовалось сравнительно мало. Существуют различные варианты правил игры для трех игроков. Третий игрок (далее мы будем называть его Зеленым) получает свою «конечную» вершину ( $g$ ) и начальный капитал  $\mathbf{G}$ . (Как и в игре двух игроков предполагается, что из любой вершины графа есть путь, ведущий в вершины  $r, g, b$ .) По сути в игре меняется лишь алгоритм проведения аукциона. По одной из версий игрок, сделавший наибольшую ставку, отдает ее тому игроку, ставка которого вторая по величине; по другой — его ставка делится между остальными пропорционально их ставкам.

Здесь мы рассмотрим только первую версию правил. В качестве пространства финансовых состояний игроков используем равносторонний треугольник с вершинами  $\mathbf{B}$ ,  $\mathbf{R}$  и  $\mathbf{G}$ . Состояние игроков будет изображаться в нем точкой, расстояния от которой до сторон равны капиталам игроков, которым принадлежат вершины, противоположные этим сторонам.

Таким образом, начальные условия игры определяются начальной вершиной графа и точкой в треугольнике состояний. В отличие от случая двух игроков, могут существовать целые области в треугольнике состояний, в которых ни у одного из игроков нет беспроигрышной стратегии. Ниже мы приводим такой пример.

Начнем с нескольких предварительных замечаний общего характера.

Мы интересуемся лишь выигрышными или беспроигрышными стратегиями для одного из игроков, скажем, Синего. Тогда можно считать, что Красный и Зеленый действуют в коалиции. Рассматривая коалиционную игру, можно конечные вершины  $r$  и  $g$  отождествить (обозначать ее будем  $r$ ). Будем называть такое преобразование графа *коалиционной заменой*. При этом нельзя забывать, что если ставки и Красного, и Зеленого больше по величине ставки Синего, то Красный передает свою ставку Зеленому (или наоборот).

Один из возможных вариантов стратегии для коалиции состоит в том, что один из коалиционеров делает все время нулевые ставки. Тогда игра будет эквивалентна обычной игре для двух игроков. Обозначим функцию Ричмана в этой игре через  $R(v)$ . Из классической теории Ричмана следует, что если  $\frac{\mathbf{B}}{\mathbf{B} + \mathbf{R}} < R(v)$  ( $\frac{\mathbf{B}}{\mathbf{B} + \mathbf{G}} < R(v)$ ), то Синий проигрывает. Есть и специфический для игры трех игроков случай: если  $\mathbf{R} > 2\mathbf{B}$  и  $\mathbf{G} > \mathbf{B}$ , то у коалиции есть выигрышная стратегия. Красный и Зеленый ставят по  $\mathbf{B} + \varepsilon$ , Синий не может их перекупить. Ставка переходит к Зеленому, у которого станет  $\mathbf{B} + \mathbf{G} + \varepsilon > 2\mathbf{B}$ , а у Красного  $\mathbf{R} - \mathbf{B} > -\varepsilon\mathbf{B}$ . Условие сохраняется, и коалиция бесплатно сделала ход. Действуя таким образом, коалиционеры могут без потерь суммарного капитала прийти в  $r$  и выиграть.

Рассмотрим теперь игру трех игроков на графе, изображенном на рисунке 3а) (начальная вершина отмечена). Области существования беспроигрышных (как выяснится, даже выигрышных) стратегий отмечены для каждого из игроков на рисунке 4. Из этого рисунка видно, что есть область (выпуклый пятиугольник) в которой ни у одного из игроков нет беспроигрышной стратегии.



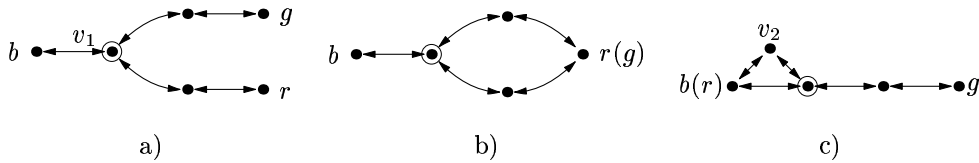


Рис. 3.

После коалиционных замен (отождествления конечных вершин коалиционеров) из него получаются графы на рисунках 3b) и 3c). Приводя их к случаю игры Синего против коалиции Красного и Зеленого, как описано выше, получаем в обоих случаях коалиционную игру на графе, изображенном на рисунке 2, но с разными начальными вершинами.

Рассмотрим эту игру (Синий против коалиции Красного и Зеленого на графе рисунка 2). Пусть начальная вершина —  $v_1$ . Из сказанного выше ясно, что если  $\mathbf{R} > 2\mathbf{B}$  ( $\mathbf{G} > 2\mathbf{B}$ ), то коалиция выигрывает (так как  $R(v_1) = 1/3$ ). Далее, если  $\mathbf{R} > \mathbf{B}$  и  $\mathbf{G} > \mathbf{B}$ , то коалиция тоже выигрывает: сначала Красный  $\mathbf{B}$ , а Зеленый — ничего. Тогда у Синего станет  $2\mathbf{B}$ , у Зеленого —  $\mathbf{G} > \mathbf{B}$  и фишка передвинется, и, так как  $R(v_2) = 2/3$ , Синий проиграет.

Предположим, что ни одно из этих условий не выполнено и ни одно из указанных неравенств не обращается в равенство, т. е. справедливы следующие условия:  $\mathbf{R} < 2\mathbf{B}$ ,  $\mathbf{G} < 2\mathbf{B}$ , и  $\mathbf{R} < \mathbf{B}$  или  $\mathbf{G} < \mathbf{B}$  (невывуклая область на рисунке 4). Тогда у Синего есть выигрышная стратегия. Первым ходом Синий ставит на аукцион весь свой капитал. Если его никто не перекупит, то он выиграет. Без ограничения общности считаем, что его перекупит Красный. Тогда  $\mathbf{R} > \mathbf{B}$ , откуда следует, что  $\mathbf{G} < \mathbf{B}$ . После сдвига фишки капиталы и Красного, и Зеленого стали меньше, чем  $\mathbf{B}$ . Теперь Синий подбирает  $\varepsilon > 0$  так, чтобы  $s = \max(\mathbf{R}_1, \mathbf{G}_1) + \varepsilon < \mathbf{B}$ , где  $\mathbf{R}_1$  и  $\mathbf{G}_1$  — новые капиталы Красного и Синего соответственно, и делает ставку  $s$ . Он выиграет аукцион, сдвинет фишку и отдаст свою ставку одному из коалиционеров. Так как  $s < \mathbf{B}$ , у него останется больше его начального капитала. Далее он ставит  $2\mathbf{B} - s$ . Перекупить его сможет лишь тот, кому он отдал деньги, а значит, после очередного сдвига фишки у Синего станет  $4\mathbf{B} - 2s$ , у одного из его противников капитал понизится, а у другого останется неизменным. Итого, капитал Синего повысится на  $2\mathbf{B} - 2s$ . Так как на следующем цикле ставка Синего не уменьшится, его капитал будет все быстрее

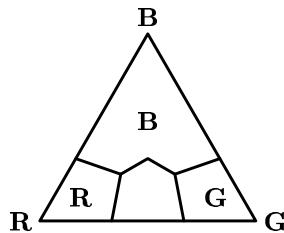


Рис. 4.

расти, и в какой-то момент он сможет, находясь в вершине  $v_1$ , перекупить обоих противников.

Для начальной вершины  $v_2$  аналогичным рассуждением доказывается, что для выигрыша Синего необходимо и достаточно, чтобы  $\mathbf{B} > 2\mathbf{R}$  и  $\mathbf{B} > 2\mathbf{G}$ . Для исходного графа на рисунке 3а) получаем четырехугольные области, отмеченные на рисунке 4 буквами  $\mathbf{R}$  и  $\mathbf{G}$ .

### БЛАГОДАРНОСТИ

Авторы благодарны М. Н. Вялomu за предложение написать эту статью для «Математического просвещения»; А. Фроленкову за обсуждения; А. Акопяну за многочисленные обсуждения и полезные предложения. Авторы также хотят поблагодарить Н. Н. Константинова за замечательную идею летних конференций Турнира Городов.

### СПИСОК ЛИТЕРАТУРЫ

- [1] Материалы XII Летней Конференции Турнира Городов (электронная версия: [www.mcsme.ru/olympiads/lktg](http://www.mcsme.ru/olympiads/lktg)).
- [2] *A. J. Lazarus, D. E. Loeb, J. G. Propp, D. Ullman* Richman games. Cambridge University Press, 1995.

# Олимпиады

---

---

## Математические соревнования имени Уильяма Лоуэлла Путнама

Е. М. Бронштейн

Мистер Уильям Лоуэлл Путнам в 1882 году получил в Гарварде степень бакалавра по математике и в течение ряда лет возглавлял гостевой департамент математического факультета Гарварда. Он был организатором соревнований, получивших название путнамовских. Не будет преувеличением сказать, что в них он увековечил свое имя.

Путнамовские соревнования являются одними из самых интересных интеллектуальных конкурсов. Идеи, положенные в основу их проведения, Путнам изложил в статье, опубликованной в 1921 году. Основная идея — командный подход, при котором участники защищают цвета своих университетов и колледжей. Во исполнение завещания супруга, мадам Путнам в 1927 году учредила фонд для поддержки проведения соревнований. Центром организационной деятельности стал Гарвард, в то время Президентом Гарварда был Л. Лоуэлл, брат мадам Путнам. Надо сказать, вначале речь не шла о проведении лишь математических соревнований, проводились экспериментальные соревнования по английскому языку, но, видимо, под влиянием выдающихся математиков отца и сына Биргофов, Пойа, Радо и других было решено проводить именно математические соревнования. Джордж Дэвид Биргоф сформулировал несколько основных принципов их проведения на которых базируется вся эта работа:

1. Состязания должны быть лично-командными. Команда каждого университета (колледжа) состоит из трех студентов, но участие могут принимать и другие студенты любого университета (колледжа) США и Канады. При этом подводятся как командные, так и личные итоги.
2. Общее руководство соревнованиями осуществляет Математическая ассоциация Америки — профессиональное объединение преподавателей математики учебных заведений разных типов.

---

Печатается с любезного разрешения Д. Альбера (МАО).

3. Призами и поощрениями отмечается значительная часть участников. Выделяются первая пятерка, первая десятка и т. д.
4. Одному из пяти лучших участников представляется возможность продолжить обучение (на магистерском уровне) в Гарварде.

Первое соревнование состоялось в 1938 году, и с тех пор они проводятся ежегодно (исключение составили только военные 1943–1945 годы). 4 декабря 1999 года студенты собрались в шестидесятый раз.

Непосредственное руководство конкурсом осуществляет Путнамовский комитет, состоящий из трех человек. В разные годы в состав комитета входили такие выдающиеся математики, как Радо, Пойа, Кац, Капланский (победитель первых соревнований), Келли, Мозер, Беллман, Рота, Коксетер. Комитет разрабатывает задания, проверяет работы участников, публикует результаты. Само соревнование проводится в одно из воскресений (в последние годы — в первое воскресенье декабря). Предварительно по университетам и колледжам рассылаются задания в соответствии с заявками. В день проведения участники собираются в своих учебных заведениях, им предлагается 6 задач (A1–A6) на три часа с утра и 6 задач (B1–B6) на три часа после обеда. Наблюдатель-преподаватель обеспечивает порядок и отправляет работы в комитет. К участию допускаются студенты первой (бакалаврской) ступени, студент не может участвовать в соревнованиях более четырех раз. Интересно, что в итоге (они публикуются в журналах Математической ассоциации Америки) включаются результаты решения каждой задачи лидерами в виде 11-мерного вектора: первая компонента — число решивших задачу полностью, далее число продвинувшихся на соответствующую долю, десятая — число представивших совершенно неверные попытки решения, одиннадцатая — число участников, не представивших решений вообще. Соревнования очень популярны, ежегодно в них принимают участие около 300 команд и свыше 2000 участников. В ряде университетов в учебный план включены специальные семинары по подготовке к соревнованиям. Успехи команд и отдельных студентов — предмет гордости университетов и факультетов.

Тематика задач ограничена стандартными математическими курсами (анализ, алгебра, геометрия, дискретная математика, дифференциальные уравнения и др.). Многие задачи опираются на знание элементарной математики.

Сложность задач весьма сильно варьируется. Наряду с утешительными легкими задачами, доступными подавляющему большинству участников, есть очень трудные, которые поддаются усилиям всего нескольких участников.

Приведем задачи некоторых соревнований разных лет.

A5 (1985). Для каких целых  $1 \leq m \leq 10$  справедливо равенство

$$I_m = \int_0^{2\pi} \cos(x) \cos(2x) \cdot \dots \cdot \cos(mx) dx = 0?$$

A6 (1985). Пусть  $\Gamma(p(x)) = a_0^2 + a_1^2 + \dots + a_n^2$  — функция полинома  $p(x) = a_0 + a_1x + \dots + a_nx^n$ . Найдите полином  $g(x)$  с вещественными коэффициентами такой, что  $g(0) = 1$ ,  $\Gamma(g(x)^n) = \Gamma((3x^2 + 7x + 2)^n)$  для всех натуральных  $n$ .

В3 (1985). Пусть

$$\begin{array}{cccc} a_{1,1} & a_{1,2} & a_{1,3} & \dots \\ a_{2,1} & a_{2,2} & a_{2,3} & \dots \\ a_{3,1} & a_{3,2} & a_{3,3} & \dots \\ \dots & \dots & \dots & \dots \end{array}$$

дважды бесконечный массив натуральных чисел, в котором каждое натуральное число встречается ровно 8 раз. Докажите, что для некоторой пары индексов  $(m, n)$  выполняется неравенство  $a_{m,n} > mn$ .

В6 (1985). Группа по умножению  $\{M_i\}$  вещественных квадратных матриц порядка  $n$  содержит  $r$  элементов. Пусть выполняется равенство  $\sum_{i=1}^r \text{tr}(M_i) = 0$ , где  $\text{tr}$  — след матрицы. Докажите, что матрица  $\sum_{i=1}^r M_i$  — нулевая.

А3 (1986). Вычислите сумму ряда  $\sum_{n=0}^{\infty} \text{arctg}(n^2 - n + 1)$ .

В4 (1986). Пусть  $G(r) = \min_{m,n \in \mathbb{Z}} \left| r - \sqrt{m^2 + 2n^2} \right|$ . Докажите или опровергните утверждение  $\lim_{n \rightarrow +\infty} G(r) = 0$ .

В6 (1986). Пусть  $A, B, C, D$  — квадратные матрицы порядка  $n$  с элементами из поля  $\mathbb{F}$ , такие, что матрицы  $AB', CD'$  — симметрические и  $AD' - BC' = I$ . Здесь  $B'$  — матрица, транспонированная  $B$ ,  $I$  — единичная матрица. Докажите равенство  $A'D - C'B = I$ .

А1 (1987). На плоскости определены множества

$$\begin{aligned} A &= \left\{ (x, y) : x^2 - y^2 = \frac{x}{x^2 + y^2} \right\}; & B &= \left\{ (x, y) : 2xy + \frac{y}{x^2 + y^2} = 3 \right\}; \\ C &= \left\{ (x, y) : x^3 - 3xy^2 + 3y = 1 \right\}; & D &= \left\{ (x, y) : 3yx^2 - 3x - y^3 = 0 \right\}. \end{aligned}$$

Докажите, что  $A \cap B = C \cap D$ .

А5 (1987). Дано векторное поле  $G(x, y) = \left( -\frac{y}{x^2 + 4y^2}, \frac{x}{x^2 + 4y^2}, 0 \right)$ . Докажите или опровергните существование векторного поля:  $F(x, y, z) = (M(x, y, z), N(x, y, z), P(x, y, z))$  со свойствами

- (1) функции  $M, N, P$  имеют непрерывные частные производные при  $(x, y, z) \neq (0, 0, 0)$ ;
- (2)  $\text{rot } F = 0$  при  $(x, y, z) \neq (0, 0, 0)$ ;
- (3)  $F(x, y, 0) = G(x, y)$ .

В4 (1987). Пусть  $x(1) = 0,8, y(1) = 0,6$  и при  $n = 1, 2, 3, \dots$

$$\begin{aligned} x(n+1) &= x(n) \cos(y(n)) - y(n) \sin(y(n)), \\ y(n+1) &= x(n) \sin(y(n)) + y(n) \cos(y(n)). \end{aligned}$$

Вычислите пределы  $\lim_{n \rightarrow \infty} x(n), \lim_{n \rightarrow \infty} y(n)$  или опровергните их существование.

В6 (1987). Пусть  $\mathbb{F}$  — конечное поле с  $p^2$  элементами, где  $p$  — нечетное простое число. Пусть  $S$  — множество, состоящее из  $(p^2 - 1)/2$  различных ненулевых элементов поля, такое, что для всякого ненулевого элемента  $a \in \mathbb{F}$  в точности один из двух элементов  $a, -a$  входит в  $S$ . Пусть  $N$  число элементов в пересечении  $S$  с множеством  $\{2a : a \in S\}$ . Докажите, что число  $N$  четное.

А6 (1988). Пусть линейный оператор  $A$  в  $n$ -мерном векторном пространстве имеет  $(n + 1)$  собственных векторов, любые  $n$  из которых линейно независимы. Следует ли отсюда, что  $A = aI$ , где  $I$  — тождественный оператор,  $a$  — скаляр?

В4 (1988). Пусть ряд  $\sum_{n=1}^{\infty} a_n$  ( $a_n \geq 0$ ) сходится. Докажите сходимость ряда  $\sum_{n=1}^{\infty} (a_n)^{n/(n+1)}$ .

А3 (1989). Докажите, что если  $11z^{10} + 10iz^9 + 10iz - 11 = 0$ , то  $|z| = 1$ .

А6 (1989). Пусть  $a_n = 1$ , если в двоичном разложении числа  $n$  каждый блок идущих подряд нулей состоит из четного числа цифр и 0 в противном случае. Например,  $a_{36} = 1$ , поскольку  $36 = 100100_2$ , а  $a_{20} = 0$ , поскольку  $20 = 10100_2$ . Пусть  $\alpha = 1 + a_1x + a_2x^2 + \dots$  — формальный степенной ряд по модулю 2. Докажите, что  $\alpha^3 + x\alpha + 1 = 0$  по модулю 2.

В6 (1989). Пусть  $(x_1, x_2, \dots, x_n)$  — точка, случайно выбираемая из  $n$ -мерной области  $0 < x_1 < x_2 < \dots < x_n < 1$  в соответствии с равномерным распределением, и  $f(x)$  — функция, непрерывная на  $[0, 1]$ , у которой  $f(1) = 0$ . Положим  $x_0 = 0, x_{n+1} = 1$ . Докажите, что математическое ожидание суммы Римана  $\sum_{i=0}^n (x_{i+1} - x_i)f(x_{i+1})$  равно  $\int_0^1 P(x)f(x) dx$ , где  $P(x)$  — многочлен степени  $n$ , не зависящий от функции  $f(x)$  и такой, что  $0 < P(x) < 1$  при  $0 < x < 1$ .

А3 (1990). Докажите, что любой выпуклый пятиугольник, вершины которого имеют целочисленные координаты, и никакие три из которых не лежат на одной прямой, имеет площадь, не меньшую  $5/2$ .

А5 (1990). Если  $A$  и  $B$  — квадратные матрицы одного порядка, для которых справедливо равенство  $ABAB = 0$ , то обязательно ли  $BABA = 0$ ?

В4 (1990). Пусть  $G$  — конечная группа порядка  $n$ , порожденная элементами  $a$  и  $b$ . Докажите или опровергните следующее утверждение: существует последовательность  $g_1, g_2, \dots, g_{2n}$  такая, что

- (1) каждый элемент группы встречается в последовательности 2 раза;
- (2)  $g_{i+1} = g_i a$  или  $g_{i+1} = g_i b$  при  $i = 1, 2, \dots, 2n$ . (Считать, что  $g_{n+1} = g_1$ .)

А3 (1991). Найдите все вещественные многочлены  $p(x)$  степени  $n \geq 2$ , для каждого из которых существует последовательность вещественных чисел  $r_1 < r_2 < \dots < r_n$  такая, что

- (1)  $p(r_i) = 0$  при  $i = 1, 2, \dots, n$ ,
- (2)  $p'((r_i + r_{i+1})/2) = 0$  при  $i = 1, 2, \dots, n - 1$ .

А5 (1991). Найдите максимальное значение величины  $\int_0^y \sqrt{x^4 + y^2(1-y)^2} dx$  при  $y \in [0, 1]$ .

В3 (1991). Существует ли число  $L$  такое, что если  $m$  и  $n$  целые числа, большие  $L$ , то прямоугольник размеров  $m \times n$  можно разрезать на прямоугольники размеров  $4 \times 6$  и  $5 \times 7$ ?

В4 (1991). Пусть  $p$  — простое нечетное число. Докажите, что

$$\sum_{j=0}^p \binom{p}{j} \binom{p+j}{j} = (2^p + 1) \pmod{p^2}.$$

А3 (1992). Для данного натурального числа  $m$  найдите все тройки натуральных чисел  $(n, x, y)$ , где числа  $m, n$  взаимно простые, для которых справедливо равенство  $(x^2 + y^2)^m = (xy)^n$ .

А5 (1992). Пусть  $a_n = 0(1)$ , если число единиц в двоичном разложении натурального числа  $n$  соответственно четное (нечетное). Докажите, что не существует натуральных  $k$  и  $m$  таких, что  $a_{k+j} = a_{k+m+j} = a_{k+m+2j}$  при  $0 \leq j \leq m-1$ .

### РЕШЕНИЯ ЗАДАЧ

А5 (1985). Применяя формулу Эйлера  $\cos \varphi = \frac{e^{i\varphi} + e^{-i\varphi}}{2}$ , получим

$$\begin{aligned} I_m &= \int_0^{2\pi} \cos(x) \cos(2x) \cdot \dots \cdot \cos(mx) dx = \\ &= \frac{1}{2^m} \int_0^{2\pi} (e^{ix} + e^{-ix}) \cdot \dots \cdot (e^{imx} + e^{-imx}) dx = \\ &= \frac{1}{2^m} \int_0^{2\pi} \sum e^{ix(\pm 1 \pm 2 \dots \pm m)} dx. \end{aligned}$$

Поскольку  $\int_0^{2\pi} e^{ikx} dx = 0$  при целом  $k \neq 0$  и  $\int_0^{2\pi} dx = 2\pi$ , то  $I_m \neq 0$ , если при некоторой расстановке знаков справедливо равенство  $(\pm 1 \pm 2 \pm \dots \pm m) = 0$ .

Докажем, что это возможно тогда и только тогда, когда  $m \equiv 0 \pmod{4}$  или  $m \equiv 3 \pmod{4}$ . Четности чисел  $\pm 1 \pm 2 \dots \pm m$  совпадают при любых расстановках знаков, следовательно, для выполнения равенства  $\pm 1 \pm 2 \dots \pm m = 0$  необходима четность числа  $1 + 2 + \dots + m = m(m+1)/2$ . Это равносильно делимости на 4 одного из чисел  $m, m+1$ . Пусть теперь число  $k = m(m+1)/4$  целое и  $s = \min\{t : t + (t+1) + \dots + m \leq k\}$ . Если  $s + (s+1) + \dots + m = k$ , то  $1 + \dots + (s-1) - s - (s+1) - \dots - m = 0$ . Если  $s + (s+1) + \dots + m < k$ , то число  $p = k - (s + (s+1) + \dots + m)$  удовлетворяет неравенствам  $0 < p < s-1$ . Если плюсы взять у чисел  $p, s, (s+1), \dots, m$  и минусы у остальных, то полученная сумма будет равна нулю. Среди чисел  $1, 2, \dots, 10$  найденным свойством обладают числа  $3, 4, 7, 8$ . Значит, ответ задачи — множество  $\{1, 2, 5, 6, 9, 10\}$ .

A6 (1985). Пусть  $p(x) = a_0 + a_1x + \dots + a_nx^n$  ( $a_n \neq 0$ ). Определим многочлен  $p^T(x) = x^n p(1/x) = a_0x^n + a_1x^{n-1} + \dots + a_n$ . Очевидно, что  $(p^k(x))^T = (p^T(x))^k$ .

Докажем, что для любых многочленов  $p(x) = a_0 + a_1x + \dots + a_nx^n$ ,  $q(x) = b_0 + b_1x + \dots + b_nx^n$  выполняется равенство  $\Gamma(p(x)q(x)) = \Gamma(p^T(x)q(x))$ . Перемножая многочлены  $p(x), q(x)$ , получим

$$\Gamma(p(x)q(x)) = \sum a_i^2 b_j^2 + 2 \sum_{\substack{i \neq j \\ i+n-s=j+n-t}} a_i a_j b_i b_j.$$

То же получим и для

$$\Gamma(p^T(x)q(x)) = \sum a_i^2 b_j^2 + 2 \sum_{\substack{i \neq j \\ i+n-s=j+n-t}} a_i a_j b_i b_j.$$

Положим  $p(x) = x + 2$ ,  $q(x) = 3x + 1$ . Тогда  $p(x)q(x) = 3x^2 + 7x + 2$ ,  $p^T(x)q(x) = 6x^2 + 5x + 1$ . При этом по доказанному ранее  $\Gamma(p(x)q(x))^n = \Gamma(p^T(x)q(x))^n$ . Таким образом, трехчлен  $6x^2 + 5x + 1$  обладает нужными свойствами.

B3 (1985). Пусть, напротив, при всех  $(m, n)$  выполняется неравенство  $a_{m,n} \leq mn$ . Определим множество  $T_k = \{(m, n) : a_{m,n} \leq k\}$ . По условию  $|T_k| = 8k$ . С другой стороны, по предположению  $S_k = \{(m, n) : mn \leq k\} \subset T_k$ . Но

$$|S_k| = k + \left\lfloor \frac{k}{2} \right\rfloor + \left\lfloor \frac{k}{3} \right\rfloor + \dots + 1 \geq \frac{k}{2} \left( 1 + \frac{1}{2} + \dots + \frac{1}{k} \right)$$

при достаточно больших  $k$ . Тем самым при достаточно больших  $k$  должно выполняться неравенство  $\frac{k}{2} \left( 1 + \frac{1}{2} + \dots + \frac{1}{k} \right) \leq 8k$ , что противоречит расходимости гармонического ряда.

B6 (1985). Пусть  $S = \sum_{i=1}^r M_i$ . По условию  $\text{tr}(S) = 0$ . Имеем  $SM_j = \sum_{i=1}^r M_i M_j = S$ , поскольку  $\{M_i\}$  — конечная группа по умножению. Далее,  $S^2 = \sum_{j=1}^r SM_j = \sum_{j=1}^r S = rS$ . Отсюда следует, что собственные числа матрицы  $S$  могут равняться только 0 и  $r$ . Из равенства  $\text{tr}(S) = 0$  следует, что единственное собственное значение матрицы  $S$  равно 0, поскольку  $\text{tr}(S)$  есть сумма собственных значений матрицы  $S$ . А тогда матрица  $S - rI$  не имеет нулевых собственных значений, т. е. обратима. Поэтому из равенства  $S^2 - rS = S(S - rI) = 0$  следует равенство  $S = 0$ , что и требовалось.

A3 (1986). Воспользовавшись формулой  $\text{ctg}(\alpha - \beta) = \frac{1 + \text{ctg} \alpha \text{ctg} \beta}{\text{ctg} \beta - \text{ctg} \alpha}$ , получаем  $\text{ctg}(\text{arcctg} n - \text{arcctg}(n + 1)) = n^2 + n + 1$ . Отсюда

$$\text{arcctg} n - \text{arcctg}(n + 1) = \text{arcctg}(n^2 + n + 1).$$

Следовательно,  $\sum_{n=0}^k \text{arcctg}(n^2 + n + 1) = \text{arcctg} 0 - \text{arcctg} k$ , что дает

$$\sum_{n=0}^{\infty} \text{arcctg}(n^2 + n + 1) = \pi/2.$$



В4 (1986). Докажем справедливость сформулированного утверждения.

Прежде всего, проверим, что для положительного  $k$  справедлива оценка  $k - \lfloor \sqrt{k} \rfloor^2 \leq 3\sqrt{k}$ . Пусть  $p = \lfloor \sqrt{k} \rfloor$ . Тогда  $k = p^2 + s$ , где  $s < 2p + 1$ . Следовательно,  $k - \lfloor \sqrt{k} \rfloor^2 = s < 2p + 1 \leq 3\sqrt{k}$ , что и требовалось.

Пусть  $r > 0$ ,  $m = \lfloor r \rfloor$ . Тогда  $r = m + d$  ( $0 \leq d < 1$ ). Положим  $n = \lfloor \sqrt{md} \rfloor$ . Оценим значение  $r^2 - m^2 - 2n^2$ . Получаем

$$\begin{aligned} (m + d)^2 - m^2 - 2n^2 &= 2md + d^2 - 2\lfloor \sqrt{md} \rfloor^2 = \\ &= 2(md - \lfloor \sqrt{md} \rfloor^2) + d^2 \leq 6\sqrt{md} + d^2 \leq 6\sqrt{r} + 1. \end{aligned}$$

Окончательно имеем

$$G(r) \leq r - \sqrt{m^2 + 2n^2} = \frac{r^2 - m^2 - 2n^2}{r + \sqrt{m^2 + 2n^2}} \leq \frac{6\sqrt{r} + 1}{r}.$$

Поскольку  $\lim_{r \rightarrow \infty} \frac{6\sqrt{r}}{r} = 0$ , нужное утверждение доказано.

В6 (1986). Симметричность матрицы  $AB'$  означает, что  $(AB')' = BA' = AB'$ , симметричность матрицы  $CD'$  — что  $(CD')' = DC' = CD'$ . Из равенства  $AD' - BC' = I$  следует равенство  $(AD' - BC')' = I'$ , т. е.  $DA' - CB' = I$ . Рассматривая произведение блочных матриц  $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ ,  $\begin{pmatrix} D' & -B' \\ -C' & A' \end{pmatrix}$ , получаем

$$\begin{pmatrix} AD' - BC' & -AB' + BA' \\ CD' - DC' & -CB' + DA' \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}.$$

Но тогда справедливо и равенство

$$\begin{pmatrix} D' & -B' \\ -C' & A' \end{pmatrix} \times \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} D'A - B'C & D'B - B'D \\ -C'A + A'C & -C'B + A'D \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}.$$

Отсюда,  $A'D - C'B = I$ , что и требовалось установить.

А1 (1987). Рассмотрим функцию  $f(z) = z^2 - 1/z - 3i$  комплексного переменного  $z = x + iy$ . Имеем

$$\operatorname{Re} f(z) = x^2 - y^2 - \frac{x}{x^2 + y^2}, \quad \operatorname{Im} f(z) = 2xy + \frac{y}{x^2 + y^2} - 3.$$

Пересечение  $A \cap B$  в естественном смысле совпадает с множеством корней уравнения  $f(z) = 0$ .

Аналогично, пересечение  $C \cap D$  совпадает с множеством корней уравнения  $g(z) = z^3 - 3iz - 1 = 0$ . Поскольку  $z = 0$  не является корнем ни одного из этих уравнений, они равносильны.

Разумеется, эту задачу можно решить и без привлечения комплексных чисел, но такое решение выглядит изящнее.

А5 (1987). Такого поля не существует. Действительно, если такое поле существует, то его циркуляция по любому замкнутому контуру, являющемуся краем ориентируемой поверхности, не содержащей начало координат, равна нулю. Рассмотрим половину эллипсоида  $\frac{x^2}{4} + y^2 + z^2 = 1$ ,  $z \geq 0$ . Его границей является плоский эллипс  $C$ , задаваемый уравнениями  $\frac{x^2}{4} + y^2 = 1$ ,  $z = 0$ . Вычисляя ин-

теграл  $\oint_C G_x dx + G_y dy$ , убеждаемся, что он отличен от нуля. При вычислении интеграла удобно использовать параметризацию  $x = 2 \cos \varphi$ ,  $y = \sin \varphi$ .

V4 (1987). Поскольку  $x^2(1) + y^2(1) = 1$  и  $x^2(n+1) + y^2(n+1) = x^2(n) + y^2(n)$  при  $n = 1, 2, 3, \dots$ , то  $x^2(n) + y^2(n) = 1$  при тех же  $n$ . Отсюда, при любом  $n$  существует угол  $\theta(n)$  такой, что  $\cos \theta(n) = x(n)$ ,  $\sin \theta(n) = y(n)$ . Из данных равенств следует, что можно принять  $\theta(n+1) = \theta(n) + y(n) = \theta(n) + \sin \theta(n)$ , при этом в качестве  $\theta(1)$  можно выбрать острый угол. Рассмотрим функцию  $y = x + \sin x$ . Эта функция возрастающая, причем  $y(x) < \pi$  при  $x < \pi$ . Отсюда следует, что последовательность  $\{\theta(n)\}$  сходится. Переходя к пределу в равенстве  $\theta(n+1) = \theta(n) + \sin \theta(n)$ , получим:  $\sin \lim_{n \rightarrow \infty} \theta(n) = 0$ , т.е.  $\lim_{n \rightarrow \infty} \theta(n) = \pi k$ . Из предыдущих рассуждений следует, что  $\lim_{n \rightarrow \infty} \theta(n) = \pi$ . Тем самым  $\lim_{n \rightarrow \infty} x(n) = -1$ ,  $\lim_{n \rightarrow \infty} y(n) = 0$ .

V6 (1987). При нечетном  $p$  число  $(p^2 - 1)/2$  четное. Обозначим через  $P_1$  произведение всех элементов множества  $S$ , и через  $P_2$  произведение элементов вида  $2a$  ( $a \in S$ ). Отсюда  $P_2 = 2^{|S|} P_1$ . С другой стороны, элементы вида  $2a$  обладают теми же свойствами, что и элементы  $S$ , т.е. для всякого ненулевого элемента  $b \in \mathbb{F}$  либо  $b$ , либо  $-b$  имеет вид  $2a$ , ( $a \in S$ ). Отсюда,  $P_2 = (-1)^{|S|-N} P_1 = (-1)^N P_1$  в силу четности  $|S|$ . Осталось доказать, что  $2^{|S|} = 1$ . Поскольку характеристика поля  $\mathbb{F}$  равна  $p$  и  $2$  принадлежит минимальному подполю  $\mathbb{F}$ , то по малой теореме Ферма  $2^{p-1} = 1$ . Тогда  $2^{|S|} = 2^{(p-1)(p+1)/2} = 1^{(p+1)/2} = 1$ , что и требовалось.

A6 (1988). Ответ положительный. Пусть  $x_1, \dots, x_{n+1}$  — собственные векторы оператора  $A$ ;  $\lambda_1, \dots, \lambda_{n+1}$  — соответствующие собственные значения. Достаточно доказать, что все собственные значения совпадают. Поскольку собственные значения — корни полиномиального уравнения степени  $n$ , среди них есть совпадающие — пусть  $\lambda_n = \lambda_{n+1}$ . Тогда все элементы двумерной линейной оболочки  $L_2$  векторов  $x_n, x_{n+1}$  — собственные векторы оператора  $A$  с собственным значением  $\lambda_n$ . Обозначим через  $L_{n-1}$   $(n-1)$ -мерную линейную оболочку векторов  $x_1, \dots, x_{n-1}$ . Из соотношения Грассмана следует, что  $\dim(L_{n-1} \cap L_2) = 1$ . Пусть  $y$  — ненулевой вектор из этого пересечения. Векторы  $x_1, \dots, x_{n-1}, y$  линейно зависимы. Пусть  $\mu_1 x_1 + \dots + \mu_{n-1} x_{n-1} + \mu_n y = 0$ , где среди коэффициентов есть отличные от нуля. Применив к обеим частям этого равенства оператор  $A$ , получаем:  $\lambda_1 \mu_1 x_1 + \dots + \lambda_{n-1} \mu_{n-1} x_{n-1} + \lambda_n \mu_n y = 0$ . Если среди собственных значений есть различные, то вычтя из последней линейной комбинации предыдущую, умноженную на  $\lambda_n$ , приходим к выводу, что векторы  $x_1, \dots, x_{n-1}$  — линейно зависимые, а это противоречит условию.

V4 (1988). Разобьем множество натуральных чисел на два класса:

$$N_1 = \left\{ n : (a_n)^{1/(n+1)} \leq 1/2 \right\}, \quad N_2 = \left\{ n : (a_n)^{1/(n+1)} > 1/2 \right\}.$$

Для первого множества  $(a_n)^{n/(n+1)} \leq (1/2)^n$ , для второго —  $(a_n)^{n/(n+1)} < 2a_n$ . Тем самым сходятся оба ряда  $\sum_{n \in N_1} (a_n)^{n/(n+1)}$  (всегда) и  $\sum_{n \in N_2} (a_n)^{n/(n+1)}$  (вследствие сходимости ряда  $\sum_{n=1}^{\infty} a_n$ ). Поскольку члены ряда  $\sum_{n=1}^{\infty} (a_n)^{n/(n+1)}$  положительные, отсюда следует его сходимость.

А3 (1989). Поскольку 0 не является корнем данного уравнения, его можно разделить на  $z^5$ . Получаем уравнение  $11(z^5 + z^{-5}) + 10i(z^4 + z^{-4}) = 0$ . Рассмотрим корни уравнения, расположенные на единичной окружности  $z = e^{i\varphi}$ . Подставляя и используя формулы Эйлера для синуса и косинуса, получим уравнение  $11 \cos 5\varphi + 10 \sin 4\varphi = 0$ . Рассмотрим функцию  $F(\varphi) = 11 \cos 5\varphi + 10 \sin 4\varphi$ . Имеем  $F(0) > 0$ ,  $F(\pi/5) < 0$ ,  $F(2\pi/5) > 0$ , ...,  $F(9\pi/5) < 0$ . Тем самым на единичной окружности расположено 10 корней данного уравнения. Поскольку оно имеет не более 10 различных корней, утверждение доказано.

А6 (1989). Поскольку число  $4n$  получено из  $n$  добавлением двух нулей, то  $a_{4n} = a_n$ . Аналогично  $a_{2n+1} = a_n$ . При этом, поскольку число  $4n + 2$  получено из  $n$  добавлением цифр 10, имеем  $a_{4n+2} = 0$ . Разобьем формальный ряд  $\alpha(x) = 1 + a_1x + a_2x^2 + \dots$  на две части:  $\alpha_1(x) = 1 + a_4x^4 + a_8x^8 + \dots = 1 + a_1x^4 + a_2x^8 + \dots = \alpha(x^4)$ ;  $\alpha_2(x) = a_1x + a_3x^3 + a_5x^5 + \dots = x + a_1x^3 + a_2x^5 + \dots = x\alpha(x^2)$ .

Далее, для двоичных переменных  $a, b$  выполнено  $a^k = a$ ,  $(a + b)^2 = a^2 + b^2$  по модулю 2. Поэтому  $\alpha(x^4) = (\alpha(x))^4$ ,  $\alpha(x^2) = (\alpha(x))^2$ . Отсюда,  $\alpha(x) = \alpha(x^4) + x\alpha(x^2) = (\alpha(x))^4 + x(\alpha(x))^2$ . Тем самым,  $\alpha(x)(\alpha^3(x) + x\alpha(x) - 1) = 0$ . Поскольку у формального ряда  $\alpha = 1 + a_1x + a_2x + \dots$  свободный член 1, то по индукции убеждаемся, что  $(\alpha^3(x) + x\alpha(x) - 1) = 0$ . Но  $(\alpha^3(x) + x\alpha(x) - 1) = (\alpha^3(x) + x\alpha(x) + 1)$  по модулю 2.

В6 (1989). Пусть  $M$  — описанная в условии задачи  $n$ -мерная область. Ее  $n$ -мерный объем равен

$$\int_M dX = \int_0^1 dx_n \int_0^{x_n} dx_{n-1} \dots \int_0^{x_2} dx_1 = \frac{1}{n!}.$$

Здесь  $dX$  — элемент объема множества  $M$ . Обозначим через  $U(f, x_1, \dots, x_n)$  риманову сумму  $\sum_{i=1}^n (x_i - x_{i-1})$ . Имеем

$$A(f) = \mathbf{E}[U(f, x_1, \dots, x_n)] = n! \int_M U(f, x_1, \dots, x_n) dX.$$

Рассмотрим интеграл от слагаемого  $(x_i - x_{i-1})f(x_i)$ . Перейдем к повторному интегралу таким образом, чтобы интеграл по  $x_i$  являлся внешним. Имеем

$$\int_0^1 f(x_i) dx_i \int_0^{x_i} (x_i - x_{i-1}) dx_{i-1} \int_0^{x_{i-1}} dx_1 \int_{x_1}^{x_{i-1}} dx_2 \dots \int_{x_{i-3}}^{x_{i-1}} dx_{i-2} \int_{x_i}^1 dx_{i+1} \dots \int_{x_{n-1}}^1 dx_n.$$

Этот интеграл равен  $\int_0^1 f(x) P_n^{(i)} dx$ , где  $P_n^{(i)} = \frac{(1-x)^{n-i}}{(n-i)!} \cdot \frac{x^i}{i!}$ . Таким образом,

$$P_n(x) = n! \sum_{i=1}^n P_n^{(i)}(x) = \sum_{i=1}^n \binom{n}{i} (1-x)^{n-i} x^i = 1 - (1-x)^n.$$

Здесь использовано биномиальное разложение функции  $((1-x) + x)^n$ . Очевидно, что полином  $P_n(x)$  обладает нужными свойствами.

А3 (1990). Докажем вначале, что площадь невырожденного треугольника с целочисленными координатами не меньше  $1/2$ . Опишем вокруг треугольника прямоугольник, стороны которого параллельны координатным осям. Тогда площадь треугольника равна разности площади прямоугольника с целочисленными сторонами и площадей нескольких прямоугольных треугольников с целочисленными катетами. Поскольку площадь прямоугольника целая и площади треугольников полуцелые, то площадь данного треугольника полуцелая и поэтому не меньше  $1/2$ .

Докажем, что в пятиугольнике найдется точка с целочисленными координатами, отличная от вершины. Среди пяти точек с целочисленными координатами найдутся две, у которых совпадают четности обеих координат. Но тогда середина соответствующего отрезка также имеет целочисленные координаты. Если такая точка находится внутри пятиугольника, то пятиугольник разбивается на пять треугольников, вершины которых целочисленные, то есть площадь пятиугольника не менее  $5/2$ . Если же точка находится на границе, то площадь пятиугольника не менее 2.

Проанализируем дополнительно последний случай. Пусть  $v_1, \dots, v_5$  — вершины данного пятиугольника в порядке обхода,  $v_0$  — середина отрезка  $[v_4, v_5]$  с целочисленными координатами. Исходный пятиугольник распадается на треугольник с вершинами  $v_3v_4v_0$  и пятиугольник  $v_1v_2v_3v_0v_5$ . По доказанному площадь треугольника не меньше  $1/2$ , площадь пятиугольника не меньше 2. Тем самым, площадь данного пятиугольника не менее  $5/2$ , что и требовалось.

А5 (1990). Пусть  $ABAB = 0$ . Тогда  $(BA)^3 = B(ABAB)A = 0$ . Докажем, что в таком случае при  $n = 2$  ( $n$  — порядок матрицы) выполняется равенство  $(BA)^2 = 0$ . Обозначим через  $S$  матрицу  $BA$ .  $S$  — матрица линейного оператора, который будем обозначать так же. Если образ оператора  $S$  двумерен, то  $S$  — линейный изоморфизм и поэтому  $S^3 \neq 0$ . Если образ оператора  $S$  нульмерен, то  $S^2 = S^3 = 0$ . Пусть теперь образ  $L$  оператора  $S$  одномерен. Если  $S^2 \neq 0$ , то  $SL = L$ , а тогда и  $S^3L = L$ , т.е.  $S^3 \neq 0$ . То же самое можно установить, используя жорданову форму матриц. Таким образом, матриц второго порядка с такими свойствами не существует.

Иначе обстоит дело при  $n > 2$ . Приведем пример таких матриц при  $n = 3$ :

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

В4 (1990). Построим ориентированный граф следующим образом. Его вершины являются элементами группы; ребро  $(p, q)$  существует тогда и только тогда, когда  $q = pa$  или  $q = pb$ . Этот граф обладает следующими свойствами:

- (1) из всякой вершины исходят два ребра, поскольку из равенства  $pa = pb$  следует равенство  $a = b$ ;
- (2) во всякую вершину входят два ребра. Обоснование то же;
- (3) граф является сильно связным. Действительно, из всякой вершины есть путь в единичный элемент и обратно, поскольку группа порождена элементами  $a, b$ .

Известно, что граф с такими свойствами является эйлеровым, т. е. существует цикл, в котором каждое ребро встречается один раз. Выписывая вершины в порядке прохождения этого цикла, получим нужную последовательность вершин.

A3 (1991). Справедливо представление  $p(x) = c(x - r_1)(x - r_2) \cdot \dots \cdot (x - r_n)$ . Отсюда

$$p'(x) = c[(x - r_2)(x - r_3) \cdot \dots \cdot (x - r_n) + (x - r_1)(x - r_3) \cdot \dots \cdot (x - r_n) + \dots + (x - r_1)(x - r_2) \cdot \dots \cdot (x - r_{n-1})].$$

Тогда

$$p'(\frac{r_1 + r_2}{2}) = c[(\frac{r_1 + r_2}{2} - r_2)(\frac{r_1 + r_2}{2} - r_3) \cdot \dots \cdot (\frac{r_1 + r_2}{2} - r_n) + (\frac{r_1 + r_2}{2} - r_1)(\frac{r_1 + r_2}{2} - r_3) \cdot \dots \cdot (\frac{r_1 + r_2}{2} - r_n) + \dots + (\frac{r_1 + r_2}{2} - r_1)(\frac{r_1 + r_2}{2} - r_2) \cdot \dots \cdot (\frac{r_1 + r_2}{2} - r_{n-1})].$$

Поскольку сумма первых двух слагаемых равна нулю, а при  $n > 2$  все остальные слагаемые отличны от нуля и имеют один и тот же знак, то равенство  $p'((r_1 + r_2)/2) = 0$  возможно только при  $n = 2$ . Очевидно при этом, что любой квадратный трехчлен с различными вещественными корнями обладает нужными свойствами.

A5 (1991). Воспользовавшись неравенством

$$\sqrt{x^4 + y^2(1 - y)^2} \leq x^2 + y(1 - y),$$

получаем оценку

$$\int_0^y \sqrt{x^4 + y^2(1 - y)^2} dx \leq \int_0^y (x^2 + y(1 - y)) dx = y^2 - 2y^3/3.$$

Максимальное значение функции  $y^2 - 2y^3/3$  на отрезке  $[0, 1]$  равно  $1/3$ . Тем самым, данный интеграл не превосходит  $1/3$ . Но при  $y = 1$  интеграл имеет вид  $\int_0^1 x^2 dx = 1/3$ . Отсюда следует ответ:  $1/3$ .

V3 (1991). Ответ положительный.

Докажем вначале такое утверждение. Пусть натуральные числа  $a_1, a_2, \dots, a_k$  взаимно простые. Любое натуральное число, большее, чем  $ka_1a_2 \cdot \dots \cdot a_k$ , можно представить в виде  $t_1a_1 + t_2a_2 + \dots + t_ka_k$ , где числа  $t_1, t_2, \dots, t_k$  неотрицательные.

Поскольку числа  $a_1, a_2, \dots, a_k$  взаимно простые, то 1, а значит и любое число, представимо в виде целочисленной линейной комбинации чисел  $a_1, a_2, \dots, a_k$ . Пусть  $m > ka_1a_2 \cdot \dots \cdot a_k$  и  $m = t_1a_1 + t_2a_2 + \dots + t_ka_k$ , причем среди чисел  $t_1, t_2, \dots, t_k$  есть отрицательные (пусть для определенности  $t_k < 0$ ). Среди положительных слагаемых есть превосходящее  $a_1a_2 \cdot \dots \cdot a_k$ , пусть для определенности  $t_1a_1 > a_1a_2 \cdot \dots \cdot a_k$ . Тогда  $t_1 > a_2 \cdot \dots \cdot a_k$ . Определим новые значения  $t'_1 = t_1 - a_k > 0$ ,  $t'_k = t_k + a_1 > t_k$ . Остальные значения  $t_i$  оставим без изменения.

Тогда

$$t'_1 a_1 + t_2 a_2 + \dots + t'_k a_k = t_1 a_1 + t_2 a_2 + \dots + t_k a_k = m.$$

Продолжая аналогично, получаем нужное утверждение.

Построим теперь из данных прямоугольников пары прямоугольных блоков, имеющих общую сторону.

1.  $20 \times 6$  — из прямоугольников  $4 \times 6$  и  $20 \times 7$  — из прямоугольников  $5 \times 7$ ;
2.  $35 \times 7$  — из прямоугольников  $5 \times 7$  и  $35 \times 5$  — из тех же прямоугольников (при изменении ориентации);
3.  $42 \times 4$  — из прямоугольников  $5 \times 7$  и  $42 \times 4$  — из прямоугольников  $4 \times 6$ .

Применяя доказанное утверждение, получаем: при достаточно больших  $n$  из данных прямоугольников можно сложить прямоугольники размеров  $20 \times n$ ,  $35 \times n$ ,  $42 \times n$ . Поскольку числа 20, 35, 42 взаимно простые, то из того же утверждения следует, что при достаточно большом  $m$  из данных прямоугольников можно сложить прямоугольник размеров  $m \times n$ , что и требовалось.

В4 (1991). Рассмотрим полином  $(1+(1+x))^p(1+x)^p = (2+x)^p(1+x)^p$ . Найдем коэффициент этого полинома при  $x^p$  с использованием каждого из представлений.

Левую часть можно представить в виде  $\sum_{j=0}^p \binom{p}{j} (1+x)^{p+j}$ . Отсюда, коэффициент при  $x^p$  равен

$$\sum_{j=0}^p \binom{p}{j} \binom{p+j}{p} = \sum_{j=0}^p \binom{p}{j} \binom{p+j}{j}.$$

С другой стороны, коэффициент при  $x^p$  равен

$$\sum_{j=0}^p 2^j \binom{p}{p-j} \binom{p}{j} = \sum_{j=0}^p 2^j \binom{p}{j}^2.$$

При  $0 < j < p$  биномиальный коэффициент

$$\binom{p}{j} = \frac{p(p-1) \cdot \dots \cdot (p-j+1)}{j!}$$

делится на  $p$ , поскольку в силу простоты числа  $p$  знаменатель взаимно прост с  $p$ . Тогда имеем  $\sum_{j=0}^p 2^j \binom{p}{j}^2 \equiv (2^p + 1) \pmod{p^2}$ , что и требовалось доказать.

А3 (1992). Применяя неравенство  $x^2 + y^2 \geq 2xy$ , получаем  $(xy)^n \geq (2xy)^m$ , откуда  $n > m$ .

Пусть  $p$  — простое число,  $a$  и  $b$  высшие степени  $p$ , на которые делятся  $x$  и  $y$ , причем  $a < b$ . Тогда левая часть данного равенства делится на  $p^{2am}$ , а правая на  $p^{(a+b)n}$  и не делится на более высокие степени  $p$ . Но тогда  $2am = (a+b)n > 2an$ , что противоречит неравенству  $n > m$ . Таким образом,  $a = b$  и потому  $x = y$ . Подставляя в исходное уравнение  $x = y$ , получаем  $2^m x^{2m} = x^{2n}$ , т.е.  $2^m = x^{2(n-m)}$ . Отсюда,  $x = 2^t$ . Имеем  $m = 2t(n-m)$ . Отсюда, число  $m$  четное. Пусть  $m = 2s$ . Имеем  $s = t(n-2s)$ . Если  $n-2s > 1$ , то простой делитель числа  $n-2s$  является общим делителем  $n$  и  $s$ , что противоречит взаимной простоте  $n$

и  $m$ . Отсюда,  $n = 2s + 1 = m + 1$ . При этом  $t = s$ ,  $x = 2^s = 2^{m/2}$ . Окончательно, равенство выполняется только при  $m$  четном,  $n = m + 1$ ,  $x = y = 2^{m/2}$ .

A5 (1992). Очевидны равенства  $a_{2n} = a_n$ ,  $a_{2n+1} + a_{2n} = a_{2n+1} + a_n = 1$ .

Пусть такие пары чисел  $k, m$  существуют. Выберем какую-либо из таких пар с минимальным значением  $m$ .

Пусть число  $m > 1$  нечетное. Можно считать, что  $a_k = a_{k+m} = a_{k+2m} = 0$ , противоположный случай рассматривается аналогично. Поскольку одно из чисел  $k, k + m$  четное, то  $a_{k+1} = a_{k+m+1} = a_{k+2m+1} = 1$ . Продолжая, получим  $a_{k+2} = a_{k+m+2} = a_{k+2m+2} = 0$ , и т. д. Поскольку число  $m - 1$  четное, то  $a_{k+m-1} = a_{k+2m-1} = a_{k+3m-1} = 0$ . Одно из чисел  $k + m - 1, k + 2m - 1$  четное, поэтому  $a_{k+m} = 1$  или  $a_{k+2m} = 1$ , что противоречит исходному предположению.

Пусть число  $m$  четное. Рассмотрим равенства  $a_{k+j} = a_{k+m+j} = a_{k+m+2j}$  только для четных индексов  $k + j$ . По предыдущему  $a_{(k+j)/2} = a_{(k+m+j)/2} = a_{(k+m+2j)/2}$ . Тем самым, справедливы равенства  $a_{[k/2]+i} = a_{[k/2]+m/2+i} = a_{[k/2]+m/2+2i}$  при  $0 \leq i \leq (m/2) - 1$ . (В предыдущих равенствах при  $k$  четном следует принять  $j = 2i$ , при  $k$  нечетном —  $j = 2i - 1$ ). Но это противоречит минимальности  $m$ .

Наконец, при  $m = 1$  должны выполняться равенства  $a_k = a_{k+1} = a_{k+2}$ . Но это невозможно, поскольку одно из чисел  $k, k + 1$  четное.

Утверждение доказано.

## Венгерская задача о квадратах

С. А. Дориченко

Автору этой заметки стала известна от венгерских математиков задача об оценке отношения периметра к площади для фигур, образованных объединением одинаковых квадратов. Этот сюжет был положен в основу следующей задачи, предложенной школьникам 10 – 11 классов на осеннем туре 22-го Турнира Городов, который состоялся 29 октября 2000 года:

а) *Несколько черных квадратов со стороной 1 см прибиты к белой плоскости одним гвоздем толщины 0,1 см. Образовалась многоугольная черная фигура. Может ли периметр этой фигуры быть больше, чем 1 км? (Гвоздь не задевает границ квадратов.)*

б) *Та же задача, но гвоздь имеет толщину 0 (то есть точка).*

в) *Несколько черных квадратов со стороной 1 лежат на белой плоскости, образуя многоугольную черную фигуру (возможно, состоящую из нескольких кусков и имеющую дырки). Может ли отношение периметра этой фигуры к ее площади быть больше 100 000?*

Ответ во всех трех случаях отрицательный. Конечно, пункты а) и б) следуют из пункта в), они были добавлены, чтобы облегчить решение последнего, весьма трудного, пункта.

Пункт а) этой задачи был решен некоторыми школьниками, решение пункта в) пока не встретилось (проверка еще не закончена, так как получены еще не все работы).

Заметим также, что для рассматриваемых фигур известны и гораздо лучшие (на несколько порядков) оценки отношения периметра к площади (см. подробности в статье автора исходной задачи: Tamas Keleti, "A Covering Property of Some Classes of Sets in  $\mathbb{R}^n$ ", Acta Univ. Carol. — Math. Phys. Vol. 39 (1998) No. 1–2, p. 111–118. Corollary 12 and Remark 14), но точная оценка сверху до сих пор неизвестна. Может быть, ее удастся получить кому-нибудь из наших читателей?

РЕШЕНИЕ ЗАДАЧИ О КВАДРАТАХ. Пункт а). Гвоздь пересекается с плоскостью, к которой прибиты квадраты, по кругу, обозначим его центр через  $O$ , а радиус — через  $r$ . Получившаяся черная фигура является многоугольником, обозначим его  $A_1A_2 \dots A_n$ . Соединим точку  $O$  с вершинами этого многоугольника. Тогда наша фигура разобьется на непересекающиеся треугольники  $OA_1A_2, OA_2A_3, \dots, OA_nA_1$  (это следует из того, что точка  $O$  принадлежит всем квадратам). Обозначим соответственно через  $\alpha_1, \dots, \alpha_n$  углы этих треугольников при вершине  $O$ .

Рассмотрим один из треугольников, скажем  $OA_1A_2$ , и оценим длину стороны  $A_1A_2$ .

Отрезок  $A_1A_2$  лежит на стороне одного из исходных квадратов. Так как по условию гвоздь лежит целиком внутри этого квадрата, расстояние от точки  $O$



до прямой, содержащей  $A_1A_2$ , больше  $r$ . Поэтому площадь треугольника  $OA_1A_2$  больше  $\frac{1}{2} \cdot r \cdot A_1A_2$ . С другой стороны, эта площадь равна  $\frac{1}{2} \cdot OA_1 \cdot OA_2 \cdot \sin \alpha_1$ , и поскольку  $OA_1$  и  $OA_2$  меньше диагонали квадрата (то есть  $\sqrt{2}$ ), площадь треугольника  $OA_1A_2$  меньше  $\sin \alpha_1$ . Получаем неравенство:

$$\frac{1}{2} \cdot r \cdot A_1A_2 < \sin \alpha_1, \quad \text{откуда } A_1A_2 < \frac{2}{r} \cdot \sin \alpha_1.$$

Записав для каждой стороны фигуры аналогичное неравенство и просуммировав их, получим:

$$A_1A_2 + \dots + A_nA_1 < \frac{2}{r} \cdot (\sin \alpha_1 + \dots + \sin \alpha_n) < \frac{2}{r} \cdot (\alpha_1 + \dots + \alpha_n) = \frac{4\pi}{r}.$$

Так как  $r = 1/20$  см, мы получаем, что периметр фигуры не превосходит  $80\pi$  см, что меньше  $80 \cdot 3,15 = 252$  см и подавно меньше 1 км.

Пункт в). Оказывается, общую задачу можно свести к предыдущей. Для этого на плоскость квадратную сетку с шагом  $1/2$ , в узлах которой вобьем гвоздики радиуса  $1/20$  (то есть для каждого узла сетки нарисуем круг радиуса  $1/20$  с центром в этом узле).

*ЛЕММА. Для каждого из исходных квадратов найдется гвоздик сетки, целиком содержащийся в этом квадрате.*

**ДОКАЗАТЕЛЬСТВО.** Впишем в каждый из исходных квадратов окружность радиуса  $1/2$ . Достаточно доказать, что для каждой такой окружности найдется гвоздик сетки, целиком в ней содержащийся. Пусть  $\omega$  — одна из этих окружностей,  $M$  — ее центр. Точка  $M$  лежит внутри одного из квадратиков сетки, расстояние от  $M$  до ближайшей из вершин этого квадратика не больше половины диагонали квадратика, то есть не больше  $1/2\sqrt{2}$ . Тогда гвоздик с центром в этой вершине будет целиком содержаться в окружности  $\omega$ , так как ее радиус равен  $1/20$ , что больше  $1/2\sqrt{2} + 1/20$ . Лемма доказана.

Разобьем теперь исходные квадраты на группы по правилу: в одну группу входят квадраты, целиком содержащие конкретный гвоздик сетки (возможно, некоторые квадраты попадут сразу в несколько групп).

Применив для каждой группы неравенство из пункта а), получим, что периметр фигуры, образованной квадратами одной группы, не больше  $80\pi$ .

Пусть всего групп  $N$ . Тогда периметр всей нашей фигуры не больше  $80\pi N$ . С другой стороны, так как площадь каждого гвоздика равна  $\pi/400$ , и гвоздики не перекрываются, общая площадь нашей фигуры не меньше  $\pi N/400$ . Поэтому отношение периметра нашей фигуры к ее площади не превосходит  $32\,000$ , что меньше  $100\,000$ .

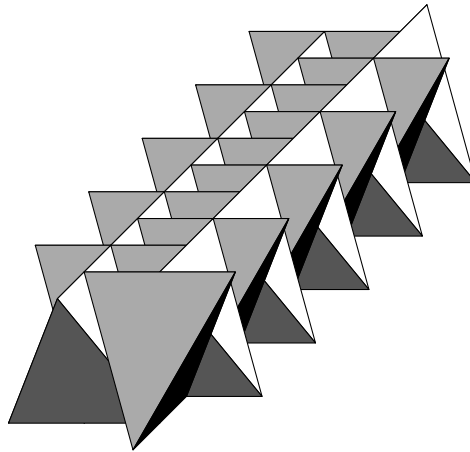
## История одной олимпиадной задачи

А. Я. Канель

На 63-й Московской Городской Математической олимпиаде (2000 г., 11 класс, задача №6) была предложена следующая задача:

*Можно ли так расположить в слое между двумя плоскостями семейство одинаковых выпуклых тел так, чтобы ни одно из них нельзя было сдвинуть, не задевая остальные?*

На плоскости, в слое между двумя прямыми, такого расположения выпуклых фигур нет (это довольно содержательная задача). А в пространстве ответ – да! (см. рис.).



Докажем, что ни один из тетраэдров нельзя выдвинуть. Заменяем один из них вписанным в него шаром  $S$  и покажем, что этот шар нельзя выдвинуть. Вращать шар бесполезно, поэтому ограничимся рассмотрением параллельных переносов.

Рассмотрим плоскость грани  $\Gamma$  другого тетраэдра, которая касается  $S$ . Множество векторов сдвига, разрешенных  $\Gamma$ , задается множеством векторов, идущих из  $\Gamma$  в сторону  $S$ . Такие вектора образуют полупространство направлений, а множество возможных сдвигов есть пересечение множеств направлений (полупространств), разрешенных гранями тетраэдров укладки, касающихся  $S$ .

Легко видеть, что оно пусто. Поэтому шар, а значит, и тетраэдр, в который он вписан, не вынимается.

Я встретился с этой задачей так. В свое время меня не взяли в аспирантуру МГПИ им. В. И. Ленина (помимо обычных проблем была и такая — тогдашний проректор института заявил, что специалисты по истории КПСС против) и я

поступил в аспирантуру московского Горного Института. Там я занялся изучением равновесия блочных массивов скальных пород, прочел книгу R. Googman and Shi-gen-Hua “Introduction to Rock Mechanics”. В этой книге была изложена концепция *ключевого* блока. Ключевые блоки — это те блоки, которые не заклинены остальными и могут двигаться внутри выработки. Для обеспечения безопасности выработки достаточно осуществить крепление только таких блоков. Там же содержатся рассуждения с пересечением пространств направлений, задаваемых гранями. Вопросы, связанные с возможным отсутствием ключевых блоков там даже не поднимались.

После защиты диссертации я к горному делу не возвращался. Однако, узнав, что из выпуклых тел можно составить выпуклое тело так, чтобы ни одно из них нельзя было выдвинуть, придумал ту самую задачу для Московской олимпиады.

История имела занятное продолжение. Когда я был в Израиле в мае 2000 года и имел трудное финансовое положение, то вспомнил, что за доклады платят (50\$) и нашел бывшего научного руководителя по аспирантуре. На какую тему сделать доклад?

Пришла идея. В малом зерне не развивается трещина. С другой стороны, зерна могут клинить друг друга, если их сделать, скажем, в форме правильных тетраэдров и расположить так, как это сделано в решении задачи Московской Олимпиады. Поэтому возникает возможность создавать композиты, в которых плохо распространяются трещины. Они могут выдерживать высокое давление и обладают другими интересными свойствами.

Данная идея заинтересовала специалистов. Так был изобретен новый способ укладки кирпичей (как мне кажется, красивый).

Были предложены другие расположения зерен. Оказалось, что вопросы, связанные с пространственной организацией структурных элементов при расположении композитов (в плане создания пространственной структуры материала) в должной мере не исследованы, не говоря уже об эффекте самозаклинивания зерен. В данный момент ведется работа вместе с материаловедами и специалистами по механике; уже принята к печати статья (A. V. Dyskin, Y. Estrin, A. J. Kanel-Belov, E. Pasternak *A new concept in design of materials: assembles of interlocked tetrahedron-shaped elements*, to appear in Scripta Materialia).

Таким образом, я выступил в роли своего рода «культурного агента» (ведь ничего по сути изобретено мною не было). Эффективным оказался другой, чуть более широкий взгляд на ситуацию, чем просто решение олимпиадной задачи.

# Нам пишут. . .

---

---

## Программа курса лекций на двухгодичном потоке ФМШ им. М. А. Лаврентьева, СУНЦ НГУ

В. В. Войтишек

Вацлав Вацлавович Войтишек, доцент кафедры математики СУНЦ НМУ, предлагает сравнить приводимую ниже программу с программой «Матшкольник», опубликованной в №2 третьей серии «Математического просвещения».

### ДЕСЯТЫЙ КЛАСС. ПЕРВОЕ ПОЛУГОДИЕ

#### МАТЕМАТИЧЕСКИЙ АППАРАТ НАУКИ

1. Комплексные числа. Формы записи. Основные действия, их геометрический смысл (включая инверсию и стереографическую проекцию). Формула Муавра. Корни из единицы.
2. Интеграл. Формулы бруса, трапеции, параболы. Метод удвоения. Натуральный логарифм: площадь подграфика  $0 \leq xy \leq 1$  при  $x \geq 1$ . Вычисление  $\pi$ .
3. Первая, вторая,  $n$ -я производная. Формула Тэйлора для квадратного трехчлена и основных функций. «Правило дождя». Задача о ракете. Формула Ньютона – Лейбница. Криволинейный интеграл.
4. Этапы анализа функции  $y = f(x)$ . Список основных функций и соответствий в  $\mathbb{R}^3$  (85 названий). Метод линий уровня.
5. Основные отображения  $w = f(z)$ . Понятие о римановой поверхности (15 примеров).
6. Матрицы, определители, характеристические уравнения (второй, третий порядок). Преобразование координат. Углы Эйлера. Обратная матрица.
7. Элементы анализа в  $\mathbb{R}^3$ . Параметрическое задание кривых и поверхностей. Градиент, касательная плоскость, якобиан.
8. Элементы топологии. Эйлерова характеристика. Существование правильных многогранников. Классификация поверхностей.

9. Гиперболические поворот и функции. Преобразование Лоренца (по книге Л. И. Головиной «Линейная алгебра и некоторые ее приложения», изд. М.: Наука, 1985).
10. Решение систем уравнений. Симметрические полиномы. Орбиты. Формула Варинга.

### ДЕСЯТЫЙ КЛАСС. ВТОРОЕ ПОЛУГОДИЕ

#### ГОД СПОКОЙНОЙ МАТЕМАТИКИ

11. Способы приближения функций другими функциями. Полиномы и ряды Тэйлора и Фурье. Многочлены Чебышева, наименее отклоняющиеся от нуля.
12. Интерполяция. Формулы Лагранжа и Ньютона. Сплаины.
13. Коэффициенты Фурье. Сложение гармоник. Среднее квадратичное отклонение.
14. Целые числа и многочлены. НОД, НОК, схема Горнера, алгоритм Эвклида, цепные дроби. Золотое сечение, соизмеримость отрезков.
15. Многочлены над  $\mathbb{Z}$ . Нули целые, рациональные. Границы нулей. Метод касательных для их нахождения. Сжимающие отображения. Неподвижные точки.
16. Существование комплексных корней многочленов над  $\mathbb{Z}$  ( $\mathbb{R}$ ,  $\mathbb{C}$ ). Метод Лобачевского для нахождения всех нулей многочлена.
17. Конические сечения. Шары Данделена. Различные определения и свойства конических сечений. Квадратура. Выпуклые части. Биполярные и эллиптические координаты.
18. Углубленное изучение этапов анализа функции. Четность. Периодичность. Монотонность. Обратимость. Выпуклость.
19. Поведение функции на краю  $\text{Dom}(y)$  — характер разрывов, асимптоты.
20. Построение графиков уравнений и неравенств. Фазовые портреты функции  $x(t)$ .

### ВЫПУСКНОЙ КЛАСС. ПЕРВОЕ ПОЛУГОДИЕ

21. Суммирование. Константа Эйлера. Классические последовательности и суммы.
22. Азбука пределов ( $x_n$ ) и  $\lim f(x)$ . Теоремы Штольца. Условие Липшица. Сжатие.
23. Основные правила комбинаторики. Конечные группы.
24. Формула Муавра – Стирлинга для  $n!$
25. Вероятность. Выборки. Схема и теорема Я. Бернулли. Аксиомы А. Н. Колмогорова. Математическое ожидание, дисперсия, условная вероятность. Неравенство П. Л. Чебышова. Что такое математическая статистика?
26. Элементы теории информации. Мера энтропии. Неравенство Йенсена.
27. Барцентрические координаты. Средние величины. Классические неравенства.

28. Число в планиметрии (умножение вектора на число, меры углов, длин, площадей).
29. Число в стереометрии. Объем, площадь поверхности. Расстояния между точками, прямыми, плоскостями.
30. Элементы теории множеств. Счетность. Континуум. Канторово множество. Начальные сведения о мере Лебега. Фигуры Серпинского. Понятие области.
31. Профессия математика (перспективы). Математические центры России, зарубежья.
32. Институт математики СО РАН.

### ВЫПУСКНОЙ КЛАСС. ВТОРОЕ ПОЛУГОДИЕ

#### АЛЬТЕРНАТИВНЫЙ КУРС В. В. ВОЙТИШЕКА

33. Развитие идей математики (с Ньютоном 357 лет, без Гильберта — 57). (январь.)
34. Элементы проективной геометрии. Модель проективной плоскости. Теоремы Паппа, Дезарга, Паскаля, Бриансона. Конечные проективные пространства (по А. И. Ширшову). (февраль.)
35. Подход к модели плоскости Н. И. Лобачевского (Э. Бельтрами, Ф. Клейн, А. Пуанкаре). (март) Топологии 105 лет.
36. Эквидистанты, орициклы, гиперциклы, расстояния, угол параллелизма Лобачевского (по книге Б. В. Шабата «Введение в комплексный анализ», часть I, изд. 3. М.: Наука, 1985).
37. Теорема Пифагора в плоскости Лобачевского. Пентаграмма Непера. (март-апрель.)
38. Как вычисляли интеграл: задачи Архимеда, яблоко Кеплера, пика Торичелли, подход Лейбница, Коши, Римана, Лебега. (апрель-май)

#### ДРУГИЕ АЛЬТЕРНАТИВНЫЕ КУРСЫ

39. Основные сведения о дифференциальных уравнениях. Графическое интегрирование, изоклины, метод ломаных. Фазовые плоскости.
40. Метод разделения переменных. Однородные уравнения. Знакомство с групповым анализом дифференциальных уравнений по Л. В. Овсянникову.

### ЛИТЕРАТУРА ПО КУРСУ В. В. ВОЙТИШЕКА

- [ВИСКУ] Смирнов В. И. Курс высшей математики. Том 1, том 2.
- [Кл] Клейн Ф. Элементарная математика с точки зрения высшей. Том 1, том 2. М.: Наука, 1987.
- [ФаНиСо] Фаддев Д. К., Никулин М. С., Соколовский И. Ф. Элементы высшей математики для школьников. М.: Наука, 1987.
- [КиР] Курант Р., Робинс Г. Что такое математика? М.: Просв., 1967.
- [ПоГеВу] Погорелов А. В. Геометрия для вузов. М.: Наука, 1984.
- [Болт] Болтянский В. Г. Элементарная геометрия. М.: Просв., 1985.

- [Ш] Пашкин Ю. А. Неподвижные точки. М.: Наука, 1989. (сер. «Популярные лекции по математике», вып. 60)
- [бибК] Библиотечка «Квант» — математические выпуски 17, 21, 22, 23, 56, 61, 64, 77, 83.
- [К] Журнал «Квант» (РАН, РАО) — годовая подписка.
- [Э] Энциклопедии — БСЭ, ДЭ, МЭ, МЭС, ЭЭМ, ЭСКМ.
- [ЛаШа] Лаврентьев М. А., Шабат Б. В. Методы ТФКП. Изд. 5. М.: Наука, 1987.
- [БеГ] Берже М. Геометрия. Том 1, 2. М.: Мир, 1984.
- [АВРы] Александров А. Д., Вернер А. П., Рыжик В. И. Геометрия. М.: Просв., 1988.
- [ВГ] Вейль Г. Математическое мышление. М.: Наука, 1989.
- [БиС] Бронштейн И. Н., Семендяев К. А. Справочник по математике. М.: Наука, 1985.

ПРОГРАММА ПРАКТИЧЕСКИХ ЗАНЯТИЙ (ПО КУРСУ  
В. В. ВОЙТИШЕКА)

1. Повторение — сентябрь, октябрь.

Алгебра — решение типичных систем неравенств ( $\geq 1$ ) между элементарными функциями [БАНА, гл. 6]; [ДПР]; [ЯкПе-3, гл. 1, 3, 7, 8]. Геометрический смысл систем и решений в  $\mathbb{R}$ ,  $\mathbb{R}^2$ ,  $\mathbb{R}^3$ .

Планиметрия — теорема Пифагора (12 доказательств) [КиС], [АГ], [Пр, часть 1]; окружность и задача Аполлония [Пр, часть 1, гл. 9]; задачи на построение [ПоГеВу, гл. 22]; площадь круга.

2. Текущие занятия — октябрь – май.

Анализ — нахождение производных порядка  $n \geq 2$  и построение графиков основных функций [Ку, том 1], [Вин, гл. 1], [БАНА]. Нахождение первообразных для функций из физики, экономики. [Ку, том 2, гл. 1, 2]. Использование разложений по формуле Тэйлора [Ку, том 1, §19]. Построение графиков соответствий (и неявных функций) в декартовой, полярной, цилиндрической системах координат, включая и параметрически заданные кривые ([Ку, том 1, 7.1–7.283; 24.1–24.70], [Вин, гл. 1]) и поверхности [ПоГеВу].

Решение задач Архимеда, используя принцип Кавальери [Ку, том 2, 8.63–8.70].

Геометрия — стереометрия — занимательные задачи по черчению, построение сечений [ЯкПе-3, гл. 16, 17]; параллельность и перпендикулярность прямых и плоскостей [КиС], [АГ], [ЯкПе-3]; теорема о трех перпендикулярах, коэффициент искажения площади при ортогональном проектировании [ПоГеВу].

Геометрия — планиметрия — инверсия и стереографическая проекция [К, №7, 86], [ПоГеВу], [ВИСКУ, том 2], [Кл, том 1], [КиР], [ЛаШа], [Пр, часть 2, гл. 28].

Алгебра — формулы сложения круговых функций [БАНА], [ДПР], [ЯкПе-3]. Обратные круговые функции [ДПР], [Ку, том 1]. Действия с комплексными числами [Ку, том 1], [ФиС]. Корни из единицы.

Действия с матрицами, определителями, характеристическими уравнениями [ФиС].

### ИТОГ

На первом экзамене учащийся решает, например, такие задачи:

1. В каком отношении делит плоскость (прямая) объем (площадь) правильного многогранника (многоугольника), если она проходит через вершину, середину ребра и центр грани?
2. Пусть  $M$  — множество всех действительных чисел на  $z$ -плоскости и тех  $z \in \mathbb{C}$ , что расположены на расстоянии 2 от числа  $1 + 2i$ . Перечислить дробно-линейные преобразования, переводящие  $M$  в себя.
3. Вычислить ([Ку, том 2, 8.70.4]) объем тела, если  $by = x(a - z)$ ,  $bx = x(z - a)$ ,  $x = b$ ,  $z = 0$  — граница тела.

### НОВЫЕ ТЕМЫ

(ГОД СПОКОЙНОЙ МАТЕМАТИКИ), (янв.)

1. Алгоритмы, связанные с многочленами [ФиС].
2. Ряды Фурье. Простейшие разложения [Ку, том 2].
3. Конические сечения (пособие В. В. Войтишека «Введение в математику», НГУ, 1978).
4. Этапы анализа функции (нахождение нулей, четность, периодичность, монотонность, обратимость) [Ку, том 1].

### В ВЫПУСКНОМ КЛАССЕ

5. Этапы анализа (окончание): выпуклость, асимптоты, фазовые портреты [Ку, том 1].
6. Комбинаторика, подсчет вероятностей [Ку, том 1].
7. Нахождение сумм [Ку, том 2], вычисление пределов последовательностей, функций [Ку, том 1].
8. Использование алгебраических формул (включая определители) для нахождения величин, связанных с геометрическими фигурами на плоскости и в пространстве [БиС].
9. Градиент, касательная плоскость, якобиан.
10. Решение задач вступительного экзамена за последние три года в НГУ, МГУ, МФТИ, МИФИ, МГТУ, ПГУ, по журналу «Квант» и задачнику В. С. Белоносова и М. В. Фокина, изд. 5, СО РАН, НГУ, 2000.

ПРИМЕЧАНИЕ. Остальные темы школьной математики (см. [БАНА], [АГ]) учащиеся и преподаватели выбирают, затем изучают самостоятельно.

### СПИСОК ЛИТЕРАТУРЫ

- [КиС] Киселев А. П. Элементарная геометрия. Книга для учителей. М.: Просв., 1980.



# Задачный раздел

---

---

В этом разделе вниманию читателей предлагается подборка задач разной степени сложности, в основном трудных. Некоторые из этих задач (не обязательно самые сложные!) требуют знания «неэлементарной» математики — анализа, линейной алгебры и т. п.

Составителям этой подборки кажется, что предлагаемые ниже задачи окажутся интересными как для сильных школьников, интересующихся математикой, так и для студентов-математиков.

Мы обращаемся с просьбой ко всем читателям, имеющим свои собственные подборки таких задач, присылать их в редакцию. И, разумеется, мы с удовольствием будем публиковать свежие авторские задачи. Ждем ваших писем (как с вновь предлагаемыми задачами, так и с решениями опубликованных задач).

В скобках после условия задачи приводится фамилия автора (уточнения со стороны читателей приветствуются). Если автор задачи нам неизвестен, то в скобках указывается «фольклор».

1. В телесериале «Тайна Санты-Барбары» 15 действующих лиц. Серия называется *содержательной*, если в ней происходит одно из следующих событий. Либо кто-то узнает тайну, либо кто-то узнает, что кто-то знает тайну, либо кто-то узнает, что кто-то не знает тайну. Каково максимально возможное число содержательных серий? (тайна одна и первоначально ее не знает никто).

(С. В. Колягин)

2. Али-Баба делит с разбойником 10 куч золотого песка. Али-Баба может в любой момент взять три кучи и уйти, либо он может выбрать 4 любые кучи и разделить каждую из куч на правую и левую часть. Разбойник образует из этих частей 4 новые кучи, объединяя каждую правую часть с одной из левых. Сможет ли Али-Баба унести с собой свыше 49 кг золотого песка, если всего было 50 кг?

(А. Я. Белов)

3.  $A_1, \dots, A_n$  — ненулевые матрицы. Докажите, что найдется матрица  $B$  такая, что

$$BA_1BA_2 \dots BA_nB \neq 0. \quad (А. Я. Белов)$$

4. Отрезок  $SP$  соединяет точку  $P$  на границе и фокус  $S$  эллипса. Точка  $Q$  лежит на касательной к эллипсу в точке  $P$  (достаточно близко к  $P$  в направлении приближения к фокусу  $S$ ). Параллельная  $SP$  прямая, проходящая через  $Q$ , пересекает эллипс первый раз в точке  $R$ . Точка  $T$  — основание

перпендикуляра, опущенного из  $R$  на отрезок  $SP$ . Найдите

$$\lim_{Q \rightarrow P} \frac{QT^2}{RQ}.$$

Длины полуосей  $a, b$  эллипса считать известными.

(А. Сендеричин)

5. Докажите, что пересечение 10 правильных тетраэдров, вписанных в додекаэдр (вершины тетраэдров являются вершинами додекаэдра), есть икосаэдр.  
(А. Я. Белов)
6. а) Все коэффициенты многочлена  $P(x)$  целые и среди них есть нечетные. Докажите, что найдется многочлен  $Q(x)$ , имеющий ровно два нечетных коэффициента и делящийся на  $P(x)$ .  
б) Существует ли многочлен  $P(x, y)$  с целыми коэффициентами такой, что всякий многочлен  $R(x, y)$ , делящийся на  $P(x, y)$ , имеет более миллиона нечетных коэффициентов?  
(А. Я. Белов)
7.  $G$  — группа без кручения, т. е. нет неединичных элементов конечного порядка. Известно, что для некоторого  $n > 1$  выполняется тождество  $(ab)^n = a^n b^n$ . Докажите, что группа абелева, т. е. для любых  $a, b$  выполнено  $ab = ba$ .  
(Г. А. Карасев)
8. Разбейте плоскость на непересекающиеся отрезки равной длины.  
(Фольклор)
9. а) В клетках бесконечной клетчатой ленты записаны положительные числа. Известно, что каждое число не меньше среднего арифметического трех соседей слева и трех справа. Докажите, что числа равны.  
б) На клетчатой плоскости в клетках расставлены положительные числа так, что каждое записанное число равно среднему арифметическому 4 своих соседей. Докажите, что все числа равны.  
в) Верно ли аналогичное утверждение для пространственной решетки?  
(И. Ф. Шарыгин)
10. Случайные величины  $X, Y, Z$  равномерно распределены на единичном отрезке. Докажите, что величина  $(XY)^Z$  также равномерно распределена.  
(М. Кельберт)

## Решения задач из предыдущих выпусков

1.2. УСЛОВИЕ. Доказать, что существует бесконечно много натуральных чисел  $N$  таких, что  $2^N$  оканчивается на  $N$ .

РЕШЕНИЕ. Прежде всего, нетрудно проверить, что

$$2^{36} \text{ оканчивается на } 36. \quad (\text{i})$$

Пусть

$$2^{36} = 10^4 N + 10^3 K + 36. \quad (\text{ii})$$

Покажем, что  $2^{10^3 K + 36}$  оканчивается на  $10^3 K + 36$ . По ходу дела нам придется воспользоваться тем, что

$$2^{4 \cdot 5^3} - 1 \text{ делится на } 5^4, \quad (\text{iii})$$

это следует из биннома Ньютона:  $(15 + 1)^{5^3} - 1 = 5^3 \cdot 15 \cdot N = 5^4 \cdot N_1$ . Имеем:

$$\begin{aligned} 2^{10^3 K + 36} - (10^3 K + 36) &\stackrel{(\text{ii})}{=} 2^{10^3 K} (10^4 N + 10^3 K + 36) - (10^3 K + 36) = \\ &= 2^{10^3 K} \cdot 10^4 N + (10^3 K + 36)(2^{10^3 K} - 1) \stackrel{\text{Id}}{=} 10^4 N_1 + N_2 \left( \left( (2^4)^{5^3} \right)^{2K} - 1 \right). \end{aligned}$$

Второе слагаемое делится на  $5^4$  в силу (iii). С другой стороны, в силу (i)  $10^3 K + 36$  делится на  $2^4$ , значит  $2^{10^3 K + 36} - (10^3 K + 36)$  делится на  $10^4$ , что и требовалось.

Мы совершили первый шаг индукции. Общее построение таково. Положим  $a_{m+1} = 2^{a_m}$ ,  $b_m$  — число, образованное последними  $m$  цифрами числа  $a_m$ ,  $m \geq 2$ . Тогда  $2^{b_m} - b_m$  делится на  $10^m$ . Предоставим читателю провести  $m$ -й шаг индукции самостоятельно. (В. М. Тихомиров)

1.4. УСЛОВИЕ. Дан треугольник  $ABC$ .  $K$  — точка касания вписанной в него окружности и стороны  $BC$ . Рассмотрим две окружности, касающиеся прямой  $BC$ , луча  $AK$  и окружности, описанной вокруг  $\triangle ABC$  (на рис. 1 изображены пунктиром). Доказать, что их радиусы равны.

РЕШЕНИЕ. Пусть  $E$  — центр вневписанной окружности, касающейся стороны  $BC$  и продолжений сторон  $AB$  и  $AC$ , радиус этой окружности обозначим  $r_a$ , а  $\gamma$  — окружность, касающаяся продолжений отрезков  $AK$ ,  $BK$  и также касающаяся описанной вокруг треугольника  $ABC$  окружности с внешней стороны.

ЛЕММА 1. Радиус  $\gamma$  равен  $r_a$ .

Утверждение задачи сразу следует из леммы 1. Оказывается, обе пунктирные окружности на рис. 1 не только равны между собой, но и равны вневписанной окружности.

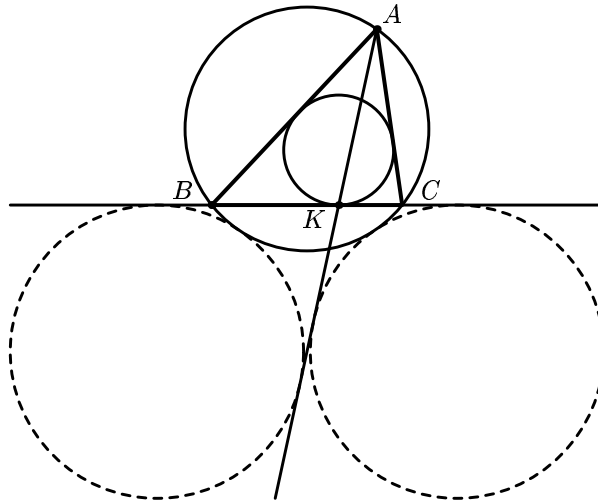


Рис. 1.

Лемма 1 выводится из следующего утверждения. Пусть  $K'$  — произвольная точка на стороне  $BC$ ,  $\gamma'$  — окружность, касающаяся продолжений отрезков  $AK'$  и  $BK'$  (точки касания обозначим  $M$  и  $L$  соответственно) и также касающаяся описанной вокруг треугольника  $ABC$  окружности с внешней стороны.

ЛЕММА 2. Для любой точки  $K' \in BC$  прямая  $ML$  проходит через точку  $E$ .

Доказательство этого факта (и аналогичного факта для вписанной окружности) можно найти в книге И. Ф. Шарыгина «Задачи по геометрии (планиметрия)», библиотечка «Квант», №7, 1986 и в книге решений четвертой Соросовской Олимпиады (10 класс).

Покажем, как из леммы 2 выводится лемма 1, а значит, и утверждение задачи.

ДОКАЗАТЕЛЬСТВО ЛЕММЫ 1. Опустим перпендикуляр  $EV$  на сторону  $BC$  и продолжим его за точку  $E$  на его длину. Получим отрезок  $VW$ , перпендикулярный  $BC$  и такой, что  $VW = 2VE = 2r_a$ . Легко показать (из гомотетии окружностей), что отрезок  $BW$  проходит через  $K$ . Применяя лемму 2 для  $K' = K$ , заключаем, что точки  $E, M, L$  лежат на одной прямой. Возьмем точку  $H$  на стороне  $BC$  так, чтобы  $WH \parallel ML$ . Обозначим  $O$  середину  $WH$ . Треугольник  $MKL$  — равнобедренный, значит, и треугольник  $WKN$  также равнобедренный. Поэтому  $KO$  — биссектриса угла  $WKN$ . Более того, отрезок  $EL$  является средней линией в треугольнике  $WVH$ . Поэтому  $WELO$  — параллелограмм. Это означает, что  $OL = r_a$ ,  $OL \perp BC$ . Таким образом,  $O$  — центр окружности  $\gamma$  (как единственная точка пересечения перпендикуляра  $LO$  с биссектрисой угла  $WKN$ ). Следовательно, радиус  $\gamma$  равен  $r_a$ , что и требовалось доказать.

(В. Протасов)

АНАЛИТИЧЕСКОЕ РЕШЕНИЕ. Докажем, что радиусы окружностей, указанных в условии задачи, равны радиусу  $r_a$  вневписанной окружности  $C_a$  треугольника

$ABC$ , касающейся стороны  $BC$ . Пусть  $l$  — вторая касательная к  $C_a$ , параллельная  $BC$ . Точку касания  $l$  и  $C_a$  обозначим  $K'$ . Легко видеть, что эта точка лежит на прямой  $AK$ .

Введем прямоугольную систему координат, ось абсцисс которой совпадает с  $l$ , а ось ординат проходит через  $A$ . Координаты  $A$  равны  $(0, L)$ , абсциссы точек пересечения прямых  $AK$ ,  $AB$ ,  $AC$  с осью абсцисс обозначим  $\varkappa, \xi_1, \xi_2$  соответственно. Найдём соотношение между ними. Алгебраически условие касания окружности  $C_a$  и прямой  $AB$  ( $AC$ ) означает, что у уравнения

$$(x - \varkappa)^2 + (y - r_a)^2 = r_a^2,$$

где  $y = L - \frac{L}{\xi_i}x$ , — кратный корень, т.е. дискриминант трехчлена равен 0. Подставляя выражение для  $y$ , получаем условие

$$\left(2\varkappa + \frac{2L(L - r_a)}{\xi_i}\right)^2 - 4\left(1 + \frac{L^2}{\xi_i^2}\right)(\varkappa^2 + L(L - 2r_a)) = 0,$$

из которого после упрощения следует, что  $\xi_1, \xi_2$  — корни квадратного уравнения (относительно  $\xi$ )

$$(L - 2r_a)\xi^2 - 2(L - r_a)\varkappa\xi + L\varkappa^2 - Lr_a^2 = 0. \quad (1)$$

Абсциссы  $\eta_1, \eta_2$  точек касания (и центров) окружностей, касающихся  $BC$ ,  $l$ ,  $AK$ , являются корнями аналогичного (1) квадратного уравнения (но относительно  $\eta$ )

$$(L - 2r_a)\varkappa^2 - 2(L - r_a)\varkappa\eta + L\eta^2 - Lr_a^2 = 0. \quad (2)$$

Теперь нужно проверить, что эти окружности касаются описанной вокруг  $\triangle ABC$  окружности, т.е. совпадают с пунктирными окружностями на рисунке к условию задачи.

Координаты вершин  $B$  и  $C$  равны  $(\frac{\xi_1}{L}(L - 2r_a), 2r_a)$  и  $(\frac{\xi_2}{L}(L - 2r_a), 2r_a)$  соответственно. Поэтому из (1) получаем, что абсцисса центра  $O$  описанной вокруг  $\triangle ABC$  окружности равна

$$\tilde{\eta} = \frac{1}{2}\left(\frac{\xi_1}{L}(L - 2r_a) + \frac{\xi_2}{L}(L - 2r_a)\right) = \frac{L - 2r_a}{2L}(\xi_1 + \xi_2) = \frac{L - r_a}{L}\varkappa = \frac{\eta_1 + \eta_2}{2}, \quad (3)$$

ординату  $O$  обозначим  $h$ , а радиус описанной окружности —  $R$ . Тогда абсциссы  $B$  и  $C$  являются корнями квадратного уравнения

$$(x - \tilde{\eta})^2 + (2r_a - h)^2 = R^2,$$

которое поэтому можно переписать в виде

$$x^2 - \frac{L - 2r_a}{L}(\xi_1 + \xi_2)x + \left(\frac{L - 2r_a}{L}\right)^2 \xi_1 \xi_2 = 0.$$

Приравнивая свободные члены этих уравнений и используя (1), получаем

$$\tilde{\eta}^2 + (2r_a - h)^2 - R^2 = \left(\frac{L - 2r_a}{L}\right)^2 \frac{L(\varkappa^2 - r_a^2)}{L - 2r_a} = (\varkappa^2 - r_a^2) \frac{L - 2r_a}{L}. \quad (4)$$

Теперь упростим левую часть (4), учитывая уравнение  $\tilde{\eta}^2 + (h - L)^2 = R^2$  (вершина  $A$  лежит на описанной окружности). Заменяя  $\tilde{\eta}^2 - R^2$  на  $-(h - L)^2$  и рас-

кладывая разность квадратов на множители, получаем выражение для  $h$

$$2h = L + 2r_a + \frac{\varkappa^2 - r_a^2}{L} \quad (5)$$

и для  $R^2$

$$R^2 = \tilde{\eta}^2 + \frac{1}{4} \left( 2r_a - L + \frac{\varkappa^2 - r_a^2}{L} \right)^2. \quad (6)$$

Осталось проверить, что окружности с центрами  $(\eta_i, r_a)$  и радиусом  $r_a$  касаются окружности с центром  $(\tilde{\eta}, h)$  и радиусом  $R$ , т. е.

$$(\tilde{\eta} - \eta_i)^2 + (h - r_a)^2 = (R + r_a)^2. \quad (7)$$

Это условие, с учетом (1), (2), (3) и (6) можно записать в виде

$$\left( \frac{\eta_1 - \eta_2}{2} \right)^2 + \frac{1}{4} \left( L + \frac{\varkappa^2 - r_a^2}{L} \right)^2 = r_a^2 + \tilde{\eta}^2 + \frac{1}{4} \left( 2r_a - L + \frac{\varkappa^2 - r_a^2}{L} \right)^2 + 2r_a R,$$

откуда после упрощений представляем его в виде выражения для  $R$ :

$$R = \frac{L(L - 2r_a) + \varkappa^2 + r_a^2}{2L}. \quad (8)$$

Непосредственным вычислением проверяется, что квадрат правой части (8) тождественно равен правой части (6).

(М. Н. Вялый)

1.6. УСЛОВИЕ. а) Дан многочлен  $P(X)$ . Для любого  $X > 0$ :  $P(X) > 0$ . Доказать, что  $P = Q/T$ , где  $Q$  и  $T$  — многочлены с неотрицательными коэффициентами.

б)\* Пусть  $P$  — квадратный трехчлен,  $\alpha$  — аргумент его комплексного корня. Тогда степень  $Q$  не меньше  $\pi/\alpha$ .

РЕШЕНИЕ. а) Прежде всего заметим, что любой многочлен с вещественными коэффициентами разлагается в произведение линейных и квадратных многочленов с вещественными коэффициентами, причем дискриминант всех этих квадратных трехчленов отрицателен. Из условия следует, что все линейные множители не имеют положительных корней и потому имеют вид  $x + c$ , где  $c > 0$ . Поэтому достаточно решить задачу для квадратного трехчлена. Далее, делая замену  $x \rightarrow \lambda \cdot x$  и умножая квадратный трехчлен на подходящее  $\lambda'$ , можно привести его к виду  $P(x) = x^2 - \delta x + 1$ . При этом  $\delta < 2$  и нетривиальным является случай, когда  $\delta > 0$ . Это условие равносильно положительности вещественных частей корней многочлена  $P$ . Теперь приступим к решению.

Домножив многочлен  $P(x) = x^2 - \delta x + 1$  на многочлен  $\bar{P}(x) = x^2 + \delta x + 1$  получим выражение  $P_1(y) = y^2 + (2 - \delta^2)y + 1$ , где  $y = x^2$ . При этом аргументы корней многочлена  $P_1$ , как легко видеть, будут в два раза больше аргументов соответствующих корней  $P$ .

Аналогичным образом, построим по многочлену  $P_1$  многочлен  $P_2$  и т. д. Изначально аргумент одного из корней заключен между 0 и  $\pi/2$ . Поскольку аргументы корней всякий раз будут удваиваться, то на некотором шаге возникнет многочлен  $P_k$ , аргументы корней которого будут заключены между  $\pi/2$  и  $\pi$ . Все

его коэффициенты будут неотрицательны. Поскольку многочлены  $\bar{P}_i$  имеют неотрицательные коэффициенты, многочлены  $Q = \prod_{i=1}^{k-1} \bar{P}_i(x^{2^{i-1}})$  и  $T = P_k(x^{2^{k-1}})$  удовлетворяют требованиям задачи.

Отметим, что можно было бы работать с последовательностью  $\delta_i = 2 - \delta_{i-1}^2$  и убедиться явно в наличии неположительного члена.

Утверждение пункта б) означает, что предыдущая конструкция близка к оптимальной. Как и выше, рассматривается многочлен вида  $P(x) = x^2 - \delta x + 1$ , причем  $\delta > 0$ . Тогда  $\delta = 2 \cos \varphi$  и величина  $\varphi$  ( $0 < \varphi < \pi/2$ ) есть аргумент одного из корней  $P$ . Заметим, что коэффициент  $\delta$  монотонен относительно положительного аргумента корня. Кроме того, если коэффициенты многочленов  $PT$  и  $T$  неотрицательны, а коэффициенты  $P_1$  не меньше соответствующих коэффициентов  $P$ , то все коэффициенты  $P_1T$  тоже неотрицательны. Поэтому достаточно доказать, что если  $\varphi = \frac{\pi}{2k}$ , то минимальная степень  $Q$  равна  $k$  и больше  $k$ , если  $\varphi < \frac{\pi}{2k}$ . В дальнейшем мы будем считать степень многочлена  $T$ , удовлетворяющего условию задачи, минимально возможной.

Поскольку многочлен  $P$  фиксирован, коэффициенты многочлена  $Q = PT$  будут линейно зависеть от коэффициентов  $t_k$  многочлена  $T$ . Будем двигать параметры  $t_k$  с постоянными скоростями, т. е. рассматривать семейства параметров  $t'_k = t_k + v_k \tau$  ( $\tau$  — «время»). Обращению в нуль коэффициентов многочленов  $Q$  или  $T$  отвечают линейные уравнения на параметры.

Пусть в процессе движения некоторый коэффициент обратится в нуль. Тогда рассмотрим новые начальные значения  $t_i$ , а на скорости  $v_i$  наложим ограничение, означающее, что этот коэффициент не меняется. Тогда он остается нулем и соответствующее ограничение будет иметь вид линейного уравнения с нулевой правой частью:  $\sum \alpha_i v_i = 0$ . В самом деле: линейность уже установлена, а утверждение про нулевую правую часть следует из того, что все коэффициенты сохраняются, когда все  $v_i = 0$ .

Будем до тех пор, пока это возможно, стремиться обратить в нуль как можно большее число коэффициентов многочленов  $T$  и  $R$ . Следующая лемма позволяет ограничиться рассмотрением только коэффициентов многочлена  $Q = RT$ .

**ЛЕММА 1.** Пусть  $T(x) = 1 + t_1x + \dots + t_kx^k + \dots + t_r x^r$ ;  $t_r > 0$ . Тогда многочлен  $T_1(x) = 1 + t_1x + \dots + t_{k-1}x^{k-1}$  удовлетворяет условию задачи (т. е. все коэффициенты  $Q_1 = PT_1$  неотрицательны) и имеет меньшую степень.

Количество свободных параметров равно степени  $\deg(T)$  многочлена  $T$ , а количество коэффициентов  $Q$ , которым отвечают уравнения, равно  $\deg(Q) - 1 = \deg(T) + 1$  (свободный член не меняется). Мы воспользуемся следующим утверждением из линейной алгебры:

**ЛЕММА 2.** Любая однородная система  $n$  линейных уравнений от  $k > n$  неизвестных имеет нетривиальное решение (в котором не все значения неизвестных нули).

Доказательство этой леммы — индукция по числу уравнений и неизвестных.

В силу данной леммы можно добиться того, чтобы  $\deg(T)$  коэффициентов многочлена  $Q$  обратились в нуль. Нулем не окажется свободный член и старший

коэффициент. Поэтому ненулевым кроме этих двух может оказаться еще только один коэффициент.

Покажем, как добиться того, чтобы этот коэффициент был предпоследним (т. е. при  $x^{\deg(Q)-1}$ ).

Обозначим  $k$ -й коэффициент (при  $x^k$ ) многочлена  $Q$  через  $q_k$ . Тогда  $q_k = t_k + t_{k-2} - 2 \cos \varphi \cdot t_{k-1}$ . Пусть  $t_k > 0$  для  $k \leq \deg(Q) - 2$  и  $t_k = 0$  для остальных  $k \neq 1, \deg(Q)$ . Покажем, как построить многочлен  $T'$ , удовлетворяющий условию задачи, у которого первые  $k$  коэффициентов — нули.

Положим  $t'_i = t_i$  при  $i \leq k$  и  $t'_i = \lambda t_i$  если  $i > k$ . При этом  $0 \leq \lambda \leq 1$ . Ясно, что при таком варьировании у коэффициентов  $q_s$  для  $s \leq k - 1$  или  $s \geq k + 2$  сохраняются положительность, а также свойство «быть равным нулю».

Поскольку  $q'_{k+1}/\lambda = q_{k+1} - (1/\lambda - 1)t_{k+1}$ , то неотрицательность  $t_{k+1}$  также сохраняется. Что касается члена  $q_k$ , то он уменьшается и в какой-то момент становится равным нулю. Таким образом, осуществлен переход к ситуации, когда первые  $k$  коэффициентов у многочлена  $Q$  (не считая свободного члена) нулевые.

В этом случае коэффициенты  $t_k$  будут удовлетворять следующим рекуррентным соотношениям:  $t_0 = 1, t_1 = \cos \varphi, t_s = 2 \cos \varphi \cdot t_{s-1} - t_{s-2}; s = 2, \dots, \deg(T)$ . Поэтому многочлен  $T$  будет образован первыми  $k$  членами ряда Тейлора функции

$$\frac{1}{1 - 2 \cos \varphi x + x^2} = \frac{1}{(z - \bar{z})} \cdot \left( \frac{1}{(x - z)} - \frac{1}{(x - \bar{z})} \right),$$

где  $z = \cos \varphi + i \sin \varphi$ .

$$\text{Откуда } t_k = \frac{1}{2i \sin \varphi} (z^k - \bar{z}^k) = \sin(k\varphi) / \sin(\varphi).$$

Если величина  $t_k$ , определенная предыдущим соотношением, перестает быть положительной, то ряд  $\sum t_n x^n$ , начиная с  $(k+1)$ -го члена, можно оборвать. Легко видеть, что верно и обратное.

Для завершения решения задачи остается отметить, что условие  $\sin(k\varphi) = 0$  равносильно условию  $k = \pi / \arg(\varphi)$ . (А. Я. Белов)

**1.9. УСЛОВИЕ.** Известно, что при любых действительных  $A, B$  ряд  $\sum \frac{1}{|Ax_n + By_n|}$  расходится. Обязан ли расходиться ряд  $\sum \frac{1}{|x_n| + |y_n|}$ ? А что если  $A$  и  $B$  могут быть комплексными?

**РЕШЕНИЕ.** В условии задачи не оговорено, как трактовать члены рядов с  $Ax_n + By_n = 0$ . Мы будем полагать, что речь идет о рядах

$$L(A, B) = \sum_{n: Ax_n + By_n \neq 0} \frac{1}{|Ax_n + By_n|}$$

и

$$S = \sum_{|x_n| + |y_n| \neq 0} \frac{1}{|x_n| + |y_n|} = \sum_{f(n) \neq 0} \frac{1}{f(n)}, \quad f(n) \stackrel{\text{def}}{=} |x_n| + |y_n|.$$

Для упрощения записи считаем далее, что во всех формулах из сумм исключаются члены, в которых один из знаменателей обращается в 0.

**ОТВЕТ:** «нет» в действительном случае и «да» в комплексном.



КОМПЛЕКСНЫЙ СЛУЧАЙ. Пусть  $S$  сходится. Докажем, что тогда сходится и один из рядов  $L(q, -1)$ . Поскольку

$$\sum_{n: x_n=0} \frac{1}{|y_n|} \leq \sum \frac{1}{f(n)} = S,$$

сходимость ряда  $L(q, -1)$  не меняется при отбрасывании членов с  $x_n = 0$ . Обозначим ряд с отброшенными членами, для которых  $x_n = 0$ , через  $\tilde{L}(q)$ , введём также обозначение  $\lambda_n = -y_n/x_n$ . Тогда

$$\tilde{L}(q) = \sum \frac{1 + |\lambda_n|}{f(n)|q - \lambda_n|}. \quad (1)$$

При построении  $q$  будем использовать меру

$$\mu(X) = \frac{\sum_{\lambda_n \in X, x_n \neq 0} f(n)^{-1}}{\sum_{x_n \neq 0} f(n)^{-1}}.$$

Обозначим  $l_0 = 1$ ,  $m = 30$ . Разобьём квадрат  $Q_0 = \{z : |\operatorname{Re} z| \leq l_0/2, |\operatorname{Im} z| \leq l_0/2\}$  прямыми, параллельными действительной и мнимой осям, на  $m^2$  квадратов со стороной  $l_1 = l_0/m$  (принадлежность граничных точек произвольная). Для каждого из  $(m-2)^2$  внутренних квадратов  $Q$  определим «окрестность»  $Q^+$  как объединение этого квадрата и 8 смежных с ним. Поскольку каждый квадрат разбиения входит не более чем в 25 «окрестностей», то найдётся квадрат разбиения  $Q_1$  такой, что  $\mu(Q_1^+) \leq 25\mu(Q_0)/(m-2)^2$ . Применяв ту же процедуру к  $Q_1$ , получим квадрат  $Q_2$  и т. д. Последовательность квадратов  $Q_i$  определяет некоторую точку  $q$ .

Докажем, что ряд (1) для  $q$  сходится. Обозначим  $N_0 = \{n : \lambda_n \notin Q_1^+\}$ ,  $N_1 = \{n : \lambda_n \notin Q_2^+\} \setminus N_0$  и т. д. Из процедуры построения следуют оценки

$$\sum_{n \in N_k} \frac{1}{f(n)} \leq \frac{25^k}{(m-2)^{2k}} S, \quad (n \in N_k) \implies |q - \lambda_n| \geq \frac{1}{m^k}. \quad (2)$$

Выберем теперь такое  $R$ , чтобы при  $|z| > R$  выполнялось  $(1 + |z|)/|q - z| < 2$ . Оценивая по отдельности члены с  $\lambda_n > R$  и  $\lambda_n \leq R$ , получаем

$$\sum_{n \in N_k} \frac{1 + |\lambda_n|}{f(n)|q - \lambda_n|} \leq (2 + (R+1)m^k) \sum_{n \in N_k} \frac{1}{f(n)} \leq (R+2) \left( \frac{25m}{(m-2)^2} \right)^k S,$$

откуда при  $25m/(m-2)^2 < 1$  следует сходимость ряда  $\tilde{L}(q)$ .

ДЕЙСТВИТЕЛЬНЫЙ СЛУЧАЙ. Используя уже введённые обозначения, полагаем  $f(n) = n \ln^2 n$  (этого достаточно для сходимости  $S$ ). Последовательность  $\lambda_n$  построим следующим образом. Разделим её на две подпоследовательности:  $\lambda_n = \lambda_s^{(t)}$ , где  $n = t + 2s$ ,  $t \in \{0, 1\}$ . Зададим  $\lambda_n^{(1)} = n$ . Последовательность  $\lambda_n^{(0)}$  определим следующим образом. Занумеруем натуральными числами пары целых чисел  $(u, v)$ ,  $v > 0$ , в порядке возрастания  $\max\{|u|, v\}$  (в остальном нумерация произвольная). Обозначим  $n$ -ю пару  $(u_n, v_n)$ , тогда  $\lambda_n^{(0)} = u_n/v_n$ . Заметим, что пара  $(u, v)$  встретится не позже, чем на  $(\max\{|u|, v\})^2$ -м месте.

Проверим расходимость рядов  $L(A, B)$  при таком выборе  $\lambda_n$ .

Пусть  $B = 0$ . Возьмём подпоследовательность  $n_k = 1 + 2k$ . Для неё

$$\sum_k \frac{1}{|Ax_{n_k}|} = \frac{1}{|A|} \sum_k \frac{1 + |\lambda_{n_k}|}{f(n_k)} = \frac{1}{|A|} \sum_k \frac{1 + k}{(1 + 2k) \ln^2(1 + 2k)},$$

так что ряд  $L(A, 0)$  расходится.

При  $B \neq 0$  обозначим  $a = A/B$ . Оценим вклад  $L_v$ , который даёт в ряд подпоследовательность  $n_k = 2s_k$ , где  $s_k$  нумеруют те пары  $(u, v)$ , для которых

$$|a - \frac{u}{v}| < 1.$$

Для таких пар  $\max\{|u|, v\} \leq v \max\{1, |a| + 1\} = Cv$ , поэтому имеем

$$\begin{aligned} L_v &= \sum_k \frac{1}{|Ax_{n_k} + By_{n_k}|} = \frac{1}{|B|} \sum_k \frac{1 + |u_k/v_k|}{f(n_k)|a - u_k/v_k|} \geq \\ &\geq \frac{1}{4C^2|B|} \sum_{k=1}^{v/2} \frac{v}{kv^2 \ln^2(4C^2v^2)} \geq \frac{\ln v/2}{4C^2|B|v \ln^2(4C^2v^2)} = \Omega\left(\frac{1}{v \ln v}\right). \end{aligned}$$

Так как  $L(A, B) \geq \sum_v L_v$ , ряд  $L(A, B)$  расходится.

(М. Вялый)

1.10. УСЛОВИЕ. Функция, заданная на всей вещественной прямой, бесконечно дифференцируема. В каждой точке некоторая производная (номер производной может зависеть от точки) равна нулю. Докажите, что эта функция — многочлен.

РЕШЕНИЕ 1<sup>1</sup>). Пусть  $F_n = \{x \mid f^{(n)}(x) = 0\}$ ,  $G_n = \mathbb{R} \setminus F_n$ . Докажем вспомогательное утверждение (А): Если  $f$  не является многочленом на некотором открытом интервале  $\Delta$ , то для любого  $n$  найдется составляющий интервал непустого открытого множества  $G_n \cap \Delta$ , на котором  $f$  не является многочленом.

Из (А) легко выводится утверждение задачи: предположим, что  $f$  не многочлен на  $\mathbb{R}$ , построим последовательность вложенных интервалов  $\Delta_n = (a_n; b_n) \subset \subset G_n$ ,  $[a_{n+1}; b_{n+1}] \subset \Delta_n$ , пересечение которых непусто, что противоречит условию.

Предположим, что (А) не выполняется, т. е. что для некоторых  $n$  и  $\Delta$  функция  $f$  — не многочлен на  $\Delta$ , но ее ограничение на любой составляющий интервал множества  $G_n \cap \Delta$  является многочленом. Обозначим символом  $\sigma(x)$  ( $x \in \Delta$ ) максимальный промежуток, обладающий свойствами:  $x \in \sigma(x) \subset \Delta$ ,  $f|_{\sigma(x)}$  — многочлен. Проверьте, что если ограничения  $f$  на два примыкающих промежутка — многочлены, то  $f$  — многочлен на их объединении. Покажите, что если  $a$  — граничная точка  $\sigma(x)$ , то: 1) существует последовательность  $x_k \in F_n$ ,  $x_k \neq a$ ,  $x_k \rightarrow a$ ; 2)  $f^{(m)}(a) = 0$  для всех  $m \geq n$ ; 3)  $f^{(n)}$  обращается в нуль на  $\sigma(x)$ . Тогда, в частности,  $f^{(n)}(x) = 0$  для всех  $x \in G_n \cap \Delta$ , что невозможно.

(Д. Ю. Бураго)

РЕШЕНИЕ 2. Пусть  $F_n$  — множество нулей функции  $f^{(n)}$ . Объединение замкнутых множеств  $F_n$  покрывает прямую, а доказать надо, что одно из этих множеств совпадает со всей прямой. Предположим противное:  $F_n \neq \mathbb{R}$  при всех  $n$ .

<sup>1</sup>Из книги Б. М. Макаров, М. Г. Голузина, А. А. Лодкин, А. Н. Подкорытов «Избранные задачи по вещественному анализу». М.: Наука, 1992. С. 294–295.

Пусть  $E_n = F'_n$  — множество предельных точек множества  $F_n$ . Достаточно доказать следующие три леммы:

ЛЕММА 1. *Последовательность  $(E_n)$  возрастает.*

ЛЕММА 2. *Если непустой открытый интервал  $(a, b)$  не пересекается с некоторым  $E_n$ , а один из его концов  $a, b$  принадлежит  $E_n$ , то  $(a, b)$  не содержится в  $E_m$  ни при каком  $m > n$ .*

ЛЕММА 3. *Пусть  $(G_n)$  — убывающая последовательность непустых открытых множеств на прямой. Предположим, что для всех  $n < m$  каждый максимальный интервал  $(a, b)$  в  $G_n$  пересекается с  $G_m$ , причем  $a = \inf(a, b) \cap G_m$ , если  $a$  конечно, и  $b = \sup(a, b) \cap G_m$ , если  $b$  конечно. Тогда множество  $\bigcap G_n$  несчетно.*

Предположим, что эти леммы установлены. Положим  $G_n = \mathbb{R} \setminus E_n$ . Согласно леммам 1 и 2, последовательность  $(G_n)$  удовлетворяет условиям леммы 3. По лемме 3 множество  $A := \bigcap G_n = \mathbb{R} \setminus \bigcup E_n$  несчетно. С другой стороны, так как  $\bigcup F_n = \mathbb{R}$ , то  $A \subset \bigcup (F_n \setminus E_n)$ . Множество  $F_n \setminus E_n$  — это множество изолированных точек множества  $F_n$ . У любого множества на прямой множество изолированных точек счетно. (Указание: рассмотрите все интервалы с рациональными концами, каждый из которых содержит ровно одну точку данного множества.) Из включения  $A \subset \bigcup (F_n \setminus E_n)$  вытекает, таким образом, что  $A$  должно быть счетным, и мы приходим к противоречию.

Остается доказать три леммы, сформулированные выше.

Доказательство леммы 1. Достаточно установить следующее: всякий неизолированный нуль дифференцируемой функции является неизолированным нулем ее производной. Это вытекает из теоремы Ролля: между всякими двумя нулями функции лежит нуль ее производной.

Доказательство леммы 2. Допустим, что  $(a, b)$  не пересекается с  $E_n$ ,  $a \in E_n$  и  $(a, b) \subset E_m$  при некотором  $m > n$ . На интервале  $(a, b)$  функция  $f$  совпадает с многочленом степени  $k$ , где  $n \leq k < m$ . Отсюда  $f^{(k)}(a) \neq 0$ . Однако по лемме 1  $a \in E_n \subset E_k$ , так что  $f^{(k)}(a) = 0$ . Противоречие.

Доказательство леммы 3. Если множество  $\bigcap G_n$  содержит интервал, то оно несчетно. Предположим, что  $\bigcap G_n$  не содержит интервала. Если  $(a, b)$  — максимальный интервал в  $G_n$ , то число компонент (максимальных интервалов) множества  $(a, b) \cap G_m$  неограниченно растет с ростом  $m$ , так что при больших  $m$  можно выбрать два максимальных интервала в  $(a, b) \cap G_m$  с непересекающимися замыканиями. Используя это замечание, легко построить по индукции ветвящуюся систему открытых интервалов  $I_s^k$ ,  $1 \leq s \leq 2^k$ , и числа  $n_1 < n_2 < \dots$ , такие, что: (1) при фиксированном  $k$  интервалы  $I_s^k$  имеют попарно непересекающиеся замыкания; (2) каждый интервал  $I_s^k$  является компонентой множества  $G_{n_k}$ ; (3)  $I_{2s-1}^{k+1} \cup I_{2s}^{k+1} \subset I_s^k$ . Пусть  $B_k$  — объединение интервалов  $I_s^k$ ,  $1 \leq s \leq 2^k$ , а  $C_k$  — объединение их замыканий. Множество  $\bigcap C_k$  находится в естественном взаимно однозначном соответствии с множеством всех бесконечных последовательностей нулей и единиц и гомеоморфно канторову совершенному множеству. Таким образом,  $\bigcap C_k$  несчетно. Множество  $\bigcap B_k$  также несчетно, поскольку разность  $\bigcap C_k \setminus \bigcap B_k$  содержится в счетном множестве  $\bigcup (C_k \setminus B_k)$  кон-

цов всех интервалов  $I_s^k$ . Так как  $\bigcap B_k \subset \bigcap G_{n_k} = \bigcap G_n$ , то множество  $\bigcap G_n$  несчетно. (В. В. Успенский)

2.2. УСЛОВИЕ. Пусть  $P(x)$  и  $Q(x)$  — многочлены, причем  $Q(0) = 0$ . Докажите, что если  $P(Q(x))$  — чётная функция, то  $Q(x)$  — чётная или нечётная функция.

РЕШЕНИЕ. Утверждение задачи непосредственно вытекает из следующих фактов про многочлены:

ЛЕММА 1. Пусть  $Q$  — многочлен. Тогда  $||Q(x)| - |Q(-x)||$  есть либо константа, либо стремится к бесконечности при  $x \rightarrow \infty$ .

ЛЕММА 2. Пусть  $P \neq \text{const}$  — многочлен и  $||x_n| - |y_n|| \rightarrow \infty$ . Тогда  $||P(x_n)| - |P(y_n)|| \rightarrow \infty$ .

Доказательство этих утверждений несложно и опускается. (А. Я. Белов)

3.3. УСЛОВИЕ. (а) Существует ли такая непрерывная функция  $f(x)$ , что  $f(f(f(x))) = e^{-x}$ ?

(б) Тот же вопрос для разрывной функции.

(в) Тот же вопрос для функции с конечным числом точек разрыва.

РЕШЕНИЕ.

Ответ: а) не существует, б) существует, в) не существует.

(а) Не существует. Пусть непрерывная функция  $f(x)$  удовлетворяет уравнению  $f(f(f(x))) = e^{-x}$ . Так как  $e^{-x}$  взаимно однозначна, то и  $f(x)$  взаимно однозначна. Будучи непрерывной, она тогда строго монотонна. Поскольку  $e^{-x}$  убывает, то и  $f(x)$  убывает. Если ее множество значений — вся вещественная прямая, то это верно и для  $f(f(f(x)))$ , тогда как  $e^{-x}$  всюду положительна. Значит,  $f(x)$  ограничена сверху или снизу. Если  $f(x)$  ограничена снизу числом  $c$ , то  $f(f(x))$  ограничена снизу тем же числом, а сверху — значением  $f(c)$ . Противоречие. Аналогично, если  $f(x)$  ограничена сверху, то  $f(f(x))$  также ограничена с обеих сторон. Как следствие, это верно и для  $f(f(f(x)))$ . Однако  $e^{-x}$  не ограничена сверху. Значит, искомой функции не существует.

(б) Таких функций бесконечно много. Действительно, луч  $(-\infty; 0]$  под действием  $e^{-x}$  отображается на луч  $(+\infty; 1]$ . Тот, в свою очередь, отображается на полуинтервал  $(0; 1/e]$ , и т. д. При бесконечном продолжении этого процесса заполняется вещественная прямая, за исключением некоторого отрезка  $[A; B]$ , причем  $0 < A \leq B < 1$ ,  $e^{-A} = B$ ,  $e^{-B} = A$ . Заметим теперь, что производная функции  $g(x) = e^{-e^{-x}} - x$  равна  $e^{-e^{-x}} - 1$  и при положительных  $x$  заведомо отрицательна (нетрудно показать, что она отрицательна при всех вещественных  $x$ ). Так как  $g(A) = g(B) = 0$ , то  $A = B$ . Как следствие,  $e^{-A} = A$ .

Разобьем теперь все отрицательные числа на тройки. Если  $x_1, x_2, x_3$  — одна из таких троек, то положим  $f(x_1) = x_2, f(x_2) = x_3$ , а при  $n > 3$   $x_n = e^{-x_{n-3}} = f(x_{n-1})$ . Числа в этой последовательности не повторяются, поскольку функция  $e^{-x}$  взаимно однозначна и всюду положительна. По той же причине две последовательности для разных троек не пересекаются. Положим также  $f(A) = A$ . Из доказанного выше следует, что  $f(x)$  определена на всей вещественной прямой, причем  $f(f(f(x))) = e^{-x}$ .

(в) Такой функции не существует. Пусть  $f(f(f(x))) = e^{-x}$  для всех вещественных  $x$ . Пусть также  $A$  — (единственный) корень уравнения  $e^{-x} = x$ . Если  $C$  — точка разрыва для  $f(x)$ , то  $C$  является точкой разрыва и для  $e^{-f(x)}$ . Но  $e^{-f(x)} = f(f(f(f(x) = f(e^{-x})))$ , поэтому  $e^{-x}$  — точка разрыва для  $f(x)$ . Аналогично, точкой разрыва является  $e^{-e^{-x}}$ , и т. д. Как показано в решении пункта (б), при  $C \neq A$  элементы полученной последовательности не повторяются. Значит, если  $f(x)$  имеет хотя бы один разрыв при  $x \neq A$ , то таких разрывов бесконечно много.

Пусть теперь  $f(x)$  имеет единственный разрыв при  $x = A$ . Поскольку  $f(A) = f(e^{-A}) = f(f(f(f(x)))) = e^{-f(A)}$ , то  $f(A) = A$ . Слева от  $A$  функция непрерывна; если в этой области она принимает значения и больше, и меньше  $A$ , то там встречается и значение, равное  $A$ . Но это противоречит взаимной однозначности  $f(x)$ , которая следует из взаимной однозначности  $e^{-x}$ . Если  $f(x) < A = f(A)$  при  $x < A$ , то это выполнялось бы для  $f(f(x))$  и  $e^{-x}$ , но последнее неверно. Значит,  $f(x) > A$  при  $x < A$ . Аналогично  $f(x) < A$  при  $x > A$ .

Любое число, большее чем  $A$ , содержится в области значений  $e^{-x} = f(f(f(x)))$ . Поэтому любое число, меньшее чем  $A$ , содержится в области значений  $f(f(x))$  (с учетом взаимной однозначности  $f(x)$ ). Все эти числа являются значениями  $f(x)$ , и с учетом равенства  $f(A) = A$  эта функция отображает вещественную прямую на себя. Но тогда тем же свойством обладает и  $f(f(f(x)))$ , в то время как  $e^{-x}$  всюду положительна. Значит, искомой функции не существует.

(А. Канель)

4.8. УСЛОВИЕ. Решите функциональное уравнение для непрерывных вещественных функций вещественного переменного:

$$F(x+y) = A(x) + B(x)C(y).$$

РЕШЕНИЕ.

Постараемся свести исходное уравнение к частному случаю. Положим

$$\begin{aligned} F_1(x) &= F(x) - F(0); & C_1(y) &= C(y) - C(0); \\ A_1(x) &= A(x) - F(0) + B(x)C(0). \end{aligned} \quad (1)$$

Тогда  $F_1(0) = C_1(0) = 0$ ,

$$F_1(x+y) = A_1(x) + B(x)C_1(y). \quad (2)$$

Положим в (2)  $y = 0$ :

$$F_1(x) = A_1(x). \quad (3)$$

Если  $F_1(x) \equiv 0$ , то (ввиду (2) – (3)) или  $B(x)$ , или  $C_1(y)$  является тождественным нулем. Используя (1) ( $F(0)$  и  $C(0)$  можно взять произвольными), получаем классы функций, указанные в конце решения под номерами 1 и 2.

Пусть теперь  $F_1(x) \not\equiv 0$ . Положим в (2)  $x = 0$ . С учетом (3)  $F_1(y) = B(0)C_1(y)$ . Как следствие,  $B(0) \neq 0$ . Положим  $B_1(x) = B(x)/B(0)$ . Тогда

$$F_1(x+y) = F_1(x) + B_1(x)F_1(y). \quad (4)$$

Используем ассоциативность сложения:

$$\begin{aligned} F_1(x+y+z) &= F_1(x+y) + B_1(x+y)F_1(z) = \\ &= F_1(x) + B_1(x)F_1(y) + B_1(x+y)F_1(z); \\ F_1(x+y+z) &= F_1(x) + B_1(x)F_1(y+z) = \\ &= F_1(x) + B_1(x)F_1(y) + B_1(x)B_1(y)F_1(z). \end{aligned}$$

Отсюда  $(B_1(x+y) - B_1(x)B_1(y))F_1(z) = 0$ . Поскольку  $F_1(z) \neq 0$ , то  $B_1(x+y) = B_1(x)B_1(y)$ . Таким свойством среди непрерывных функций обладают только экспоненты и тождественный нуль. Но  $B(0) \neq 0$ , и (4) принимает вид  $F_1(x+y) = F_1(x) + e^{ax}F_1(y)$  для некоторого  $a$ . Если  $a = 0$ , то  $F_1(x+y) = F_1(x) + F_1(y)$ . Среди непрерывных функций таким свойством обладают только однородные линейные ( $y = kx$ ), что приводит к классу функций 3. Если же  $a \neq 0$ , то ввиду коммутативности сложения  $F_1(x) + e^{ax}F_1(y) = F_1(x+y) = F_1(y) + e^{ay}F_1(x)$ . Отсюда при  $x, y \neq 0$   $\frac{F_1(x)}{e^{ax} - 1} = \frac{F_1(y)}{e^{ay} - 1} = c$  для некоторого  $c$ , и  $F_1(x) = c(e^{ax} - 1)$ . Это выполнено и при  $x = 0$  (поскольку  $F_1(0) = 0$ ). Отсюда возникает класс функций 4.

В итоге получаем следующие решения:

1)  $F(x) \equiv A(x) \equiv K$  ( $K$  — произвольная константа);  $B(x) \equiv 0$ ;  $C(y)$  — произвольная непрерывная функция.

2)  $F(x) \equiv K_1$ ;  $B(x)$  — произвольная непрерывная функция;  $C(y) \equiv K_2$ ;  $A(x) = K_1 - B(x)K_2$  ( $K_1, K_2$  — произвольные константы).

3)  $F(x) = K_1x + K_2$ ;  $A(x) = K_1x + K_2 - K_3K_4$ ;  $B(x) = K_3$ ;  $C(y) = \frac{K_1}{K_3}y + K_4$  ( $K_1, K_2, K_3, K_4$  — произвольные константы,  $K_3 \neq 0$ ).

4)  $F(x) = K_1(e^{ax} - 1) + K_2$ ;  $A(x) = e^{ax}(K_1 - K_3K_4) + K_2 - K_1$ ;  $B(x) = K_3e^{ax}$ ;  $C(y) = \frac{K_1(e^{ay} - 1)}{K_3} + K_4$  ( $a, K_1, K_2, K_3, K_4$  — произвольные константы,  $a \neq 0$ ,  $K_3 \neq 0$ ).

(А. Канель, Б. Френкин)